Nuclear Safety
NEA/CSNI/R(2020)16
May 2023
www.oecd-nea.org

SAPIUM: Development of a Systematic Approach for Input Uncertainty Quantification of the Physical Models in Thermal-Hydraulic Codes

Good Practice Guidance

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

Nuclear Energy Agency

NEA/CSNI/R(2020)16

Unclassified English text only

30 May 2023

NUCLEAR ENERGY AGENCY
COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

SAPIUM: Development of a Systematic Approach for Input Uncertainty Quantification of the Physical Models in Thermal-Hydraulic Codes

Good Practice Guidance

This document is available in PDF format only.

JT03520061

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 38 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1 February 1958. Current NEA membership consists of 34 countries: Argentina, Australia, Austria, Belgium, Bulgaria, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, Norway, Poland, Portugal, Romania, Russia (suspended), the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, the United Kingdom and the United States. The European Commission and the International Atomic Energy Agency also take part in the work of the Agency.

The mission of the NEA is:

– to assist its member countries in maintaining and further developing, through international co-operation, the scientific, technological and legal bases required for a safe, environmentally sound and economical use of nuclear energy for peaceful purposes;

– to provide authoritative assessments and to forge common understandings on key issues as input to government decisions on nuclear energy policy and to broader OECD analyses in areas such as energy and the sustainable development of low-carbon economies.

Specific areas of competence of the NEA include the safety and regulation of nuclear activities, radioactive waste management and decommissioning, radiological protection, nuclear science, economic and technical analyses of the nuclear fuel cycle, nuclear law and liability, and public information. The NEA Data Bank provides nuclear data and computer program services for participating countries.

COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

The Committee on the Safety of Nuclear Installations (CSNI) addresses Nuclear Energy Agency (NEA) programmes and activities that support maintaining and advancing the scientific and technical knowledge base of the safety of nuclear installations.

The Committee constitutes a forum for the exchange of technical information and for collaboration between organisations, which can contribute, from their respective backgrounds in research, development and engineering, to its activities. It has regard to the exchange of information between member countries and safety R&D programmes of various sizes in order to keep all member countries involved in and abreast of developments in technical safety matters.

The Committee reviews the state of knowledge on important topics of nuclear safety science and techniques and of safety assessments, and ensures that operating experience is appropriately accounted for in its activities. It initiates and conducts programmes identified by these reviews and assessments in order to confirm safety, overcome discrepancies, develop improvements and reach consensus on technical issues of common interest. It promotes the co-ordination of work in different member countries that serve to maintain and enhance competence in nuclear safety matters, including the establishment of joint undertakings (e.g. joint research and data projects), and assists in the feedback of the results to participating organisations. The Committee ensures that valuable end-products of the technical reviews and analyses are provided to members in a timely manner, and made publicly available when appropriate, to support broader nuclear safety.

The Committee focuses primarily on the safety aspects of existing power reactors, other nuclear installations and new power reactors; it also considers the safety implications of scientific and technical developments of future reactor technologies and designs. Further, the scope for the Committee includes human and organisational research activities and technical developments that affect nuclear safety.

Acknowledgements

The NEA Secretariat wishes to express its gratitude to Jean Baccou from the IRSN and Jinzhao Zhang from Tractebel (ENGIE) for initiating and co-ordinating the SAPIUM project.

Lead authors:

- Jean Baccou, IRSN, France
- Jinzhao Zhang, Tractebel (ENGIE), Belgium

The following experts contributed to the drafting of this report:

- Philippe Fillion, CEA, France
- Guillaume Damblin, CEA, France
- Alessandro Petruzzi, N.IN.E, Italy
- Rafael Mendizábal, CSN, Spain
- Francesc Reventos, UPC, Spain
- Tomasz Skorek, GRS, Germany
- Mathieu Couplet, EDF, France
- Bertrand Iooss, EDF, France
- Deog-Yeon Oh, KINS, Korea
- Takeshi Takeda, NRA, Japan.

The following experts also provided valuable input to various chapters or reviewed drafts of the report:

- Andriy Kovtonyuk, Tractebel (ENGIE), Belgium
- Dominique Bestion, CEA, France
- Estelle Nouy, CEA, France
- Pierre Gaillard, Framatome, France
- Camille Charignon, EDF, France
- Jean-Luc Vacher, EDF, France
- Fabrice Fouet, IRSN, France
- Pierre Probst, IRSN, France
- Jordi Freixa, UPC, Spain
- Óscar Cabellos, UPM, Spain

Many thanks are also due to members of the NEA Working Group on the Analysis and Management of Accidents (WGAMA) for reviewing the final draft of the present report. Additional thanks are extended to Nils Sandberg in support of the organisation of the project and Martina Adorni and Christèle Tephany-M'Pania in support of editing of this report.

Table of contents

List of figures

List of tables

Executive summary

The computer codes currently used for analysis of system-level thermal-hydraulic processes in nuclear reactors are based on models (correlations) that effectively describe physical phenomena such as flow resistance and heat transfer. In recent years, the assessment of uncertainties associated with models and parameters in these system codes has become a key issue in best-estimate plus uncertainty (BEPU) nuclear safety analyses. Forward uncertainty propagation of input uncertainty can be used when a physical parameter is independently known, or can be estimated, and was addressed in several previous Nuclear Energy Agency (NEA) projects such as the uncertainty methods study (UMS) or best-estimate methods uncertainty and sensitivity evaluation (BEMUSE). While the associated output uncertainty quantification (UQ) methods have now been widely used for industrial applications, it was also emphasised that special attention should be paid to input uncertainty quantification (IUQ) on the physical models, or correlations. In general, such model input uncertainty cannot be obtained independently, but must be quantified based on separate effect tests or combined effect tests, which typically involves solving an inverse problem. The NEA's post-BEMUSE reflood models input uncertainty methods (PREMIUM) project was organised as a first step towards the development and the application of model IUQ methods. However, even if the PREMIUM project has been a useful activity to test the different available IUQ approaches, the analysis of PREMIUM phases III and IV results has shown a large dispersion of results between participants. Moreover, the results were even more unsatisfactory when moving from the experiment for quantification to the experiment for validation. One main reason was the lack of common consensus and practices in the followed IUQ process and method.

In the current project, a new systematic approach has been developed, called SAPIUM (systematic approach for input uncertainty quantification methodology) for transparent and rigorous model IUQ. This approach is based on inverse propagation of the information associated with the discrepancy between simulation results and experimental data, using verified and validated simulation models. The starting point of the work performed in SAPIUM is the available state of knowledge from previous related NEA projects and current practices in regulation, industries and research.

The SAPIUM project was organised as a writing group with contributions from ten organisations, including technical support organisations, industry and universities. By clearly compiling the existing approaches for IUQ, as well as the methodological tools to handle each step, the SAPIUM report offers a shared and better understanding about appropriate "good practices" for model input uncertainty quantification. This should improve the reliability of the BEPU analysis and facilitate progress on the validity of extrapolation of its results to the nuclear power plant case. The SAPIUM report does not describe a ready-to-use methodology for model input uncertainty quantification, but rather identifies a number of generic steps and requirements that need to be covered in order to successfully address the key issues identified in previous benchmark studies for IUQ.

The SAPIUM report can be used in two ways, depending on the reader's interests and needs. Firstly, it can be used as a good practice guidance for users of best-estimate codes and BEPU methods to ensure that appropriate modelling choices are made and that important steps are not omitted. Secondly, it can be used at a more detailed technical level to orientate BEPU code and method developers and users in the development or application of IUQ methods.

The SAPIUM approach is split into five elements that are described in this report:

- Chapter 2: Element 1, specification of the problem and requirements;
- Chapter 3: Element 2, establishment[1] and assessment of the experimental database;
- Chapter 4: Element 3, selection and assessment of the simulation model;
- Chapter 5: Element 4, model input uncertainty quantification;
- Chapter 6: Element 5: model input uncertainty validation.

In addition, Chapter 7 deals with scaling considerations and predictive capability assessment, which needs further development before being included in the SAPIUM methodology (see Section 8.3, "Main recommendations and open issues").

Each chapter includes a review of existing methods for each element. The field of applications mainly concerns thermal-hydraulic simulation models, but the generic procedure as well as some of the described tools can be used for other types of models (e.g. neutronics, fuel rod thermal mechanics or coupled multiphysics). The current report clearly emphasises that IUQ should not be reduced to the application of appropriated inverse methods, as was the case during the PREMIUM activity. It is a more general process that involves considering the different elements listed above. The description of each element also revealed the need to use mathematical tools integrating the physical knowledge in order to be fully rigorous, transparent and reproducible.

The SAPIUM project has also identified the following open issues for a complete application of the proposed IUQ approach:

The adequacy evaluation of an experimental database. The experimental database is fundamental to the application of input uncertainty quantification through back-propagation of information. Therefore, it is recommended to follow a structured and transparent approach for the development and assessment of the experimental database. The use of representativeness/completeness indices is discussed in the current report but their application in the framework of IUQ remains an open issue. Moreover, there is no practical solution to deal with the situation when the number of experiments is so restricted that it does not allow constructing an adequate experimental database.

The integration of the state of knowledge coming from Elements 1 (problem specification), 2 (experiments) and 3 (simulation model) in the construction of inverse methods. The majority of the common inverse methods discussed in the current report are advanced statistical tools and rely on a set of assumptions that ensures their rigorous construction. However, the degree of reliability of the information provided by each pair of experimental/simulation values coming from Elements 2 and 3 of the SAPIUM approach is hardly taken into account and further development is needed in the systematic weighting and combination of this information. Moreover, most of the available inverse methods are constructed in the probabilistic framework and adaptation of inverse methods in presence of incomplete knowledge, requiring that the use of alternative uncertainty models also be investigated.

The acceptability of the validation results. Significant efforts should be devoted to the analysis of the validation results in the validation domain with the objective of extrapolation

1. Depending on the type of users, this step consists of compiling a database from available experiments or performing new ones to improve the assessment of the database.

to the application domain. It includes the evaluation of the predictive maturity of the whole IUQ process in order to exploit it in full nuclear power plant applications. Moreover, the acceptability of the validation results strongly depends on the quantification of other uncertainty sources (rather than uncertainties on model parameters) and further work is necessary to understand and evaluate the impact of each category of uncertainty sources on the validation results.

The predictive assessment through the development of new strategies to integrate scaling effects and allow extrapolation to the application domain. System codes used for nuclear power plant safety analysis are validated against integral effect tests, in addition to separate or combined effect tests. These integral effect tests are typically performed on scaled-down models of power reactors; any applications of the validated system codes to the safety analysis of power reactors require a scaling analysis to be carried out, and this analysis needs to account for scaling distortions that inevitably occur. Before the BEPU methodology can be applied to the full nuclear power plant with confidence, how scaling distortions affect the uncertainty quantification should be investigated.

As a first good practice guidance document, the SAPIUM report does not address a full application of the approach to a practical case study whose lessons learnt could be exploited to progress on the above open issues and to confirm or revise some recommendations. Therefore, to complement this work, some demonstration exercises must be performed. Further activities based on progressive test cases (in terms of complexity) are recommended in the conclusion (Chapter 8).

List of abbreviations and acronyms

AA	Average amplitude
ACC	Accumulator
AE	Acceptable error
AEAT	AEA Technology
AEKI	Hungarian Academy of Sciences KFKI Atomic Energy Research Institute
AHP	Analytical hierarchical process
AI	Artificial intelligence
AIAA	American Institute of Aeronautics and Astronautics
AM	Accident management
ASAP	Adjoint sensitivity analysis procedure
ASME	American Society of Mechanical Engineers
ATHLET	Analysis of thermal-hydraulics of leaks and transients
ATWS	Anticipated transients without scram
BE	Best-estimate
BEMUSE	Best estimate methods – uncertainty and sensitivity evaluation
BEPU	Best-estimate plus uncertainty
BIC	Boundary and initial conditions
BWR	Boiling water reactor
CANDU	Canada deuterium uranium (reactor)
CASUALIDAD	Code with the capability of adjoint sensitivity and uncertainty analysis by internal data adjustment and assimilation
CATHARE	Code for analysis of thermalhydraulics during an accident of reactor and safety evaluation
CCFL	Countercurrent flow limitation
CCVM	CSNI Code Validation Matrix
CDF	Cumulative distribution function
CEA	Commissariat à l'Énergie Atomique et aux Énergies Alternatives (French Alternative Energies and Atomic Energy Commission)
CET	Component or combined effect test
CFD	Computational fluid dynamics
CFT	Critical flow test
CIAU	Code with the capability of internal assessment of uncertainty

CIRCÉ	Calcul des incertitudes relatives aux corrélations élémentaires
CPR	Critical power ratio
CSAU	Code scaling, applicability and uncertainty evaluation
CSN	Consejo de Seguridad Nuclear (Nuclear Safety Council, Spain)
CSNI	Committee on the Safety of Nuclear Installations (NEA)
CWO	Core wide oxidation
DA	Data assimilation
DAA	Data adjustment/assimilation
DGSM	Derivative-based global sensitivity measures
DIPE	Determination of input parameter empirical properties
DNB	Departure from nucleate boiling
DNBR	Departure from nucleate boiling ratio
DoE	Design of experiment
DP	Direct or forward problem
DPA	Derivative parameter
ECCS	Emergency core cooling system
ECME	Expectation/conditional maximisation either algorithm
ECR	Equivalent cladding reacted
EDF	Électricité de France
ELECTRE	ÉLimination et choix traduisant la réalité
EM	Evaluation methodology
EMDAP	Evaluation model development and assessment process
ENUSA	Empresa Nacional del Uranio SA
FFTBM	Fast fourier transform based method
FLB	Feedwater line break
FSA	Fractional scaling analysis
FSAR	Final safety analysis report
FT	Fourier transform
GASAP	Global adjoint sensitivity analysis procedure
GP	Gaussian process
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit GmbH
GSA	Global sensitivity analysis
H2TS	Hierarchical 2-tiered scaling
HPI	High-pressure injection

HSIC	Hilbert-Schmidt independence criterion
IAEA	International Atomic Energy Agency
IBLOCA	Intermediate break loss-of-coolant accident
ICAPP	International Congress on Advances in National Power Plants
IE	Improved estimation
IET	Integral effect test
IMSE	Integrated mean square error
IOECCS	Inadvertent operation of ECCS
IP	Inverse problem
IPA	Integral parameter
IPREM	Input parameter range evaluation methodology
IPSN	Institut de Protection et de Sûreté Nucléaire (Protection and Nuclear Safety Institute, France; former name of IRSN)
IR	Importance ratio
IRSN	Institut de Radioprotection et de Sûreté Nucléaire (Radioprotection and Nuclear Safety Institute, France)
ISP	International standard problem
IT	Intermediate test
ITF	Integral test facility
IUQ	Input uncertainty quantification
JAEA	Japan Atomic Energy Agency
KAERI	Korea Atomic Energy Research Institute
KEPRI	Korea Electric Power Research Institute
KF	Kalman Filter
KR	Knowledge ratio
KINS	Korea Institute of Nuclear Safety
KOH	Kennedy and O'Hagan methodology
KREM	KEPRI Realistic Evaluation Methodology
LBLOCA	Large break loss-of-coolant accident
LHS	Latin hypercube sampling
LMO	Local maximum oxidation
(LB/SB) LOCA	(Large break/small break) Loss-of-coolant accident
LOFW	Loss of normal feedwater flow
LOL	Loss of external electric load
LPI	Low-pressure injection

LR	Locked rotor
LS	Least square
LSTF	Large scale test facility
LTCC	Long-term core cooling
LUB	Lower uncertainty bound
LWR	Light water reactor
M&S	Modelling and simulation
MAP (estimator)	Maximum A posteriori (estimator)
MCDA	Model calibration through data assimilation
MCDM	Multi-criteria decision making
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
MLE	Maximum likelihood estimation
NDP	Non dimensional parameter
NEA	Nuclear Energy Agency
NEMM	NINE evaluation model methodology
NINE	Nuclear and industrial engineering S.R.L.
NRA	Nuclear regulation authority
NRI	Nuclear research institute
OAT	One-at-a-Time
OECD	Organisation for Economic Co-operation and Development
P-box	Probability box
PCMM	Predictive capability maturity models
PCT	Peak cladding temperature
PDF	Probability density function
PhW	Phenomenological window
PII	Primary influential inputs
(Q-)PIRT	(Quantitative) Phenomena Identification and Ranking Table
PLI	Perturbed law based indices
PMI	Predictive maturity index
PML	Principle of the Maximum Likelihood
PORV	Power-operated relief valve
PRCC	Partial rank correlation coefficient
PREMIUM	Post-BEMUSE reflood models input uncertainty methods

PRZ	Pressurizer
PSI	Paul Scherrer Institute
PWR	Pressurised water reactor
QoI	Quantity of interest
RCC	(Spearman) Rank correlation coefficient
RCP	Reactor coolant pump
RCS	Reactor coolant system
RDA	Rod drop accident
REA	Rod ejection accident
RELAP	Reactor excursion and leak analysis programme
RIA	Reactivity insertion accident
RLBLOCA	Realistic large break LOCA
ROSA	Rig-of-safety Assessment
RTA	Relevant thermal-hydraulic aspect
RV	Reference value
SA	Sensitivity analysis
SAPIUM	Systematic approach for input uncertainty quantification methodology
SCC	(Pearson) Simple correlation coefficient
SCCRED	Standardised consolidated calculated and reference experimental database
S&E	Science and Engineering
SEM	Stochastic expectation-maximisation algorithm algorithm
SET	Separate effects test
SETF	Separate effects test facility
SG	Steam generator
SGTR	Steam generator tube rupture
SIAM	Society for Industrial and Applied Mathematics
SI	Safety injection
SLB	Steam line break
SM	Simulation model
SPDF	Subjective probability distribution functions
SRC	Standard Regression Coefficient
SRQ	System response quantities
SRRC	Standard rank regression coefficient
SVP	Single valued parameters

TQV	Target quantity for validation
TRACE	TRAC/RELAP advanced computational engine
TSE	Time sequence of events
TSO	Technical support organisation
UA	Uncertainty analysis
UMAE	Uncertainty method based on accuracy extrapolation
UMS	Uncertainty method study
UNIPI	University of Pisa
UPC	Universitat Politècnica de Catalunya (Spain)
UQ	Uncertainty quantification
USNRC	United States Nuclear Regulatory Commission
UUB	Upper uncertainty bound
V&V	Verification and validation
VVER	Water water energy reactor
VVUQ	Verification, validation and uncertainty quantification
WF	Weighted frequency
WGAMA	Working Group on Analysis and Management of Accidents (NEA)
WLS	Weighted least squares

Glossary

This Glossary is constructed from state-of-the-art definitions available in the literature from different frameworks (statistical, V&V, thermohydraulic, etc.). Modifications or extensions of some definitions were required for application to input uncertainty quantification. Therefore, the use of this Glossary is limited to the SAPIUM project.

Acceptability criteria for validation of simulation models: set of predefined requirements that must be met in order to mark the qualification and validation of simulation models as complete.

Acceptability criteria for validation of IUQ: set of predefined requirements that must be met in order to mark the validation of IUQ as complete.

Accuracy (between a simulation model calculation and a given experiment): closeness of an agreement between calculation and experimental value.

Adequacy (of an experimental database): the state of being sufficient for the problem under study. It combines the representativeness of each experiment and the completeness of the database.

Aleatory uncertainty: uncertainty due to the natural variability or randomness of an observed phenomenon and irreducible by arrival of new information.

Applicability of simulation model: capability of a simulation model to calculate a nuclear power plant scenario.

Application domain: operating envelope different from the validation domain.

Bayesian statistical inference: the statistical formalism based on the assumption that uncertain quantities to be estimated are random variables. Prior distribution of such quantities is combined with evidence (e.g. from a sample) via Bayes theorem, producing the posterior distribution.

Best-estimate (calculation): simulation that attempts to predict realistic system response quantities by using the current state of the art for nodalisation.

Bias: estimate of a systematic correction.

Calibration indicator: type of validation indicator checking the agreement between simulation results after input uncertainty propagation and experimental results.

Completeness (of an experimental database): ability of a set of experiments to fulfil the physical conditions of the problem under study.

Confirmation (of an input uncertainty quantification): the process of determining that input uncertainties are compatible with the experimental database used in the quantification.

Consistency indicator: mathematical operator that compares the simulation model calculation with experimental data.

Cumulative distribution function (of a continuous scalar random variable V): the function that gives the probability that V does not exceed a given value, $F_V(v) \equiv PR\{V \leq v\}$.

Epistemic uncertainty (imprecision): uncertainty resulting from a lack of knowledge of information and reducible by arrival of new information.

Evaluation model: calculation framework for evaluating the behaviour of the reactor system during a postulated event, which includes one or more simulation models and all other information needed for use in the target application.

Frequentist statistical inference: the statistical formalism based on the assumption that uncertain quantities to be estimated are fixed but unknown quantities.

Geometrical fidelity of the nodalisation: evaluation and comparison of the geometrical data of the real world design with the numerical values implemented in the nodalisation.

Global sensitivity analysis: the process to study how the variation in the output of a simulation model can be apportioned, qualitatively or quantitatively, to different sources of variation in the model inputs. It provides sensitivity/importance measures/indices.

Important/influential parameter: parameter whose uncertainty contributes substantially to the uncertainty of the system response quantity.

Informativeness indicator: type of validation indicator checking the dispersion of the information obtained after propagation of input uncertainties.

Input deck: necessary data and information to run a simulation.

Input uncertainty (scientific computing): uncertainties related to all uncertainties influencing the simulation model.

Likelihood: given a continuous random variable V having a pdf which depends on a parameter θ, the likelihood function of V is the pdf considered as function of θ, $L(\theta; v) \equiv pdf(v; \theta)$.

Given a simple random sample of V of size N, (v_1,..,v_2), the likelihood of the sample is the product $L(\theta; v_1, \dots, v_N) = \prod_{i=1}^{N} pdf(v_i; \theta)$.

Maximum a posteriori: mode of a posterior distribution.

Mode (of a random variable V): value of V where the pdf of the variable attains a local maximum.

Model input uncertainties: uncertainties related to model inputs.

Model input uncertainty quantification: estimation of the uncertainty associated with model inputs that does not include model calibration.

(Physical) Model: representation of a particular physical phenomenon within a computer code or procedure.

Model calibration: the process of adjusting model selected parameters in order to adapt the model predictions to a set of experimental data.

Model discrepancy/inadequacy: difference between the true value of the real world process and the simulation model output at the true values of the inputs.

Nodalisation: spatial discretisation of the field in which model equations are solved.

Posterior probability distribution (of an uncertain quantity): in Bayesian statistical inference, it is the probability distribution of this quantity, taking into account the prior distribution and the evidence (e.g. sample).

Predictive capability: part of the extended validation process of input uncertainties that includes extrapolation beyond the existing experimental database and acceptability checking for the intended use.

Predictive maturity (of the IUQ results): qualitative or quantitative measure of the applicability of the IUQ results for the intended use in the application domain.

Prior probability distribution (of an uncertain quantity): in Bayesian statistical inference, it is the probability distribution expressing one's beliefs about this quantity before some evidence (e.g. sample) is taken into account.

Probability density function (of a continuous scalar random variable V): the derivative (when it exists) of the cumulative distribution function.

Qualification (of simulation model): the process to determine the degree to which a simulation model is compatible with an intended use when it is not possible to perform a comparison between the simulation model results with respect to a reference.

Representativeness (of an experiment): ability (of an experiment) to provide relevant information for model uncertainty quantification and validation.

Sensitivity analysis: the computation of the effect of changes in input values or assumptions on the system response quantities.

Sensitive parameter: parameter that has a significant influence on the system response quantity.

Simulation model: part of the evaluation model including code, nodalisation and algorithms to approximate the solution of physical equations.

System response quantity: output of interest of a simulation model.

Uncertainty analysis (or quantification): the process to evaluate the overall uncertainty (including statistical uncertainty coming from finite sample size effect) associated with the system response quantity of a simulation model because of uncertainties in the inputs.

Validation (of input uncertainties): the process involving a comparison between the results of input uncertainty propagation and experimental data to determine the degree to which input uncertainties are compatible with an intended use.

Validation (of simulation model): the process involving a comparison between the results of a simulation model and the experimental data to determine the degree to which a simulation model is compatible with an intended use.

Validation domain: subset of points of the input space where experiments have been conducted and validation indicators computed.

Validation indicator: mathematical operator that compares the system response quantity uncertainty coming from the quantified input uncertainties (uncertainty analysis) with experimental data.

1. Introduction

1.1. Context of the SAPIUM project

The assessment of uncertainties associated with best-estimate (BE) calculations has become of prime importance in best-estimate plus uncertainty (BEPU) nuclear safety analyses. From a methodological point of view, the treatment of uncertainties can be split into two main topics that require different approaches. The first one is focused on the quantification of uncertainty sources and mainly model input uncertainties quantification (IUQ), i.e. uncertainties associated to model input parameters. It is based on the discrepancy between simulation and experiment and its inverse propagation to derive input uncertainties (Tarantola, 2005). In the second one, it is assumed that all uncertainty sources have been previously determined, and the objective is to estimate their impact on uncertainties associated to system response quantities (SRQs). It is referred to as uncertainty analysis (UA) or uncertainty quantification (UQ) and is usually based on input uncertainty (forward) propagation (Smith, 2014).

The question of input uncertainty propagation has already been addressed by several NEA projects such as UMS (NEA, 1998) or BEMUSE (NEA, 2011). While it appears that uncertainty analysis methods have now become mature for industrial applications, a special attention should be devoted to the input uncertainty quantification on the physical models. Therefore, following this recommendation, the PREMIUM (NEA, 2013) benchmark (2012-2015) was organised as a first step towards the development and the application of model IUQ methods.

However, even if this project has been a useful activity to test the different available IUQ approaches, the analysis of PREMIUM phases III and IV has shown a large dispersion between participants. Moreover, the results were not satisfactory when moving from the experiment used for quantification (FEBA) to the experiment used for validation (PERICLES). One main reason could be attributed to the lack of common consensus and practices in the followed process and method (NEA, 2017).

A main recommendation from the PREMIUM benchmark was that a systematic approach devoted to model input uncertainty quantification and validation should be developed to improve the reliability of the analysis and to ensure the extrapolation of its results to the nuclear power plant case. Therefore, following a first investigation by IRSN, Tractebel and CEA (Baccou et al., 2017) that led to the identification of five key generic elements that should be considered in the construction of a systematic approach, the NEA SAPIUM project was proposed to progress on the issue of the quantification and validation of the uncertainty of the physical models in thermal-hydraulic codes.

1.2. Scope, limitation and organisation of the SAPIUM project

1.2.1. Scope and limitation

SAPIUM (systematic approach for input uncertainty quantification methodology) is devoted to model input uncertainty quantification based on inverse propagation of the information associated to the discrepancy between simulation results and experimental data, using verified and validated simulation models. In the literature, the term "quantification" can be similar to the term "calibration". In this project, the two terms are clearly distinguished: calibration is related to adjusting input parameter value whereas quantification is focused on uncertainty estimation without recalibration (as recommended in the PREMIUM benchmark) since the simulation model is assumed to be verified and validated. The objectives of the project are:

- to construct a systematic procedure (following different elements in a step by step approach) to perform a meaningful model input uncertainty quantification and validation;
- to address the identified key issues for each step;
- to provide "good practice guidelines" recommendations on the tools and methods to be used for each step;
- to provide, if needed and possible, simple illustration examples for application of the tools and methods;
- to identify the open issues for future development (including a more complete demonstration of the tools and methods in the next step).

This work can be used to minimise (or at least allow to understand) the user effect identified in previous projects. By clearly compiling the different approaches as well as the methodological tools to handle each step, the SAPIUM framework offers a shared understanding about "appropriate" practices for model input uncertainty quantification in order to improve the reliability of the analysis and to progress on the validity of extrapolation of its results to the nuclear power plant case. Therefore, the main outcome of the project is a first "good practice guidance" document that can be exploited for safety studies to increase agreement among experts on recommended practices as well as on open issues. The SAPIUM approach is developed mainly for thermal-hydraulic models but the generic procedure as well as some of the described tools can be used for other types of models. End users are research institutes and universities, manufacturers, utilities and safety authorities. In other words, they are the developers and the users of BEPU approaches, as well as the organisations in charge of evaluating these approaches.

This project can be considered as a follow-up of PREMIUM. However, it is not a benchmarking of available methods for model input uncertainty quantification but provides a methodological document (only simple additional studies will be considered to get reliable insights into methodological key steps). It is also important to mention that it is not intended to develop in this project a unique (statistical) method for model input uncertainty quantification but to provide the description of the different generic steps and requirements that a method has to successfully address the key issues identified in previous benchmark studies. Therefore, it is not expected to derive at the end of the project certified model input uncertainties to be used in nuclear power plant studies. Before using the SAPIUM approach in nuclear safety studies, a full application would be necessary to better specify some key

issues of its development and to confirm or revise the SAPIUM recommendations. This task is also out of the scope of this project and could be the topic of further projects.

In scientific computing, there exist different sources of uncertainties that can be categorised in three classes (Roy and Oberkampf, 2011):

- 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 such as space-time discretisation or iterative convergence errors.
- The last one concerns model form and includes all assumptions, conceptualisations, abstractions, approximations and mathematical formulations on which the model relies.

As mentioned previously, in SAPIUM, the focus is on the first class and more precisely on input parameters involved in the physical models implemented in the code.

1.2.2. Organisation

The SAPIUM project is organised as a writing group. The contributors are from ten organisations and include technical support organisations (TSO), industry and universities. Each contributor provided its own expertise on input uncertainty quantification and validation through the writing of assigned sections and the review of the final report. Moreover, if deemed necessary, they performed simple illustrating applications to facilitate the discussions and reach a consensus on the SAPIUM process.

The SAPIUM project lasted two and a half years (January 2017–September 2019). The milestones of the project are given in the Annex A (Table A.1).

The starting point of the SAPIUM project was the available knowledge from previous relevant NEA projects as well as current practices in regulation, industry and research. They are briefly reviewed in the next section.

1.3. Review of previous projects and current practices

1.3.1. NEA UMS and BEMUSE Projects

Several international initiatives such as the uncertainty methods study (UMS) (NEA, 1998) and the best-estimate methods – uncertainty and sensitivity evaluation programme (BEMUSE) (NEA, 2011) have been organised to progress on the issue of BEPU approach with the objectives of stimulating discussions, developing a common understanding and consolidating methods. The UMS study was the first international study on the uncertainty analysis methodologies applied to a 5% break SBLOCA experiment (LSTF SB-CL-18), in the Japanese large scale test facility (LSTF). The UMS group, following a mandate of the CSNI, has compared five methods for evaluating the uncertainty in the predictions of advanced “best-estimate” thermal-hydraulics codes: four of them (GRS, ENUSA, AET and IPSN) were based on a propagation of input uncertainties through the simulation model; one (UMAE) was based on accuracy extrapolation. For some methods, it was the first full application and it constitutes a pilot study that allowed a better understanding of the underlying assumptions at the basis of each method.

The participants calculated uncertainty ranges for experimental parameters including pressurizer pressure, primary circuit inventory and clad temperature as function of time. Following the objective of the study, the participants compared:

- the methods, step by step, when applied to LSTF SB-CL-18;
- the uncertainties predicted with each other;
- the uncertainties predicted with measured values.

The probabilistic methods (ENUSA, GRS and IPSN) share a same set of assumptions, i.e. i) the knowledge of the values of the quantities of interest, which shall be consistent with the available evidence; ii) the use of subjective probability distribution functions (SPDF) to describe the input uncertainties; and iii) the uniform distributions for describing the input uncertainty variations, particularly when little knowledge is available. The AEA Technology (AEAT) and UMAE methods, in contrast, have unique assumptions. For example, the AEAT method is based on a bounding analysis performed to predict uncertainty ranges based on the available evidence and the UMAE method requires satisfaction of a number of statistical conditions.

Where the predictions of the methods differ, the differences have been accounted for in terms of the assumptions of the methods and the input data used. The calculated ranges bound the experimental results with some exceptions and the possible causes of these discrepancies have been identified.

The BEMUSE programme – promoted by the Working Group on Analysis and Management of Accidents (WGAMA) and endorsed by the Committee on the Safety of Nuclear Installations (CSNI) – represented an important step towards reliable application of high-quality best-estimate and uncertainty and sensitivity evaluation methods. The main goals of the programme were:

- to evaluate the practicability, quality and reliability of best-estimate (BE) methods including uncertainty evaluations in applications relevant to nuclear reactor safety;
- to develop a common understanding in this domain;
- to promote/facilitate the use of these methods by the regulatory bodies and the industry.

The BEMUSE programme was divided into two main steps, each consisting of three phases. The first step is to perform an uncertainty and sensitivity analysis related to the LOFT L2-5 test, and the second step is to perform the same analysis for a nuclear power plant large break loss-of-coolant accident (LBLOCA). The programme started in January 2004.

- First step (Phases I, II and III):
    - Phase I: presentation “a priori” of the uncertainty and sensitivity analysis methodology to be used by the participants (NEA, 2005).
    - Phase II: re-analysis of the ISP-13 exercise, post-test analysis of the LOFT L2-5 large cold leg break test calculation (NEA, 2006).
    - Phase III: sensitivity and uncertainty analyses of the L2-5 test calculations, first conclusions on the methods and suggestions for improvement (NEA, 2007).

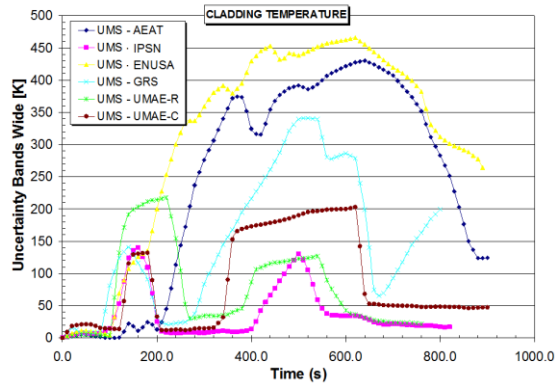
- Second step (Phases IV, V and VI):
  - Phase IV: best-estimate analysis of a nuclear power plant -LBLOCA (NEA, 2008).
  - Phase V: sensitivity and uncertainty analyses for the nuclear power plant - LBLOCA, with or without methodology improvements resulting from Phase III (NEA, 2009).
  - Phase VI: status report on the area, classification of the methods, conclusions and recommendations (NEA, 2011).

The BEMUSE activity compared the applications of essentially two UA methods. The first method uses a probabilistic approach, which propagates input uncertainties to the SRQ uncertainties (Glaeser et al., 1994). The method is associated with order statistics and the use of Wilks' formula (Wilks, 1942). That method was first proposed by GRS, and applied by the majority of participants. The second method is the uncertainty method based on accuracy extrapolation (UMAE) (D'Auria and Galassi, 1995) and its extension to the code with the capability of internal assessment of uncertainty (CIAU) (D'Auria and Giannotti, 2000; Petruzzi et al., 2005). A comparison of calculation results with data from integral experiments investigating the same course of events for which the uncertainty analysis is to be performed is a necessary basis for applying this method.

