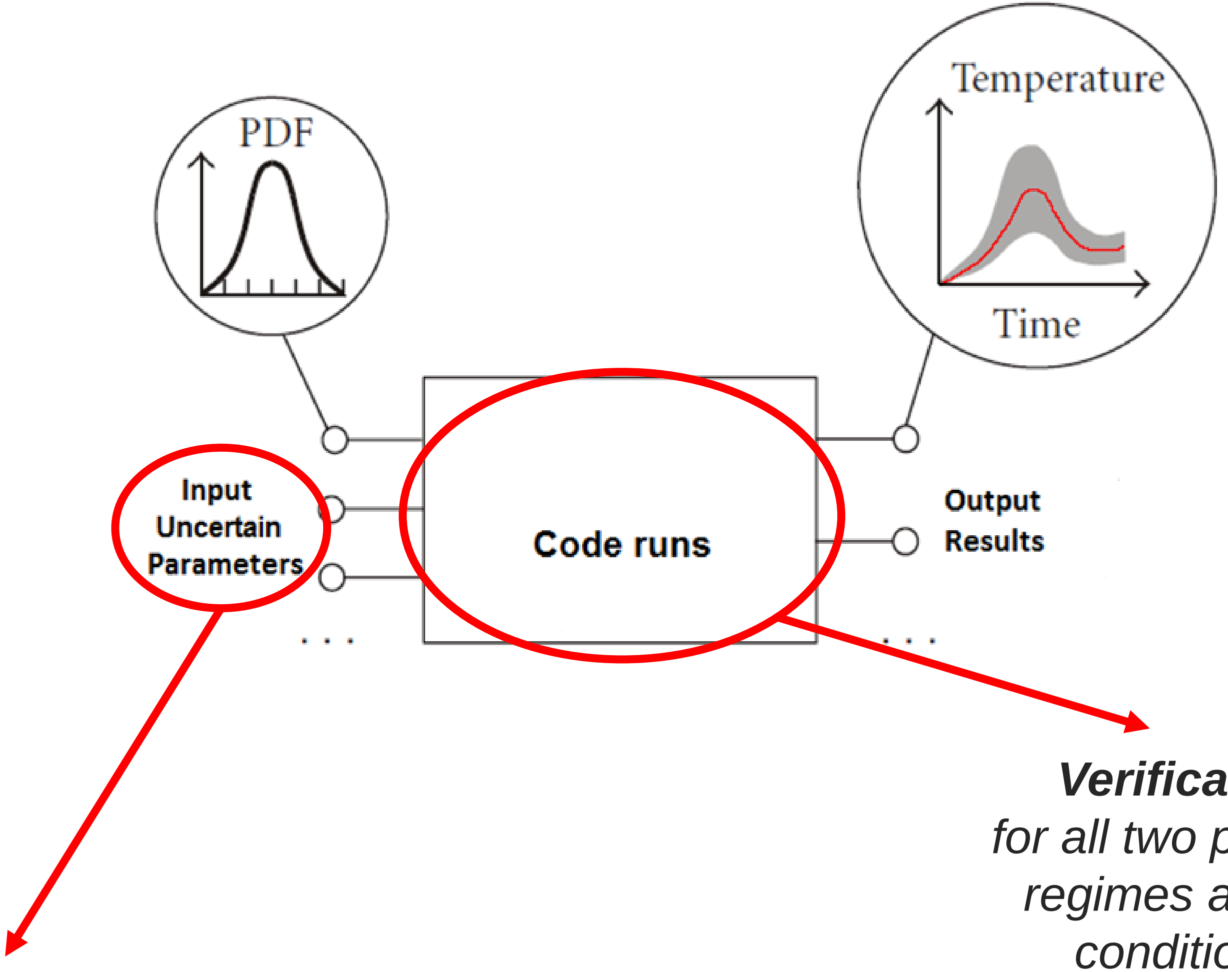
How to quantify uncertainties in thermal-hydraulic simulations for nuclear safety?
The international ATRIUM benchmark

A. Ghione, L. Sargentini, G. Damblin

Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), Saclay, France

Workshop "Expériences et simulations : comment les planifier et les exploiter au mieux ?", Avignon Université – November 6th – 7th, 2025

Best Estimate Plus Uncertainty (BEPU)



Verification & Validation (V&V)
for all two phase flow and heat transfer regimes and for large variety of flow conditions and flow geometries

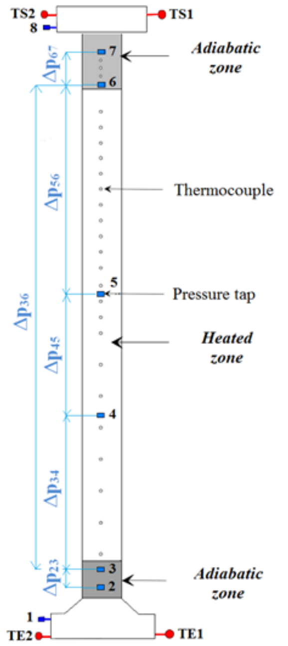
Biggest issue: input uncertainties?

Which ones? How to quantify? How accurate?

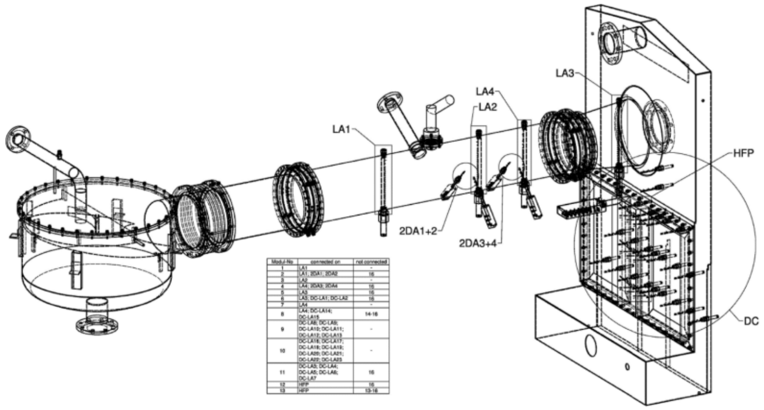
Verification and Validation

- **Verification** = Determine if the equations are correctly solved from a numerical point of view
- **Validation** = Determine if the physical models and the whole code are consistent/reliable/good comparing with experimental data

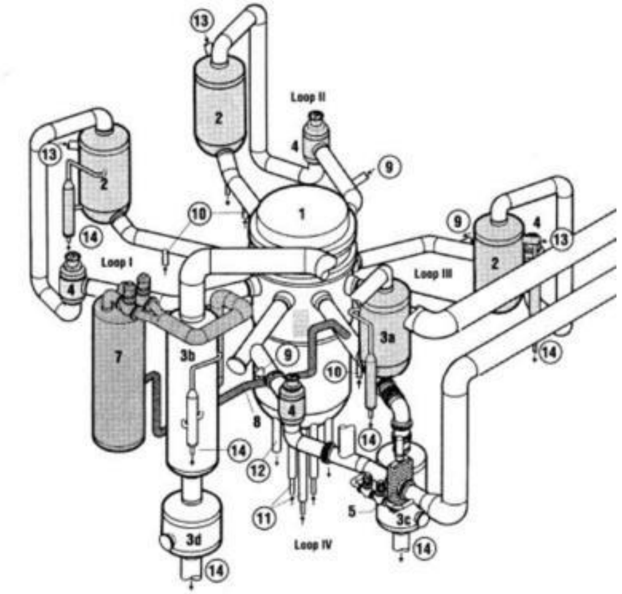
Separated Effect Tests (SETs)



Combined Effect Tests (CETs)



Integral Effect Tests (IETs)



Uncertainty sources in T-H simulations

- Plant uncertainties (initial and boundary conditions, geometry, material data)
- Representation uncertainties (discretization)
- Scaling uncertainties
- User effect (modelling choices)
- Code and model uncertainties: Numerical scheme, Physical models of the code

Inverse Uncertainty Quantification (IUQ)

- Determined by expert judgement or dedicated statistical methods (e.g. CIRCE)
- Often questioned by safety authorities

Main ideas behind IUQ methods

- Based on the comparison between exp. Quantity of Interest (QoI) and its simulation
- Exp. data comes from Separate (SETs) or Combined (CETs) Effect Tests
- Same data that can be used for Validation → Validation = IUQ prerequisite
- Uncertainty λ usually modeled as multiplicative → $\boldsymbol{M_i \rightarrow \lambda_i M_i}$
- Goal of IUQ: Estimate *pdf*s of λ_i
- Discrepancies between exp. and calc. values of the QoI are modelled as:

$$\boldsymbol{\varepsilon_{\mathrm{i}} = y_{\mathrm{exp,i}} - y_{\mathrm{calc},i}(\lambda_i, x_i) \qquad 1 \le i \le n}$$

