POLITECNICO DI MILANO

Chemical Reaction Engineering and Chemical Kinetics

Curve Matching: a generalized framework for model comparison with large sets of experiments

A. Stagni[(1)], M.S. Bernardi[(2)], M. Pelucchi[(1)], L.M. Sangalli[(2)], A. Cuoci[(1)], A. Frassoldati[(1)], P. Secchi[(2)], T. Faravelli[(1)]

[(1)] *CRECK - Department of Chemistry, Materials and Chemical Engineering*
[(2)] *MOX - Department of Mathematics*

Outline

1. Introduction and motivations

→ Experiments vs models
→ Kinetic model update
→ The NO_x case
→ Aim of the work

2. The Curve Matching framework

→ Functional data analysis
→ Dissimilarity measures
→ Integrated index
→ Output analysis

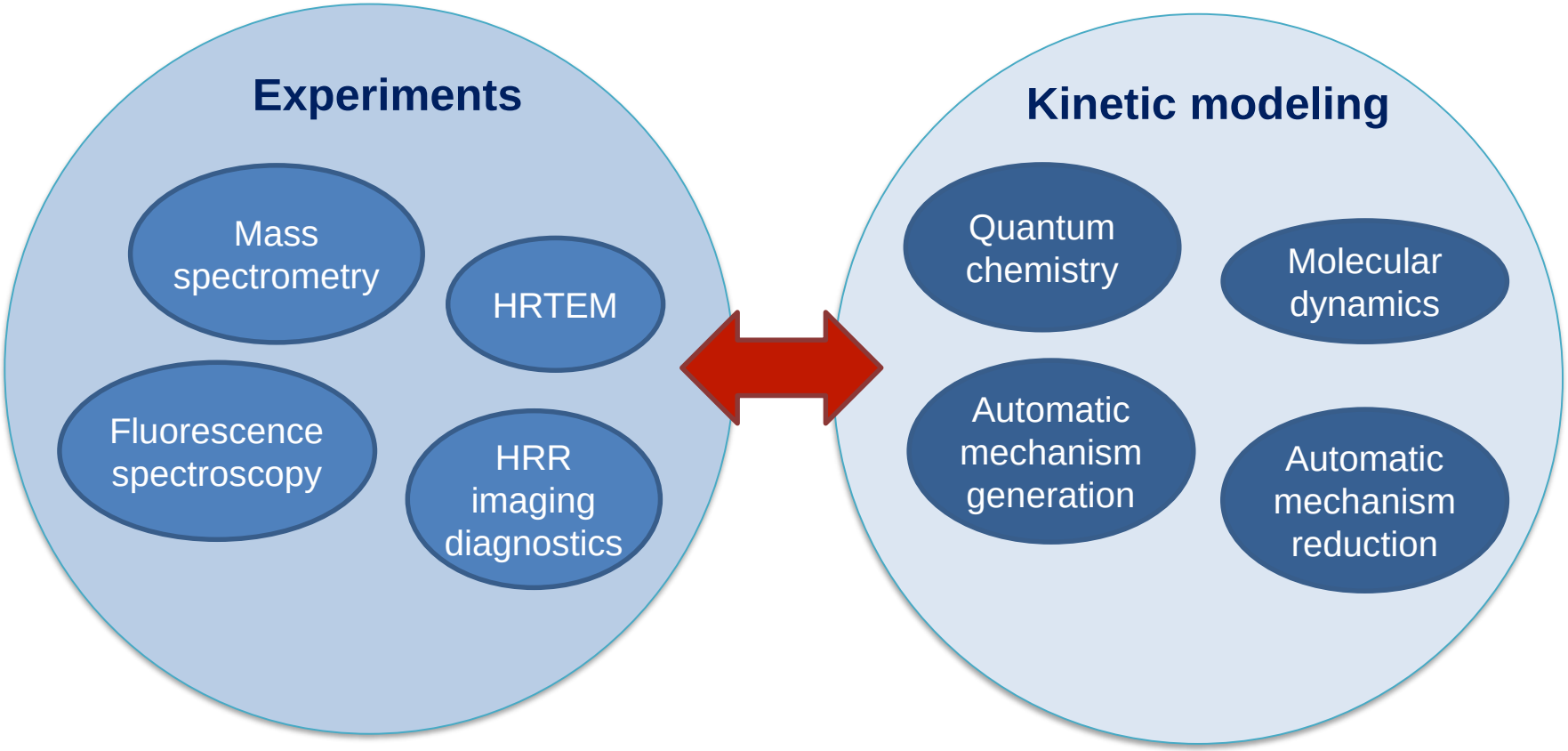
3. Integration with experimental database

→ Automatic Procedure
→ Software interface

4. Conclusions and next steps

Introduction

Combustion kinetics

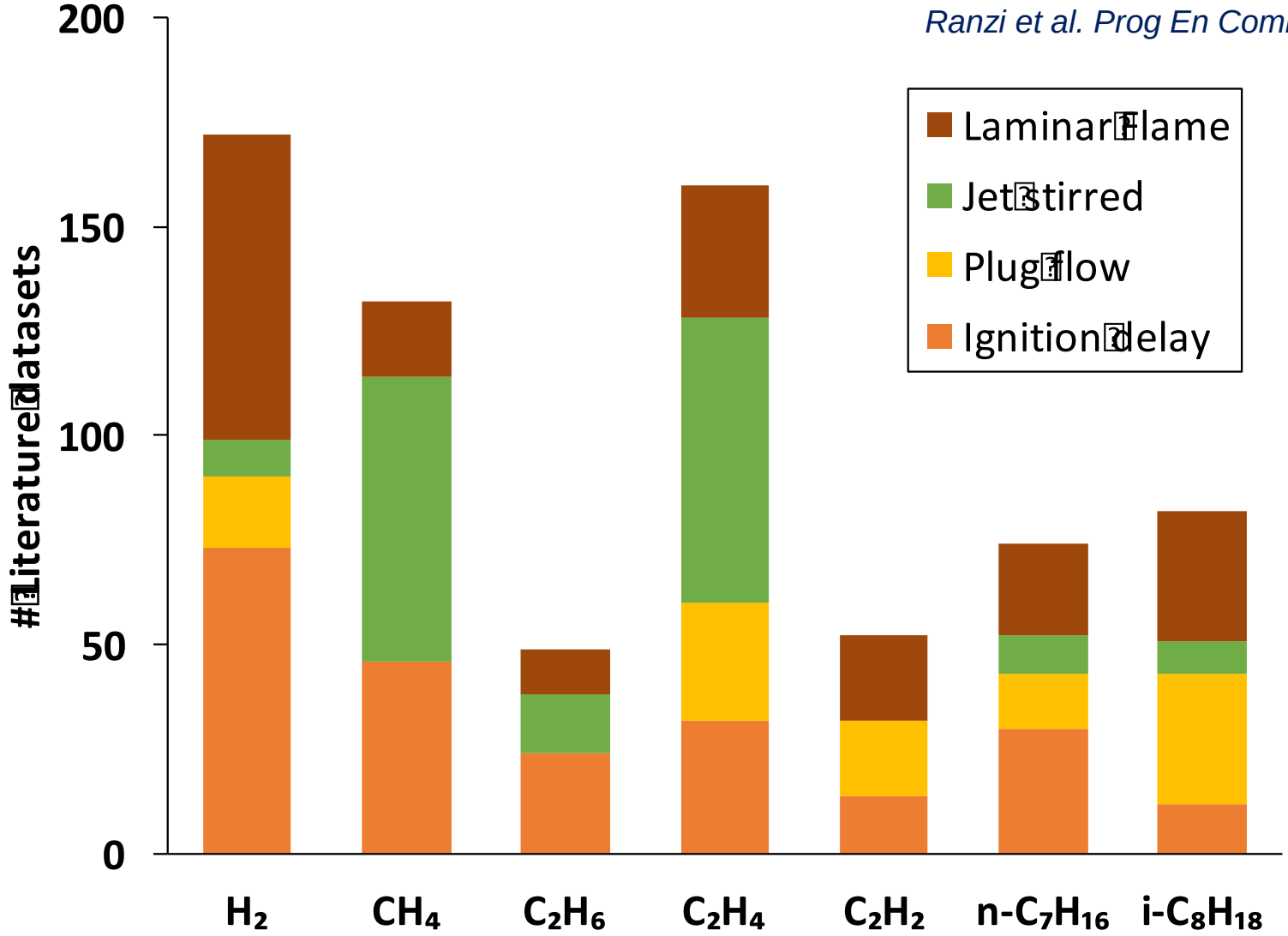


Higher accuracy
Wider capabilities
Faster sampling frequency

Higher computing power
Improved numerical techniques
Standardized kinetics formulation

Experimental datasets

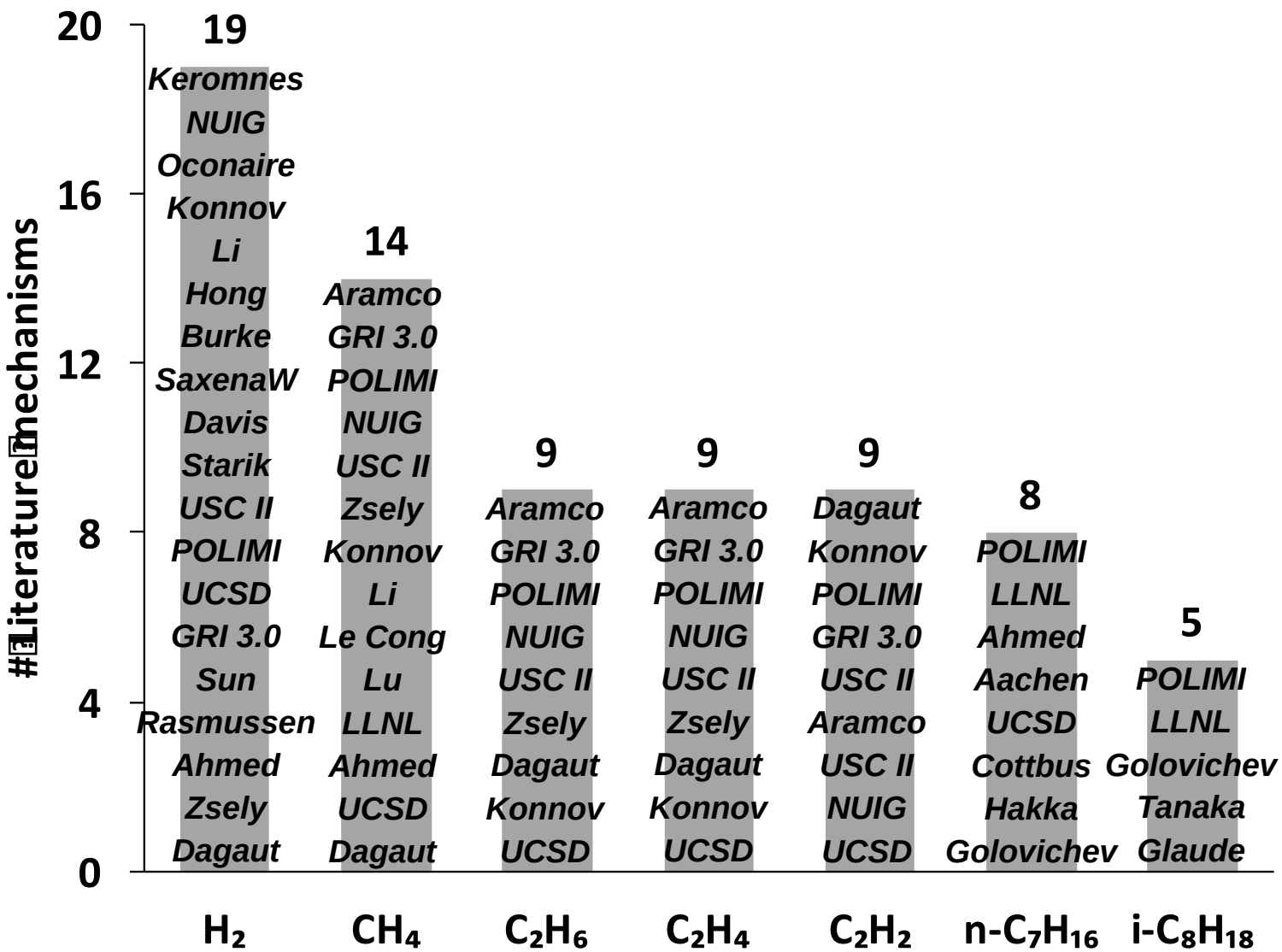
Curran et al. Comb Flame, 114 (1998)
Curran et al. Comb Flame, 129 (2002)
Ranzi et al. Prog En Comb Sci, 38 (2012)



