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

J. Baccou[1], T. Glantz[1], A. Ghione[2], L. Sargentini[2], P. Fillion[2], G. Damblin[3], R. Sueur[4], B. Iooss[4], J. Fang[5], J. Liu[5], C. Yang[5], Y. Zheng[5], A. Ui[6], M. Saito[6], R. Mendizábal Sanz[7], A. Bersano[8], F. Mascari[8], T. Skorek[9], L. Tiborcz[9], Y. Hirose[10], T. Takeda[10], H. Nakamura[10], C. Choi[11], J. Heo[11], A. Petruzzi[12], K. Zeng[12], Z. Xie[13], X. Wu[13], H. Eguchi[14], F. Pangukir[15], P. Breijder[15], S. Franssen[15], G. Perret[16], I. D. Clifford[16], T. M. Coscia[17], F. Di Maio[17], E. Zio[17,18], N. Pedroni[19], J. Zhang[20], J. Freixa[21], F. Rizzo[22], C. Ciurluini[22], F. Giannetti[22], M. Adorni[23]

[1]Institut de Radioprotection et de Sûreté nucléaire (IRSN), Centre de Cadarache, 13115 St Paul-Lez-Durance, France; jean.baccou@irsn.fr, tony.glantz@irsn.fr

[2]Université Paris-Saclay, CEA, Service de Thermohydraulique et de Mécanique des Fluides, 91191, Gif-sur-Yvette, France ; alberto.ghione@cea.fr, lucia.sargentini@cea.fr, philippe.fillion@cea.fr

[3]Université Paris-Saclay, CEA, Service de Génie Logiciel pour la Simulation, 91191, Gif-sur-Yvette, France ; guillaume.damblin@cea.fr

[4]EDF R&D, Département PRISME, 6 Quai Watier,78401 Chatou, France; roman.sueur@edf.fr, bertrand.iooss@edf.fr

[5]China Nuclear Power Engineering Co., Ltd. (CNPE), Beijing 100840, China; fangjun@cnpe.cc, liujt@cnpe.cc, yangcs@cnpe.cc, zhengyt@cnpe.cc

[6]Nuclear Risk Research Center (NRRC), Central Research Institute of Electric Power Industry (CRIEPI), 2-6-1, Nagasaka, Yokosuka-shi, Kanagawa, 240-0196 Japan; ui@criepi.denken.or.jp, mi-saito@criepi.denken.or.jp

[7]Consejo de Seguridad Nuclear (CSN), 28040 Madrid, Spain; rmsanz@csn.es

[8]ENEA, FSN-SICNUC-SIN, via Martiri di Monte Sole 4, Bologna, 40129, Italy ; andrea.bersano@enea.it, fulvio.mascari@enea.it

[9]Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) GmbH, Forschungszentrum , 85748 Garching, Germany; tomasz.skorek@grs.de, livia.tiborcz@grs.de

[10]Japan Atomic Energy Agency (JAEA), Tokai-mura, Naka-gun, Ibaraki-ken 319-1195, Japan; hirose.yoshiyasu@jaea.go.jp, takeda.takeshi@jaea.go.jp, nakamura.hideo@jaea.go.jp

[11]Korean Atomic Energy Research Institute (KAERI), 989-111 Daedeok-Daero, Yuseong-Gu, Daejeon 305-353, Republic of Korea; cwchoi@kaeri.re.kr, jheo@kaeri.re.kr

[12] N.IN.E. - Nuclear and INdustrial Engineering S.r.l., Via della Chiesa XXXII, 759 - 55100 Lucca, Italy; a.petruzzi@nineeng.com, k.zeng@nineeng.com

[13]Department of Nuclear Engineering, North Carolina State University, Burlington Engineering Laboratories, 2500 Stinson Drive, Raleigh, NC 27695, USA; xwu27@ncsu.edu, zxie22@ncsu.edu

[14]Nuclear Regulation Authority (NRA), 1-9-9, Roppongi, Minato-ku, 106-8450 Tokyo, Japan; eguchi_hiroshi_v6d@nra.go.jp

[15] Nuclear Research and Consultancy Group (NRG), Westerduinweg 3, 1755 LE Petten, Netherlands; pangukir@nrg.eu, breijder@nrg.eu, franssen@nrg.eu

[16]Paul Scherrer Institute (PSI), 111 Forschungstrasse, 5232 Villigen, Switzerland; gregory.perret@psi.ch, ivor.clifford@psi.ch

[17]POLIMI, Energy Department, Politecnico di Milano, Milano, Italy; thomasmatteo.coscia@polimi.it, francesco.dimaio@polimi.it, enrico.zio@polimi.it

[18]MINES Paris-PSL, Centre de Recherche sur les Risques et les Crises (CRC), Sophia Antipolis, France

[19]POLITO, Energy Department, Politecnico di Torino, Torino, Italy; nicola.pedroni@polito.it

[20]Tractebel (ENGIE), Boulevard Simon Bolivar 36, 1000 Brussels, Belgium; jinzhao.zhang@tractebel.engie.com
[21]Universitat Politècnica de Catalunya (UPC), Avda. Diagonal 647, 08028 Barcelona, Spain; jordi.freixa-terradas@upc.edu
[22]Sapienza University of Rome, Department of Astronautical, Electrical and Energy Engineering
Corso Vittorio Emanuele II, 244, 00186 Rome, Italy; fr.rizzo@uniroma1.it, cristiano.ciurluini@uniroma1.it, fabio.giannetti@uniroma1.it
[23]OECD Nuclear Energy Agency (NEA), 46 quai Alphonse Le Gallo, 92100 Boulogne-Billancourt, France; martina.adorni@oecd-nea.org

Abstract

In the Best-Estimate Plus Uncertainty (BEPU) framework, the use of best-estimate code requires to go through a Verification, Validation and Uncertainty Quantification process (VVUQ). The relevance of the experimental data in relation to the physical phenomena of interest in the VVUQ process is crucial. Adequacy analysis of selected experimental databases addresses this problem. The outcomes of the analysis can be used to select a subset of relevant experimental data, to encourage designing new experiments or to drop some experiments from a database because of their substantial lack of adequacy. The development of a specific transparent and reproducible approach to analyze the relevance of experimental data for VVUQ still remains open and is the topic of this contribution.
In this paper, the concept of adequacy initially introduced in the OECD/NEA SAPIUM (Systematic APproach for model Input Uncertainty quantification Methodology) activity is formalized. It is defined through two key properties, called representativeness and completeness, that allows considering the multifactorial dimension of the adequacy problem. A new systematic approach is then proposed to analyze the adequacy of a set of experimental databases. It relies on the introduction of two sets of criteria to characterize representativeness and completeness and on the use of multi-criteria decision analysis method to perform the analysis. Finally, the approach is applied in the framework of the new OECD/NEA ATRIUM activity which includes a set of practical IUQ exercises in thermal-hydraulics to test the SAPIUM guideline in determining input uncertainties and forward propagating them on an application case . It allows evaluating the adequacy of eight experimental databases coming from the Super Moby-dick, Sozzi-Sutherland and Marviken experiments and identifying the most adequate ones.

Keywords

Adequacy analysis, experimental databases, representativeness analysis, completeness analysis, multicriteria decision problem, ATRIUM activity

1. Introduction

Experimental data play a key role in the Verification, Validation and Uncertainty Quantification (VVUQ) process [1] for scientific computing and for the application of artificial intelligence (IA). In the framework of nuclear applications, the adequacy of experimental data (i.e., the relevance of the information they bring) for the intended use is important for Best Estimate Plus Uncertainty (BEPU) simulations [2] adopted for safety assessment. A careful analysis and selection of experimental databases appears as a key step in various systematic approaches used in regulatory requirements and industrial applications for BEPU methodology development ([3], [4]). The importance of experiments has been also recently shown for input uncertainty quantification of the physical models implemented in thermal-hydraulics codes. The OECD/NEA PREMIUM (Post BEMUSE Reflood Models Input Uncertainty Methods) activity [5] first illustrated the strong influence of the selection of the experimental databases used for quantification and validation of input uncertainties. The lack of a structured and clear approach to explain the selection of experimental databases was identified as one of the main reasons that prevented from reaching a consensus on the quantified input uncertainties. To further emphasize the crucial role of experimental database selection, the OECD/NEA SAPIUM guideline [6] introduced a new systematic approach for input uncertainty quantification in which a specific step is devoted to the experimental data and their adequacy analysis.

An adequacy analysis should rely on a transparent and reproducible methodology with the objective to support the analysts in the choice of the experimental databases for the study of interest. It can help to summarize as objectively as possible the different steps of the analysis in order to reproduce the results and understand the choices made by the analysts. It can also contribute to identifying the sources of discrepancies between the results obtained by different analysts.

One of the first attempts to use an adequacy analysis in the literature is described in the OECD/NEA CSNI Code Validation Matrix (CCVM) of Separate Effect Tests (SET) and Integral Effect Tests (IET). For example, the methodology to establish a collection of SET, i.e., the SET matrix, relevant for Loss of Coolant Accident (LOCA) application and transients in Light-Water Reactors (LWRs) is summarized in [7]. The IET matrix [8] was constructed by considering several factors including the similarity of the SET/IET with a selected reactor case and the capability of the SET/IET to cover the physical phenomena relevant in the reactor case. These factors[1] can be used to define the adequacy criteria. Several authors have also developed objective criteria to evaluate the predictive maturity of computational modeling and simulation efforts ([9], [10]). Predictive maturity indexes can then be used to design new experiments [11] or combined with the Analytical Hierarchical Process (AHP, [12]) - as illustrated in [13] - to construct a systematic approach for code validation assessment. However, these methodologies are not specifically focused on the adequacy of experimental database. Other contributions on the topic of adequacy analysis can be found in metrology [14] where the concepts of measurement precision and trueness are central. However, the adequacy only refers here to the quality of measurement and not the quality or relevance of a database for an intended use. Finally, classical statistical indicators used for the design of numerical experiments ([15], [16]) could bring some insights on the adequacy of a database. More precisely, they are well suited to evaluate the distribution of a set of points in an input space associated to the parameters of a simulation model. Several of them exploit a quantity called the discrepancy [17] that measures how far the point set distribution is from a target uniform distribution that allows covering the whole input space without over/under weighting some specific regions. Interpreting an experimental database as a collection of experimental tests, i.e. a set of points in a physical space, it is then straightforward to apply these indicators to evaluate if the tests cover uniformly the physical space of interest, which contributes to the adequacy analysis. However, these indicators remain limited to the analysis of the spatial distribution of a set of experimental tests and do not allow integrating qualitative features and physical expertise in the selection of experimental tests. Therefore, up to now, there is no fully transparent and reproducible approach to perform the adequacy assessment of a set of experimental databases in BEPU applications.

In this paper, an adequacy analysis approach is proposed. Its construction first requires defining how to characterize the adequacy of an experimental database. Then, specific mathematical tools are introduced to handle the analysis of a set of databases. These tools come from the so-called Multicriteria Decision Analysis

[1] Interested readers can find more details on these factors p 392 of [7] and p 29 of [8]

field (MCDA) [18] and are used in a wide range of domains (e.g., reference examples in Chapter 3 of [18]) including the nuclear industry ([19], [20], [21]). The originality of our approach stands in the following points:

- it is constructed for the analysis of experimental databases and combines both an objective description of the adequacy and a numerical procedure that allows aggregating expert's knowledge on adequacy;
- it relies on two generic key properties of a database (called representativeness and completeness). Therefore, the approach can be extended to different types of applications including the training of machine learning models for example;
- it is flexible enough to allows using a large variety of MCDA methods in order to integrate the expert's knowledge.

The paper is organized as follows. In Section 2, the concept of adequacy is presented and a systematic approach to analyze the adequacy of a set of experimental databases is described. Section 3 is devoted to the application of this approach for input uncertainty quantification as performed in the OECD/NEA ATRIUM (Application Tests for Realization of Inverse Uncertainty quantification and validation Methodologies in thermal-hydraulics) activity. It is, to our knowledge, the first time that a systematic approach for a rigorous adequacy analysis is applied and shared among engineers in the BEPU domain. Some further developments are proposed in Section 4, and Section 5 provides some synthesis and conclusions as well as the remaining open issues.

2. The methodology of adequacy analysis

In this section, the definition of the adequacy of a database and a summary of the steps to obtain a systematic approach for its analysis are described. For the remaining of the paper, a database has to be understood as a set of experimental tests i.e., a set of points in the physical space. These points can come from one experiment or multiple ones.