$\boldsymbol{\varepsilon_{\mathrm{i}}}$ includes experimental uncertainties
(often neglected → conservative hypothesis)

URANIE & CIRCE training at CEA-Saclay : 27-28 Nov. 2025

1. The ATRIUM project

Context

Former OECD-NEA activities on uncertainties in thermal-hydraulics

2004 - 2010 **BEMUSE**

(Best Estimate Methods plus Uncertainty and Sensitivity Evaluation)

Uncertainty propagation for a LB-LOCA ➔ *what about the input uncertainties?*

2012 - 2015 **PREMIUM**

(Post BEMUSE Reflood Models Input Uncertainty Methods)

Inverse uncertainty quantification (IUQ) for reflooding ➔ *important user-effect*

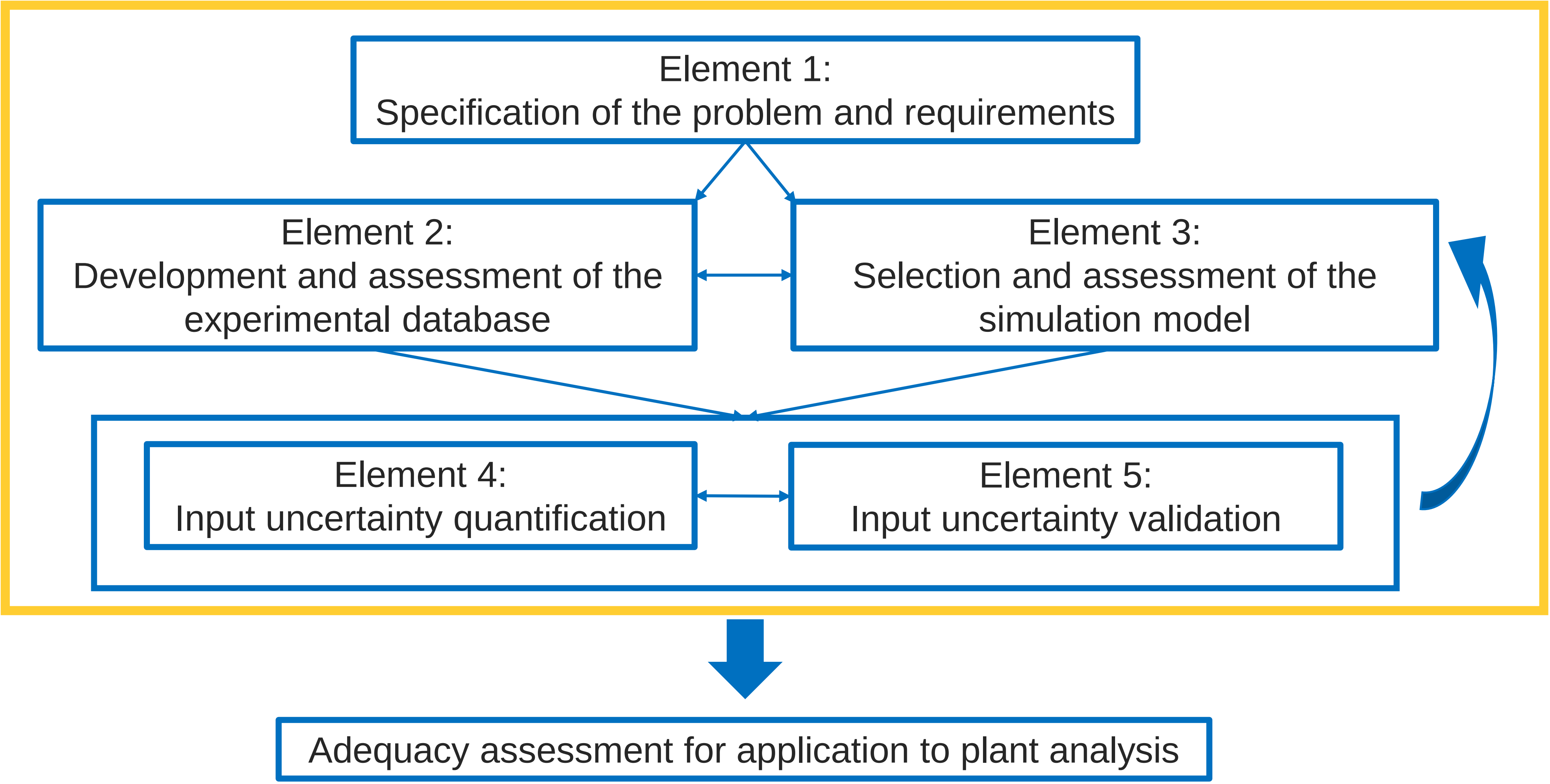
2017 - 2019 **SAPIUM**

(Systematic APproach for Input Uncertainty quantification Methodology)

Guideline for IUQ ➔ *theory but no application*

SAPIUM guideline

IUQ as a global process!

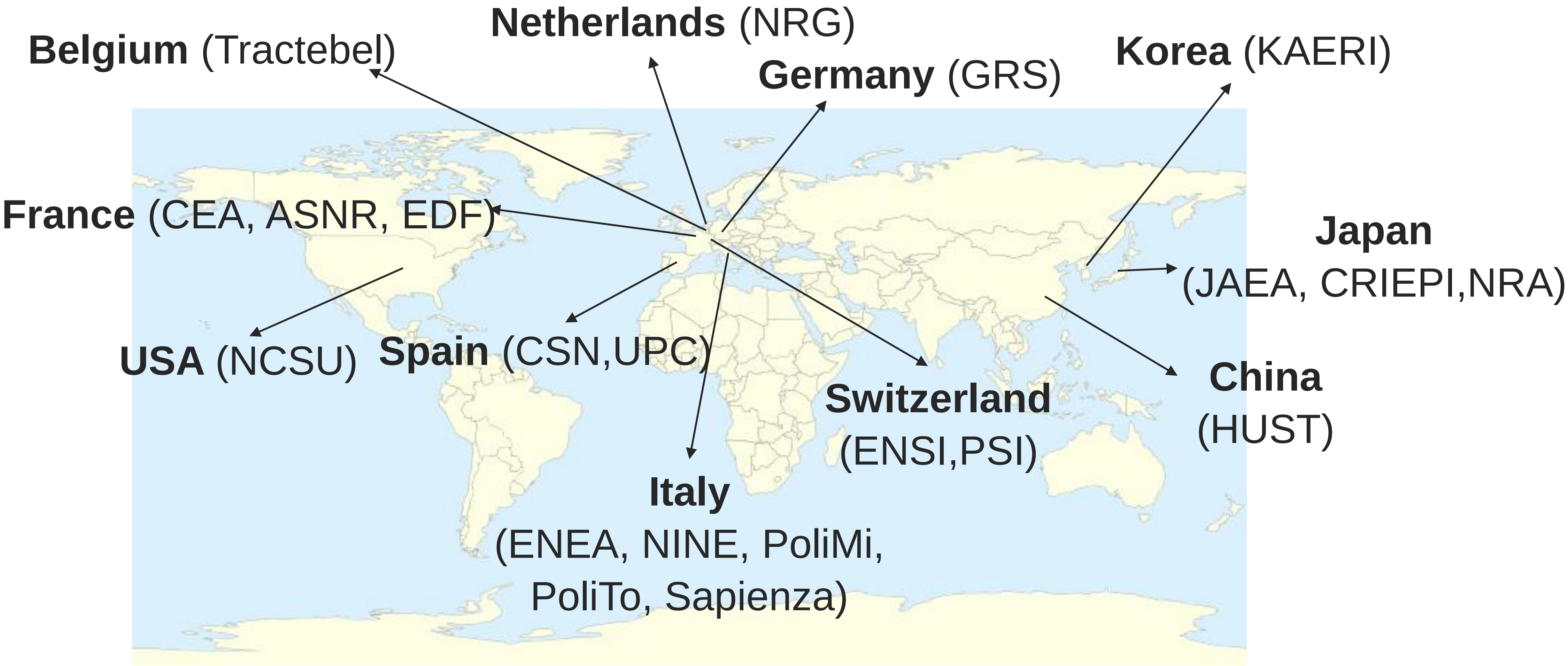


Goal of the ATRIUM project (2022 – 2025)

- Perform practical IUQ exercises in order to:
  - Demonstrate the applicability of the SAPIUM guideline;
  - Identify and try to solve potential issues coming from its practical application;
  - Summarize the lessons learnt from participants and possibly update the guideline
  - Progress on BEPU (Best-Estimate Plus Uncertainty) methodology

- Focus on Adequacy and IUQ for Intermediate Break Loss Of Coolant Accident (IB-LOCA)

Participants

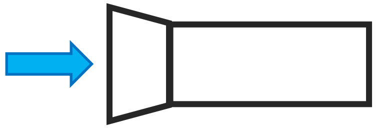


21 institutions from 11 countries hosted by OECD-NEA!

