Nuclear Engineering and Design 419 (2024) 112971

Contents lists available at ScienceDirect

Nuclear Engineering and Design

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

Data Adequacy by an Extended Analytic Hierarchy Process for Inverse Uncertainty Quantification in Nuclear Safety Analysis

Francesco Di Maio [a,*], Thomas Matteo Coscia [a], Enrico Zio [a,b]

[a] Energy Department, Politecnico di Milano, Via La Masa 34, Milano 20156, Italy
[b] Centre for Research on Risk and Crises (CRC), MINES ParisTech, PSL Research University, 06904 Sophia Antipolis, France

ARTICLE INFO

Keywords:
Best Estimate Plus Uncertainty (BEPU)
Data Adequacy (DA)
Inverse Uncertainty Quantification (IUQ)
Multi-Criteria Decision Making (MCDM)
Analytic Hierarchy Process (AHP)
Monte Carlo Simulation (MCS)

ABSTRACT

Data Adequacy (DA) assessment of experimental databases must be performed to control the impact of user effects on the results provided by the Thermal-Hydraulic (T-H) codes employed for the safety assessment of Nuclear Power Plants (NPPs). The activity is typically based on expert judgement, which, however, lacks a rigorous treatment of the uncertainties. With the objective to overcome this limitation, we propose a Multi-Criteria Decision Making (MCDM) approach to consider the Representativeness (R) and Completeness (C) of the databases by an Analytic Hierarchy Process (AHP) combined with Interval Analysis (IA) and Monte Carlo Simulation (MCS) to quantify the uncertainty. The approach for DA is exemplified on the databases made available to the participants of the ATRIUM (Application Tests for Realization of Inverse Uncertainty quantification and validation Methodologies in thermal hydraulics) project promoted by the WGAMA of the OECD-NEA, whose ultimate objective is the systematic application of Inverse Uncertainty Quantification (IUQ) methodologies to assess the uncertainties affecting the T-H model of an Intermediate Break Loss Of Coolant Accident (IBLOCA) of a Light Water Reactor (LWR). The outcomes of the application show that the proposed approach allows overcoming some of the limitations of expert-based approaches, reducing the reliance on subjective evaluations through the incorporation of quantitative metrics in the analysis and via the proper quantification of the uncertainty.

1. Introduction

Nuclear safety analysis relies on the application of analytical and computational models to describe the physical evolution of nuclear systems during various accidental scenarios. Historically, these models have evolved from approaches based on lumped parameters and over-conservative assumptions to the detailed and realistic transient Thermal-Hydraulic (T-H) calculations adopted nowadays, i.e., Best Estimate (BE) T-H codes. The main objective of safety assessment consists in the verification of the safety margins related to some safety-relevant Quantity of Interest (QoI) (e.g., fuel element peak cladding temperature), to demonstrate the capability of Nuclear Power Plants (NPPs) to withstand even the most severe accidental conditions (Marquès et al., 2005). Nevertheless, the computation of the safety margins, requires the identification of the main sources of uncertainty affecting the T-H code responses. Hence, the development of BE models has been extended to the Best Estimate Plus Uncertainty (BEPU) framework, that lays on two fundamental pillars (D'Auria and Galassi, 1998); (D'Auria et al., 2012); (D'Auria et al., 2022) the capability of building an accurate (or as most realistic as possible) representation of the physical phenomena under study through a numerical model, and the ability of quantifying the uncertainties that arise in the model predictions due to the inherent variability of the phenomena (*aleatory uncertainty*) and the limited knowledge on its nature (*epistemic uncertainty*). Such Uncertainty Quantification (UQ) is carried out by forward propagation of the uncertainties on relevant input parameters through the simulation model so as to estimate the uncertainties in the QoI. Thanks to the research activity carried out in the last decades (Nea, 1998; Nea, 2011; Nea, 2016), the (forward) UQ is somewhat a state of practice in the nuclear industry. On the contrary, the characterization and quantification of the uncertainty in the model input parameters still needs attention (IAEA., 2014). Indeed, the starting point of the UQ is the assumption of distributions for the input parameters identified as most relevant by sensitivity analysis (Saltelli et al., 2004). In the current practice, the distributions of the input parameters of a T-H code are generally guessed by expert judgment (Hou et al., 2020); (Bersano et al., 2020); (D'Onorio et al., 2022). To limit the inherent subjectivity, Inverse Uncertainty

* Corresponding author.
E-mail address: francesco.dimaio@polimi.it (F. Di Maio).

https://doi.org/10.1016/j.nucengdes.2024.112971
Received 2 November 2023; Received in revised form 25 January 2024; Accepted 25 January 2024
Available online 5 February 2024

Nomenclature

Acronyms

AHP	Analytic Hierarchy Process
ATRIUM	Application Tests for Realization of Inverse Uncertainty quantification and validation Methodologies in thermal-hydraulics
BE	Best Estimate
BEPU	Best Estimate Plus Uncertainty
C	Completeness
CEA	Commissariat à l'énergie atomique et aux énergies alternatives
CI	Consistency Index
CR	Consistency Ratio
CSNI	Committee on the Safety of Nuclear Installation
DA	Data Adequacy
ENEA	Ente Nazionale Energie Alternative
IA	Interval Analysis
IBLOCA	Intermediate Break Loss Of Coolant Accident
IETs	Integral Effect Tests
IUQ	Inverse Uncertainty Quantification
LWR	Light Water Reactor
Marv	Marviken experiment
MCDM	Multi-Criteria Decision Making
MCS	Monte Carlo Simulation
NEA	Nuclear Energy Agency
NPPs	Nuclear Power Plants
OECD	Organisation for Economic Cooperation and Development
PDF	Probability Density Function
PIRT	Phenomena Identification Ranking Table
PREMIUM	Post-BEMUSE Reflood Model Input Uncertainty Methods
QoI	Quantity of Interest
R	Representativeness
RI	Random Index
SAPIUM	Systematic Approach for Input Uncertainty quantification Methodology
SETs	Separate Effect Tests
SMD	Super Moby Dick experiment
S-S	Sozzi-Sutherland experiment
T-H	Thermal Hydraulics
UQ	Uncertainty Quantification
V&V	Verification and Validation
WGAMA	Working Group on the Analysis and Management of Accidents

Symbols

$\overline{W}$	Eigenvector of AHP procedure
$\overline{x}_n^d$	Vector of experimental decision variables
$\overline{y}^{obs}$	Vector of experimental observed quantities
A_{max}	IA max matrix
A_{mean}	IA mean matrix
A_{min}	IA min matrix
$\overline{x}_n^v$	Vector of chosen experimental decision variables
a_{lk}^m	Entry of comparison matrix
A^m	Generic comparison matrix
λ_{max}	Maximum eigenvalue of AHP procedure
P	Probability (information theory)
N_p	Number of experimental points
N_s	Number of MCS trials
$f(\bullet)$	Probability density function
$\varnothing$	Pipe diameter
m	Index of expert
i	Index of Representativeness criteria
j	Index of Completeness criteria
l	Index of comparison matrix row
k	Index of comparison matrix column
g	Index of MCS trials
n	Index of experimental database
d	Index of experimental decision variables
v	Index of chosen experimental decision variables
M	Number of experts
D	Number of experimental decision variables
V	Number of chosen experimental control variables
$\mathscr{R}$	Representativeness score distributions
$\mathscr{C}$	Completeness score distributions
$\mathscr{DA}$	Adequacy score distributions
$\overline{\mathscr{R}}$	Vector of Representativeness scores
$\overline{\mathscr{C}}$	Vector of Completeness scores
$\overline{\mathscr{DA}}$	Vector of Adequacy scores
I	Number of Representativeness criteria
J	Number of Completeness Criteria
N	Number of experimental databases
$\mu(\bullet)$	Distribution mean
$\sigma(\bullet)$	Distribution standard deviation
ε	Gaussian noise
$min[\bullet]$	Minimum operator
$max[\bullet]$	Maximum operator
ϕ	Quantitative Completeness metric
P_0	Pressure at stagnation point
T_0	Temperature at stagnation point
X_0	Quality at stagnation point
L	Discharge line length
$L/\varnothing$	Nozzle length over diameter ratio
$\dot{\Gamma}$	Steady state critical mass flux
H	Order of square matrix
U	Uniform distribution
T	Triangular distribution
EXP	Experimental database
$C.R$	Representativeness criteria
t	Representativeness sub-criteria index
N_t	Number of Representativeness sub-criteria
$C.C$	Completeness criteria
h	Completeness sub-criteria index
N_h	Number of Completeness sub-criteria
p	Rejected samples index
z	Generic matrix index
Δ	Database extension
χ^2	Chi-square
S	Shannon entropy
O	Observed frequency
E	Expected frequency
u	Chi-square bin index
q	Shannon entropy feature index
N_u	Number of Chi-square bins
Q	Number of Shannon entropy features
s	Generic hierarchy level
δ	Generic s-level element
β	Generic (s-1)-level element
$\mathfrak{C}$	Matrix of Completeness criteria scores
$\mathscr{C}_n^j$	Completeness score for the j-criteria and n-experiment
$\overline{\mathscr{C}}_n^{dis}$	Euclidean distance vector
$\overline{\mathscr{C}}_{(n^*)}$	Vector corresponding to n-row of $\mathfrak{C}$
$\overline{\mathscr{C}}^{opt}$	Completeness optimal point
G_M	Generic group of M experts
$\widehat{\overline{x}}_n^v$	Vector of perturbed chosen experimental decision variables
$\widehat{a}_{lk}$	Sampled entry of comparison matrix

Quantification (IUQ) methods relying on experimental data are being considered (Wu et al., 2021) to find the input distributions that have generated the data (Roma et al., 2021; Roma et al., 2022). Specifically with regards to this, the PREMIUM project (Nea, 2016) is one of the first benchmark activity devoted to the development of IUQ methods. More recently, the SAPIUM project (Baccou et al., 2020; Baccou, 2023), has proposed the following steps of analysis for performing a rigorous IUQ:

1. Experimental databases adequacy assessment
2. T-H model development
3. Input Uncertainty Quantification

In this paper, we present an innovative method for performing the first step above, i.e., the Data Adequacy (DA) of the experimental databases. In the research community concerned with the development of nuclear T-H codes, the problem of DA has been addressed under various perspectives. For example in (D'Auria and Galassi, 1998); (Oberkampf and Trucano, 2007; Oberkampf, et al., 2007); (Oberkampf, et al., 2007; Oberkampf and Trucano, 2007); (Petruzzi and D'Auria, 2008); (Unal et al., 2011); (Nea, 2016), DA is presented as the procedure to be used for Verification and Validation (V&V) of Best Estimate models with respect to experimental data. In (Lin et al., 2020), the attention is, instead, devoted to the adequacy of numerical models of smoothed particle hydrodynamics with respect to previously qualified experimental data. In (Nusret, et al., 1993) and (Mascari et al., 2015), DA is referred to the choice of suitable experimental facilities as source of data for building T-H models considering scaling issues. Actually, even if the specific meaning of the objective of DA varies with the different perspectives, we claim that it is still possible to frame it as a Multi-Criteria Decision Making (MCDM) problem, in which the decision alternatives are the experimental databases (or the numerical models), whose adequacy is evaluated by qualitative/quantitative criteria. Many different approaches exist for solving MCDM problems (see for example (Malczewski and Rinner, 2015) and (Roy, 1996)). Specifically in the nuclear field, the Phenomena Identification and Ranking Table (PIRT) (Wilson and Boyack, 1998) and the Analytical Hierarchy Process (AHP) (Zio et al., 2003); (Yu et al., 2010) have been adopted for incorporating expert judgment in the decision process in a concise and simple way. A quantitative PIRT has been proposed in (Yurko and Buongiorno, 2012), but its application is limited to specific cases for which well-established knowledge regarding the analytical model of the occurring phenomena is available. When it comes to DA, the criteria for the evaluation of the datasets are Completeness (C) and Representativeness (R) (Baccou et al., 2019). Their evaluation is challenged by the need of:

- minimizing the reliance on expert judgment, given the related subjectivity;
- accounting for the uncertainty that arises at various levels of the analysis (e.g., by PIRT or AHP);
- propagating the uncertainty to allow measuring the confidence in the outcomes of the assessment of the experimental dataset for IUQ;

In this paper, we propose a novel approach for DA based on a modified (extended) AHP (Saaty, 1980) that incorporates heterogeneous (i.e., objective and subjective) information to evaluate the C and R criteria by both qualitative and quantitative metrics for the assessment of the T-H DA and the quantification of uncertainty.

The developed approach has been applied to a case study made available to the ATRIUM (Application Tests for Realization of Inverse Uncertainty quantification and validation Methodologies in thermal hydraulics) project participants. The ATRIUM project (Ghione, 2023) has been promoted by OECD-NEA to investigate the applicability of the guidelines developed in the previous SAPIUM project (Baccou et al., 2020) for the IUQ of Intermediate Break Loss Of Coolant Accidents (IBLOCA) of a Light Water Reactor (LWR). The DA is performed on a set of experimental databases related to Separate Effect Test facilities (SETs), i.e., Sozzi-Sutherland (S-S) (Sozzi and Sutherland, 1975), Super Moby Dick (SMD) (Rousseau, 1987), Marviken (Marv) (Sokolowski and Kozlowski, 2012), to identify those most suitable to perform the IUQ of the parameters used to model the IBLOCA in an Integral Effect Test experiment (IETs), i.e., LSTF IB-HL-01 (OECD., 2011). The need of using SETs in substitution of the IET is related to the impossibility of directly performing the IUQ on it, both due to the complexity of the model and the scarcity of data. The QoI investigated by all these SETs is the critical mass flowrate at the break section in steady state conditions and, correspondingly, the objective of this work is focused on the uncertainties affecting this phenomenon in LSTF IB-HL-01. The results obtained with the proposed method are compared with those provided by a traditional AHP, as suggested in (Baccou et al., 2019). It is shown that:

1) To avoid the systematic bias related to the subjective nature of experts opinions (when compiling the comparison matrix of an AHP (see Appendix A for details), typically with the qualitative criteria used for the R assessment), the proposed AHP allows to accounting for the aggregation of experts opinions so as to limit the effects of biased judgements on the results and to allow quantifying the uncertainty related to the outcomes. Indeed, the dispersion of the AHP comparison matrices is quantified by Interval Analysis (IA) and propagated by Monte Carlo Simulation (MCS) to quantify the uncertainty.
2) The metrics proposed to quantitatively evaluate the C have allowed considering the effect of the uncertainties by geometric considerations (i.e., the Euclidean norm).

The structure of the paper is organized as follows: Section 2 illustrates the proposed framework for performing DA; Section 3 presents the case study; Section 4 shows the results of the application of the approach to the case study; also, a comparison of the results with those obtained with a standard AHP is described. Finally, in Section 5 conclusions are drawn.

2. The novel approach for DA

We define the DA problem as a MCDM problem (Baccou, 2023), where N experimental databases EXP_n, $n = 1, 2, \dots, N$, are analysed with respect to the objectives of:

- Representativeness (R): the capability of each single record of the database (i.e., local property) to correctly represent the target physical phenomena (Baccou, 2023).
- Completeness (C): the capability of the whole database (i.e., global property) to cover the physical domain of the target numerical simulation (Baccou, 2023).