2.1. Definition of adequacy

Following the SAPIUM guideline [22], adequacy is defined by two main properties of a given database. The first one is related to the ability of each experimental test to provide relevant information for the problem of interest and is called *representativeness*. The definition of the information relevance relies on the introduction of a set of criteria that characterizes the different aspects of the relevance and depends on the problem of interest. Examples of these criteria are the geometrical fidelity with respect to a target application facility (considering for example the similarity in the diameter, length… using a scaling approach), the capability to cover a physical phenomenon of interest or the measurement quality. We refer to Section 3.3.1 for a more exhaustive list of criteria in the case of input uncertainty quantification. Moreover, once the relevance has been specified, important questions are related to its evaluation: when is information relevance considered high/low/ or medium? How should it be assessed? They can be addressed by exploiting so-called MCDA [18] methods that are described and illustrated in Sections 2.2.2 and 3.3.2. Finally, it is important to mention that representativeness analysis can be performed separately for every single experimental test of a database. However, in order to reduce the number of tests to analyze, representativeness can be also evaluated for the whole database. In this paper, the second approach is used.

The second property refers to the ability of a set of experimental tests to cover the physical conditions of the problem under study, i.e., to include experimental tests allowing to explore the whole physical space of interest. It is referred as *completeness*. Therefore, it can only be evaluated for a group of experiments and is typically performed for the whole database. The strict definition of completeness also depends on the problem of interest and relies on the introduction of criteria. For example, in the case of a bi-dimensional physical domain defined by two variables (T,P), assuming that a database includes four tests represented by four points in the physical domain (T_1,P_1), (T_2,P_2), (T_3, P_3) and (T_4,P_4), these criteria can be focused on the spatial distribution of these points (are they lying on a grid? Are they clustered in a specific region?) but also on the surface that they cover. We refer to Sections 2.2.1 and 3.3.1 for more details on completeness criteria. Similarly to the representativeness analysis, the completeness analysis can then be performed with MCDA methods. Since it operates in the physical space, statistical indicators coming for the design of numerical experiments ([15], [16]) can be more appropriate for a quantitative analysis (see Section 3.3.2 for an example). It is also important to keep in mind that this type of analysis relies on a correct definition of the postulated physical domain. Section 4.1 provides a discussion on this topic.

2.2. Systematic approach for adequacy analysis

Based on the adequacy definition introduced in the previous section, a systematic approach is proposed to perform the adequacy analysis of a set of databases. Its development follows the two main steps that are displayed in Figure 1.

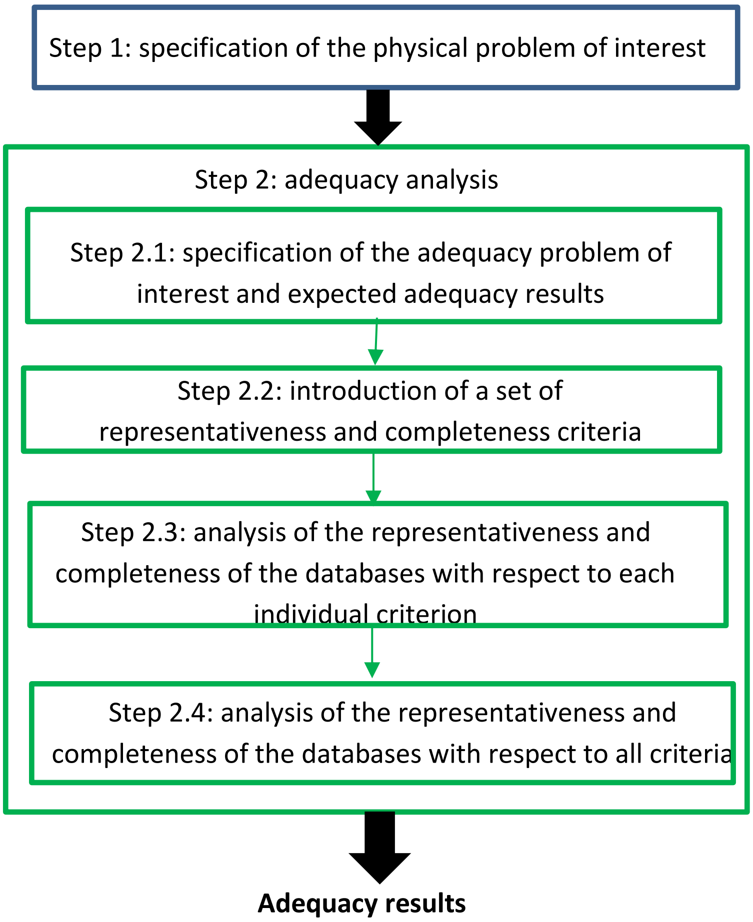

Figure 1: Sketch of the systematic approach for adequacy analysis of a set of databases.

The first step is common to any type of studies. The adequacy analysis is carried out in the second step. It concerns the specification of the practical problem, which includes the identification of the transient or

accident scenario of interest[2]. When the problem of interest involves simulations, the objective for which experimental databases are used and analyzed such as computer code validation or input uncertainty quantification should also be described.
The second step focuses on adequacy analysis. Step 2.1 first establishes the list of experimental databases to analyze and specifies the target application, e.g., a nuclear reactor or an integral test facility that will be the reference for the adequacy analysis. This sub-step also includes the definition of the adequacy analysis objectives. Among classical objectives, one can mention the identification of the most adequate database in the set of databases, the ranking of the databases within the set with respect to their adequacy or the computation of an adequacy score associated with each database. Obviously, an analysis based on the second, or third, objective can be used to address the previous or two previous ones, respectively. However, as it will become clear in the next section, an adequacy analysis method can be developed to specifically address each objective[3].

Steps 2.2 to 2.4 are devoted to the computation of the adequacy and are described in the following sections.

2.2.1. Representativeness and completeness criteria (Step 2.2)

Step 2.2 relies on the introduction of a set of representativeness and completeness criteria, which are denoted by C^r_i and C^c_i respectively. These criteria should possess some specific properties to represent the multicriteria nature of the problem [23]:

- each criterion should be qualitatively or quantitatively measured;
- each criterion should be meaningful for the analysis to perform;
- the set of criteria should cover all aspects of the problem;
- the set of criteria should be non-redundant to avoid the problem of double counting;
- the set of criteria should be minimal (i.e. kept as small as possible) to reduce the complexity of the analysis.

Once the representativeness and completeness criteria have been introduced, the analysis steps 2.3 and 2.4 refer to MCDA problem.

2.2.2. Multicriteria decision analysis problem (Steps 2.3 and 2.4)

The starting point of Step 2.3 is to establish a table called *decision matrix* that summarizes the analyst's preferences related to each database with respect to each criterion. Denoting by C_i the i^{th} criterion of representativeness or completeness, a score $s_{i,bj}$ is given to quantify the representativeness or the completeness of the database b_j with respect to criterion C_i. Moreover, an importance weight w_i is associated to each criterion C_i to represent the relative weight of the criterion i with respect to the others.

For instance, in the case of three criteria and four experimental databases, a decision matrix can formally be represented as in Table 1.

[2] Depending on the problem of interest, a scenario can involve more than one phenomenon that intervene in specific condition, or a single phenomenon or a group of phenomena over a diversity of conditions.
[3] For example, dedicated methods can be proposed to exhibit a ranking without computing adequacy scores.

Table 1: Example of a decision matrix in the case of 3 criteria and 4 experimental databases.

Criteria / Experimental databases	C_1	C_2	C_3
b_1	$s_{1,b1}$	$s_{2,b1}$	$s_{3,b1}$
b_2	$s_{1,b2}$	$s_{2,b2}$	$s_{3,b2}$
b_3	$s_{1,b3}$	$s_{2,b3}$	$s_{3,b3}$
b_4	$s_{1,b4}$	$s_{2,b4}$	$s_{3,b4}$
Importance weights	w_1	w_2	w_3

Different methods can be followed [23] to compute the scores and the importance weights. Two families of methods are considered in this work: rating and AHP [12] methods.

The rating method requires the analyst to estimate scores and importance weights based on a given scale: for example, 1 for low representativeness/completeness/importance, 2 for medium and 3 for high. When the rating scale is based on the database (or criterion) ranks, one speaks about ranking method. For example if N databases (or criteria) are available, the score of the most representative/complete/important is N and that of the least representative/complete/important is 1. Following the discussions of Section 2.1, it might be more appropriate in the case of completeness to perform a quantitative analysis. The score is then given by the value computed by the quantitative indicator and does not exploit a discrete scale chosen by expert judgement.

The AHP method relies on the theory of consistent matrices as proposed by Saaty ([12], [24]). It is based on pairwise comparisons to evaluate the importance of every criterion and the representativeness (completeness) of each experimental database. More precisely, the expert is asked to consider each pair of criteria or of databases and to provide associated importance or representativeness/completeness ratio following a numerical scale from 1 to 9 proposed by T. Saaty [12]. Then, a matrix calculus allows computing a vector of importance weights and a vector of representativeness (completeness) scores associated to each criterion.

Step 2.4 requires the aggregation of the information provided by the decision matrix. There exist several methods to perform such an aggregation. A first category is based on the computation of a unique synthesis indicator and leads to a score for each database. Different synthesis operators can be used such as the weighted average and are often used in the case of the decision matrix obtained by rating. The aggregation step of the AHP method also exploits the weighted average but operates on a decision matrix obtained by pairwise comparisons between databases/criteria. A second category of aggregation methods includes more flexible approaches representing the analysts' preferences but is not used in this paper. They rely on outranking relations. A well-known method of this category is the ELECTRE approach [25] that can be used to identify the most representative/complete database or to provide a ranking of databases without computing a score. An example of this last application is given in [26] where the representativeness of the validation databases of the IRSN SCANAIR code has been studied.

3. Adequacy analysis in the OECD/NEA ATRIUM activity

The systematic approach for adequacy analysis is illustrated in the framework of the ATRIUM activity. This section describes the activity focusing on the target application and experimental databases before applying the systematic approach on the first exercise of the activity.

3.1. The ATRIUM activity

In the past few decades, there has been an increasing interest in the use of BEPU methodologies for the safety analyses of Nuclear Power Plants (NPPs). The behavior of the reactor cooling system during operational and accidental scenarios is simulated with a best-estimate thermal-hydraulic system code (e.g., ATHLET, CATHARE, RELAP, SPACE, TRACE) considering realistic assumptions [27] and the impact of the uncertainties needs to be quantified. There exist three different sources of uncertainties [1]: the first is related to model inputs that include model parameters of closure laws, geometry, initial and boundary conditions, the second is associated to the numerical approximation error (space-time discretization or iterative convergence errors), the last one concerns model form and includes all assumptions, conceptualizations, abstractions, approximations and mathematical formulations on which the model relies. In this work, we focus on the first category. Such a quantification can be performed by comparison with experimental data and back-propagation of the discrepancy between experimental and simulation values. In such a case, it is usually referred to as Input (or Inverse) Uncertainty Quantification (IUQ).

Several activities have been devoted to the applicability and improvement of IUQ approaches. Some of them focus on the mathematical inverse methods that are used to back-propagate the information associated to the discrepancy between simulation results and experimental data ([28], [29], [30]). Recent contributions also rely on the interpretation of IUQ as a global approach that involves different elements to consider before applying inverse methods. This was the topic of the SAPIUM activity ([6], [31]). Such activity provided a generic framework similar to systematic approaches developed in regulatory requirements and industrial applications for BEPU methodology but specifically devoted to quantification and validation of model input uncertainty.

At the end of the SAPIUM activity, the need to prove the applicability of its guideline on practical IUQ exercises was highlighted. In such a context, the ATRIUM activity was proposed ([32], [33]). The scope of the activity is to perform practical IUQ exercises using the SAPIUM guideline and to:

- demonstrate the applicability of the guideline,
- resolve some identified open issues and identify possible new issues,
- summarize the lessons learned from the different participants and possibly update the recommendations of the guideline based on the results of the activity.