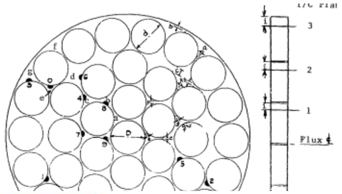
Key steps of the ATRIUM project

1. Uncertainty quantification of the physical models relevant for **critical flow** and **PCT** → Two IUQ exercises with increasing complexity (SETs → CETs):

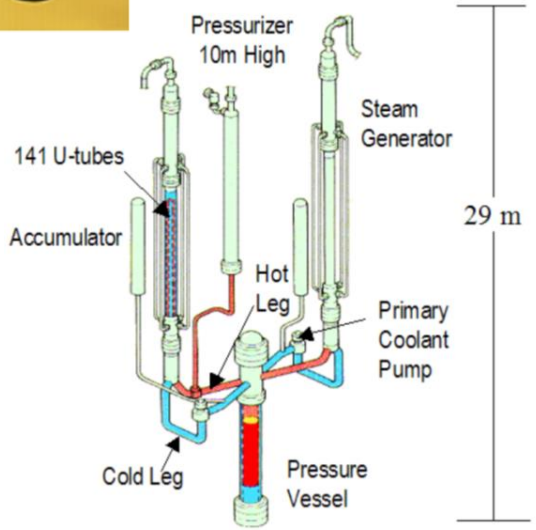
1st EXERCISE on CRITICAL FLOW



2nd EXERCISE on POST-CHF



2. Forward propagation of uncertainties on LSTF IB-HL-01 (IB-LOCA)



How to do validation and IUQ?

For each exercise, the SAPIUM guideline will be applied:

1. Database analysis
2. Adequacy
3. Physical models choice
4. Creation of the file for code for BE simulations (choice of nodalization)
5. IUQ with the own methodology
6. Analysis and validation based on the envelope band on the Figure of Merit (critical mass flow rate and cladding temperature)

Experimental analysis (1–2)

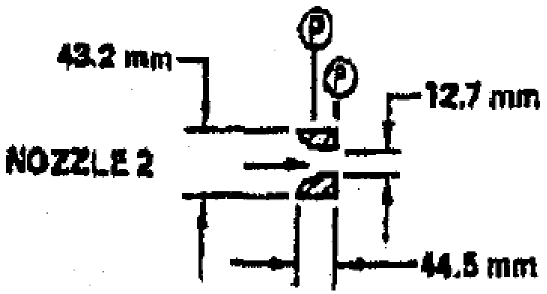
Validation (3–4)

IUQ (5–6)

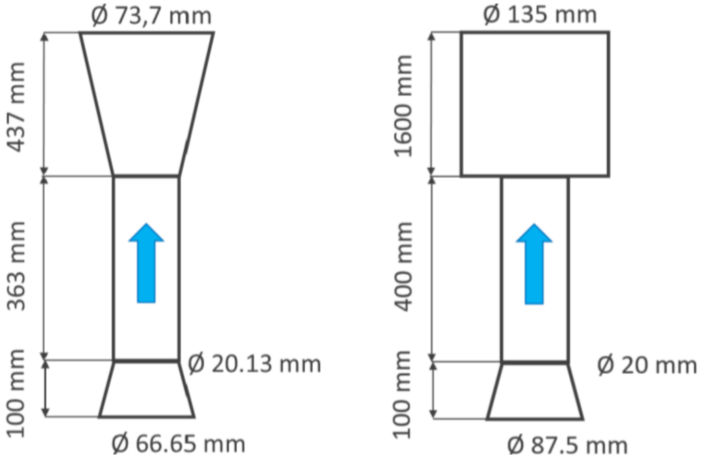
Participants must describe and justify the adequacy of the experimental databases, the modelling choices and the IUQ methodologies.

2. 1st exercise on critical flow: main results

1st exercise: Critical flow experiments

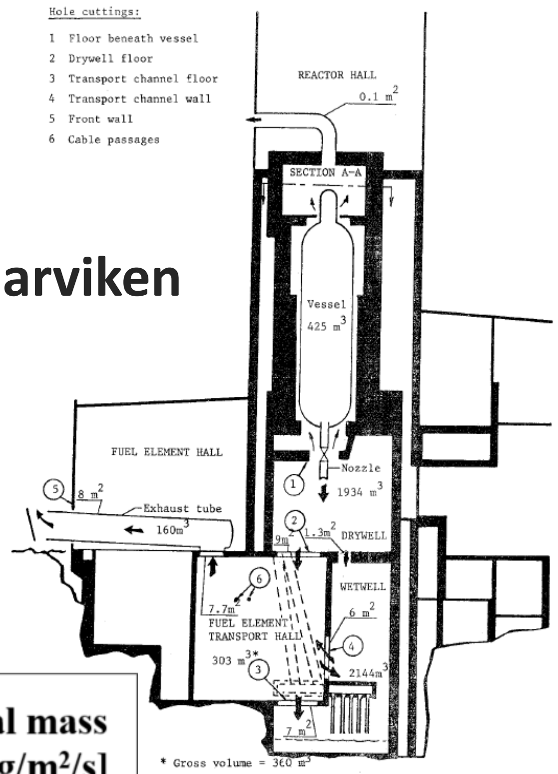


Sozzi-Sutherland



SuperMoby Dick

Marviken



		D [mm]	L/D	N° tests	Inlet pressure [bar]	Inlet thermodynamic quality [-]	Critical mass flux [kg/m²/s]
Sozzi - Sutherland	N2	12.7	0 – 140	358	56.0 – 71.3	-0.1439 – 0.0065	17528 - 75824
	N3	12.7	0	58	42.7 – 69.0	-0.2192 – 0.0060	33161 - 61226
	N4	19	0	23	56.0 – 66.3	-0.0095 – 0.0100	29295 - 51266
SMD	Div.	20.13	18	27	20.0 – 120.1	-0.0991 – -0.0005	15300 - 62200
	Exp.	20.0	20	12	20.0 – 120.1	-0.0991 – -0.0005	16100 - 61800
Marviken	13	200	3	1	40 – 52.8	< 0	23650 - 89200
	17	300	3.7	1	31 – 51.4	< 0	23600 - 61700
	24	500	0.33	1	25 – 51.7	< 0 (t < 30 s) > 0 (30 < t < 54 s)	18000 - 59750
Target domain in LSTF IB-HL-01							
LSTF IB-HL-01		41	12	1	10 – 155	-0.153 – 1.0	1500 – 46000

2-1. Adequacy analysis

Definition of the adequacy

Is the experimental database adequate w.r.t. the transient of interest?

Adequacy = representativeness + completeness

representativeness: ability of each experimental test to provide relevant information for the problem of interest

completeness: ability of a set of experimental tests to cover the physical conditions of the problem of interest

Representativeness criteria

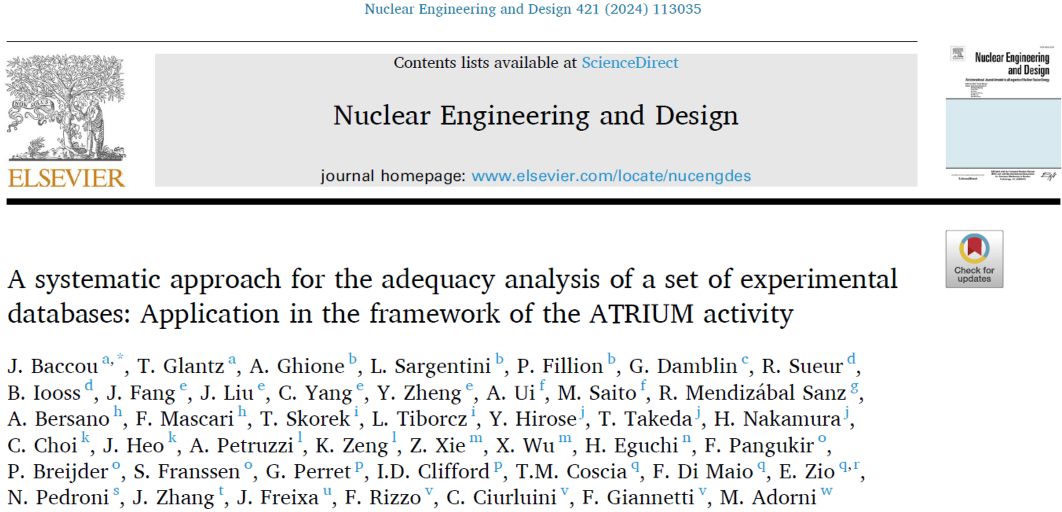
Criteria should be generic and applicable to other cases of study

Criteria	Sub-criteria
C^r_1: Fidelity with the target application facility for the accidental transient of interest	$C^r_{1\text{-}1}$: Fidelity experimental facility geometry/target application facility geometry
	$C^r_{1\text{-}2}$: Fidelity of thermal-hydraulic conditions between experiment and target application facility
C^r_2: Control of experimental data	*Dedicated Check-list*
C^r_3: Modelling of the physical phenomena for their implementation in the system code	$C^r_{3\text{-}1}$: Capability to cover physical phenomena of interest required for the simulation
	$C^r_{3\text{-}2}$: Separability
	$C^r_{3\text{-}3}$: Capability of the simulation tool to reproduce the experimental data

Consensus of the participants on the criteria!