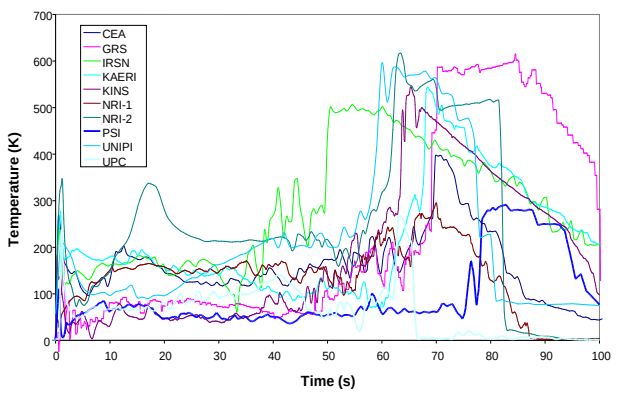
The participants using these methods applied different computer codes. A direct comparison between both applications of the above-mentioned methods (CIAU and GRS) can only be seen for applications of the CATHARE and RELAP5 codes. However, each participant used different nodalisations and partly different code options. Therefore, comparisons of both uncertainty methods in this programme are influenced by user-specific applications of a computer code or by using different computer codes.

Before summarising the lessons learnt from the two international projects, it might be useful to compare some of the results of the BEMUSE Project Phase III (NEA, 2007) with the UMS (NEA, 1998). Figure 1.1 provides a sense of the magnitude of the discrepancy among the results predicted by participants to the NEA UMS and BEMUSE benchmark, respectively, adopting very similar UA methods (mostly of them applied methods based on propagation of input uncertainty parameters) and in some cases using also the same thermal-hydraulic code. It can be noted that the spread in the size of the uncertainty bands (for cladding temperatures in Figure 1.1) predicted by the UMS participants in 1995 was not reduced by the participants to the BEMUSE benchmark, which was held more than ten years later.

Figure 1.1. Spread of uncertainty bands for cladding temperatures predicted by UMS and BEMUSE participants



UMS Participants (1995)



BEMUSE Participants (2006)

Sources: NEA, 1998 and NEA, 2007

The reasons that explain the differences between the results of users of the same codes and (almost) same UA methods are several and coincide with some of the main outcomes from the BEMUSE project (NEA, 2011):

- Differences between applications of statistical methods may mainly be due to procedures adopted:
  - to select the input parameters;
  - to properly quantify input parameter uncertainties;
  - to properly estimate the probability distribution functions.
- Differences in reference calculations.
- Some participants using statistical methods specified few important parameters or overly narrow uncertainty ranges for important input uncertainties based on expert judgement, and not on sufficient code validation experience. Therefore, skill, experience and knowledge of the users about the applied suitable computer code as well as the used UA method are important for the quality of the results (an uncertainty methodology user-effect was characterised).
- An increased number of calculations for the statistical methods may be advisable because it decreases the dispersion of the tolerance limits.
- Differences between the methods based upon propagation of input uncertainties and methods based upon propagation of output errors may come from different experimental data bases used for the analysis.

As a conclusion, as mentioned in the final BEMUSE report, significant efforts are required on the quantification of input uncertainties, which is a compulsory step when performing, for example, a probabilistic uncertainty analysis. More precisely, high requirements are needed in the determination and the justification of the uncertainty range associated with uncertain model input parameters. This quantification is often performed by subjective engineering judgement and therefore requires further development to provide a common

understanding on this key issue and to reduce as much as possible the uncertainty methodology user-effect.

1.3.2. NEA PREMIUM benchmark

PREMIUM ("Post-BEMUSE reflood models input uncertainty methods") was a benchmark endorsed by the WGAMA Group of NEA/CSNI. Its main goal was the development, study, comparison and application of methods for quantification of the uncertainty of the physical model parameters contained in thermal-hydraulic codes used in nuclear safety. It was addressed to model input uncertainty quantification based on the so-called "intermediate" tests (ITs), which are relatively simple experiments with few phenomena and models involved. The application of these methods may help to reduce the dependency on expert judgement in model input uncertainty quantification.

The benchmark application was focused on the physical models involved in the prediction of the core reflood, which is a fundamental stage in the loss-of-coolant accident (LOCA) scenario.

Sixteen organisations participated in the PREMIUM benchmark. A majority of them were involved in all the phases of the benchmark. The participants quantified the model input uncertainties using results of selected reflood tests in FEBA facility, and applying different methods. The resulting uncertainties were verified (or confirmed if the PREMIUM terminology is followed) and validated by propagating them in simulations of all FEBA Series I tests (the same ones used for quantification), and reflood tests in the PERICLES facility, and comparing the results with the experimental data. The results of these model input uncertainty propagation exercises were analysed with two different procedures: a qualitative procedure established by CEA and a quantitative procedure developed by IRSN.

PREMIUM was organised in five consecutive phases:

- In Phase I (co-ordinated by UPC and CSN), the participants presented and described their methods of model input uncertainty quantification.
- In Phase II (co-ordinated by University of Pisa) participants identified influential input parameters, from the point of view of reflooding, and made a preliminary quantification of their variation range.
- In Phase III (co-ordinated by GRS), the uncertainty of influential input parameters (identified in Phase II) was quantified.
- In Phase IV (co-ordinated by CEA and IRSN), quantified model input uncertainties of Phase III were propagated through the simulation model for the selected tests of FEBA and PERICLES experiments, and compared to the experimental results. The objective was to confirm (in FEBA) and validate (in PERICLES) the model input uncertainties.
- In Phase V (co-ordinated by CSN and UPC), the main conclusions and lessons learnt of the benchmark were drawn, lines of future work were proposed, and the final report of PREMIUM (the present report) was compiled and written.

The PREMIUM benchmark has been a valuable exercise on methods of uncertainty quantification of physical model parameters, and their application to the models involved in the reflooding prediction. Different methods and thermal-hydraulic codes were used along the benchmark. Results were more dependent on the quantification method than on

the code employed. Furthermore, the results of quantification showed a strong dependency on factors such as:

- the set of selected SRQs used in the quantification;
- the set of selected input parameters to be quantified;
- selected tests for quantification;
- the simulation models, which, in general depend on the TH code being used.

Nonetheless, it was noticed that there was still a lack of clear "best practice" guidelines on these topics. Indeed, participants in PREMIUM tackled the issue by following different approaches according to their own experience or procedures. It was concluded that the IUQ methods used in PREMIUM showed a strong user effect. As a final outcome, the results of quantified model input uncertainties in PREMIUM showed a large variability and discrepancy among participants.

PREMIUM has been useful as a test bed for inverse uncertainty quantification methods. Some of the used methods were developed or improved in the course of the participation in the benchmark. Nevertheless, the application conducted in PREMIUM did not allow for a deep and general assessment of the quantification methods, because it was limited to a number of experimental tests concerning a specific scenario (the PWR reflooding).

The benchmark has revealed the necessity of further work on inverse methods, and on the development of guidelines for quantification and validation of model input uncertainties. In the case of methods having the option of performing calibration in addition to the model input uncertainty quantification, better results were obtained when such recalibration was omitted.

The propagation of the quantified model input uncertainties to FEBA tests has given better results than the analogous exercise for PERICLES tests, in the sense that the calculated uncertainty bands for selected SRQs enveloped the real data in a larger percentage of cases.

PREMIUM has also been useful in testing a methodology, developed by IRSN, for analysis of uncertainty bands calculated for selected SRQs, based on the computation of two features: informativeness (depending on the width of the band) and calibration[2] (depending on the closeness of predictions to experimental values).

Despite the strong user effect revealed in the exercise, the benchmark identified some lessons learnt and recommendations that are highlighted hereafter. Methods for input uncertainty quantification of the physical models in system TH codes must be further studied and developed, so that their different performances can be understood. The selection of input parameters, SRQs and experimental database are fundamental parts of IUQ methods. Guidelines and procedures should be established for such processes. The quantified uncertainty obtained for a specific parameter strongly depends on the total set of parameters being simultaneously quantified.

In addition, the outcome of PREMIUM shed light on the direction of future advances in this area. The most important recommendation was that future work in the issue of BEPU

2. The term *calibration* is used here for uncertainty assessment as in *Experts in Uncertainty* (Cooke, 1991). It has therefore a different meaning from *model calibration*, which refers to the adjustment of the model parameters to match better the experimental data.

analyses should be focused on the development of a systematic approach, including common "best practice" guidelines on:

- construction of the numerical methods to infer model input uncertainties from the comparison between simulations and experiments;
- selection of the SRQs and the experimental database used for model input uncertainty quantification;
- selection of uncertain input parameters;
- assessment and selection of the simulation model.

More detailed information on the PREMIUM benchmark can be found in the reports produced during the exercise. There are two levels of detail. A first approach can be found in two CSNI reports (NEA, 2016) and (NEA, 2017) while a comprehensive view is shown in the phase reports that were issued as WGAMA documents (NEA, 2015; NEA, 2014a; NEA, 2014b).

1.3.3. USNRC CSAU and EMDAP processes

The development of systematic approaches has been already widely adopted in regulatory requirements and industrial applications for best-estimate plus uncertainty (BEPU) methodology development.

The code scaling, applicability and uncertainty (CSAU) methodology (Boyack et al., 1989) is a structured, traceable and practical approach to quantify uncertainty. It addresses in a unified and systematic manner questions related to the scaling applicability of best-estimate code, to its applicability to scenarios of interest to nuclear power plant safety studies and to the quantification of uncertainties associated to selected SRQs when the code is used to perform a calculation for a specified scenario and nuclear power plant design. It was recommended by the group of experts as an acceptable methodology to develop a BEPU methodology for best-estimate loss-of-coolant accident (LOCA) analysis that complies with the USNRC regulatory guide RG-1.157 (USNRC, 1988).

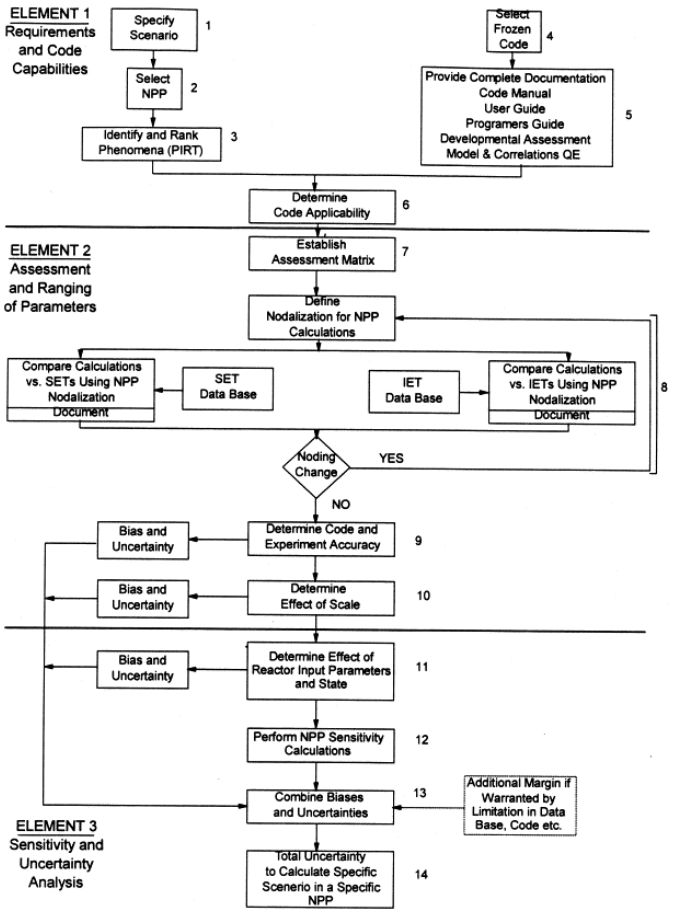
As shown in Figure 1.2, the CSAU methodology consists of 14 steps organised into three major elements:

- Element 1 - Requirements and code capabilities[3]: Scenario-modelling requirements are identified in a Phenomena Identification and Ranking Table (PIRT), and compared against code capabilities to determine the codes applicability to the particular scenario and to identify potential limitations.
- Element 2 - Assessment and ranging of parameters: code capabilities to calculate processes important to the scenario are assessed against experimental data to determine code accuracy and scale-up capability and to specify ranges of input parameter variations needed for sensitivity studies.
- Element 3 - Sensitivity and uncertainty analysis: the effects of individual contributors to total uncertainty are obtained and the propagation of uncertainty through the transient is properly determined.

3. In the CSAU framework, code capabilities and accuracy assessment require the definition of a nodalisation strategy and therefore are related to what is called in SAPIUM simulation model.

The early USNRC-approved best-estimate LOCA evaluation methodologies (EMs), such as Westinghouse's BELOCA (Young et al., 1998) and ASTRUM (Frepoli, 2008) and AREVA's RLBLOCA (Martin and O'Dell, 2005), followed the CSAU approach. Similar methodologies have been developed and approved in other countries, such as the ESM-3D in France (Sauvage and Keldenich, 2005) or KREM in Korea (Ban et al., 2004).

Figure 1.2. Code scaling, applicability and uncertainty (CSAU) evaluation methodology



Source: Boyack et al., 1989

The CSAU approach described above was recently endorsed as an acceptable structured process in the RG-1.203 "transient and accident analysis methods" (USNRC, 2005). It describes a process that the USNRC considers acceptable for use in development and assessment of evaluation models (i.e. collection of codes and procedures) that may be used to analyse transient and accident behaviour within the design basis of a nuclear power plant. This process is called the evaluation model development and assessment process (EMDAP).

The EMDAP essentially follows the same main principles of the CSAU methodology described above, but with more emphasis on the evaluation model development process. The EMDAP consists of four elements, each consisting of several steps:

- Element 1 – Establish requirements for evaluation model capability[4]: In this first element, the exact application envelope for the evaluation methodology is

4. In the EMDAP framework, the evaluation model refers to the computer code and the input model, including the nodalisation strategy and model options. It is called simulation (or computational) model in SAPIUM.

determined. Furthermore, the importance of constituent phenomena processes and key parameters within this envelope has been agreed upon.

- Element 2 – Develop assessment base: In this second element, the purpose is to provide the basis for development and assessment of the evaluation methodology. This includes acquiring appropriate experimental data relevant to the scenario being considered and ensuring the suitability of experimental scaling.
- Element 3 – Develop evaluation model: In this third element, the evaluation model is developed and organised to meet the requirements defined in element 1.
- Element 4 – Assess evaluation model adequacy: In this fourth and last element, the adequacy and capability of the evaluation model are assessed and documented. In the final step of this last element, it is to be decided whether the evaluation model is adequate or not.

The recently USNRC-approved best-estimate LOCA evaluation methodologies (EMs), such as Westinghouse's FSLOCA (Frepoli, 2013), followed the EMDAP approach.

1.3.4. Industrial VVUQ formal procedure

The development of formal procedures for code verification, validation and uncertainty quantification (VVUQ) remains an active research field of interest. A comprehensive framework for verification, validation and uncertainty quantification in scientific computing was proposed in "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing" (Roy and Oberkampf, 2011).

In VVUQ procedure, the validation process can be performed by way of comparison of simulated results with available experimental measurements. It is performed at the conditions where experimental data are available and the validation is addressed by defining and computing the difference, i.e. validation metrics associated to SRQs (Oberkampf and Barone, 2006). If the calculated validation metric result meets the accuracy requirements, the computational (or simulation) model is considered as adequate to the application of interest (i.e. used for BEPU methodology development). If the accuracy requirements are not met, one may need to either update (or recalibrate) the computational (or simulation) model or improve or add experimental measurements. Note that recalibration should be made in the computational (or simulation) model development process, which is out of the scope of SAPIUM.

In this VVUQ framework, *uncertainty quantification* defines the set of tools and formalisms that are both related to input parameters and SRQs. It includes the following key steps:

1. the identification of all sources of uncertainty associated to model inputs, numerical approximation and model form;
2. the characterisation of uncertainties including the quantification of input uncertainties;
3. the estimation of the uncertainty due to numerical approximation;
4. the propagation of quantified model input uncertainties through the computational model[5] to obtain uncertainties in the SRQs;

5. In the VVUQ framework, the computational model includes the computer code, nodalisation strategy and model options, which is called simulation model in SAPIUM.

5. the estimation of the model form uncertainty;
6. the determination of the total uncertainty in the SRQs at the application conditions of interest.

The SAPIUM project concerns the quantification of model input uncertainties in step (2), but they must be validated through the process of uncertainty estimation coming from different sources in steps (3) to (5). The final objective is to extrapolate the total uncertainty to the application conditions of interest for where no experimental data is available (step (6)). The extrapolated uncertainty is included in the prediction of the computational (or simulation) model at the conditions of interest.

Concept of predictive capability has also been developed over the last two decades (Oberkampf et al., 2004; Oberkampf et al., 2007; Rider et al., 2015), which focuses on predicting the performance, safety and reliability of systems that have not been tested, based on the high-fidelity physics simulations, including estimates of all sources of uncertainty and separated mathematical representation of aleatory and epistemic uncertainties. Different predictive capability maturity models (PCMM) have been developed for assessment of the maturity of the simulation model (Oberkampf et al., 2007). PCMM provides a structured breakdown of the component work within an engineering simulation, which contains six elements: geometry/representation fidelity, model fidelity, code verification, solution verification (numerical error estimation), validation and uncertainty quantification/sensitivity analysis. Each of these elements entails significant complexity and contributes to the overall quality. The PCMM differs from traditional approaches (like CSAU and EMDAP) used in nuclear power engineering, as it covers much broader scope than propagation of input to output uncertainties and does not use static or dynamic scaling approaches.

Various professional bodies, such as the AIAA and ASME, have also developed standards for verification and validation of computer codes for thermal hydraulics simulations (AIAA, 1998; ASME, 2009) for example). In addition, the development of new procedures for the treatment of model input uncertainty in thermal-hydraulic codes for nuclear power plants remains an active research field of interest (among various works, one can cite for example (Pourgol-Mohamad et al., 2011; Hemez et al., 2010; or Unal et al., 2011). Therefore, the SAPIUM project also exploits the current state of knowledge in R&D to develop the systematic approach. This point is developed further in each chapter.

1.4. Key elements and major steps of the SAPIUM approach

Following the main conclusion of the PREMIUM activity, and based on the experience feedback from these previous benchmarks, it has been proposed to develop a systematic approach for model input uncertainty quantification that could improve the reliability of the analysis and the confidence on the extrapolation of its results to the nuclear power plant case. This led to the definitions of five key elements that provide a general framework for the construction of IUQ methods and that is the starting point of the SAPIUM activity (Baccou et al., 2017). They are shown in Figure 1.3.

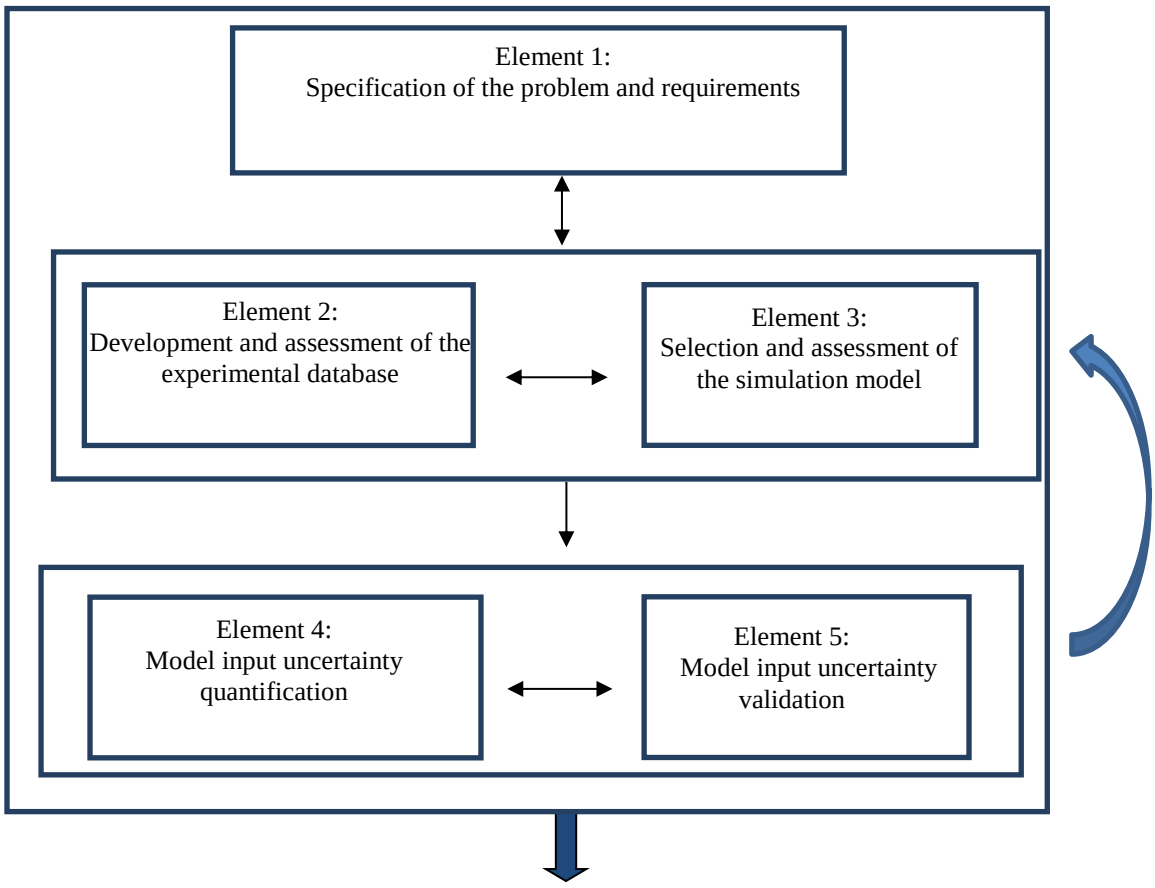
Element 1 (specification) is common to any kind of nuclear safety analysis studies. Elements 2 (experimental database) and 3 (simulation model) provide the information for the model input uncertainty quantification and validation. Interactions between them (depicted by the two-sided arrow on Figure 1.3) are required for their construction. Element 1 might also require using code and sensitivity analysis tools (to confirm a PIRT for example) addressed in Element 3. Since IUQ methods are based on the comparison

between simulation model results and experimental values, Elements 2 and 3 are crucial for their development. They will control the reliability of the final model input uncertainties and the capability of the method to extrapolate the results to real situations.

Element 4 consists in inferring, from the comparison between simulation model calculations and experimental values, the information related to model input uncertainties. The experimental knowledge is here associated with a subset of the database constructed in Element 2 (the remaining subset will be used for model input uncertainty validation).

Finally, the validation performed in Element 5 is based on the propagation of all input uncertainties (integrating the quantified model input uncertainties obtained in Element 4) through the simulation model. It also exploits the experimental database identified in Element 2. This last element is interacting with Element 4 (depicted by the two-sided arrow on Figure 1.3) and might be used in an iteration process with Elements 2 to 4 (as indicated by the blue arrow) in case of non-acceptable validation results.

Figure 1.3. The five key elements of the IUQ framework



The SAPIUM elements are restricted to the methodological steps to validate quantified uncertainties in the validation domain. The acceptability of the results for application to plant analysis requires one extra step related to the assessment of the scaling adequacy and of the predictive capability of the simulation model and the quantified model input uncertainty for extrapolation to application of interest (full-scale nuclear power plants). Although application to nuclear power plants is out of the scope of this work, Chapter 7 describes some good practices to progress on the assessment of the scaling adequacy, which may need further development in future projects.

As described by Figure 1.3, the SAPIUM approach is related closely to VVUQ procedure. Although part of the process is common (Elements 1-3), the systematic approach is focused on the quantification of model uncertainty itself (Element 4), and the IUQ validation (Element 5) as well the scaling issue and predictive capability assessment addressed in Chapters 7 and 8 are related to the adequacy assessment for nuclear power plant applications. The conclusions and results of previous work can, however, be adapted to the framework of IUQ methods. As an example, the sequel describes how the background recalled in Section 1.2 can contribute to the development of each element. Elements 1-3 are common to any BEPU methodology focusing on the application of fully verified and validated simulation models for accident analysis and can therefore benefit from the CSAU and EMDAP practices. The PREMIUM benchmark has been devoted to Elements 4 and 5 and its conclusions are valuable to progress on these parts of the methodology.

It is important to notice that the sources of discrepancy between participants recalled in Section 1.3 that prevented a consensus from being reached on the final model input uncertainties are included in Elements 1, 2, 3 and 4, respectively. They will be integrated in the list of open questions that could be addressed during the SAPIUM project. Moreover, the development of Elements 3 and 5 can exploit the BEMUSE contributions that focused on several issues such as nodalisation strategy and model options or selection of input parameters and input uncertainty propagation. Finally, the large literature on VVUQ approaches is of prime importance especially for Element 5 and the construction of validation metrics.

A questionnaire was sent to participants[6] before the beginning of the SAPIUM project in order to establish a list of key issues to describe each element of Figure 1.3. All the organisations were asked also to provide a knowledge level ranking on existing methods or practices to handle them. The list of key issues and the analysis of the answers are given in Annex A (Table A.2 and Figure A.1). This work makes it possible to identify the major steps (Table 1.1) of the systematic approach for IUQ to be addressed in each element.

Table 1.1. Major steps of the SAPIUM approach

Key elements	**Major steps**
Element 1: Specification of the problem and requirements (Chapter 2)	Step 1: Specification of the IUQ purpose Step 2: Selection of system response quantities Step 3: Identification of important phenomena (PIRT)
Element 2: Development and assessment of the experimental database (Chapter 3)	Step 4: Establishment of a list of the available experiments and standardised description of each experiment including evaluation of experimental uncertainties Step 5: Assessment of the representativeness of each experiment and of the completeness of the database Step 6: Selection of the experimental database for the quantification and the validation according to the representativeness and completeness assessment
Element 3: Selection and assessment of the simulation model (Chapter 4)	Step 7: Selection of code based on capability assessment Step 8: Development of the simulation model (SM) and assessment of applicability Step 9: Selection and specification of uncertain input parameters and confirmation by sensitivity analysis
Element 4: Model input uncertainty quantification (Chapter 5)	Step 10: Aggregation of the information coming from the experiments of the database and from the simulation model to be used in the "inverse propagation" Step 11: Quantification of model input uncertainties by appropriate "inverse propagation" methods Step 12: Combination of model input uncertainties if several quantifications are performed Step 13: Confirmation by counterpart tests
Element 5: Model input uncertainty validation (Chapter 6)	Step 14: Determination of numerical approximation and other input data uncertainties for each validation case Step 15: Propagation of all input uncertainties through the simulation model Step 16: Computation of validation indicators for the comparison between the simulation model output uncertainty and experimental data used for validation Step 17: Analysis of the validation results and iteration quantification/validation in a loop-approach if the acceptability is not reached

6. Bel V also answered the questionnaire, but did not participate in SAPIUM.

Table 1.2. Major steps of the SAPIUM approach (Continued)

Key elements	Major steps
Scaling issues (Chapter 7)	Good practices for assessment of the scaling adequacy of the simulation model and the quantified model input uncertainty for extrapolation to application of interest (full-scale nuclear power plants)

In the following chapters of this document, the above-identified major steps will be addressed in detail, aiming to answer the questions raised in PREMIUM:

- What is the objective of the IUQ? Chapter 2 specifies the problem and provides guidelines for identification and ranking of important phenomena.
- Why are the chosen FEBA and PERCICLES test data not sufficient or adequate for quantification of the model input uncertainties related to the reflood heat transfer? Chapter 3 will provide guidelines for assessing the adequacy of a database (SETs/IETs).
- What are the best practices for modelling of SETs/IETs to reduce the dispersions? Chapter 4 provides guidelines for the assessment of the applicability of the codes for modelling of the identified important phenomena.
- Why are the quantification results methods dependent? Chapter 5 provides practical guidelines for the choice of IUQ method.
- What are the criteria for accepting the IUQ results? Chapter 6 provides practical guidelines for validation of the quantified uncertainties.
- Is it possible to extrapolate the quantified and validated uncertainties to nuclear power plant applications? Chapter 7 provides some good practices for the assessment of the scalability of the quantified uncertainties from SETs to IETs and from SETs/IETs to nuclear power plant applications.

References

AIAA (1998), Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, American Institute of Aeronautics and Astronautics, AIAAG-077-1998.

ASME (2009), Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer, ASME V&V 20-2009.

Baccou, J., J. Zhang and E. Nouy (2017), "Towards a systematic approach to input uncertainty quantification methodology", in Proceedings of the NURETH-17 Conference, Xi'an, China.

Ban, C.-H., S.-Y. Lee and C.-K. Sung (2004), "Development and Application of KEPRI Realistic Evaluation Methodology (KREM) for LB-LOCA", in Proceedings of the International meeting on updates in best estimate methods in nuclear installation safety analysis (BE-2004), pp. 14-18, Washington, DC, United States.

Boyack, B., R. Duffey, G. Wilson, P. Griffith, G. Lellouche, S. Levy, U. Rohatgi, W. Wulff, W. and N. Zuber (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, United States.

Cooke, R.M. (1991), *Experts in Uncertainty*, Oxford University Press.

D'Auria, F. and W. Giannotti (2000), "Development of Code with capability of Internal Assessment of Uncertainty", in *Journal of Nuclear Technology*, 131(1), pp. 159-196.

D'Auria, F. and G. Giorgio (1995), "Outline of the Uncertainty Methodology based on Accuracy Extrapolation (UMAE)", in *Journal of Nuclear Technology*, 109 (1), pp. 21-38.

Frepoli, C. (2013), "The development of a Realistic LOCA Evaluation Model Applicable to the Full range of Break sizes: Westinghouse Full Spectrum™ LOCA (FSLOCA™) Methodology", Regulatory Information Conference (RIC), USNRC.

Frepoli, C. (2008), "An Overview of Westinghouse Realistic Large Break LOCA Evaluation Model", in *Science and Technology of Nuclear Installations*, dx.doi.org/10.1155/2008/498737.

Glaeser, H., E. Hofer, M. Kloos and T. Skorek (1994), "Uncertainty and Sensitivity Analysis of a Post-Experiment Calculation in Thermals Hydraulics", in *Reliability Engineering and System Safety*, 45, pp. 19-33.

Hemez, F., H. Sezer Atamturkur and C. Unal (2010), "Defining predictive maturity for validated numerical simulations", in *Computers & Structures*, 88(7-8), pp. 497-505.

Martin, R.P. and L.D. O'Dell (2005), "AREVA's realistic large break LOCA analysis methodology", in *Nuclear Engineering and Design*, 235(16), pp. 1713-1725.

NEA (2017), "Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark: Final Report", NEA/CSNI/R(2016)18, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19752.

NEA (2016), "PREMIUM: A Benchmark on the Quantification of the Uncertainty of the Physical Models in System Thermal-hydraulic Codes: Methodologies and Data Review", NEA/CSNI/R(2016)9, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19720.

NEA (2015), "Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark Phase II: Identification of Influential Parameters", NEA/CSNI/R(2014)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19582.

NEA (2014a), "PREMIUM Phase III report: Quantification of Model Uncertainties on the Basis of FEBA/SEFLEX Reflood Experiments" (internal document).

NEA (2014b), "PREMIUM Phase IV report: Confirmation/validation of the uncertainties found within Phase III" (internal document).

NEA (2013), "PREMIUM – Benchmark on the quantification of the uncertainty of the physical models in the system thermal-hydraulic codes", in Proceedings of CSNI Workshop on Best Estimate Methods and Uncertainty Evaluations, Barcelona, Spain, NEA/CSNI/R(2013)8/PART2, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19377.

NEA (2011), "BEMUSE Phase VI Report, Status report on the area, classification of the methods, conclusions and recommendations", NEA/CSNI/R(2011)4, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19020.

NEA (2009), "BEMUSE Phase V Report, Uncertainty and Sensitivity Analysis of a LB-LOCA in Zion Nuclear Power Plant", NEA/CSNI/R(2009)13, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_18866.

NEA (2008), "BEMUSE Phase IV Report, Simulation of a LB-LOCA in Zion Nuclear Power Plant", NEA/CSNI/R(2008)6, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_18716.

NEA (2007), "BEMUSE Phase III Report, Uncertainty and Sensitivity Analysis of the LOFT L2-5 experiment", NEA/CSNI/R(2007)4, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_18442.

NEA (2006), "BEMUSE Phase II Report, Re-Analysis of the ISP-13 Exercise, post test analysis of the LOFT L2-5 experiment", NEA/CSNI/R(2006)2, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_18346.

NEA (2005), "Presentation a priori of the uncertainty evaluation methodology to be used by the participants" (internal document).

NEA (1998), "Report on the Uncertainty Methods Study", NEA/CSNI/R(97)35, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_16220.

Oberkampf, W.L., T.G. Trucano and C. Hirsch (2004), "Verification, Validation, and Predictive Capability in Computational Engineering and Physics", in *Applied Mechanics Reviews*, 57(5), pp. 345-384.

Oberkampf, W.L. and M.F. Barone (2006), "Measures of agreement between computation and experiment: validation metrics", in *Journal of Computational Phyics*, 217, pp. 5-36.

Oberkampf, W.L., T.G. Trucano and M.M. Pilch (2007), *Predictive Capability Maturity Model for Computational Modeling and Simulation*, SAND2007-5948.

Petruzzi, A., F. D'Auria, W. Giannotti and K. Ivanov (2005), "Methodology of Internal Assessment of Uncertainty and Extension to Neutron-Kinetics/Thermal-Hydraulics Coupled Codes", in *Nuclear Science and Engineering*, 149, pp. 1-26.

Pourgol-Mohamad, M., A. Mosleh and M. Modarres (2011), "Structured treatment of model uncertainty in complex thermal-hydraulics codes: Technical challenges, prospective and characterisation", in *Nuclear Engineering and Design*, 241, pp. 285-295.

Rider W. J., W.R. Witkowski and V.A. Mousseau (2015), *UQ's Role in Modeling and Simulation Planning, Credibility and Assessment Through the Predictive Capability Maturity Model*, SAND2015-20747J.

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

Sauvage, J.-Y. and M. Keldenich (2005), "ESM-3D: a fully extended statistical method based on CATHARE 3D for loss of coolant accident transients analysis", in Proceedings of the 13th International Conference on Nuclear Engineering (ICONE '05), Beijing, China.

Smith, R. (2014), *Uncertainty Quantification: Theory, Implementation, and Applications, Computational Science and Engineering*, SIAM.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.

Unal, C., B. Williams, F. Hemez, S.H. Atamturkur and P. McClure (2011), "Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors", in *Nuclear Engineering and Design*, 241, pp. 1813-1833.

USNRC (2005), "Transient and accident analysis methods", RG-1.203, USNRC.

USNRC (1988), *Best-Estimate Calculations of Emergency Core Cooling System Performance*, RG- 1.157, USNRC.

Wilks, S.S. (1942), "Statistical Prediction with Special reference to the Problem of Tolerance Limits", in *Annals of Mathematical Statistics*, 13(4), pp. 400-409.

Young, M.Y., S.M. Bajorek, M.E. Nissley and L.E. Hochreiter (1998), "Application of code scaling applicability and uncertainty methodology to the large break loss of coolant accident", in *Nuclear Engineering and Design*, 186(1-2), pp. 39-52.

2. Specification of input uncertainty quantification (Element 1)

2.1. Introduction

As part of the BEPU methodology, the quantification of model input uncertainty should start with a clear and concise specification of the problem. It consists in the identification and definition of the system responses quantities (SRQs), and important physical phenomena (or key model input parameters[7]) for the intended applications, which are the first steps of the BEPU methodology for transient and accident analyses (IAEA, 2009; Boyack et al., 1989; USNRC, 2005), such as those included in the final safety analysis report (FSAR) (USNRC, 1978).