Given the specificity of the nuclear safety problem, DA has to deal with the combination of qualitative and quantitative information. AHP (see Appendix A for details) has been chosen for addressing the MCDM problem (Zeng et al., 2017). In Fig. 1, the hierarchical structure of the proposed AHP for DA assessment is sketched. The goal is the Data Adequacy "DA" of the set of N alternative experimental databases "EXP_n", $n = 1, 2, \dots, N$, which are judged in terms of the criteria of Representativeness "R" and Completeness "C", through a set of sub-criteria "$C.R_i$" and "$C.C_j$", with $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$, respectively. Each n-th EXP_n can be seen as an empirical mapping Eq. (1) between a vector of $d = 1, 2, \dots, D$ decision variables $\overline{x}_n^d$ and a vector of observed values of the Quantity of Interest (QoI) $\overline{y}^{obs}$:

$$\overline{x}_n^d \rightarrow \overline{y}^{obs} \tag{1}$$

In general for each n-th EXP_n it is possible to identify a subset of variables $v = 1, 2, \dots, V$ among $\overline{x}_n^d$, where $V < D$, relevant for the

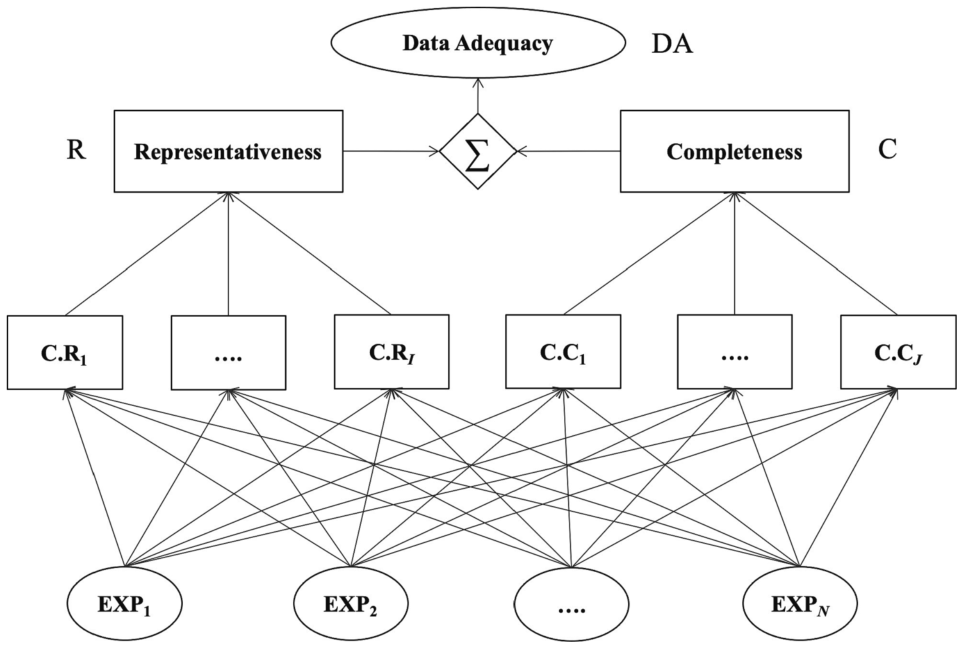


Fig. 1. The AHP hierarchy used to solve the MCDM problem of DA for an arbitrary number N of alternatives (EXP_n) and $I+J$ sub-criteria for R and C, respectively. Adapted from (Baccou et al. 2023).

evaluation of each $C.C_j$, $j = 1, 2, \ldots, J$ (the length V of $\overline{x}_n^v$ can change depending on the j-th $C.C_j$).

To avoid the bias that might be introduced by a single expert filling the comparison matrices of the AHP in Fig 1:

1) The subjective sub-criteria $C.R_i$, $i = 1, 2, \ldots, I$, are evaluated by M different experts and, then, aggregated as explained in Section 2.1, to represent the uncertainty related to the expert-based judgments.
2) The quantitative sub-criteria $C.C_j$, $j = 1, 2, \ldots, J$, described in Section 2.2, are used in place of the expert-based evaluations and, then, aggregated with a geometric method based on the Euclidean norm (presented in Section 2.2.1).
3) Finally, in Section 2.3, R and C are aggregated in one single DA measure.

2.1. Representativeness assessment

Let us assume that the m-th expert, $m = 1, 2, \ldots, M$, contributes to the assessment of R with the comparison matrix $A^m = \left[a_{lk}^m\right]_{H\times H}$ where a_{lk}^m ($a_{kl} = 1/a_{lk}$ $\forall l,k|l \neq k$ and $a_{lk} = 1$ for $l = k$) is the pairwise comparison value by the m-th expert, considering the element in the l row and k column of A^m, with $l = 1, 2, \ldots, H$ and $k = 1, 2, \ldots, H$, respectively. The same procedure is iterated for each level of the AHP hierarchy. In the case of DA, since at least a two-level hierarchy is adopted for R (Fig. 1), at the bottom level are collected I matrices of order $H = N$ for the comparisons of each EXP_n, $n = 1, 2, \ldots, N$ with respect the i-th sub-criterion $C.R_i$, $i = 1, 2, \ldots, I$, whereas in the upper level a single matrix of order $H = I$ is collected, containing the values of the relative importance of each sub-criterion $C.R_i$, $i = 1, 2, \ldots, I$, for the final R value. Once all the matrices are collected and tested for consistency (Saaty, 1980), the combination of their dominant eigenvectors (through matrix multiplication) is used to assess the final R ranking (Saaty, 2003). The spread on the pairwise comparison values a_{lk}^m expressed by the M alternative experts is used to quantify the uncertainty in the assessment of R (Sajjad Zahir, 1991) and to avoid inconsistencies, as shown in the following Section, by aggregating the multiple expert evaluations using IA and MCS.

2.1.1. Interval analysis (IA)

When $M \longrightarrow 1$ the arithmetic mean A_{mean} Eq. (2) can be used as a central estimator of the variability on the pairwise comparison values a_{lk}^m (Ramanathan and Ganesh, 1994); (Saaty and Vargas, 2012):

$$A_{mean} = \begin{bmatrix} 1 & \frac{\sum_{m=1}^{M} a_{12}^m}{M} & \cdots & \frac{\sum_{m=1}^{M} a_{1H}^m}{M} \\ \frac{M}{\sum_{m=1}^{M} a_{21}^m} & 1 & \frac{\sum_{m=1}^{M} a_{2k}^m}{M} & \vdots \\ \vdots & \frac{M}{\sum_{m=1}^{M} a_{l2}^m} & \ddots & \frac{\sum_{m=1}^{M} a_{lH}^m}{M} \\ \frac{M}{\sum_{m=1}^{M} a_{H1}^m} & \cdots & \frac{M}{\sum_{m=1}^{M} a_{Hk}^m} & 1 \end{bmatrix} \tag{2}$$

In these circumstances, the minimum and maximum operators can be used to estimate the intervals of a_{lk}^m i.e., the lower A_{min} and upper A_{max} bounds of the entry a_{lk}^m, as shown in Eq. (3) and Eq. (4), respectively:

$$A_{min} = \begin{bmatrix} 1 & \min_{\forall m} a_{12}^m & \cdots & \min_{\forall m} a_{1H}^m \\ \frac{1}{\min_{\forall m} a_{21}^m} & 1 & \min_{\forall m} a_{2k}^m & \vdots \\ \vdots & \frac{1}{\min_{\forall m} a_{l2}^m} & \ddots & \min_{\forall m} a_{lH}^m \\ \frac{1}{\min_{\forall m} a_{H1}^m} & \cdots & \frac{1}{\min_{\forall m} a_{Hk}^m} & 1 \end{bmatrix} \tag{3}$$

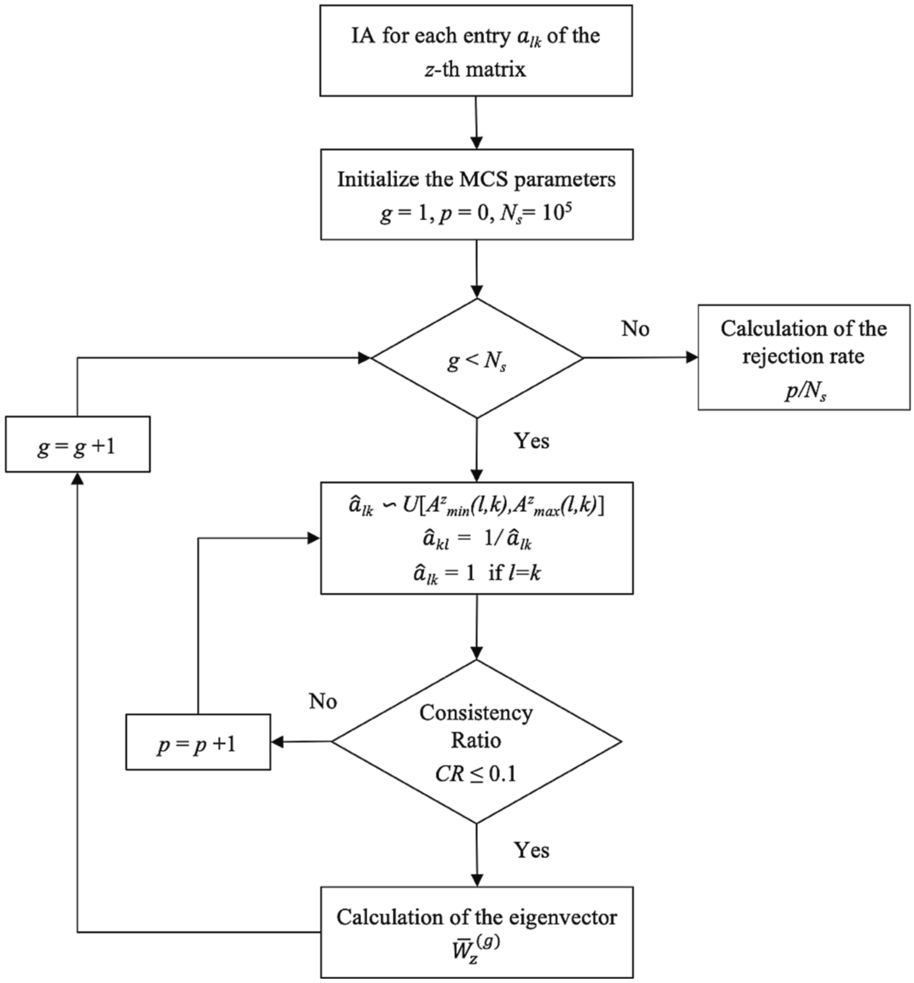


Fig. 2. MCS flowchart for the AHP framework.

$$A_{max} = \begin{bmatrix} 1 & \max\limits_{\forall m} a_{12}^m & \cdots & \max\limits_{\forall m} a_{1H}^m \\ \dfrac{1}{\max\limits_{\forall m} a_{21}^m} & 1 & \max\limits_{\forall m} a_{2k}^m & \vdots \\ \vdots & \dfrac{1}{\max\limits_{\forall m} a_{l2}^m} & \ddots & \max\limits_{\forall m} a_{lH}^m \\ \dfrac{1}{\max\limits_{\forall m} a_{H1}^m} & \cdots & \dfrac{1}{\max\limits_{\forall m} a_{Hk}^m} & 1 \end{bmatrix} \tag{4}$$

It is worth mentioning that the comparison matrices A_{mean}, A_{min}, A_{max}, are not required to satisfy the multiplicative reciprocity condition of the comparison matrix A^m i.e., $a_{kl} = 1/a_{lk}$ $\forall l, k | l \neq k$ (Fedrizzi et al., 2020).

2.1.2. Monte Carlo simulation (MCS)

MCS is used to propagate, through the AHP model, the Probability Density Function (PDF) $f(a_{lk})$ of the preference scores that arise from the corresponding entries of A_{min} and A_{max}, i.e., the min and max bounds of each pairwise preference a_{lk} (Eq. (3) and Eq. (4)), as suggested in (Cagno et al., 2000); (Cagno et al., 2001), for each generic z-th matrix of the hierarchy, $z = 1, 2, \ldots, I+1$. At each g-th simulation, $g = 1, 2, \ldots, N_s$, a comparison matrix of random samples $\hat{a}_{lk} \sim U\left[A_{min}^z(l,k), A_{max}^z(l,k)\right]$ for $l < k$ is compiled and checked for consistency (Rosenbloom, 1997). If the consistency test is passed, the z-th matrix is accepted, otherwise the algorithm rejects it and resample its entries. The accepted $I+1$ matrices are solved following the standard AHP procedure (Saaty and Vargas, 2012): thus, the principal eigenvectors $\overline{W}_z^{(g)}$, $z = 1, 2, \ldots, I+1$, are calculated and properly combined in a vector $\overline{\mathscr{R}}_n^{(g)}$ of size $(N \times 1)$ (Appendix A), to rank the N EXP_n, $n = 1, 2, \ldots, N$, for each g-th MCS trial. The resulting empirical distributions of the R scores $\mathscr{R} = \left[\overline{\mathscr{R}}_n^{(g)}\right]_{N \times N_s}$, $n = 1, 2, \ldots, N$ and $g = 1, 2, \ldots, N_s$, are collected by as customary done by MCS for a forward uncertainty propagation procedure, repeating a random sampling of the entries of the comparison matrix until the consistency check is passed, and then, by calculating the corresponding eigenvectors, that are propagated through the AHP. In Fig 2, a scheme flowchart of the MCS adopted is sketched for a single matrix z, where p is the MCS index variable of sampled matrices that have been rejected.

2.2. Completeness assessment

For each sub-criteria $\mathrm{C.C}_j$, $j = 1, 2, \ldots, J$, we introduce a quantitative metric $\phi_j : \mathbb{R}^V \longrightarrow \mathbb{R}_*^+$, where $\mathbb{R}_*^+ = \mathbb{R} \backslash \{(-\infty, 0]\}$, that maps the experimental variables $\overline{x}_n^v$, $v = 1, 2, \ldots, V$, related to the n-th experiment EXP_n into its C score $\mathscr{C}_n^j$:

$$\mathscr{C}_n^j = \phi_j\left(\widehat{\overline{x}}_n^v\right) \tag{5}$$

$$\widehat{\overline{x}}_n^v = \overline{x}_n^v + \overline{\varepsilon}_n^v \tag{6}$$

Eq. (6) gives the experimental variables vector $\widehat{\overline{x}}_n^v$, whose values are affected by experimental measurement uncertainty, modelled as gaussian noise $\overline{\varepsilon}_n^v$. A final score $\mathscr{C}_n^j$, that represent the performance of the n-th experimental database with respect to the j-th criterion, is

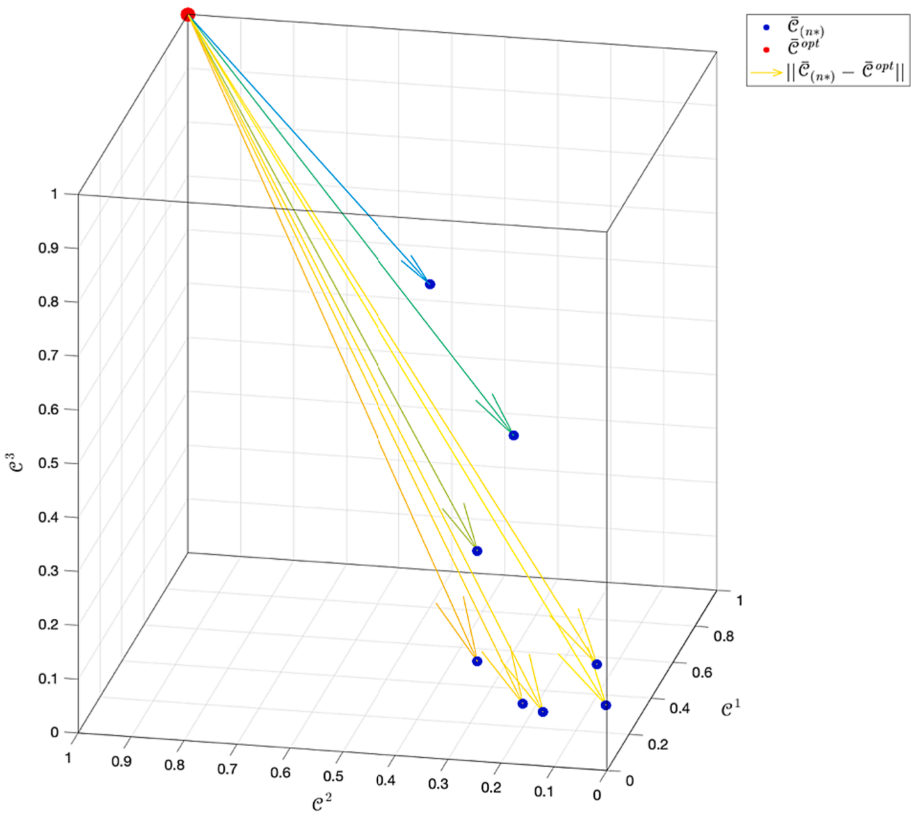


Fig. 3. An heuristic representation in a tridimensional geometrical setting of the C assessment evaluation by means of the proposed Euclidean norm minimization procedure.