Summary of the adequacy

- Definition of a systematic approach for adequacy analysis with generic criteria for representativeness and completeness
- First application of a transparent and reproducible approach for adequacy analysis (multi-criteria analysis)
- Consensus reached on the most and the least representative database
- Several insights on SAPIUM guideline:
  - Adequacy (especially completeness) should be repeated/complemented before the IUQ step
  - Representativeness is easier to define than completeness


Nuclear Engineering and Design 421 (2024) 113035

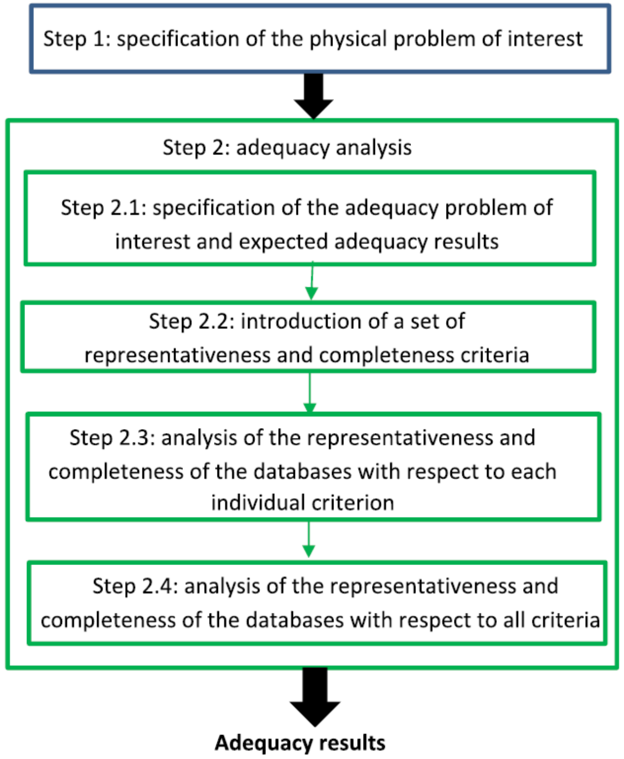
Contents lists available at ScienceDirect

Nuclear Engineering and Design

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

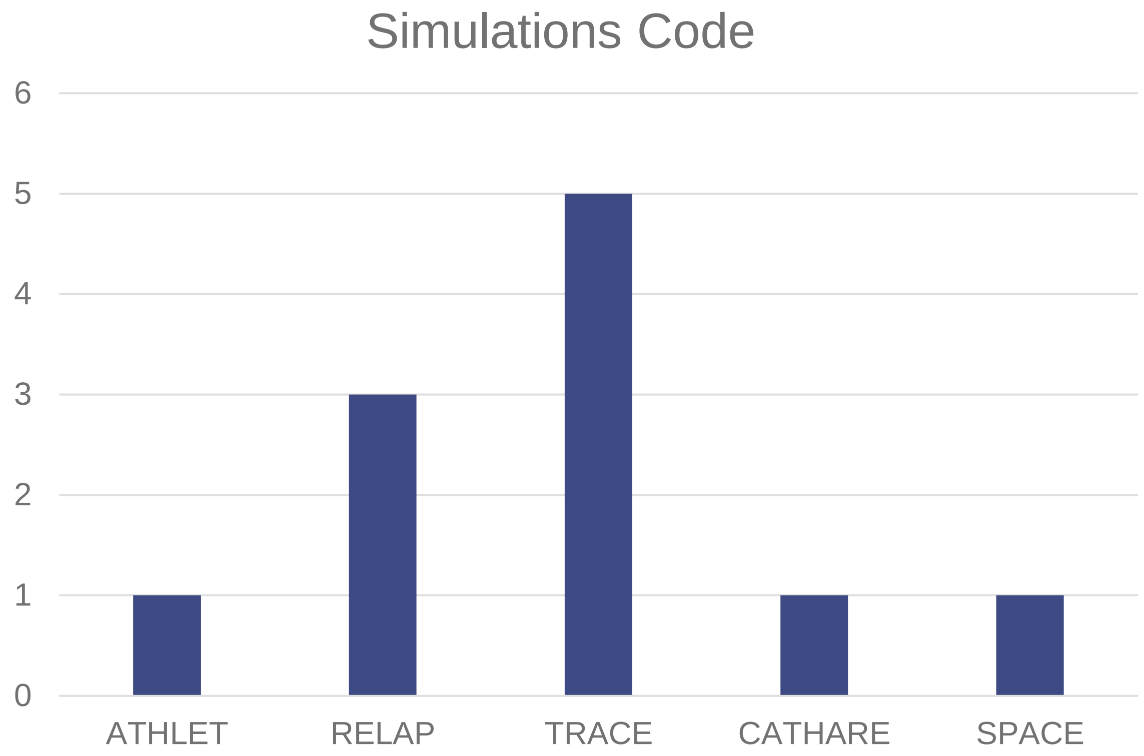
A systematic approach for the adequacy analysis of a set of experimental databases: Application in the framework of the ATRIUM activity

J. Baccou[a,*], T. Glantz[a], A. Ghione[b], L. Sargentini[b], P. Fillion[b], G. Damblin[c], R. Sueur[d], B. Iooss[d], J. Fang[e], J. Liu[e], C. Yang[e], Y. Zheng[e], A. Ui[f], M. Saito[f], R. Mendizábal Sanz[g], A. Bersano[h], F. Mascari[h], T. Skorek[i], L. Tiborcz[i], Y. Hirose[j], T. Takeda[j], H. Nakamura[j], C. Choi[k], J. Heo[k], A. Petruzzi[l], K. Zeng[l], Z. Xie[m], X. Wu[m], H. Eguchi[n], F. Pangukir[o], P. Breijder[o], S. Franssen[o], G. Perret[p], I.D. Clifford[p], T.M. Coscia[q], F. Di Maio[q], E. Zio[q,r], N. Pedroni[s], J. Zhang[t], J. Freixa[u], F. Rizzo[v], C. Ciurluini[v], F. Giannetti[v], M. Adorni[w]



2.2. BE simulation and validation

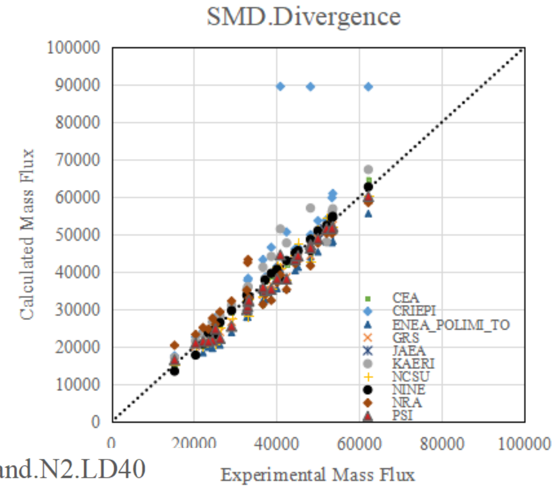
Simulation codes

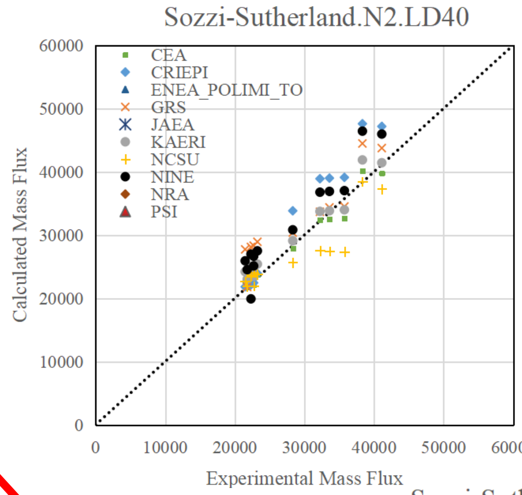


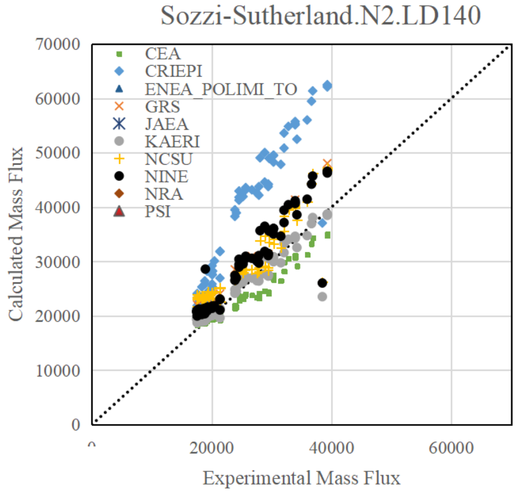
100% participants use the critical mass flow rate as QoI