In this process, the type of nuclear power plant and the transient or accident scenario of interest are specified first, the relevant SRQs are defined according to the acceptance criteria (ANSI/ANS, 1973), and the concerned reactor components and the involved physical phenomena or input parameters are then identified. Since many physical phenomena and input parameters may be involved in any thermal-hydraulic analysis, and they are not modelled in a simulation model (or computer code)[8] at the same fidelity, it may be impractical to quantify the uncertainty for each phenomenon and each input parameter. Therefore, it is essential to rank the importance of the involved physical phenomena or input parameters.

The selected or developed simulation model should be able to predict adequately the SRQs for the transients or accidents of interest, and the uncertainties on the important phenomena or input parameters should be quantified based on their importance to the modelling of the scenario and their impact on the SRQs for the simulation model calculation. The physical models included in the simulation model, and, their degree of fidelity in predicting physical phenomena must be consistent with the results of this process, namely, for a physical phenomenon that is ranked important to the scenario under consideration, the simulation model must have a relatively accurate model for that phenomenon and an IUQ of that model must be provided. Less accurate models may represent lower ranking phenomena with larger inherent uncertainty, for which an IUQ may not be necessary. The formality and complexity of this process should be coherent with the complexity and importance of the scenario under consideration.

The nuclear power plant accident scenario identification and IUQ definition process can rely heavily on expert opinion and can be subjective. Therefore, iteration of the process, based on experimentation and analysis, is important.

It is recommended to use the Phenomena Identification and Ranking Table (PIRT) technique (Wilson and Boyack, 1998; Diamond, 2006) or any other techniques, such as modified PIRT (Pourgol-Mohamad et. al. 2006), based on Analytical Hierarchical Process (AHP) (Saaty, 2001), or the most recently developed quantitative-PIRT (or Q-PIRT) based

7. The *key model input parameters* should be selected according to the recommendations in section 4.3. They are mentioned here only for the discussion related to Q-PIRT.

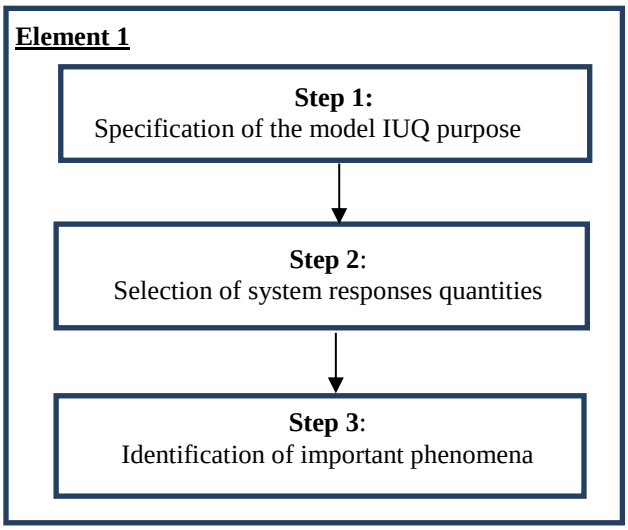
8. The term *simulation model* is used for *computer code* and associated *nodalisation and model options*.

on global sensitivity analysis (GSA) (Luo et al. 2010; Martin, 2011; Yurko and Jacopo, 2012) or data-driven methodology (Dinh et al. 2017).

This chapter provides recommendations and good practices for specifying the input uncertainty quantification and validation problem. They consist in three steps displayed in Figure 2.1 and aim to answer the following questions:

- Which problem should be considered? In addition, for which nuclear power plant and which scenario? (Step 1)
- What are the system response quantities of interest? (Step 2)
- Which physical phenomena are involved? What are the key physical phenomena (or model input parameters) affecting the model SRQs? (Step 3)

Figure 2.1. Steps of Element 1



Note that this element is part of the BEPU methodology, and is common to all VVUQ processes. Therefore, the available best practices remain applicable except that the focus is on the model input uncertainty quantification, rather than on the uncertainty analysis of the calculated SRQs of the simulation model.

2.2. Step 1: Specification of the IUQ purpose

The first step of SAPIUM is to specify the IUQ objective and identify the nuclear power plant type and transient scenarios for which the input uncertainty quantification and validation need to be performed.

Specification of the IUQ objective is important because any given nuclear power plant type or transient scenario may be analysed for different reasons. The specification influences the whole process of simulation model development, assessment and specific analysis for each specific transient scenario. For example, a larger break loss-of-coolant accident (LBLOCA) may be analysed to verify the emergency core cooling system (ECCS) performance by demonstrating compliance with the ECCS acceptance criteria (e.g. USNRC 10 CFR 50.46 [USNRC, 1988]), or to verify the integrity of the containment and of the reactor vessel or core component. The BEPU analysis has been widely developed, and applied for the ECCS performance verification in FSAR Chapter 15 accident analysis (Young et al. 1998; Ban et

al., 2014; Ban et al., 2004; Sauvage and Keldenich, 2005; Martin and O'Dell, 2005; Frepoli, 2008), which requires IUQs for all relevant and complicated multi-phase flow and heat transfer models during the LOCA blowdown, refill and reflood phases (Frepoli, 2008). A conservative approach has usually been applied for the containment integrity analysis, which only requires a bounding estimation of the uncertainties related to the major phenomena and models (break flow, stored energy release, overall heat transfer from the core and steam generator, and condensation in the containment, etc.). However, recent trends show that BEPU methods are also being developed and applied to the containment integrity analysis (Abdelghany and Martin, 2010) and other non-LOCA transient analyses (Kawamura and Hara, 2000; da Cruz et al., 2014; Brown et al., 2016; Walters et al., 2018).

The application domain or applicability of the IUQ depends on the transient scenario because the dominant safety parameters and acceptance criteria differ from one scenario to another. Therefore, the transient scenario determines also the key phenomena and input uncertainties that must be quantified and validated. However, a complete scenario definition is specific to the nuclear power plant type (e.g. PWRs, BWRs, or CANDU) or sometimes even to the plant itself, because the dominant physical phenomena and their interactions differ in various reactor designs or specific plant configurations. For example, the fuel design, core loading pattern, number and design of steam generators, number and design of reactor coolant loops, safety injection system design, and control systems can be significantly different from plant to plant and will significantly influence scenario behaviour.

It is thus recommended to start the specifications according to the transient classification as documented in the final safety analysis report (FSAR) of the targeted nuclear power plant (USNRC, 1978), or in case of a new plant design, to specify the application domain for the newly identified transient scenarios.

In order to reduce the IUQ efforts for a simulation model for a specific application, it is recommended to group different transient scenarios into a category of transients with common SRQs, and to make a generic input uncertainty quantification and validation for that class of transients. For example, one can group the transients and accidents in terms of the following phenomenology (or transient class) for light water reactors (USNRC, 1978; ANSI/ANS, 1973):

1. Cool-down: increase in heat removal from the primary system, such as steam system piping failures inside and outside of containment in a PWR (SLB).
2. Heat-up: decrease in heat removal by the secondary system, such as loss of external electric load (LOL), loss of normal feedwater flow (LOFW), or feedwater piping break (FLB).
3. RCS flow reduction: decrease in reactor coolant system (RCS) flow rate, such as reactor coolant pump shaft seizure (e.g. Locked Rotor or LR).
4. Reactivity insertion accident (RIA): reactivity and power distribution anomalies, such as rod ejection accidents in a PWR (REA) or rod drop accident in a BWR (RDA).
5. RCS volume increase: increase in reactor coolant inventory, such as inadvertent operation of ECCS during power operation (IOECCS);
6. RCS volume reduction: decrease in reactor coolant inventory, such as steam generator tube failure (SGTR), LBLOCAs or SBLOCA;

7. ATWSs: anticipated transients without scram, such as loss of A.C. power or turbine trip.

The following steps can be performed according to each transient scenario class for a specific nuclear power plant type.

2.3. Step 2: Selection of system response quantities

This step first involves selection of the SRQs according to the objective of the IUQ study. In most cases, this selection is straightforward, as the SRQs are related directly to the acceptance criteria as specified in the applicable rules, guides, design codes or standards for the transient scenario of interest.

In nuclear power plant final safety analysis reports (FSAR) (USNRC, 1978), the acceptance criteria for the events of interest are defined in terms of quantitative fuel and reactor system *design limits*[9], such as reactor coolant system pressure or temperature limits, departure from nucleate boiling ratio (DNBR) limits, cladding or fuel temperature limits (ANSI/ANS, 1973). For example, five specific criteria must be met for LOCA analysis (USNRC, 1988), namely, the peak cladding temperature (PCT), maximum local oxidation (LMO) or equivalent cladding reacted (ECR), core wide oxidation (CWO), coolable geometry and long-term core cooling (LTCC). For non-LOCA transients, however, the most common acceptance criteria are the maximum pressure, minimum DNBR or fraction of fuel rods experiencing departure from nucleate boiling (DNB). Thus, for FSAR accident analysis, SRQs are generally synonymous with criteria directly associated with the regulations, and their selection is usually a simple matter.

During simulation model development and assessment, a *surrogate*[10] SRQ may be of value in evaluating the importance of phenomena and processes. For example, in small-break LOCA (SBLOCA) or feedwater line break (FLB) analyses, the vessel coolant inventory (or core collapsed level) may be deemed more valuable in defining and assessing simulation model capability. In such a case, justification for using a surrogate SRQ should be provided.

In line with the surrogate SRQ, it is also important to consider other related performance measures in conjunction with the principle objectives. Because *compensating errors*[11] in the simulation model can unintentionally lead to correct answers, additional performance measures serve as physical tracking points and additional proof of accuracy. For example, in LBLOCA analysis, while the simulation model may calculate the correct peak cladding temperature (PCT), incorrect or physically impossible parameter values could evolve in other areas of the calculation, such as the quench front level or pressure losses. It is thus important to assess the accuracy of the related models.

For IUQ purpose, the chosen SRQs should be the parameters that are measured directly and accurately in the experiments, and will be used for verification of the design limits in the

9. The term *design limits* is used for all limits set by the designers on physical parameters, the respect of which ensures that the safety limit or acceptance criteria are met.

10. The term *surrogate* is used for replacement of the safety limit or acceptance criteria.

11. The term *compensating errors* means the errors in one model are off-set by the errors in other models by chance, which is difficult to find by statistical analysis.

nuclear power plant. Typical SRQs and their design limits or surrogates used in the nuclear power plant accident analysis are:

- Reactor coolant system (RCS), steam generator (SG) or containment (Cont.) pressure;
- Reactor coolant system or containment temperature;
- Reactor coolant system inventory or coolant level;
- Reactor core power/criticality/reactivity;
- DNBR or critical power ratio (CPR);
- SG secondary side mass inventory (in terms of mass or level);
- Containment pressure or temperature, or mass and energy release into the containment;
- Fuel temperature or enthalpy;
- Cladding temperature;
- Cladding oxidation or hydrogen pickup;
- Cladding stress, strain or deformation;
- Fuel rod internal pressure;
- Fuel or cladding failure fraction (melting fraction, burst fraction…).

In summary, the SRQs should be chosen for the typical transient scenario class and for a specific nuclear power plant type of interest, as illustrated in Table 2.1[12].

Table 2.1. Typical SRQs for PWR transients

SRQs/ Transients	RCS/SG/ Cont. Pressure	DNBR	Fraction of Rods in DNB	Fuel Centreline Temp.	Fuel Enthalpy Deposition	PCT	ECR	CWO
Cooldown	√ **(Cont.)**	√						
Heat-up	√	√	√					
RCS flow reduction		√						
Reactivity Insertion	√	√	√	√	√	**(√)**		
RCS Volume Increase	√							
RCS Volume Reduction	√ **(Cont.)**					√	√	√
ATWSs	√	√						

12. The table comes from an exercise performed by the authors based on review and summary of several publications.

2.4. Step 3: Identification and ranking of important phenomena (PIRT)

2.4.1. Definition of PIRT process

The behaviour of a specific plant and scenario is not equally influenced by all the processes, and phenomena that occur during a transient period. The most cost-effective but sufficient analysis reduces all potential phenomena to a manageable set by identifying and prioritising phenomena according to their influence on SRQs. Each phase of the transient scenario and the system components are investigated separately. The processes and phenomena associated with each component are examined. The cause and the effect are differentiated. Once the processes and phenomena are identified, they are ranked according to their impact on the relevant SRQs. The main product of the process described above is the so-called Phenomena Identification and Ranking Table (PIRT).

The development and assessment of a simulation model, including the IUQ activities, should be based on a credible PIRT. The PIRT should be used to determine requirements for physical model or methodology development, scalability, validation, and sensitivity studies. In the end, the PIRT is used to guide any uncertainty analysis or to assess the overall adequacy of the model. The PIRT is not an objective in itself; it is rather a tool to guide the next steps. Thus, the PIRT process is a systematic approach to prioritise the phenomena involved and identify the ones on which to concentrate efforts.

A typical PIRT development process consists in the following steps (Wilson and Boyack, 1998; Diamond, 2006):

1. Define objectives of the PIRT process (PIRT process is conditioned on the objectives: IUQ in this case).
2. Identify the plant design and the scenario type (as described in Step 1).
3. Define the SRQs of interest as physical phenomena have different impacts on different SRQs (as described in Step 2).
4. Define high-level basic system processes.
5. Partition transient scenario into convenient time phases and plant design into subsystems and components.
6. Identify plausible physical phenomena by phase and component.
7. Define screening criteria if only important phenomena should remain for the next steps of IUQ.
8. Develop a ranking for identified phenomena (for each component) and associated rationale for each phenomenon, either by expert judgement and discussions, or by using pair-wise AHP methodology.
9. Perform sensitivity analysis, if necessary and possible, to confirm the results from the previous steps.
10. Assess the level of knowledge (or uncertainty) regarding each phenomenon.
11. Document PIRT for subject scenarios and plant designs.

Each step will be described in the following paragraphs.

2.4.2. Identification of processes and phenomena

Once the plant design and the scenario type have been chosen and the SRQs have been identified (Steps 1-3), the processes and phenomena occurring in the area of interest are identified (Steps 4-6).

It is recommended to first identify high-level system processes and then divide the scenario into temporal and spatial categories. The processes and phenomena of each period and component are identified by examining the available experimental data, the design of the plant and its protection systems, operating experience and model simulations related to the scenario under consideration. The following procedure is recommended (USNRC, 2005):

- The scenario is divided into operationally characteristic time periods in which the dominant processes and phenomena remain essentially constant.
- For each time period, processes and phenomena are identified for each component, following a closed circuit throughout the system, to differentiate cause from effect.
- Starting with the first time period, the activities continue, component by component, until all potentially significant processes have been identified.
- The procedure is repeated sequentially, from time period to time period, until the end of the scenario.

This provides a matrix of time period (phase) and space (component or subsystem) for which all plausible phenomena and processes can be identified.

For example, LOCA transients are divided into phases in which dominant processes do not change significantly, and the reactor systems are split into components or subsystems, which are expected to spatially isolate some key phenomena. The large break LOCA can be divided in blowdown, refill, reflood and long-term core cooling phases. The small break LOCA can be divided into blowdown, natural circulation, loop seal clearance, boil-off and core recovery and long-term core cooling. For each time period (or phase), all possible processes and phenomena are identified for each component or subsystem covering fuel rods, core, upper plenum, hot leg, pressurizer (PRZ), steam generator or SG, reactor coolant pump (RCP), cold leg and ECCS including accumulator (ACC), safety injection (SI), down-comer, lower plenum, break, etc. An example of the processes and identified phenomena for a PWR LBLOCA reflood phase is given in Table 2.2[13].

Table 2.2. Processes and phenomena identified for each component for a PWR LBLOCA reflood phase

Component /Phenomenon	Fuel rods	Core	Upper Plenum	Hot Leg	PRZ	SG	RCP	Cold Leg/ ACC/ SI	Down-comer	Lower Plenum	Break
Stored energy	√										
Oxidation/Hydrogen Pickup	√										
Decay Heat	√										
Gap Conductance	√										
Nucleate Boiling		√							√		

13. The table comes from an exercise performed by the author based on review and summary of several publications.

Table 2.3. Processes and phenomena identified for each component for a PWR LBLOCA reflood phase (Continued)

DNB		√									
Post-CHF Heat Transfer		√									
Reflood Heat Transfer/ Quenching		√									
Rewet		√									
1 Phase Vapor Heat Transfer		√									
3D flow		√							√	√	
Void Gen./Distribution		√		√							
Droplet Entrainment/ De-entrainment		√	√	√					√		
Flow Reversal and Stagnation		√									
Phase Separation			√								
Countercurrent Flow and Limitation			√	√		√			√		
2 Phase Convention			√	√					√		
Flashing/Steam Expansion					√						√
Steam Binding						√					
Pressure Drop, Form Loss						√	√				
Condensation/Oscillation								√	√		
Non-condensable Gases								√			
ECC Water Mixing								√			
Hot Wall Effect									√	√	
Liquid Level Oscillation									√	√	
Critical Flow											√
Containment Pressure											√

Based on the NEA CSNI work on the matrix of separate effects and integral effects tests suitable for validating thermal-hydraulic codes (NEA, 1994a and b; NEA, 1996), 116 thermal-hydraulic phenomena have been identified for water cooled reactors, including new reactors (NEA, 2018), which cover virtually all LOCA and non-LOCA thermal-hydraulic transients.

Although a few non-LOCA transients (SLB or REA for example) exhibit many of the same challenging phenomena as observed in LOCA's, the primary coolant system remains principally single phase outside of the pressurizer for most PWR non-LOCA transients. Therefore, only less than one-third of the identified phenomena are applicable to non-LOCA transients. Since the non-LOCA transients are typically less challenging and less complex events than LOCA, and many events are also shorter in duration, it is reasonable to use only a single time interval for all components for most of the events. An example of the processes and phenomena identified for typical PWR non-LOCA transients is given in Table 2.3[14].

14. The table comes from an exercise performed by the authors based on review and summary of several publications.

Table 2.4. Processes and phenomena identified for typical PWR non-LOCA transients

Transients /Phenomena	Cooldown (e.g. MSLB)	Heat-up (e.g. FLB)	RCS flow Reduction (e.g. LR)	Reactivity Insertion (e.g. REA)	RCS Volume Increase (e.g. IOECCS)	RCS Volume Reduction (e.g. LOCAs, SGTR)	ATWSs
Fuel Rod Heat Transfer	√	√	√	√		√	√
Kinetics Feedback	√	√	√	√		√	√
Decay Heat		√				√	
PRZ Insurge/ Outsurge (RCS Heat/Mass Transfer)	√	√	√	√	√	√	√
SG Primary Side Heat Transfer	√	√				√	√
SG Secondary Side Heat Transfer	√	√					√
RCS Flow Coastdown (Pump, Pressure Drop, Form Loss)	√	√	√		√	√	√
Natural Circulation	√	√					
Core Mixing	√	√				√	
Boron Tracking	√					√	
Critical Flow	√	√				√	
Containment Pressure	√	√				√	

2.4.3. Ranking of importance

Once the identification is complete, the ranking process begins. Ranking by importance (Steps 7 to 9) is at the heart of the PIRT process and depends on the lists of phenomena identified in the last step. Each phenomenon, process, parameter, factor or characteristic is evaluated according to its relative importance for one or more SRQs.

In principle, the ranking process is not mandatory in the case where all the processes and phenomena that occur in a transient period of interest are processed. However, the ranking result helps to focus and allocate valuable resources for in-depth analysis, qualification of code models, evaluation of model sensitivities and uncertainties. With a well-defined ranking of important processes, model capabilities and calculated results, it is easier to prioritise modelling improvements. An important principle is the recognition that higher-order phenomena and processes require greater modelling fidelity.

As noted in Step 2, it may be possible to show that a surrogate SRQ other than the applicable acceptance criterion is more appropriate as a standard for identifying and ranking phenomena. This is acceptable if it can be demonstrated that, for all scenarios considered

for the specific classification and identification activity, the surrogate SRQ is compatible with plant safety.

It is imperative that all processes and phenomena be ranked according to their relevance to the scenario. It is important to note that importance determination should also apply to high-level system processes, which can be omitted if the focus is only on specific components. High-level system processes, such as depressurisation and inventory reduction, are very often closely linked to SRQs. Focusing on such processes can also help identify the importance of individual component behaviours.

In the importance ranking process, a scale of high (H), medium (M), low (L) and not applicable (N/A) is usually used due to its ease of application (see Table 2.4).

Table 2.5. Definition of importance scale

Importance scale	Definition
High (H)	The phenomenon/model/parameter has a controlling impact on the SRQs, which must be simulated by experiments and/or analytic modelling with a high degree of accuracy. IUQ is mandatory.
Medium (M)	The phenomenon/model/parameter has a moderate impact on the SRQ and only an appropriate moderate degree of accuracy is required for analytic modelling or measurements. IUQ is recommended.
Low (L)	The phenomenon/model/parameter has a minimal or zero impact on the SRQs. The phenomena need to be model led in the simulation model or explained in adequate detail in the methodology, but accuracy in modelling the process is not considered very influential to the analysis of the whole transient. No IUQ is needed.
Not Applicable (N/A)	The phenomenon is considered not to occur at all.

The different phenomena that are identified are ranked according to the influence they can or could have on the SRQs. Different techniques are available to perform the evaluation of importance of the processes and phenomena:

- Expert opinion: each phenomenon is assigned an importance rank by a panel of experts with a discussion about the rationales (Wilson and Boyack, 1998; Diamond, 2006):
- Decision-making methods, such as the Analytical Hierarchical Process (AHP) (Pourgol-Mohamad, et. al. 2006; Saaty,2001): expert judgements can be assisted by such analysis in order to reduce the subjectivity;
- Fuzzy set theory (Kljenak et al. 2001);
- Sensitivity analysis with One-at-a-Time (OAT) single parameter variation method (Kovtonyuk et al., 2017);
- Quantified PIRT (Q-PIRT) (Luo et al. 2010; Martin, 2011; Yurko and Jacopo, 2012) based on global sensitivity analysis (GSA) methods.

The original PIRT for LBLOCA in the CSAU methodology (Boyack et al., 1989) was developed by a group of nine experts with many years' experience in designing LWRs, or test facilities, performing tests or analysing test results, performing simulation model calculations and analysing calculated results from national laboratories, academia, industry and the NRC. When considering the importance of each phenomenon to each SRQ in a specific scenario, panellists voted from nine (the highest) to one (the lowest), and the phenomena importance ranking was then consolidated by averaging the votes. Final scores of seven to nine were considered high importance, four to six medium, and one to three low.

To avoid the difficulty of dispersed votes, a global index, importance ratio (IR), can be defined to provide a weighted assessment of the importance vote results:

$IR = 100 \times (H + M/2)/(H+M+L)$,

where H, M and L stand for the number of high, medium and low votes, and a value of 1 is assigned to a "high" vote, a value of 0.5 to "medium" vote and a value of zero to a "low" vote. As an example, an $IR \geq 75$ % can be considered as high importance, $75\% > IR > 25\%$ medium, and $IR \leq 25\%$ low.

The AHP method was adopted as part of the PIRT process in CSAU methodology by an independent group of thermal-hydraulic analysts to serve as a supplement and confirmatory counterpart to the expert judgements (Boyack et al., 1989). AHP is a systematic, logical approach that was developed to reduce complex issues into manageable pieces (Saaty, 2001). An attractive feature of AHP is that the relative importance of any given process is determined only through pairwise comparison with other processes taken one at a time. The AHP then includes matrix arithmetic to assimilate all individual decisions in a global conclusion. Thus, AHP combines features of the bottom-up deductive approach, which focuses on the phenomena for each component, with the top-down systematic approach that concentrates on the important phenomena for whole system.

Another attempt has been made on the use of the fuzzy set theory (Kljenak et al. 2001). In this approach, expert opinions about the importance of different basic phenomena, expressed in the form of grades, were aggregated by first mapping grades into fuzzy sets and then multiplying point-wise corresponding membership functions. Phenomena were ranked by comparing the centre-of-gravity values of the resulting functions. Although the use of fuzzy set theory did not result in significantly different ranking of basic phenomena, the method is still more appropriate than arithmetic or weighted averaging of grades, as it allows the experts' impreciseness and uncertainty to be implicitly included in the evaluations.

The initial phases of the PIRT process described above can rely heavily on expert opinion, which can be subjective. Therefore, it is important to validate the PIRT using experimentation and analysis. Although still limited, other less subjective initial importance determination methods based on sensitivity studies have been developed (Luo et al. 2010; Martin, 2011; Yurko and Jacopo, 2012), which can help determine the relative influence of phenomena identified early in the PIRT development and for final validation of the PIRT as the simulation model development and assessment process is iterated.

The OAT based on simulation model calculations is a coarse sensitivity analysis process where each model input parameter for a phenomenon is varied once (for example from a minimal value to a maximal value), and its effect on the model output value is analysed (Iooss and Lemaître, 2015). Some threshold values can be defined to determine the relative influence of the model input parameters, e.g. for LBLOCA reflood heat transfer phenomena, $\Delta PCT \geq 50$ °K for high influence, $10\ °K < \Delta PCT < 50\ °K$ for medium influence, and $\Delta PCT < 10$ °K for low influence.

The Q-PIRT approach (Luo et al. 2010; Martin, 2011; Yurko and Jacopo, 2012) consists of two steps: a *"top-down"* step focusing on identifying the important physical phenomena controlling the SRQ, and a *"bottom-up"* step which focuses on determining the correlations from those key physical phenomena that significantly contribute to the SRQ uncertainty. The top-down step evaluates phenomena using the governing equations of the system code at nominal parameter values, providing a "fast" screening step. The bottom-up step then

analyses the correlations and models for the phenomena identified from the top-down step to find which parameters to sample.

The GSA based on statistical analysis of the simulation model calculations may provide an additional "quantitative" insight into the impact of the identified phenomena and related model input parameters on the SRQs (Luo et al. 2010; Martin, 2011; Yurko and Jacopo, 2012). A GSA technique based on random sampling can be adopted when the perturbations in the potentially important model input parameters are propagated through a simulation model (McKay, 2005; Helton et al., 2006). Statistical treatment of responses provides some quantitative sensitivity measures for the relation between the SRQs and the analysed input parameters (Luo et al. 2010; Martin, 2011; Yurko and Jacopo, 2012).

Correlation coefficients (e.g. Pearson simple correlation coefficient or Spearman rank correlation coefficients) are the most important measure of the degree of correlation between two variables, as they standardise the covariance (a measure of the amount of association between two variables) by eliminating the dependency on scale of measurement for a particular data set. A correlation coefficient does not provide any conclusions regarding cause and effect; rather, it indicates the degree (or strength) of statistical relationship between one or more phenomena or model input parameters and the SRQs. For example, the Pearson simple correlation coefficient (SCC) provides a measure of the linear relation between the SRQs and one of the model input parameters, while the Spearman rank correlation coefficient (RCC) provides a measure of the monotonic relation between the SRQs and one of the model input parameters (McKay, 2005; Helton et al., 2006). It is recommended to use the partial rank correlation coefficients (PRCC), which remove trends associated with other input parameters, and hence improve the resolution of the sensitivity measures.

In general, the sensitivity measures (ρ) can be interpreted as follows (McKay, 2005; Helton et al., 2006):

- $|\rho| \in [0.75, 1]$: strong linear or monotonic relationship;
- $|\rho| \in [0.25, 0.75]$: moderate linear or monotonic relationship;
- $|\rho| \in [0, 0.25]$: weak linear or monotonic relationship;
- $|\rho| \approx 0$: no consistent pattern that allows for prediction of one variable's values based upon knowledge of the other variable's values.

Therefore, it is fairly intuitive to define the importance of a phenomenon or model input parameter by comparison with the strength of a correlation coefficient, for example, as shown in Table 2.5.

Table 2.6. Definition of importance of phenomena by the absolute values of the correlation coefficients

Correlations coefficients $\|\rho\|$	Phenomenon Importance Level
$0.75 \leq \|\rho\| \leq 1$	H
$0.25 \leq \|\rho\| < 0.75$	M
$0.01 < \|\rho\| < 0.25$	L
$\|\rho\| < 0.01$	N/A

It should be noted that the threshold for the level of influence (or importance) depends on the number of samples and other parameters, and should be accompanied by some statistical significant testing (e.g. by ensuring that the significance level p value < 0.1). However, the threshold values given in Table 2.5 can be used as practical "rule of thumb" guidance.

As the phenomena and simulation models are very complex and there are certain interactions between model input parameters, these coefficients can only be considered as qualitative and as a relative index for screening the non-important model input parameters. Other sensitivity measures, such as the Sobol's indexes, could be obtained by using the variance-based decomposition method (Iooss and Lemaître, 2015). The Sobol's indexes provide quantitatively the contribution of the uncertainty of each input parameter to the target output parameter uncertainty. However, the variance-based decomposition method requires much more calculations effort.

A detailed description and recommendation of the sensitivity studies can be found in Section 4.3.

Finally, it is worth mentioning the innovative data-driven model development and assessment process that effectively utilises data from physical and numerical experiments (Dinh et al., 2017). It is expected that new artificial intelligence (AI) techniques can be used for analysing the experimental data and to identify and rank the phenomena.

2.4.4. Ranking of knowledge level or uncertainty

Another important part of the PIRT process is scoring the knowledge level (or uncertainty) for each phenomenon to reflect how well the phenomenon/model/parameter is measured (for test data) or calculated (for simulation model) in determining the SRQ. The rationale for the scoring is an important product of the expert elicitation. When a phenomenon is identified as being important but the corresponding knowledge level is low, it is an indication that more effort must be applied, e.g. more research support is needed. A qualitative scale of high (H), medium (M) and low (L) is also adopted due to its ease of application, as defined in Table 2.6.

Table 2.7. Definition of knowledge level scale

Knowledge Level Scale	Definition
High (H) or Known (K)	The phenomenon/model/parameter is fully, or almost fully known (more than 75% of what one could expect to know). Additional research on this phenomenon is not necessary even if the importance level is high.
Medium (M) or Partial Known (P)	The knowledge base is moderate (25-75% of the knowledge base is established). Research is suggested if the phenomenon is of high importance.
Low (L) or Unknown (U)	The knowledge base is low (less than 25% of the knowledge base is established). This phenomenon is a priority for additional research, particularly if the importance level is high, but also if the importance is only medium.

To avoid the difficulty of dispersed votes, a global index, knowledge ratio (KR), can be defined to provide a weighted assessment of the knowledge vote results:

$$KR=100 \times (H + M/2)/(H+M+L)$$

where H, M and L stand for the number of known, partially known and unknown votes respectively, and a value of 1 to a "high" (known) vote, a value of 0.5 to "medium" (partially known) vote and a value of zero to a "low" (unknown) vote. A $KR \geq 75$ % can be considered as high knowledge (known), $75\% > KR > 25\%$ medium (partially known), and $KR \leq 25\%$ low (unknown).

It is important to consider all sources of information regarding the expected influential phenomena: plant design and its protection systems, experiments and code calculations. For example, experiments provide only well-measurable information to the analyst while the simulation model calculation may provide insight on the phenomena that are difficult to measure but important nevertheless. The importance of a phenomenon itself may be different between an experiment and simulation model. Furthermore, experimental rigs do not model all the plant features, which may be important for certain scenarios.

This is especially the case for non-LOCA transients in actual nuclear power plants, for which the moment of reactor trip on protection signal is often determinant. Thus, the phenomena related to correct simulation of the protection signals become important: e.g. in case of asymmetric transients, the loop-to-loop mixing ration determines the accuracy of calculation of over-temperature or over-power delta-temperature protection signals, or the pressure losses and recirculation ratio in the steam generator influence the correct simulation of narrow range level measurement and related protection set-points. Such information seldom can be derived from the experimental tests, as the mock-up rigs do not reproduce such measurements and signals.

In the IUQ process, attention should be paid to the phenomena whose importance is ranked high or medium while knowledge level is ranked medium or low. Either new experiments or model improvements have to be made to adequately quantify the input uncertainties. Otherwise, appropriate provisions must be considered to cover the insufficient knowledge.

2.4.5. Documentation of PIRT

The last, but not least important, step is the sufficient documentation of the PIRT development process. It should be comprehensive, up to date, transparent and traceable in order to follow quality assurance protocols and to enable efficient and exhaustive peer review (Wilson and Boyack, 1998; Diamond, 2006).

It has to be noted that the above procedure establishes the general approach to develop a PIRT. However, the actual process will vary as the purpose of PIRT varies, and some steps may be modified or omitted. Indeed, PIRT has been applied in thermal-hydraulics, nuclear physics, fuels and severe accidents analyses, and new reactor design, and each application has a different objective and modified the basic PIRT approach in some way to fit its own needs (Shaw et al. 1988; Boyack et al., 2001a and b; MHI, 2007; Greene, 2001; Song et al., 2004).

Top ranked phenomena identified by different expert groups for each phase of a large break LOCA in a PWR are summarised in Table 2.7.

Table 2.8. Important phenomena identified for a PWR LBLOCA

Phase	Component and Phenomenon	Original CSAU	Westinghouse	Framatome
Blowdown	Fuel rod stored energy & response (ballooning, burst)	√	√	√
	Core power distribution		√	√
	Core heat transfer (DNB, post-CHF, rewet)	√	√	√
	Loop pressure loss (in particular broken)		√	
	PRZ (quenching, flashing, CCFL)	√		
	RCP degradation (2-phase performance)	√	√	
	Break flow (critical flow, break path resistance)	√	√	√
Refill	Fuel rod stored energy & response (swell & burst)	√	√	√
	Core power distribution		√	√
	Core heat transfer (post-CHF, rewet)	√	√	√
	Loop pressure loss (in particular broken)		√	
	Cold leg/Acc condensation	√	√	√
	Accumulator resistance		√	
	Downcomer Droplet Entrainment/De-entrainment, condensation, 3-D effect	√		
	Lower plenum level swell	√		
	Break flow (critical flow, break path resistance)	√	√	√
Reflood	Fuel rod response (Oxidation/Hydrogen Pickup, gap conductance)	√	√	√
	Decay Heat & Core power distribution	√	√	√
	Core heat transfer (Reflood Heat Transfer/ Quenching)	√	√	√
	Spacer effects			√
	Core/Upper plenum 3D flow (Void Gen./Distribution, Droplet Entrainment/ De-entrainment)	√	√	√
	Droplet Entrainment/ De-entrainment & SG steam binding	√	√	√
	RCP Pressure Drop, Form Loss	√		
	Cold leg/Acc Non-condensable Gases/Nitrogen effects	√	√	√
	Downcomer Hot Wall Effect	√		
	Lower plenum Liquid Level Oscillation	√		

The most common important phenomena identified for all LBLOCA phases are:

- Break flow (critical flow, break path resistance).
- Initial stored energy/fuel rod response (ballooning, burst).
- Core power distribution.
- Core heat transfer (DNB, post-CHF, rewet, reflood heat transfer, quenching).

- Core/upper plenum 3D flow (void gen./distribution, droplet entrainment/de-entrainment).
- Delivery and bypass of ECCS water.
- Steam binding/entrainment.
- Condensation in cold leg and down-comer.
- Non-condensable gases/accumulator nitrogen effects.

An example of PIRT for non-LOCA heat-up transient is shown in Table 2.8.