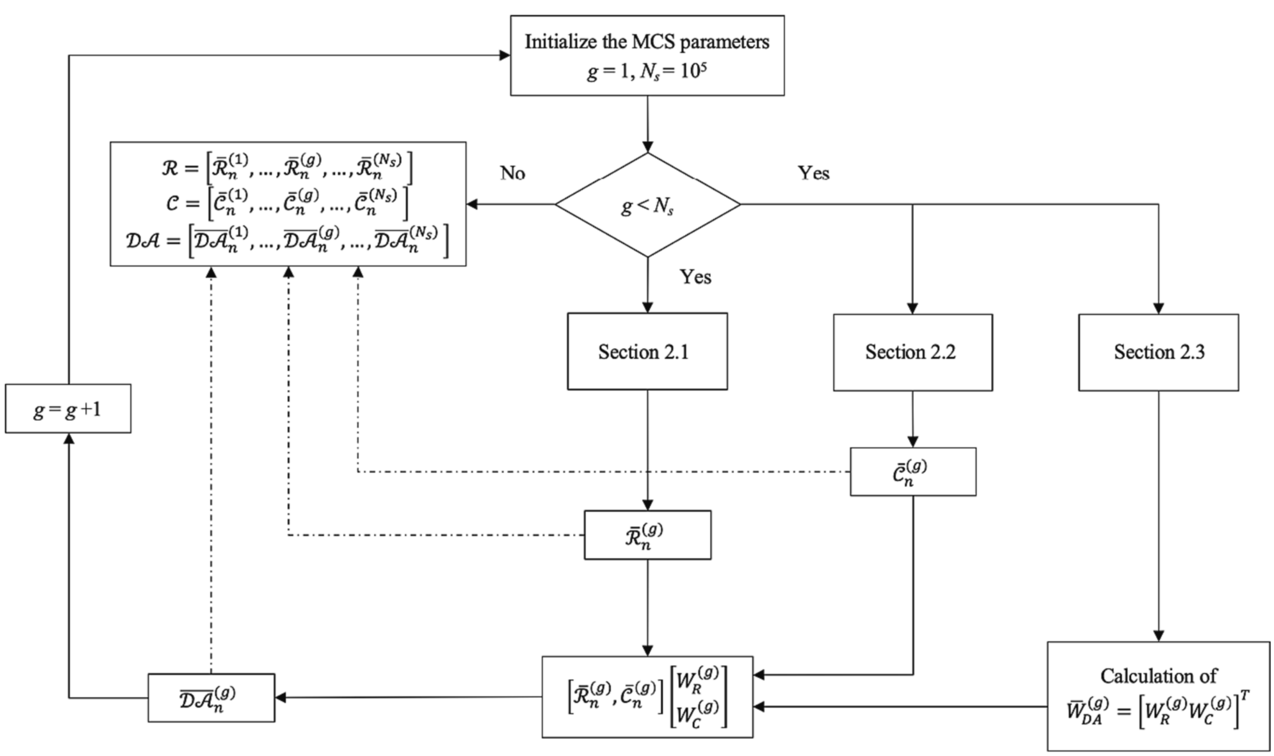


Fig. 4. Flowchart for the evaluation of the DA distributions.

Table 1
Experimental data of IET for exercise 1 of ATRIUM (Baccou et al. 2023; Ghione et al. 2023).

Experiment Name	Label	N_p	$L/\varnothing$ (-)	$\varnothing$ (mm)	P_0 (bar)	T_0 (°C)	X_0 (-)	$\dot{\Gamma}$ (kg/m^{2}s)
LSTF IB-HL-01	LSTF	–	15	–	10–155	15	–	1500–46000

Table 2
Available experimental datasets of SETs for exercise 1 of ATRIUM (Baccou et al. 2023; Ghione et al. 2023).

Experiment Name	Label	N_P	$L/\varnothing$ (-)	$\varnothing$ (mm)	P_0 (bar)	T_0 (°C)	X_0 (-)	$\dot{\Gamma}$ (kg/m^{2}s)
Sozzi Sutherland	S-S N2	358	0–140	12.7	56.0–71.3	232–286	−0.0044–0.0065	17528–75824
	S-S N3	58	0	12.7	42.7–69.0	212–285	−0.0059–0.0060	33161–61226
	S-S N4	23	0	19	56.0–66.3	271–282	−0.0003–0.0099	29295–51266
Super Moby Dick	SMD Div	27	18	20	20–120.1	192.3–324.4	<0	15300–62200
	SMD Exp	12	20	20	20–120.1	191.5–323.6	<0	16100–61800
Marviken	Marv 13	1	3	200	~ 50	ΔT_{sub}	<0	<89200
	Marv 17	1	3,7	300	transient	~ 31 (°C)		<61700
	Marv 24	1	0,33	500				<59750

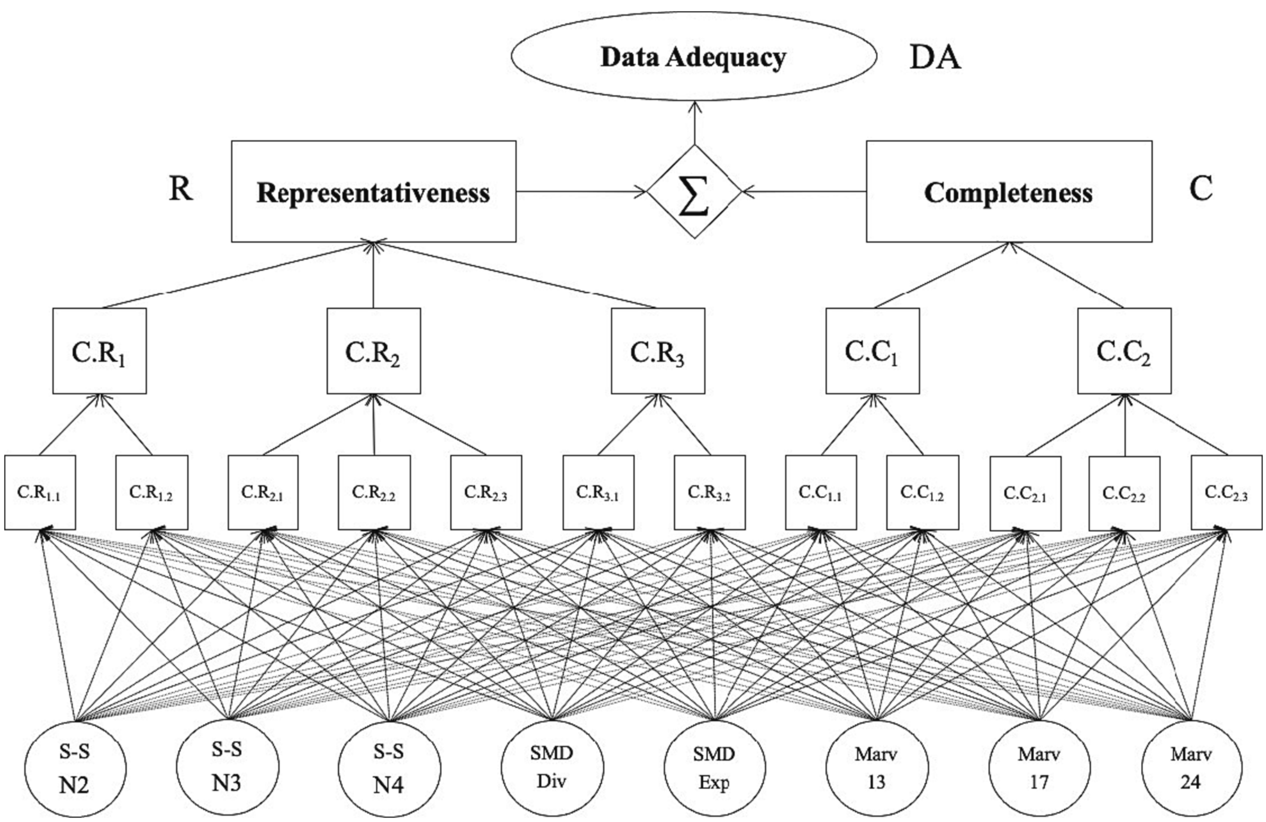


Fig. 5. The employed AHP structure. Adapted from (Baccou et al. 2023).

obtained after a normalization procedure (Zeng et al., 2017):

$$\mathscr{C}_n^j = \frac{\mathscr{C}_n^j}{\sum_{n=1}^{N} \mathscr{C}_n^j} \quad \text{If the scores are sorted in descending order} \tag{7}$$

$$\mathscr{C}_n^j = \frac{\left(\mathscr{C}_n^j\right)^{-1}}{\sum_{n=1}^{N} \left(\mathscr{C}_n^j\right)^{-1}} \quad \text{If the scores are sorted in ascending order} \tag{8}$$

The obtained vector of normalized values $\overline{\mathscr{C}}_n^j = \left[\mathscr{C}_1^j, \cdots, \mathscr{C}_n^j, \cdots, \mathscr{C}_N^j\right]$ ($N \times 1$) carries the information about the C performance of all experimental databases EXP_n with respect to the j-th sub-criterion. By iterating the aforementioned procedure $\forall j = 1, 2, \ldots, J$, the complete evaluation of any n-th EXP_n with respect to all $\mathrm{C.C}_j$ is represented by the matrix $\mathfrak{C} = \left[\mathscr{C}_n^j\right]_{N \times J}$:

$$\mathfrak{C} = \begin{bmatrix} \mathscr{C}_1^1 & \cdots & \mathscr{C}_1^J \\ \vdots & \mathscr{C}_n^j & \vdots \\ \mathscr{C}_N^1 & \cdots & \mathscr{C}_N^J \end{bmatrix} \tag{9}$$

Then, for the aggregation of the single criteria scores collected in $\mathfrak{C}$, in one single value of C for each EXP_n, i.e., the vector $\overline{\mathscr{C}}_n$ ($N \times 1$), assuming that all $\mathrm{C.C}_j$ are equally important, a fully quantitative approach based on the calculation of the Euclidean distance can be adopted. It is worth mentioning that the adoption of the Euclidean distance can lead to distortions in the case that the C criteria are not equally informative; to overcome this limitation, the adoption of outranking methods

Table 3
Criteria for R, [*] suggested by ATRIUM project coordinators (Baccou et al. 2023).

Criteria	Meaning	Sub Criteria	Meaning
C.R$_1$	Fidelity with LSTF for the accidental transient of interest	C.R$_{1.1}$ [*]	Fidelity of the experimental facility geometry with respect to LSTF geometry
		C.R$_{1.2}$ [*]	Fidelity of thermal–hydraulic conditions with respect to LSTF
C.R$_2$	Control of experimental data	C.R$_{2.1}$[*]	Availability of exhaustive documentation
		C.R$_{2.2}$ [*]	Replicability of experimental data
		C.R$_{2.3}$ [*]	Availability of information on experimental uncertainties
C.R$_3$	Modelling of the physical phenomena for their implementation in the system code	C.R$_{3.1}$ [*]	Capability to cover the physical phenomena of interest
		C.R$_{3.2}$ [*]	Separability

(Bouyssou, 2001) or not symmetric distances, e.g., Mahalanobis distance (Brereton and Lloyd, 2016), could be explored.

2.2.1. Completeness assessment (Euclidean)

In the metric space $\mathbb{R}^J$ in which matrix $\mathfrak{C}$ lays, it is convenient to define the optimal point $\overline{\mathscr{C}}^{opt}(1 \times J)$ as the vertex of the hypercube valued one (i.e., $\mathscr{C}_n^j$ scores are normalized in [0, 1]) in all directions i.e., the absolute maximum possible C score value. Intuitively the n-th experiment EXP$_n$ closest to $\overline{\mathscr{C}}^{opt}$ is the best one among all in terms of C score, i.e., it has the lowest Euclidean distance $\overline{\mathscr{C}}_n^{dis}$ $(N \times 1)$ from $\overline{\mathscr{C}}^{opt}$:

$$\overline{\mathscr{C}}_n^{dis} = \left\| \overline{\mathscr{C}}_{(n^*)} - \overline{\mathscr{C}}^{opt} \right\| \tag{10}$$

where $\overline{\mathscr{C}}_{(n^*)}$ is the vector $(1 \times J)$ corresponding to the n-th row of matrix $\mathfrak{C}$ Eq. (9). To obtain a measure of C consistent with the R, $\overline{\mathscr{C}}_n$ scores must be normalized and ranked in descending order. Hence, since the rank provided by $\overline{\mathscr{C}}_n^{dis}$ is in ascending order, the same normalization procedure of Eq. (8) has been applied. Thus, the final C score $\overline{\mathscr{C}}_n$ for any EXP$_n$, $n = 1, 2, \dots, N$, is obtained as:

$$\overline{\mathscr{C}}_n = \frac{\left(\overline{\mathscr{C}}_n^{dis}\right)^{-1}}{\sum_{n=1}^{N}\left(\overline{\mathscr{C}}_n^{dis}\right)^{-1}} \tag{11}$$

It is worth noting that, the Euclidean distance Eq. (10) can be calculated also by setting $\overline{\mathscr{C}}^{opt}(1 \times J)$ as the vertex of the hypercube valued zero in every direction and by employing Eq. (7) for the normalization of the C score $\overline{\mathscr{C}}_n$. The two approaches are equivalent due to the symmetry of the Euclidean distance and the fact that are geometrically complementary (absolute minimum-direct distance or absolute maximum-inverse distance). Similarly to what done in Section 2.1.2, also for the uncertainty propagation of the C score $\overline{\mathscr{C}}_n$, a direct MCS scheme can be adopted: for each MCS trial $g = 1, 2, \dots, N_s$, a vector of gaussian noise $\overline{\varepsilon}_n^{v(g)}$ for each EXP$_n$, is sampled and added to the experimental control variables $\overline{x}_n^v$ before the evaluation of the metrics ϕ_j, $\forall j = 1, 2, \dots, J$; then, following the aforementioned approach based on the Euclidean distance, the C results $\overline{\mathscr{C}}_n^{(g)}$ related to the g-th simulation are calculated. Finally, a distribution $\mathscr{C} = [\overline{\mathscr{C}}_n^{(g)}]_{N \times N_s}$, $n = 1, 2, \dots, N$ and $g = 1, 2, \dots, N_s$, is obtained.

As we shall see in the case study, this approach for C assessment is fully quantitative in nature and allows reducing the user effect on the analysis (see Fig 3).