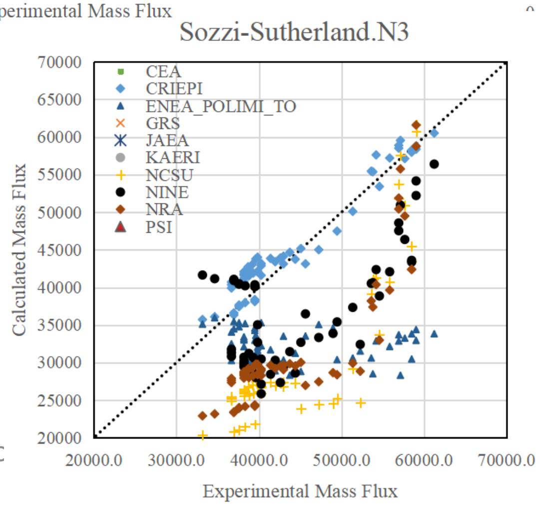
Results from participants

- Good agreement in Super Moby Dick experiments by most of the participants
- Decent results in Sozzi with long nozzles
- Large spread of results in Sozzi with short nozzles
- In RELAP: Henry Fauske model performs better than Ransom-Trapp model at short nozzle, while Ransom-Trapp model performs better at long nozzle









2.3. Inverse Uncertainty Quantification

IUQ and Validation Database

(black)	completely used
(grey)	partially used
(white)	not used

IUQ database	CEA	CRIEPI	ENEA - PoliMi - PoliTo	GRS	KAERI	NCSU	NINE	NRA	PSI	UPC	N° participants
Sozzi N2 L/D = 0				partially used	completely used						2
Sozzi N2 L/D = 1		completely used		partially used	completely used						3
Sozzi N2 L/D = 3	completely used			partially used	completely used		partially used			completely used	5
Sozzi N2 L/D = 5	completely used	completely used	partially used	partially used	completely used		partially used			completely used	7
Sozzi N2 L/D = 9	completely used	completely used	partially used	partially used	completely used		partially used		partially used	completely used	8
Sozzi N2 L/D = 15	partially used	completely used	partially used		completely used		partially used		partially used	completely used	7
Sozzi N2 L/D = 18	partially used	completely used	partially used	partially used	completely used		partially used		partially used	completely used	8
Sozzi N2 L/D = 25	completely used	completely used	partially used	partially used	completely used		partially used		partially used	completely used	8
Sozzi N2 L/D = 40	completely used	completely used		partially used	completely used		partially used			completely used	6
Sozzi N2 L/D = 50	completely used	completely used		partially used	completely used		partially used	completely used		completely used	7
Sozzi N2 L/D = 140	completely used	completely used		partially used	completely used		partially used			completely used	6
Sozzi N3		completely used									1
Sozzi N4				partially used	completely used						2
SMD Div.	partially used	completely used	partially used	completely used	completely used	completely used	completely used		partially used	completely used	9
SMD Exp.		completely used	partially used			completely used			partially used	completely used	5

Experiments used for IUQ and validation ≠ final step of the adequacy

Most popular reasons:

- Simulations were not considered best-estimate (or not enough)
- Reducing time of simulation

No reason given by some participants

Nobody uses Marviken to assess IUQ

Short nozzles usually discarded

SMD and Sozzi N2 L/D > 3 are the most used

1 participant ~ 1 list of experiments

Input uncertain factors

TRACE	
ENEA – PoliMi - PoliTo	• CHM12 (subcooled flow rate) • CHM22 (two-phase flow rate) • Relaxation Constant (SUB) • Relaxation Constant (MP)
NRA	• CHM12 (subcooled flow rate) • CHM22 (two-phase flow rate)
PSI	• CHM12 (subcooled flow rate) • CHM22 (two-phase flow rate)
NCSU	• CHM12 (subcooled flow rate) • CHM22 (two-phase flow rate)

RELAP	
CRIEPI	• Wall roughness • Discharge coefficient (subcooled) • Discharge coefficient (two-phase) • Discharge coefficient for Henry-Fauske • Thermal non-equilibrium term for Henry-Fauske • Energy loss coefficient
NINE	• Wall roughness • Discharge coefficient (subcooled) • Discharge coefficient (two-phase)
UPC	• Discharge coefficient for Henry-Fauske • Thermal non-equilibrium term for Henry-Fauske • Wall roughness

ATHLET	
GRS	Critical mass flow rate

CATHARE	
CEA	• Flashing • Wall-to-liquid friction factor

SPACE	
KAERI	12 input parameters: Inlet pressure, temperature, and void fraction, discharge coefficients for liquid, two-phase, and vapor phase, non-equilibrium factor in Henry-Fauske, heat transfer coefficients of bubbly, slug, bubbly-annular, bubbly-slug, and flashing

IUQ methodology: Modeling of the uncertainty

Uncertainty Modelling of Input Factors + Technical estimation

Statistically reducible or statistically non-reducible?

One-Level Parameter Uncertainty (Standard approach)

Or

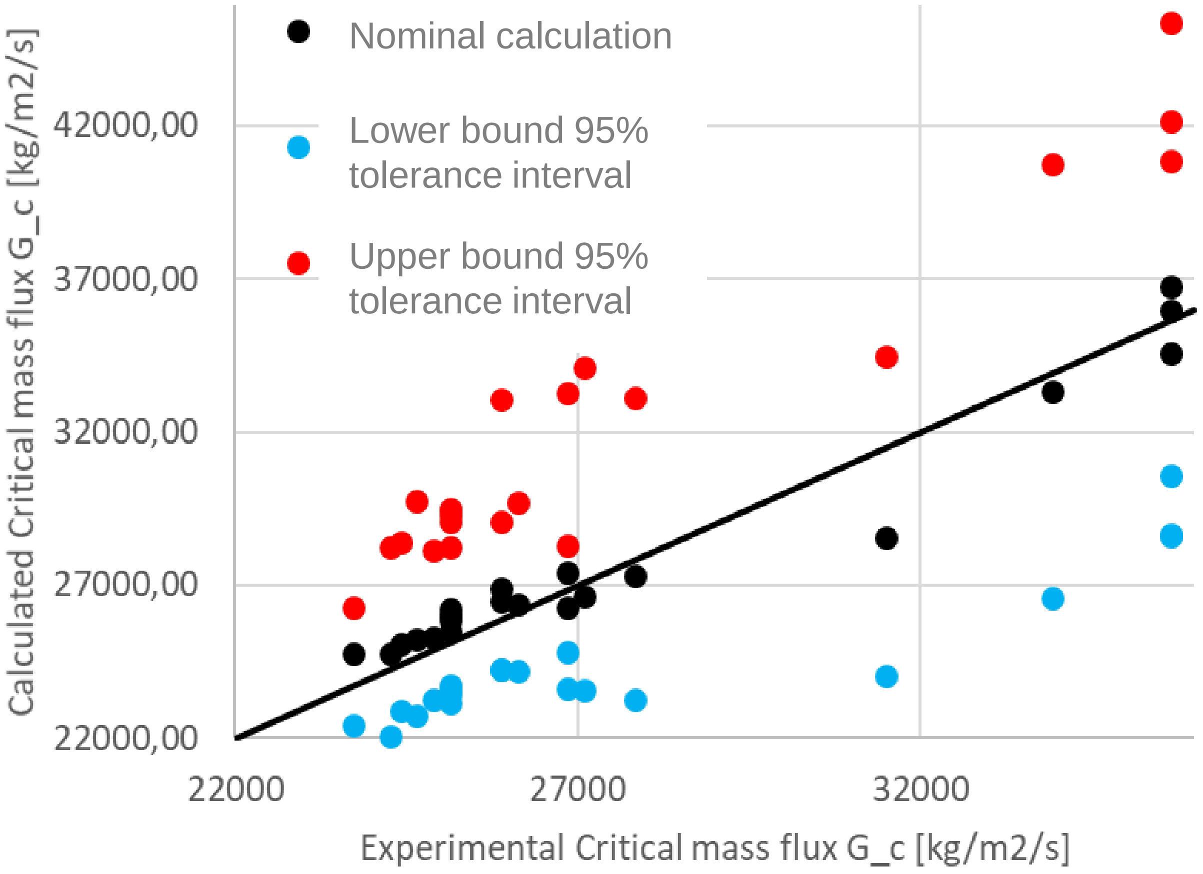
Hierarchical Uncertainty (Two-Level approach)

DoE, Bayesian, Frequentist, Data Assimilation

CEA used CIRCE, CIRCE 2-Steps and hierarchical Bayesian method (Bayesian CIRCE)

Confirmation/validation of the IUQ

Envelope calculations: cover the experimental value through the propagated uncertainty interval



Confirmed/Validated

The calculated value and the experimental value are in between the interval

IUQ methodology: Confirmation/validation

- Quantitative indicator: we checked that 95% of the experimental points are enveloped by the 95-95 tolerance interval

<table>
<tr><th></th><th>Hierarchical</th><th>One-level</th></tr>
<tr><td>DoE</td><td></td><td></td></tr>
<tr><td>Frequentist</td><td>CEA; GRS</td><td></td></tr>
<tr><td rowspan="2">Bayesian</td><td>UPC, NCSU2</td><td rowspan="2">ENEA; NRA; NCSU1</td></tr>
<tr><td>PSI</td></tr>
<tr><td rowspan="2">Data Assimilation</td><td rowspan="2"></td><td>NINE</td></tr>
<tr><td>KAERI</td></tr>
</table>

	confirmed
	partial confirmed
	not confirmed

- Most important result: **<u>Only hierarchical approach can confirm</u>** → To be used in the IUQ framework!
- Currently working on an article to explain this important point.