The activity aims at quantifying the uncertainties associated with selected relevant physical phenomena during an intermediate break LOCA. Two main IUQ exercises with increasing complexity are included in the ATRIUM activity. In the first exercise, experimental databases are composed of SETs that allows isolating the physical phenomenon of interest, simplifying the IUQ process. The selected phenomenon is the critical flow at the break, which is of significant relevance during a LOCA ([34], [35]). In the second exercise, several influential phenomena are involved at the same time in the experiments. This kind of experiments are called Combined Effect Tests (CETs). In this category, the post-CHF heat transfer phenomena will be studied. Finally, the obtained input model uncertainties will be propagated on a suitable IET, such as OECD/NEA ROSA-2

Project Large Scale Test Facility Intermediate Break-Hot Leg-01 (LSTF IB-HL-01) ([36], [37]) to validate their application in experiments at a larger scale. The final goal in ATRIUM is then to calculate reliable input uncertainties for the application in LSTF IB-HL-01. The use of an IET was preferred over an existing NPP in order to have access to real experimental data and avoid any limitation on the sharing of data among participants of the activity. LSTF was chosen because it is a well-instrumented and well-documented facility and it was already available to share for NEA participants. Its data were already used to validate the thermal-hydraulic system codes and for this test the PIRT (phenomena identification and ranking table) analysis was performed, which is the first step of the SAPIUM guidelines.

Adequacy analysis with respect to LSTF IB-HL-01 (which is the target application) is the content of the first task of the first exercise.

3.2. Experimental databases

In this section, we briefly describe the target application and the experimental databases of the first IUQ exercise on critical flow.

3.2.1. The LSTF IB-HL-01 test

The Large Scale Test Facility is a full-pressure and full-height two-loop IET reproducing a 1100 MWe four-loop Westinghouse-type Pressurized Water Reactor (PWR) with a volumetric scaling of 1/48 (see Figure 2). It was built by the Japanese Atomic Energy Agency (JAEA) in 1985 for the fourth-phase of the Rig-of-Safety Assessment (ROSA-IV) program. The aim was to investigate thermal-hydraulic responses during reactor accidents and abnormal transients after the TMI-2 (Three Mile Island - Unit 2) accident. Following scaling considerations, the full size rod bundles are electrically heated. The core power is reduced to 14% volumetric core power w.r.t. 1100 MWe power. The flow area in the hot and cold legs was scaled to conserve the ratio of the length to the square root of pipe diameter [38], i.e., $L/\sqrt{D}$ of the reference PWR. This approach was taken to better simulate the flow regime transitions in the primary loops.

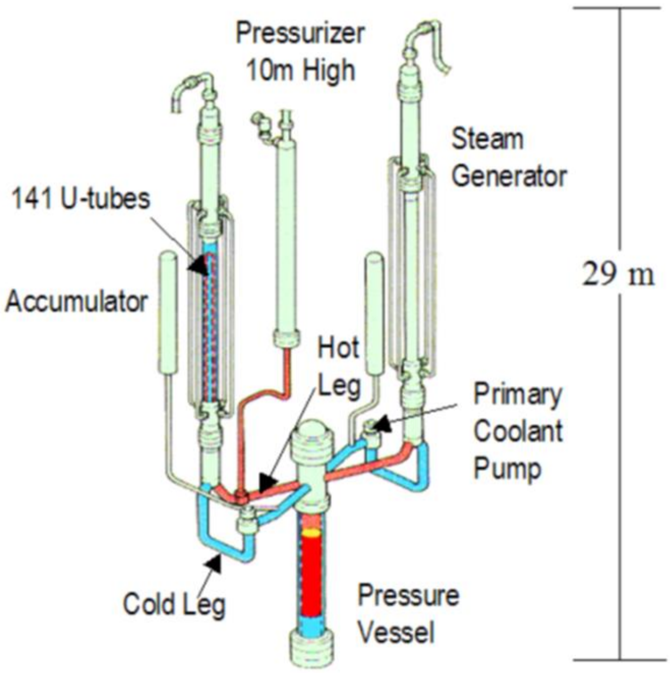

Figure 2 : Schematic view of LSTF [37].

The LSTF IB-HL-01 test ([36], [37], [39], [35],) simulates an intermediate break LOCA due to a double-ended guillotine break of the pressurizer surge line connected to the hot leg. The break is represented by an upward vertical long nozzle (the ratio *L/D* between the straight length *L* and the inner diameter *D* is equal to 12) placed at the top of the hot leg in the loop without pressurizer. The area of the break is equal to 17% of the cold leg, which corresponds to a nozzle throat diameter *D* of 41 mm. A Venturi flowmeter is installed before the break valve to measure the break mass flowrate.

The initial conditions are representative of a typical PWR ([37]). The scenario starts when the break valve is opened. The emergency shutdown of the simulated core (SCRAM) is triggered after 1 second. As shown in Figure 3, the primary pressure decreases rapidly. Due to the leakage at the break, the hot leg empties. At the break, the critical flow occurs and lasts roughly 200 seconds. The flow at the break switches from two-phase flow regime to single-phase steam flow regime in a short time. Actually, at the very beginning of the transient, the thermodynamic quality at the break is negative and liquid goes through the break. The two-phase flow appears very sudden and the thermodynamic quality becomes positive. The water level in the core starts to decrease and eventually flashing occurs. The cladding temperature arises sharply at 165 s because the Critical Heat Flux occurs, just 10 s after the accumulators opening. The occurrence of the Loop Seal Clearing leads to full core quenching at approximately 190 s.

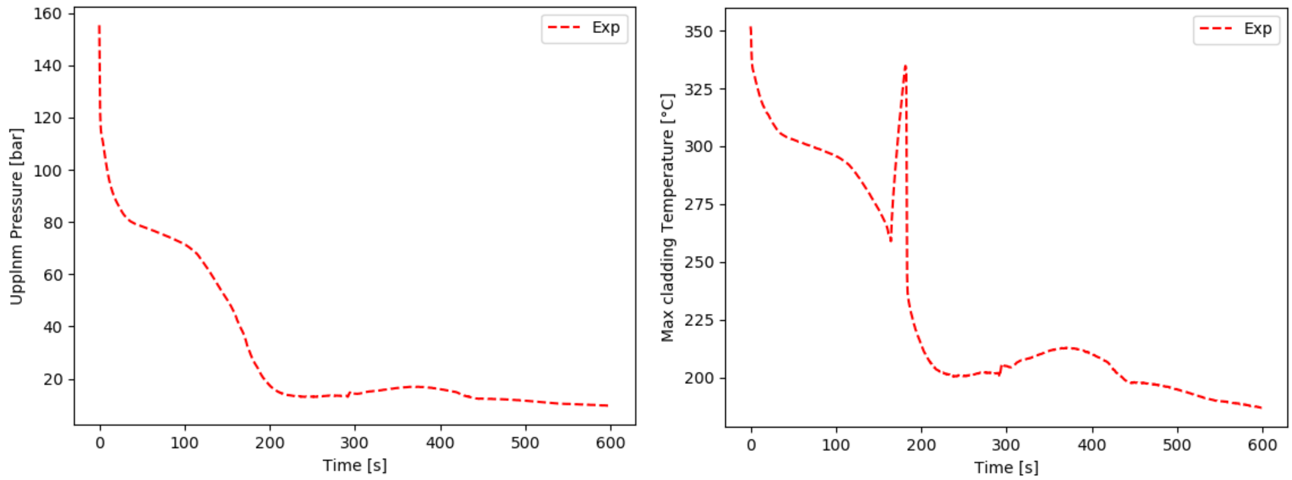

Figure 3 : LSTF IB-HL-01 – Time evolution of the primary pressure and of the cladding temperature with the highest peak temperature.

For this kind of transient, Phenomena Identification and Ranking Table (PIRT) analysis ([34], [35], [40]) revealed that the critical mass flow is one of the most influential phenomena for the depressurization of the reactor. For that reason, the first exercise of the ATRIUM activity focuses on the quantification of the uncertainty of physical models related to the critical flow. However, the activity is interested to quantifying the uncertainties for the target application, i.e., LSTF IB-HL-01. Therefore, we define a target application domain, which represents the space of physical and geometrical variables in which uncertainties should be quantified to apply in the target application. The target application domain is presented in Table 2.

Table 2: LSTF IB-HL-01 - Target application domain for the critical flow.

Nozzle throat diameter D [mm]	41
L/D [-]	12
Inlet pressure [bar]	10 – 155
Inlet thermodynamic quality [-]	-0.153 – 1.0
Critical mass flux [kg/m^2/s]	1500 – 46000

3.2.2. Critical flow experimental databases

An initial list of experiments was proposed to the participants of the benchmark to simplify comparisons and to have the same starting database when applying the SAPIUM framework. However, the final database used for IUQ may vary among participants due to their adequacy and experimental analysis, as detailed in the rest of the paper.

The initial set of databases come from three experimental facilities: Sozzi-Sutherland (Nozzle 2, 3 and 4) [41], Super Moby-Dick ([42], [43], [44]) and Marviken CFT [45] (tests 13 [46], 17 [47] and 24 [48]). The aim of these facilities was to obtain critical flow experimental data at conditions representative of LOCA transients.

The Sozzi-Sutherland facility experiments were carried out in 1975 by General Electric. A vessel of 0.28 m^3 filled with subcooled or saturated water at approximately 70 bar was emptied into the atmosphere through various horizontal nozzles and the critical flowrate was measured. In ATRIUM, only experiments performed in Nozzle 2, 3 and 4 are considered. Nozzle 2 (N2) is a 44.5 mm long well-rounded inlet convergent. The nozzle diameter of N2 geometry is smoothly reduced from 43.2 mm to 12.7 mm. Downstream the throat, either a straight pipe of length *L* or an abrupt expansion (*L* = 0) is present. The ratio *L*/*D* varies between 0 and 140 across tests. Nozzle 3 (N3) is a 12.7 mm sharp-edge orifice, 4.7 mm long. Nozzle 4 (N4) is a 44.5 mm long inlet convergent with abrupt expansion downstream (L = 0). The pipe diameter is progressively reduced from 43.2 mm to 19 mm. Overall, a total number of 439 averaged steady-state test values is available.

The Super Moby-Dick (SMD) facility experiments were performed in CEA between 1982 and 1983. In ATRIUM, two configurations considered in these experiments are studied. The first test section (indicated as "SMD with divergent") is a 100 mm long smooth convergent (diameter reduction from 66.65 to 20.13 mm), followed by a 363 mm straight pipe and a 437 mm long diverging outlet (conical expansion with an angle of 6°57'). The inlet convergent was designed particularly smooth to minimize two-dimensional effects. The other test section ("SMD with expansion") aimed at evaluating the impact of the outlet geometry on the critical flow. It is a 100 mm long smooth convergent (diameter reduction from 87.5 to 20.0 mm), followed by a 400 mm straight pipe and an abrupt expansion (instead of a divergent). In ATRIUM, 39 steady-state SMD tests (keeping constant the pressure at the outlet of the nozzle once the critical flow reached) are employed.

The Swedish Marviken reactor was never put into service due to technical and economic reasons. It was therefore reconverted into a full-scale reactor experimental facility that operated between 1972 and 1982. Twenty-seven Critical Flow Tests (CFT) were carried out in 1978-79. These transient experiments consisted in a discharge of subcooled water or steam-water from the reactor vessel (net volume of 425 m^3, 24.55 m high and 5.22 m in diameter) to the drywell atmosphere. The vertical smooth converging nozzle, placed at the bottom of the vessel, is followed by a straight pipe and equipped with a ball valve whose opening started the tests. In ATRIUM, three Marviken facility CFT tests (13, 17 and 24) are studied.

A summary of each experimental database is reported in Table 3, Figure 4 and Figure 5. The table contains the main geometric features of the nozzle (the throat diameter D and the length-to-diameter ratio L/D), the number of tests, the pressure and thermodynamic quality ranges at the inlet of the nozzle and the range of measured critical mass fluxes.

Table 3: Summary of each experimental database and comparison with the target domain.

		***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 LSTF		41	12	1	10 – 155	-0.15 – 1.0	1500 – 46000

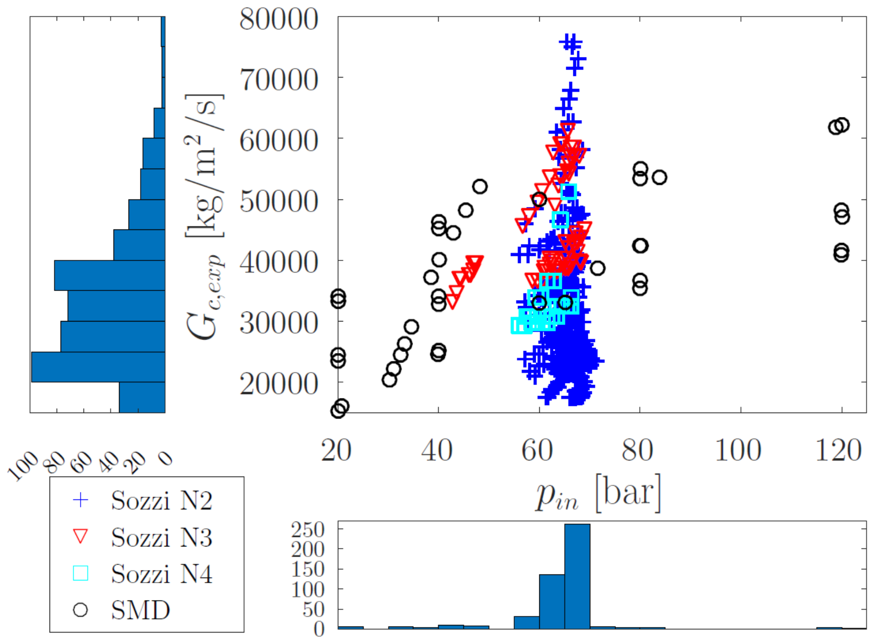

Figure 4 : Experimental critical mass flux vs inlet pressure (Sozzi-Sutherland and SMD).