Olm et al. Comb Flame, 161 (2014)

Metcalfe et al. Int J Chem Kin, 45 (2013)

Chemical mechanisms



Adapted from Olm et al. Comb Flame, 161 (2014)

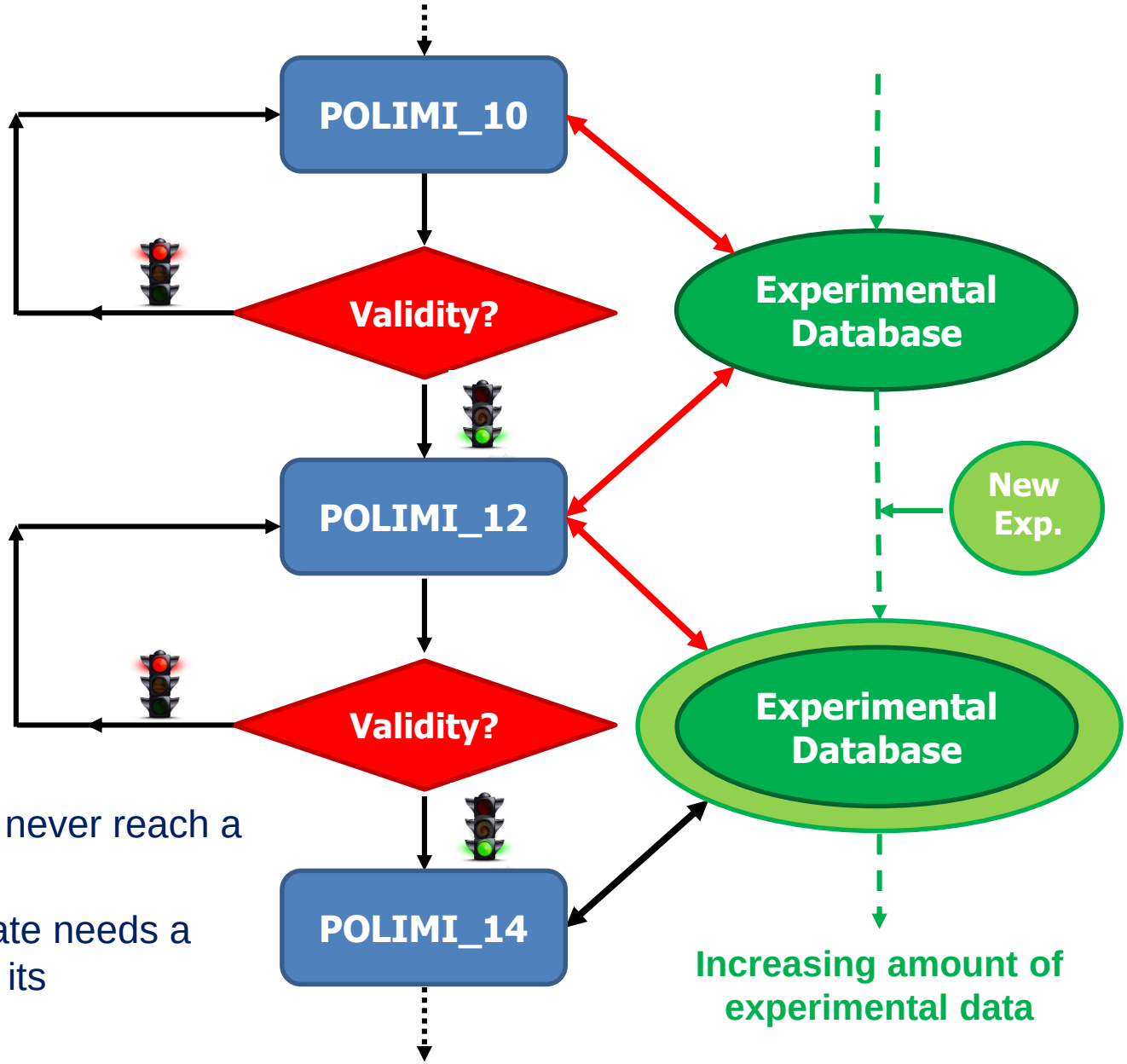
Updating a kinetic model

New reaction classes

More accurate theoretical rate constants

More accurate thermodynamics

New Fuels



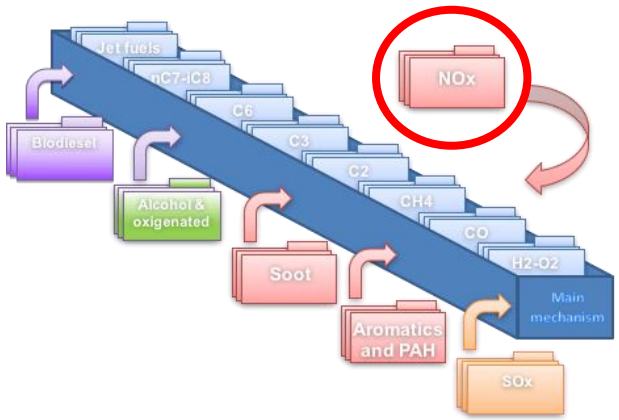
- Chemical mechanisms never reach a **final** version
- Every mechanism update needs a complete **validation** of its performance

The NO_x case

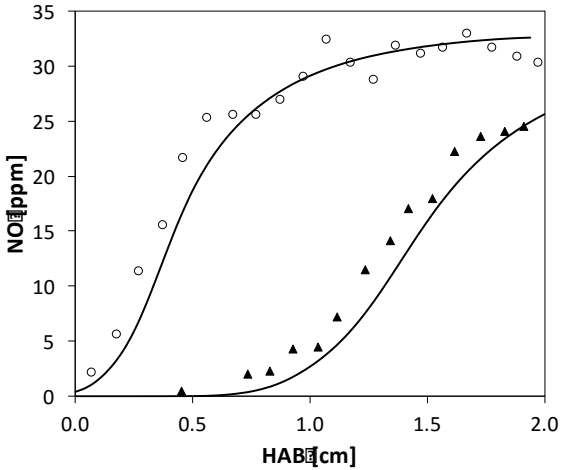
- Hierarchical dependencies strongly affect a kinetic mechanism
- E.g. **NO_x formation** depends on the core C_0-C_3 submechanism

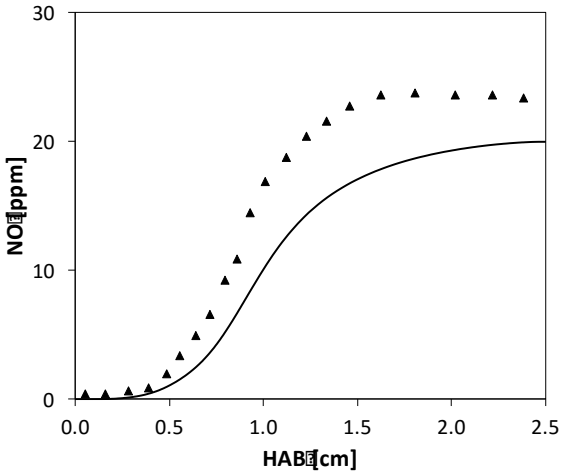
What can happen?

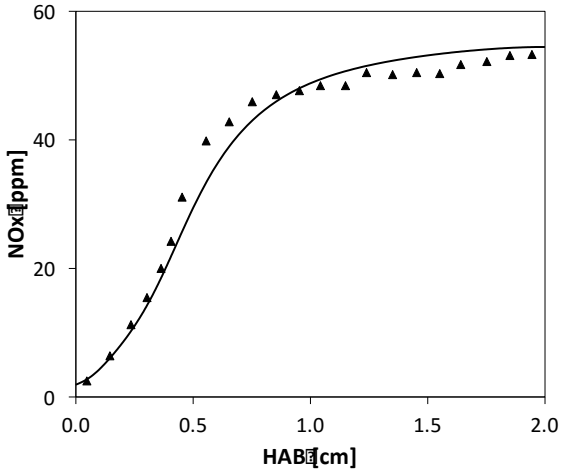
- January 2012: **critical update** to C_0-C_3

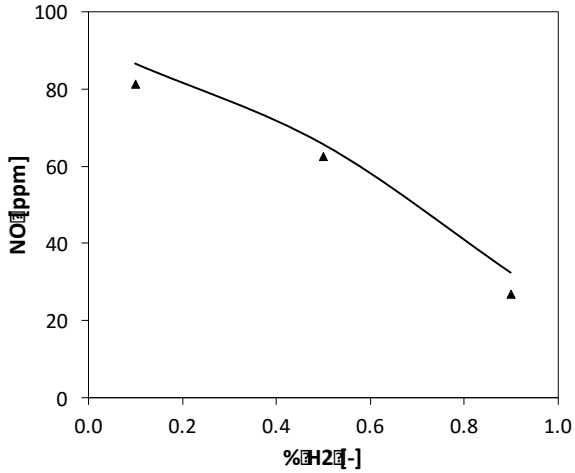


Before 2012 update









van Essen et al. Comb Flame, 153 (2008)
Sutton et al. Comb Flame 159 (2012)