2.3. Adequacy assessment

The objective of DA assessment is to combine the distributions of the R and C scores, i.e., $\mathscr{R} = [\overline{\mathscr{R}}_n^{(g)}]_{N \times N_s}$, and $\mathscr{C} = \left[\overline{\mathscr{C}}_n^{(g)}\right]_{N \times N_s}$, $n = 1, 2, \dots, N$ and $g = 1,2, \dots, N_s$, in one final distribution of values of adequacy scores $\mathscr{DA} = \left[\overline{\mathscr{DA}}_n^{(g)}\right]_{N \times N_s}$. Since there is not an objective way to aggregate $\mathscr{R}$ and $\mathscr{C}$ without relying on subjective expert-based evaluations, the same MCS approach described in Section 2.1 can be adopted. Hence, we have

Table 4
Criteria for C, [*] suggested by ATRIUM project coordinators (Baccou et al. 2023), where [#] means in this work originally proposed.

Criteria	Meaning	Sub Criteria	Quantitative Metric	Meaning
C.C$_1$	Coverage of the application domain	C.C$_{1.1}$	$\phi_{C.C_{1.1}} = \frac{\Delta P_0^n}{\Delta P_0^{LSTF}}$ [*]	Ratio between the volumes of the experimental and LSTF input domains (P_0)
		C.C$_{1.2}$	$\phi_{C.C_{1.2}} = \frac{\Delta T_0^n}{\Delta T_0^{LSTF}}$ [*]	Ratio between the volumes of the experimental and LSTF input domains (T_0)
C.C$_2$	Spatial distribution of the experiments in the experimental domain	C.C$_{2.1}$	$\phi_{C.C_{2.1}} = \sum_{u=1}^{N_u} \frac{\left(O_{P_u}^n - E_u\right)^2}{E_u}$ [#]	Uniformity of the distribution of the input variable (P_0)
		C.C$_{2.2}$	$\phi_{C.C_{2.2}} = \sum_{u=1}^{N_u} \frac{\left(O_{T_u}^n - E_u\right)^2}{E_u}$ [#]	Uniformity of the distribution of the input variable (T_0)
		C.C$_{2.3}$	$\phi_{C.C_{2.3}} = \sum_{q=1}^{Q} \log_Q\left(\frac{1}{P_q^n}\right) P_q^n$ [#]	Uniformity of the distribution of single and multiphase flow regimes in the experimental domain (X_0)

collected M (2 × 2) expert-compiled comparison matrices. The m-th expert matrix looks like:

$$A^m_{DA} = \begin{bmatrix} 1 & a^m_{RC} \\ \frac{1}{a^m_{RC}} & 1 \end{bmatrix} \tag{12}$$

As it can be seen in Eq. (12), a single pairwise comparison a^m_{RC} is needed to define A^m_{DA}, expressing the m-th expert opinion about the importance of the R score with respect to the C score for the DA. As seen in Section 2.1.1, by IA it is possible to determine the matrices A^{DA}_{min}, A^{DA}_{max} and define a PDF $f(a_{RC})$ to perform the MCS: for each simulation $g = 1,2, \dots, N_s$, the matrix $A^{(g)}_{DA}$ is compiled by sampling from $f(a_{RC})$; since its order is $H = 2$, it can be directly solved following the standard AHP procedure, without the need of performing the consistency test (Saaty and Vargas, 2012). The resulting normalized dominant eigenvector $\overline{W}^{(g)}_{DA}$ of size (2 × 1) carries the quantitative information related to the relative importance between R and C scores in shaping the final DA ranking for the g-th MCS trial. Finally, collecting $\overline{\mathscr{R}}^{(g)}_n = \left[\mathscr{R}^{(g)}_1, \cdots, \mathscr{R}^{(g)}_n, \cdots, \mathscr{R}^{(g)}_N\right]^T$ and $\overline{\mathscr{C}}^{(g)}_n = \left[\mathscr{C}^{(g)}_1, \cdots, \mathscr{C}^{(g)}_n, \cdots, \mathscr{C}^{(g)}_N\right]^T$ in a (N × 2) matrix, the vector (N × 1) $\overline{\mathscr{DA}}^{(g)}_n = \left[\mathscr{DA}^{(g)}_1, \cdots, \mathscr{DA}^{(g)}_n, \cdots, \mathscr{DA}^{(g)}_N\right]^T$ carrying the DA scores for the g-th simulation, is obtained as:

$$\overline{\mathscr{DA}}^{(g)}_n = \begin{bmatrix} \mathscr{DA}^{(g)}_1 \\ \vdots \\ \mathscr{DA}^{(g)}_n \\ \vdots \\ \mathscr{DA}^{(g)}_N \end{bmatrix} = \begin{bmatrix} \mathscr{R}^{(g)}_1 & \mathscr{C}^{(g)}_1 \\ \vdots & \vdots \\ \mathscr{R}^{(g)}_n & \mathscr{C}^{(g)}_n \\ \vdots & \vdots \\ \mathscr{R}^{(g)}_N & \mathscr{C}^{(g)}_N \end{bmatrix} \begin{bmatrix} W^{(g)}_R \\ W^{(g)}_C \end{bmatrix} = \begin{bmatrix} \mathscr{R}^{(g)}_1 W^{(g)}_R + \mathscr{C}^{(g)}_1 W^{(g)}_C \\ \vdots \\ \mathscr{R}^{(g)}_n W^{(g)}_R + \mathscr{C}^{(g)}_n W^{(g)}_C \\ \vdots \\ \mathscr{R}^{(g)}_N W^{(g)}_R + \mathscr{C}^{(g)}_N W^{(g)}_C \end{bmatrix} \tag{13}$$

By solving the matrix product of Eq. (13), the adequacy scores $\overline{\mathscr{DA}}^{(g)}_n = \left[\mathscr{DA}^{(g)}_1, \cdots, \mathscr{DA}^{(g)}_n, \cdots, \mathscr{DA}^{(g)}_N\right]^T$ can be seen as a weighted average of vectors $\overline{\mathscr{R}}^{(g)}_n$ and $\overline{\mathscr{C}}^{(g)}_n$, where the weight is the eigenvector $\overline{W}^{(g)}_{DA} = \left[W^{(g)}_R, W^{(g)}_C\right]^T$. At the end of the g-th simulation, the distributions of adequacy scores $\mathscr{DA} = \left[\overline{\mathscr{DA}}^{(g)}_n\right]_{N \times N_s}$, $n = 1, 2, \dots, N$ and $g = 1, 2, \dots, N_s$, are obtained. A sketch of the full procedure of aggregation is presented in Fig 4. It is worth noting that from a computational point of view, thanks to the linear dependency between all the elements of the AHP hierarchy, the MCS evaluations needed for the assessment of R and C Eq. (13) can be parallelized, to further reduce the computational time.

3. Case study

3.1. The ATRIUM Project

The ATRIUM project has been promoted by NEA/CSNI/WGAMA for advancing the methodologies of IUQ in the framework of BEPU modelling for NPPs safety assessment (Ghione, 2023). The final goal of the project is performing the IUQ of BE models for an IBLOCA that might occur in the Integral Effect Test facility (IET) LSFT IB-HL-01 (see Table 1). As said in the Introduction DA assessment is necessary to identify the Separate Effect Test facilities (SETs) and their experimental datasets to be used to model the phenomena of interest (Baccou et al. 2023). The available experimental datasets are listed in Table 2, in which the name of the set, the label utilized throughout the paper, the number of experimental points N_P and the geometric (pipe diameter ∅, length over diameter ratio $L/\varnothing$) and T-H (pressure P_0, temperature T_0, quality X_0, all at stagnation point, and the QoI, critical mass flux.

$\dot{\Gamma}$ in steady state conditions) properties are also given. The three SET facilities, namely S-S, SMD and Marv, are characterized by a similar experimental layout, i.e., a pressure vessel connected through a discharge line of variable length (L) to an open nozzle simulating the break. The differences in nozzle geometry are used to define the experimental datasets adopted in the analysis for each experiment. Both S-S and SMD collect direct measurements of the steady state mass flow rate at nozzle exit whereas for Marv tests the complete discharge transient is recorded. For the interested reader, a detailed description of S-S, SMD and Marv can be found in (Sozzi and Sutherland, 1975), (Rousseau, 1987), (Sokolowski and Kozlowski, 2012), respectively.

The approach presented in Section 2 has been implemented to perform the DA assessment for the SETs presented in Table 2. The hierarchy for the AHP is sketched in Fig 5: the evaluation of R and C is performed with two layers of criteria. We refer to the upper level simply as criteria and to the lower one as sub-criteria. Moreover, for the C assessment just the sub-criteria C.C$_{j.h}$, $j = 1, 2$ and h, variable depending on j, are evaluated, accordingly to the approach presented in Section 2.2, since the layer of C criteria C.C$_j$, $j = 1, 2$, is adopted just for coherence respect to the hierarchical structure of the problem. The following assumptions hold:

- a set of $M = 5$ alternative experts are asked to provide M independent judgements on:
  - the qualitative criteria and sub-criteria, listed in Table 3,i.e., the C.R$_i$ and C.R$_{i.t}$, $i = 1, 2, 3$, and $t = 1, \dots, N_t$, with the value of N_t depending on i, e.g., if $i = 1$or 3 then $N_t = 2$, if $i = 2$ then $N_t = 3$;
  - the relative importance of R versus C Eq. (12);
- every m-th expert opinion is considered equally important (Zio, 1996);
- the only source of uncertainty for the C.R$_i$ and C.R$_{i.t}$, is related to the disagreement between experts on the pairwise preference scores a_{lk};
- for the MCS procedure, the PDF $f(a_{lk})$ is assumed either uniform or triangular, centered on A_{mean} Eq. (2). This allows comparing the effects of the assumption of equally important judgments across the M experts (e.g., uniform distribution U) with that of a probability mass centered on the empirical mean of the judgements (e.g., triangular distribution T). The interested reader can refer to (O'Hagan et al., 2006); (Oberkampf et al., 2004) for details on the process of building PDFs from expert judgement elicitation processes;
- the adopted criteria for R assessment are shown in Table 3;
- the only source of uncertainty for the quantitative evaluations of C.C$_{j.h}$, $j = 1, 2$, and $h = 1, \dots, N_h$, with the value of N_h dependent on j (e.g., if $j = 1$ then $N_h = 2$, if $j = 2$ then $N_h = 3$), is related to the experimental uncertainty (modelled as gaussian noise ε if not available) of the experimental decision variables $\overline{x} = [P_0, T_0, X_0]$ (pressure (P_0), temperature (T_0) and quality (X_0), at the stagnation point). The adopted quantitative metrics $\phi_{j.h}$ are listed in Table 4 (see Appendix B for details) and are defined for each variable independently. For future developments, the application of more sophisticated metrics capable to properly consider the multidimensional nature of the SETs physical space could be explored. It should be noted that the adoption of the Euclidean distance for the aggregation of the C criteria scores makes the results of the analysis dependent on the number of C criteria;

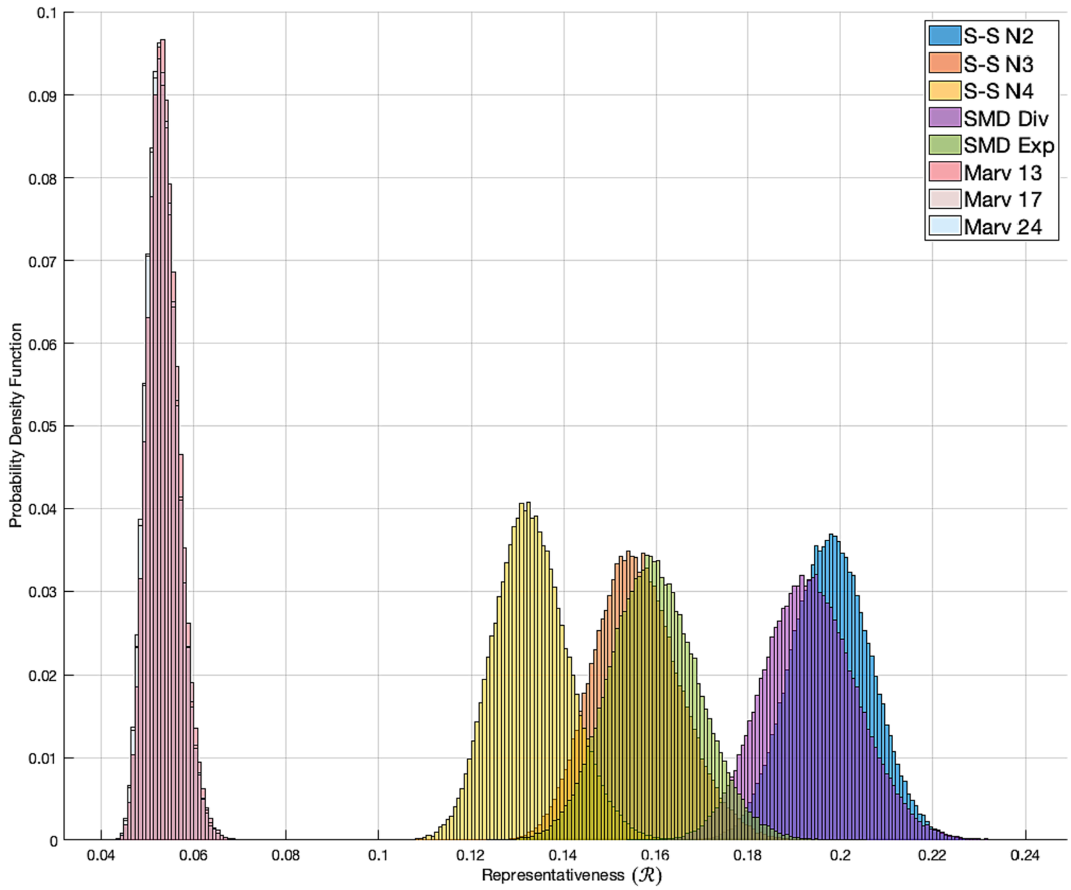


Fig. 6. R score distributions using MCS with uniform distributed a_{lk}.

Table 5
R MCS results using uniform (U) and triangular (T) distributions.

Experiment Name	$\mu_U(\mathscr{R})$	$\sigma_U(\mathscr{R})$	$\mu_T(\mathscr{R})$	$\sigma_T(\mathscr{R})$
S-S N2	0.1968	0.0092	0.1979	0.0061
S-S N3	0.1568	0.0096	0.1529	0.0065
S-S N4	0.1332	0.0084	0.1309	0.0059
SMD Div	0.1949	0.0104	0.1976	0.0071
SMD Exp	0.1609	0.0108	0.1648	0.0068
Marv 13	0.0542	0.0033	0.0548	0.0023
Marv 17	0.0534	0.0033	0.0535	0.0023
Marv 24	0.0534	0.0033	0.0535	0.0023

Table 6
R MCS results using uniform (U) distributions with groups of $M = 10$ and $M = 15$ experts (groups are not independent).

Experiment Name	$\mu_{U10}(\mathscr{R})$	$\sigma_{U10}(\mathscr{R})$	$\mu_{U15}(\mathscr{R})$	$\sigma_{U15}(\mathscr{R})$
S-S N2	0.2005	0.0126	0.2103	0.0164
S-S N3	0.1447	0.0133	0.1428	0.0155
S-S N4	0.1266	0.0110	0.1318	0.0122
SMD Div	0.1952	0.0145	0.1986	0.0178
SMD Exp	0.1622	0.0106	0.1528	0.0146
Marv 13	0.0574	0.0045	0.0586	0.0060
Marv 17	0.0567	0.0041	0.0528	0.0050
Marv 24	0.0567	0.0041	0.0523	0.0051

- since the Marviken database is composed by three time series (full transients of the discharge phenomena), for the evaluation of C.C$_{j.h}$, a global assessment on the whole database has been performed, just considering as data the time series steady states. Then, the obtained C result is assigned equally to each member of the database (Marv 13, Marv 17, Marv 24);

Table 7
R results using score aggregation (Zio et al., 2003) with groups of $M = 5$, $M = 10$, $M = 15$ experts (groups are not independent).