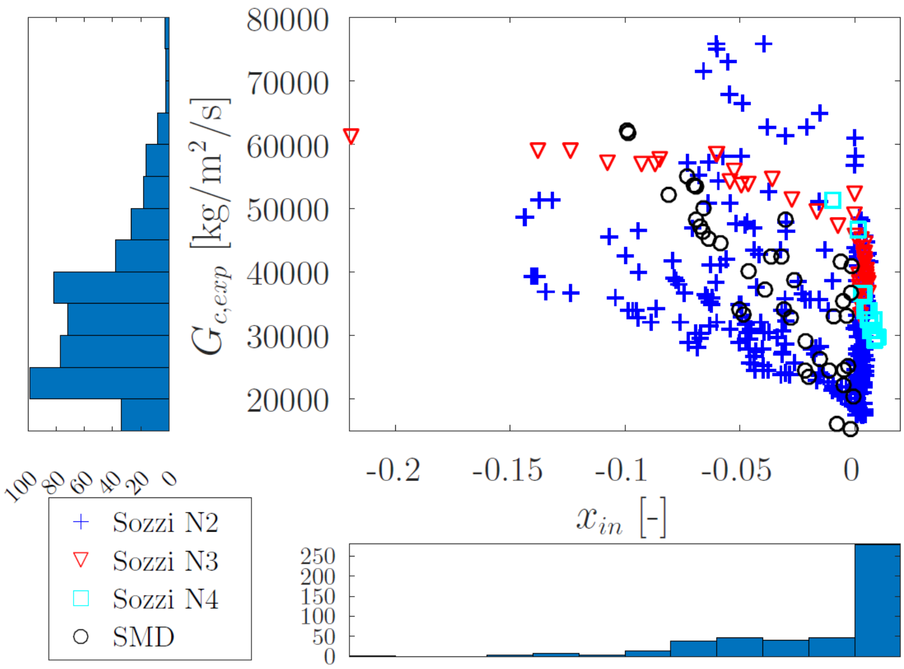

Figure 5 : Experimental critical mass flux vs inlet thermodynamic quality (Sozzi-Sutherland and SMD).

3.3. Application of the systematic adequacy analysis approach

ATRIUM participants applied the approach of Figure 1 to analyze the adequacy of the eight experimental databases summarized in Table 3: Sozzi nozzle 2, Sozzi nozzle 3, Sozzi nozzle 4, SMD divergent, SMD expansion, Marviken CFT 13, Marviken CFT 17 and Marviken CFT 24. Note that in this application, for each given database, the set of experimental tests is collected during an individual experiment.

Concerning the first step of the systematic approach related to the specification of the problem of interest, we recall that the scenario of interest is an intermediate break LOCA and the objective for which experimental databases are used and analyzed is the input uncertainty quantification. Moreover, the adequacy analysis problem (Step 2.1) considers LSTF IB-HL-01 as target application and the objective of the analysis is the ranking of the eight experimental databases.

The next section is devoted to the methodological development of the adequacy analysis (Steps 2.2 to 2.4).

3.3.1. Representativeness and completeness criteria (Step 2.2)

Table 4 provides the list of representativeness criteria and sub-criteria proposed by the ATRIUM group.

Table 4: List of representativeness criteria and sub-criteria.

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

Three criteria are proposed. For two of them, sub-criteria are introduced in order to improve their understanding by the analyst.

The first criterion aims at evaluating the representativeness of the experimental databases with respect to the target application (e.g. the transient of interest) in terms of geometry and thermal-hydraulic conditions. A scaling analysis based on the computation of quantitative indicators focusing on the distortion between experimental and target application facilities may be performed to evaluate this criterion.

The second criterion deals with the quality of the experimental data for the intended IUQ. The following questions should be considered for the analysis with respect to this criterion:

- Is the description of the facility well-documented?
- Are the description and location of instrumentation available in the documentation?
- Are experimental uncertainties (e.g., measurement uncertainties) available?
- Are the experiments repeated in the same conditions having the same results?
- Is the quality of the experimental measurements good enough for the purpose of the analysis (low uncertainties, good positioning, etc.)?
- Do you have all the measurements you need for evaluating the phenomena of interest?

The last criterion assesses the capability of the experiments to represent the physical phenomena of interest and to feed the necessary boundary conditions to validate the code response. The first sub-criterion investigates whether the physical phenomena of interest occur and are easily observable/quantifiable for IUQ purposes. The second one addresses whether the physical phenomena can easily be separated to obtain

SETs (the higher degree of separability, the higher relevance for IUQ). The third one is introduced to evaluate whether the experimental data can be reproduced by Best-Estimate (BE) simulations and therefore can be used in the next elements of the IUQ approach.

The completeness criteria are summarized in Table 5.

Table 5: List of completeness criteria.

Criteria
C^c_1: Coverage of the target application domain[4]
C^c_2: Spatial distribution of the experimental tests in the domain resulting from the intersection between the experimental and the target application domains

Both criteria of Table 5 focus on the position of experimental tests in the target application domain. The experimental domain is built as the convex hull of the experimental test points. Then, the first criterion is associated to the volume of the domain resulting from the intersection between the experimental domain and the target application, while the second criterion deals with the spatial location of the test points within this domain (a uniform repartition of the points is to be preferred).

3.3.2. Representativeness and completeness analyses (Steps 2.3 and 2.4)

Participants followed different approaches to handle the multi-criteria decision analysis problems. These approaches can be categorized in different families, as described in Section 2.2.2. They are briefly recalled in the next section and participants' results are summarized in Section 3.3.2.2.

3.3.2.1. Methods used by participants for the adequacy analysis in the ATRIUM activity

Table 6 summarizes the methodologies used by participants for the adequacy analysis. Concerning the representativeness, the used methods fall into two categories. The first category is the AHP approach ([12], [24]). The second category is based on a rating of each database and each criteria importance weight from a pre-determined scale defined by each participant. The final score is obtained by computing a unique synthesis indicator as recalled in Section 2.2.2. The synthesis indicator is obtained as the weighted average for the majority of participants. Few participants replace the weighted average score by 0 (i.e., the lowest score) if a database is not representative at all with respect to a given criterion. For sake of simplicity, we mention "Unique synthesis indicator" in Table 6 to indicate this type of aggregation.

Table 6: List of participants and methods used for representativeness and completeness analyses. The symbol "-" indicates that the participant did not perform the full completeness analysis of each database.

Participants	Method for representativeness analysis	Method for completeness analysis
CEA	AHP	-

[4] We recall that the target application domain represents the space of physical and geometrical variables in which uncertainties should be quantified to apply in the target application. An example of target application domain is given by Table 2.

CNPE	Rating + Unique synthesis indicator	Rating + Unique synthesis indicator
CRIEPI	Rating + Unique synthesis indicator	Rating + Unique synthesis indicator
EDF	AHP	Spatial quantitative indicators
GRS	Rating + Unique synthesis indicator	-
IRSN	AHP	Spatial quantitative indicators
JAEA	AHP	AHP
KAERI	Rating + Unique synthesis indicator	Rating + Unique synthesis indicator
NCSU	Rating + Unique synthesis indicator	Spatial quantitative indicators
NRA	Modified AHP [49]	Spatial quantitative indicators
NRG	AHP	Spatial quantitative indicators
ENEA/POLIMI/POLITO	AHP	AHP
PSI	Rating + Unique synthesis indicator	-
UPC	Rating + Unique synthesis indicator	Rating + Unique synthesis indicator
UROMA	Rating + Unique synthesis indicator	Spatial quantitative indicators

For completeness analysis, some participants have followed the same methods as for representativeness analysis: AHP or rating with aggregation based on the computation of a unique synthesis indicator. The other participants have computed quantitative indicators coming from spatial data analysis such as volume ratio or discrepancy indicator [16]. Some participants did not perform a full completeness analysis for each individual database, since they claimed that completeness can be evaluated only for the set of databases (combining all selected databases) resulting from the representativeness analysis.

3.3.2.2. Participants' results

This section is devoted to the summary of participants' results. Each contribution is a ranking of the eight experimental databases according to their representativeness and completeness. The objective of this summary is to investigate whether a partial consensus between participants is reachable on the representativeness and completeness of the studied databases. Eventually it is not expected to exhibit a final ranking shared between all the participants. Such objective would require a confrontation of participants' point of view. The combination of several analyses refers to group-decision making problem to which several multi-criteria decision analysis methods such as AHP [50] have been extended. Since this paper is focused on a first illustration of the adequacy analysis systematic approach, the application of a method to handle group-decision making problem is out of the scope of this work.

Representativeness:

As shown in Table 6, all participants have performed the representativeness analysis. The majority of participants have considered all eight databases, except CEA. Since the SMD tests with expansion have very similar inlet conditions and results to the ones with divergent, CEA decided not to include SMD with expansion in their analysis, in order to avoid dealing with duplicated experimental points that could bias the IUQ performed in the subsequent elements of the SAPIUM methodology. Moreover, two participants (NRA and PSI) split the Sozzi nozzle 2 database with respect to the nozzle length and analyzed the resulting parts separately. For NRA, eleven different types of nozzles ranging from $L = 0$ mm to $L = 1778$ mm were introduced while PSI considered short ($L/D \leq 8$) and long ($L/D > 8$) nozzles as the ratio L/D may significantly impact the critical flow phenomenon, especially the presence of non-equilibrium phenomena in the nozzle [51].

For the sake of comparison among participants, we integrate in the following synthesis exclusively the part of Sozzi nozzle 2 database that was considered by these two participants for input uncertainty quantification. It was also checked that discarding their contributions related to Sozzi nozzle 2 does not change the conclusions of the analysis.

Figure 6 displays the representativeness ranking of each database for each participant. The rank can go from least (1) to most representative (8). Due to equal representativeness between databases, not all ranks vary over the whole scale between 1 and 8.

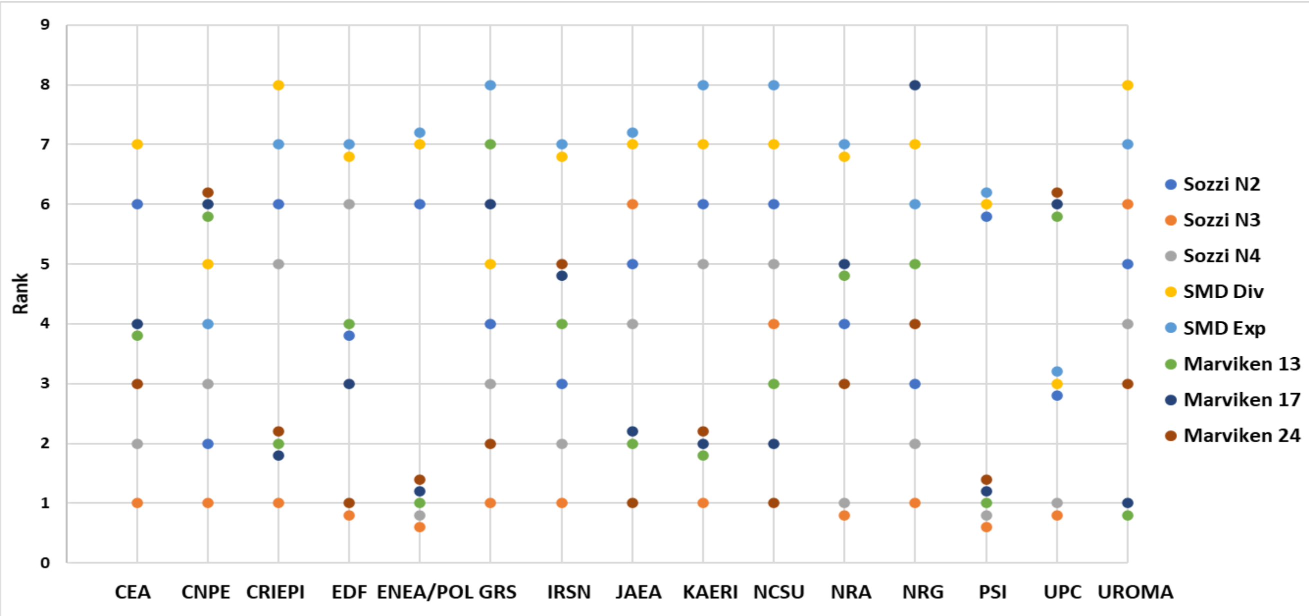

Figure 6 : Representativeness ranking of the eight experimental databases.