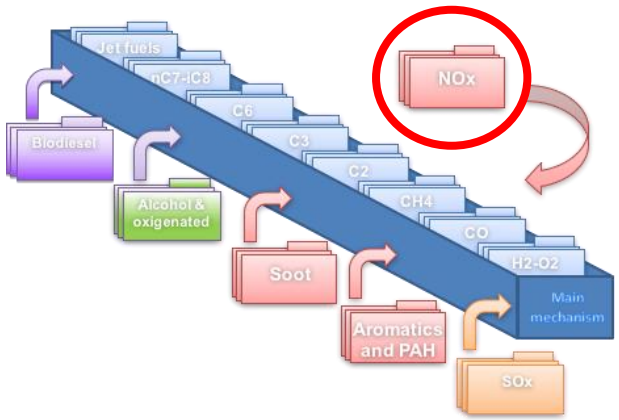
Lamoureux et al. Comb Flame, 157 (2010)
Rørtveit, G.J., et al., 4 (2003)

The NO_x case

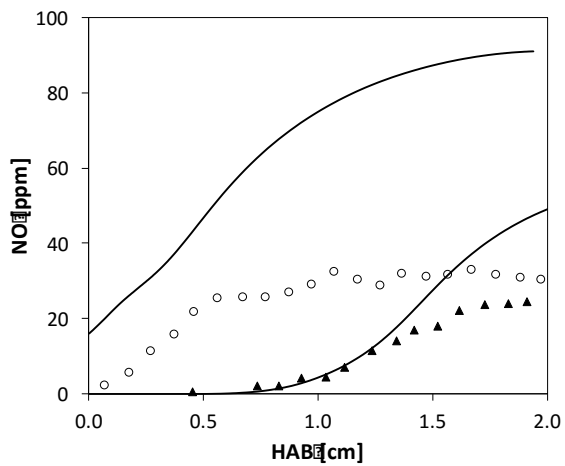
- Hierarchical dependencies strongly affect a kinetic mechanism
- E.g. **NO_x formation** depends on the core C_0-C_3 submechanism

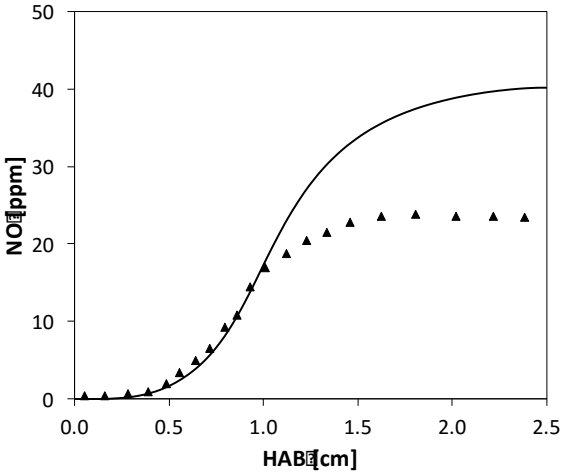
What can happen?

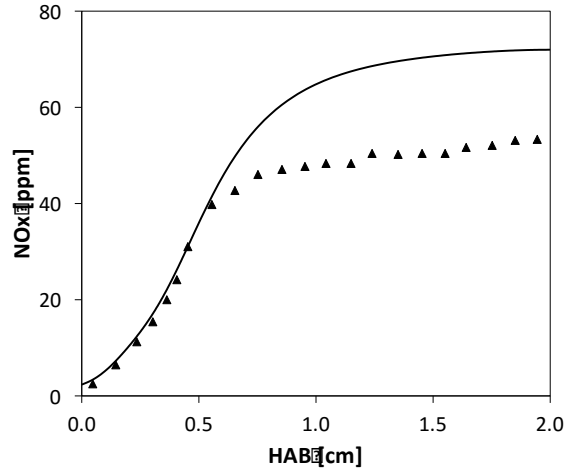
- January 2012: **critical update** to C_0-C_3

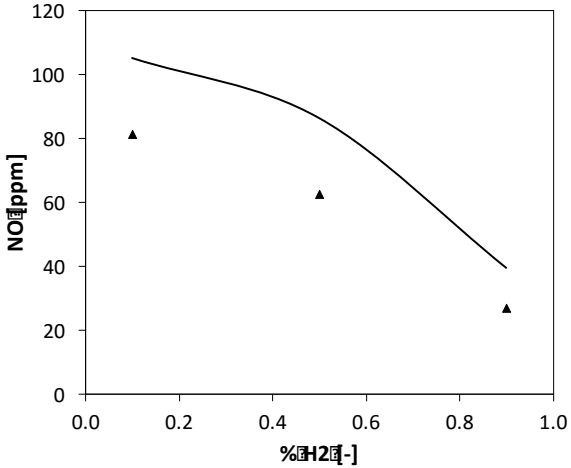


After 2012 update





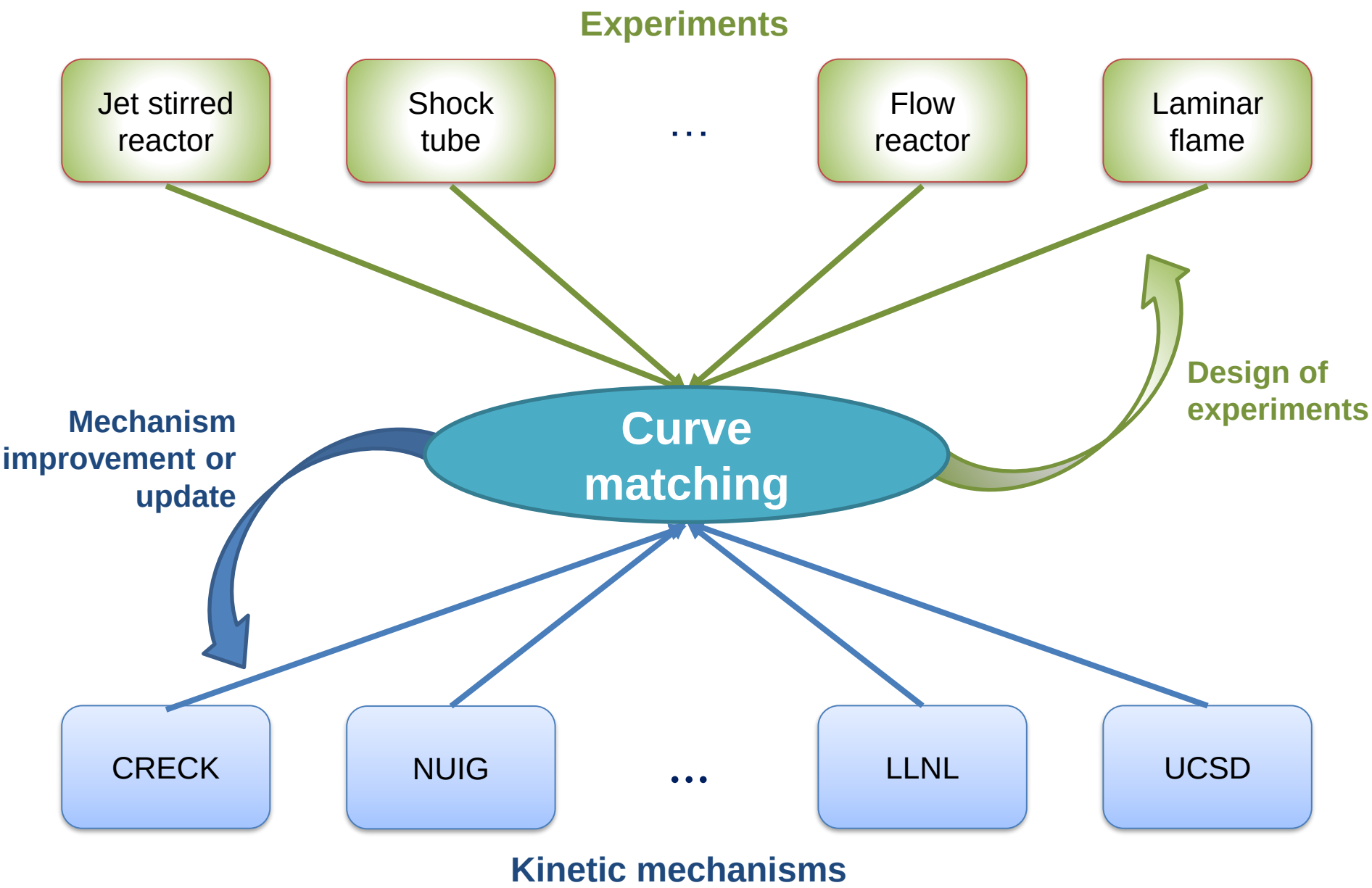




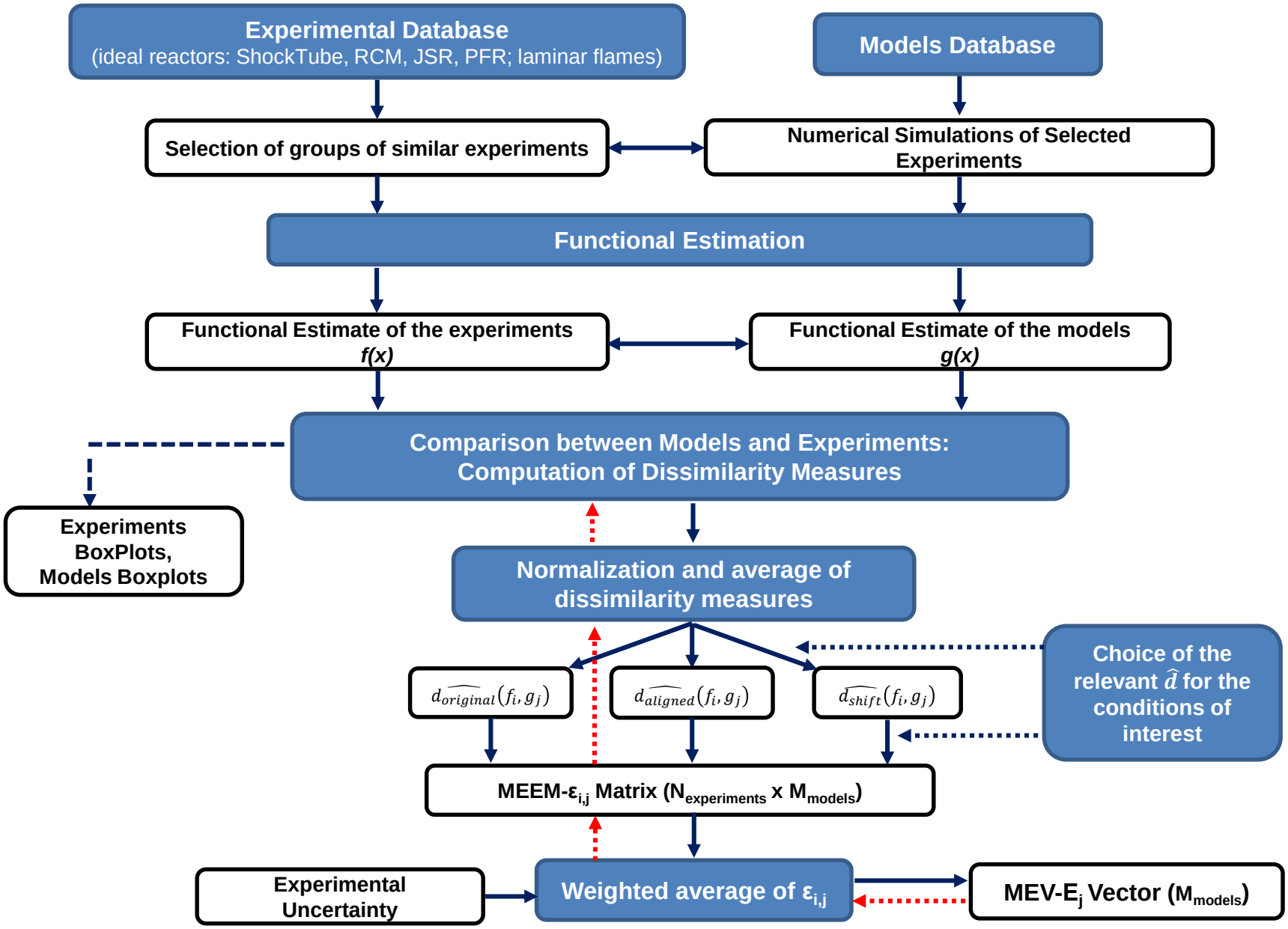
van Essen et al. Comb Flame, 153 (2008)
Sutton et al. Comb Flame 159 (2012)