Experiment Name	$\mu_5(\mathscr{R})$	$\sigma_5(\mathscr{R})$	$\mu_{10}(\mathscr{R})$	$\sigma_{10}(\mathscr{R})$	$\mu_{15}(\mathscr{R})$	$\sigma_{15}(\mathscr{R})$
S-S N2	0.1788	0.0251	0.1792	0.0348	0.1868	0.0392
S-S N3	0.1335	0.0225	0.1360	0.0240	0.1386	0.0254
S-S N4	0.1165	0.0248	0.1259	0.0232	0.1359	0.0248
SMD Div	0.2103	0.0469	0.2030	0.0404	0.1953	0.0370
SMD Exp	0.1893	0.0264	0.1812	0.0377	0.1700	0.0369
Marv 13	0.0574	0.0036	0.0599	0.0046	0.0605	0.0072
Marv 17	0.0568	0.0031	0.0579	0.0034	0.0570	0.0047
Marv 24	0.0573	0.0039	0.0569	0.0045	0.0560	0.0050

4. Results and discussion

4.1. Representativeness

In Fig 6, the results of the MCS approach presented in Section 2.1.2 for the R assessment in the case of uniform PDF $f(a_{lk})$ and a group of $M = 5$ experts are presented. Assuming $N_s = 10^5$ MCS trials and the evaluation of all the criteria and sub-criteria listed in Table 3, following the hierarchical structure presented in Fig 5, the normalized histograms of R for each n-th EXP are plotted. The resulting R distributions mean values μ are clearly separated in a high values cluster (S-S N2, SMD Div, SMD Exp, S-S N3, S-S N4) and a low value one (Marv 13, Marv 17, Marv 24). The estimated uncertainty in terms of the results standard deviation σ,

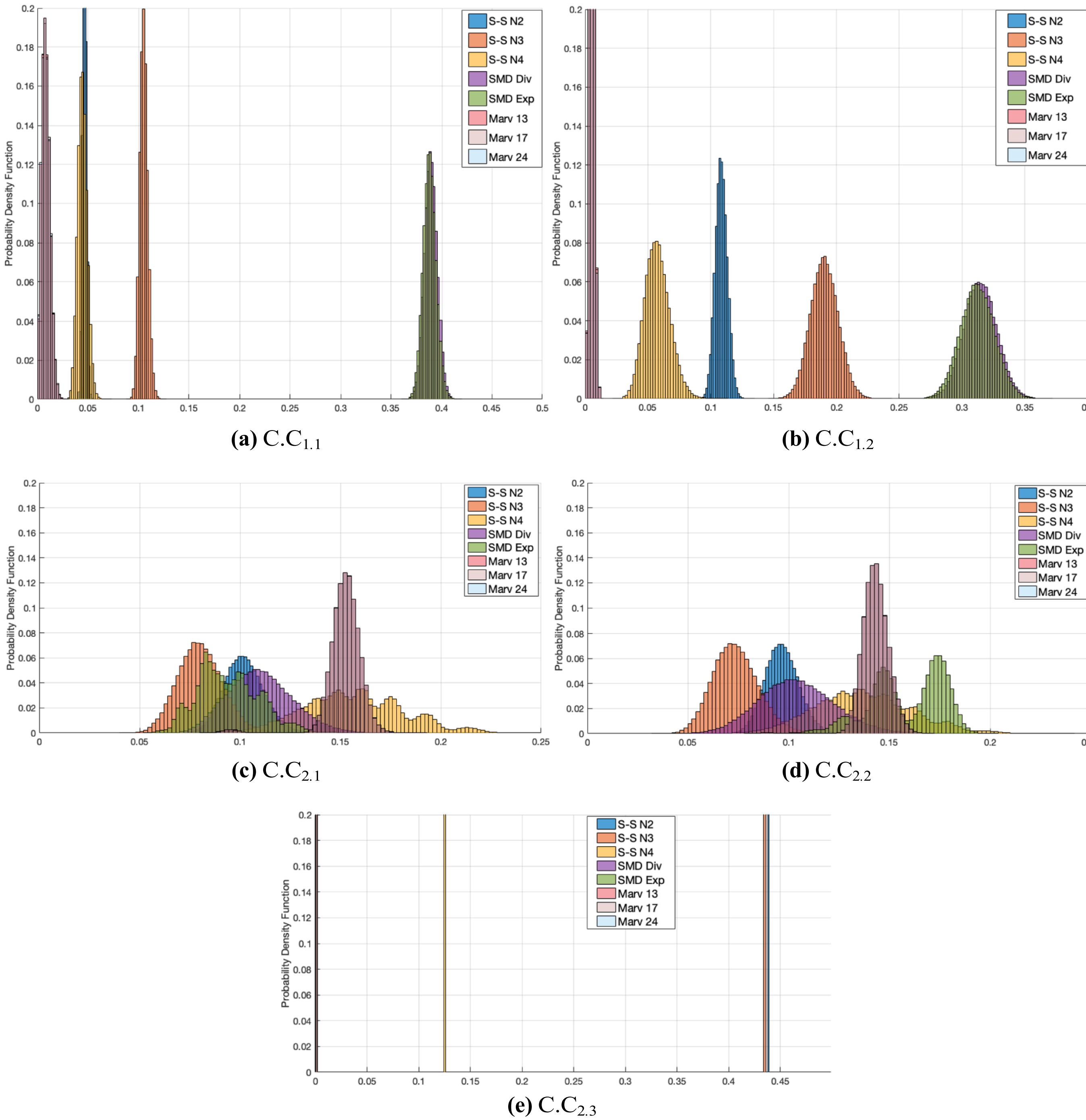


Fig. 7. Resulting distributions from the quantitative evaluation of C criteria presented in Table 4.

Table 8
Quantitative evaluation of C criteria, mean value μ of estimated distributions.

Experiment Name	μ(C.C$_{1.1}$)	μ(C.C$_{1.2}$)	μ(C.C$_{2.1}$)	μ(C.C$_{2.2}$)	μ(C.C$_{2.3}$)
S-S N2	0.0468	0.1084	0.1020	0.0966	0.4393
S-S N3	0.1051	0.1904	0.0802	0.0738	0.4349
S-S N4	0.0448	0.0575	0.1569	0.1372	0.1256
SMD Div	0.3895	0.3145	0.1102	0.1041	0
SMD Exp	0.3883	0.3123	0.0942	0.1586	0
Marv 13	0.0084	0.0055	0.1521	0.1431	0
Marv 17	0.0083	0.0055	0.1521	0.1431	0
Marv 24	0.0084	0.0055	0.1521	0.1431	0

Table 9
Quantitative evaluation of C criteria, standard deviation σ of estimated distributions.

Experiment Name	σ(C.C$_{1.1}$)	σ(C.C$_{1.2}$)	σ(C.C$_{2.1}$)	σ(C.C$_{2.2}$)	σ(C.C$_{2.3}$)
S-S N2	0.0019	0.0048	0.0098	0.0084	0
S-S N3	0.0040	0.0110	0.0111	0.0113	0
S-S N4	0.0046	0.0098	0.0253	0.0242	0
SMD Div	0.0063	0.0131	0.0159	0.0188	0
SMD Exp	0.0063	0.0135	0.0153	0.0183	0
Marv 13	0.0041	0.0019	0.0093	0.0068	0
Marv 17	0.0041	0.0019	0.0093	0.0068	0
Marv 24	0.0041	0.0019	0.0093	0.0068	0

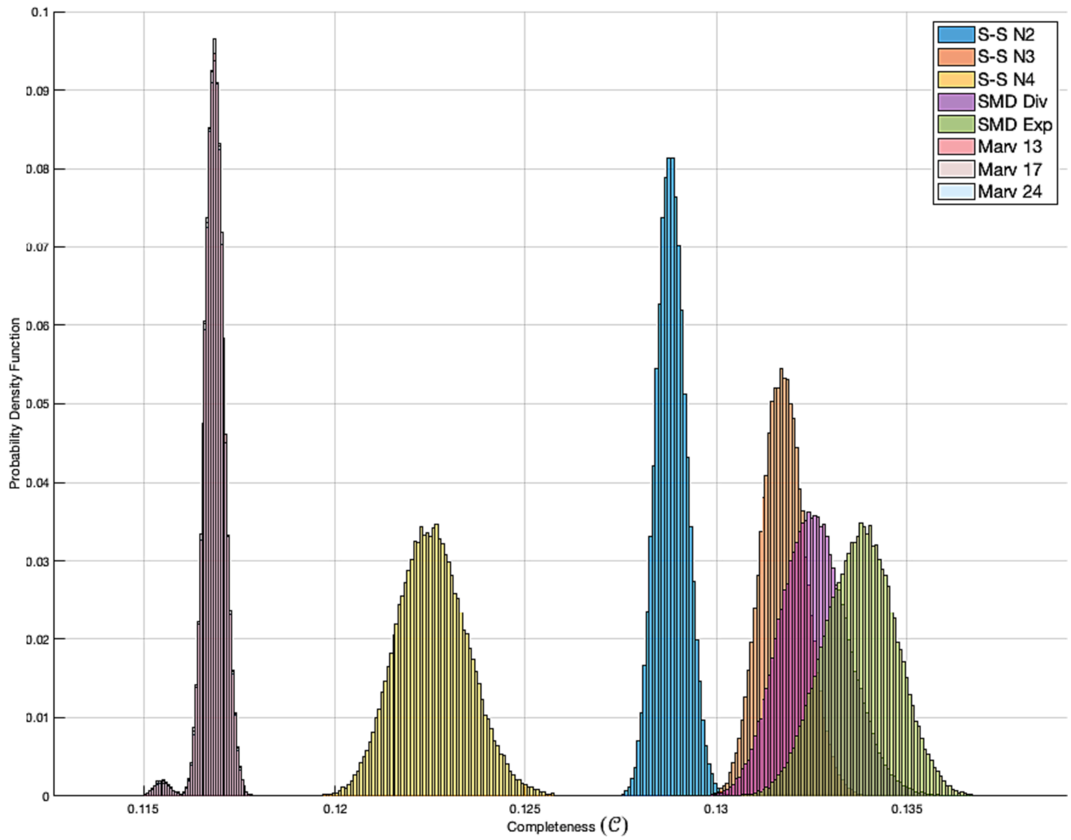


Fig. 8. C results using the fully quantitative approach based on the Euclidean norm.

Table 10
C results.

Experiment Name	$\mu(\mathscr{C})$	$\sigma(\mathscr{C})$
S-S N2	0.1288	0.0004
S-S N3	0.1318	0.0006
S-S N4	0.1226	0.0009
SMD Div	0.1326	0.0009
SMD Exp	0.1338	0.0009
Marv 13	0.1168	0.0003
Marv 17	0.1168	0.0003
Marv 24	0.1168	0.0003

combined with the shape of the distributions, is useful to identify as clearly more representative the three highest ranked experiments (S-S N2, SMD Div, SMD Exp) among the cluster of high R scores. Indeed, the high-μ experiments, have quite similar σ but (S-S N3 and S-S N4) appear more positively skewed and more peaked on their μ, than (S-S N2, SMD Div, SMD Exp), which tend to be more symmetric. The latter qualitative considerations on high distributions moments enforce the capability to assess a reasonable ranking from the results μ. In Table 5, the numerical values of the first two moments (sample mean μ and sample standard deviation σ) of the R distributions are listed, when both uniform and triangular PDFs $f(a_{lk})$, with same range of the uniform PDF and maximum centered on A_{mean}, are adopted. It should be noticed that the results are not sensible to the two different PDFs assumed in terms of μ, but they are in terms of σ: as suggested in (Cagno et al., 2000); (Cagno et al., 2001), the adoption of uniform PDFs is taken to conservatively maximize the uncertainty on the obtained results, as usually done in safety–critical applications like NPPs safety assessment. Moreover, uniform PDFs enable the exploration of a large number of different combinations of a_{lk} values, overcoming the limitations related to the typically scarce information available from expert elicitation (Cagno et al., 2001). In general, the number of experts M might affect the results. In our case, Table 6 shows (for uniform distributions of $f(a_{lk})$ and three groups of experts G_M) that by changing M only minor modifications occur in μ whereas σ increases with M, as expected because of the increased variability of expert judgements which leads to the widening of the support of $f(a_{lk})$. In Table 7, the results of a simplified methodology based on single experts score aggregation (Zio et al., 2003) is reported, for the sake of comparison. As it can be seen, by comparing the results of Table 7 with that of Table 5, the overall rankings are similar in terms of μ, but different in terms of σ (due to the dependence of σ from the number N_s of MCS trials). In addition, increasing the number of experts, the R scores tend to approximate better those provided by MCS. This means that the MCS scores can be seen as the limit value for the number of artificially generated expert elicitations going to infinite: thus, the main advantage of the proposed MCS approach is that it allows exploring the full range of variability of expert-based judgements, which is typically limited by the time consuming elicitation process.

4.2. Completeness

In Fig 7, the results of the approach based on quantitative metrics (presented in Section 2.2) for the evaluation of the C criteria listed in Table 4 are presented. The C scores distributions show the effect of the propagation of experimental noise on the metric input variables. Indeed, the results of the first two criteria (C.C$_{1.1}$, C.C$_{1.2}$), related to the coverage of the application domain, are much more stable under perturbed inputs (lower σ) than the results obtained from the metrics related to uniformity (C.C$_{2.1}$, C.C$_{2.2}$) (much higher σ). Then, we can claim that the choice of different metrics can greatly affect the overall uncertainties introduced in the analysis, hence the incorporation of experimental noise should be a mandatory element to obtain robust results, even when quantitative methods are employed. Moreover, for the first two criteria (C.C$_{1.1}$, C.C$_{1.2}$) the results are clearly separated by their μ values,

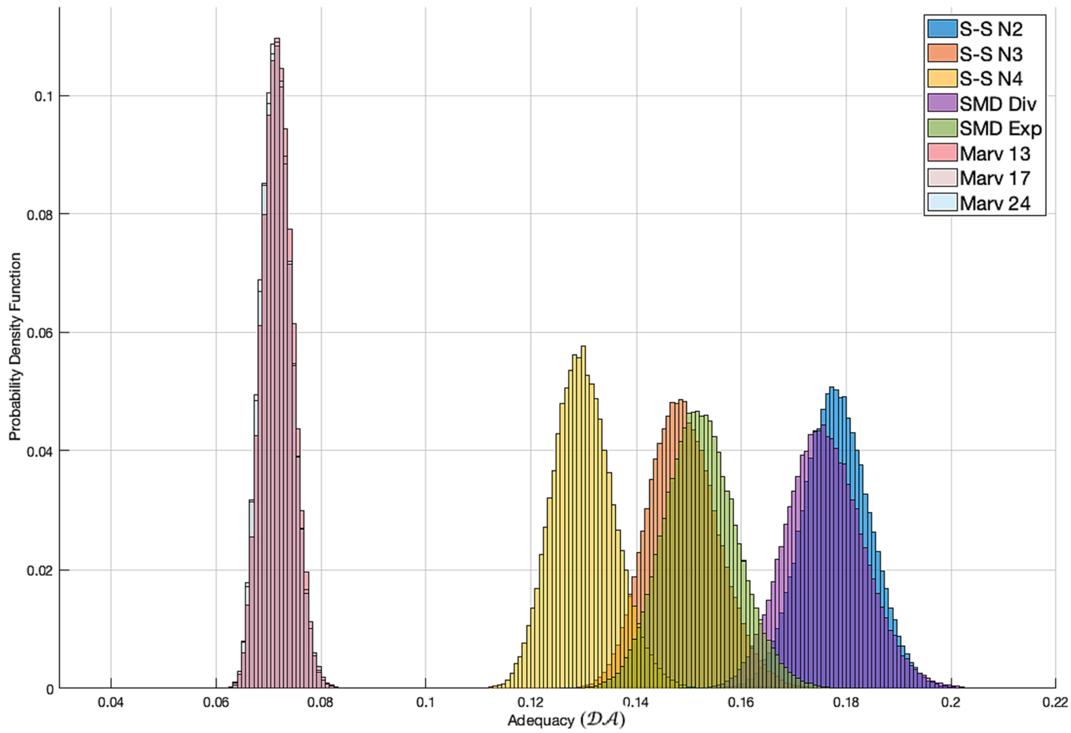


Fig. 9. Resulting distributions for the DA assessment.

whereas for the criteria related to uniformity ($C.C_{2.1}$, $C.C_{2.2}$) their μ tend to be much closer. Thus, due to the large σ it is very difficult, if not impossible, to assess an objective preference for the EXP_n just relying on $C.C_{1.1}$ and $C.C_{1.2}$. Finally, the last criteria $C.C_{2.3}$ based on the entropy measurement of the dataset in terms of initial flow conditions is not affected by experimental uncertainties (since we have neglected the uncertainties in the identification of single or multiphase inlet conditions). The numerical values of the distribution's μ and σ, for each criteria, are listed in Table 8 and Table 9, respectively.