This visual representation of the results emphasizes that Sozzi nozzle 3 (orange point) is the least representative and SMD (yellow and light blue points) the most representative databases for a large majority of participants. Among the justifications provided for this ranking, the participants formulated mainly considerations on:

- the nozzle geometry (e.g., the orifice geometry of Sozzi nozzle 3 ($L/D = 0$), which is very different compared to the convergent nozzle in LSTF $L/D = 12$),
- the experimental conditions (e.g. the SMD experiments were performed in steady-state conditions with a dedicated loop, while the Sozzi and Marviken tests featured the discharge of water from a big tank to the atmosphere),

- the possibility to control and measure precisely the relevant physical quantities (e.g. Marviken and SMD allow measuring the inlet pressure in the nozzle),
- the modeling in system codes (e.g. difficulties to model non-equilibrium phenomena).

Most participants have considered the possibility offered by the SMD facility to perform steady state tests and to access detailed measurements on the physical conditions during the test as a major asset of the SMD databases. This mostly explains the relative consensus about the high ranking of SMD.

To go further in the analysis, Table 7 (a) and (b) are focused on the pairwise comparison between databases. The (i,j)-cell of Table (a) corresponds to the percentage of participants considering the i^{th} database as strictly more representative than the j^{th} database, i.e. rank(i^{th} database) > rank(j^{th} database). Table (b) deals with equal representativeness between the i^{th} and the j^{th} databases i.e. rank(i^{th} database) = rank(j^{th} database).

Table 7: Pairwise comparison between databases expressed as the percentage of participants considering a given database as (a) strictly more representative than or (b) equally representative to another one.

>	Sozzi N2	Sozzi N3	Sozzi N4	SMD Div	SMD Exp	Marviken 13	Marviken 17	Marviken 24
Sozzi N2	0	87	87	0	0	53	60	73
SozziN3	13	0	13	0	0	20	20	20
Sozzi N4	13	60	0	0	0	40	40	47
SMD Div	87	100	100	0	29	80	73	87
SMD Exp	86	100	100	21	0	86	79	86
Marviken 13	40	67	47	20	14	0	20	47
Marviken 17	40	67	47	27	21	13	0	47
Marviken 24	27	60	40	13	14	13	7	0

(a)

=	Sozzi N2	Sozzi N3	Sozzi N4	SMD Div	SMD Exp	Marviken 13	Marviken 17	Marviken 24
Sozzi N2	100	0	0	13	14	7	0	0
SozziN3	0	100	27	0	0	13	13	20
Sozzi N4	0	27	100	0	0	13	13	13
SMD Div	13	0	0	100	50	0	0	0
SMD Exp	14	0	0	50	100	0	0	0
Marviken 13	7	13	13	0	0	100	67	40
Marviken 17	0	13	13	0	0	67	100	47
Marviken 24	0	20	13	0	0	40	47	100

(b)

Highlighted cells correspond to a percentage strictly larger than 50%. Several conclusions can be drawn from these tables. A consensus is reached for several pairwise comparisons with a percentage of participants strictly larger than 50%. In these cases, the visual analysis of Figure 6 is confirmed concerning the highest representativeness of SMD and the lowest one of Sozzi nozzle 3. Moreover, Sozzi nozzle 2 appears to be the second most representative since it is more representative than Sozzi nozzle 3, Sozzi nozzle 4, Marviken CFT 17, Marviken CFT 24 and Marviken CFT 13. Finally, Marviken CFT 13 and CFT 17 are likely to have the same level of representativeness and are at least as representative as Marviken CFT 24 for more than 85% of the participants. There is no significant conclusion concerning the comparison between Sozzi nozzle 4 and Marviken.

It is also interesting to study the representativeness results with respect to each criterion of Table 4. Figure 7 focuses on the criterion weights provided by each participant and displays their associated ranks. The modified AHP [49] used by NRA did not assign any weight. Their contribution is therefore left blank in this figure.

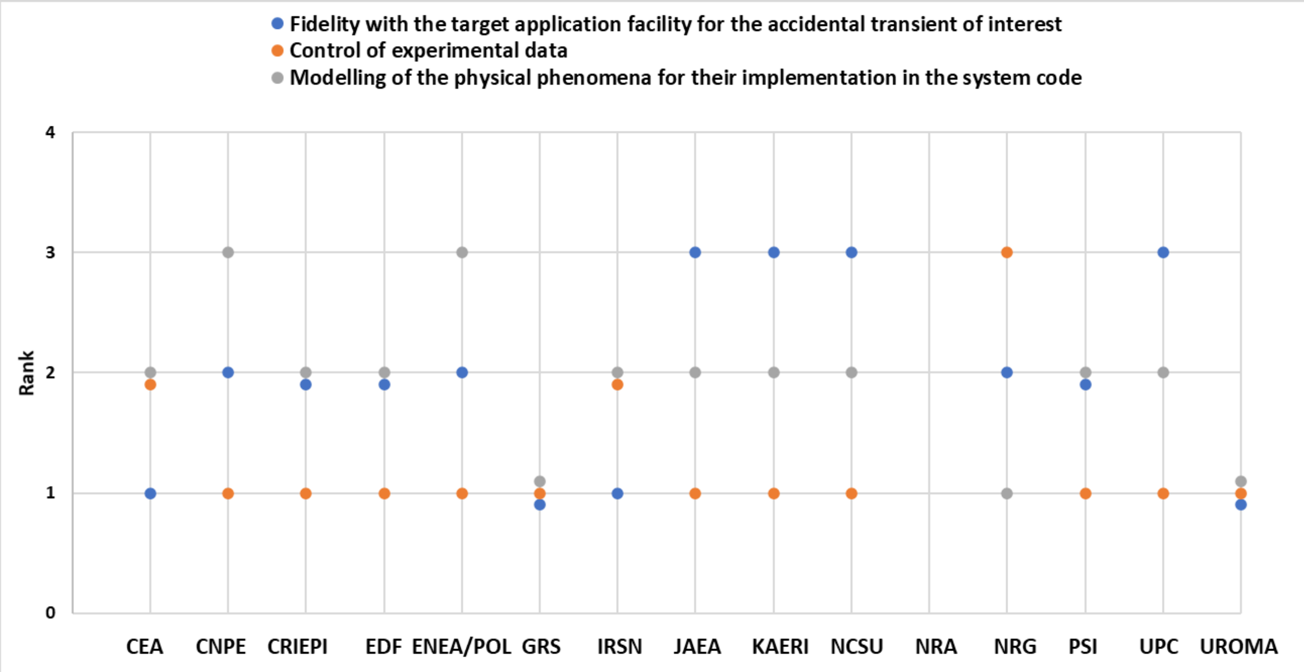

Figure 7 : Representativeness criterion weight ranking. Due to equal importance between criteria, not all rankings vary between 1 and 3.

A majority of participants has identified the fidelity with LSTF (C^r_1) and the physical phenomena modelling (C^r_3) as the most important criteria. However, as already mentioned, a strong consensus among participants could not be identified.

The main point of divergence between participants concerns the relative importance of criteria C^r_1 and C^r_3 (see Figure 7). Indeed, some participants give prominence to the fidelity of experimental settings to the LSTF case. This choice is sometimes justified by the need to control the specific phenomenon to be modeled and hence the physical quantities involved in the model. Some other participants use similar arguments about the ability of the data to serve adequately the physical modeling to give more weight to criterion C^r_3, which includes a separability sub-criterion ($C^r_{3\text{-}2}$). This choice is however often supported by the higher emphasis put on the ability of a test facility to reproduce the physical phenomena actually occurring in the target application, rather than the correspondence between the physical domain covered by the data and the target one.

We finally investigate the representativeness per criterion. The analysis of the representativeness with respect to C^r_1 leads to the same conclusions as for the global representativeness. Therefore, the ranking of the databases given by each participant is not provided in this paper.

Concerning the second criterion, Figure 8 shows that the large majority of participants cluster the quality of the experimental data basing on the name of the experimental facility. For example, all the 3 Marviken databases are considered with the same quality by each participant. It can occur since documents for the description of the experimental apparatus can be the same for more than one experimental campaign.

This is confirmed by Table 8 (b), in which blue cell corresponds to a percentage strictly larger than 50%. It is also interesting to notice from Table 8 (a) that SMD and Marviken databases are the most representative when considering this criterion.

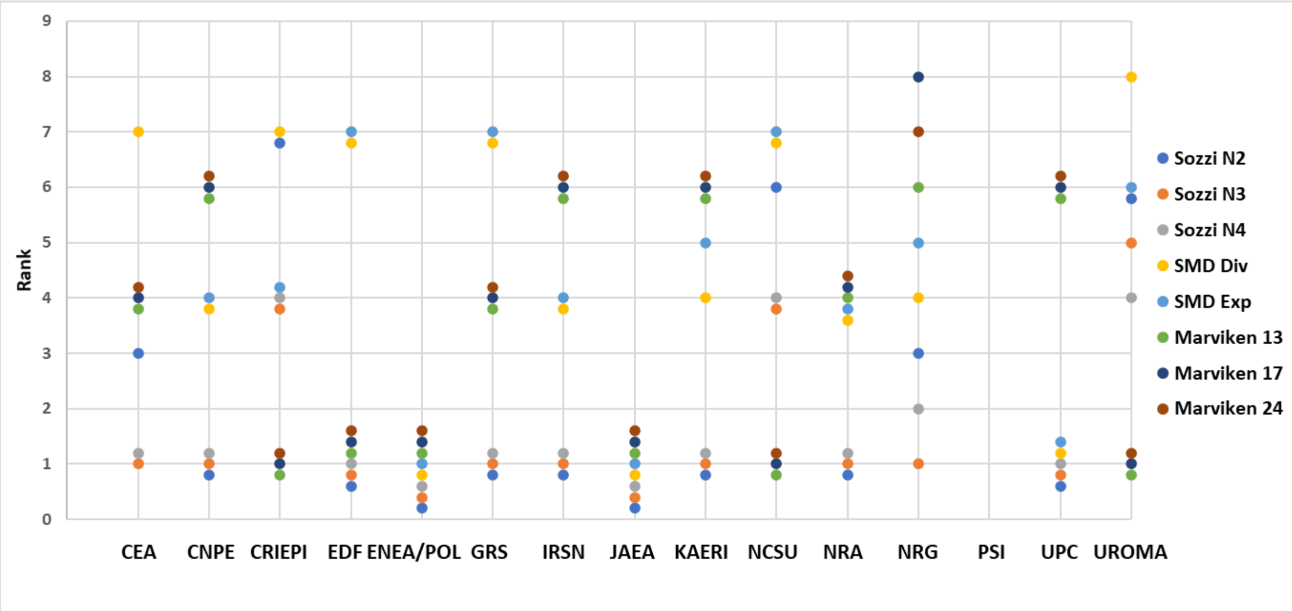

Figure 8 : Representativeness ranking of the eight experimental databases with respect to the control of experimental data (C^{r}_{2}). Due to equal representativeness between databases, not all rankings vary between 1 and 8.

Table 8: Pairwise comparison between databases with respect to the control of experimental data (C^{r}_{2}) expressed as the percentage of participants considering a given database as (a) strictly more representative than or (b) equally representative to another one.