Lamoureux et al. Comb Flame, 157 (2010)
Rørtveit, G.J., et al., 4 (2003)

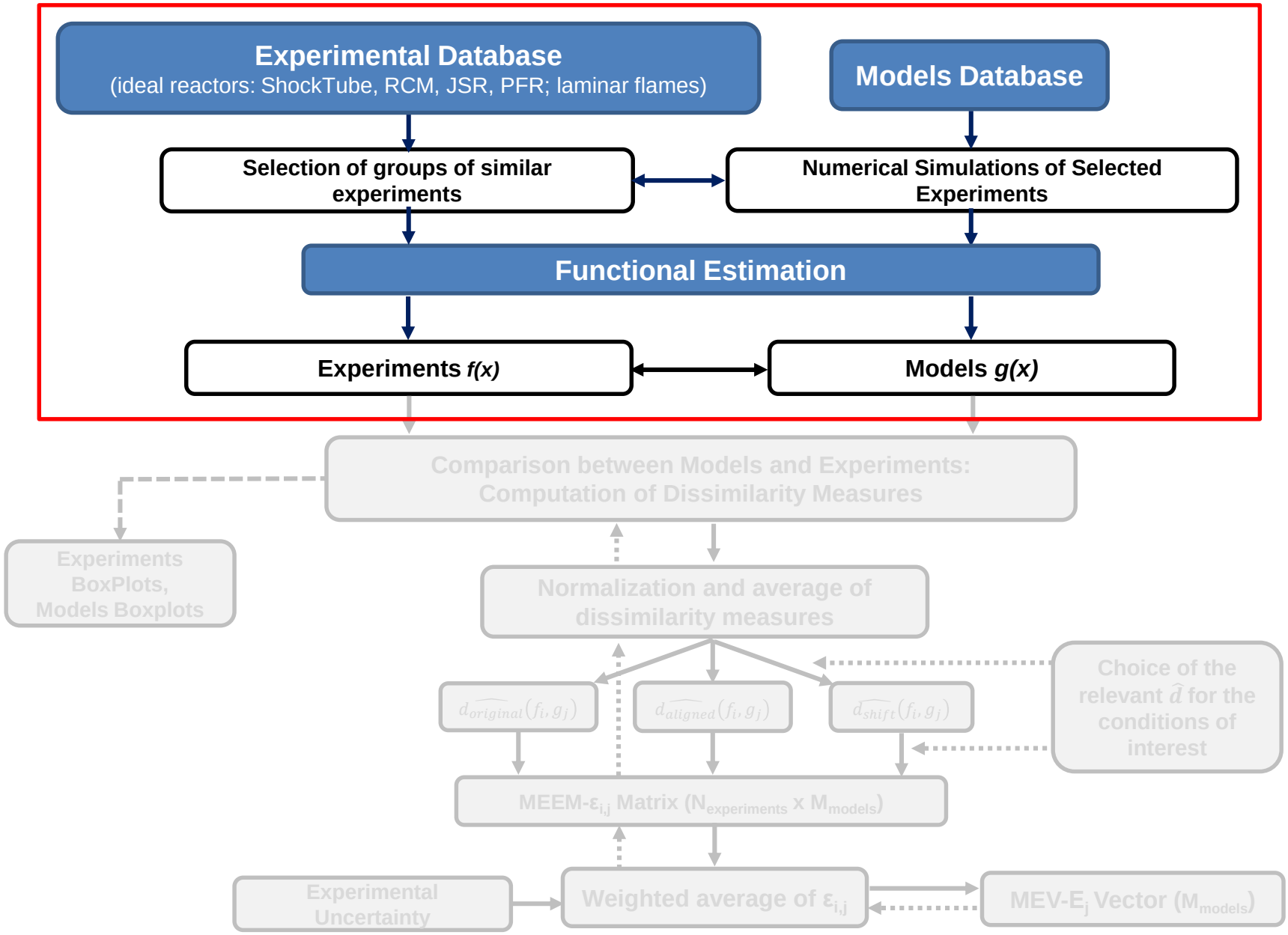
Aim of the work



The Curve Matching framework



Curve Matching methodology



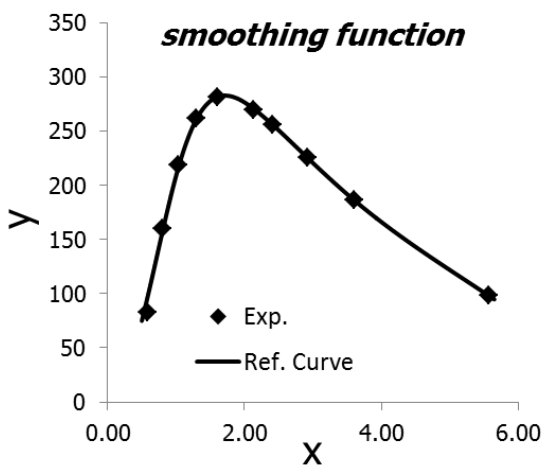
Functional estimation

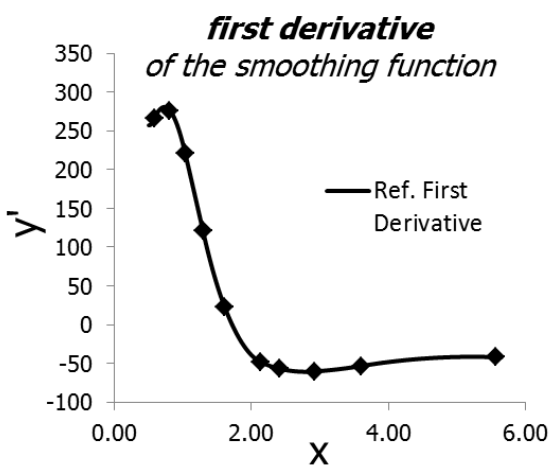
Experimental data and models are *«regular processes»* described by a **functional estimate** (5th-degree polynomial). The curves have to be square integrable and have square integrables first derivatives

$$\hat{f} = \underset{f}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \int \left(f''(x)\right)^2 dx \right]$$

Roughness penalty parameter

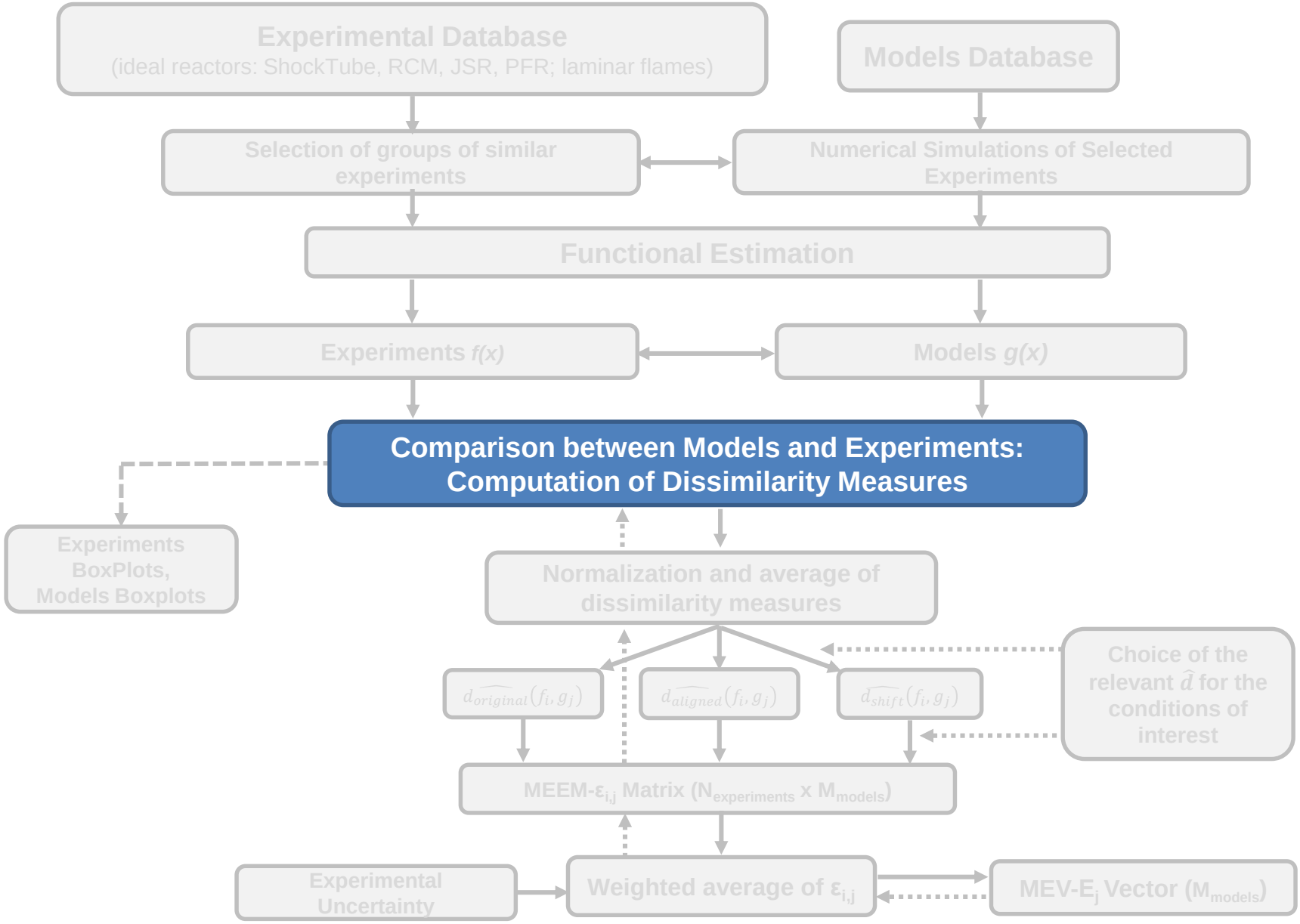
Functional representatives of the experimental measurements (***f(x),* reference curve**) and the models (***g(x)***) are obtained, together with their **first derivatives (*f'(x), g'(x)*).**





Series of data are converted into functions, and can be stored as functions!