2.4. CEA's results

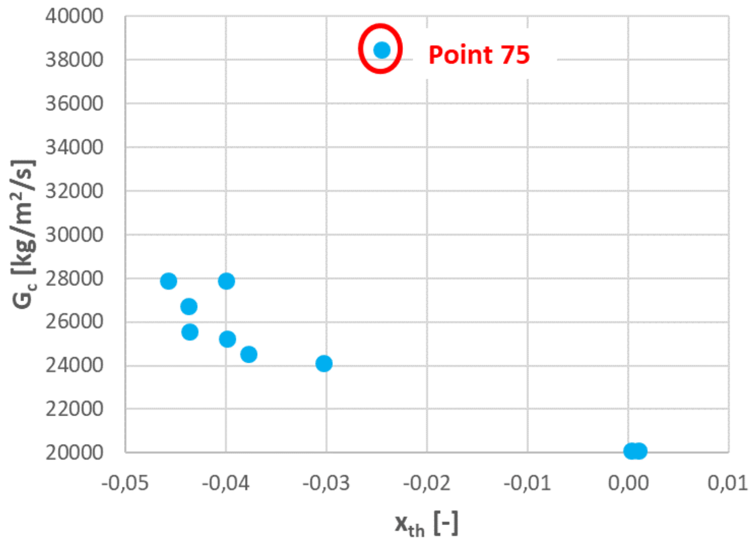
Main conclusions of the adequacy analysis

- Adequacy interpreted as a preliminary selection of experiments → possible modifications of the IUQ database after experimental analysis and simulation

- Adequate experimental databases:
  - Quantification:
    - Super Moby Dick (SMD) with divergent
    - Sozzi-Sutherland N2
    - Sozzi-Sutherland N4
  - Validation:
    - Marviken 13, 17 and 24 (less representative)

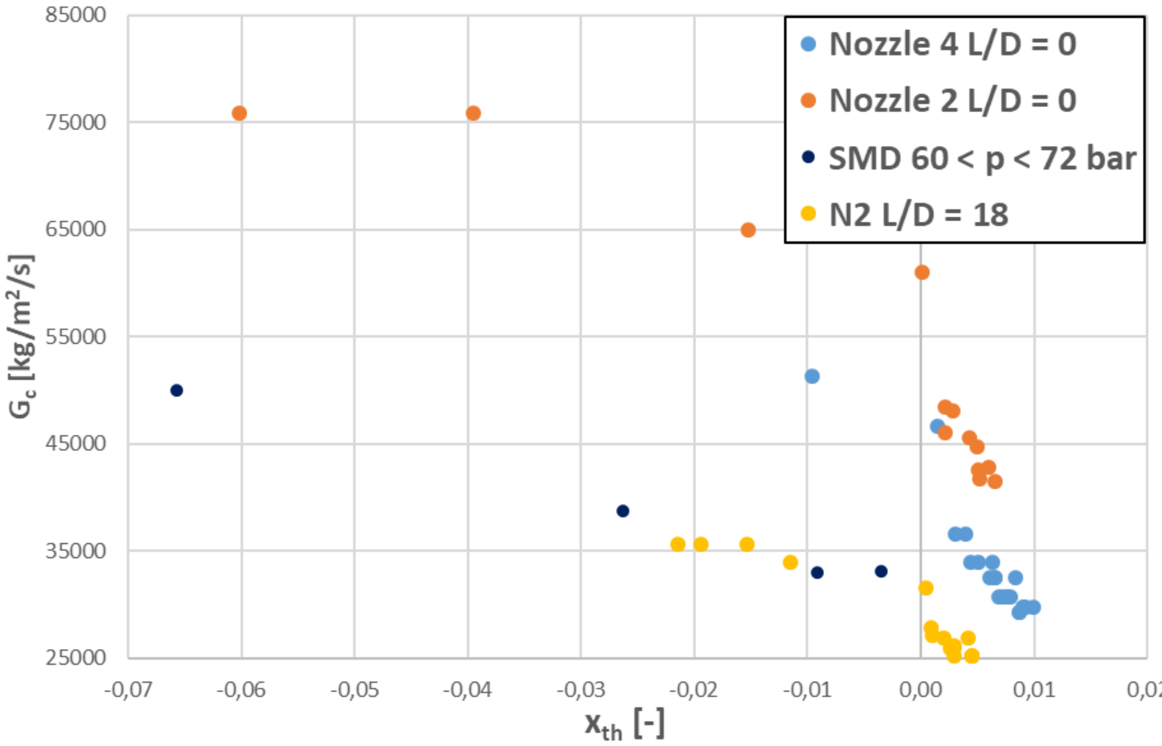
Experimental analysis

- Plot of the experimental data to identify outliers and study dependencies
- Repeated tests: Points 66 and 67 of Sozzi N2 *L*/*D* = 140
- 1 outlier identified: point 75 of Sozzi N2 *L*/*D* = 140



Experimental analysis

- For short nozzles (*L*/*D* < 3), the critical mass flux is highly sensitive to:
  - Throat diameter
  - Nozzle form loss
- Difficult to model in CATHARE
- Eliminate all the tests with *L*/*D* < 3
- Already ranked lowest by the adequacy
- 323 experimental points retained for IUQ:
  - SMD
  - Sozzi *L*/*D* ≥ 3