>	Sozzi N2	Sozzi N3	Sozzi N4	SMD Div	SMD Exp	Marviken 13	Marviken 17	Marviken 24
Sozzi N2	0	36	36	0	8	21	21	21
SozziN3	0	0	7	0	0	21	21	21
Sozzi N4	0	7	0	0	0	21	21	21
SMD Div	71	79	79	0	15	43	43	43
SMD Exp	62	69	69	15	0	38	38	38
Marviken 13	57	57	57	36	38	0	0	0
Marviken 17	57	57	57	36	38	7	0	7
Marviken 24	57	57	57	36	38	7	0	0

(a)

=	Sozzi N2	Sozzi N3	Sozzi N4	SMD Div	SMD Exp	Marviken 13	Marviken 17	Marviken 24
Sozzi N2	100	64	64	29	31	21	21	21
SozziN3	64	100	86	21	31	21	21	21
Sozzi N4	64	86	100	21	31	21	21	21
SMD Div	29	21	21	100	69	21	21	21
SMD Exp	31	31	31	69	100	23	23	23
Marviken 13	21	21	21	21	23	100	93	93
Marviken 17	21	21	21	21	23	93	100	93
Marviken 24	21	21	21	21	23	93	93	100

(b)

Finally, the influence of the facility is also noticeable when focusing on the modeling of physical phenomena (C^r_3) with several databases of equal representativeness (Figure 9). The SMD databases still appear as the most representative due to the steady-state conditions and smooth inlet convergent nozzle which reduces 3D effects, and no clear conclusion can be drawn from this figure for the pairwise comparison between Sozzi and Marviken databases.

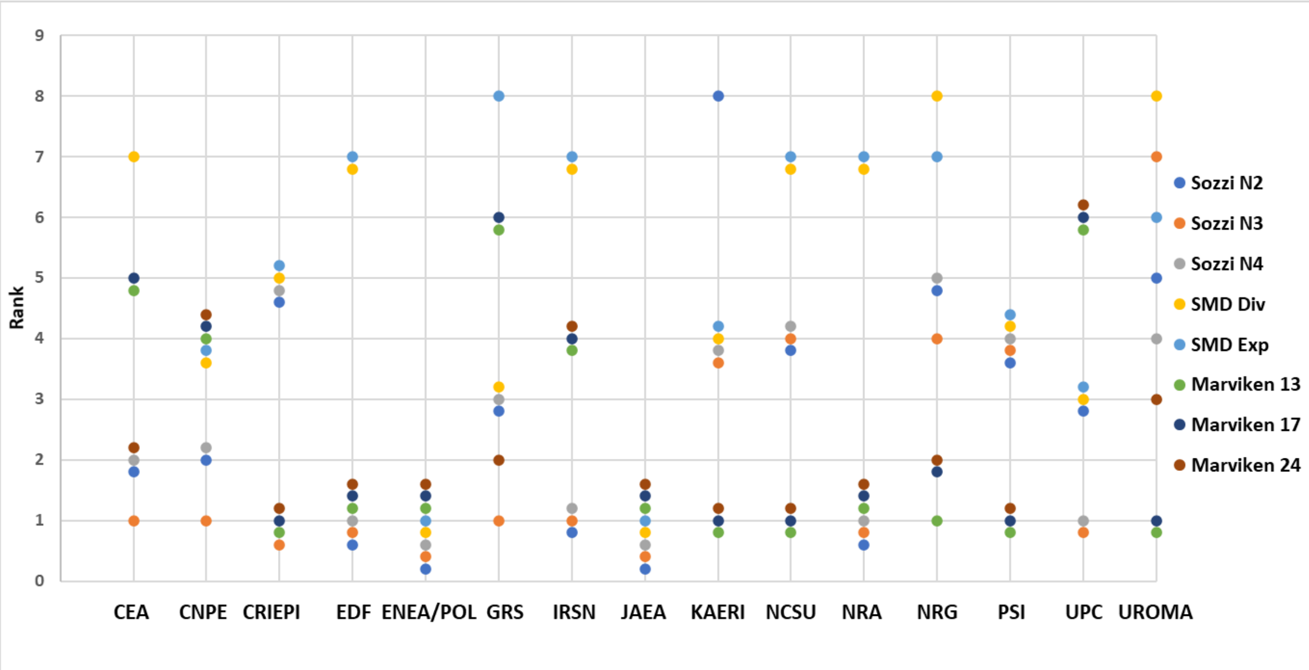

Figure 9 : Representativeness ranking of the eight experimental databases with respect to modelling of the physical phenomena for their implementation in the system code (C^r_3). Due to equal representativeness between databases, not all rankings vary between 1 and 8.

<u>Completeness:</u>

Most participants (12 participants among 15) have performed the full completeness analysis. CEA and GRS did not perform the completeness evaluation of the height databases separately but of the data set resulting from the combination of the most representative databases. This strategy provides relevant information to the next elements of the SAPIUM approach (see Section 4.2). UPC did not evaluate the completeness of Sozzi nozzle 3 and 4. Moreover, similarly to the representativeness analysis, NRA split the Sozzi nozzle 2 database with respect to the nozzle length and analyzed the resulting parts separately. We only provide in the summary the length selected for the following task of the activity.

Figure 10 displays the ranking of each database for each participant and Table 9 (a) and (b) are focused on the pairwise comparison between databases.

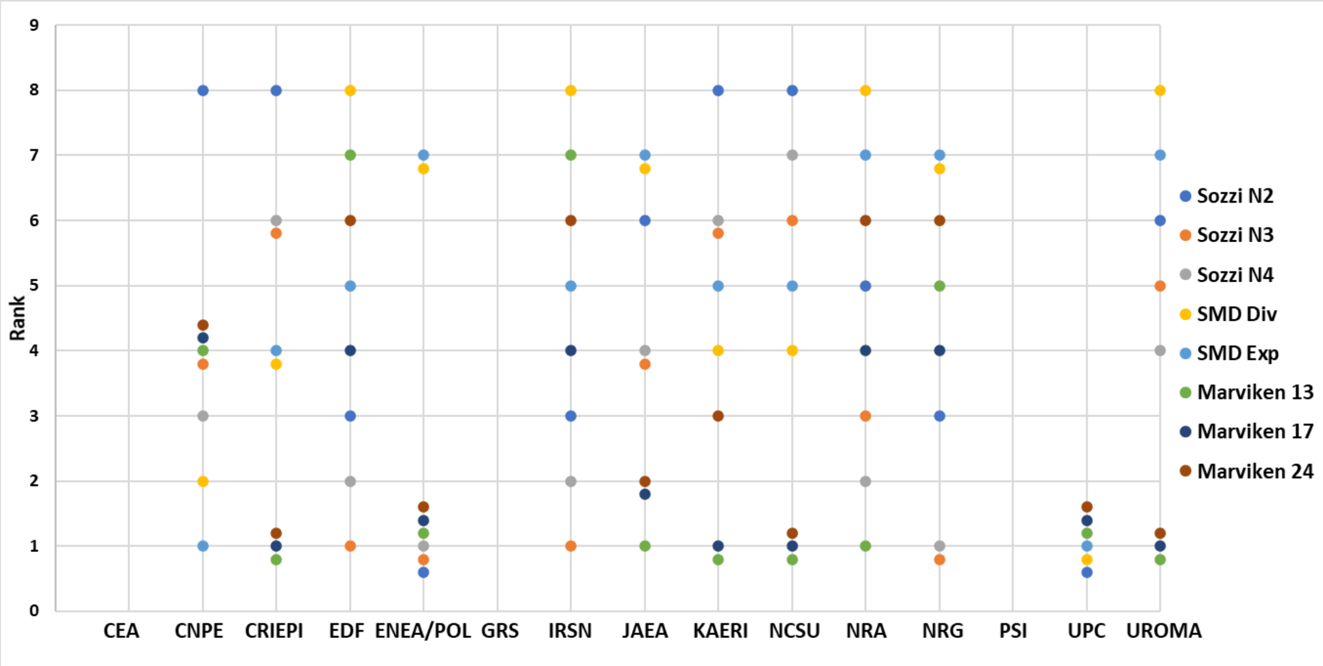

Figure 10 : Completeness ranking of the eight experimental databases. Due to equal completeness between databases, not all rankings vary between 1 and 8.

Table 9: Pairwise comparison between databases expressed as percentage of participants considering a given database as (a) strictly more complete than or (b) equally complete as another one. Blue cell corresponds to a percentage strictly larger than 50%.

>	Sozzi N2	Sozzi N3	Sozzi N4	SMD Div	SMD Exp	Marviken 13	Marviken 17	Marviken 24
Sozzi N2	0	90	90	36	36	64	64	55
SozziN3	0	0	30	40	40	60	50	50
Sozzi N4	0	20	0	40	40	60	50	50
SMD Div	55	60	60	0	36	82	82	82
SMD Exp	55	60	60	18	0	73	82	73
Marviken 13	18	20	30	9	18	0	18	9
Marviken 17	18	30	40	9	9	18	0	0
Marviken 24	27	30	40	9	18	36	36	0

(a)

=	Sozzi N2	Sozzi N3	Sozzi N4	SMD Div	SMD Exp	Marviken 13	Marviken 17	Marviken 24
Sozzi N2	100	10	10	9	9	18	18	18
SozziN3	10	100	50	0	0	20	20	20
Sozzi N4	10	50	100	0	0	10	10	10
SMD Div	9	0	0	100	45	9	9	9
SMD Exp	9	0	0	45	100	9	9	9
Marviken 13	18	20	10	9	9	100	64	55
Marviken 17	18	20	10	9	9	64	100	64
Marviken 24	18	20	10	9	9	55	64	100

(b)

A majority of participants considers SMD and Sozzi nozzle 2 as the most complete (see rows 1, 4 and 5 of Table 9 (a)) whereas Marviken databases are the least complete. Similar conclusions can be drawn from the completeness analysis with respect to the criteria introduced in Table 5. Therefore, we reduce the summary of these results to the criteria ranking given by each participant (Figure 11). The contributions associated to participants that did not perform the full analysis and to NRA whose method did not assign weight are left blank on this figure.

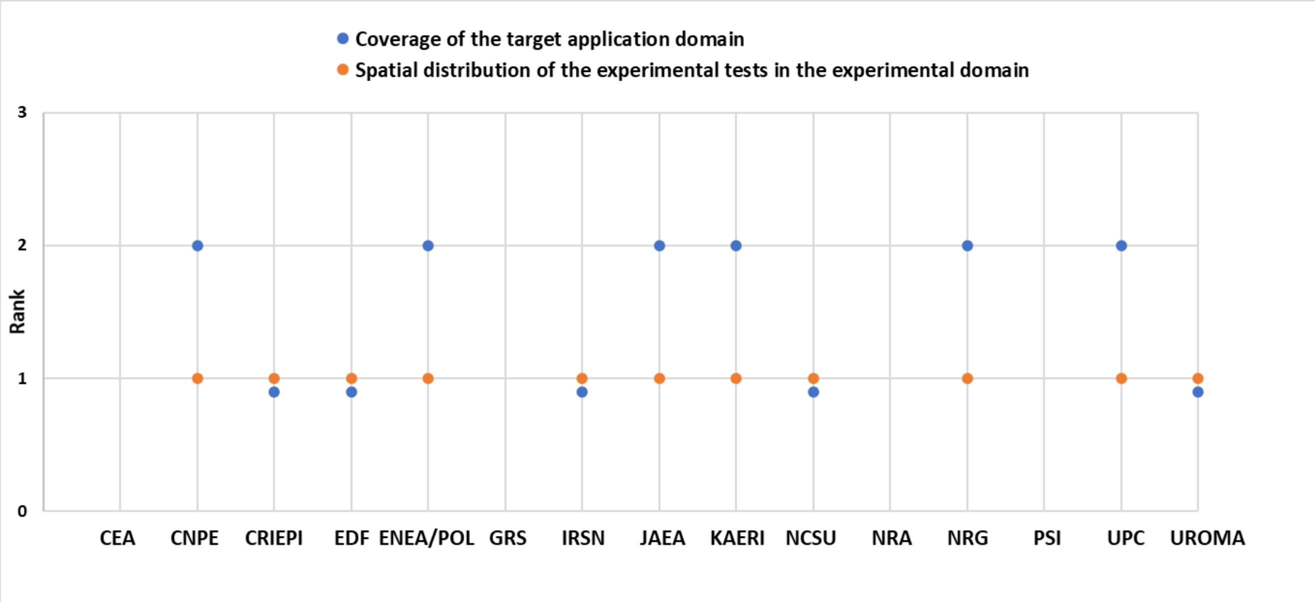

Figure 11 : Completeness criterion weight ranking.

Excluding the participants that provide equal importance to both criteria, the coverage of the application domain (C^c_1) appears to be more important than the spatial distribution of the experimental tests.