Curve Matching methodology



Dissimilarity measures

Let $\|f\| = \sqrt{\int_{\mathbb{R}} f(x)^2 dx}$, $\|g\| = \sqrt{\int_{\mathbb{R}} g(x)^2 dx}$ and $\langle f, g \rangle = \int_{\mathbb{R}} f(x) g(x) dx$

be the norms and the scalar product in the L^2 space of the functions f(x) and g(x)

Vertical distance

$$d^0_{L^2}(f,g) = \frac{\|f-g\|}{|D|} \in (0,+\infty)$$

$$d^1_{L^2}(f,g) = \frac{\|f'-g'\|}{|D|} \in (0,+\infty)$$

→ **L_2-norm** of the difference between the two functions (**0**) or their first derivatives (**1**). **D** is the intersection between the domains of the two functions

$$d^0_{Pearson}(f,g) = \frac{1}{2}\left\|\frac{f}{\|f\|} - \frac{g}{\|g\|}\right\| \in (0,1)$$

$$d^1_{Pearson}(f,g) = \frac{1}{2}\left\|\frac{f'}{\|f'\|} - \frac{g'}{\|g'\|}\right\| \in (0,1)$$

→ measure the cosine of the angle between the two curves, *i.e. how the two curves correlate in* **shape**

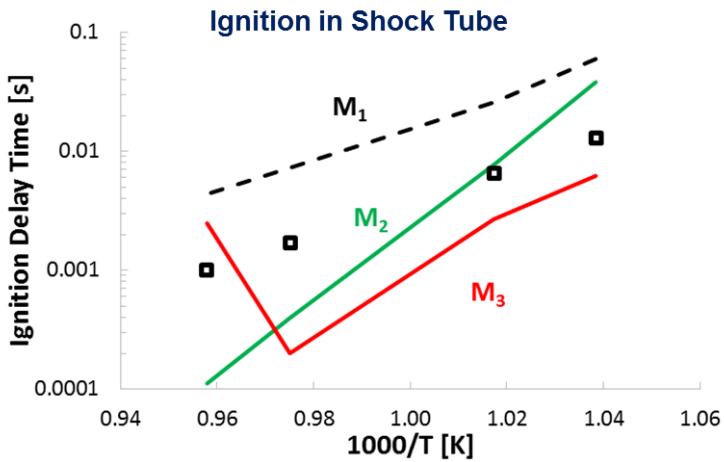
Horizontal distance

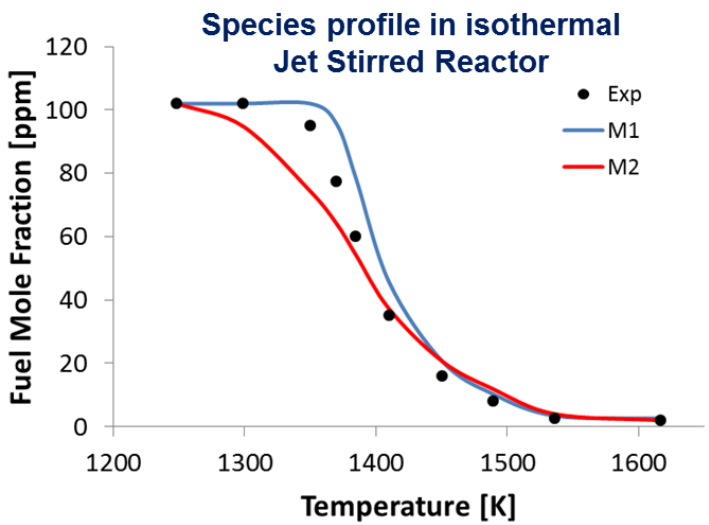
$$\delta_{opt} = \mathrm{argmin}_\delta\, d\big(f(x+\delta), g(x)\big)$$

→ **alignment** between the two functions to evaluate the horizontal distance

if d→0, g(x)→ f(x)

Dissimilarity measures

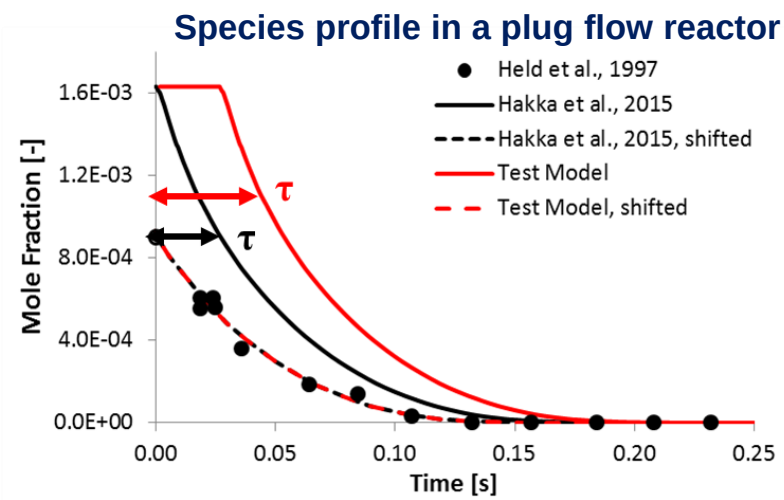




original	$d^0_{L^2,orig}$	$d^1_{L^2,orig}$	$d^0_{P,orig}$	$d^1_{P,orig}$
M_1	0.54	0.13	0.01	0.02
M_2	0.42	4.30	0.07	0.00
M_3	0.37	2.60	0.02	0.39

original	$d^0_{L^2,orig}$	$d^1_{L^2,orig}$	$d^0_{P,orig}$	$d^1_{P,orig}$
M_1	0.0035	0.0080	0.0442	0.1517
M_2	0.0036	0.0083	0.0405	0.1676
shift	$\delta^0_{L^2}$	$\delta^1_{L^2}$	δ^0_P	δ^1_P
M_1	0.0405	0.0345	0.0365	0.0365
M_2	0.0265	0.0225	0.0025	0.0305
aligned	$d^0_{L^2,align}$	$d^1_{L^2,align}$	$d^0_{P,align}$	$d^1_{P,align}$
M_1	0.0013	0.0055	0.0188	0.0928
M_2	0.0032	0.0080	0.0404	0.1543

Dissimilarity measures



original	$d^0_{L^2,orig}$	$d^1_{L^2,orig}$	$d^0_{P,orig}$	$d^1_{P,orig}$
Hakka *et al.*, 2015 [23]	0.31	0.64	0.04	0.12
Test Model	0.67	0.86	0.09	0.36
shift	$\delta^0_{L^2}$	$\delta^1_{L^2}$	δ^0_P	δ^1_P
Hakka *et al.*, 2015 [23]	0.06	0.09	0.03	0.05
Test Model	0.13	0.15	0.11	0.13
aligned	$d^0_{L^2,align}$	$d^1_{L^2,align}$	$d^0_{P,align}$	$d^1_{P,align}$
Hakka *et al.*, 2015 [23]	0.03	0.12	0.04	0.10
Test Model	0.03	0.12	0.04	0.10

Issues:

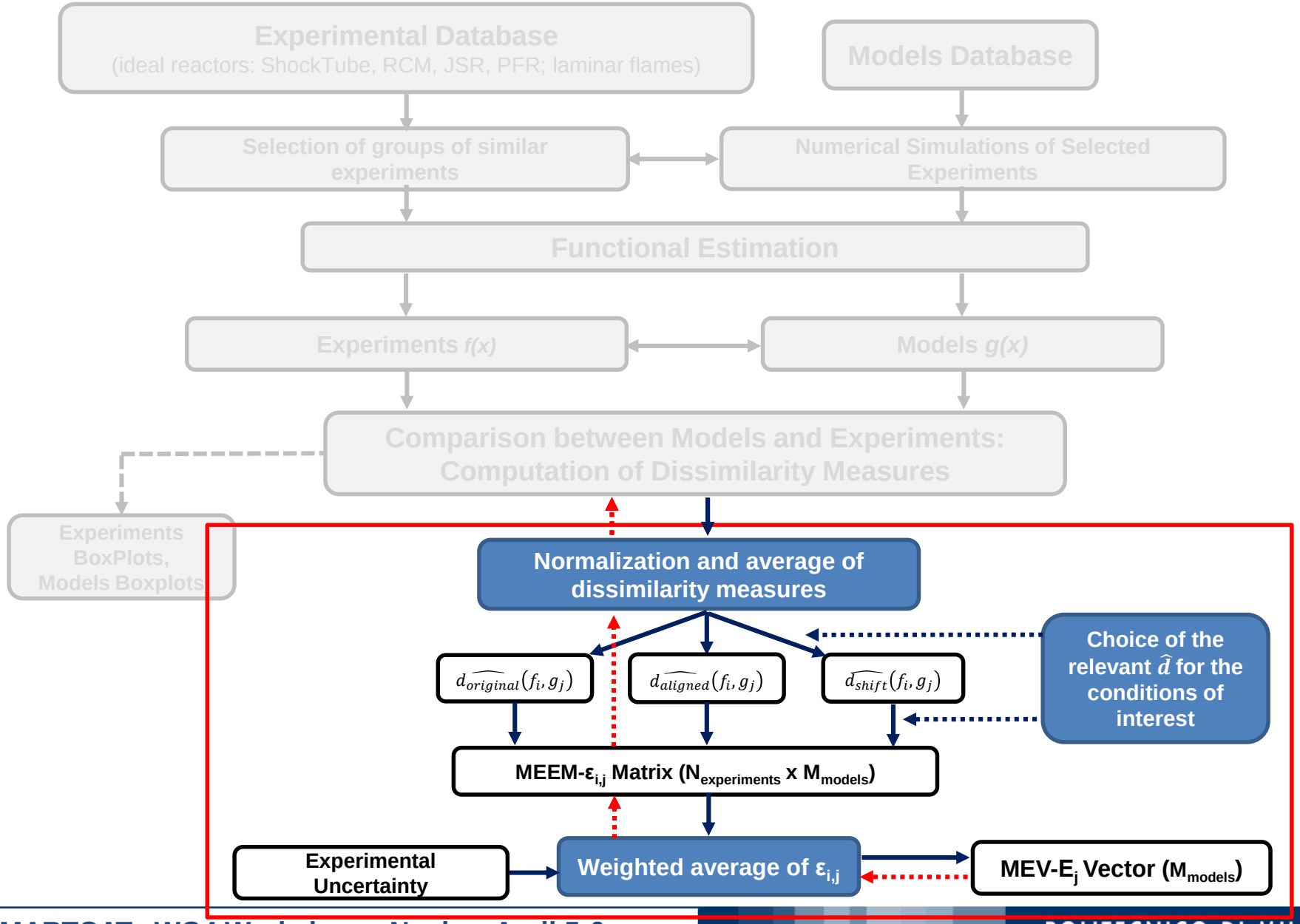
- for each ***model curve g(x), 4*** $d_{i,\,orig}(f,g)$ + ***4*** $d_{i,\,align}(f,g)$ + ***4*** $\delta_{i,opt}$ are obtained

$$\mathbf{12}\,\frac{parameter}{model\cdot dataset}\ \text{x}\ \mathbf{10}\ model\ \text{x}\ \mathbf{10\text{-}10^2}\ dataset = \underline{\mathbf{\sim 10^4}}\,\frac{parameter}{fuel}$$

- The **values of the dissimilarity measures** are **not directly comparable**: they refer to **different quantities** and are **dimensionally not homogenous**

How to extract useful information?