Table 2.9. An example of PIRT for non-LOCA heat-up transient (loss of normal feedwater)

Component	Phenomena	Importance	Knowledge Level
Core	Fuel Rod Heat Transfer	L	H
	Kinetics Feedback	L	M
	Decay Heat	H	H
Pressurizer	PRZ Insurge/Outsurge (RCS Heat/Mass Transfer)	H	H
Coolant Pumps	RCS Flow Coastdown (Pump, Pressure Drop, Form Loss)	L	H
	Natural Circulation	L	L
Steam Generator	Primary side Heat Transfer	H	H
	Secondary Heat Transfer	M	M
	Separator	L	L

2.5. Recommendations and open issues

2.5.1. Recommendations

This element is part of the BEPU methodology, and is common to all VVUQ processes. Therefore, the available best practices remain applicable, except that the focus is on the model input uncertainty quantification (i.e. IUQ), rather than on the uncertainty analysis of the calculated SRQs of the simulation model.

It is recommended to clearly specify the model input uncertainty quantification (IUQ) problem for the developed or selected simulation model, according to the transient classification as documented in the final safety analysis report (FSAR) of the targeted nuclear power plant, or in case of a new plant design, to specify the application domain of the IUQ for the newly identified transient scenarios.

It is recommended to choose the SRQs based on physical parameters that are directly and accurately measured in the experiments, and will be used for verification of the design limits or acceptance criteria (or surrogates) in the nuclear power plant accident analysis.

It is recommended to group different accident scenarios (e.g. RCS heat-up or cool-down accidents, reactivity-initiated accidents) into a single IUQ problem with common SRQs of interest, and make a “generic” input uncertainty quantification and validation for the developed or selected simulation model.

It is recommended to use the Phenomena Identification and Ranking Table (PIRT) technique to identify and rank the important physical phenomena, and use the sampling-based global sensitivity analysis (GSA) to confirm the PIRT.

2.5.2. Open issues

The only open issue is the subjectivity of the highly ranked phenomena, which could be subject to the limitations of the knowledge of the expert, the experimental databases or simulation models. An iteration with other elements may be necessary if such limitations are identified during the SAPIUM process.

References

Abdelghany, J.M. and R.P. Martin (2010), "Uncertainty analysis for containment response of U.S. EPR TM reactor to large break loss-of-coolant accidents," in *Proce. of the 2010 International Congress on Advances in Nuclear Power Plants - ICAPP '10*, ANS.

Aksan, N., F. D'Auria and H. Glaeser (2018), "Thermal-hydraulic phenomena for water cooled nuclear reactors", in *Nuclear Engineering and Design*, 330, pp. 166-186.

ANSI/ANS (1973), *Nuclear Safety Criteria for the Design of Stationary Pressurized Water Reactor Plants,* ANSI/ANS N18.2-1973, ANSI.

Ban, C.H., S.-Y. Lee and C.-K. Sung (2004), "Development and Application of KEPRI Realistic Evaluation Methodology (KREM) for LB-LOCA", in *Best Estimate-2004: Proceedings of International Meeting on Updates in Best Estimate Methods in Nuclear Installations Safety Analysis,* Washington, DC, United States.

Boyack, B., R. Duffey, G. Wilson, P. Griffith, G. Lellouche, S. Levy, U. Rohatgi, W. Wulff, W. and N. Zuber (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, United States.

Boyack, B.E., A.T. Motta and K.L. Peddicord (2001a), *Phenomenon Identification and Ranking Tables (PIRTs) for Rod Ejection Accidents in Pressurized Water Reactors Containing High Burnup Fuel,* NUREG/CR-6742, USNRC, United States.

Boyack, B.E., A.T. Motta, K.L. Peddicord, C.A. Alexander and J.G.M. Andersen (2001b), *Phenomenon identification and ranking tables (PIRTs) for loss-of-coolant accidents in pressurized and boiling water reactors containing high burnup fuel,* NUREG/CR-6744, USNRC, United States.

Brown, S., H. Zhand, V. Kucukboyaci and Y. Sung (2016), "Best Estimate Plus Uncertainty Analysis of Departure from Nucleate Boiling Limiting Case with CASL Core Simulator VERA-CS in Response to PWR Main Steam Line Break Event," in *Nuclear Engineering and Design*, 309, 8-22.

da Cruz, D.F., D. Rochman and A.J. Koning (2014), "Propagation of Nuclear Data Uncertainty for A Control Rod Ejection Accident Using the Total Monte-Carlo Method", in *Proc. PHYSOR 2014 - The Role of Reactor Physics toward a Sustainable Future,* Kyoto, Japan.

Diamond, D.J. (2006), "Experience Using Phenomena Identification and Ranking Technique (PIRT) for Nuclear Analysis," in *Proc. PHYSOR-2006 Topical Meeting*, Vancouver, Canada (BNL-76750-2006-CP).

Dinh, N., L. Lin and P. Athe = (2017), "Validation of RISMC Models", in *Multiphysics Model Validation Workshop,* Raleigh, NC, United States.

Frepoli, C. (2008), "An Overview of Westinghouse Realistic Large Break LOCA Evaluation Model", in *Science and Technology of Nuclear Installations*, 2008, Article ID 498737.

Greene, K.R., C.D. Fletcher, R.C. Gottula, T.R. Lindquist and B.D. Stitt (2001), "Utilising Elements of the CSAU Phenomena Identification and Ranking Table (PIRT)

to Qualify a PWR Non-LOCA Transients System Code", in *Proc. International Conference on Nuclear Engineering,* Nice Acropolis, France.

Helton, J.C., J.D. Johnson, C.J. Sallaberry and C.B. Storlie (2006), "Survey of sampling-based methods for uncertainty and sensitivity analysis," in *Reliability Engineering and System Safety*, 91(10-11), pp. 1175-1209.

IAEA (2019), *Deterministic Safety Analysis for Nuclear Power Plants.* IAEA Specific Safety Guide No. SSG-2 Rev. 1. IAEA, Vienna.

Iooss, B. and P. Lemaître (2015), "A review on global sensitivity analysis methods," in *Uncertainty management in Simulation-Optimisation of Complex Systems: Algorithms and Applications (C. Meloni and G. Dellino, Eds),* Springer.

Kawamura, S. and T. Hara (2000), "Best Estimate Methods for Licensing Analysis", in *Proc. Int. Mtg. BE-2000,* Washington, DC, United States.

Kljenak, I., R. Jordan Cizelj and A. Prosek (2001), "Development of a Phenomena Identification and Ranking Table Using Fuzzy Set Theory", in *Proc. Ninth International Conference on Nuclear Engineering ICONE 9,* Nice, France.

Kovtonyuk, A., S. Lutsanych, F. Moretti and F. D'Auria (2017), "Development and assessment of a method for evaluating uncertainty of input parameters", in *Nuclear Engineering and Design*, 321, pp. 219-229.

Luo, H., Q. Wu and V. Mousseau (2010), "Quantified PIRT for Reactor System Safety Analysis Code Validation Using Characteristic Time Ratios", in *Proc. NUTHOS-8,* Shanghai, China.

Martin, R.P. (2011), "Quantifying Phenomenological Importance In Best-Estimate Plus Uncertainty Analyses," in *Nuclear Technology*, 175, pp. 652-662.

Martin, R.P. (2016), "Best-Estimate-Plus-Uncertainty-Informed Deterministic Safety Evaluation," in *Transactions of the American Nuclear Society*, 115, pp. 1136-1139.

Martin, R.P. and L.D. O'Dell (2005) "AREVA's realistic large break LOCA analysis methodology," in *Nuclear Engineering and Design*, 235(16), pp. 1713-1725.

McKay, M.D. (1998), "Sensitivity and uncertainty analysis using a statistical sample of input values," Ch. 4, in *Uncertainty Analysis, (Ronen, Y. Editor),* CRC Press, Florida, United States.

MHI (2007), "*Small Break LOCA Methodology for US-APWR",* MUAP-07013-NP(R0), Non-Proprietary Version, Mitsubishi Heavy Industries, Ltd.

NEA (1994a), "Separate effects test matrix for thermal-hydraulic code validation - Volume I: Phenomena characterisation and selection of facilities and tests", NEA/CSNI/R(1993)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_15968.

NEA (1994b), "Separate effects test matrix for thermal-hydraulic code validation - Volume II: Facility and experiment characteristics", NEA/CSNI/R(1993)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_15968.

NEA (1996), "CSNI Integral Test Facility Validation Matrix for the Assessment of Thermal-Hydraulic Codes for LWR LOCA and Transients", NEA/CSNI/R(1996)17, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_16134.

Pourgol-Mohamad, M., M. Modarres and A. Mosleh (2006), "Modified Phenomena Identification and Ranking Table (PIRT) for Uncertainty Analysis", In *Proc. 14th Int. Conf. Nuclear Engineering (ICONE 14),* Miami, FL., United States.

Saaty, T.L. (2001), *Models, methods, concepts & applications of the analytic hierarchy process*, Kluwer Academic Publishers, Boston.

Sauvage, J.-Y. and M. Keldenich (2005), "ESM-3D: a fully extended statistical method based on CATHARE 3D for loss of coolant accident transients analysis," in *Proceedings of the 13th International Conference on Nuclear Engineering (ICONE '05),* Beijing, China.

Shaw, R.A., T.K. Larson and R.K. Dimenna (1988), *Development of a Phenomena Identification and Ranking Table (PIRT) for Thermal-Hydraulic Phenomena During a PWR LBLOCA,* NUREG/CR-5074, USNRC.

Song, J.H., B.D. Chung, J.J. Jeong, W.P. Baek, S.Y. Lee, C.J. Choi, C.S. Lee, S.J. Lee, K.S. Um, H.G. Kim and Y.S. Bang (2004), "Phenomena Identification and Ranking Table for the APR-1400 Main Steam Line Break", in *Nuclear Engineering and Technology*, 36(5), pp. 88-402.

USNRC (1978), *Standard Format and Content of Safety Analysis Reports for Nuclear Plants,* Regulatory Guide RG-1.70, Rev. 2, USNRC.

USNRC (1988), Section 50.46, *"Acceptance criteria for emergency core cooling systems for light water nuclear power reactors*" and Appendix K to Part 50, "ECCS evaluation models", 10 CFR Part 50, USNRC.

USNRC (1989), *Best estimate calculations of emergency core cooling system performance,* Regulatory Guide RG-1.157, USNRC.

USNRC (2005), *Transient and accident analysis methods,* Regulatory Guide RG-1.203, USNRC.

Walters, W., R. Bobolea, K. Maupin, R. Deveney, J. Parker, K. Segard and R. Barner (2018), "ARTEMIS / RELAP5 Integrated Transient Analysis Application To Non-LOCA Transients", in *Proc. TOPFUEL 2018,* Prague, Czech Republic.

Wilson, G.E. and B.E. Boyack (1998), "The role of the PIRT process in experiments, code development and code applications associated with reactor safety analysis," in *Nuclear Engineering and Design* Volume 186(1–2), pp. 23-37.

Young, M.Y., S.M. Bajorek, M.E. Nissley and L.E. Hochreiter (1998), "Application of code scaling applicability and uncertainty methodology to the large break loss of coolant", in *Nuclear Engineering and Design,* 186(1-2), pp. 39-52.

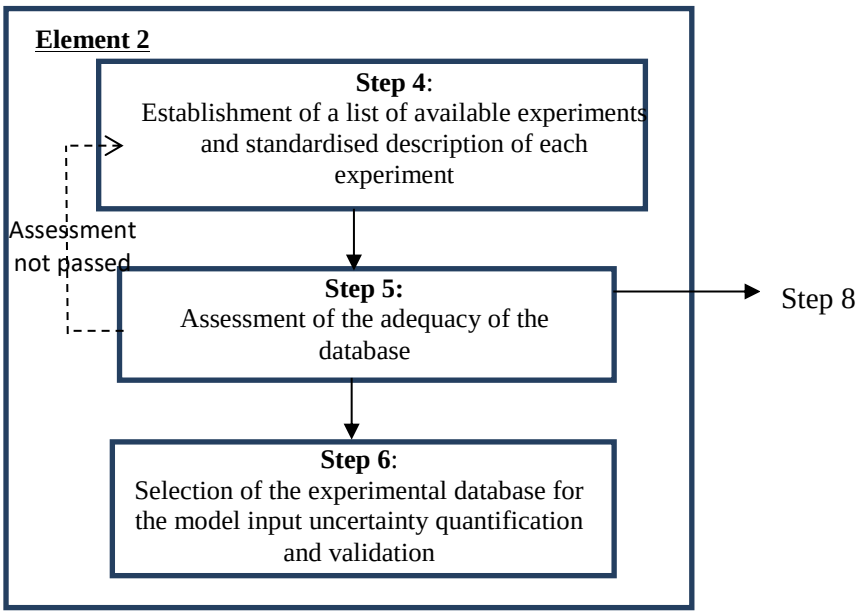
Yurko, J.P. and B. Jacopo (2012), "Quantitative Phenomena Identification and Ranking Table (QPIRT) for Bayesian Uncertainty Quantification." In *Proc. 2012 International Congress on Advances in National Power Plants (ICAPP '12),* Chicago, United States.

3. Development and assessment of the experimental database (Element 2)

3.1. Introduction

The objective of Element 2 is to construct a representative experimental database for the problem specified in Element 1. The key steps that should be taken into account according to the SAPIUM approach are outlined in Figure 3.1.

Figure 3.1. Steps of Element 2



These steps include the following topics to address for a transparent and reliable IUQ, as described in the sections of this chapter:

1. Level of detail in the construction of an experimental database

A fine description of the problem requires a certain amount of experiments to cover each different situation that can be encountered (different model/phenomenon, different geometry, different flow regime…). This is not always affordable in practice due to the lack of a sufficient number of experiments with the required detailed and accurate measurements available. Therefore, a compromise has to be found.

2. Type of experimental database

The database may be specific to a singular investigated transient; it can be general (generic), i.e. covering all code models and addressing all possible transients. It can also be in between, i.e. related to a class of similar transient and addressing models, for which uncertainties have to be considered.

3. Description of an experimental database

For an efficient selection procedure, a complete detailed description of both experiments and associated tests should be provided. It requires the definition of a set of criteria leading to a common framework for the description.

4. Analysis of an experimental database to drive the construction

An evidence-based analysis is performed to evaluate if a given experimental database is representative for the intended use. It relies on two important properties (adequacy of each experiment and completeness of the database) that should be quantified. This analysis can be exploited to establish a ranking between experiments and select the most relevant ones.

5. Consideration of the scaling problem

It is mostly a problem of consideration of scale-up effects. The constitutive models of the code are often developed based on small-scale experiments. Uncertainty analyses are performed frequently for large-scale facilities (in particular nuclear reactors). Some limited data exist at reactor scale coming from reactor data or from UPTF tests, which help at the validation of the scalability of some models. The lack of large-scale experimental data requires special approaches that can compensate for this deficiency. The related approaches and methods are described in Chapter 7, dedicated to scaling issues.

6. Selection of experiments for quantification and validation database

The full IUQ procedure involves both quantification and a validation step that should exploit two different subsets of experiments extracted from the experimental database. However, in practice, it might happen that the number of experiments is too limited and the selection strategy should be adapted.

3.2. Step 4: Establishment of a list of the available experiments and standardised description of each experiment including evaluation of experimental uncertainties

3.2.1. Level of detail in the construction of an experimental database

An experimental database should be constructed to cover all the models and correlations describing identified important physical phenomena that are going to be considered and quantified in the uncertainty analysis. It is preferable to build a detailed database for phenomenon-specific quantification, where for each phenomenon and its specific features the available experiments are assembled. For example, the evaporation models should have different features for low void fraction (bubble flow) and for high void fraction (droplet flow). This should be considered for the construction of the experimental database in order to perform a reliable uncertainty quantification.

To develop an experimental database, several aspects have to be taken into account. In particular, how detailed should uncertainty quantification be? Should the uncertainty be quantified for each correlation? Should the uncertainty of more than one important model input parameter of one correlation even be considered (if available in input)? On the other hand, should the uncertainty of the model be determined as a whole? In particular cases, the uncertainty of several closure laws may also be determined as a whole. In this part, answers to these questions are suggested.

The experimental database should be divided into sub-groups according to each model/phenomenon. If required and possible, it should be divided according to geometry, flow regime and heat transfer regime, and, if necessary, according to range of physical

parameters. For instance, for each wall heat transfer mode and for each transition region, a separate set of experiments is preferred. Also, a further subdivision by thermal-hydraulic conditions is of interest, in particular when different correlations are available in the code, e.g. film boiling for forced and natural convection (pool boiling).

A serious problem in building the experimental database is the lack of adequate experiments. For some phenomena, e.g. interfacial friction between phases, only a very limited number of suitable tests for uncertainty quantification exists. The problem arises when there are not enough experiments available or when measurements are not accurate enough. It means measurement error is comparable to the estimated modelling error. In such a situation, a very detailed subdivision of experimental database is not reasonable. It would be better to build a less detailed database while ensuring sufficient information, meaning the database would contain a sufficient number of test data for an accurate quantification of uncertainties.

The optimal level of detail both depends on the availability of an experimental database and on the accuracy of physical modelling. It can be adjusted during the input uncertainty quantification procedure. The technical possibilities of model parameters variation in the code are usually known at the beginning of the analysis and can be considered when selecting and assembling experimental data. It means it is not reasonable to select experiments with data related to the models that cannot be varied in the code and consequently cannot be considered by an uncertainty analysis.

3.2.2. Developing a database designed for a given scenario: suggested approach

The aim of this section is to present an approach to building an experimental database that depends on the target scenario, and to suggest recommendations to help in this procedure.

The classification of approaches in the framework of model uncertainty quantification is a problem of model uncertainties quantification as a whole. It considers not only the availability of experimental data but also the problems with implementation of phenomenon and component related uncertainties, selection of the most effective solution, and the effort required to apply the chosen approach.

3.2.2.1. *Input uncertainties quantification (and experimental database) for special case vs. general approach* (and database)

This is a classification at the level of the application of the model uncertainties quantification approach. The general approach (and the database required) can be understood as quantification of model uncertainties for the application field of the code. The quantification of model uncertainties for the code application field is the domain of code developers and is conducted in the verification and validation (V&V) and uncertainty quantification process. The model uncertainties quantification done by users for particular applications is always a special case. Even if code developers perform an uncertainty analysis, it is the applied approach, and developed experimental database performed by the users. However, code developers may have at their disposal the general database, which would be rather case specific. It can be expected in such a situation that a case specific database would be selected and applied. The case specific analyses offer better accuracy in the uncertainty analysis.

3.2.2.2. *The problem of specific vs. generic database (and approach)*

The generic approach does not mean a quantification of model uncertainties for the whole application field of the code. This is another level of classification compared with case

specific vs. general. The common understanding is that a specific approach is related to the phenomenon and component. The generic approach is related to the code/model (possibly also geometry). Therefore, generic approach uncertainties quantification and the validation of the model uncertainties would also be done for the application (particular) case and not generally.

A good example of a component specific and a generic approach is the blowdown and reflooding phase of LOCA. In the specific approach, uncertainties for blowdown and reflooding phase would be evaluated separately and different input uncertainties, for instance interfacial friction and evaporation rates, would be determined and applied for both phases of the transient. In the generic approach, each code correlation is evaluated and its uncertainty range is applied for the whole transient according to the flow pattern and geometry, if different correlations for different geometries are implemented in the code.

A further step in the specific approach could be, for instance, the differentiation between cold leg and hot leg or U-tubes and pipe geometry. In a specific approach, model uncertainties for cold and hot legs and for U-tubes would be determined. In a generic approach, uncertainty of the correlation for pipe geometry would be quantified, if such a correlation is used in the code for all three components.

A related issue in the specific and generic approach is scaling. The specific approach should deal with reactor scale, but since there are not enough related experiments, frequently the experiments in smaller scale must be taken into account. The generic approach is applied usually by V&V and uncertainty quantification of the codes is part of development and validation. In this case, when the approach is regarded as generic, sometimes the correlations are quantified separately according to the scale (application), even if the correlations are applied for all scales. It is due to large differences of accuracy by small and large-scale applications.

With the specific approach, the possibility of variation of input uncertainties has to be kept in mind. Changing the model output during simulation according to the part of the transient is not an easy task and could lead to serious problems. Similarly, the introduction of different uncertainties for different components may be impossible in the standard input and would require modification in the source codes.

The experience with ATHLET is that for the introduction of component (but also phenomenon) related uncertainties, modification of the code is always necessary. For instance, the introduction of separate uncertainties of phase relative velocities for particular geometries in the uncertainty analysis of OMEGA rod bundle test (Glaeser et al., 1994) required code modification, with the aim of identifying the distinguished objects in the subroutines related to modelling of phase relative velocity. There may be similar situations with other codes. They would require source modifications for this purpose. This requires additional effort and knowhow regarding code models and programming. These could exceed the ability of many users performing uncertainty analyses. Therefore, a purely specific, but also a purely generic, approach is not really practicable. This is the case of component and phenomena-specific approaches, mainly due to implementation in the code source, and of the generic approach because of the need to consider scaling effects.

The best solution seems to be a balanced mix of specific and generic approaches, according to the application and capabilities.

3.2.3. Description of available experiments from the point of view of the objective of the designed database

3.2.3.1. Summary and type of the available experiments

The experiments in nuclear thermal hydraulics can be split into several categories: basic tests, separate effect tests (SETs), integral effect tests (IETs) and combined effect tests (CETs). These categories have been defined in several reports or publications. This report reuses some of the terms given in (NEA, 2016) for the basic tests, SETs and IETs.

Basic tests aim at understanding the phenomena and do not necessarily refer to the geometry or the actual ranges of operating parameters in power plants (NEA, 2016). These tests may have analytical solutions or may use correlations or data derived from experiments (Glaeser, 2017).

A separate effects test facility (SETF) is designed to investigate:

- The reactor component behaviour (SETF-Component test) by characterising the component responses that are typical of the design function.
- The local phenomena (SETF-Basics test) in order to validate models. Separate effects tests deal usually with one phenomenon. Tests where several combined phenomena are investigated are usually called combined effect tests.

SETs make it possible to observe phenomena in selected zones in a nuclear power plant's system, or in specific plant components and some specific process in a particular period of a given transient. The main role of SETs is to provide experimental data to develop and validate the physical models and/or empirical correlations under prototypical or simulated conditions. Recently, heavily instrumented SETFs were built to produce spatially and temporally fine resolution data for validating the computational fluid dynamics (CFD) codes (called CFD-grade experiments) (NEA, 2016).

The integral test facility (ITF) is a scaled-down test facility designed to investigate:

- the overall system behaviours and the related phenomena and processes;
- the interaction of two or more components;
- the local phenomena: those are typical of the overall system design target function.

Performed in such a facility, an IET provides a similar thermal-hydraulic dynamic response to a postulated accident, and/or abnormal transient in a reference reactor. The data obtained from scaled ITF experiments are not directly considered applicable to full-scale conditions due to scale distortions. They are used mostly for understanding accident phenomena and validating the system codes (NEA, 2016). Transient data from real nuclear power plants should also be considered due to the absence of scale-down or limitations of ITFs.

Intermediate tests are also identified. These tests can be component or combined effect tests (CETs) where phenomena occur in an identified part (component) of the circuit (e.g. hot leg, core) (Mascari et al., 2016). A combined effect test usually represents several components of a reactor but not the whole system, like the IETs. It may have a lower capability to give information on models than SETs and a lower capability to simulate all system effects but it may address coupled (combined) phenomena in a prototypical geometry.

In 1996, the CSNI established a list of 177 PWR and BWR specific integral test facilities for code validation (NEA, 1996a). A list of SETs was compiled in 1994 and includes more

than 1 000 tests from 185 experimental facilities (NEA, 1994a; NEA, 1994b). For VVER nuclear power plants, a validation matrix has also been built (NEA, 2001). These lists have been established as databases for model validation, but are not initially intended for uncertainty analysis or the quantification of input uncertainties.

Few published collections of basic tests exist. In the context of the ECUME project (Minouni and Serre, 2001) CEA and EDF compiled a list of 27 tests, aimed at testing the different potentialities required for the numerical methods for the system-scale and component-scale codes. The tests are split into two categories defined as: 1) test with known analytical solution, and 2) test with unknown analytical solution with experimental reference. For the second category, some of these tests are clearly SETs (Super Canon, Super Moby Dick…).

3.2.3.2. Description of an experiment

In order to help select an experiment, a proposal for the description of a test is given in Table 3.1.

Such a test description exists in other documents. See for example "Separate effects test matrix for thermal-hydraulic code validation – Volume I: Phenomena characterisation and selection of facilities and tests" (NEA, 1994a). It gives information on the type of test: covered phenomena, experimental conditions, instrumentation, available measurements, and so on, which can be used as criteria for the selection. Annex B gives examples of test descriptions.

Table 3.1. Description of a test

Criteria/items	Test XXX
Type (SET, IET, CET…) Component and/or reactor if interest	
Working fluid (steam water, simulant fluid…), Material properties	
Range of main parameters Pressure, mass flux, quality or void fraction, heat flux…	
Geometry	
Scale (vertical scale, volume scale wrt component or nuclear power plant)	
Covered phenomena	
Covered model	
Validate complete system, subsystem, component	
Available measurements	
Instrumentation (tool used for data measurement)	
Data access condition	
Available documentation/reports	
Publications	

Hereafter the items are described:

- *Working fluid*: the fluid used in the experiment can be a criterion in the selection of the test. For example, air/water experiments are limited. These experiments are usually performed at low pressure, and may not transposable for high-pressure conditions. Experiments with simulant fluid (e.g. Freon) need a transposition for similar steam/water conditions. This transposition is a source of uncertainties.
- *Material properties*: this item is relative to material properties of the heated elements (thermal properties (density, specific heat, thermal conductivity), which can have for example an impact on the rewetting in case of LOCA, or the surface quality of the solids [roughness]).
- *Geometry*: the geometrical representativeness of an experiment is one of the main criteria in selecting an experiment.
- *Scale*: see Chapter 7.
- *Covered phenomena, range of main parameters*: these criteria make it possible to judge if the test is able to represent one or several phenomena. SETs are designed to covered dedicated phenomena, and conditions and data available in IETs need a judgement if they covered a phenomenon.
- *Covered model*: this criterion makes it possible to judge the relevance of the experiment with respect to the code. It can be split into several criteria, addressing the thermal hydraulics, thermal, radiochemical or mechanical models implemented in the code.
- *Available measurements, instrumentation*: these criteria are key for the selection of the test. Sources of experimental uncertainties are described in Section 3. In relation to instrumentation and measurement ability, the SET CNSI report (NEA, 1994a) distinguishes two kinds of phenomena:
  - Macroscopic phenomena like depressurisation rate, which can be characterised by the available instrumentation (pressure, temperature transducers);
  - Microscopic phenomena like droplet coalescence. In most case, an average signal is provided, without precise indication about the local value of the reference quantity.

It is important to point out that the available measurements or the instrumentation techniques are often not able to give directly the measurement of the phenomenon or the quantity of interest. For example, in the case of the interfacial mass transfer relative to the flashing, the interfacial flux is deduced from other measurements only (namely the pressure profile) according simplification assumptions.

- *Available documentation/reports*: this criterion is mentioned in the selection of tests in the CSNI Code Validation Matrix.
- *Publications*: this criterion gives some information on the validation of other codes against the experiment. It can also give some feedback about the user effect (how the tests are simulated).

Ideas for the standardisation of information of the tests can be found in the CSNI reports or other literature. For example, Petruzzi and D'Auria (2016) developed the idea of consolidating qualified databases (both experimental and code calculation results) through

standardisation, aiming to support the V&V activities of system codes and uncertainty methodologies. The database, called SCCRED (Standardised consolidated calculated and reference experimental database), includes documentation such as the reference data set of the facility (description of the facility, instrumentation, physical properties of the material, evaluation of the pressure losses, etc.) and of the associated tests (main phenomena investigated during the test, configuration of the facility, etc.). Further details on SCCRED are given in Section 4.1.4.

3.2.3.3. Evaluation of experimental uncertainties

Among the different sources of uncertainties, the experimental uncertainties have to be identified. They can be divided in different sources relative to each measurement. Some of the main sources of such uncertainties are listed below:

- Inherent limitations of the measurement technique.
- Sensitivity of the sensor.
- Possible electronic noise.
- Known biases (systematic error, estimation of the accuracy).
- Lack of calibration.
- Intrusive method (e.g. probe) or non-intrusive method (e.g. X-ray).
- Space and time resolution of the measurement (e.g. time sampling).
- Space or time reconstruction methods. Associated to the space and time resolution of the measurement, the method of space or time reconstruction may be the cause of uncertainties. For example, in a rod-bundle geometry with non-transparent rods, the PIV (particle image velocimetry) technique makes it possible to measure the velocity in the central region of the sub-channel, but not close to the rods where the laser beam cannot penetrate. To reconstruct the average velocity in the sub-channel, the assumption on the velocity profile within the sub-channel and the associated uncertainty in this profile is the cause of uncertainties in this quantity.

The best source of experimental uncertainty evaluation is estimation based on information in the experimental report provided by experimenters.

The repeatability of the tests is important for the estimation of experimental uncertainties and the qualification of measured data.

It is useful to prove consistency in the measured data. This can be done by comparing the measured data using different techniques or comparing local measurements with global balance analyses, e.g. mass or heat balance.

In the worst case, when no estimation of uncertainty is available, rough estimations of every type of sensor can be applied. However, in such a case, it is strongly recommended to use different parallel experiments and compare with other experimental data. For example, this was done in the case of FEBA, one of the experiments selected in the PREMIUM benchmark. The description of the FEBA experiment seemed to provide very little information about uncertainties of the experimental data. In particular, the accuracy of the pressure drop measurement, not reported during the FEBA runs, was estimated from similar bundle configurations (NEA, 2017).

The issue of large experimental uncertainties, i.e. when measurement error is comparable to the estimated modelling error, is discussed in Section 3.2.1.

3.3. Step 5: Assessment of the adequacy of the database

3.3.1. Adequacy of a database

The adequacy of an experimental database includes two main properties that should be checked in the analysis. The first is related to the ability of an experiment to provide relevant information for model input uncertainty quantification and validation. It is called *representativeness*. The second concerns the ability of a set of experiments to fulfil the physical conditions of the problem under study, e.g. cover the physical space of interest. It is referred to as *completeness*.

The CSNI Code Validation Matrix (CCVM) of SET and IET gives part of the methodologies for the assessment of the representativeness of an experiment and of the completeness of an experimental database. The methodology to establish the SET matrix relevant for LOCA application and transients in LWRs is summarised in "Separate effects test matrix for thermal-hydraulic code validation – Volume I: Phenomena characterisation and selection of facilities and tests" (NEA, 1994a) as:

1. Identification of phenomena relevant for these transients.
2. Characterisation of phenomena, in terms of a short description of each phenomenon.
3. Setting up a catalogue of information sheets on the experimental facilities, as a basis for the selection of the facilities and specific tests.
4. Forming a separate effects test facility cross-reference matrix through the classification of the facilities in terms of the phenomena they address.
5. Identification of the relevant experimental parameter ranges in relation to each facility that addresses a phenomenon and selection of relevant facilities related to each phenomenon.
6. Establishing a matrix of experiments (the SET matrix) suitable for the developmental assessment of thermal-hydraulic transient system computer codes, by selecting individual tests from the selected facilities, relevant to each phenomenon.

The CSNI IET matrix was constituted considering several factors, including (NEA, 1996a):

1. Companion matrices, which relate phenomena of interest, test facility and test type.
2. Typicality of facility and experiment to expected reactor conditions.
3. Quality and completeness of experimental data (measurement and documentation).
4. Relevance to safety issue.
5. Tests selected must clearly exhibit phenomena.
6. Each phenomenon should be addressed by tests of different scaling (at least one test if possible).
7. High priority to ISPs (International Standard Problems), counterpart and similar tests.
8. Challenge to system code.

System thermal-hydraulics codes have developed SET and IET matrices dedicated to target scenarios in the validation process. For example, for LBLOCA, such a methodology was

adopted during the development and assessment of the CATHARE code (Barré and Bernard, 1990; Geffraye et al., 2011). The validation matrix includes integral tests performed in large test facilities such as BETHSY. Phenomena occurring during the reflooding phase (see Table 2.2, Chapter 2) are assessed using different tests, including BETHSY 6.7c. For this specific phase of the transient, several experiments, covering a wide range of operation parameters, are used to validate models in a separate way. This SET matrix includes FEBA and PERICLES experiments, dedicated to assess the reflood heat transfer and quenching phenomena in the core, with and without fuel ballooning. In a continuous way, the test matrix is enhanced and re-evaluated for a better account of the phenomena. This is the case for steam binding, oscillatory and top-down reflooding, which are expected to be improved by the use of a three-field model, allowing improvements in the simulation of the role of the droplets in the core (Valette, et al., 2011).

Based on the previous practice, the aim is to identify objective criteria to characterise the adequacy of a set of experiments from their detailed description and with respect to the scenario. Both individual criteria and the full set of criteria should possess some properties to represent the multi-criteria nature of the problem (Malczewski and Rinner, 2015). It especially includes:

- The measurability of each criterion.
- The operationality of each criterion, i.e. meaningfulness for the analysis to perform.
- The capability of the set of criteria to cover all aspects of the problem.
- The non-redundancy of the set of criteria to avoid the problem of double counting.

The set of criteria should be also minimal (i.e. kept small as much as possible) to reduce the complexity of the analysis.

3.3.2. Representativeness

3.2.3.1. Criteria for representativeness

In the framework of IUQ, priority for representativeness characterisation should be given to:

- The separable nature of the experiment.
- The quality of the experimental data such as density of measurements, quality of boundary and initial conditions characterisation, quality of measured flow parameters, availability and amount of measurement uncertainty (see Paragraph 3.2.3.3.), capability to cover important physical phenomena/models of interest, sensitivity of measured parameters to phenomena and models of interest, repeatability.
- The agreement between experimental and case study conditions with a focus on the geometry and the boundary and initial conditions.
- The capability of the experiment to address different simulation scales (CFD, component-scale, system scale).

The major part of these criteria is covered by the methodology used in the selection and evaluation of the CCVM (NEA, 1996a; NEA, 1994a; NEA, 1994b; NEA, 1996b).

In practice, not all the previous priorities are always reachable. It is the case when there are only combined effect tests in the experimental database and no separated effect tests

available (like for heat transfer enhancement at the quench front). This problem was investigated in the PREMIUM project and the answer comes from the development and application methodologies suitable for quantification of model uncertainties based on CETs. An example of such quantification of physical model uncertainties based on FEBA and PERICLES experiments, considered as CETs, can be found in "Input uncertainties in uncertainty analyses of system codes: Quantification of physical model uncertainties on the basis of CET (combined effect tests)" (Skorek, 2017). Moreover, it can happen that experimental database does not contain enough varied data. In such a case, the use of integral effect tests together with separated effect tests could be discussed.

3.3.2.2 Evaluation of the representativeness of an experiment

In the framework of multi-criteria problem, the evaluation of representativeness can be handled by exploiting multi-criteria decision analysis or making (MCDM)[15] approaches.

Given an experimental database, it can be recalled how MCDM can be used to evaluate the representativeness of each experiment and provide a ranking between them for a specific problem/objective defined in Element 1 of the SAPIUM methodology.