4. Proposed further developments on adequacy analysis

This section is focused on two main issues. The first one is related to the construction and the use of completeness indicators. The second one deals with the update of the adequacy results when following the IUQ SAPIUM framework.

4.1. Completeness indicators

By definition, the completeness analysis of a given database relies on the spatial location of the associated tests in the physical domain. Therefore, compared to the representativeness problem, it seems easier to introduce quantitative indicators from spatial data analysis such as volume ratio or discrepancy indicator [16]. However, in order to construct relevant indicators, several key elements should be considered.

The first one is the construction of the target application domain (associated to LSTF in the ATRIUM activity). It involves the specification of the physical quantities of interest, as well as their range of variation, which define the target application domain. This is usually done by expert judgement. In the ATRIUM activity, it was recommended to consider inlet pressure, inlet thermodynamic quality (based on enthalpies) and critical mass fluxes. However, some participants focused on inlet temperature and pressure instead. Furthermore, other parameters (e.g. *L/D*) could also have been added to the analysis. Depending on the code, some parameters defining the application domain might play a more important role than others. This should be taken into

account in the construction of the completeness indicators. The variation ranges of each physical quantity of interest should be then used to define the shape of the application domain to be covered, which is multidimensional. A possible way is to build the associated hypercube as done by many participants in the ATRIUM activity. However, this shape might not be valid since some regions could be uninteresting for the application or even unphysical. For example, some combinations of input parameters may not result in a critical flow condition. The definition of a reliable application domain is therefore difficult to construct a priori. A possible solution is to consider the hyperspace that envelops the application experimental points such as the convex hull.
The second key issue is related to the use of spatial analysis indicators to study the spatial distribution of a set of tests (criterion C^c_2 in Table 5). Even if the application domain is a hypercube, its intersection with the convex hull of the experimental tests has a more complex shape. There exists in the literature statistical tools such as discrepancy indicators [16] that measure the deviation between the empirical and the theoretical uniform distribution. However, they are limited to hypercube domains. Other techniques are therefore required to compute the distance between an empirical distribution associated with the experimental tests and a target distribution in a complex domain. Some elements to solve this issue can be found in [52] and [53] but the extension to domains where unphysical regions have been removed remains open.

Once completeness indicators are constructed, a third key issue concerns its use to evaluate if a given database is complete enough or should be improved. This relies on the introduction of threshold values to decide if a completeness is acceptable or not. In case of an uncomplete database, improvement requires designing new experimental tests or dropping some tests from a database. It could be achieved by plugging our approach to adaptive experimental design methods [11] where experimental tests can be iteratively added or dropped by maximizing the completeness.

4.2. Adequacy analysis in the IUQ SAPIUM framework

The objective of Section 3.3 was to illustrate the application of the adequacy analysis systematic approach to study the experimental databases of the first exercise of the ATRIUM activity. It should be kept in mind that this analysis is performed in the SAPIUM framework whose elements are recalled by Figure 12 taken from [6]. The adequacy analysis is performed in Element 2 and it is therefore important to discuss on the relevance of the adequacy results with respect to the next elements of the IUQ methodology.

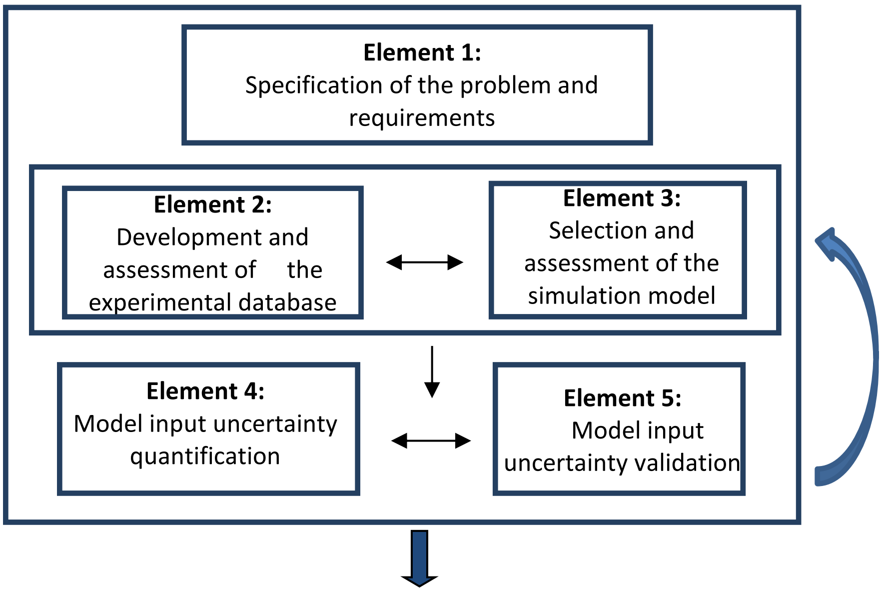

Figure 12 : Elements of the SAPIUM approach.

4.2.1. Update of the representativeness analysis

The SAPIUM approach emphasizes the strong interaction between the choice of adequate experimental databases (Element 2) and the selection of an accurate simulation model (Element 3). The list of criteria of Table 4 was established assuming that some information is available on best-estimate simulations when starting Element 2. However, since Element 3 focuses on the system code, it is recommended to update the representativeness assessment after this element. In particular, the adequacy with respect to the third sub-criterion associated with C^r_3 allowing a feedback of the best-estimate simulations on the adequacy analysis should be reevaluated at this stage.

4.2.2. Adequacy of a combined set of experimental databases

Quantification and validation (Elements 4 and 5 of the SAPIUM approach) require the selection of several databases according to their adequacy. An important question to address is therefore related to the combination of databases. More precisely, the analysis performed in Section 3.3 showed that SMD and Sozzi nozzle 2 are the most representative and complete for a majority of participants. Is this statement still true when focusing on the database resulting from their combination?

It is true in the case of representativeness. For example, if two databases are representative in term of fidelity with the target application, their combination is still representative with respect to the same target application. This point might not be always valid for completeness analysis when focusing on the uniformity of the spatial distribution. Combining two sets of tests exhibiting a high uniformity does not always lead to a resulting set of tests of high uniformity. To illustrate that, a simple toy example is given in Figure 13. It is assumed that two databases (Figure 13 (a) and (b)) are available and that the experimental domain characterized by two quantities (x_0 and x_1) is the unit square for both. Their combination (Figure 13 (c)) produces a database with clusters of points and the resulting database is therefore of lower uniformity.

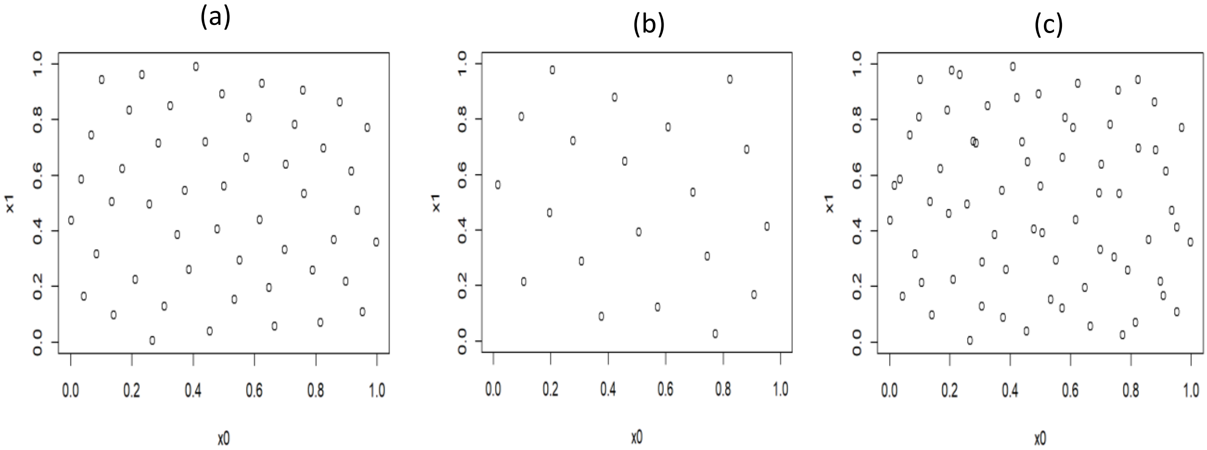

Figure 13 : Example of combination of two databases ((a) and (b)) exhibiting high uniformity in the test distribution but leading to a combined database of lower uniformity ((c)).

Therefore, in the IUQ framework, it is recommended to update the completeness analysis after the selection of candidate subsets of representative databases. The final assessment of the completeness is then best to be performed on the set of databases to be used for IUQ and not on each individual database.

5. Synthesis and conclusions

This paper is devoted to the adequacy analysis of databases that plays an important role in the VVUQ process. After defining adequacy through the introduction of two key properties, which are representativeness and completeness, a new systematic (transparent and reproductible) approach exploiting MCDA methods for analyzing the adequacy of a set of experimental databases is proposed. It is then applied in the OECD/NEA ATRIUM activity which includes a set of practical IUQ exercises in thermal-hydraulics to test the SAPIUM guideline in determining input uncertainties and forward propagating them on an application case. The application case is an intermediate break LOCA and the phenomenon of interest for the first IUQ exercise is the critical flow at the break. The adequacy of eight experimental databases coming from the Super Moby-dick, Sozzi-Sutherland and Marviken experiments are analyzed by fifteen participants.

This work first reveals that a consensus can be reached on the introduction of two sets of criteria to characterize representativeness and completeness. Moreover, based on these criteria, the adequacy analysis of the eight databases performed by all 14 participants shows that Super Moby-dick and Sozzi nozzle 2 are the most adequate (representative and complete) for a large majority of participants. It also emphasizes that Sozzi nozzle 3 is the least representative and Marviken is the least complete. However, as it could be expected, there was no consensus between participants on a common full ranking of the 8 databases. The use of MCDA approaches devoted to group-decision making problem (such as [50]) can be a solution to improve the results.

This exercise is a first and promising application of a transparent and reproducible approach for adequacy analysis. It also emphasizes several improvements to ensure the reliability of the adequacy results. The first one deals with the construction of completeness indicators which requires the specification of the target application domain and the use of spatial data analysis tools. The second one focuses on the use of the adequacy results in the IUQ framework. In order to improve the representativeness analysis with respect to the capability of the best-estimate simulations to reproduce experimental data, it is proposed to update the analysis after the third element of the SAPIUM guideline devoted to the selection and assessment of simulation model. Moreover, since high completeness is not always guaranteed when combining highly complete databases, it is also recommended to perform the completeness analysis after Element 3 in the SAPIUM guideline on the whole final database to be used for IUQ. These two points will lead to a revision of the SAPIUM guideline.

Further works in the ATRIUM activity concern the integration of the adequacy results in the next elements of the SAPIUM guideline (input uncertainty quantification by inverse methods and validation). A special attention will be devoted to the to the evaluation of the effect of a rigorous adequacy analysis exploiting this new systematic approach on the whole IUQ process. Moreover, the adequacy analysis approach will be applied in a second exercise where the post-CHF (Critical Heat Flux) heat transfer phenomena will be studied. From a methodological point of view, new developments should also consolidate the completeness analysis by focusing on the adaptation of spatial indicators to complex experimental domains.