Curve Matching methodology



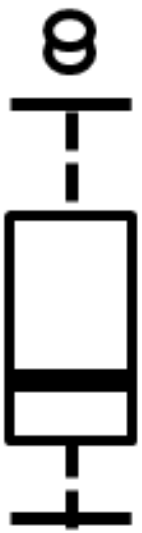
Integrated index

The **values of the dissimilarity measures** are **not directly comparable**: they refer to **different quantities** and are **dimensionally not homogenous**

$$\underbrace{\hat{d}(f_i, g_j)}_{} = \frac{1}{4}\left(\frac{d^0_{L^2}(f_i, g_j) - m^0_{L^2}}{IQR^0_{L^2}/2} + \frac{d^1_{L^2}(f_i, g_j) - m^1_{L^2}}{IQR^1_{L^2}/2} + \frac{d^0_P(f_i, g_j) - m^0_P}{IQR^0_P/2} + \frac{d^1_P(f_i, g_j) - m^1_P}{IQR^1_P/2}\right)$$

Dimensionless index measuring the relative goodness of the **model *j***, in representing a given set of **experimental data, *i*, with respect to the population of models** considered

m and IQR are computed over the dissimilarity measures of the population of models



Ref. Curve f(x)

↑ **[2, +∞] = significantly bad performances**

[1, 2] = bad performances

[-1, 1] = average performances

[-1, -2] = good performances

↓ **[-2, -∞] = significantly good performances**

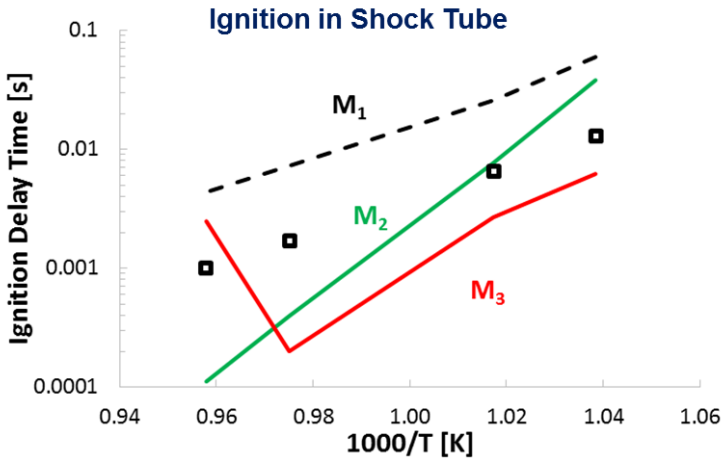
$$\hat{d}(f_i, g_j) \rightarrow \widehat{d_{origin}}(f_i, g_j),\ \widehat{d_{align}}(f_i, g_j),\ \widehat{d_{shift}}(f_i, g_j)$$

Choice of relevant $\hat{d}$

$$\hat{d}(f_i, g_j) \rightarrow \begin{cases} \cancel{\widehat{d_{origin}}(f_i, g_j)} \\ \widehat{d_{align}}(f_i, g_j) \\ \widehat{d_{shift}}(f_i, g_j) \end{cases} \Rightarrow \varepsilon_{i,j} = \frac{\sum_{i=1}^{n} \hat{d}}{n}$$

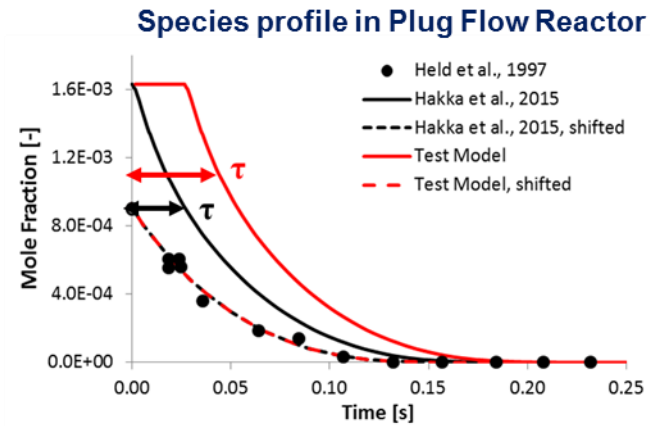
$n = 1, 2, 3$, depending on the specific case

$$\varepsilon_{i,j} = \widehat{d_{origin}}(f_i, g_j)$$



- **x=T$_{reflected}$**, defined by the experimentalist
- On a standard Ignition Delay Arrhenius plot **an horizontal shift corresponds to a vertical shift**

$$\varepsilon_{i,j} = \widehat{d_{aligned}}(f_i, g_j)$$



- Species profiles are usually time-shifted to match 50% fuel conversion, therefore we care about the aligned values, keeping memory of $\widehat{d_{shift}}(f_i, g_j)$

Model-Experiment Error Matrix

Significantly good

Significantly bad

Experiments

Models

	37	38	39	40	41	42	…	61	62	63	64	65	66
M1	-0.03	0.97	0.52	0.20	0.55	2.52	…	-0.31	-0.35	-1.37	-0.55	0.27	0.34
M2	-0.8	0.15	0.03	0.15	0.31	0.00	…	-1.16	-0.29	-0.20	-0.23	-0.38	0.23
M3	-0.68	0.31	-0.14	-0.69	0.68	0.32	…	-0.78	-0.79	-0.33	-1.13	-0.38	-0.36
M4	0.58	0.41	-0.83	-0.01	-0.58	0.81	…	-0.41	-0.59	-0.35	-0.09	0.18	-0.32
M5	0.43	0.66	-0.28	1.07	-0.61	0.43	…	-0.32	0.18	-0.38	0.62	0.68	0.37
M6	-1.05	-2.14	-2.56	-6.46	-1.19	0.19	…	-2.99	-2.34	0.84	-3.85	-0.76	-3.27
M7	-1.89	0.13	1.04	0.49	-0.16	-0.36	…	1.19	0.67	1.53	0.63	1.43	0.39
M8	-0.88	-2.5	1.85	1.03	-0.09	0.03	…	0.62	1.40	-0.87	0.86	1.17	1.00
M9	-1.26	-3.72	-4.04	-7.08	-0.68	-0.73	…	-3.46	-0.79	2.93	-2.77	-1.24	-3.69
M10	2.03	1.45	3.54	1.80	2.39	1.77	…	3.31	10.38	-1.16	5.65	4.91	0.70
M11	1.97	-0.28	-3.79	-3.26	3.20	2.35	…	0.88	1.22	1.18	-2.66	-4.94	-2.78
M12	4.35	0.28	5.63	7.95	-0.38	-1.17	…	3.10	1.90	0.41	6.74	5.55	30.96

Model-Error Vector

- Each matrix contains ~***10*** model x ~***10-10²*** dataset = ***10²-10³*** values. Too many parameters for performance analysis!
- Matrix values are **homogeneous**, and can be directly combined.

$$E_j = \frac{1}{\left(\sum_{i=1}^{N} \frac{1}{\sigma_i}\right)} \sum_{i=1}^{N} \frac{\varepsilon_{ij}}{\sigma_i}$$

Experimental uncertainty needs to be accounted for

N_{exp} x M_{models}, MEEM
$\varepsilon_{i,j}$ for $i = 1 \dots N$

→

Weighted average of
$\varepsilon_{i,j}$ for $i = 1 \dots N$

→

Model Error Vector, MEV
E_j for $j = 1 \dots M$

Experimental Uncertainty:

- **σ of the experimental measurements (*if available from the experimental study*)**
- **10% error for Ignition Delay Times***
- **2 cm/s error for Laminar Flame Speed***
- **1% of the highest measured concentration or the lowest measured concentration***

$Model_j$	E_j	$Model_j$	E_j
M1	0.15	***M7***	0.02
M2	1.34	***M8***	-0.17
M3	0.01	***M9***	-0.93
M4	0.43	***M10***	1.38
M5	0.66	***M11***	0.30
M6	-1.08	***M12***	**2.43**

* as in Olm et al., 2015, Comb Flame 161 (2014)

How to leverage such information?

Example: updating *n*-heptane mechanism

Speciation in Jet Stirred Reactors

MEV

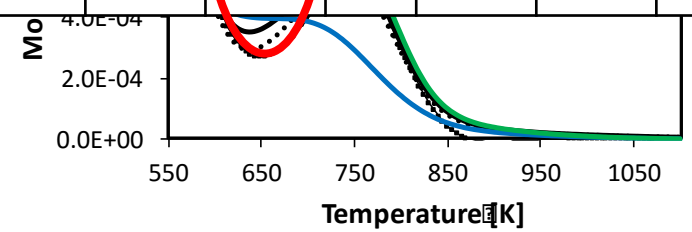
Model$_j$	E_j
POLIMI_1412	-0.56
POLIMI_1604_a	-1.36
POLIMI_1604_b	1.50
POLIMI_1604_c	0.56

MEEM – Case 04

Exp. Data from Dagaut et al., Comb Flame, 1995

Experiments

Models	01	02	03	04	…	28	29	30
POLIMI_1412	0.02	-0.21	-0.67	-0.42	…	-0.82	1.62	-0.62
POLIMI_1604_a	-1.67	-2.05	-1.75	-2.48	…	-1.32	-3.00	-1.50
POLIMI_1604_b	1.65	1.70	3.00	3.53	…	0.65	1.65	-0.45
POLIMI_1604_c	0.13	0.45	0.82	0.23	…	0.23	0.03	0.75