The reliability of MCDM results first relies on the introduction of a set of criteria to characterise the representativeness. It is assumed that this set (denoted $\{C_i\}_{i=1,\ldots,p}$ in the sequel) has been already constructed following the recommendations of Sections 3.3.1 and 3.3.2.1.

MCDM methods require handling the two following steps.

(1) Construction of the decision matrix

A decision matrix is a table summarising the analyst's preferences. It first includes a characterisation of the representativeness of each experiment according to each criterion. For each experiment e, a vector $(g_1(e), g_2(e), \ldots, g_p(e))$ is defined so that $g_i(e)$ is the valuation (i.e. a score characterising the representativeness) of the experiment relative to the criterion i. Moreover, an importance coefficient w_i is associated to each criterion C_i, and represents the relative weight of the criterion i with respect to the others.

Therefore, a decision matrix can be formally represented as follows:

Criteria / Experiments	C_1	C_2	C_3
e_1	$g_1(e_1)$	$g_2(e_1)$	$g_3(e_1)$
e_2	$g_1(e_2)$	$g_2(e_2)$	$g_3(e_2)$
Importance Coefficient	w_1	w_2	w_3

In the framework of experiment ranking where criteria are usually qualitative, there is no need to use standardisation methods (Malczewski and Rinner, 2015) to transform the criteria to comparable units as it is classically performed in the generic application of

15. In this document, the acronym MCDM (multi-criteria decision making) has been used to avoid confusion with MCDA (Model Calibration through Data Assimilation) method recalled in Chapter 5.

MCDM tools. There exist different approaches to define unitless scales and assess the experiment valuation as well as the importance weights.

One can mention the ranking method, which consists in ranking the experiments with respect to each criterion (resp. the criteria) in the order of the analyst's preferences, i.e. the most representative/important = 1, the second more representative/important = 2. Then, the valuation and the importance weight can be computed as:

$$g_i(e_k) = \frac{n - r_k^{e,i} + 1}{\sum_{i=1}^{n}(n - r_i^{e,i} + 1)}$$

$$w_k = \frac{p - r_k^{w} + 1}{\sum_{i=1}^{p}(p - r_i^{w} + 1)}$$

Where n (resp. p) denotes the number of experiments (resp. criteria) and $r_k^{e,i}$(resp. r_k^{w}) stands for the rank position of the experiment e_k for the i^{th} criterion (resp. the rank position of the k^{th} criterion).

A second strategy can exploit a rating method that requires the analyst to estimate valuations and weights based on a predetermined scale (e.g. from one to 100).

Both ranking and rating methods are empirical approaches. An alternative method relying on the theory of consistent matrices has been proposed by Saaty (Saaty, 1982; Mu and Pereyra-Rojas, 2017) in the context of the Analytical Hierarchical Process (AHP). It is based on pairwise comparisons involving a rating scale from one to nine. It leads to the construction of a matrix exhibiting a particular structure (called reciprocal) whose eigenvector associated to the largest eigenvalue can be used to derive the valuations and the importance weights. A consistency checking has, however, to be performed to assess the reliability of the results.

(2) Construction of the decision rule

A decision rule is a procedure for ordering the set of experiments. It requires combining (or aggregating) the information provided by the decision matrix to evaluate the representativeness and establish the final ranking.

This rule can rely on the computation of a unique synthesis index as in Table 3.2 below.

Table 3.2. Example of decision matrix

Criteria / Experiments	C_1	C_2	C_3
e_1	0.875	0.167	0.1
e_2	0.125	0.833	0.9
Importance Coefficient	0.669	0.088	0.243

If Table 3.2 gives the decision matrix, this decision rule leads to the following final score for each experiment:

- Score for e_1: 0.875*0.669+0.167*0.088+0.1*0.243 = 0.624
- Score for e_2: 0.125*0.669+0.833*0.088+0.9*0.243 = 0.376

And the first experiment appears to be the most representative according to this method. A unique synthesis index defined as a weighted sum can be sensitive to the numerical values of the decision matrix, e.g. if $g_1(e_1)$=0.6 and $g_1(e_2)$=0.4 in the previous table (e_1 is still more adequate than e_2 for the first criterion) then the new scores are:

- Score for e_1: 0.6*0.669+0.167*0.088+0.1*0.243 = 0.44
- Score for e_2: 0.4*0.669+0.833*0.088+0.9*0.243 = 0.56

And the second experiment becomes the most representative even if the number of criteria for which e_1 is most representative than e_2 is the same as in the previous example.

Outranking methods (such as ELECTRE [Roy, 1996]) can be an interesting alternative. An experiment *e* is said to outrank an experiment *e'* if, taking into account all information given by the valuation, the analyst can conclude that *e* is at least as good as *e'*. The construction of the outranking relation is based on the computation of two indices called concordance and discordance. The first one measures the strength of support in the information given, for the hypothesis "*e* is at least as good as *e'* ". More precisely, for each pair of experiments (e_i,e_k), this index is computed as the sum of the relative weights associated to the criteria where the hypothesis "e_i is at least as good as e_k" is satisfied. Coming back to the example of Table 3.2,

$$c(e_1, e_2) = \frac{w_1}{\sum_{i=1}^{3} w_i} = 0.669$$

On the contrary, the discordance index measures the strength of opposition to the hypothesis "*e* is at least as good as *e*' " and to which extent the outranking hypothesis can still be accepted. For each pair of experiments (e_i,e_k) and each criterion C_j such that $g_j(e_i)<g_j(e_k)$, this index is defined by:

$$d_j(e_i, e_k) = g_j(e_k) - g_j(e_i)$$

Which is written in the case of Table 3.2:

$$d_2(e_1, e_2) = g_2(e_2) - g_2(e_1) = 0.666$$
$$d_3(e_1, e_2) = g_3(e_2) - g_3(e_1) = 0.8$$

The values of each index are then compared with several thresholds to conclude if there is enough evidence to support that e_i outranks e_k. The thresholds are chosen empirically. For the concordance index, they are usually larger than 0.5. For the discordance one and for a given criterion C_j, they depend on the difference between the maximal and the minimal values of the valuation function g_j when taking into account all experiments. For each index, the different thresholds make it possible to accept the outranking hypothesis with a strong or weak certainty. Even though some recommendations exist, the robustness of the results of the choice of the empirical thresholds associated to concordance and discordance indices is important to ensure the reliability of the final ranking.

An illustration of the use of this type of MCDM method is provided in "On the methodological treatment of input uncertainty quantification: illustration in the RIA framework" (Baccou et al., 2018) in the framework of RIA (reactivity insertion accident) applications.

Remark 1: In the case of a unique synthesis index, the representativeness of a set of experiments can be evaluated. For example, if three experiments {e_1, e_2, e_3} are available with respective representativeness scores {0.15,0.8,0.05}, then the representativeness of the set {e_1,e_3} is 0.15+0.05=0.2.

3.3.3. Completeness of a database

The idea is to develop quantitative criteria that could help estimate in a practical way the completeness of the experimental database. Currently, such a procedure is not practised but it could mean:

- Building an adequate matrix of sensitivities of measured parameters with respect to models of interest.
- Using methods for checking whether the database contains enough information.

Criteria for assessing the maturity level of various issues within the V&V UQ procedure for nuclear licensing (Unal et al., 2011) have already been proposed under the terminology predictive maturity indexes (PMI). In "Defining predictive maturity for validated numerical simulations" (Hemez et al., 2010), where the PMI is applied, the authors point out that such indexes should include three features, the goodness of fit, complexity and coverage. They then studied the mathematical properties that a PMI should satisfy. What emerges is that a high level of maturity could contribute to prove the completeness of a database. However, since this SAPIUM element is restricted to the experimental database and not to the whole quantification process, the existing works on PMI need to be adapted.

It is therefore advisable to first integrate in the construction of completeness criteria the ratio between the convex hull areas of the validation and application domains. A ratio close to one will reduce extrapolation when moving to the application domain. The spatial distribution of the experiments within the validation domain can also be interesting to avoid overweighting the influence of specific regions in the input uncertainty quantification. Several mathematical tools exist to check the uniformity of a spreading (e.g. Pronzato and Muller, 2012 or Damblin et al., 2013).

3.3.4. Adequacy index

If the representativeness and the completeness (denoted respectively α and κ) have been evaluated, an adequacy index can be defined as:

$$\rho=F(\alpha,\kappa)$$

Where F is an aggregation operator chosen to satisfy desired properties. In practice, it is often a function whose main expected property is to ensure the increase of ρ when a new experiment contributes to a strong improvement of the adequacy. A database could then be considered as adequate as possible (i.e. taking into account the available experiments) once extra physical tests would not enhance the index.

Starting from an initial experimental database with few experiments and from a large one[16], which is the initial set of candidates for the new experiment, the adequacy index can be used to enrich the initial design by iterating the three following steps:

- Find the experiment e of the set of candidates that maximises ρ;
- Update the database by adding e to the current database;
- Remove e from the previous set of candidates; the resulting set defines the new set of candidates.

16. Coming for example from the test matrices used for thermal-hydraulic codes validation and assessment.

This type of approach is called a sequential design of experiment (DoE) strategy. The iteration is stopped once the new experiment does not increase the adequacy index.

3.4. Step 6: Selection of the experimental database for the quantification and the validation, according to the adequacy assessment

This selection is based on the PIRT (Chapter 2) and on the adequacy analysis described in the previous section. It is also important to recall that the database should integrate various scales of IETs, which is mandatory for the application to the reactor case.

The definition of model uncertainties is often characterised by the application of separate effect tests for quantification of model uncertainties and validation of quantified uncertainties based on integral experiments. The combined effect tests were used in the past rather for validation than for quantification of model uncertainties. This was mainly due to the fact that quantification of model uncertainties on the basis of combined effect tests, when several phenomena occur simultaneously and several model uncertainties have to be quantified in the same procedure, is a difficult task. However, with increasing accuracy of uncertainty analysis and improving methodologies, CETs are used more and more also for uncertainties quantification (Skorek, 2017).

The current praxis is to use CETs also for quantification of model uncertainties. CETs are applied for model uncertainties quantification mainly for complex phenomena for which separate effect tests (SETs) do not exist, e.g. critical discharge or heat transfer enhancement at the quench front. This development was reflected by investigation of methodologies for quantification of model uncertainties based on combined effect tests in the frame of the PREMIUM project.

The standard approach is still to use SETs for quantification of model uncertainties. The quantification methodology is simple and effective. However, there are no fundamental reasons not to use SETs also for the validation step, assuming that a sufficient number of experiments is available. Validation on separate effects tests can be even more precise, since influence of other uncertainties and resulting compensation errors does not take place. On the other side, validation based on integral experiments is more effective, with all or at least the majority of models validated simultaneously. In this way, interaction of the models is proved.

Therefore, the integral experiments are preferred for the validation. However, it is also highlighted that in the original CSAU, the integral tests were used also to quantify the so-called “separate biases” due to limitation in experimental data bases or code. Those biases are expected to be included in the model uncertainties. However, the biases were applied to compensate code limitations (uncertainties) according to phenomena (and models) not simulated in the code. Such code limitations usually cannot be expressed by input uncertainty of the physical code model. Section 6.5 (Step 17 of the SAPIUM methodology) discusses the analysis of the validation results. Perhaps the most important reason for applying SETs for the quantification step is the shortage of a sufficient number of experiments. For successful quantification of model uncertainty, at least several experiments are necessary. Ideally a completely investigated application area should be covered by the experimental data but in practice it is generally not the case due to a lack of experimental data (this issue is discussed, for example, in “Scaling issues for the experimental characterisation of reactor coolant system in integral test facilities and role of system code as extrapolation tool” [Mascari et al, 2016]). For each part of the application field (range of thermal-hydraulic parameters and geometry), at least two adequate

experiments (ideally more) should be applied. Quantification based on a single experiment is not recommended. The chosen single experiment might be very specific and not wholly relevant to the investigated case, or experimental measurements might be disturbed by a systematic error that was not realised by experimenters, and as such not reported. Therefore, it is recommended to apply more than one experiment for the same range of parameters.

In the face of a limited number of suitable experiments, it can be assumed that all SETs would be applied for the quantification. For example, for a particular phenomenon such as the critical discharge, where many tests were performed, the tests can be applied for quantification as well as for validation. In fact, critical discharge experiments are combined effect tests, though they are classified frequently as SETs. This is mainly the case when classification considers only integral and separate effects tests.

Regarding the selection of experiments for the quantification and validation, it can happen in practice that the number of available experiments is too limited to split between quantification and validation. In this case, all available experiments should be considered for the quantification step and the validation step should be adapted (a cross-validation approach can be followed; more details are available in Chapter 6). Moreover, experiments from chemical processing or the oil industry can be of interest. The interesting experiments from these fields are solely separate and combine effect tests. It is advisable to select for the validation step the experiments performed in the context of nuclear reactor thermal hydraulics. The available adequate integral tests, however, are practically only from the field of nuclear reactor science.

3.5. Recommendations and open issues

3.5.1. Recommendations

The construction of the experimental database is a key element of the input uncertainty quantification process since, together with the simulation model, it provides the mathematical methods of IUQ (described in Chapter 5) and the information to back-propagate in order to derive input uncertainties. It strongly contributes to the reliability of the quantified input uncertainties. Therefore, it is recommended to follow a structured and transparent approach by developing and assessing an experimental database.

This first requires collecting all available experiments from SETs, IETs and CETs. Various scales of IETs are mandatory for the application to the reactor case. Extra experiments might also be required if the adequacy of the database is not sufficient. Special attention should be devoted to the question of dependency of the experimental database with respect to the reactor transient. The best solution seems to be a balanced mix of specific and generic approaches to derive the database, according to actual application and capabilities.

Each experiment should also establish a standardised description. The description elements include the covered phenomena, the geometry, the scaling effect of the experimental facility and the available measurements with the associated uncertainties.

An important step is the assessment of the adequacy of a database. Two properties of the database contribute to its adequacy: the representativeness of each experiment and the completeness of the database for the intended use. For a transparent and reproducible assessment, it is recommended to use mathematical tools to quantitatively perform the analysis. A multi-criteria decision-making approach can be used to objectively and automatically evaluate the representativeness leading to a ranking of experiments. The application of this type of approach relies on the introduction of a set of criteria to

characterise the representativeness. The evaluation of the completeness of the experimental database is based on a completeness index. This index should be restricted to the database and not be applied to the whole quantification/validation process as it is classically done in VVUQ.

Finally, special attention should be devoted to the splitting of the experimental database in two parts, one for input uncertainty quantification and one for input uncertainty validation. If the number of available experiments is too limited to perform this splitting, all available experiments should be considered for the quantification step.

3.5.2. Open issues

The main issue is related to the quantitative evaluation of the adequacy. To achieve that evaluation, both representativeness and completeness should be ensured. If representativeness can exploit a multi-criteria decision-making approach, the determination of the completeness index still requires further work to integrate and quantify relevant properties of the database to check (such as the spatial distribution of the experiments, the ratio between areas of the validation and application domains). Moreover, the combination with the representativeness index to evaluate the overall adequacy involves the use of aggregation operators (such as min, max, product) and the definition of a relevant strategy to perform this aggregation remains an open issue.

The SAPIUM methodology proposes to improve the quantification by expert judgement. It is important to keep in mind that problematic situations might exist where the expert judgement could be used:

- Situations where available measured parameters are not sufficient to determine separate uncertainties. This is usually the case with integral experiments. This problem has to be treated by a quantification and validation step; e.g. by quantification of model uncertainties, experts could be consulted and results proved by validation.
- Treatment of indirect links between measures and input parameters. In such situations, the problem needs to be solved by a quantification procedure. A frequently applied solution is the iteration of calculations and comparison with experimental data. If this is not possible, a solution could again be expert consultation and to prove resulting uncertainties with a validation step.

References

Barré, F. and M. Bernard (1990), "The CATHARE code strategy and assessment", in *Nuclear Engineering and Design*, 124, pp. 257-284.

Baccou, J., E. Chojnacki, V. Georgenthum and S. Marais (2018), "On the methodological treatment of input uncertainty quantification: illustration in the RIA framework", Proceedings of BEPU-2018, Lucca, Italy.

Damblin, G., M. Couplet and B. Iooss (2013), "Numerical studies of space-filling designs: optimization of Latin Hypercube Samples and subprojection properties", in *Journal of Simulation*, 7(4), pp. 276-289.

Geffraye, G., O. Antoni, M. Farvacque, D. Kadri, G. Lavialle, B. Rameau and A. Ruby (2011), "CATHARE 2 V2.5 2: A single version for various applications", in *Nuclear Engineering and Design*, 241, pp. 4456-4463.

Glaeser, H., E. Hofer, M. Kloos and T. Skorek (1994), "Uncertainty and Sensitivity Analysis of a Post-Experiment Calculation in Thermals Hydraulics", in *Reliability Engineering and System Safety*, 45, pp. 19-33.

Glaeser, H. (2017), "Verification and validation of system thermal-hydraulic computer codes, scaling and uncertainty evaluation of calculated code results" In *Thermal-Hydraulics of Water Cooled Nuclear Reactors*, by F. D'Auria, pp. 831-803, Woodhead Publishing.

Hemez, F., H. Sezer Atamturktur and C. Unal (2010), "Defining predictive maturity for validated numerical simulations", in *Computers and Structures*, 88, pp. 497-505.

Malczewski, J. and C. Rinner (2015), *Multicriteria Decision Analysis in Geographical Information Science*, Springer.

Mascari, F., H. Nakamura, K. Umminger, F. De Rosa and F. D'Auria (2016), "Scaling issues for the experimental characterisation of reactor coolant system in integral test facilities and role of system code as extrapolation tool", in Proceedings of NURETH-16, Chicago, United States.

Mu, E. and M. Pereyra-Rojas (2017), *Practical Decision Making*, Springer.

Minouni, S. and G. Serre (2001), "List of benchmarks for simulation tools of steam-water two-phase flows", in Proceedings of ICONE 9, Nice, France.

NEA (1994a), "Separate effects test matrix for thermal-hydraulic code validation – Volume I: Phenomena characterisation and selection of facilities and tests", CSNI/OECD/GD(94)82, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_15968.

NEA (1994b), "Separate effects test matrix for thermal-hydraulic code validation – Volume II: Facility and experiment characteristics", CSNI/OECD/GD(94)83, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_15968.

NEA (2017), "Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark - Final report", NEA/CSNI/R(2016)18, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19752.

NEA (2016), "Scaling in System Thermal-hydraulics Applications to Nuclear Reactor Safety and Design: A State-of-the-art Report", NEA/CSNI/R(2016)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19744.

NEA (2001), "Validation matrix for the assessment of thermal-hydraulic codes for VVER LOCA and transients", NEA/CSNI/R(2001)4, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_17492.

NEA (1996a), "CSNI Integral Test Facilicity validation matrix for the assessment of thermal-hydraulic codes LWR LOCA and transients", NEA/CSNI/R(96)17, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_16134.

NEA (1996b), "Evaluation of the separate effect test (SET) validation matrix", OECD/GD(97)9, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_36360.

Petruzzi. A. and F. D'Auria (2016), "Standardized Consolidated Calculated and Reference Experimental Database (SCCRED): A Supporting Tool for V&V and Uncertainty Evaluation of Best-Estimate System Codes for Licensing Applications", in *Nuclear Science and Engineering*, 182, pp. 13-53.

Pronzato, L. and W. Muller (2012), "Design of computer experiments: space filling and beyond", in *Statistics and Computing*, 22(3), pp. 681-701.

Roy, B. (1996), *Multicriteria Methodology for Decision Aiding*, Kluwer Academic Publishers.

Saaty, T.L. (1982), *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World*, Analytic hierarchy process series, RWS Publications.

Skorek, T. (2017), "Input uncertainties in uncertainty analyses of system codes: Quantification of physical model uncertainties on the basis of CET (combined effect tests)", in *Nuclear Engineering and Design*, 321, pp. 301-317.

Unal, C., B. Williams, F. Hemez, S.H. Atamturkur and P. McClure (2011), "Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors", in *Nuclear Engineering and Design*, 241, pp. 1813-1833.

Valette, M., J. Pouvreau, D. Bestion and P. Emonot (2011), "Revisiting large break LOCA with the CATHARE-3 three-field model", in *Nuclear Engineering and Design*, 241, pp. 4487-4496.

4. Selection and assessment of simulation model (Element 3)

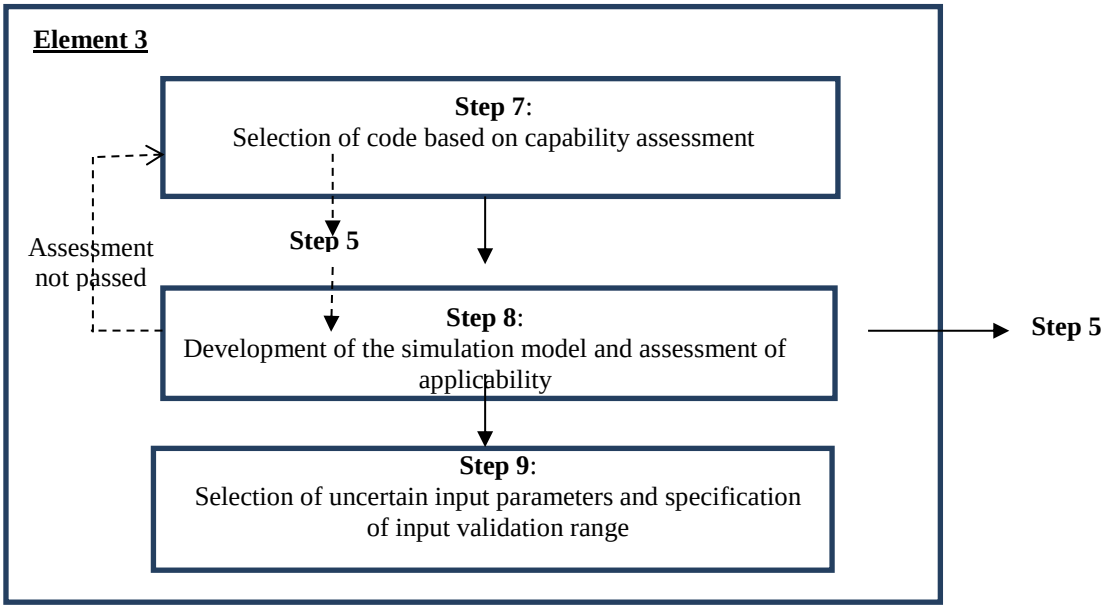
4.1. Introduction

In order to quantify model input uncertainty, the simulation model (SM) should be first developed and assessed so as to adequately predict the SRQs for the transients, or accidents of interest for the test facility (or the plant) from Element 2, and the key input parameters should be identified based on their importance to the modelling of the scenario and their impact on the SRQs for the SM calculations.

The physical models included in the SM and their degree of fidelity in predicting physical phenomena must be consistent with the results of the PIRT process in Element 1. In particular, a process to assess the validation of the developed SM to simulate the considered SETs/IETs shall be considered. It requires establishment of an adequate nodalisation strategy and the selection of physical model options that should be consistent between the experimental facility and similar components in the nuclear power plant .Special attention shall also be devoted to the construction of consistency indicators and the definition of a scale of accuracy to evaluate the accuracy code/experiment. A review of sensitivity analysis methods is provided to identify key model input parameters whose uncertainties shall be quantified for the application to the nuclear power plant.

This Element 3 consists in the three steps displayed in Figure 4.1 and discussed in the sections below.

Figure 4.1. Steps of Element 3



Note that this element is common to the element required in any BEPU methodology, and hence the corresponding good practice guidelines should apply here. This document will focus on the aspects related to the input uncertainty quantification.

4.2. Step 7: Selection of code based on capability assessment

4.2.1. Requirements, features and limitations of thermal-hydraulic system codes

The code for safety analysis is chosen based on the investigated transient after a PIRT has identified the important phenomena for the given scenario and plant (see Element 1).

Based on the code user manuals, the code applicability is determined: the code's formulation, model, and correlations are reviewed to assess if the code has the model/correlations needed to simulate the important phenomena, and if the code has scale-up capability. It is recognised that, while the formulation may be general, the correlations or constitutive relationships are empirical. The code's scalability will depend on these correlations and the underlying tests. If the tests are able to scale up the plant for the phenomena of interest, then correlation is applicable. However, if correlation was derived from tests that do not scale up the plant, the code may not scale up the plant. Code validation with scaled tests or counterpart tests is another way of assessing the code's ability to scale and applicability. For practical purposes, hereafter the attention is given to the system thermal-hydraulics codes.

The typical two-fluids two-fields system thermal-hydraulic codes (such as RELAP5, TRACE, CATHARE2) are based on the solution of six balance equations for liquid and steam that are supplemented by a suitable set of constitutive equations. The balance equations are coupled with conduction heat transfer equations and with neutron kinetics equations (typically point kinetics). The two-phase flow field is organised in a number of lumped volumes connected with junctions. Thermal-hydraulic components such as valves, pumps, separators, annulus, accumulators, etc. can be defined in order to represent the overall nuclear power plant system configuration. In the following sections, the main aspects of a thermal-hydraulic system code – from the point of view of the code user – are highlighted.

System nodalisation

All major existing light water reactor (LWR) safety thermal-hydraulics system codes follow the concept of a "free nodalisation", i.e. the code user has to build up a detailed noding diagram which maps the whole system to be calculated into the frame of a one-dimensional thermal-hydraulic network. To do this, the codes offer a number of basic elements like single volumes, pipes, branches, junctions and heat structures. This approach provides not only a large flexibility with respect to different reactor designs, but also makes it possible to predict separate effect and integral test facilities that might deviate considerably from the full-size reactor.

Because of this rather open strategy, a large responsibility is passed to the user of the code in order to develop an adequate nodalisation scheme, which makes best use of the various modules and the prediction capabilities of the specific code. Due to the existing code limitations and to economic constraints, the development of such a nodalisation always represents a compromise between the desired degree of resolution and an acceptable computational effort. It is not possible here to cover all the aspects of the development of an adequate nodalisation diagram; however, two crucial issues will be briefly mentioned which illustrate the basic problem.

Spatial convergence

As has been often misunderstood, a continuous refinement of the spatial resolution (e.g. a reduction of the cell sizes) does not automatically improve the accuracy of the prediction. There are two major reasons for this behaviour:

(1) The large number of empirical constitutive relations used in the codes has been developed based on a fixed (in general coarse) nodalisation;

(2) The numerical schemes used in the codes generally include a sufficient amount of artificial viscosity, needed to provide stable numerical results. A reduction of the cell sizes below a certain threshold value might result in severe non-physical instabilities.

From those considerations, it can be concluded that no a priori optimal approaches for the nodalisation scheme exist.

Mapping of multidimensional effects

Multidimensional effects, especially with respect to flow splitting and flow merging processes (e.g. the connection of the main coolant pipe to the pressure vessel), exist also in relatively small-scale integral test facilities. The problem might become even more complicated due to the presence of additional bypass flows and a large re-distribution of flow during the transient. It is left to the code user to determine how to map these flow conditions within the frame of a one-dimensional code, using the existing elements like branch components, multiple junction connections or cross-flow junctions. These two examples show how the limitations in the physical modelling and the numerical method in the codes have to be compensated by an "engineering judgement" of the code user, which, at best, is assessed based on results of detailed sensitivity studies. However, in many cases, due to a lack of time or appropriate experimental data, the user is forced to make ad hoc decisions.

Code user options: physical model options

Even though the number of user options has been largely reduced in the advanced system TH codes, there are various possibilities for the code to physically model specific phenomena. They include:

- Choice between engineering type models for choking flow or use of code implicit calculation of critical two-phase flow conditions.
- Flow multipliers for subcooled or saturated choked flow.
- Type of correlations and constants for the countercurrent flow limitation (CCFL).
- Tube or bundle interphase friction or heat transfer models.
- The efficiency of separators.
- Two-phase flow characteristics of main coolant pumps.
- Pressure loss coefficient for pipes, pipe connections, valves, branches, etc.

Since in many cases direct measured data are not available, or at least not complete, the user is left to his engineering judgement or limited assessment to specify those parameters.

Input parameter related to specific system characteristics

The assessment of LWR safety codes is mainly performed based on experimental data coming from scaled integral or separate effects test facilities. Typically, in these scaled-down facilities, specific effects that might be small or even negligible for the full-size reactor case can become as important as the major phenomena to be investigated. Examples include the release of heat from the structures to the coolant, heat losses to the environment, or small bypass flows. The quality of the prediction often depends largely on the correct description of those effects, which needs a very detailed representation of the structural materials and a good approximation of the local distribution of the heat losses. However, many times the importance of those effects is largely underestimated and, consequently, wrong conclusions are drawn from results based on incomplete representation of a small-scale test facility.

Input parameters needed for specific system components

The general system thermal-hydraulic behaviour is described in the codes by the major code modules based on a one-dimensional formulation of the mass, momentum and energy equations for the separated phases. However, for a number of system components, this approach is not adequate and consequently additional, mainly empirical models have to be introduced, e.g. for pumps, valves and separators. In general, these models require a large amount of additional code input data that often are not known since they are largely scaling dependent.

A typical example is the input data needed for the homologous curves, which describe the pump behaviour under single and two-phase flow conditions, which in general are known only for a few small-scale pumps. In all these cases, the code user has to extrapolate from existing data obtained for different designs and scaling factors, which introduces further uncertainty in the prediction.

Specification of initial and boundary conditions

Most of the existing codes do not provide a steady state option. In these cases, pseudo-steady state runs have to be performed using more or less artificial control systems in order to drive the code towards the specified initial conditions. The specification of stable initial and boundary conditions and the setting of related controllers require great care and detailed checking. If this is not done correctly, there is a large risk that even small imbalances in the initial data will overwrite the following transient, especially for slow transients and small break LOCA calculations.

Specification of state and transport property data

The calculation of state and transport properties is usually done implicitly by the code. However, in some cases, for example in RELAP5, the code user can define the range of reference points for property tables and, therefore, can influence the accuracy of the prediction. This might be of importance especially in more “difficult regions”, e.g. close to the critical point or at conditions near atmospheric pressure. Another example is constituted by the fuel materials property data: the specification of fuel rod gap conductance (and gap thickness) is an important parameter, affecting core dry-out and rewet occurrences that must be selected by the user.

Selection of parameters determining time step sizes

All the existing codes use automatic procedures for the selection of time step sizes in order to provide convergence and acceptable accuracy in the prediction. Experience shows,

however, that these procedures do not always guarantee stable numerical results and, therefore, the user might often force the code to take minor steps in order to pass through trouble spots. In some cases, if this action is not taken, very large numerical discretisation errors can be introduced in the evolution of any transient scenario and are not always checked by the code user.

Code input errors

In order to prepare a complete input data deck for a large system like a nuclear power plant, the code user has to provide a huge number of parameters, which must be typed one by one. Even if all the codes provided consistency checks, the probability for code input errors is relatively high and can be reduced only by extreme care following clear quality assurance guidelines.

4.2.2. Classification of input uncertainty parameters

The majority of the aspects of the system codes application listed in the previous subchapter are source of simulation model uncertainty. They are defined in uncertainty analysis as input uncertainties.

There are different input uncertainties, which are related to:

- physical models;
- initial and boundary conditions;
- time integration;
- spatial discretisation;
- material properties;
- data concerning facility construction and operation.

The different classes of uncertain input parameters are quantified in different ways using appropriate sources of information. The basis for the state of knowledge quantification of facility description and operation are construction plans and engineering judgement. For the initial and boundary condition uncertainties quantification, the accuracy of measurements and control devices, the fabrication tolerances, and the composition and materials of technical components are important.

Physical model uncertainties are the subject of this guide. In the uncertainty analyses they are represented by suitable uncertain input parameters. The model input uncertainties are expressed by variation of the model output in the code. Generally, there are three possibilities to introduce a model output variation associated with its input uncertainty to the code:

- Correction of the model output by a corrective term:
  - multiplication of the model output by a correction factor;
  - addition of a correction factor to the model output.
- Variation of a key parameter of the model (frequently available in the code input) or a coefficient in the correlation. It can be done by:
  - introduction of the multiplier or additive term for the selected key parameter (coefficient);
  - variation of the absolute value of the parameter.

- Selection of correlations from a set of alternative model formulations.

Model input uncertainty can be expressed by selecting alternative model formulations, e.g. selecting a different correlation of wall heat transfer coefficient for the same flow pattern or defining the application frequency of different model formulations (correlations). However, this option is used in connection with expert judgement based on experience from code application rather than in the quantification based on the comparison with experimental data.

Key model parameters, such as the number of nucleation sites (ATHLET) or characteristic minimum droplet diameter (RELAP), are frequently available in the code input. It is the reason why they are often used for expression of the model input uncertainty. Coefficients in the models (correlations) are usually not available in the input. However, if they are available, e.g. the multiplication factor according to Sideman used in the calculation of evaporation (ATHLET), sometimes they can be used for expression of model input uncertainties.

The most used methods adopt as expression of model input uncertainties the multiplication of the model output and correction of the model key parameters. These methods can be easily applied through determination of model input uncertainties based on comparisons with experimental data using mathematical methods described in the Element 6 (Chapter 5).

The preferred option is the application of a multiplier for the model output. There are two ways of applying the multiplier:

- Variation of output of singular correlations, e.g. correction of the output of interfacial friction correlation for each flow pattern, as dispersed droplet, bubble flow and so on.
- Global multiplier for the output of the whole model, e.g. correction of the output of interfacial friction model as a whole.

Using the model output multiplier has the advantage that it can always express the complete range of model variation associated with determined model uncertainty. In the case of model parameters, it can happen that selected model parameters are not sufficient for expression of the completely detected model uncertainty.

4.3. Step 8: Development of the SM and assessment of applicability

4.3.1. Background and motivation

The main elements of a best-estimate plus uncertainty (BEPU) evaluation model (EM) include a) the best-estimate computer codes; b) the nodalisations, including the procedures for the development and the validation/qualification[17] of the calculation results; and c) the uncertainty quantification methodology, including the procedures for the validation/qualification[1] of the uncertainty evaluations.

17. The term “validation process” is used any time the process involves a comparison between the results of a SM and the experimental data (this activity is typically performed when comparing SM results against SETF/ITF experiments). The term “qualification process” is used when it is not possible to perform a comparison between the results of a SM respect to a reference (this activity is performed when the SM refers to a nuclear power plant simulation).

While the computer code and the uncertainty quantification methodology are respectively the fundamental element of an evaluation methodology (EM) and the characterising element of a BEPU approach, the nodalisations are the result of a brainstorming process by the code user which connect the code with the physical system to be simulated. The importance of the nodalisation in a BEPU framework is acknowledged by CSAU as similar to any code model or correlation in the code: i.e. system nodalisation presents an inherent code uncertainty and like code models and correlations, quantification of nodalisation-based code uncertainty is deemed to be equally important to quantify model accuracy and uncertainty.

It shall be emphasised that the nodalisation features depend upon the objective and the scope of the EM. Here, the focus is on nuclear power plant safety relevant issues, i.e. FSAR Chapter 15 applications. Two broad groups of nodalisations shall be distinguished when applying a BE code-nodalisation to the analysis of nuclear power plant safety relevant issues:

a) The code focus is a component or a "simple" system having parameters directly recognised by the code structure (limited or no user interpretation is needed): in this case the effort requested of the code user is to implement the system or the component properties (typically geometric and thermodynamic) into the nodalisation with limited or no need for engineering judgement;

b) The code has a modular nature and the system to be modelled is complex. In this case, expertise is requested of the user "to make readable" the system peculiarities of the code. This is achieved throughout the nodalisation, whose development does require engineering judgement.