In Fig 8, the results of the aggregation of the quantitative criteria evaluated before, in one single C measure through the procedure based on the Euclidean norm (Section 2.2.1) are presented. As we can observe in Fig 8, the low ranked experiments in terms of C are exactly correspondent to the lowest ranked ones for R (Marv 13, Marv 17, Marv 24). Instead, for the high scored experiments (SMD Exp, SMD Div, S-S N3, S-S N2) the same considerations on σ and skewness are again useful to identify with a high confidence the best experiments (SMD Exp, SMD Div) in terms of C score. In Table 10, the numerical results for the aggregated C score distributions are listed in terms of μ and σ. Moreover, it is worth noting that, thanks to the application of a fully quantitative procedure (without expert-based evaluations), the overall uncertainty on the C results (in terms of σ) is one order of magnitude lower than the one obtained for the R assessment in Section 4.1. The assumption of C criteria as equally important and the definition of specific quantitative metrics for their evaluation can still be regarded as an activity affected by expert-judgement; nevertheless, the employment of quantitative information can reduce the impact of expert judgement on the overall analysis.

Table 11
DA results.

Experiment Name	$\mu(\mathcal{DA})$	$\sigma(\mathcal{DA})$
S-S N2	0.1782	0.0064
S-S N3	0.1489	0.0066
S-S N4	0.1300	0.0058
SMD Div	0.1760	0.0074
SMD Exp	0.1522	0.0069
Marv 13	0.0718	0.0029
Marv 17	0.0715	0.0029
Marv 24	0.0715	0.0029

4.3. Adequacy

In Fig 9, the final results in terms of DA distributions of the EXP_n are presented. Since the DA score is obtained as an expert-based weighted average Eq. (13) between R and C, as described in Section 2.3, it is not surprising that the DA ranking clearly identifies as best experiments (S-S N2, SMD Div), followed by (SMD Exp, S-S N3, S-S N4) in the medium range and as worst alternatives the Marviken experiments (Marv 13, Marv 17, Marv 24). As reported in Table 11, the information about

Table 12
Comparison of results for R, C and DA from a single-expert, standard AHP and the approach proposed in Section 2.

Experiment Name	$\mathcal{R}^{AHP}$	$\mu(\mathcal{R}) \pm 3\sigma$	$\mathcal{C}^{AHP}$	$\mu(\mathcal{C}) \pm 3\sigma$	$\mathcal{DA}^{AHP}$	$\mu(\mathcal{DA}) \pm 3\sigma$
S-S N2	0.142	0.1968 ± 0.0276	0.094	0.1288 ± 0.0012	0.130	0.1782 ± 0.0192
S-S N3	0.105	0.1568 ± 0.0288	0.094	0.1318 ± 0.0018	0.102	0.1489 ± 0.0198
S-S N4	0.105	0.1332 ± 0.0252	0.094	0.1226 ± 0.0027	0.102	0.1300 ± 0.0174
SMD Div	0.166	0.1949 ± 0.0312	0.219	0.1326 ± 0.0027	0.179	0.1760 ± 0.0222
SMD Exp	0.166	0.1609 ± 0.0324	0.219	0.1338 ± 0.0027	0.179	0.1522 ± 0.0207
Marv 13	0.105	0.0542 ± 0.0099	0.094	0.1168 ± 0.0009	0.102	0.0718 ± 0.0087
Marv 17	0.105	0.0534 ± 0.0099	0.094	0.1168 ± 0.0009	0.102	0.0715 ± 0.0087
Marv 24	0.105	0.0534 ± 0.0099	0.094	0.1168 ± 0.0009	0.102	0.0715 ± 0.0087

$\sigma(\mathscr{DA})$ is fundamental to assess the validity of the obtained final ranking, especially for experiments with similar $\mu(\mathscr{DA})$ values, like the (SMD Exp, S-S N3) and (SMD Div, S-S N2), for which just qualitative judgments related to the distributions shapes (higher moments), combined with quantitative information about their σ, can confirm the superiority of one experiment instead of another. For example, in the case of S-S N2 compared with SMD Div, the lower variance of S-S N2 acts as a favorable element to support its superiority with respect to SMD Div. On the other hand, in the case of SMD Exp and S-S N3, since the σ values are also comparable, the consideration of the fact that SMD Exp distribution is more positively skewed than S-S N3 supports its higher ranking.

4.4. Comparison of the results with the standard AHP method

This Section presents the comparison of the results obtained by the application of the classical single- expert AHP methodology (Saaty, 1980) and those presented in Section 4. As can be seen in Table 12, where the results of the single-expert AHP (Ghione, 2023) are listed as ($\mathscr{R}^{AHP}$, $\mathscr{C}^{AHP}$, $\mathscr{DA}^{AHP}$) and compared with the mean (μ) and standard deviation (σ) of the distributions obtained with the here proposed approach, one of the classical limitations of the standard AHP methodology lies in the impossibility to discriminate the experiments when their score is quite similar. In this case, it is even worse because a lot of single-expert scores are basically identical, so that without a proper uncertainty quantification procedure it is impossible to assess an objective final ranking for the experiments. Instead, following our proposed method, it is possible to obtain a finer discrimination between the results due to the additional information on their σ and the distributions higher moments (skewness and kurtosis). Moreover, the single-expert results for R ($\mathscr{R}^{AHP}$) and DA ($\mathscr{DA}^{AHP}$) are consistent with the value of μ of the obtained distributions for what regards the identification of the three highest scored experiments (SMD Div, SMD Exp, S-S N2). Only the results for C ($\mathscr{C}^{AHP}$) differ greatly (in magnitude), due to the adoption of quantitative metrics presented in Section 2.2; nevertheless, the overall C ranking is still comparable. As a final consideration, we could state that, even if the single-expert deterministic AHP is capable to correctly assess the presence of clusters of different values, i.e., identify the set of best experiments, it is not capable to prove in an objective way that the obtained ranking is effectively sound (Rosenbloom, 1997). This feature could be an obstacle when the decision problem stakes and uncertainties are large (Saltelli et al., 1999), like in many applications of the nuclear industry. In these cases, the adoption of the proposed approach could be helpful in terms of added robustness on the analysis results.

5. Conclusion

In this paper, a novel approach for performing DA assessment of experimental databases in light of an application to nuclear T-H codes is presented. The proposal is rooted on established methods and tailored on the peculiar needs of a IUQ methodology accordingly to (Baccou, 2023). The developed approach consists in the reformulation of the decision problem with a hierarchical structure, where qualitative evaluations are performed with multi-expert AHP and quantitative evaluations are carried out through quantitative metrics. The coupling of different sources of information (qualitative and quantitative) is performed exploiting a direct MCS scheme for uncertainty propagation. The quantitative evaluation of the C sub-criteria has reduced the bias of expert judgement on the analysis results. This approach has been applied to a case study taken from the first exercise of the ATRIUM project, focused on uncertainties related to critical flow modelling during IBLOCA accidents, whose objective is the selection of adequate SETs experimental databases to perform the IUQ of a large scale facility model (IET). The underlying conjecture is that the most adequate SETs will provide the best results when the uncertain parameters estimated by IUQ using their data are propagated to the target model, i.e., LSTF. The results have been compared with the ones obtained through a standard AHP methodology showing that the extended AHP overcome the traditional AHP, allowing for an objective ranking, in light of the different sources of uncertainty affecting the decision problem, arising both from qualitative and the quantitative criteria evaluations. The computational demand for the case study (seven (8×8) matrices, two (3×3), three (2×2) and the full quantitative evaluation of five C sub-criteria) has been of 500 seconds on an Intel I9 2.8 GHz processor for the complete simulation of 10^5 MCS trials. We can, therefore, claim that this approach is a good compromise between results precision, analysis completeness and required computational time. Moreover, this framework based on MCS could be easily extended to account for other MCDM methods outside the AHP. Additionally, the robustness of this approach suggest that it could be effectively adopted as a benchmarking method for testing the effectiveness and performances of other alternative methods of reduced computational complexity. Considering future developments, research efforts should be focused on the minimization of expert-based reliance for the R assessment, possibly by rethinking if some of the most qualitative features of those criteria could be reformulated with an equivalent quantitative formulation.

CRediT authorship contribution statement

Francesco Di Maio: Conceptualization, Methodology, Supervision, Writing – review & editing. **Thomas Matteo Coscia:** Methodology, Software, Data curation, Writing – original draft. **Enrico Zio:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors thank Dr. Andrea Bersano (ENEA, Italy), Dr. Fulvio Mascari (ENEA, Italy) and Prof. Nicola Pedroni (Politecnico di Torino, Italy) for insightful discussions that have helped developing the approach and testing it on real data. Also, the authors would like to acknowledge the contribution of all those individuals who had a key role and leadership in the conduct of the activity of the Application Tests for Realisation of Inverse Uncertainty quantification and validation Methodologies in thermal-hydraulics (ATRIUM) of the NEA Working Group on the Analysis and Management of Accidents (WGAMA), especially the task leaders L. Sargentini and A. Ghione (CEA, France).

Appendix A

In Fig. A1 is reported a schematic example of a hierarchical decomposition of a MCDM problem, for further details on the subject, the interested reader should consult (Saaty, 1980; Saaty, 1990); (Vargas, 1990) and (Zio, 1996); (Ramanathan and Ganesh, 1994), for an application to expert judgment in group decision making. The base of AHP consists in the building of a hierarchical representation of the problem under analysis following

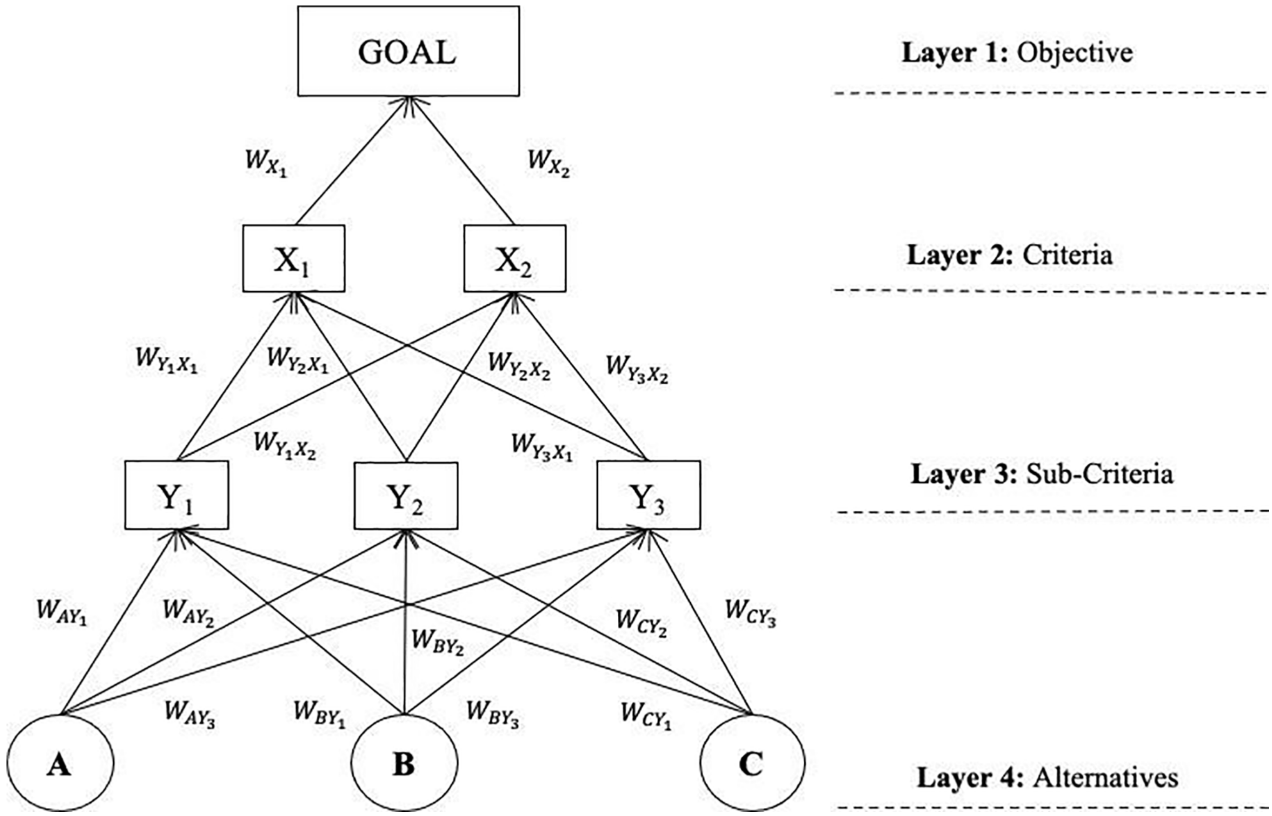


Fig. A1. A four level hierarchy.

the hereafter listed steps:

1. Define the object of the analysis and place it at the top of the hierarchy.
2. Follow an up-down logic, starting from the top objective define the lower levels of the hierarchy by aggregating in the same level all the factors that are directly influencing the layer above and are influenced by the layer below. Each interconnection between the elements is specified by a single arrow.
3. In the lowest level define the decision problem alternatives.

The hierarchical form of the problem is able to represent how interaction between different elements affect each other through pairwise importance judgments (preferences) a_{lk}, with $l = 1, 2, \ldots, H$ and $k = 1, 2, \ldots, H$, that can be collected following this procedure:

1) For each element of each level build a pairwise comparison matrix to assess the relative importance on the influence of the entries of the level below in relation to the element under analysis. The pairwise comparisons can be performed directly with reference to numerical scale or in a qualitative fashion (linguistic statement) and then translated into a numerical scale. Typically, the scale of integer numbers from 1 to 9 is used and the values a_{lk} obtained from the comparisons are organized in a square matrix. For example, performing qualitatively the comparison of element A with element B, the scale is the following:

1 = A and B equally important.
3 = A slightly more important than B.
5 = A strongly more important than B.
7 = A very strongly more important than B.
9 = A absolutely more important than B.