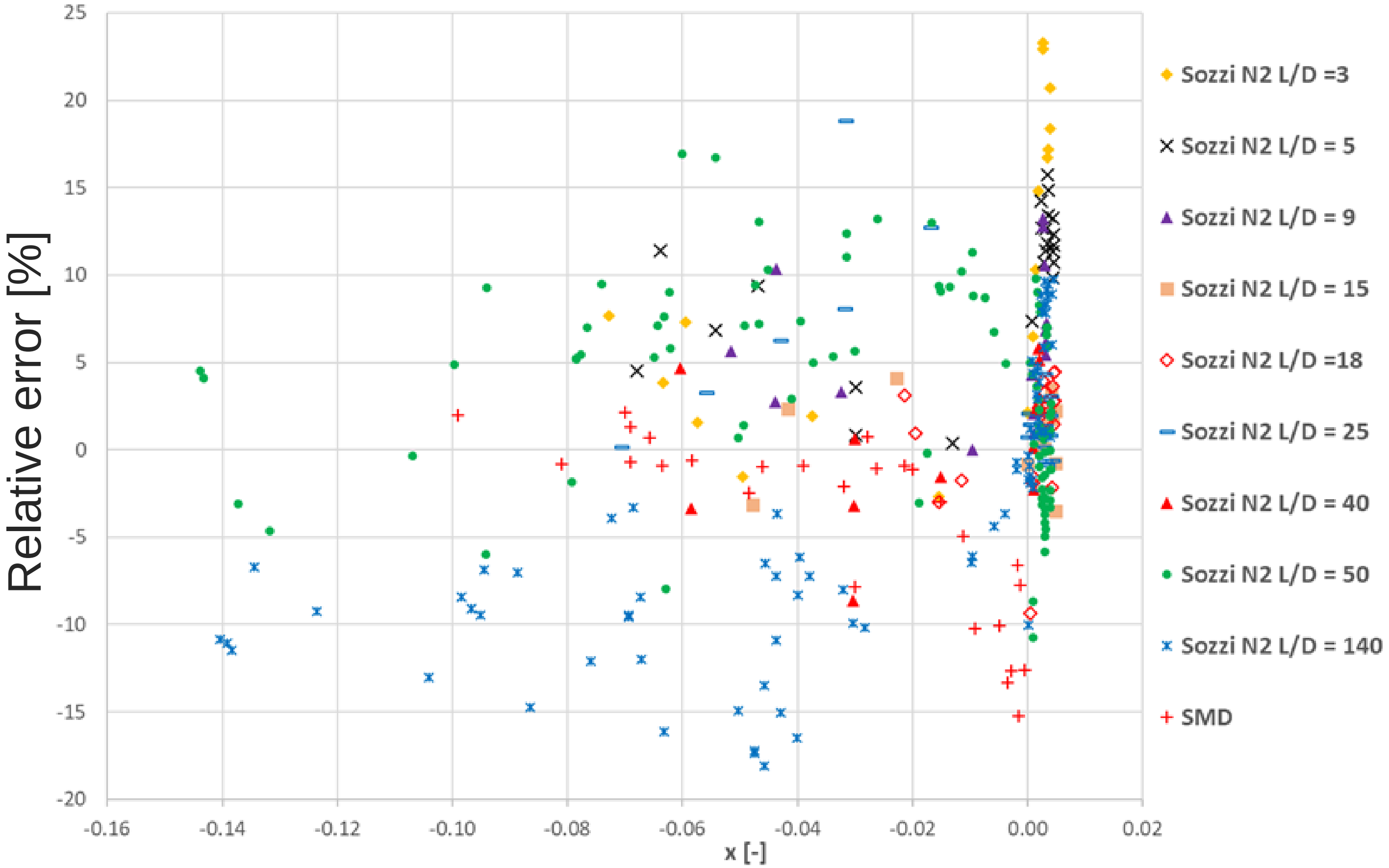


CATHARE simulation

14th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics, Operation and Safety (NUTHOS-14)
Vancouver, BC, Canada, August 25 – 28, 2024

Validation of the system code CATHARE3 on critical flow experiments in the framework of the OECD-NEA ATRIUM project

Alberto Ghione, Lucia Sargentini, Guillaume Damblin, Philippe Fillion



All tests [323 tests]	
Mean error [%]	1.6
σ [%]	7.5
Min error [%]	-18.1
Max error [%]	23.3

- Good agreement were obtained between CATHARE and experiments (L/D ≥ 3)
- CATHARE ➔ only code without a 0D correlation for critical flow

CIRCE Methodology: Brief description

$$M \rightarrow \lambda\, M$$

$$\lambda \sim N(m, \sigma^2) \qquad \lambda \sim \mathcal{L}N(m, \sigma^2)$$

- Estimation of the (log-)normal distribution of the uncertain model parameters
- Discrepancies between experimental and calculated values of the QoI are modelled as:

$$\epsilon_i = z_i - Y(\lambda_i, \mathbf{x}_i), \qquad 1 \leq i \leq n$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\epsilon_i})$$

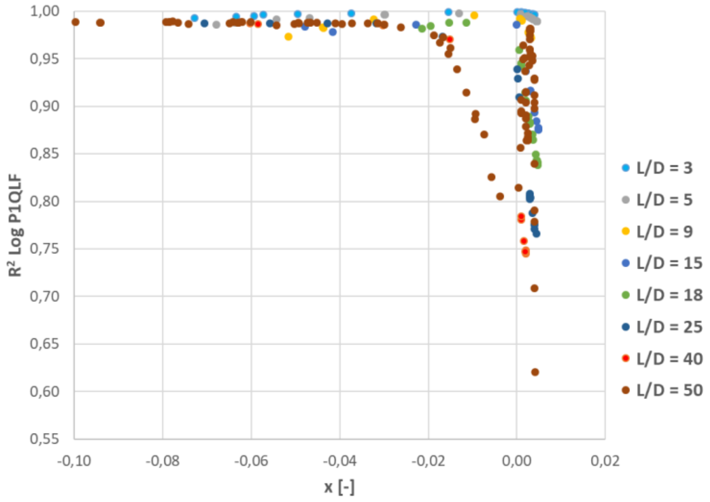
Conservative hypothesis

- Hierarchical modelling of input factors + Frequentist IUQ estimation (MLE)

- Hyp: (log-)linearity of code output to input variation + normality of standardized residuals

IUQ preparatory work

- A priori sensitivity analysis:
  - Linear response of the code (sum of Pearson squared ~ 1) → Same results with Sobol
  - Flashing model (P1QLF) is dominant for shorter nozzles (L/D < 18)
  - Wall-to-liquid friction (SP1CL) is dominant for long nozzles (L/D = 140)
- Study of the log-linearity of the response:
  - Always respected for SP1CL ($R^2 > 0.98$)
  - Some deviations for P1QLF around saturation for Sozzi *L*/*D* > 15
  - 48 tests with no log-linearity → moved to Validation database



- Remaining Quantification database is split in:
  - CIRCE for P1QLF → 68 tests (Sozzi L/D < 18 and SMD) with log-linearity OK
  - CIRCE 2-Steps for SP1CL → 207 tests (Sozzi L/D > 25)

IUQ results

- Standard CIRCE for Flashing → $\mathrm{P1QLF} \sim \mathcal{LN}(0.342, 0.246)$

m	σ^2	Median	F.I. 95%	p-value K-S	NEC
0.342	0.246	1.408	[0.53, 3.72]	0.294	< 0.17

- CIRCE 2-Steps for Wall-to-liquid friction → $\mathrm{SP1CL} \sim \mathcal{LN}(-0.014, 0.117)$

m	σ^2	Median	F.I. 95%	p-value K-S	NEC
-0.014	0.117	0.986	[0.50, 1.93]	0.281	< 0.11

- Kolmogorov-Smirnov test (p-value > 0.05 → normality hypothesis OK)
- Smaller NECs = better statistical identifiability of CIRCE estimates

- Other approaches tested (including joint CIRCE and joint Bayesian CIRCE)

→ Results remained consistent

Confirmation/validation of the obtained uncertainties

- Envelop calculations → For each test in the Quantification/Validation database:
  - Forward propagation with 93 CATHARE simulations (Wilks order 2)
  - Simultaneous variation of SP1CL and P1QLF based on their *pdf*s
  - Estimation of the 95%-95% two-sided confidence intervals (i.e. min and max calculated values) of the QoI

- Confirmation:
  - 3 tests over 275 not enveloped
  - 98,5% of the experimental points are enveloped by the confidence intervals
  - Percentage > 95% → Confirmation OK !

- Validation: 100% envelop for the validation database → OK!

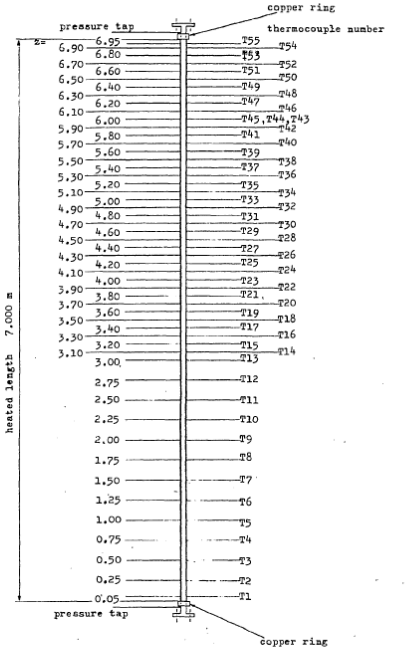
Conclusions from CEA

- All the SAPIUM steps (and their justification) are crucial for a successful IUQ:
  - Adequacy and experimental analysis (one of the most important steps)
    - Understanding of the physics, experimental facilities and errors
    - Careful selection of experimental data
  - Simulation
    - Correct modelling of the facility
    - Definition of Best-Estimate? – Which validation metrics? (open question)
  - IUQ
    - Understanding/verification of hypotheses of the IUQ method
    - Confirmation seems trivial but very important to verify
  - Validation
    - Better understanding of uncertainty extrapolability and justifiability

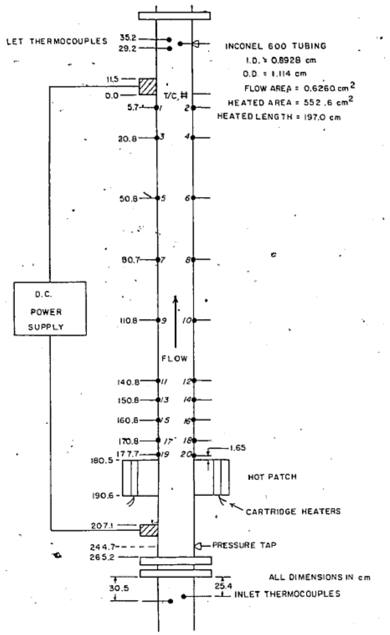
3. 2^{nd} exercise on post-critical heat flux: main results