In the former case, different expert users develop the same (or a very similar) nodalisation for an assigned purpose, provided that best practice guidelines are followed whenever available. Sensitivity tests can be performed to demonstrate the nodalisation quality and the achievement of mesh-independence of the results, which means that varying the node density (or the number of nodes) does not make the results change significantly.

In the latter case, a code-user strategy (or "nodalisation techniques") is needed to develop a nodalisation and different groups of expert users may adopt different strategies. In this case, suitable nodalisations may have very different structures and convergence of results when the number of node density is varied, is not expected or cannot be achieved.

The distinction between categories a) and b) is relevant to the present framework. In the former case, the nodalisation can be presumed as embedded into the code architecture and the code validation implies the validation of the nodalisation approach. In the latter case, the system nodalisation is developed by dividing the real plant/facility component volumes into a set of control volumes that are essentially stream-tubes having inlet and outlet flow path connections. It is clear that subdivision of such a complex system can be done in a number of ways. The simplest subdivision of a real plant/facility model would be into a set of control volumes or nodes that are equally sized, but for a successful solution in the case of the analysis, a number of factors must be satisfied: numerical stability, run time and spatial convergence. In addition, engineering judgement is normally used to a wide extent to develop the system nodalisation.

Experience with code assessment case studies and International Standard Problems have shown the dominant effect of the code-user-effect on the predicted system behaviour. How the code user influences the predicted system, behaviour is a crucial point with respect to the quantitative evaluation of the code uncertainties for the "best estimate" code use in nuclear reactor safety. In other terms, the nodalisation structure is subjected to the "code

user-effect" to quite a large extent and the nodalisation itself constitutes an "independent" computational tool, which shall undergo an assessment of applicability. The importance of establishing a procedure for the nodalisation set-up and the assessment of its applicability as part of the evaluation model (EM) is a consequence of the above-mentioned complexity and subjectivity of the process.

In the present chapter, the focus will be on the SM, which is that part of the EM that includes the code, the nodalisation and the algorithms to approximate the solution of physical equations.

The assessment of the applicability of the SM is the process of determining the degree to which the SM is an accurate representation of the reality. The approach to assess the applicability is to measure the agreement between model predictions and experimental data from appropriately designed and conducted experiments. Agreement is measured by quantifying the discrepancy between the experimental data and the SRQ adopting a selected consistency indicator.

Complex model simulations generate an enormous amount of information from which to select the SRQs of interest. Firstly, the selection of the simulation outcome should be driven by application requirements. For example, if a design requirement is that the pressure at a specified location should not exceed some value, then the model validation should focus on the comparison of measured and computed pressure at that location.

SRQs from experimental data and model outputs must be carefully selected. A response may be simple, such as the maximum response's value for all times at a specific location in the computational domain, or more complex, such as the complete response history at a specific location, modal frequencies or peak amplitudes. In other words, a response can be used directly to derive a consistency indicator; in other cases, the responses must be processed further into a form more suitable for comparison against validation requirements.

A consistency indicator is the basis for comparing SRQs from experimental data with model predictions (Petruzzi and Cherubini, 2019a). Consistency indicators must be established during the setting-up of the validation requirements phase of the SM and the primary consideration should be what the model must predict in conjunction with what types of data available from the experiment. Additionally, the indicators should provide a measure of agreement that includes uncertainty requirements, i.e. includes estimates of the numerical and experimental errors.

If the error, e, between experimental data, y, and model prediction, y^* , is given by $e = y - y^*$, a simple indicator could be the expected value of the error, $E(e)$, or the variance of the error, $V(e)$. Other indicators could be, for example: $P(e_{95\%}>x)$, which is the probability, that the 95th percentile associated to e (e is considered as a random variable with a known, e.g. estimated, probability distribution) is larger than x; or, a hypothesis test where the consistency indicator can be exploited for a decision of whether or not the model is contradicted by the data.

The final step in the validation process is to compare values of the selected indicators used to measure the agreement between model outputs with the experimental data, with respect to selected threshold-value requirements. Following the example below, this step implies defining the threshold values for the expected value of the error, E*, or the variance of the error, V^* or the value of x and the significance level for the probability P or the hypothesis test, respectively.

The assessment of the model accuracy following the consistency indicator is inherently achieved at this point.

In carrying out the assessment of applicability of the SM, a range of tests (with different boundary and initial conditions and at different scales) will be employed to demonstrate that the SM has not been tuned to a single test. For integral behaviour assessment, counterpart tests (similar scenarios and transient conditions) in different experimental facilities at different scales should be selected. Assessments using such tests lead to information concerning scale effects on the models used for a particular SM.

The predictability of SM therefore is validated against integral data from test facilities (IET) to confirm their applicability to the transient within the complicated geometry of reactor system, which, in most cases, is outside the range of conditions for separate effect tests (SET) used to develop the models and correlations for the computer codes of concern in the SM. The capability of extrapolation/interpolation from the scaled facility, whether SETFs or ITFs, to the prototype shall therefore be addressed to provide that a good amount of experimental data cover the prototypic conditions that are required for safety analyses.

Distortions in the IET or SET database may arise from scaling compromises (missing or atypical phenomena) in sub-scale facilities or atypical initial and boundary conditions in all facilities. The effects of the distortions should be evaluated in the context of the experimental objectives determined in Element 2. If the effects are important, a return to this step is needed.

The scaling issue arises from the impossibility of obtaining transient data from the prototype system under off-nominal conditions. Solving the scaling issue implies developing approaches, procedures and data suitable for predicting the prototype's performance utilising small-scale models.

Scaling analyses should be conducted to ensure that the data, and the models based on those data, would be applicable to the full-scale analysis of the plant transient. Scaling analyses are employed to demonstrate the relevance and sufficiency of the collective experimental database for representing the behaviour expected during the postulated transient and to investigate the scalability of the SM and its component codes for representing the important phenomena. The need is to demonstrate that the experimental database is sufficiently diverse that the expected plant-specific response is bounded, and the SM calculations are comparable to the corresponding tests in non-dimensional space. This demonstration allows extending the conclusions related to code capabilities, drawn from the assessment activity, comparing calculated and measured test data in the validation domain, to the prediction of plant-specific transient behaviour (application domain).

The code validation process and scaling in code validation process are so intimately related that the words "validation for scaling" (or validation with respect to scaling) has been introduced by some authors (NEA, 2017).

A short description about the process for assessing the applicability of the SM is provided in Section 4.3.2, the methodology to define the consistency indicator is defined in Section 4.3.3, and the available methodologies, procedures and tools are presented in Section 4.3.4.

4.3.2. The process for assessing the applicability of the SM

The process for assessing the applicability of the SM has the goal to demonstrate that the SM results – obtained by the application of the best-estimate code with the nodalisation

developed following the nodalisation techniques – constitute a realistic approximation of the reference plant behaviour (a facility or a full-size nuclear power plant).

The process shall take into account the effect of many different sources of approximation:

- The data of the reference plant or facility available to the code-user are typically non-exhaustive to reproduce a perfect nodalisation of the reference plant.
- The code-user derives from the available data an approximated nodalisation-schematisation of the plant or facility reducing the level of detail of the simulated hardware.
- The code capability to reproduce the hardware, the plant systems and the actuation logic of the systems further reduce the level of detail of the nodalisation-schematisation.

The needs for assessing the SM applicability, i.e. the code capability and the nodalisation features to predict the expected phenomena behaviour, derive from the following statements:

- The code options must be adequate.
- The nodalisation solutions must be adequate.
- The simulation of some systems can be tested only under transient conditions performances that are not involved in the normal plant operation, i.e. during steady-state code condition.
- The capability of the "code-nodalisation" of the system to reproduce relevant thermal-hydraulic phenomena expected in the transient must be tested.

The development of the process for assessing the applicability of the SM shall include:

a) the establishment of necessary requirements for the different aspects mentioned above;

b) the determination of whether or not those requirements are met by the SM for all selected experiments in the database.

The goal of the process for assessing the applicability of the SM is to develop and obtain a qualified SM considering the comparison with the hardware data, the BIC and the time trends of relevant quantities. The process shall distinguish at least between three main steps:

- Step 1: demonstration of the geometrical fidelity of the nodalisation;
- Step 2: demonstration of the achievement of the steady state;
- Step 3: qualitative and quantitative transient analysis.

Criteria for selecting relevant quantities in each of the above three steps and the assessment applicability requirements shall be defined. If any requirement in any of the three steps is not fulfilled, the process of assessment of applicability of the SM is not passed and the main elements of the SM shall be improved. A new process shall be then applied in full, i.e. all three steps, to the modified SM.

In addition, it shall be emphasised that the three steps process for assessing the applicability of the SM applies to all tests of the experimental database used for the validation, and if the process fails for any of the test, the SM shall be modified and improved and a new process shall be applied to the modified SM for all tests of the experimental database. In other terms,

the nodalisation techniques adopted to develop the nodalisation and the code shall be valid over the set of all experimental tests.

4.3.3. Construction of a consistency indicator

Several approaches have been proposed to quantify the consistency indicator of a given code calculation (Ambrosini and Bovalini, 1990; Kunz et al., 2002; Petruzzi and D'Auria, 2008). Even though these methods were able to give some information about the consistency indicator, they were not considered satisfactory because they involved some empiricism and lacked a precise mathematical meaning. Besides, there is subjective engineering judgement at various levels in the proposed methods.

Generally, the starting point of each method is an error function $\Delta F(t)$, by means of which the consistency indicator is evaluated. Some requirements were fixed which an objective error function should satisfy:

1. At any time of the transient, this function should remember the previous history.
2. Engineering judgement should be avoided, or reduced.
3. The mathematical formulation should be simple.
4. The function should be non-dimensional.
5. It should be independent of the transient duration.
6. Compensating errors should be taken into account, or pointed out.
7. Its values should be normalised.

The information contained in the time dependent function $\Delta F(t)$, which is continuously varying, should be condensed to give a limited number of values which could be taken as indexes for quantifying the degree of consistency. This is allowed because the complete set of instantaneous values of $\Delta F(t)$ is not necessary to draw an overall judgement about accuracy. Integral approaches satisfy this requirement since they produce a single value based on the instantaneous trend of a given function of time. On the other hand, searching for functions expressing all the information through a single value, some interesting details could be lost. Therefore, it would be preferable to define methodologies leading to more than one value in order to characterise the code calculation accuracy. Information that comes from the time-trend of a certain parameter, whether physical or derivate, may not be sufficient for a deep comprehension of the concerned phenomenon; in such a case, it may be useful to study the same phenomenon from other points of view, free of its time dependence. In this context, the complete behaviour of a system in periodic regime conditions (periodic conditions due to instability phenomena are excluded explicitly) can be shown by the harmonic response function that describes it in the frequency domain. Furthermore, the harmonic analysis of a phenomenon can point out the presence of perturbations otherwise hidden in the time domain.

4.3.4. Available methodologies/procedures/tools

Among the existing methodologies or tools, the "Validation Procedure[18]" which is part of the NEMM (NINE Evaluation Model Methodology, (Petruzzi and Cherubini, 2019a; Petruzzi and Cherubini, 2019b; D'Auria et al., 1995) seems to be the more rigorous and

18. The Element 3 of SAPIUM deals with the SM of SETF/ITF, thus the "validation procedure" is of interest in this respect.

systematic approach for carrying out the process of assessing the applicability of SM. A brief summary of each step of the process, depicted in Figure 6.2, is provided in Annex C with the goal of constituting a demonstrative example.

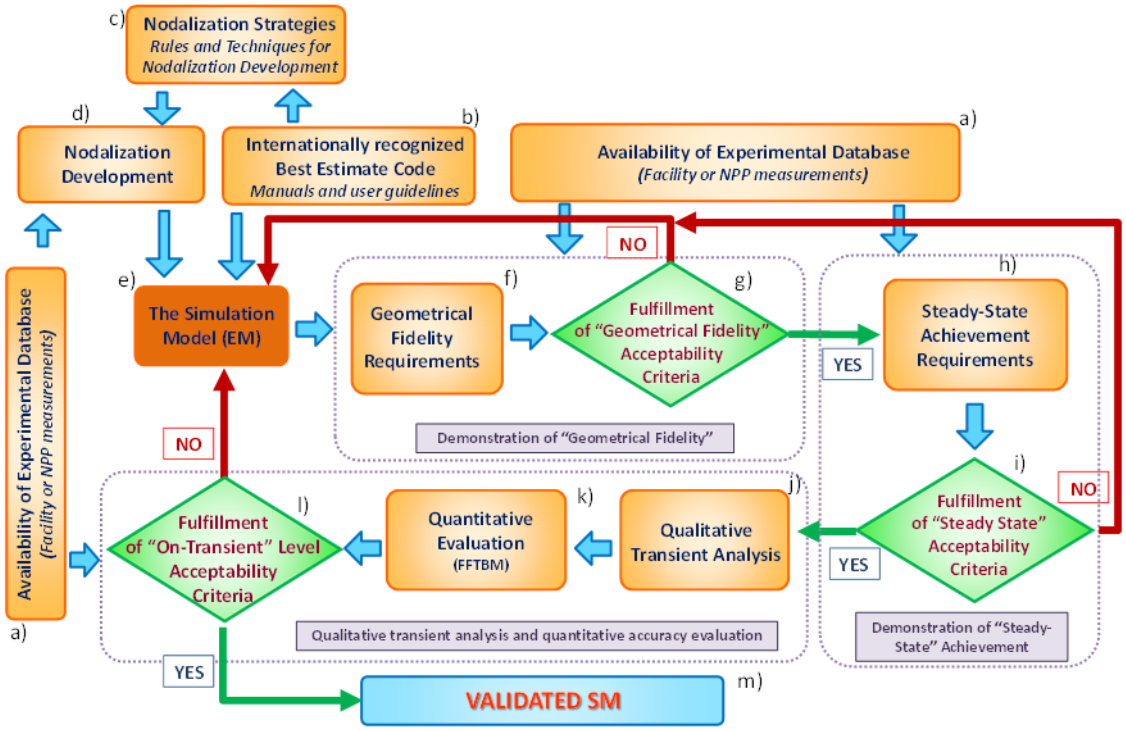
The quantification of the consistency between experimental data and results of the SM constitutes the final step of a validation process. Several approaches have been proposed to quantify this degree of consistency (Ambrosini and Bovalini, 1990; Kunz et al., 2002; Petruzzi and D'Auria, 2008) and hereafter a short summary is provided for two approaches:

- The fast fourier transfer based method (FFTBM) (Ambrosini and Bovalini, 1990), adopted in NEMM (Petruzzi and Cherubini, 2019a) that uses the fast fourier transform to analyse in the frequency domain the relation between two time dependent quantities without losing any information;
- (a part of) The best-estimate results with reduced uncertainties predictive modelling (BERRU-PM) methodology (Cacuci, 2018) that allows for providing a quantitative indicator, constructed from parameter and response covariance and sensitivity matrices, for quantifying in the corresponding metric the consistency (agreement or disagreement) among the experimental data and the SM results.

The consistency indicator provided by BERRU-PM constitutes the most advanced tool nowadays available to perform in a rigorous way from a mathematical (i.e. not subjective) point of view the validation process of a SM simultaneously over all experimental tests of the database. However, its derivation depends on the availability of response sensitivities to system parameters that can be computed efficiently and accurately only if the adjoint sensitivity analysis procedure (ASAP) (Cacuci, 2003) is implemented in the thermal-hydraulics code. As this is not the case for any current system thermal-hydraulics code, only the FFTBM tool is discussed in the following.

The FFTBM allows for a quantitative judgement of a given simulation of an experimental test by the developed SM. Each set of two curves constituted by a calculated and a measured time-trend can be processed by FFTBM. The transformation from time to the frequency domain avoids the dependence of the error from the transient duration. Weight factors are attributed to each time-trend to make possible the summing up of the error and the achievement of a unique threshold for accepting a calculation. The quantification of the consistency between experimental and results of the SM must be carried out following demonstration that the SM results are qualitatively acceptable. The same time trends selected at Item "e" in Figure 6.2 above for carrying out the qualitative transient analysis shall be utilised as input to the FFTBM. More details about the FFTBM can be found in Annex C.

Figure 4.2. Flow chart of the validation procedure of a SM



4.4. Step 9: Selection and specification of uncertain input parameters and confirmation by sensitivity analysis

4.4.1. Background and motivation

The models of complex physical systems, like the BE SM for conducting the safety analysis of a nuclear power plant, are law-driven models characterised by the presence of balance equations and several correlations. Consequently, those models are customarily over-parametrised (e.g. thousands of input parameters are needed to build a typical SM for a nuclear power plant), as they may include more relevant laws than the amount of available data would support for the validation. For the same reason, those models may have also a greater capacity to describe the system under unobserved circumstances (outside the ranges of derivation of the correlations – see the list of "sources of uncertainty" – i.e. outside the validation domain), even though the crucial question stays on the uncertainty of those predictions with respect to reality.

In a perfect world, all model parameters are estimated from the data. The estimation can be reached in different ways: usually it is achieved by minimising, e.g. by least squares, some measure of distance between the model's prediction and the data. At the end of the estimation step, "best" parameter values as well as their "errors" are known. At this point, the model can be considered validated with the "best" parameter values and an uncertainty analysis can be carried out by propagating the "errors" in the parameters through the model, all the way to the model output.

This situation does not apply to the complex physical models, like the BE SM for conducting the safety analysis of a nuclear power plant for which the large part of the parameters and associated "errors" cannot be estimated from the data.

At this point, it is important to highlight a distinction between "important" parameters as those whose uncertainty contributes substantially to the uncertainty of the output results, and "sensitive" parameters as those which have a significant influence on the output results. Indeed, the models are sensitive to input parameters in two distinct ways: 1) the variability, or uncertainty, associated with a sensitive input parameter is propagated through the model, resulting in a large contribution to the overall output variability; and 2) output results can be highly correlated with an input parameter so that small changes in the input value result in significant changes in the output. The necessary distinction between important and sensitive parameters is in the type of analysis conducted: global sensitivity (that addresses parameter importance) and local sensitivity analysis (that addresses the parameter sensitivity). An important parameter is not always sensitive because parameter variability will not appear in the output results if the model is not sensitive to that input parameters. A sensitive parameter, however, is not necessarily important because it may be known precisely, thereby having little variability to add to the output results (Cacuci, 2003).

In practice, the sensitivity analysis (SA) process is therefore an invaluable tool. It allows studying how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input (Saltelli et al., 2004). In engineering studies, numerical model users and modellers have shown strong interest in SA that takes full advantage of the development of computing equipment and numerical methods. The objectives of SA are numerous; one can mention model verification and understanding, model simplification and factor prioritisation. Finally, the SA is an aid in the validation of a computer code, the guidance of research efforts, or the justification of a system design in terms of safety.

Specifying clearly the objectives of a study before performing a sensitivity analysis is essential, as done for example by Saltelli et al. (2004). In the following, it is proposed to take a new look at these sensitivity analysis settings by defining them from an engineering point of view (note that in this section the terms "factors", "inputs" and "model input parameters" are synonyms, as "model" and "simulation model"):

1. The *model exploration* setting aims at understanding the behaviour of the model by trying to decompose the inputs-output relationships.
2. The *factors fixing* setting aims at reducing the number of uncertain inputs by fixing unimportant factors. Unimportant factors are the ones that, if fixed to any value, would not lead to a significant reduction of the output uncertainty.
3. The *factors prioritisation* setting aims at identifying the most important factors. The most important factor is the one that, if fixed, would lead to the greatest reduction in the uncertainty of the quantity of interest (QoI). Let us recall that the QoI is the statistical quantity, applied to the SRQ, that is important in the context of a specific study. For example, a global uncertainty analysis can focus on the variance, as the QoI, while a safety study should concern a high-level quantile or a probability of threshold exceedance. The variance-based sensitivity analysis (where the QoI is the variance) is the most classical approach, but when looking at important inputs in a specific domain of the output values, methods have to be adapted (Raguet and Marrel, 2018).
4. The *input probability density function robustness* setting aims at analysing the QoI variations with respect to uncertainties on the inputs' probability density functions.

At the end of a sensitivity analysis, the analysts will hold a list, or "sensitivity ranking", of the input parameters sorted by the amount of influence each has on the model output. Several sensitivity techniques are available and each of them would result in a slightly

different sensitivity ranking. The actual ranking is not as important as is the specification of which parameters are near the top of the list; in other terms, disagreement among rankings by the various methods for parameters less sensitive is not of practical concern since these variables have little or no influence on model output. This last statement – "little or no influence on model output" - might be false, or completely misleading when an importance analysis (or better known as uncertainty analysis) has to be considered: in this situation, given the practical difficulties/impossibilities to identify the uncertainty ranges (and even more PDF) of input parameters (for over-parametrised models as discussed above), the conclusions that can be drawn from the actual ranking – determined by the sensitivity analysis – might not be appropriate to estimate the uncertainty of the output model responses (it is for instance the case of a very low sensitive input parameter with a large range of uncertainty).

Because of the above discussion, a fundamental step in the application of BE SM to the safety analysis of a nuclear power plant is the identification and characterisation of uncertainties. This is connected with the approximate nature of the codes and of the process of code applications. In other words, "*sources of uncertainty*" affect the predictions of BE codes and therefore the characterisation of the sources must be taken into account before any meaningful attempt to evaluate the uncertainty associated with the output results of the BE SM. The following sources of uncertainty are generally recognised as relevant for the applications of BE SM to the safety analysis of a nuclear power plant:

- Balance (or conservation) equations are approximate.
- Presence of different fields of the same phase.
- Geometry averaging at a cross section scale and at a volume scale.
- Presence of large and small vortex or eddy.
- The numerical solution is approximate and approximate equations are solved by approximate numerical methods.
- Extensive and unavoidable use is made of empirical correlations. Typical situations are:
  - The ranges of validity are not fully specified.
  - Relationships are used outside their range of validation.
  - Correlations are implemented approximately into the code.
  - Reference database is affected by scatter and errors.
- A paradox shall be noted: "steady state" and "fully developed" flow conditions are a prerequisite or condition adopted when deriving correlations. In other terms, all qualified correlations must be derived under the steady state and fully developed flow conditions. However, those conditions apply in almost no region of the nuclear power plant during the course of an accident.
- The state and the material properties are approximate.
- Code-user-effect exists.
- Imperfect knowledge of boundary and initial conditions.

Traditionally, the models of complex physical systems, like the BE SM for conducting the safety analysis of a nuclear power plant, involve two distinct sources of uncertainties, for input parameters, namely: i) the stochastic uncertainty which expresses that the system

under investigation can behave in many different ways; and ii) the subjective or epistemic uncertainty, which derives from the inability to specify an exact value for a parameter that is assumed to have a constant value.

As mentioned above, the uncertainty quantification of BEPU SM consists of several logical steps, one of which is the determination of parameters, which are most sensitive on model results. A "sensitivity analysis" of these parameters can be used to support this step of the uncertainty evaluation (at least for methods based on propagation of input uncertainties).

As explained before, the sensitivity analysis can be either local or global in scope. The objective of local analysis is to analyse the behaviour of the system response locally around a chosen point or trajectory in the combined phase space of parameters and state variables. On the other hand, the objective of global analysis is to determine all of the system's critical points (bifurcations, turning points, response maxima, minima, and/or saddle points) in the combined phase space formed by the parameters and dependent (state) variables, and subsequently analyse these critical points by local sensitivity analysis.

Modellers and analysts may conduct sensitivity analyses for a number of reasons, including the need to determine: 1) which parameters require additional investigation to increase the knowledge base, thereby, reducing the output uncertainty; 2) which parameters are insignificant, and can thus be eliminated from the final model; 3) which input parameters can contribute most to output uncertainty; 4) which parameters are most highly correlated with the output responses; and 5) how much change the output results from varying a given input parameter.

There are many ways of conducting sensitivity analyses and it shall be noted that the various analyses may not produce identical results to the above questions. Generally, sensitivity analyses are conducted by: a) defining the model and its independent and dependent variables; b) assigning probability density functions to each input parameter; and c) assessing the influences and relative importance of each input/output relationship.

In addition, the methods for sensitivity and uncertainty analysis are based on either deterministic or statistical procedures. In principle, both types of procedures can be used for either local or global sensitivity and uncertainty analysis, although, in practice, deterministic methods are used mostly for local analysis while statistical methods are used for both local and global analysis. It is also important to note that all of the statistical methods for uncertainty and sensitivity analysis first commence with the "uncertainty analysis" stage and only subsequently proceed to the "sensitivity analysis" stage; this path is the exact reverse of the conceptual path underlying the deterministic methods for sensitivity and uncertainty analysis, where the sensitivities are determined prior to using them for uncertainty analysis (Cacuci, 2003).

A few of the sensitivity analysis techniques that can be used for highly complex or very large models are presented in the following Section 4.4.2, whereas the characterising elements and critical points of the process for selecting input parameters and associated range of variations are discussed in Section 4.4.3.

4.4.2. Available methodologies/procedures/tools

For simplicity, in this Section, the SM input parameters (just called "inputs") X_i ($i = 1...d$) are scalar and statistically independent, while the SRQ (just called "output") $Y=G(\mathbf{X})$, with $\mathbf{X}=(X_1 \ldots X_d)$, is also a scalar. Depending on the information required by the study and the sensitivity analysis settings, different classes of methods can be defined (see Iooss and

Lemaître (2015) for a previous classification attempt). Such a classification helps the users to select the appropriate method.

4.4.2.1 Screening

First, screening techniques aim at establishing a qualitative ranking of input factors with a minimal budget of model evaluations. The goal is mainly associated to the factor fixing setting. This is often a preliminary step before IUQ using some available information (real output observations, constraints, etc.) or before building a metamodeling. The well-known Phenomena Identification Ranking Table (PIRT) process based mainly on expert and physical analysis corresponds tothis setting. Two kinds of approaches can be distinguished: local and global ones.

The historical approach of SA is deterministic and is known as the local approach. The impact of small input perturbations around input nominal values on the model output is studied. Then, the partial derivatives of the model at this point are calculated or estimated and provide sensitivity values. To be comparable, these local sensitivities have to be normalised, which leads to the so-called elasticity coefficients. To compute the derivatives, the use of adjoint-based methods makes it possible to process models with a large number of input variables. Such approaches are commonly used in solving large environmental systems as in climate modelling, oceanography, hydrology, nuclear engineering, etc. (Cacuci, 2003; Nodet and Vidard, 2017).

The other approach comes from the field of design of experiments (Montgomery, 2004), developed one century ago for physical experiments (and not computer ones). It includes several techniques that answer the screening requirements (as the well-known fractional factorial designs) and, therefore, can be considered as a global (statistical) sensitivity analysis approach. By using only minimal and maximal bounds on the inputs, it also avoids the IUQ step. When the number of inputs is large, it is sometimes necessary to have methods which can require less model runs than model inputs: Superscreening techniques as group screening and supersaturated designs are then useful (Woods and Lewis, 2017). Several techniques have also been developed in the particular framework of computer experiments, the most famous ones being the sequential bifurcation and the Morris method.

An example of a screening sensitivity analysis of a LOCA scenario simulation test case is given (Iooss and Marrel, 2019) in Annex C.

4.4.2.2. Quantitative partitioning and exploration

In the probabilistic approach of uncertainty studies, one has to consider the joint probability density function of the inputs. The sampling-based approach (Helton et al., 2006, De Rocquigny et al., 2008) makes it possible to treat uncertainty propagation and sensitivity analysis steps in the same way. From an initial sample of input and output values (e.g. a Monte Carlo sample or another numerical design of experiments which uses the input joint probability density function), quantitative sensitivity indices can be obtained by analysing the statistical relation between the inputs and the output.

First, the model exploration setting can be addressed via visualisation techniques of the inputs/output sample: tornado diagrams, scatterplots, cobweb plots, etc. (Kurowicka and Cooke, 2006). Statistical models as smoothing techniques can also be useful for visualising main effects of the inputs (Storlie and Helton, 2008). Finally, the use of a surrogate model (also called metamodeling in the mathematical community) makes it possible to predict new points at a minimal CPU time cost and gives access to a full exploration of the input space, possibly via visualisation techniques (Forrester et al., 2008).

For the factors prioritisation setting, which aims at providing a quantitative sensitivity index for each input, many methods have been proposed in the last decades. A list of the most useful and the best-known ones is provided below:

- Linear correlation coefficient and Standard Regression Coefficient (SRC) based on a linear model relating the output and the inputs (Saltelli et al., 2004). The squared SRC of each input represents the part of the output variance (Var(*Y*)) due to this input. Some significativity statistical tests for correlation coefficients can also be applied in order to reject (or not) the hypothesis of non-correlation between an input and an output.
  - Rank correlation coefficient and standard rank regression coefficient (SRRC), which are the same techniques as the previous ones after a rank transformation of the inputs and output samples (Saltelli et al., 2004).
- General variance-based importance measure, which gives, for any non-linear model *G* with finite variations (i.e. $\mathrm{Var}(Y) < \infty$), the parts of the output variance due to each input and each interaction between inputs. It gives the so-called Sobol' indices (Sobol, 1993) that can be estimated by various sampling techniques (Prieur and Tarantola, 2017).
- Derivative-based global sensitivity measures (DGSM), which extend the deterministic local approach to a probabilistic global one (averaging the sensitivities given by the derivatives overall input space) (Kucherenko et al., 2009; Kucherenko and Iooss, 2017). It has been shown that DGSM can be used for screening as well as for factors prioritisation and is of high interest when an adjoint model of *G* is available (Roustant et al., 2014; Roustant et al., 2017). The provided information is strongly linked to the Sobol' indices but with a complementary point of view.
- Density-based importance measures, which are moment-independent metrics. They provide a complementary view to the variance-based ones.
- Dependence measures and kernel-based sensitivity indices (Da Veiga, 2015; De Lozzo and Marrel, 2016) which extend the Sobol' indices to consider higher order information (than the expectation) about the output behaviour in order to provide more detailed information.

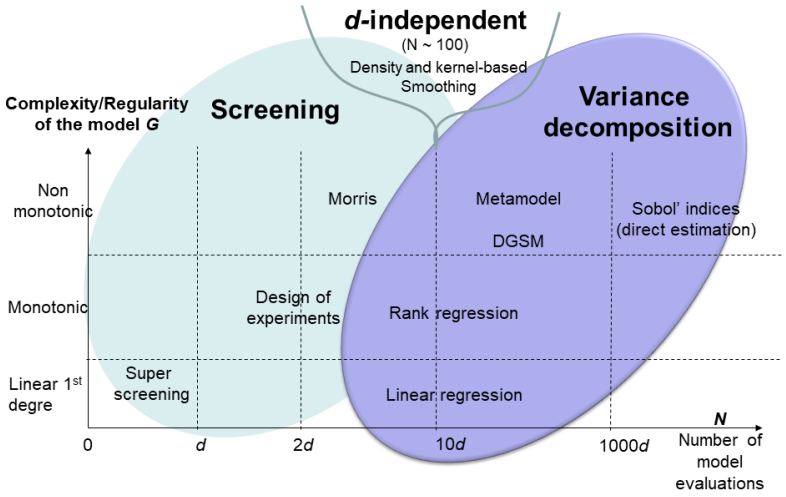
4.4.2.3. Issues for method selection

All the previously presented sensitivity analysis techniques induce a trade-off between the number of model computations required and the assumed model complexity. Figure 4.3 proposes a coarse classification of the families of the global sensitivity analysis methods, mainly related to the factors fixing and factors prioritisation settings. In this graph, the model exploration setting can be achieved through the regression and metamodeling techniques.

This figure shows how to place a method depending on its required number of computations and its underlying hypothesis on the complexity of the *G* model. For example, "non-monotonic" means that the method can be applied to non-monotonic models, as of course to monotonic and linear ones. A distinction is made between screening techniques (which provide qualitative results) and quantitative methods. As most of the methods have a dimension-dependent cost in terms of required model evaluations, another distinction is made with the few methods whose costs are dimension-independent.

Each of these techniques corresponds to different categories of problems encountered in practice. One should use the simplest method, which meets the sensitivity analysis objectives while being adapted to the number of numerical model evaluations that can be performed, and to the prior knowledge on the model's regularity. Each sensitivity analysis should include a validation step, which helps to understand if another method should be applied, if the number of SM evaluations should be increased, and so on.

Figure 4.3. Coarse classification of main global sensitivity analysis methods



As ever said, quantitative sensitivity analysis techniques require specifying a probability density function (pdf) for each input (and a probabilistic dependence structure), which is known to be a difficult task for nuclear systems. This is the main drawback of the quantitative techniques. Very often, the engineering procedure is sequential.

1. A first (a priori) sensitivity analysis is performed with coarse estimations of the pdf.
2. Results help the IUQ process, which makes it possible to define more precisely the input pdf.
3. Then, a second (a posteriori) sensitivity analysis is performed with the updated pdf.

Note that all the above tools look at the sensitivities on the overall input variation space, without any restriction on the output variation space. In most of the applicative studies, sensitivity analysis is applied by using the output variance (or another uncertainty measure) for the QoI, even if the analysis concerns a targeted region of the output variable domain. In the following section, sensitivity analysis techniques looking at more specific QoI will be underlined. In particular, it is important to use the right sensitivity analysis technique when the QoI is associated with rare events (as a high quantile or a probability of failure).

4.4.2.4 Robustness analysis

One of the most critical hypotheses in uncertainty propagation studies is the choice of the distributions of uncertain input variables which are propagated through the SM. In general, such probability density functions come from various sources (statistical inference, design or operation rules, expert judgement, etc.), and are then established with a certain level of accuracy or confidence. Hence, bringing stringent justifications to the overall approach

requires quantifying the impact of the probability density function modelling assumptions on the QoI.

Perturbed Law based Indices (PLI) perform such a robustness analysis, separately from the standard sensitivity analysis methods (Lemaître et al., 2015; Iooss and Lemaître, 2015; Perrin and Defaux, 2018). Thus, it makes it possible to fully address the input probability density function robustness setting of the sensitivity analysis. The PLI principle is to assess the influence of a perturbation on a parameter or a moment of the input probability density function on some QoI. Any QoI can be considered as the mean of the model output, its variance, a probability that the output exceeds a threshold or a quantile of the output. The case of the quantile is studied in (Sueur et al., 2017) and such quantile-PLI are applied on a thermal-hydraulic application case. One great advantage of this technique is that the number of model runs required to estimate the PLI is independent of the number of inputs.

Other solutions can be based on a probabilistic modelling of the parameters of the inputs' probability density function, then on computing sensitivity indices of these epistemic parameters (Morio, 2011). Numerous authors also propose modelling of the inputs based on alternative uncertainty theory and sensitivity tools can be associated to this more complex representation (De Rocquigny et al., 2008).

4.5. Recommendations and open issues

4.5.1. Recommendations

It is recommended that the code for safety analysis be chosen based on the investigated transient after a PIRT has identified the important phenomena for the given scenario and plant (see Element 1).