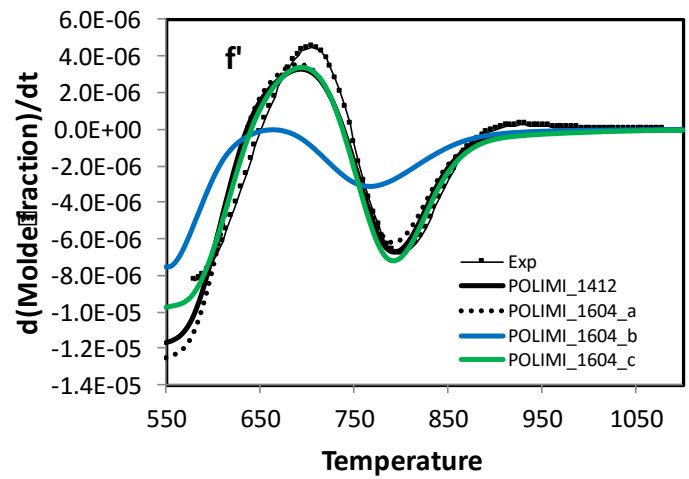


Dissimilarity Values, Original

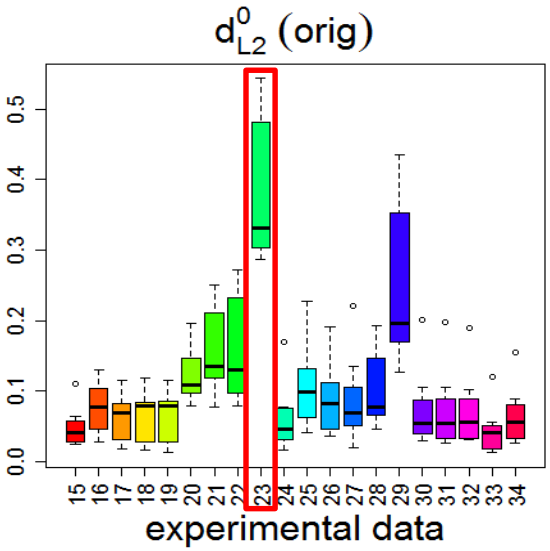
Model$_j$	$d^0_{L^2,orig}$	$d^1_{L^2,orig}$	$d^0_{P,orig}$	$d^1_{P,orig}$
POLIMI_1412	0.004	0.005	0.059	0.125
POLIMI_1604_a	0.002	0.014	0.178	0.388
POLIMI_1604_b	0.007	0.016	0.174	0.432
POLIMI_1604_c	0.006	0.004	0.063	0.108

Dissimilarity Values, Aligned

Model$_j$	$d^0_{L^2,align}$	$d^1_{L^2,orig}$	$d^0_{P,orig}$	$d^1_{P,orig}$
POLIMI_1412	0.004	0.003	0.049	0.093
POLIMI_1604_a	0.002	0.004	0.042	0.112
POLIMI_1604_b	0.006	0.011	0.146	0.213
POLIMI_1604_c	0.006	0.002	0.057	0.065



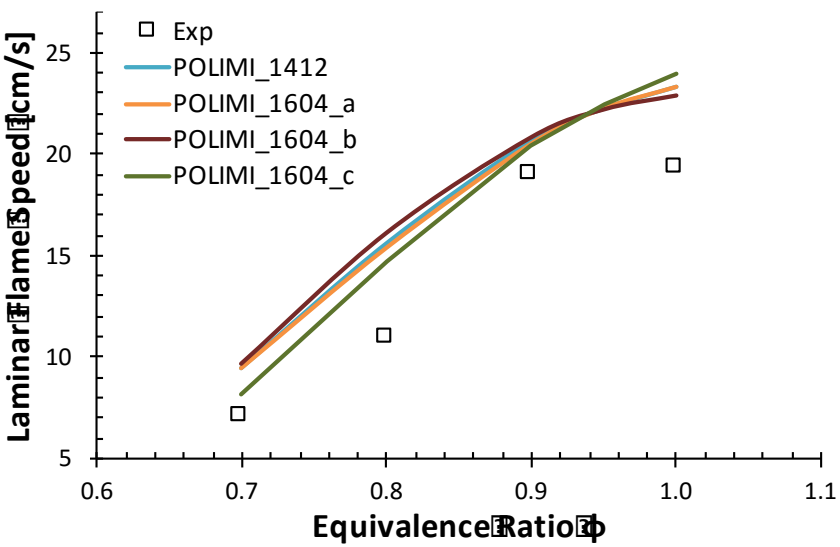
Critical experimental conditions



Boxplot visualization allows to clearly identify possible systematic errors, which might be due to:

- a shared lack of knowledge of all the models
- issues with experimental measurements

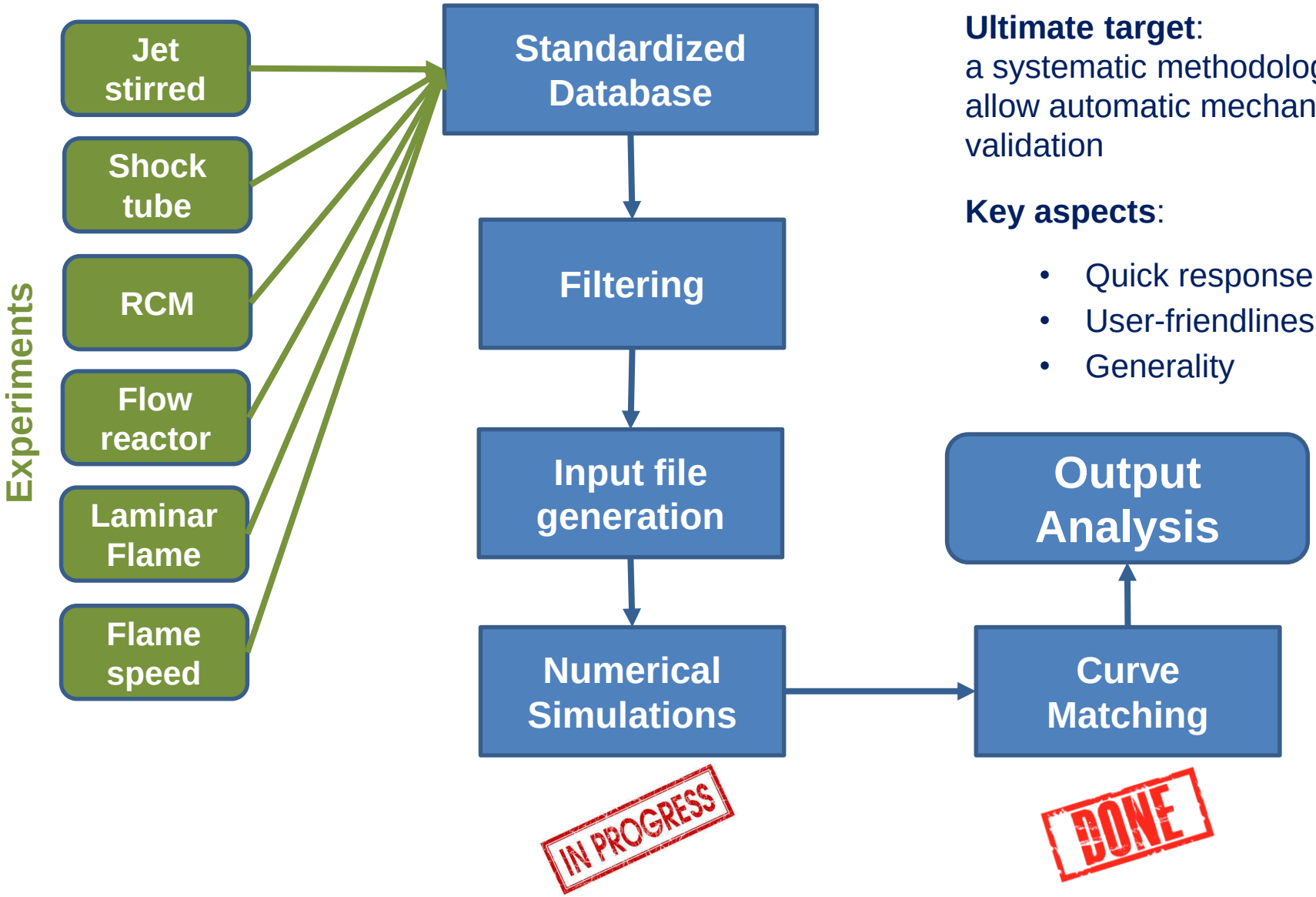
Case 23, Laminar Flame Speed, Tu=373 K, p=25 bar



Model$_j$	$d^0_{L^2,orig}$	$d^0_{L^2,shift}$	$d^0_{L^2,align}$
POLIMI_1412	0.291	0.216	0.070
POLIMI_1604_a	0.287	0.209	0.074
POLIMI_1604_b	0.315	0.238	0.071
POLIMI_1604_c	0.327	0.257	0.053

Exp. Data from Jerzembeck et al., Combust Flame, 2009

Towards an automatic procedure



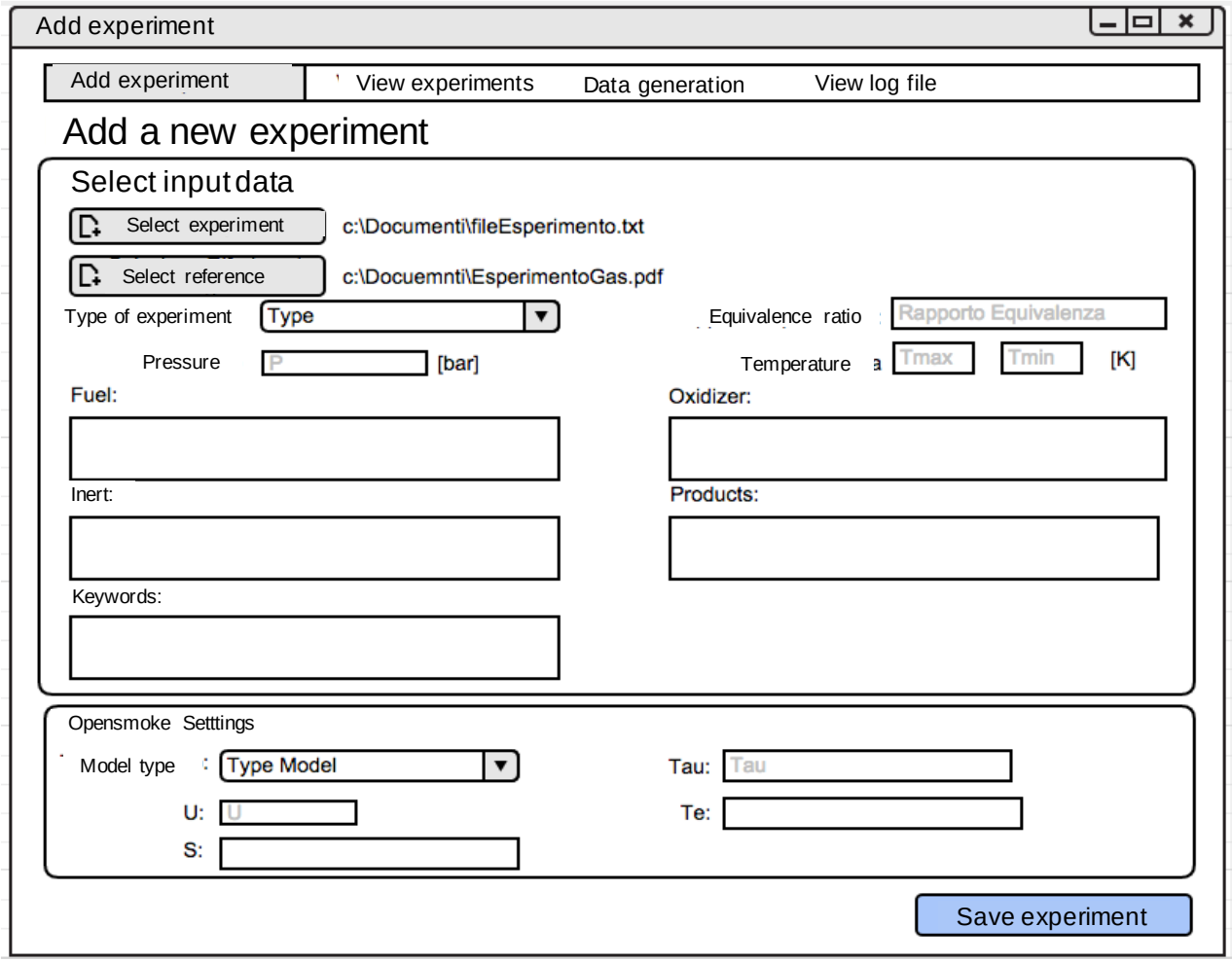
Ultimate target:
a systematic methodology to allow automatic mechanism validation