By definition, an element is equally important when compared to itself so the principal diagonal of the matrix is filled with ones. The appropriate reciprocals, 1/3, 1/5, ……., 1/9, are inserted where the reverse comparison, B vs. A, is required. The numbers 2, 4, 6, 8, and their reciprocals can be used to facilitate expressing judgments for intermediate situations. In Table A1 is presented an example of comparison matrix for the hierarchy presented in Fig. A1.

2) For each element δ in level s of the problem hierarchy, determine the weight (strength, priority) $W_{\beta\delta}$, with which each element β of level $(s-1)$ affect element δ. The priority $W_{\beta\delta}$ quantify the relative importance of an element β regarding the element δ in the next level of the hierarchy (arrows in

Table A1
An Example of comparison matrix for the alternatives (A,B,C) vs sub-criteria Y_1.

Y_1	A	B	C
A	1	6	6
B	1/6	1	3
C	1/6	1/3	1

Table A2
An Example of comparison matrix for the sub-criteria (Y_1,Y_2,Y_3) vs criterion X_1.

X_1	Y_1	Y_2	Y_3
Y_1	1	3	2
Y_2	1/3	1	3
Y_3	1/2	1/3	1

Table A3
An Example of comparison matrix for the criteria (X_1,X_2) vs top objective (GOAL).

GOAL	X_1	X_2
X_1	1	3
X_2	1/3	1

Fig. A1). Priorities can be determined by solving an eigenvector problem (Saaty, 2003). For example, $\overline{W}_{AY_1} = [W_{AY_1}, W_{BY_1}, W_{CY_1}]^T$ is the vector (3 × 1) collecting the priorities of alternatives (A,B,C) with respect to the sub-criteria Y_1 (eigenvector of the matrix presented in Table A.1). More precisely, it can be shown that given the matrix of pairwise comparisons a_{lk} for the element of interest, the normalized principal eigenvector provides the vector of priorities, and the maximum eigenvalue is a measure of consistency of the comparisons a_{lk} entered in the matrix. For complete consistency, the maximum eigenvalue $\lambda_{\max}$, should be equal to the order of the matrix H. The level of consistency of a given pairwise comparison matrix can be measured by a parameter called Consistency Ratio (CR), defined as the ratio of the Consistency Index CI = $(\lambda_{max} - H)/(H-1)$ and the Random Index (RI), which is the statistically averaged CI of randomly generated matrices of order H with entries artificially forced to be consistent. A CR of 0.10 or less is considered acceptable according to (Saaty, 1980).

3) The term consistency has a peculiar meaning; in the context of AHP it is used to define the internal degree of coherence between the expert based pairwise comparisons a_{lk}. In case of large inconsistencies in a matrix (typically when H is large), revise its entries by redoing the judgments on the individual pairwise comparisons a_{lk} or by forcing the values to be mathematically consistent. For all the details on the revision process see (Saaty, 1980). More recently (Benítez et al., 2011) proposed a novel procedure for achieving matrix consistency in a closed form through a linearization procedure.

4) Once all the priority vectors are available, multiply them appropriately through the branches of the hierarchy (by means of matrix multiplication) to determine the overall weights of the bottom-level alternatives with regards to the previously defined top goal. For example, referring to our simple example of a four-level hierarchy with three alternatives (A,B,C) (Fig. A1), the final ranking of the alternatives $\overline{W} = [W_A, W_B, W_C]^T$ is obtained solving Eq. (A.1). Where at first the matrix collecting the priorities of alternatives (A,B,C) respect the sub-criteria (Y_1,Y_2,Y_3) is multiplied with the one collecting the priorities of sub-criteria respect the criteria (X_1,X_2). The resulting matrix is multiplied with the priorities of the two criteria (X_1,X_2) respect the top goal $\overline{W}_{GOAL} = [W_{X_1}, W_{X_2}]^T$, to obtain the vector $\overline{W}$ as final ranking indicator for the alternatives (A,B,C).

$$\left(\begin{bmatrix} W_{AY_1} & W_{AY_2} & W_{AY_3} \\ W_{BY_1} & W_{BY_2} & W_{BY_3} \\ W_{CY_1} & W_{CY_2} & W_{CY_3} \end{bmatrix}\begin{bmatrix} W_{Y_1X_1} & W_{Y_1X_2} \\ W_{Y_2X_1} & W_{Y_2X_2} \\ W_{Y_3X_1} & W_{Y_3X_2} \end{bmatrix}\right)\begin{bmatrix} W_{X_1} \\ W_{X_2} \end{bmatrix} = \begin{bmatrix} W_A \\ W_B \\ W_C \end{bmatrix} \tag{A.1}$$

The major advantage of the AHP method relies in the adoption of pairwise comparisons, that are a simple and intuitive way of expressing judgments on the relative importance of the different constituents of the problem hierarchy. Moreover, it enables the possibility of checking the consistency in each expert based evaluation. This latter is a very important feature as it provides an internal qualification of the method, assuring the quality of the results after the evaluation of the input quantities. For these reasons the AHP has found extensive application in many quite different industrial sectors from company logistic organization to water resources management.

Table A2.
Table A3.

Appendix B

In this Appendix, we explain in detail the rationale behind the adopted metrics ϕ_j presented in Table 4. Without loss of generality, for clarity and simplicity sake, let us assume that only $N = 2$ experimental dataset EXP$_n$, $n = 1, 2$, are to be compared, where $\overline{x}_n$ is the set of the n-th experiment control variables that are chosen as input quantities for the metrics ϕ_j and the vector $\overline{x}^{app}$ collects the same variables in the application domain.

The coverage of the application domain (criterion C.C.1) consists in the ratio on the volumes of experimental and application input domain Eq. (B.1) as sketched in Fig. B1.

$$\Delta_n = \frac{\Delta\overline{x}_n}{\Delta\overline{x}^{app}} \tag{B.1}$$

The criterion C.C$_{2.1}$ related to the uniformity of the distribution of the experimental input domain has been evaluated using a metric based on the chi squared test Eq. (B.2):

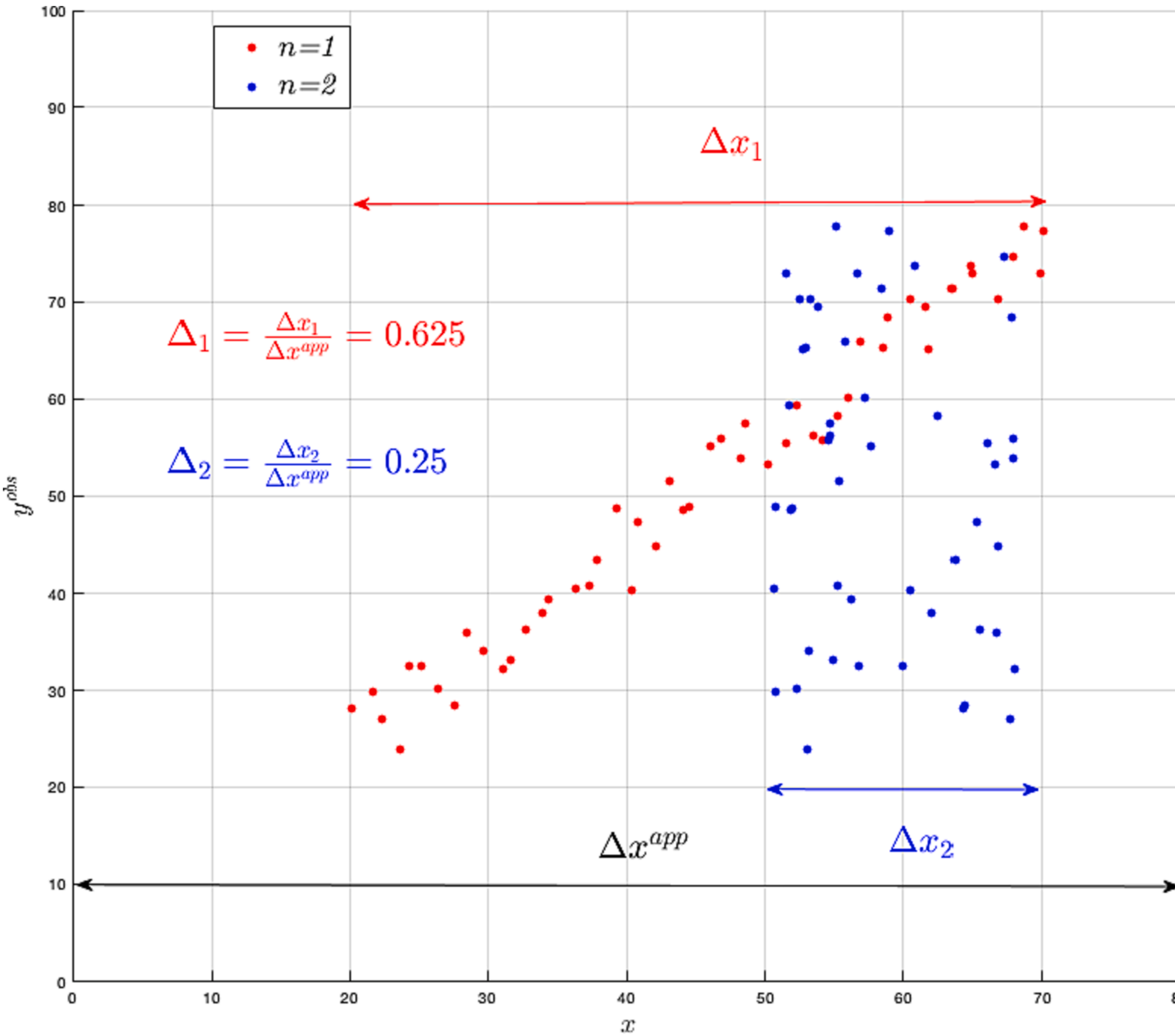


Fig. B1. An intuitive sketch of the procedure to calculate $\phi_{C.C_1}$.

$$\chi_n^2 = \sum_{u=1}^{N_u} \frac{\left(O_{x_u}^n - E_u\right)^2}{E_u} \tag{B.2}$$

where, $u = 1, 2, \ldots, N_u$ is the index that count the discretization of the experimental input domain x_n in N_u equally spaced bins (N_u could be either the number of EXP$_n$ or a fixed quantity) and $O_{x_u}^n$ is the observed frequency of experimental points falling in bin u while E_u is the expected frequency of samples falling in the same bin when the experimental points are drawn from a uniform distribution on the whole domain. As can be observed in Fig. B2 the lower the uniformity of experimental points distribution the higher is the χ^2 value, for this reason in Section 2.2 is provided a formulation Eq. (8) for sorting the scores in descending magnitude order.

Instead, for obtaining a quantitative measure on the uniformity (i.e., the number of elements having a specified characteristic) of some desired features of the n-th experimental database, we have adopted the formulation of Shannon Entropy Eq. (B.3) from Information theory (Shannon, 1948), whereby $q = 1, 2, \ldots, Q$ is the number of chosen features and P_q^n is the probability of drawing at random from the database of the n-th experiment a point belonging to the q-th feature:

$$S_n = \sum_{q=1}^{Q} \log_Q \left(\frac{1}{P_q^n}\right) P_q^n \tag{B.3}$$

The adoption of this metric based on entropy could be beneficial in many T-H applications, indeed we have just used it to evaluate the uniformity of single and two-phase inlet conditions ($Q = 2$) in the experimental databases, but more generally it could be applied also to the evaluation of flow regimes and adimensional numbers. As can be seen in Fig. B3, the highest possible value of entropy is one when the number of points belonging to each feature is equal, while the higher the imbalance the lower the entropy.

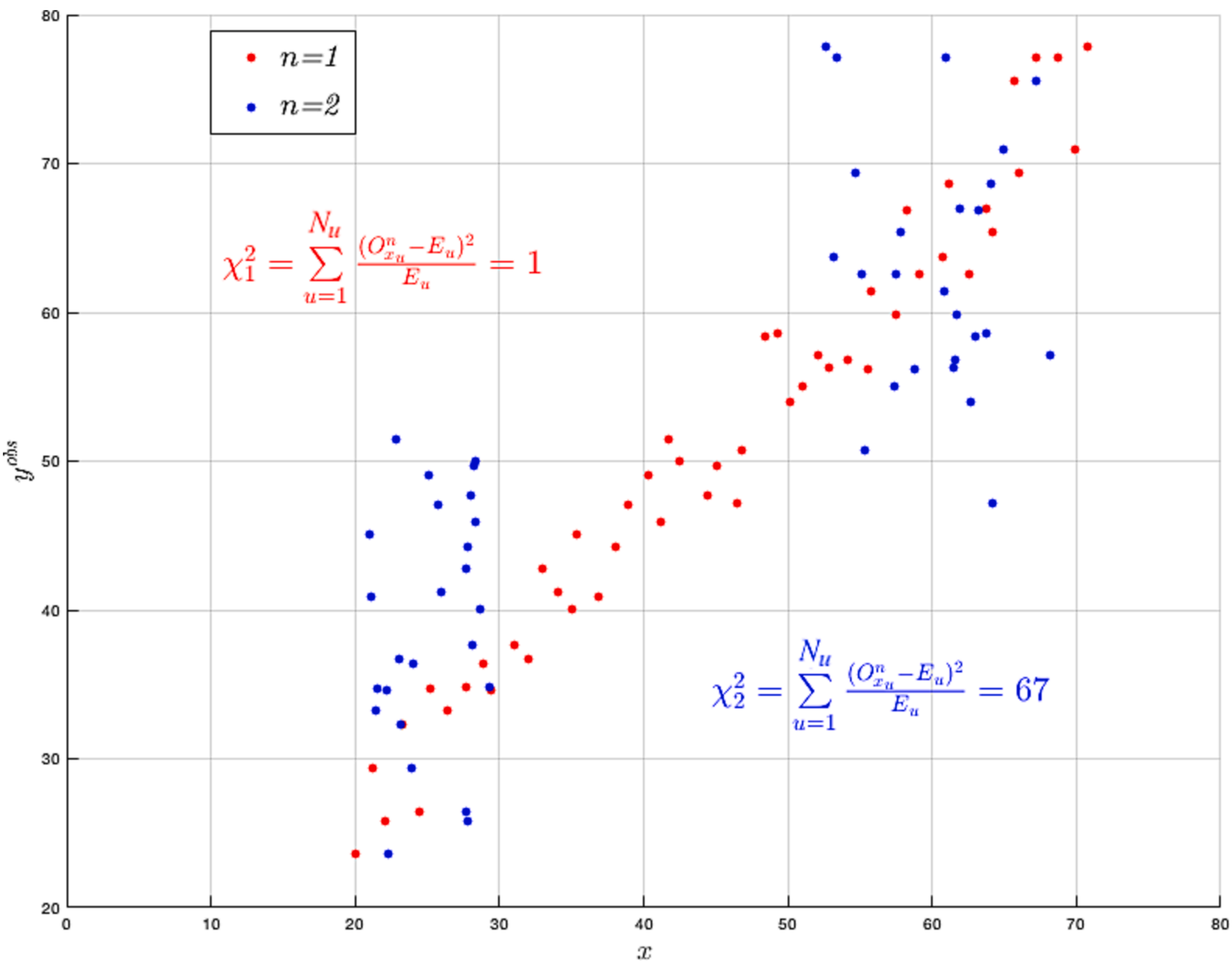
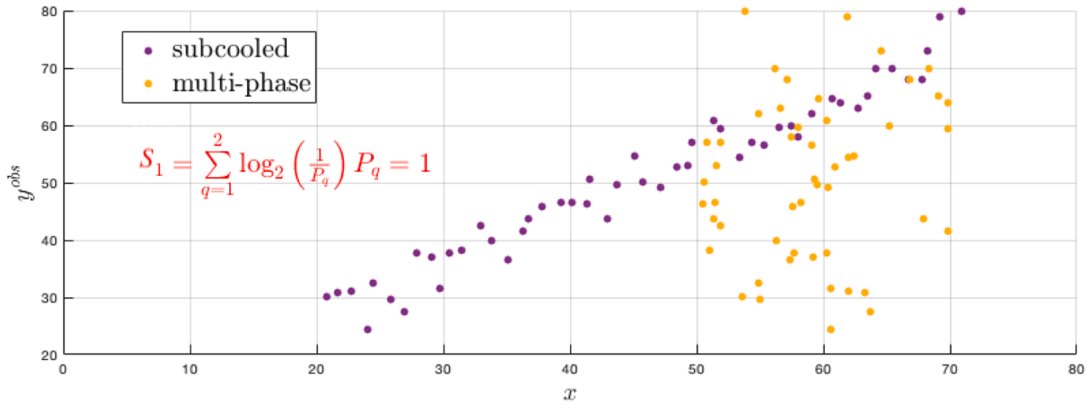
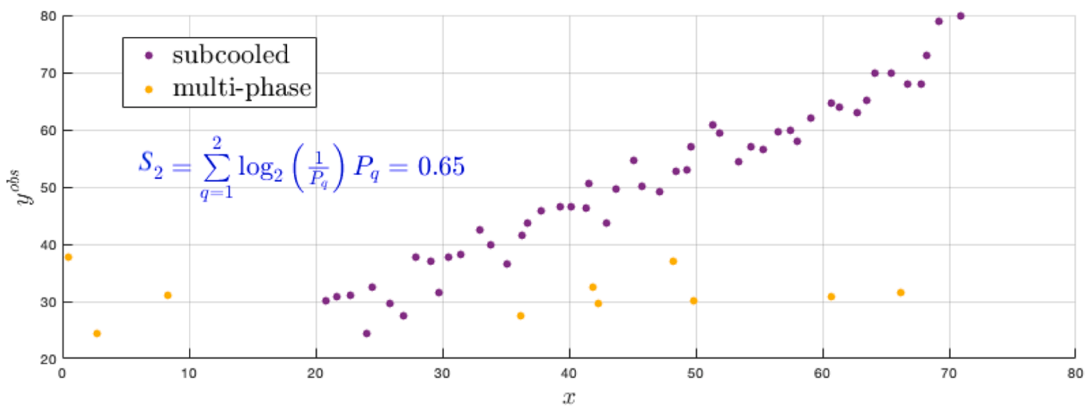