Systematic methodologies exist for the construction and assessment of a SM. The recommendations are:

- The assessment of the applicability of the SM should be an on-going activity that concludes only when acceptable agreement (with respect to the consistency indicator) between experiment results and simulation predictions is achieved.
- The assessment of the applicability of the SM should be an iterative process that applies to all experimental tests of the validation database. The possible failure of the process for one test implies the improvement of the SM (either code or nodalisation or both) and the repetition of the assessment of the applicability of an SM for all tests of the validation database.
- The validity of a model should be defined over the domain of model form, inputs, parameters and responses. This fact effectively limits use of the model to the particular application for which it was validated; use for any other purpose would require the assessment of the applicability of the SM to be performed again. In other words, the assessment of the applicability process cannot prove that an SM is correct and accurate for all possible conditions and applications, but, rather, it can provide evidence that a SM is sufficiently accurate. Therefore, the assessment of the applicability process is completed when sufficiency is reached.
- The nodalisation strategy and model option selection should be consistent between the experimental facility and similar components in the nuclear power plant. An

appropriate consistency indicator should be defined and verified for assessing the adequacy of the simulation model.

- The SM should not be tuned to a particular data set and the data used to assess the SM should not have been deliberately selected to make the SM appear more accurate than it truly is.

The selection and specification of important uncertain input parameters should be confirmed by sensitivity analysis methods in order to reduce the subjectivity by expert judgement. Much work has been done and many methods and tools are available on sensitivity analysis. It is recommended to use the simplest method in Figure 4.3, which meets the sensitivity analysis objectives while being adapted to the number of numerical model evaluations that can be performed, and to the prior knowledge on the model's regularity. Each sensitivity analysis should include a validation step.

4.5.2. Open issues

Sensitivity analysis remains difficult to apply in an industrial context due to a number of constraints listed below:

1. G is CPU time costly (from several minutes to days to compute one evaluation). In this case, metamodel-based sensitivity analysis (Le Gratiet et al., 2017; Iooss and Marrel, 2017) can be useful. Special attention can be paid to two of the most popular metamodels in computer experiments: the polynomial chaos expansion and the Gaussian process model, for which the Sobol's indices can be efficiently obtained (Le Gratiet et al., 2017).
2. Some model structural errors (also called bias) mean that G is not an exact reproduction of the reality. In this case, if a statistical model of this error has been built, sensitivity analysis can be applied on the additional model (G plus the model error).
3. d is large, which means in the UQ framework: $d > 10...100....$ In this case, screening is recommended. Recent works have studied sensitivity metrics allowing quantitative screening (as HSIC [De Lozzo and Marrel, 2016; Iooss and Marrel, 2019] and DGSM [Roustant et al., 2017]).
4. Statistical dependence/correlation between inputs. The difficulty is to know if the measured influence of an input comes from its real effect in the model or from the input to which it is dependent. In this case, the recently introduced Shapley effects (Iooss and Prieur, 2019) provide interpretable sensitivity measures at a higher computational cost than Sobol's indices (Mara, 2015). However, metamodelling can solve this problem.
5. Some input variables are not quantitative but categorical. This can happen for qualitative variables and scenario parameters. The functional inputs (e.g. stochastic temporal series or random fields) can also be treated as scenario parameters. Several authors have enlightened some methods adapted to this problem (Iooss and Ribatet, 2009; Marrel et al., 2012; Rohmer, 2014).
6. Y is not a single scalar, but a high-dimensional vector, a temporal function, a spatial field, etc. In this case, ubiquitous and aggregated indices (Terraz et al., 2017) can be developed. Visualisation tools of the ensemble of curves/fields of outputs are also important to consider (Popelin and Iooss, 2013; Ribés et al., 2018).

7. The outputs are too voluminous to be registered inside the computer storage disk. In this case, iterative sensitivity analysis techniques have to be developed (in the sense that sensitivity indices are estimated on-line, all along the numerical model calculations). For Sobol's indices, Terraz et al. (2017) have provided a first attempt on this subject.
8. The QoI of the study is not the expectation but another QoI (e.g. a probability of failure or a high-level quantile). For example, the BEPU applications often focus on the 95%-quantile of the output variable. Goal-Oriented Sensitivity Analysis have been developed (Fort et al., 2016; Maume-deschamps and Niang, 2018) in order to extend the Sobol's indices principles to any QoI. More recently, the concepts of target and conditional sensitivity analysis have been investigated (Raguet and Marrel, 2018) by using the general framework of dependence measures.

References

Ambrosini, W. and R. Bovalini (1990), "Evaluation of Accuracy of Thermal-hydraulic Codes Calculations", in *Energia Nucleare 7*.

Cacuci, D.G. (2003), *Sensitivity and uncertainty analysis – Theory*, Chapman & Hall/CRC.

Cacuci, D.G. (2018), *BERRU Predictive Modeling: Best-Estimate Results with Reduced Uncertainties*, Springer Nature, New York & Heidelberg.

D'Auria, F., N. Debrecin and G.M. Galassi (1995), "Outline of the Uncertainty Methodology based on Accuracy Extrapolation (UMAE)", in *J. Nuclear Technology*, 109(1), pp. 21-38.

Da Veiga, S. (2015), "Global sensitivity analysis with dependence measures", in *Journal of Statistical Computation and Simulation*, 85, pp. 1283-1305.

De Lozzo, M. and A. Marrel (2016). "Estimation of the derivative-based global sensitivity measures using a Gaussian process metamodel", in *SIAM/ASA Journal on Uncertainty Quantification*, 4, pp. 708-738.

De Rocquigny, E., N. Devictor and S. Tarantola (2008), *Uncertainty in industrial practice*, Wiley.

Forrester, A. et al. (2008), *Engineering design via surrogate modelling: a practical guide*, Wiley.

Fort, J.C., T. Klein and N. Rachdi (2016), "New sensitivity analysis subordinated to a contrast", in*Communications in Statistics - Theory and Methods*, 45, pp. 4349-4364.

Helton, J.C., J.D. Johnson, C.J. Sallaberry and C.B. Storlie (2006), "Survey of sampling-based methods for uncertainty and sensitivity analysis", in *Reliability Engineering and System Safety*, 91, pp. 1175-1209.

Iooss, B. and M. Ribatet (2009), "Global sensitivity analysis of computer models with functional inputs" in *Reliability Engineering and System Safety*, 94, pp. 1194-1204.

Iooss, B. and P. Lemaître (2015), "A review on global sensitivity analysis methods", in C. Meloni and G. Dellino, editors, *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, pp. 101-122.

Iooss, B. and A. Marrel (2017), "An efficient methodology for the analysis and metamodeling of computer experiments with large number of inputs", in *Proc. of UNCECOMP 2017 Conference*, Rhodes Island, Greece.

Iooss, B and C. Prieur (2018), "Shapley effects for sensitivity analysis with correlated inputs: comparisons with Sobol' indices, numerical estimation and applications", in *International Journal of Uncertainty Quantification*, In press, https://arxiv.org/abs/1707.01334.

Iooss, B. and A. Marrel (2019), "Advanced methodology for uncertainty propagation in computer experiments with large number of inputs", in *Nuclear Technology*, in press.

Kucherenko, S., M. Rodriguez-Fernandez, C. Pantelides and N. Shah (2009). "Monte Carlo evaluation of derivative-based global sensitivity measures", in *Reliability Engineering and System Safety*, 94, pp. 1135-1148.

Kucherenko, S. and B. Iooss (2017), "Derivative-based global sensitivity measures", in R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncertainty Quantification*, pp. 1241-1263.

Kunz, R.F., G.F. Kasmala, J.H. Mahaffy and C.J. Murray (2002), "On the automated assessment of nuclear reactor system code accuracy", in *Nuclear Engineering and Des*ign, 211(2-3), pp. 245-272.

Kurowicka, D. and R. Cooke (2006), *Uncertainty analysis with high dimensional dependence modelling*, Wiley.

Le Gratiet, L., S. Marelli and B. Sudret (2017), "Metamodel-based sensitivity analysis: Polynomial chaos expansions and Gaussian processes", In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncertainty Quantification*, pp. 1289–1325.

Lemaître, P., E. Sergienko, A. Arnaud, N. Bosquet, F. Gamboa and B. Iooss (2015), "Density modification based reliability sensitivity analysis", in *J. of Statistical Computation and Simulation*, 85, pp. 1200-1223.

Mara, T., S. Tarantola and P. Annoni (2015), "Non-parametric methods for global sensitivity analysis of model output with dependent inputs", in *Environmental Modeling & Software*, 72, pp. 173-183.

Marrel, A., B. Iooss, S. Da Viega and M. Ribatet (2012), "Global sensitivity analysis of stochastic computer models with joint metamodels", in *Statistics and Computing*, 22, pp. 833-847.

Maume-Deschamps, V. and I. Niang (2018), "Estimation of quantile-oriented sensitivity indices", in *Statistics and Probability Letters*, 134, pp. 122-127.

Montgomery, D.C. (2004), *Design and analysis of experiments*, John Wiley & Sons, 6th edition.

Morio, J. (2011), "Influence of input pdf parameters of a model on a failure probability estimation", in *Simulation Modelling Practice and Theory*, 19, pp. 2244-2255.

NEA (2015), "PREMIUM Phase II report: Identification of Influential Parameters", NEA/CSNI/R(2014)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19582.

NEA (2017), "A state-of-the-art report on scaling in system thermal-hydraulics applications to nuclear reactor safety and design", NEA/CSNI/R(2016)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19744.

Nodet, M. and A. Vidard (2017), "Variational methods", in R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncertainty Quantification*, pp. 1123-1142.

Perrin, G. and G. Defaux (2018), "Efficient estimation of reliability-oriented sensitivity indices", submitted, URL https://hal.archives-ouvertes.fr/hal-01689366.

Petruzzi, A. and F. D'auria (2008), "Thermal-Hydraulic System Codes in Nuclear Reactor Safety and Qualification Procedures", in Hindawi Publishing Corporation *Science and Technology of Nuclear Installations*, 2008, Article ID 460795.

Petruzzi, A. and M. Cherubini (2019a), "NINE Evaluation Model Methodology", submitted for publication in *Nuclear Engineering and Design*.

Petruzzi, A. and M. Cherubini (2019b), "NEMM - NINE Evaluation Model Methodology", in *Proc. Of NURETH-18*, Portland, United States.

Popelin, A-L. and B. Iooss (2013), "Visualization tools for uncertainty and sensitivity analyses on thermal-hydraulic transients", in *Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2013 (SNA + MC 2013)*, Paris, France.

Prieur, C. and S. Tarantola (2017), "Variance-based sensitivity analysis: Theory and estimation algorithms", in R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncer- tainty Quantification*, pp. 1217-1239.

Raguet, H. and A. Marrel (2018), "Target and conditional sensitivity analysis with emphasis on dependence measures", submitted, https://arxiv.org/abs/1801.10047.

Ribés, A., J. Pouderoux and B. Iooss (2018), *A visual sensitivity analysis for parameter-augmented ensembles of curves*. In *Journal of Verification, Validation and Uncertainty Quantification*, Under review.

Rohmer, J. (2014), "Combining meta-modeling and categorical indicators for global sensitivity analysis of long-running flow simulators with spatially dependent inputs", in *Computational Geosciences*, 18, pp. 171-183.

Roustant, O., J. Fruth, B. Iooss and S. Kohnt (2014), "Crossed-derivative-based sensitivity measures for interaction screening", in *Mathematics and Computers in Simulation*, 105, pp. 105-118.

Roustant, O., F. Barthe and B. Iooss (2017), "Poincaré inequalities on intervals - application to sensitivity analysis", in *Electronic Journal of Statistics*, 2, pp. 3081-3119.

Saltelli, A., S. Tarantola, F. Campolongo and M. Ratto (2004), *Sensitivity analysis in practice: A guide to assessing scientific models*, Wiley.

Sobol, I.M. (1993), "Sensitivity estimates for non linear mathematical models", in *Mathematical Modelling and Computational Experiments*, 1, pp. 407-414.

Sueur, R., B. Iooss and T. Delage (2017), "Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case", in *Proc. of the 10th International Conference on Mathematical Methods in Reliability (MMR 2017)*, Grenoble, France.

Terraz, T., A. Ribes, Y. Fournier, B. Iooss and B. Raffin (2017), "Large scale in transit global sensitivity analysis avoiding intermediate files", in *Proc. of the International Conference for High Performance Computing, Networking, Storage and Analysis (Supercomputing)*, Denver, United States.

Woods, D. and S. Lewis (2017), "Design of experiments for screening", in R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncertainty Quantification*, pp. 1143-1185.

5. Model input uncertainty quantification (Element 4)

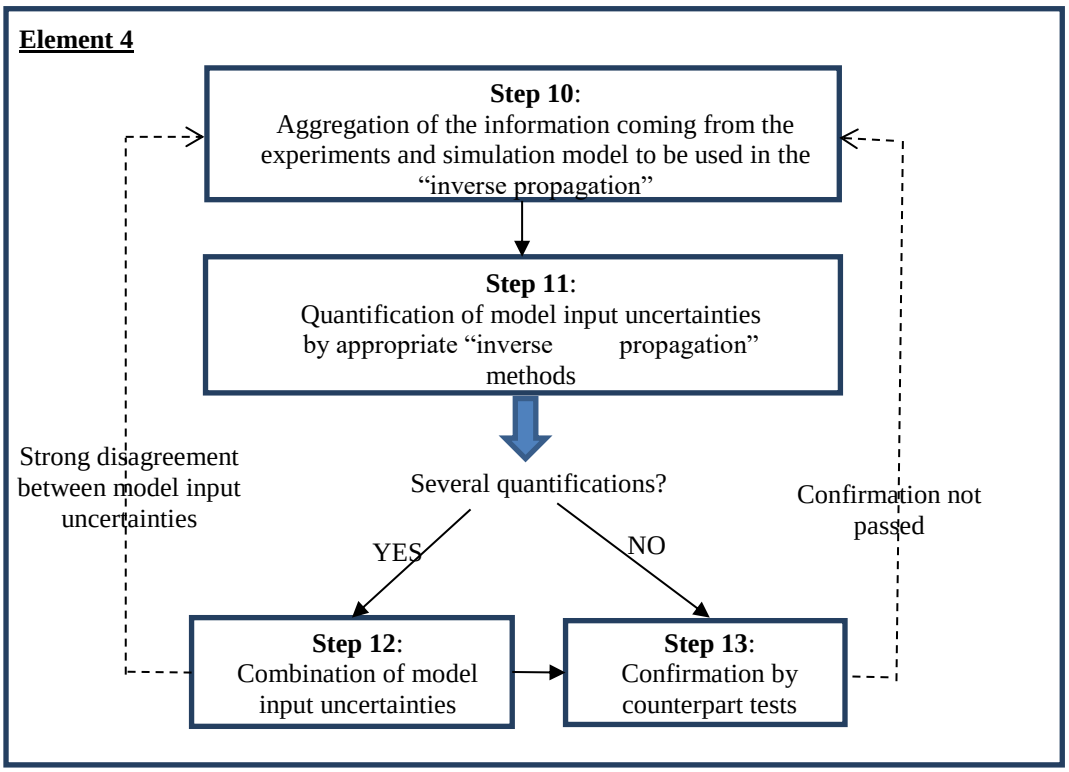
5.1. Introduction

Exploiting the experimental database (Element 2) and the simulation model (Element 3) described in the previous sections, the fourth element of SAPIUM is devoted to the uncertainty quantification of model input parameters, which is a type of inverse problem (IP).

The information coming from Elements 2 and 3 to be treated by Element 4 is a set of couples simulation/experimental values associated to different tests and experiments, components, SRQs. For example, in the PREMIUM project, the different experiments correspond to six FEBA tests while the set of SRQs includes clad temperature and quench time. Experimental uncertainties are also important to take into account when available.

Four steps are identified in this element, as shown in Figure 5.1.

Figure 5.1. Steps of Element 4



5.2. Step 10: aggregation of the information from the experiments of the database and from the simulation model to be used in the "inverse propagation"

This step consists in aggregating the information from Elements 2 and 3 before performing the inverse uncertainty quantification. In practice, it appears that different couples' simulated/experimental values can have different degrees of importance for the intended

use. This should be taken into account for the development of the quantification method to avoid under or overestimating the influence of a given couple.

The degree of importance depends on the type of SRQs (Element 1), e.g. clad temperature might be more relevant than quench time when focusing on safety studies, on the representativeness of each experiment of the database (Element 2) and on the assessment analysis performed in Element 3. It can numerically be taken into account through the assignment of weights to each data.

Along the present chapter, it is shown that, nowadays, the most used inverse methods are probabilistic (statistical). These methods are based on the use of the so-called likelihood function, which is constructed from a database of responses, typically as a product of density functions evaluated on data points. A possibility to distinguish the importance of data is to assign the different weights (in the form of exponents) to the factors of the likelihood.

It is also stated in the chapter that a classical method to solve inverse problems is least squares (LS), based on the minimisation of a sum of squared terms, one for each data point. It is possible to assign different weights to the terms of this sum, and thus obtain weighted least squares (WLS) solutions. Such solutions are in general those obtained using data assimilation approaches, which are related to Bayesian estimation under certain assumptions.

Criteria to assign weights to different databases and data points should be further studied and discussed.

5.3. Step 11: quantification of model input uncertainties by "inverse propagation"

5.3.1. Inverse problems and methods

5.3.1.1. General framework

In science and engineering, the term *inverse problems* refers to the problems focused on getting information about unknown properties of an object by using indirect and possibly noisy data. The term "indirect" means here that the data do not correspond to the unknown properties one aims to learn about, but from magnitudes that are related to them. Moreover, "noisy" means that data are afflicted by errors, for instance measurement errors.

Sometimes it is said that the object is *recovered* or *reconstructed* from indirect observations, which correspond to magnitudes called *responses*. In other words, IPs reconstruct unknown causes from known effects.

The theory on inverse problems comprises mathematical techniques for obtaining useful information about physical systems based on observations (Tarantola, 2005).

Inverse problems (IP) are ubiquitous. Strictly speaking, they are a part of mathematics, but as practical problems they arise almost in every branch of science and engineering. Nowadays, they constitute a very active field of research. Annex D.1 contains a deeper introduction to inverse problems. It also includes an overlook of inverse problems in science and engineering.

The predictive models (e.g. thermohydraulic simulation models) considered in the present chapter are *deterministic* in the sense that they assign always the same SRQ value to a given input value. Formally, the deterministic model can be represented as a function or operator F acting on the input x and producing the response y:

$y = F(x)$	(5.1)

In (5.1), F is the input-output mapping associated to a thermal-hydraulic system code (CATHARE, RELAP, etc.). In the "black box" approach, the model is represented by the associated mapping. The unknown input x and the response y can be scalar or multidimensional quantities (finite dimensional IP) or functions (infinite dimensional IP).

In (5.1) y is the response predicted by the model. The real or true value y_T of the response can be written as:

$y_T = F(x) + b$	(5.2)

b is the model bias, defined as the difference between the real value and the predictive value of the response. In simulation codes (e.g. system codes), the bias function can include various types of model imperfections, such as numerical errors, imperfect underlying physical models or scaling corrections.

The bias depends on the input x, and typically is a quantity imperfectly known (otherwise, it would be used to construct a perfect model). Therefore, it is common to model the bias as an uncertain quantity. Overlooking the model bias in the analysis of IPs may lead to deficient estimates of the unknowns input parameters.

Expression (5.2) relates the true and the predicted response. However, in the IP, the aim typically is to estimate x from measured values of the response, which can be expressed as:

$y_M = y_T + e$ $= F(x) + b + e$	(5.3)

(5.3) is equivalent to:

$y_M = F(x) + \varepsilon$	(5.4)

In (5.4) ε is the difference between the measured and predicted response. Usually, it is termed the data (or observational) noise, and is the sum of measurement error and model error. The IP seeks a solution x from the noisy responses in (5.4).

In modelling and simulation (M&S), inverse problems can be solved with or without uncertainty. When IPs are solved without uncertainty, point estimates of input parameters are obtained, and one can properly speak of "deterministic inverse problems". On the other

hand, the IP may be aimed at the uncertainty quantification (UQ) of unknown inputs. An example of IP without uncertainty is model calibration, defined in this document (see the Glossary) as the process of adjusting model selected parameters (termed calibration parameters) in order to adapt the model predictions to a set of experimental data. On the other hand, an example of IP with uncertainty is, obviously, input uncertainty quantification, defined by the Glossary as the estimation of the uncertainty associated with model inputs (that does not include model calibration). It is possible, also, to perform simultaneously the calibration and the UQ of the calibration parameters. This is actually what is often done by practitioners who are used to running a best-estimate calculation at a calibrated parameter as well as assessing the uncertainty related to it.

The uncertainties/errors leading to code imperfections come from the uncertainty of input model parameters (on which SAPIUM focuses) and all the other sources of model errors including nodalisation options, model assumptions, and so on (see Chapter 4). These sources should be taken into account as far as possible in the IP because they can significantly affect the solution.

The calibration, or the input uncertainty quantification, of the model represented by F uses a database formed by pairs (x_j, y_j), $j = 1,\ldots,N$, and obtains estimates (with or without uncertainty) of the calibration parameters. These inverse methods are "parameter estimation" methods.

When there are few calibration parameters, the problem (5.4) is typically *overestimated* (more data than unknowns did), and commonly has no solution. Then, approximate solutions, obtained via classical techniques such as regression and least squares, are adopted.

On the opposite side, when F represents a very complex model with many free parameters, the problem (5.4) is typically *underestimated* and has multiple solutions. In that case, the LS solution is not unique, and additional criteria are needed to obtain unique solutions.

Critical points

One of the distinctive features of inverse problems is ill-posedness. A problem is termed well posed (in the sense established by Hadamard) when it fulfils the three following conditions (Hadamard, 1902):

- EXISTENCE: the problem has a solution.
- UNIQUENESS: the solution is unique.
- CONTINUITY: the solution depends continuously on the data.

In the sense of Hadamard, a problem is defined as ill posed when it is not well posed.

In general, IPs do not have unique solutions; the model to the same response value can map different assignments of inputs values. Non-uniqueness may be seen because of scarcity of data and information. The general procedure to restore uniqueness is to inject more information to the problem, e.g. prior information about the solution.

Many IPs have unstable or non-robust solutions, meaning that small disturbances on data (due to e.g. the noise) can produce large changes on the solutions and make the computed solution useless. Continuity of the solution as a function of data is a necessary but not sufficient condition for stability or robustness of the solution. The solution of a well-posed IP can be very sensitive with respect to the data (i.e. small changes on data produce large

changes on solution). In this case, the IP and its solution are termed ill conditioned, and it resembles very much an ill-posed problem.

When the IP has no solution, the general procedure is to reformulate the problem and look for approximate solutions. A first possibility is to look for least squares (LS) solutions (i.e. minimising the Euclidean norm of the difference between the predicted responses and data). The LS solutions are also called pseudo-solutions. Model-fitting (including regression and least squares techniques) is an example of inverse problem (Chavent, 2009; Bates and Watts, 1988).

This reformulation may make the IP well posed. However, if the LS problem is also ill posed, the so-called regularisation methods can be applied where additional prior information about the solution and the noise is used.

A major issue in the solution of IP is the construction and analysis of regularisation methods and numerical schemes of solution.

5.3.1.2. The different approaches

This section is devoted to a brief description of the different categories of approaches used in nuclear safety applications. Examples of methods for each category are given in Section 5.3.2. The simplest and the most widespread method of inverse quantification of model uncertainties is the quantification on the basis of separate effects tests (SETs), where singular measurement representing the phenomenon can be compared with associated calculated parameter (SRQ) (Skorek, 2004), e.g. measured void fraction in the channel is associated with phase relative velocity, only. The population of point values (multipliers) obtained from the comparison can be approximated as a probability distribution representing model uncertainty (e.g. relative velocity model). Among the advanced methods addressed particularly for CETs, next will be mentioned statistical methods (frequentist and Bayesian), as well as methods based on design of experiments and forward propagation.

Frequentist and Bayesian methods:

A standard method of representing uncertainty is theory of probability, the uncertain quantities being modelled as random variables. The probabilistic framework can be used for solving inverse problems with uncertainty (it is important to recall the insufficiency of information as a characteristic of IP) (Kaipio and Someresalo, 2007; Somersalo, 2004). In this case, the unknown X, the response Y and the noise are modelled as random variables; the given data are a sample of Y. The solution of a probabilistic IP is an exercise of statistical inference, where the probability distribution of X is estimated from a sample of data of Y.

The probabilistic framework allows the modelling of the noise via its statistical properties, which are sometimes the only knowledge about it.

In probabilistic IP, the definition of well-posedness changes significantly with respect to that previously given. Now, it refers to existence, uniqueness and stability of the probability distribution that solves the IP. Enough information must be provided to allow the unambiguous determination of the distribution.

Depending on the statistical framework, two main types of methods can be distinguished: frequentist and Bayesian.

In the frequentist framework, the parameters to estimate are considered as fixed albeit unknown quantities. Then, in problem (5.1) or (5.4), the inputs x are not modelled as random variables. By contrast, the noise and the response are modelled as random variables.

Typically, the solution of statistical IP under the frequentist framework makes use of the maximum likelihood (ML) principle. Considering the density function of the data y conditioned to the value of unknown input x, the likelihood is defined as such density regarded as a function of x. Then, the ML solution of the IP is the value of x that maximises the likelihood. ML estimators have useful asymptotic properties (consistency, efficiency,…), but they are not "optimum" for finite samples.

To maximise likelihood, a possibility is to introduce "hidden" variables, which complete y so that the new likelihood is simpler to calculate iteratively and to converge to the ML solution. This is the basis of expectation-maximisation algorithm and variants as Stochastic expectation-maximisation algorithm (SEM).

The alternative to frequentist methods are Bayesian methods, based on the application of Bayes' rule, which is a procedure for updating information (Nagel, 2017; Idier, 2008; Dashti and Stuart, 2017).

In the Bayesian IP, the response Y, the unknown X and the noise ε are modelled as random variables, the randomness describing the degree of belief or information about the quantities. The information about the unknown input X before the obtainment of the response data (i.e. the a priori or prior information) is captured in a probability distribution termed the prior distribution. Bayes' rule combines the information from the response data with the prior information, simply multiplying the prior density and the likelihood of the Y sample and normalising the product. The outcome (i.e. the solution of the IP) is the posterior distribution, in the form of the density of X conditioned to the Y data. In summary, in the Bayesian framework, one learns from the data by updating prior belief about X by means of the data Y.

In the realm of IP, Bayesian methods are far more used and applied than frequentist ones. One primary reason is well known: in a problem where a major obstacle is the scarcity of information, a formalism that allows the coherent introduction of prior information should be welcomed. In the process of regularisation, previously described in this chapter, additional prior information about the solution is included. The Bayesian approach implements another way of modelling and adding the prior information.

Going a little further, there is a link between statistical IP methods and regularisation procedures. Summarily, it can be stated that:

- Maximum likelihood methods correspond to non-regularised solutions.
- Bayesian methods correspond to regularised solutions.
- This means that the Bayesian inverse problem is, in general, well posed.

The posterior distribution of X is the solution to the Bayesian inverse problem. In practice, its computation is not easy at all. Only in very special cases, closed-form expressions exist for the posterior. In most cases, the posterior must be calculated approximately, an effort that grows with the dimension of x. In this sense, point-value quantities derived from the posterior distribution are useful; typical point estimators are the posterior mean and the posterior mode, the latter also termed the maximum a posteriori (MAP), which is the x value where the posterior pdf (when it exists) attains a maximum.

Among the methods for exploring/computing the posterior distribution, the Markov Chain Monte Carlo (MCMC) method stands out as a procedure to generate samples following the posterior. The idea is to generate a Markov chain, which has the posterior as stationary limiting distribution.

There is an important type of Bayesian inverse methods: those grouped under the name "data assimilation" (DA), which are devoted to solve IP for dynamic models (i.e. models describing the time evolution of a system). In DA, observational data are combined with the dynamical model in order to estimate parameters describing the state of the system, initial conditions, model parameters, etc.

Methods based on design of experiments and forward propagation.

The previous IP methods offer a rigorous mathematical framework to treat the input uncertainty quantification. However, they rely on several assumptions that might be difficult to satisfy in presence of a poor amount of information on uncertainties.

To tackle this situation, a more empirical strategy can be to combine forward uncertainty propagation and fulfilment of requirements on specific SRQ figure of merit such as a coverage rate or a maximum allowed deviation. Then, the input parameter distribution is iteratively adjusted until the requirements are satisfied.

This type of method does not always lead to a distribution. The quantified input uncertainty can be reduced to an interval. Moreover, it can be computationally costly since it requires performing several simulations. To circumvent this limitation, design of experiments (DoEs) can be exploited to reduce the number of simulations.

5.3.1.3. Main criteria for selecting a method

The choice of a method (category) to quantify input uncertainties depends on the problem under study. It should integrate the characteristics of each method and their relevance for the intended use. The main characteristics of a method to focus on for a practical use are the following:

- Solidity:

The method should be constructed in a rigorous mathematical framework. Processes and algorithms to perform the full construction should be available in the literature and previously validated on several problems.

- Flexibility:

The method should allow taking into account different situations associated with the problem being studied. It can involve for example the capability of the method to integrate the presence of noise, to estimate bias or to combine different modellings to respect the state of knowledge on uncertainties.

- Transparency and reproducibility:

A clear and complete documentation describing the different steps of the method construction should be available. It includes the list of assumptions and a guideline to drive the analyst's choice. Uncertainty associated with the method (e.g. when using surrogate models) and its impact on the IUQ results is important information to take into account.

- Reduction of user effect/expert judgement:

All methods have user effect (prior information, acceptability thresholds, choice of a type of surrogate model, etc.). However, it is advisable to choose a method that reduces as much as possible this effect or clearly identifies the different sources in order to evaluate their impact on the uncertainty results.

- Relevance of the method assumptions for the problem under study:

The method construction is based on several assumptions that should be in agreement with the state of knowledge on the problem. It requires for example checking if the nature of uncertainties (aleatory, epistemic) and the lack of information on them are properly taken into account. It is important to assess the robustness of the method, i.e. the applicability of the method even if the assumptions are not valid for the given problem. As an example, a typical assumption is that the discrepancies between predicted and real (measured) responses constitute a simple random sample. In reality, they commonly are obtained from an experimental design, rather than from a random sampling procedure.

- Tractability:

Increasing the flexibility of the method can limit its application on industrial problems and a compromise has to be found. For example, this limitation can be due to high computational costs. Numerical treatments to overcome the limitation, such as the use of surrogate models or the efficient design of experiments, should be integrated in the method construction, provided the associated assumptions are satisfied and the numerical errors quantified.

5.3.2. Methodologies for IUQ

Nuclear engineering and science feature a large number of works on the development and application of inverse methods, referring to two main objectives:

- estimation of physical quantities;
- calibration and model uncertainty quantification.

A paradigmatic example of the estimation of physical quantities via inverse methods is the estimation of nuclear cross sections through neutron analyses. Many nuclear data adjustment methodologies exist and are either Bayesian, which can be broadly classified as deterministic, or Monte Carlo-based. Here, the term "deterministic" is somewhat misleading and should be understood as "probabilistic, but not based on Monte Carlo". Deterministic methodologies assume the linearisation of the model and normality of experimental and nuclear data. On the other hand, stochastic or Monte Carlo-based methods do not rely on linearisation or normality.

Monte Carlo-based uncertainty propagation of nuclear data is now very common, including in criticality safety analysis, reactor core analysis, depletion analysis, activation analysis, etc.

There are methodologies that combine forward uncertainty methods and Bayesian updating algorithms for the prediction of integral functions of nuclear data, e.g. reactor power distributions, neutron multiplication factors or isotopic concentrations in irradiated nuclear fuel. They can use integral experimental data to reduce the prior uncertainty of integral observables. There are methodologies that adjust nuclear data, while others are applied directly to the integral observables without previous adjustment.

The traditional method of propagating nuclear data uncertainties uses adjoint-based first-order perturbation theory. Uncertainties of integral observables are approximated as linear transformations of nuclear data co-variances, defined by sensitivities. This technique can be combined with a least-squares method, so that integral experimental data are used to update prior knowledge about nuclear data. This procedure is often called "nuclear data adjustment".

For more information on this topic see references (Hoefer et al., 2014; Hoefer et al., 2018; Chadwick et al., 2005; Castro et al., 2016) and (Capote et al., 2012; Rochman et al., 2017; Rochman et al., 2018).

In the realm of thermal hydraulics, particularly in connection with system codes, the model uncertainty quantification via inverse methods is a topic of growing interest, as evidenced by the writing of this report. The intense development and application in those fields of forward uncertainty methods, starting in the last decade of the 20th century, needed ancillary methods to calculate one of the main contributions to input uncertainty. Due to physical models imperfection, it brought into focus the uncertainty inverse methods.

This evolution is well illustrated in the sequence of NEA projects on uncertainty, mainly developed by the system codes community: UMS (NEA, 1998), BEMUSE (NEA, 2011), PREMIUM (NEA, 2016; NEA, 2017) and SAPIUM. More generally, IUQ remains an active field of research (see Wu, 2017; Wu et al., 2017; Wu et al., 2018a; Wu et al., 2018d; Wu et al., 2019; Shrestha and Kozlowski, 2016; Wu and Kozlowski, 2017; Pastore et al., 2015; Nguyen and Downar, 2017) for examples of development and application of inverse methods in the field of thermal hydraulics and nuclear fuel thermomechanics.

The next section provides an overview of methods previously applied to the nuclear safety field. The following subsections are devoted to more details on the different frameworks introduced in Section 5.3.2.1.

5.3.2.1. Overview of methods applied until now in the nuclear safety field

In the nuclear community, a growing number of calibration and model uncertainty quantification methods have been developed and applied in the last years, especially in connection with thermal-hydraulic system codes and thermo-mechanical codes. These methods rely on different frameworks explained in the following paragraphs, including data assimilation, frequentist and Bayesian methods and methods based on forward propagation and design of experiments. Below, Table 5.1 summarises the methods proposed in the nuclear safety field that have been largely applied in recent NEA projects.

Table 5.1. Inverse methods applied in the nuclear thermal-hydraulics field

METHOD	PRINCIPLE
CIRCE (CEA)	Frequentist (maximum likelihood)
IPREM (Univ PISA)	DoE/forward propagation
MCDA (KAERI)	Data assimilation
DIPE (IRSN)	DoE/forward propagation
CASUALIDAD	Data assimilation
TRACTEBEL IUQ METHOD	DoE/forward propagation
PSI METHOD	Bayesian

Most of the methods (CIRCE, IPREM, MCDA, DIPE, Tractebel and PSI methods) listed in Table 5.1 have been already described in the final PREMIUM report (NEA, 2017) and their construction is therefore not recalled in this chapter. All methods are probabilistic, except IPREM, which is described in Annex D.3 and is based on uncertainty intervals.

Despite the fact that the whole PREMIUM project could provide a first approach of methodologies comparison, stricter assessments should be found in other comparative studies that avoid circumstances that could disturb it. UPC participated in PREMIUM early phases as a CIRCÉ user, and after the completion of Phase IV decided to perform an additional study by repeating the uncertainty quantification performed in Phase III with another of the available methods. In particular, the CIRCÉ calculation of FEBA was repeated with FFTBM. Some differences were unavoidable due to the nature of each method but the comparison is included in the final PREMIUM report (NEA, 2017) and, along with some additional developments, in a post-PREMIUM journal article (Freixa et al., 2016).

In general, the obtained uncertainty bands with both methodologies enveloped well all experimental data. In addition, the bands provided by FFTBM presented a better balance between the uncertainty on the maximum cladding temperature and the quench time. The results obtained by CIRCÉ displayed narrower bands during the first part of the reflood, when the temperatures are high, and considerably wider bands during the quench.