Nomenclature

AHP	Analytical Hierarchical Process
AI	Artificial Intelligence
ATRIUM	Application Tests for Realization of Inverse Uncertainty quantification and validation Methodologies in thermal-hydraulics
BE	Best-Estimate
BEPU	Best-Estimate Plus Uncertainty
CCVM	CSNI Code Validation Matrix
CEA	Commissariat à l'Energie Atomique et aux énergies alternatives - France
CET	Combined Effect Tests
CFT	Critical Flow Tests
CNPE	China Nuclear Power Engineering - China
CRIEPI	Central Research Institute of Electric Power Industry – Japan
CSN	Consejo de Seguridad Nuclear – Spain
CSNI	Committee on the Safety of Nuclear Installations
EDF	Electricité de France – France
ENEA	Agenzia Nazionale per le Nuove tecnologie, l'Energia e lo Sviluppo economico sostenibile - Italy
GRS	Gesellschaft für Anlagen und Reaktorsicherheit - Germany
IET	Integral Effect Tests
IRSN	Institut de Radioprotection et de Sûreté Nucléaire - France
IUQ	Input (or Inverse) Uncertainty Quantification
JAEA	Japan Atomic Energy Agency – Japan
KAERI	Korea Atomic Energy Research Institute – South Korea
LOCA	Loss of Coolant Accident

LSTF IB-HL-01	Large Scale Test Facility Intermediate Break-Hot Leg-01
LWR	Light-Water Reactor
MCDA	Multicriteria Decision Analysis
NCSU	North Carolina State University - USA
NINE	Nuclear and Industrial Engineering - Italy
NPP	Nuclear Power Plant
NRA	Nuclear Regulation Authority - Japan
NRG	Nuclear Research and consultancy Group – the Netherlands
N2	(Sozzi) Nozzle 2
N3	(Sozzi) Nozzle 3
N4	(Sozzi) Nozzle 4
PIRT	Phenomena Identification and Ranking Table
POLIMI	Politecnico di Milano – Italy
POLITO	Politecnico di Torino – Italy
PREMIUM	Post BEMUSE Reflood Models Input Uncertainty Methods
PSI	Paul Scherrer institute – Switzerland
PWR	Pressurized Water Reactor
ROSA	Rig-Of-Safety Assessment (program)
SAPIUM	Systematic APproach for model Input Uncertainty quantification Methodology
SET	Separate Effect Tests
SMD	Super Moby-Dick
TMI-2	Three Mile Island - Unit 2
UPC	Universitat Politècnica de Catalunya – Spain
UROMA	Sapienza Università di Roma - Italy
VVUQ	Verification, Validation and Uncertainty Quantification process

Acknowledgments: The authors are grateful to OECD/NEA for organizing the ATRIUM activity, and to their respective organizations for supporting their participation in this activity.

References

[1] Roy C.J., Oberkampf W.L., 2011, A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, Computer Methods in Applied Mechanics and Engineering, 200(25-28), pp. 2131-2144.

[2] Wilson G.E., 2013, Historical insights in the development of best estimate plus uncertainty safety analysis, Annals of Nuclear Energy, 52, pp. 2-9

[3] Boyack B., Duffey R., Wilson G., Griffith P., Lellouche G., Levy S., Rohatgi U., Wulff W., Zuber N., 1989, Quantifying Reactor Safety Margins: Application of Code Scaling, Applicability and Uncertainty Evaluation Methodology to a Large-Break Loss of Coolant Accident, NUREG/CR-5249, USNRC.
[4] USNRC, 2005, Transient and accident analysis methods, RG-1.203, USNRC.
[5] Skorek T., de Crécy A.,Kovtonyuk A., Petruzzi A.,Mendizábal R., de Alfonso E., Reventós F., Freixa J., Sarrette C., Kyncl M., Pernica R., Baccou J., Fouet F., Probst P., Chung B-D., Tram T.T., Oh D-Y, Gusev A., Falkov A., Shvestov Y., Li D., Liu X., Zhang J., Alku T., Kurki J., Jäger W., Sánchez V., Wicaksono D., Zerkak O., Pautz A., Quantification of the uncertainty of the physical models in the system thermal-hydraulic codes - PREMIUM benchmark, Nuclear Engineering and Design 354, 110199, 2019.
[6] Baccou J., Zhang J., Fillion P., Damblin G., Petruzzi A., Mendizábal R., Reventós F., Skorek T., Couplet M., Iooss B., Oh D.-Y., Takeda T., 2019, Development of good practice guidance for quantification of thermal-hydraulic code model input uncertainty, Nuclear Engineering and Design, 354, 110173.
[7] OECD/NEA/CSNI, 1994, Separate effects test matrix for thermal-hydraulic code validation – Volume I: Phenomena characterisation and selection of facilities and tests, OECD/CSNI/GD(94)82.
[8] OECD/NEA/CSNI, 1996, CSNI Integral Test Facility validation matrix for the assessment of thermal-hydraulic codes LWR LOCA and transients, OECD/NEA/CSNI/R(96)17.
[9] Oberkampf W.L., Pilch M., Trucano T.G., 2007, Predictive capability maturity model for computational modeling and simulation, SANDIA report, SAND2007-5948.
[10] Hemez F., Atamturktur H.S., Unal C., 2010, Defining predictive maturity for validated numerical simulations, Computers & Structures, 88(7-8), pp. 497-505.
[11] Atamturktur S., Williams B. Egeberg M., Unal C., 2013, Batch sequential design of optimal experiments for improved predictive maturity in physics-based modeling, Structural and Multidisciplinary Optimization, 48, pp. 549–569.
[12] Saaty T.L., 1980, The Analytic Hierarchy Process. McGraw-Hill, New York.
[13] Athe P, Dinh N., 2019, A framework for assessment of predictive capability maturity and its application in nuclear thermal hydraulics, Nuclear Engineering and Design, 354, 110201.
[14] JCGM 200:212, 2008, International vocabulary of metrology – Basic and general concepts and associated terms (VIM).
[15] Garud, S.S., Karimi, I., Kraft, M., 2017, Design of computer experiments: A review, Computers and Chemical Engineering, 106, pp. 71-95.
[16] Damblin G., Couplet M., Iooss B., 2013, Numerical studies of space-filling designs: optimization of Latin Hypercube Samples and subprojection properties, Journal of Simulation, 7(4), pp. 276-289.
[17] Fang K.-T., Li R., Sudjianto, A., 2006, Design and modeling for computer experiments. Chapman & Hall/CRC.
[18] Roy B., 1996, Multicriteria Methodology for Decision Aiding, Kluwer Academic Publishers.
[19] Raboun O., Chojnacki E., Duffa C., Rios Insua D., Tsoukiàs A., 2020, Spatial risk assessment in case of multiple nuclear release scenarios, Socio-Economic Planning Sciences, 70, 100721.
[20] IAEA, 2019, Application of Multicriteria decision analysis methods to comparative evaluation of nuclear energy system options: final report of the INPRO collaborative project KIND, IAEA Nuclear Energy Series No. NG-T-3.20.
[21] Kiser L., Otero L.D., 2023, Multi-criteria decision model for selection of nuclear power plant type, Progress in nuclear energy, 159, 104647.
[22] OECD/NEA/CSNI, 2023, SAPIUM: Development of a Systematic APproach for Input Uncertainty quantification of the physical Models in thermal-hydraulic codes - Good Practices Guidance Report, NEA/CSNI/R(2020)16.
[23] Malczewski J., Rinner C., 2015, Multicriteria Decision Analysis in Geographical Information Science, Springer.
[24] Saaty T.L., 1990, How to make a decision: the Analytical Hierarchical Process, European Journal of Operational Research 48:9-26.
[25] Vincke Ph., 1992, Multicriteria decision aid, John Wiley and Sons, Chichester.
[26] Baccou J., Chojnacki E., Georgenthum V., Marais S., 2018, On the methodological treatment of input uncertainty quantification: illustration in the RIA framework, BEPU conference, Lucca, Italy.

[27] Bestion D., 2017, Thermal-Hydraulics of Water Cooled Nuclear Reactors – Chapter 11 - The structure of system thermal-hydraulic (SYS-TH) code for nuclear energy applications, Woodhead Publishing, pp. 639-727.
[28] OECD/NEA/CSNI, 2017, PREMIUM Phase V report: Final Report, OECD/NEA/CSNI/R(2016)18.
[29] Wu X., Xie Z., Alsafadi F., Kozlowski T., 2021, A comprehensive survey of inverse uncertainty quantification of physical model parameters in nuclear system thermal–hydraulics codes, Nuclear Engineering and Design, 384, 111460.
[30] Cocci R., Damblin G., Ghione A., Sargentini L., Lucor D., 2022, A comprehensive Bayesian framework for the development, validation and uncertainty quantification of thermal-hydraulic models, Annals of Nuclear Energy, 172, 109029.
[31] OECD/NEA/CSNI, 2023, SAPIUM: Development of a Systematic APproach for Input Uncertainty quantification of the physical Models in thermal-hydraulic codes - Good Practices Guidance Report, NEA/CSNI/R(2020)16.
[32] Ghione A., Sargentini L., Damblin G., Fillion P., 2021, Application Tests for Realization of Inverse Uncertainty quantification and validation Methodologies in thermal-hydraulics (ATRIUM), CSNI Activity Proposal Sheet (CAPS), WGAMA (2021)4.
[33] Ghione A., Sargentini L., Damblin G., Fillion P., Baccou J., Sueur R., Iooss B., Petruzzi A., Zeng K., Zhang J., Mendizábal R., Skorek T., Wu X., Freixa J., Adorni M., 2023, Application of SAPIUM guidelines to Input Uncertainty Quantification: the ATRIUM project, NURETH-20 conference, Washington DC, USA.
[34] Frepoli C., 2006, Need of a Coherent and Consistent Phenomena Identification and Ranking Table (PIRT) To Address Small, Intermediate and Large Break LOCA in PWRs, Transactions of the American Nuclear Society, 95, pp. 654 – 656
[35] Takeda T., Ohtsu I., 2018, Uncertainty analysis of ROSA/LSTF test by RELAP5 code and PKL counterpart test concerning PWR hot leg break LOCAs, Nuclear Engineering and Technology, 50, pp. 829 – 841.
[36] The ROSA-V Group, 2003, ROSA-V Large Scale Test Facility (LSTF) system description for the third and fourth simulated fuel assemblies, JAERI-Tech 2003-037, Technical report.
[37] OECD Nuclear Energy Agency, 2011, Final Data Report of ROSA-2/LSTF Test 1 (Hot Leg Intermediate Break LOCA IB-HL-01 in JAEA), Technical report.
[38] Zuber N., 1980, Problems in Modeling of Small Break LOCA, NUREG-0724.
[39] Freixa J., Kim T.-W., Manera A., 2012, Thermal-hydraulic analysis of an intermediate LOCA test at the ROSA facility including uncertainty evaluation, Nuclear Engineering and Design, 249, pp. 97-103.
[40] Freixa J., Kim T.-W., Manera A., 2013, Post-test thermal-hydraulic analysis of two intermediate LOCA tests at the ROSA facility including uncertainty evaluation, Nuclear Engineering and Design, 264, pp. 153-160.
[41] Sozzi G. L., Sutherland W. A., 1975, Critical flow of saturated and subcooled water at high pressure, Report NEDO-13418, General Electric, San Jose, USA.
[42] Rousseau J. C., 1987, Flashing flow, in Multiphase Science and Technology, 3, USA, Hemisphere Publishing Corporation, pp. 378-389.
[43] Berne P., 1983, Contribution à la modélisation du taux de production de vapeur par autovaporisation dans les écoulements diphasiques en conduite, PhD thesis, Ecole Centrale des Arts et Manufactures, Paris, France (in French).
[44] Sekri A., 1982, Cinétique d'ébullition de l'eau en décompression rapide, PhD thesis, Louis Pasteur University, Strasbourg, France (in French).
[45] Marviken Power Station, 1979, The Marviken full scale critical flow tests - Description of the test facility, Technical report MXC-101, Sweden.
[46] Marviken Power Station, 1979, The Marviken full scale critical flow tests - Results from test 13, Technical report MXC-213, Sweden.
[47] Marviken Power Station, 1979, The Marviken full scale critical flow tests - Results from test 17, Technical report MXC-217, Sweden.
[48] Marviken Power Station, 1979, The Marviken full scale critical flow tests - Results from test 24, Technical report MXC-224, Sweden.
[49] Hoshida Y., 2006, Improvement of Analytic Hierarchy Process procedures: quantification of numerous senses of values on evaluation criterion and examination of overall score matrix, Theory and Applications of GIS, 14 (2), pp.63-72 (in Japanese).

[50] Saaty, T.L., 1989, Group Decision Making and the AHP, Golden, B.L., Wasil, E.A., Harker, P.T. (eds) The Analytic Hierarchy Process. Springer, Berlin, Heidelberg.
[51] Lanfredini M., Bestion D., D'Auria F., Aksan N., Fillion P., Gaillard P., Heo J., Karppinen I., Kim K. D., Kurki J., Liu L., Shen A., Vacher J. L. and Wang D., 2020, Critical flow prediction by system codes – Recent analyses made within the FONESYS network, Nuclear Engineering and Design, vol. 366, no. 110731
[52] Szekely G.J., Rizzo M.L., 2013, Energy statistics: A class of statistics based on distances, Journal of Statistical Planning and Inference, 143, pp. 1249–1272.
[53] Fekhari, E., Iooss, B., Muré, J., Pronzato, L., Rendas, J., 2023, Model predictivity assessment: incremental test-set selection and accuracy evaluation, In: Studies in Theoretical and Applied Statistics, SIS 2021, Pisa, Italy, June 21-25, N. Salvati, C. Perna, S. Marchetti and R. Chambers (Eds), Springer Proceedings in Mathematics & Statistics, Springer.