Fig. B2. An intuitive sketch of the behavior of metric $\phi_{C.C_{2.1}}$.



(a) Entropy measure for experimental database n = 1.



(b) Entropy measure for experimental database n = 2.

Fig. B3. A graphical example for the calculation of metric $\phi_{C.C_{2.3}}$.

References

Baccou, J., Zhang, J., Fillion, P., Damblin, G., Petruzzi, A., Mendizábal, R., Reventós, F., et al., 2019. Development of good practice guidance for quantification of thermal-hydraulic code model input uncertainty. Nucl. Eng. Des. 384 https://doi.org/10.1016/j.nucengdes.2019.110173.

Baccou, J., Zhang, J., Fillion, P., Damblin, G., Petruzzi, A., Mendizábal, R., Reventos, F., et al., 2020. SAPIUM: a generic framework for a practical and transparent quantification of thermal-hydraulic code model input uncertainty. Nucl. Sci. Eng. 194 (8–9), 721–736. https://doi.org/10.1080/00295639.2020.1759310.

Baccou, J. 2023. 'A Systematic Approach for the Adequacy Analysis of a Set of Experimental Databases: Application in the Framework of the ATRIUM Project'. Nuclear Engineering and Design (Submitted).

Benítez, J., Delgado-Galván, X., Izquierdo, J., Pérez-García, R., 2011. Achieving matrix consistency in AHP through linearization. App. Math. Model. 35, 4449–4457. https://doi.org/10.1016/j.apm.2011.03.013.

Bersano, A., Mascari, F., Porfiri, M.T., Maccari, P., Bertani, C., 2020. Ingress of coolant event simulation with TRACE code with accuracy evaluation and coupled DAKOTA uncertainty analysis. Fusion Eng. Des. 159 https://doi.org/10.1016/j.fusengdes.2020.111944.

Bouyssou, D. 2001. 'Outranking Methods'. In Encyclopedia of Optimization, edited by Panos M Floudas Christodoulos A. and Pardalos, 1919–25. Boston, MA: Springer US. 10.1007/0-306-48332-7_376.

Brereton, R.G., Lloyd, G.R., 2016. Re-evaluating the role of the mahalanobis distance measure. J. Chemom. 30 (4), 134–143. https://doi.org/10.1002/CEM.2779.

Cagno, E., Caron, F., Mancini, M., Ruggeri, F., 2000. Using AHP in determining the prior distributions on gas pipeline failures in a robust bayesian approach. Reliab. Eng. Syst. Saf. 67 (3), 275–284. https://doi.org/10.1016/S0951-8320(99)00070-8.

Cagno, E., Caron, F., Perego, A., 2001. Multi-criteria assessment of the probability of winning in the competitive bidding process. International Journal of Project Management 19 (6), 313–324. https://doi.org/10.1016/S0263-7863(00)00020-X.

D'Auria, F., Camargo, C., Mazzantini, O., 2012. The best estimate plus uncertainty (BEPU) approach in licensing of current nuclear reactors. Nucl. Eng. Des. 248 (July), 317–328. https://doi.org/10.1016/j.nucengdes.2012.04.002.

D'Auria, F., Galassi, G., 1998. Code validation and uncertainties in system thermal hydraulics. Prog. Nucl. Energy 33. https://doi.org/10.1016/S0149-1970(97)00097-8.

D'Auria, F., Galassi, G., Mazzantini, O., 2022. The best estimate plus uncertainty approach in licensing of atucha II. Pressurized Heavy Water Reactors 8, 51–204.

D'Onorio, M., Maggiacomo, A., Giannetti, F., Caruso, G., 2022. Analysis of Fukushima Daiichi Unit 4 Spent Fuel Pool Using MELCOR. In: Journal of Physics: Conference Series, Vol. 2177. IOP Publishing Ltd. https://doi.org/10.1088/1742-6596/2177/1/012020

Fedrizzi, M., Brunelli, M., Caprila, A., 2020. The linear algebra of pairwise comparisons. Int. J. Approx. Reason. 118 (March), 190–207. https://doi.org/10.1016/j.ijar.2019.12.009.

Ghione, A. 2023. 'Application of SAPIUM Guidelines to Input Uncertainty Quantification: The ATRIUM Project'. In 20th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH20).

Hou, J., Avramova, M., Ivanov, K., 2020. Best-estimate plus uncertainty framework for multiscale, multiphysics light water reactor core analysis. Science and Technology of Nuclear Installations. https://doi.org/10.1155/2020/7526864.

IAEA. 2014. 'Progress in Methodologies for the Assessment of Passive Safety System Reliability in Advanced Reactors'.

Lin, L., Montanari, N., Prescott, S., Sampath, R., Bao, H., Dinh, N., 2020. Adequacy evaluation of smoothed particle hydrodynamics methods for simulating the external-flooding scenario. Nucl. Eng. Des. 365 https://doi.org/10.1016/j.nucengdes.2020.110720.

Malczewski, J., Rinner, C., 2015. Multicriteria Decision Analysis in Geographical Information Science. Springer.

Marquès, M., Pignatel, J.F., Saignes, P., D'auria, F., Burgazzi, L., Müller, C., Bolado-Lavin, R., Kirchsteiger, C., La Lumia, V., Ivanov, I., 2005. Methodology for the reliability evaluation of a passive system and its integration into a probabilistic safety assessment. Nucl. Eng. Des. 235, 2612–2631. https://doi.org/10.1016/j.nucengdes.2005.06.008.

Mascari, F., Nakamura, H., Umminger, K., De Rosa, F., D'Auria, F. 2015. 'Scaling Issues for the Experimental Characterization of Reactor Coolant System in Integral Test Facilities and Role of System Code as Extrapolation Tool'. In NURETH- 16, Chicago (IL), August 30-September 4.

NEA. 1998. 'Report On the Uncertainty Methods Study, Nuclear Safety, NEA/CSNI/R(97)35/VOLUME 1'.

NEA 2011. 'BEMUSE Phase VI Report: Status Report on the Area, Classification of the Methods, Conclusions and Recommendations, Nuclear Safety, NEA/CSNI/R(2011) 4'.
NEA 2016. 'PREMIUM, a Benchmark on the Quantification of the Uncertainty of the Physical Models in the System Thermal-Hydraulic Codes: Methodologies and Data Review, Nuclear Safety, NEA/CSNI/R(2016)9'.
Nusret, A. et al. 1993. 'Separate Effects Test Matrix for Thermal-Hydraulic Code Validation Volume 1- Phenomena Characterisation and Selection of Facilities and Tests. Volume 2-Facility and Experiment Characteristics.'.
O'Hagan, A., Buck, C.E., Alireza Daneshkhah, J., Richard Eiser, H., Paul Garthwaite, J., David Jenkinson, E., Oakley, J., Rakow, T., 2006. Eliciting and Fitting a Parametric Distribution. In: Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, pp. 121–151.
Oberkampf, W.L., Trucano, T.G. 2007. 'Verification and Validation Benchmarks'.
Oberkampf, W.L., Helton, J.C., Joslyn, C.A., Wojtkiewicz, S.F., Ferson, S., 2004. Challenge problems: uncertainty in system response given uncertain parameters. In Reliability Engineering and System Safety 85, 11–19. https://doi.org/10.1016/j.ress.2004.03.002.
Oberkampf, W.L., et al. 2007. 'Predictive Capability Maturity Model for Computational Modeling and Simulation'.
OECD. 2011. 'Final Data Report of ROSA-2/LSTF Test 1 (Hot Leg Intermediate Break LOCA IB-HL-01 in JAEA)'.
Petruzzi, A., D'Auria, F., 2008. Thermal-hydraulic system codes in nulcear reactor safety and qualification procedures. Science and Technology of Nuclear Installations. https://doi.org/10.1155/2008/460795.
Ramanathan, R., Ganesh, L.S., 1994. Group preference aggregation methods employed in AHP: an evaluation and an intrinsic process for deriving members Weightages. European Journal of Operational Research 79 (2), 249–265. https://doi.org/10.1016/0377-2217(94)90356-5.
Roma, G., Di Maio, F., Bersano, A., Pedroni, N., Bertani, C., Mascari, F., Zio, E., 2021. A bayesian framework of inverse uncertainty quantification with principal component analysis and kriging for the reliability analysis of passive safety systems. Nucl. Eng. Des. 379, 111230 https://doi.org/10.1016/j.nucengdes.2021.111230.
Roma, G., Antonello, F., Di Maio, F., Pedroni, N., Zio, E., Bersano, A., Bertani, C., Mascari, F., 2022. Passive safety systems analysis: a novel approach for inverse uncertainty quantification based on stacked sparse autoencoders and kriging metamodeling. Prog. Nucl. Energy 148, 104209. https://doi.org/10.1016/j.pnucene.2022.104209.
Rosenbloom, E.S., 1997. A probabilistic interpretation of the final rankings in AHP. Eur. J. Oper. Res. 96 (2), 371–378. https://doi.org/10.1016/S0377-2217(96)00049-5.
Rousseau, J. C. 1987. 'Flashing Flow '. In Multiphase Science and Technology, 3: 378–89. Hemisphere Publishing Coorporation, USA.
Roy, B., 1996. Multicriteria Methodology for Decision Aiding. Kluwer Academic Publishers.
Saaty, T.L., 1980. The Analytic Hierachy Process. McGraw-Hill.
Saaty, T.L., 1990. How to make a decision: the analytic hierarchy process. Eur. J. Oper. Res. 48 (1), 9–26. https://doi.org/10.1016/0377-2217(90)90057-I.
Saaty, T.L., 2003. Decision aiding decision-making with the ahp: why is the principal eigenvector necessary. Eur. J. Oper. Res. 145, 85–91 www.elsevier.com/locate/dsw.
Saaty, T.L., Luis G.V. 2012. Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, International Series in Operations Research & Management Science . Vol. 175. Springer .
Sajjad Zahir, M., 1991. Incorporating the uncertainty of decision judgements in the analytic hierarchy process. Eur. J. Oper. Res. 53 (2), 206–216. https://doi.org/10.1016/0377-2217(91)90135-I.
Saltelli, A., Tarantola, S., Chan, K., 1999. A role for sensitivity analysis in presenting the results from MCDA studies to decision makers. Decis. Anal 8, 139–145. https://doi.org/10.1002/(SICI)1099-1360(199905)8:3.
Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice. Wiley.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423. https://doi.org/10.1002/J.1538-7305.1948.TB01338.X.
Sokolowski, L., Kozlowski, T. 2012. 'Assessment of Two-Phase Critical Flow Models Performance in RELAP5 and TRACE against Marviken Critical Flow Test'. Washington, USA.
Sozzi, G.L., Sutherland, W.A. 1975. 'Critical Flow of Saturated and Subcooled Water at High Pressure, Report NEDO-13418'. San Jose, USA.
Unal, C., Williams, B., Hemez, F., Atamturktur, S.H., McClure, P., 2011. Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors. Nucl. Eng. Des. 241 (5), 1813–1833. https://doi.org/10.1016/J.NUCENGDES.2011.01.048.
Vargas, L.G., 1990. An overview of the analytic hierarchy process and its applications. Eur. J. Oper. Res. 48 (1), 2–8. https://doi.org/10.1016/0377-2217(90)90056-H.
Wilson, G.E., Boyack, B.E., 1998. The role of the PIRT process in experiments, code development and code applications associated with reactor safety analysis. Nucl. Eng. Des. 186 (1–2), 23–37. https://doi.org/10.1016/S0029-5493(98)00216-7.
Wu, X.u., Xie, Z., Alsafadi, F., Kozlowski, T., 2021. A comprehensive survey of inverse uncertainty quantification of physical model parameters in nuclear system thermal-hydraulics codes. Nucl. Eng. Des. 384 https://doi.org/10.1016/j.nucengdes.2021.111460.
Yu, Y., Tong, J., Zhao, J., Di Maio, F., Zio, E., 2010. Multi-Experts Analytic Hierarchy Process for the Sensitivity Analysis of Passive Safety Systems. In: 10th International Probabilistic Safety Assessment & Management Conference – PSAM 10. Seattle, USA, pp. 1–12.
Yurko, J.P., Buongiorno, J., 2012. 'Quantitative Phenomena Identification and Ranking Table (QPIRT) for Bayesian Uncertainty Quantification'. In International Congress on Advances in National Power Plants (ICAPP '12). American Nuclear Society, Chicago, IL.
Zeng, Z., Francesco Di M., Enrico Z., Rui K. 2017. 'A Hierarchical Decision-Making Framework for the Assessment of the Prediction Capability of Prognostic Methods'. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 231 (1): 36–52. 10.1177/1748006X16683321/ASSET/IMAGES/LARGE/10.1177_1748006X16683321-FIG9.JPEG.
Zio, E., 1996. On the use of the analytic hierarchy process in the aggregation of expert judgments. Reliab. Eng. Syst. Saf. 53, 127–138.
Zio, E., Cantarella, M., Cammi, A., 2003. The analytic hierarchy process as a systematic approach to the identification of important parameters for the reliability assessment of passive systems. Nucl. Eng. Des. 226, 311–336. https://doi.org/10.1016/S0029-5493(03)00211-5.