2nd exercise: Post-CHF heat transfer experiments

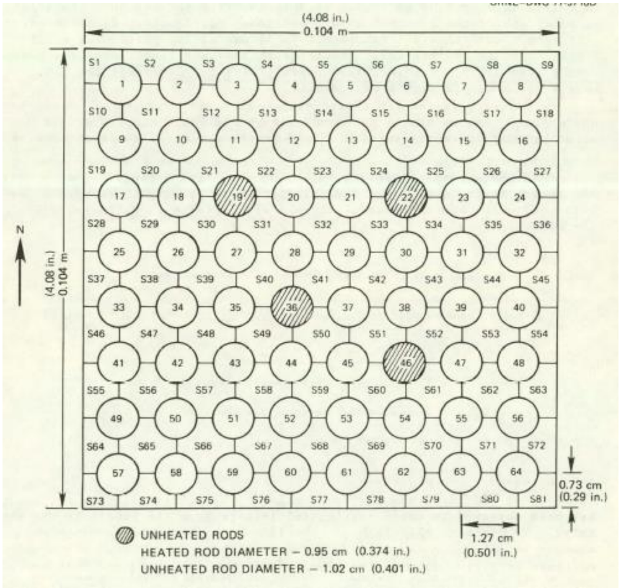
Becker



Stewart



THTF



*outlet estimated

	Type Section	Number of tests	Pressure [bar]	G [kg/m²s]	Inlet subcooling [°C]	Heat flux [kW/m²]
Becker	1-tube	420	30-200	500-3000	5-10	290 – 1250
Stewart	1-tube	312	20-90	115-2833	9-56	64-460 (°)
THTF Film Boiling	Rod-Bundle	22	40-130	226-806	8-46	320-940
THTF Uncovered Bundle	Rod-Bundle	6	40-75	3-30*	46-103	74-480
LSTF	Rod-bundle	1	20-50	0-600	~0	500-2000

Summary of the adequacy results

2 main groups *(UPC provides equal representativeness to all databases)* :

- « THTF the least representative »: CEA, CRIEPI, NCSU, NRG

- « THTF the most representative »: EDF, ASNR, KAERI

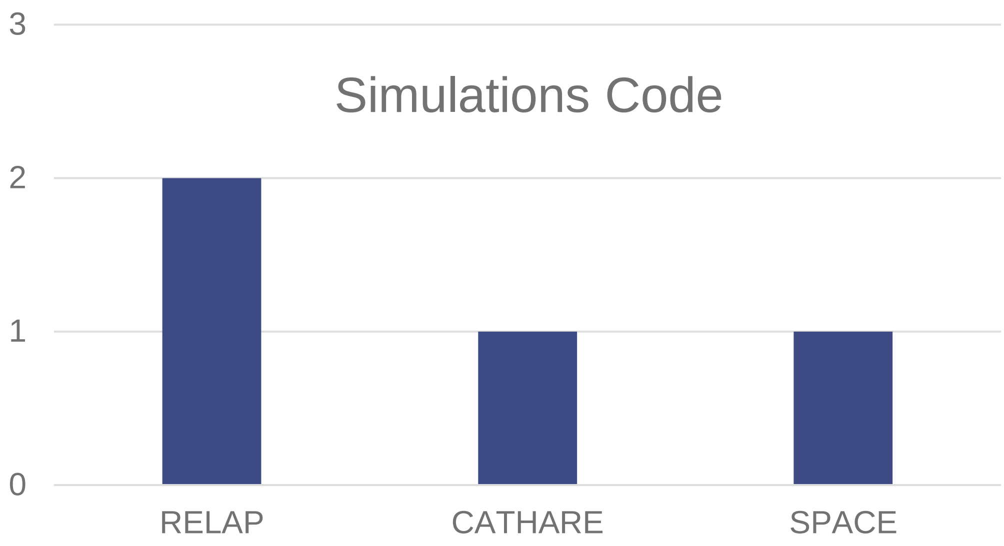
Disagreement between code "developers" and "users" !

Most significant disagreement between participants for the criteria of:
- "Separability of the effects"
- "Capability of the simulation tool to reproduce the experimental data"

CRIEPI: "*The thought process induced by the adequacy and the associated discussions may be more important than the absolute truth (that may not exist)*"

Phenomena more complex, less consensus

Assessment of simulations



Every participant used a 1D modelling

Every participant has different strategy: regulation of CHF and/or Leidenfrost temperature, optimization of both occurrence of CHF and CHF height

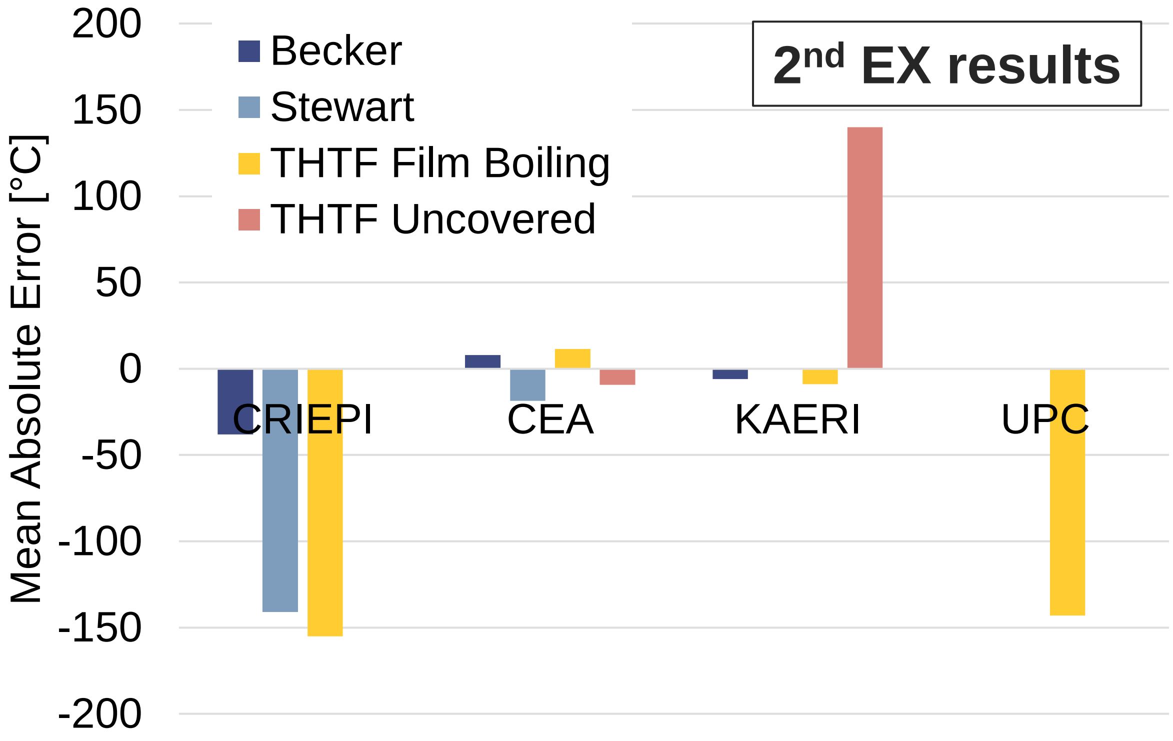
Expert analysis is crucial !

Simulated experiments

	Becker 1	Becker 2	Becker 3	Stewart	THTF FB	THTF Uncover
CEA	Y	Y	Y	Y	Y	Y
CRIEPI	Y	Y	N	Y	Y	N
KAERI	Y	Y	Y	N	Y	Y
UPC	N	N	N	N	Y	N

Key result

- Every participant :
  - Selects different databases from adequacy
  - Uses different system codes
  - Uses different modelling strategies
- This leads to different results !
- **Experience and Expert analysis are crucial!**
- However, the guideline helps to reduce expert judgement and force the expert to justify his/her choices



IUQ methodology: Confirmation/validation

- As for the first exercise → <u>**Only hierarchical approach can confirm**</u>

	Hierarchical	One-level
DoE		
Frequentist	CEA	
Bayesian	UPC	CRIEPI
Data Assimilation		KAERI

green	confirmed
yellow	partial confirmed
red	not confirmed

4. Conclusions

Conclusions

- 4-years OECD-NEA ATRIUM project ➔ improve the IUQ and BEPU methodology for safety analysis
- New insights on IUQ methodologies:
  - Systematic approach for adequacy analysis;
  - Importance of using a hierarchical approach for IUQ;
  - Ongoing discussion on validation metrics (scalar vs time-dependent data; BE vs uncertainty bands)
- Consensus between participants: expert judgement remains important but the guideline helps to reduce it and to have the good approach for IUQ problem
- Great interest and dynamism from nuclear IUQ community until the end of the project:
  - 1st exercise: 16 contributions
  - 2nd exercise: 10 contributions
  - 3rd exercise: expected 13 contributions, so far

Thanks to all contributors, task leaders, Martina Adorni and OECD-NEA !

And for your attention!

CEA SACLAY
91191 Gif-sur-Yvette Cedex
France
alberto.ghione@cea.fr
lucia.sargentini@cea.fr
guillaume.damblin@cea.fr