Key aspects:

- Quick response
- User-friendliness
- Generality

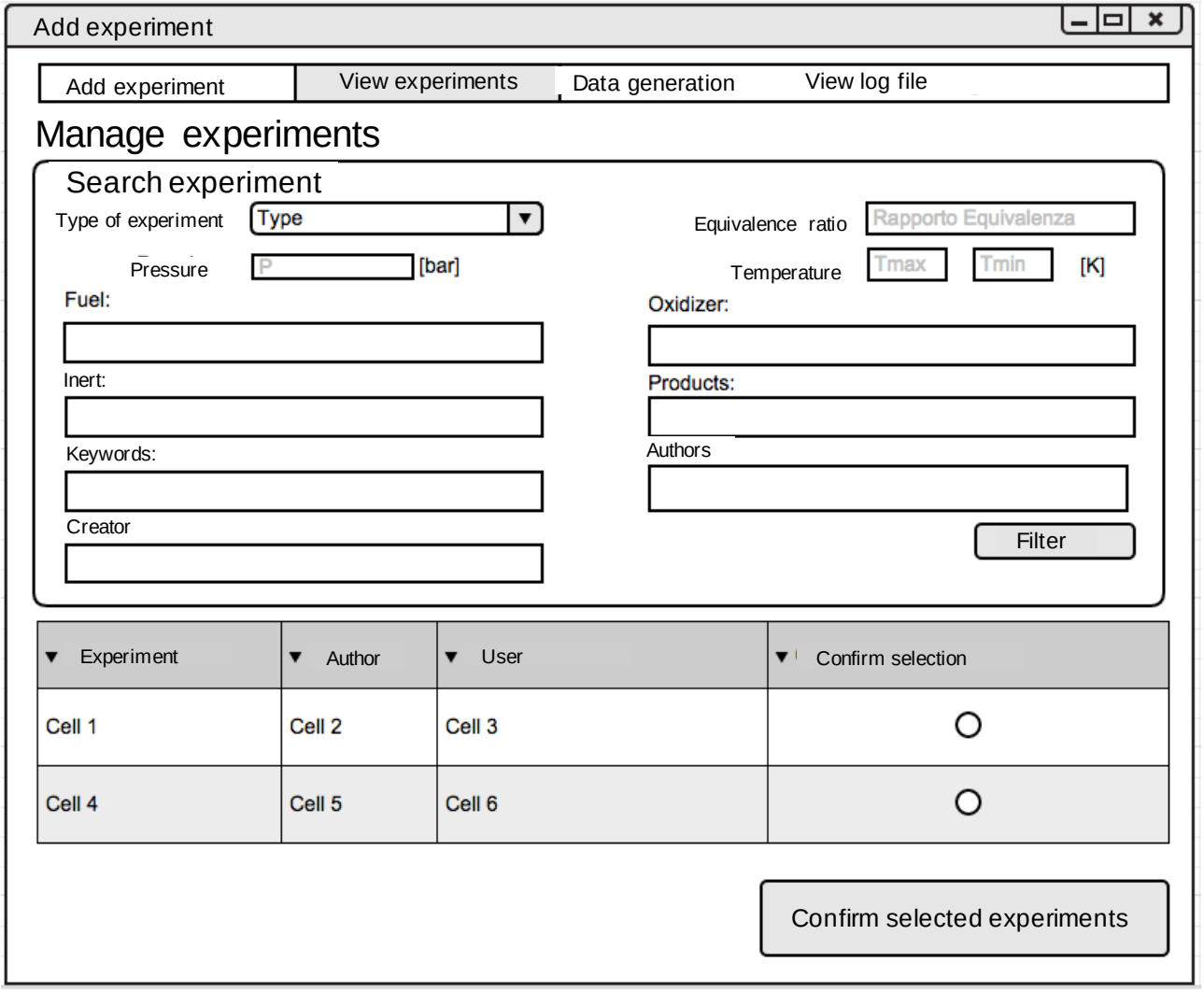
Automatic generation

1. Add experimental data to database



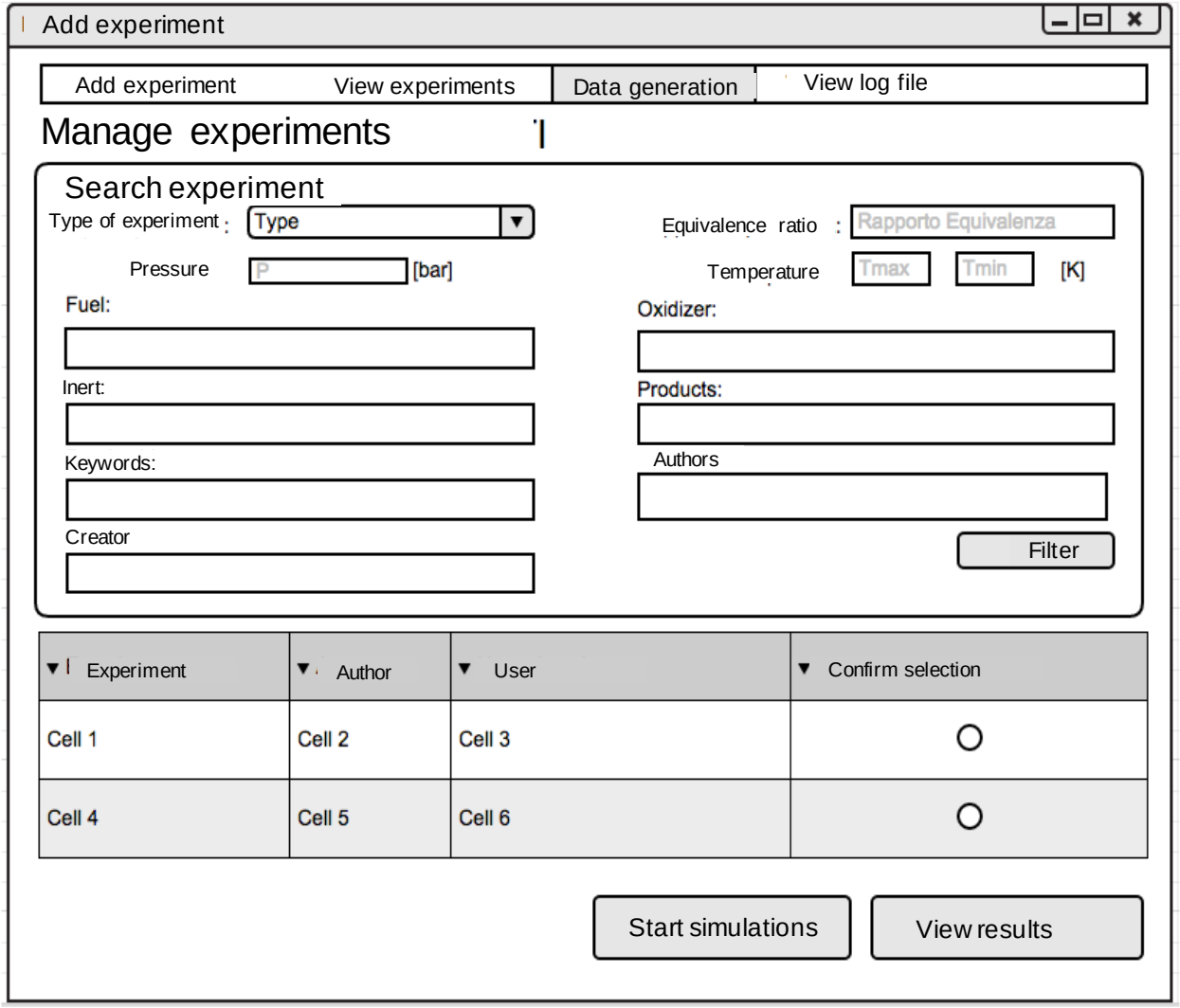
Automatic generation

2. Select experiments from database



Automatic generation

3. Start simulations



Automatic generation

4. View results

View results

Add experiment | View experiments | Data generation | View log file

View log file

Search experiment

Type of experiment: Type

Pressure: P [bar]

Fuel:

Inert:

Keywords:

Creator

Equivalence ratio : Rapporto Equivalenza

Temperature a Tmax Tmin [K]

Oxidizer:

Products:

Authors

Filter

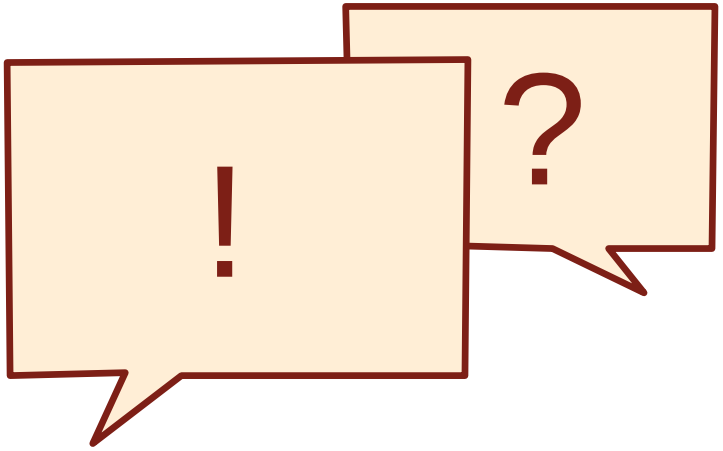
Experiment	Author	User	Error
Cell 1	Cell 2	Cell 3	err
Cell 4	Cell 5	Cell 6	err
Cell 7	Cell 8	Cell 9	err
Cell 10	Cell 11	Cell 12	err

Conclusions

- **Curve Matching**: an innovative framework applying functional data analysis and statistical analysis to chemical kinetics. Two main tasks:
  - Assess mechanism capability in reproducing experimental data
  - Compare different mechanisms
- **Multi-faceted** analysis of functional data obtained from models and experiments.
  - Distance and similarity norms
  - Functions and first derivatives
- **Pyramidal approach** in output analysis:

Model-Error Vector → **Model-Exp Error Matrix** → **Dissimilarity values**

What's next?

- Software finalization
  - Integration with standardized **database**
  - User-friendly (and online) **interface**
- **Design of Experiments** from Curve Matching output

THANK YOU
FOR YOUR ATTENTION