5.3.2.2. Frequentist approaches

This section is limited to the maximum likelihood (ML) inference applied to parametric statistical models, although other frequentist approaches are available. The advantages of ML estimates are numerous (under mild conditions): consistency, asymptotic normality and efficiency and functional invariance. Moreover, ML inference asymptotically comes to choose the closest distribution of the statistical model from the distribution, which governs the data in the sense of the relative Shannon entropy (also known as Kullback-Leibler discrepancy) (Kullback and Leibler, 1951), which is an interesting property if one considers that a statistical model can be misspecified (Marqués de Sá et al., 2013). Besides, ML inference has an interesting connection with least squares problems and Bayesian inference that is described further.

Statistical methods are always based on the hypothesis that the observed variable Y is random and that this randomness corresponds to an aleatory, not an epistemic, uncertainty. This makes it possible to define a statistical model, hereafter parameterised by θ (i.e. θ represents a parameter or set of parameters identifying a probability distribution). This comes to define a likelihood, that is the PDF of the observed variables $Y^{(i)}$ evaluated at the N available data ($y^{(1)}$, …, $y^{(N)}$) as a function of θ (supposing that this PDF exists). An aleatory measurement error is not the sole possible origin for this randomness: because the IP variable of interest X is uncertain, it may be regarded as random as well. It is important to distinguish two situations:

- (S1) When X is assumed unknown but not random (epistemic uncertainty);
- (S2) When X is assumed random, (there is then epistemic uncertainty about the value of certain parameters of the distribution of X). This is the "Probabilistic IP".

This distinction does not only matter owing to the nature of X, but also because (S2) is a more challenging IP than (S1) from a computing point of view (see next paragraphs).

Both situations are described by expression (5.4). In fact, in situation (S2), there is no unique X, but as many identically distributed $X^{(i)}$ as observations, typically:

$$Y^{(i)} = F^{(i)}\left(X^{(i)}\right) + \varepsilon^{(i)} \quad (5.5)$$

Where $\varepsilon^{(i)}$ stands for the measurement error of the i^{th} observation. Equation (5.5) is the one which is under the CIRCÉ method (de Crécy, 1997) where $X^{(i)}$ plays the role of a multiplicative factor applied to a reference closure relationship C_{ref}:

$$C^{(i)} = X^{(i)} \times C_{ref} \quad (5.6)$$

Where C_{ref} corresponds to the reference closure relationship.

Even in situation (S1), the observations $Y^{(i)}$ are generally not identically distributed because the functions $F^{(i)}$ may not be the same, e.g. because of varying experimental conditions or because of the use of different experimental mock-ups. If the distribution of the $\varepsilon^{(i)}$ is known, the parameters θ to infer are related to X according to the situation. $\theta = X^{(1)} = \ldots = X^{(N)}$ in situation (S1) and $X \sim \pi^X(.|\theta)$ in situation (S2) (i.e. X follows a distribution whose PDF $\pi^X(.|\theta)$ is parameterised by some parameters θ to estimate). Both situations have been considered in the thermal-hydraulic field.

In both situations (S1) and (S2), the mathematical expression of the likelihood can easily be written down knowing the $F^{(i)}$ and the distribution of the $\varepsilon^{(i)}$ and of the $X^{(i)}$ as a function of θ. *The $\varepsilon^{(i)}$ and $X^{(i)}$ are supposed to be mutually independent until the end of this section*. Let $\pi^{(i)}(.|F^{(i)}(x))$ denote the PDF of $Y^{(i)}$ knowing that $X^{(i)} = x$ and, in situation (S2) only, let denote $\pi^X(.|\theta)$ the PDF of the $X^{(i)}$, then the likelihood is:

$$L(\theta|y^{(1)}, \ldots, y^{(N)}) = \prod_{i=1}^{N} \pi^{(i)}(y^{(i)}|F^{(i)}(\theta)) \quad (5.7)$$

for situation (S1) and

$$L(\theta|y^{(1)}, \ldots, y^{(N)}) = \prod_{i=1}^{N} \int \pi^{(i)}\left(y^{(i)}\middle|F^{(i)}\left(x^{(i)}\right)\right) \pi^X\left(x^{(i)}|\theta\right) dx^{(i)} \quad (5.8)$$

for situation (S2).

Therefore, considering the general case of black-box functions $F^{(i)}$, an important difference reveals itself when considering the numerical evaluation of the likelihood: whereas this evaluation only demands the evaluation of the $F^{(i)}(\theta)$ in situation (S1), it requires an integration along the variables $X^{(i)}$ of a term involving $F^{(i)}(X^{(i)})$ in situation (S2), thus potentially thousands of evaluations of the $F^{(i)}$! Surrogate modelling (i.e. the replacement of the original model by a simplified model) should then be required to enable the calculation of the likelihood. Even in situation (S1), surrogate modelling, applied to the $F^{(i)}(.)$ rather

than directly to $L(.|y^{(1)}, \ldots, y^{(N)})$, may be necessary to get a result in a reasonable time if the runs of the $F^{(i)}(.)$ are CPU time consuming.

The next paragraphs provide some numerical methods to maximise the likelihood, focusing on the Gaussian framework, that is assuming Gaussian distributions for the $\varepsilon^{(i)}$ and, in situation (S2), for the $X^{(i)}$. Before, let us remark that staying in the scope of the statistical models deriving from equation (5.5) and from this Gaussian framework is not so restrictive that it may firstly appear thanks to the flexibility in defining the $F^{(i)}$:

- The estimation of some bias between the simulated response and the real physical response is possible. For instance, a systematic bias b can be estimated by replacing θ by $\theta'=(\theta, b)$ and $F^{(i)}(\theta)$ by $F^{(i)}(\theta') = F^{(i)}(\theta) + b$ in situation (S1) in the same way than equation (5.3). The statistical treatment of b is, however, made easier in a Bayesian framework (see section on Bayesian methods).
- The kth components $X_k^{(i)}$ of the $X^{(i)}$ may be assumed independent and log-normally distributed, for example by considering the variables $\log(X_k^{(i)})$ instead of the $X_k^{(i)}$.

Let us remark that some methods have already been proposed to get rid of this statistical framework (for example, see Rachdi, 2011; Kuhn, 2003; Perrin et al., 2007).

In situation (S1), it is possible to apply any generic global optimisation algorithm to minimise $L(\theta|y^{(1)}, \ldots, y^{(N)})$; however, some particular methods deserve to be mentioned.

If the measurement errors are Gaussian and the $F^{(i)}$ are linear (or approximated by a linear function of θ by means of a regression or a first-order derivation), then maximising the likelihood comes to solving an ordinary least squares problem, thus to solving a linear system. It is worth noting that the least squares estimates are relevant for a broader category of statistical problems, as stated by the Gauss-Markov theorem (non-Gaussian errors $\varepsilon^{(i)}$).

The Levenberg-Marquardt algorithm was specifically designed to solve non-linear least squares problems, hence can be efficient to carry out ML inference in the case of Gaussian measurement errors and non-linear $F^{(i)}$.

A popular global optimisation technique is *simulated annealing*. In a Bayesian perspective, this method can be seen as an adaptation of a Metropolis-Hastings algorithm to determine a MAP (maximum a posteriori) estimate.

The situation (S2) corresponds to what is called a *data missing problem* in statistics. In such a problem, the likelihood $L: \theta \mapsto L(\theta|y)$, which corresponds to the "incomplete" data y (the available ones), is expressed in function of another likelihood $\mathcal{L}: \theta \mapsto \mathcal{L}(\theta|y, x)$ which involves some so-called "complete data" (y, x), where x stands for unobserved data:

$$L(\theta|y) = \int \mathcal{L}(\theta|y, x)\, dx \qquad (5.9)$$

Specific algorithms, deriving from the expectation maximisation (EM) algorithm (Dempster et al., 1977), have been developed for a long time to take advantage of the particular structure of such a problem. They consist roughly in predicting or recovering the missing data, that is the unobserved realisations $x^{(i)}$ of the $X^{(i)}$, to iteratively update the value of θ.

The EM algorithm (Dempster et al., 1977) increases the likelihood by updating a current value θ^k thanks to the maximisation of $Q(\theta, \theta^k) = E[\log(\mathcal{L}(\theta|y, X)|\theta^k]$ with respect to θ,

where $E[f(X)|\theta^k]$ denotes the expectation of some function f(X) when X obeys the marginal distribution along x deriving from $\mathcal{L}(\theta^k|y,x)$. Therefore

$$Q(\theta,\theta k) = \int \log(\mathcal{L}(\theta|y,x))\frac{\mathcal{L}(\theta^k|y,x)}{L(\theta^k|y)}dx \qquad (5.10)$$

In practice, this algorithm cannot be implemented efficiently for any statistical model. When $(x,y) \mapsto \mathcal{L}(\theta|y,x)$ has a regular exponential-family form, the maximisation of $\theta \mapsto Q(\theta, \theta^k)$ amounts to carrying out an E (expectation) step, then a M (maximisation) step : the missing data X are "predicted" by $x^k = E[X|\theta^k]$ (E step), then this "prediction" is used to get a new estimate $\theta^{k+1} = \arg\max_\theta \mathcal{L}(\theta|y,x^k)$ (M step). For certain simple statistical models, an explicit mathematical expression of $\theta^{k+1} = \arg\max Q(\theta, \theta^k)$ in function of θ^k can be obtained. If a relevant choice of variable x is made, the EM algorithm is likely to be competitive among the most common optimisation algorithms. It insures the increase of the likelihood at each iteration; however, it may stop at a non-maximal stationary point.

An improvement of EM consists in using the ECME algorithm (expectation/conditional maximisation either) (Liu and Rubin, 1994), whose principle is splitting the parameters θ into two groups θ_1 and θ_2 ($\theta = (\theta_1\ \theta_2)^T$) so as to get, if possible, a better iterate θ^{k+1} from θ^k than the EM algorithm in the following way: ${\theta_1}^{k+1} = \arg\max_{\theta 1} Q((\theta_1\ {\theta_2}^k)^T, \theta^k)$ and ${\theta_2}^{k+1} = \arg\max_{\theta 2} L(({\theta_1}^{k+1}\ \theta_2)^T|y)$. If the $F^{(i)}$ are linear and if the Gaussian framework is assumed, an efficient optimisation of the likelihood can be carried out by the ECME algorithm with θ_1 (resp. θ_2) the covariance matrix (resp. the expectation) of the $X^{(i)}$. This approach was proposed in the CIRCE methodology (de Crécy, 1997), then the authors of (Celeux et al., 2010) suggested to iteratively linearise the functions $F^{(i)}$ at the estimated expectation ${\theta_2}^k$ (if they are not linear). CIRCÉ has been developed as part of a work programme defined in France by Areva, EDF, IRSN and CEA for the CATHARE code. In CIRCÉ, the derivatives of each code response to each parameter are calculated. The main assumptions are normality or lognormality of the multipliers associated with physical models such that $\theta \equiv (m, \sigma^2)$ with m and σ^2 being the parameters of X. Another assumption of CIRCÉ is linearity of the relation between responses and parameters. There is recently a "Bayesian counterpart" of CIRCÉ (see section on Bayesian methods).

Since the latter method cannot deal with non-linear functions $F^{(i)}$ in general, a Stochastic EM (SEM) approach, coupled with some surrogate models of the $F^{(i)}$ by kriging to circumvent their CPU cost if needed, was proposed in (Celeux et al., 2010) (see also (Kuhn, 2003) for the SAEM method) : at the k+1th iteration, assuming that $\theta = \theta^k$, some missing data x^k are randomly drawn according to the distribution of the $X^{(i)}$ conditioned by Y = y, then $\theta^{k+1} = \arg\max_\theta \mathcal{L}(\theta|y,x^k)$. The final estimate is defined as the empirical mean of the θ^k over many iterations. The sampling of the missing data can be performed by some Markov Chain Monte Carlo (MCMC) techniques (which requires some surrogate model of the $F^{(i)}$ if their global CPU cost is not negligible). A Bayesian inference was derived from this SEM method in (Fu et al., 2015); see the next section dedicated to the Bayesian methods.

Bayesian approaches

Many recent developments have used the Bayesian paradigm, modelling epistemic uncertainty with probability distributions. As discussed in Section 5.2.1.2., Bayesian estimation is convenient to regularise maximum likelihood estimation (MLE) when only a few data are available, which often happens in nuclear applications. Besides, such a framework may be more suited than frequentist methods in the context of uncertainty quantification because it can provide not only point estimates, but mostly a probabilistic assessment of θ. Through the Bayes formula, the prior distribution can be combined with the likelihood at the available data in order to derive the posterior distribution.

The prior distribution can be constructed either with physical expertise (by using relevant values such as lower and upper bounds, mean value, median value), or chosen as default when no information can be provided. The latter case corresponds to an objective Bayesian analysis where the impact of the prior on the posterior distribution is expected to be as low as possible. This is the reason that those priors are called non-informative. The most used are the Jeffreys priors and the Berger-Bernardo reference priors (see Kass and Wasserman (1996) for a comprehensive state of the art about them). These priors also satisfy relevant properties such as invariance by any bijective re-parametrisation of the statistical model. However, a critical issue is that they are sometimes improper (i.e. the PDF does not integrate to 1). Fortunately, if the likelihood is derived from both linear models and Gaussian assumptions, the posterior distribution is well defined, making the Bayesian estimation tractable.

In this section, two applications of Bayesian methods for estimation problems arising in nuclear engineering are explored. The first application deals with computer code calibration, inspired by the seminal work of Kennedy and O'Hagan (KOH) (Kennedy and O'Hagan, 2001). The second paragraph addresses the Bayesian version of the CIRCÉ method in order to assess the uncertainty of the closure relationships that are integrated in thermal-hydraulic system codes.

Bayesian calibration and uncertainty quantification of thermal-hydraulic computer codes

In the seminal work of KOH dealing with calibration of computer codes (Kennedy and O'Hagan, 2001), the statistical equation that relates the experimental data $Y^{(i)}$ to the simulations is similar to equation (5.3). It is written as:

$$Y^{(i)} = F^{(i)}(\theta) + b^{(i)} + \varepsilon^{(i)} \quad (5.11)$$

Where the function b is an *external* bias for being specified between the code outputs and the physical system. Hence, equation (5.11) corresponds to the situation (S1) studied in Section 4.2.2.2 (devoted to frequentist methods) though adding an unknown bias function.

As the simulations of $F^{(i)}$ are often time-consuming and b is an unknown function, KOH put functional priors on the two, each being a random Gaussian process (GP). Under such assumptions, a confounding effect may occur between θ and b during the estimation stage. A way to try preventing it is to assume b as a zero mean GP (Higdon et al., 2005) and (or) linearising F around a best guess of the code parameters instead of assuming a GP (Bachoc et al., 2014). The former helps for a better separation between part of uncertainty due to θ and part of uncertainty due to the b term, while the latter gives an exact expression of the θ's

posterior distribution. If the linearisation is not accurate enough, which happens if the computer code output is highly non-linear with respect to θ, then MCMC methods are required for sampling of the posterior distribution. Such algorithms can be referred to as the surrogate models-based MCMC algorithms.

The major interest of considering equation (5.11) is to avoid over-fitting. Indeed, the parametric uncertainty is realistically not the only source of mismatch between the code and the real system being modelled. In thermal hydraulics, a potential limitation in using equation (5.11) comes from the fact that extrapolating the bias from one test to another or from one scale to another may be difficult to be properly justified. In the nuclear field, a new modular approach has been recently proposed (Wu et al., 2018) in order to avoid extrapolating the b function to physical configurations that are outside of the estimation region.

A recent development: Bayesian counterpart of the CIRCE method

Instead of applying the ECME algorithm (see Section 5.2.2.2) to estimate $\theta = (m, \sigma^2)$, the Bayesian approach puts a prior distribution on it. There are two ways to do that:

1. A prior distribution exists from either a previous study or expert judgement.
2. On the other hand, there is no prior information. A prior that is as non-informative as possible should then be specified on (m,σ^2). This is referred to as the objective Bayesian approach.

In the second case, it is expected that the Bayesian inference is driven by the data as much as possible (and not by the prior). Choosing the Jeffreys prior on θ leads to an improper posterior distribution. Instead, a Gaussian-inverse gamma prior can be specified with parameters making it little informative. The posterior distribution $\theta|Y$ is now proper and can be sampled conveniently using a Gibbs algorithm based on an overall set of full conditional distributions (see Damblin and Gaillard [2018] for technical details).

The Bayesian CIRCÉ should be interpreted as an approach that is able to provide a fuller calibration and uncertainty quantification of the closure relationships than a point estimate can do. Finally, it is possible to propagate the statistical uncertainty of (m,σ^2) to X by calculating:

$\pi(X) = \iint \pi(X\|m,\sigma^2)\pi(m,\sigma^2\|Y)dm\, d\sigma^2$	(5.12)

The probability density $\pi(X)$ is the marginal density of X, which is not (log)-Gaussian (because $\pi(\theta|Y)$ is not (log)-Gaussian). At the end, if the mean or the median of X is significantly far away from one, then the closure relationship is said biased. Unlike in the KOH method, here it's important to talk about *internal* biases. The justification to extrapolate them to other scales (including the reactor scale) and tests (SET, IET) remains difficult (see PREMIUM), but this is what should be done.

5.3.2.3. Data assimilation

Data assimilation (DA) is a discipline of applied mathematics that is well suited to combine mathematical dynamical models of a system with observational data. The model collects the past knowledge of the system, while the observations represent the injection of new information. DA is an inverse method aiming to estimate the state of the system, initial conditions for a numerical forecast model, model parameters, etc. For more information, see Annex D.2 of this document.

Data assimilation can naturally deal with a special type of inverse problem, that is when the predictive model is dynamical, i.e. describing the time evolution of a system. Dynamical models commonly use the concept of *state system*. The state can be defined as a set of variables (*state variables*) fully describing the condition of the system as a function of time. Dynamical models describe the time evolution of the system state via a "state equation". The response is calculated as a function of the state, and the observed response may be noisy.

DA methods have been developed in the nuclear safety field. Some of the inverse methods used in PREMIUM are based on DA (e.g. MCDA). Another example is CASUALIDAD, which is next described.

The so-called "predictive modelling methodology" is an approach to perform uncertainty analysis, based on inverse techniques, notably on the *Data Adjustment/Assimilation* (DAA) methodology, which combines experimental observations, code predictions and their respective errors to provide an improved estimate of the system state and of the associated uncertainty. The method considers all input parameters that affect the prediction (including model parameters).

In most problems of practical interest (in large-scale systems), the number of input parameters $\boldsymbol{\alpha}$ exceeds the number of responses **R**. In this case, adjoint sensitivity analysis procedure (ASAP) is the most efficient deterministic method for computing local sensitivities **S.** First, system critical points (bifurcations, turning points, saddle points, response extrema or cliff-edge effects) are determined using a deterministic global sensitivity method (global adjoint sensitivity analysis procedure, GASAP). Subsequently, the local sensitivities of the responses at critical points are analysed by the ASAP. The sensitivity matrix **S** of the responses **R** with respect to the parameters $\boldsymbol{\alpha}$ is used in order to obtain (via the moment propagation equation) the covariance matrix $\mathbf{C_R}$ of the responses starting from the covariance matrix $\mathbf{C}_{\boldsymbol{\alpha}}$ of the system parameters.

The DAA technique results in an improved estimate of the system state, based on a Bayesian inference process. The predicted results **R** and the corresponding statistical errors $\mathbf{C_R}$ are based on the prior input parameters (i.e. $\mathbf{C}_{\boldsymbol{\alpha}}$), and are combined with experimental observations **M** of the states of a system and associated uncertainty $\mathbf{C_M}$ to generate "adjusted" values for the system parameters ($\boldsymbol{\alpha}^{IE}$, IE= improved estimate values) and the respective input covariance matrix ($\mathbf{C}_{\boldsymbol{\alpha}}^{\mathbf{IE}}$, or 'posterior' PDF). From this improved estimate analysis of the system's states, the responses $\mathbf{R}^{IE}$ and the respective covariance matrix ($\mathbf{C_R^{IE}}$) are finally derived.

The CASUALIDAD method (*Code with the capability of Adjoint Sensitivity and Uncertainty AnaLysis by Internal Data ADjustment and assimilation*) (Petruzzi, 2008) is based on the previous techniques, and can be described as a fully deterministic method for performing, internally to the thermal-hydraulic system code, the sensitivity and the uncertainty analysis. The availability of a suitable database of experiments (SETF and/or ITF) and related qualified code calculations constitutes a prerequisite for the development

and the following application of the methodology. A full description of CASUALIDAD methodology is provided in Annex D.2.

Another DAA-based method is termed as model calibration through data assimilation (MCDA). This method has been developed by KAERI (Heo and Kim, 2015). The DA is based on a linear dependence between the SRQ and uncertain parameters. The chi-square linearity test is proposed for determining the degree of nonlinearity of the simulation output with respect to the parameters. MCDA thus makes the distinction between the linear and the non-linear situation. For a linear system, a deterministic approach is used to obtain the mean value and standard deviation of the parameters. The authors point out the correspondence between DA estimates and those obtained under a linear Bayesian framework where both the parameters and observables are assumed as Gaussian (see further the paragraph devoted to Bayesian methods). Otherwise, in order to address the non-linear responses in MCDA, a sampling approach is carried out to generate the posterior distribution of the parameters. This is conducted by the Markov Chain Monte Carlo (MCMC) sampling method and the Metropolis algorithm is used for a MCMC implementation. In Annex G of the present document, MCDA is applied to quantification of the uncertainty of the critical flow mode using Marviken tests.

5.3.2.4. Methods based on forward propagation and design of experiments

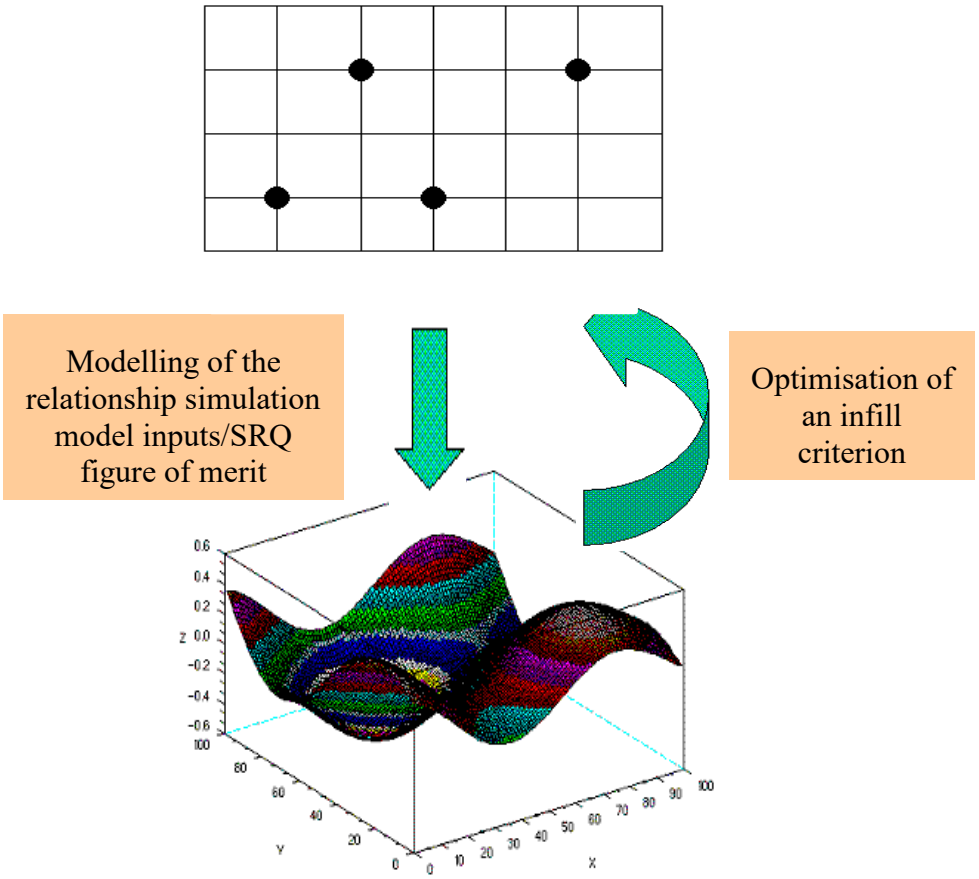
This type of method requires performing several simulations to fulfil specific requirements on SRQ figures of merit. In practice, these quantities are related to a coverage rate (in the DIPE and Tractebel methods for example), or to a maximum allowed deviation including the discrepancy between simulation (e.g. “reference” and “sensitivity” cases in the IPREM method) and experimental results. The requirements are translated through thresholds and are chosen by expert judgement even if some recommendations exist (e.g. 95% coverage rate).

When the figure of merit and threshold are specified, the key point of the mathematical treatment for this type of method is the choice of the simulations to perform which usually relies on the use of design of experiments. There exists a large literature on their construction. The objective of classical constructions is to explore the completely input space and to ensure non-redundant projection on input subspaces. Starting from the discretisation of each input parameter (also called levels), a straightforward construction consists in combining all possible levels, and it leads to full factorial designs. This approach becomes unaffordable when the number of input parameters and levels is large. A useful alternative is the fractional factorial design that allows reducing the number of evaluations (Box and Draper, 1987). Latin Hypercube Sampling (LHS) (McKay et al., 1979) is also very popular; it ensures that each of the input parameters has all the portions of its range represented.

When the method consists in identifying specific region of interest in the completely input parameter range to satisfy requirements on a figure of merit, the previous DoEs are likely to include a large set of input parameter values in order to explore the whole variation range and capture region boundaries. This is not always affordable in practice due to high computational cost. To circumvent this limitation, an adaptive strategy can be followed. Starting from an initial design, the objective is then to refine the design only in regions where the requirement is fulfilled. It leads to adaptive DoEs. Their construction (Figure 5.2) first relies on a modelling of the relationship (by a surrogate model) between simulation model inputs and a figure of merit from a DoE with few simulations, then on the optimisation of an infill criterion that provides the new simulations to perform according to

the type of regions of interest. This process is applied sequentially in order to update and improve the model at each iteration.

Figure 5.2. Sketch of the construction of adaptive DoEs



The modelling often relies on the use of Gaussian processes and kriging equations (Cressie, 1993). It provides an approximation of the figure of merit for input values that are not simulated. Then, this information is integrated in an infill criterion. A classical choice for this criterion is the weighted integrated mean square error (IMSE) criterion (Picheny et al., 2010) that provides a compromise between refinement in regions of interest (i.e. where the probability for the requirements to be fulfilled is high) and exploration of the variation range.

5.3.3. Comparison and recommendations: some practical guidelines on methods to use

Several types of inverse methods used in nuclear safety applications have been described in this chapter.

The two first categories are frequentist and Bayesian methods. They offer a theoretical framework to quantify input uncertainties. They allow including and modelling noise in the data. In many cases, the information about the noise is statistic. When the noise has a known probability distribution (e.g. Gaussian), these probabilistic methods can benefit from such

information. However, their construction relies on several assumptions related to uncertainty modelling and a careful check on the impact on the results should be performed to avoid any misleading interpretation.

It is also important to mention that Bayesian approaches provide regularised solutions to the inverse problem, while frequentist methods may not have such an advantage.

Frequentist and Bayesian methods allow both calibration and uncertainty quantification. If calibration is often performed during simulation model development, it is not recommended to recalibrate during the quantification according to the PREMIUM benchmark conclusions. The main reason is that if the experimental database is not representative enough, the recalibrated calculation might be very different from the reference one and as observed in PREMIUM (NEA, 2016; NEA, 2017), uncertainty results might not encompass the reference calculation. Moreover, if recalibration is performed, the update of the reference calculation requires going back to Element 4, which is assumed to be complete before uncertainty quantification in the SAPIUM framework.

The third category of method is more empirical and is based on forward propagation and fulfilment of requirements on specific SRQ figures of merit. By construction, contrary to the two previous ones, it does not rely on a mathematical formulation of the inverse problem but seeks to adjust input parameter uncertainty by combining simulations with an updating process in an iteration loop.

This category of methods strongly depends on the construction of the SRQ figures of merit and on the associated requirements. This last point involves expert judgement. However, these methods can be used in case of poor knowledge on input uncertainties that would prevent the validation of the underlying assumptions of frequentist and Bayesian approaches.

The choice of a method (category) to quantify input uncertainties depends on the problem under study. It should integrate the characteristics of each method and their relevance for the intended use. In Section 5.2.1.3, the main criteria for the selection of a method are given. The characteristics are solidity (rigorous theoretical basis), flexibility, transparency/reproducibility (well documented), reduction of user effect, relevance of the assumptions and tractability. It is therefore interesting to analyse the different categories of methods with respect to the main criteria (transparency is not addressed since it is related to documentation and not the method construction).

Solidity:

Frequentist and Bayesian have a solid mathematical basis, with clear assumptions. Their development/improvement is an active field of research. The third category of method is more empirical. The starting point of its construction is the choice of SRQ figures of merit as well as associated requirements, which are problem-dependent. However, once this step is achieved, it exploits the mathematical techniques of uncertainty propagation and design of experiments that are widely addressed in the scientific literature.

Flexibility:

Frequentist and Bayesian methods offer a generic framework to quantify input uncertainties, provided all the underlying assumptions associated to their construction are fulfilled. Therefore, they can be used to handle different types of problems. On the contrary, methods based on forward propagation and DoE might need to be adapted to each problem (SRQ figure of merit, thresholds).

The two first categories make it possible, for example, to include noise and bias. Even if noise can be technically integrated in the construction of a SRQ figure of merit, it is not clear in the current practice of the third category how to take into account the noise and bias components through a specific modelling.

Concerning the input uncertainty modelling, frequentist and Bayesian approaches are constructed in the probabilistic framework and allow quantification of input uncertainties represented as PDFs. Forward propagation/DoE one can also be applied to derive a PDF. However, in presence of poor knowledge, they usually lead to uncertainty intervals. A way of representing imperfectly known probability distributions are probability-boxes (p-boxes) (Ferson et al., 2003). They have been applied to estimate input and model uncertainty (He, 2019). So far in nuclear safety application, even if some strategies are introduced to partly address the treatment of aleatory and epistemic uncertainties, none of these categories can be used to combine different ways to model uncertainty in order to handle different types of uncertainty.

Reduction of user effect:

All categories include user effect.

A clear example is the prior distribution for the input parameters to estimate. This is a well-known feature of Bayesian methods. Likewise, in frequentist methods, parametric distributions may be assigned to the unknown inputs, which is also a type of prior information.

For the third category, expert judgement is required in the choice of SRQ figures of merit as well as associated requirements. This last choice is often performed according to the analyst's objectives (even if some recommendations are given in the case of the IPREM method, for example).

Relevance of assumptions:

Because they are more sophisticated, frequentist and Bayesian approaches rely on several theoretical assumptions that should be carefully checked in practice, especially when the information on uncertainties is poor. They mainly concern:

- The prior information: it is related to the input uncertainty modelling in frequentist methods (that is classically chosen as normal or lognormal) and to the choice of the prior distribution in Bayesian one.
- The connection between SRQ and input parameters, which can be linear or based on the identification of a surrogate model.

There are fewer assumptions in the application of the third category. However, special attention should be devoted to the choice of the requirements and its impact on the results.

Tractability:

Tractability is an important issue for frequentist and Bayesian methods. Both may require a huge, prohibitive number of model evaluations. This problem can be circumvented using surrogate models. It is important to choose and construct the surrogate models so that the results are not very distorted by their use. There is a growing number of references describing applications of probabilistic inverse methods and using surrogate models.

The third category of methods does not require as much computational effort as the frequentist and Bayesian ones because it usually does not seek to quantify a full PDF but

rather an interval. Numerical techniques such as surrogate modelling and DoE are, however, available to reduce the number of simulations.

5.3.4. Inclusion of experimental uncertainty in quantification

In the praxis of BEPU analyses, the experimental uncertainties are usually considered by the validation step of model uncertainty quantification as well as during validation of the code and qualification of the input data. Consideration of the uncertainty of experimental data by quantification of model uncertainties is not a standard and rather selective.

There are different types of experimental errors:

- Errors of measurement accuracy; they usually have a normal type distribution and are symmetric around the measured value.
- Errors due to the nonstable character of the measured quantity; oscillating character of the quantity due to the unstable nature of the phenomenon or due to difficulties in measurements, e.g. temperature measurements in two-phase flow, which can be measured in the liquid or gas phase.
- Systematic errors due to wrong calibration of the measurement device or positioning of the sensors. For example, the thermocouple measuring the surface temperature is usually placed in the wall and therefore does not measure exactly the surface temperature but the wall's temperature near the surface.
- Error of boundary and in the case of transient experiment initial conditions; they influence the quantification of model uncertainties indirectly through deviation/error of the calculation results, which are compared with experimental data.

Different types of experimental uncertainties are treated in different ways. However, the main question is when the experimental errors can and/or should be considered and in which way.

Consideration of measurement accuracy by the quantification of model uncertainties is sometimes performed, e.g. in CIRCÉ software (de Crécy and Bazin, 2004), (Wu et al., 2018). Consideration of the measurement error in these methods means that the variance of the difference between code prediction and experimental measurements is the sum of code uncertainty and experimental uncertainty. In this way consideration of the measurement error leads to calculation of smaller model uncertainty. This is inconsistent with principles of model uncertainties quantification, as recognised by the authors of the CIRCÉ. In the manual of the software, it is recommended not to consider accuracy of the measurements (NEA, 2016).

Theoretically, it could be possible to introduce the term representing the variance according to the measurement error as negative. It would be equivalent with an increase of simulation variance and conservative increase of the determined model uncertainty. However, it may easily lead to over-prediction of model uncertainty and excessive conservatism in the results. So, generally it can be recommended not to consider measurement accuracy.

In the case of systematic error, the situation is different. As far as the error is known, the measured results should be corrected by the error. The corrected experimental data can be used for comparison with calculated results.

The case of errors in boundary and initial conditions is the most problematic. Usually such an error is not known. Sometimes, it can be detected and estimated through analyses and

comparisons of calculations and experimental data. In the course evaluation of mass and energy balances for experiment and its simulation differences need to be analysed. Sometimes it can lead to detection of differences between the experiment and numerical simulation and in this way to the detection of inconsistency in declared boundary and/or initial conditions. For instance, in this way error concerning the axial power distribution in the FEBA experiment could be detected. In the experiment description, the power distribution was described as axial-symmetric whereas the analysis showed that in reality the power profile was top-skewed (Skorek, 2017).

However, detection of experimental errors is a difficult task and code calculations are performed without awareness of these errors. It has then a negative influence on the quantification of model uncertainties.

Problems, which occur by strongly oscillating measured data, mainly due to difficulties of comparison of measured and calculated quantities, can be solved to a large degree by the application of time averaging/smoothing procedures. Smoothing procedures are also applied in the case of strongly oscillating simulation results.

It is obvious that for the quantification of model uncertainties, accurate experimental data should be selected as far as possible. The main problem with the inaccuracy of experimental data is that with an increase in the experimental error, information carried by the experiment decreases. The critical point is when the accuracy of the experimental data is lower than the expected estimated accuracy of the physical model. In such a case, the information carried by the experiment does not contribute positively to the determination of model accuracy and such an experiment has to be rejected from the experimental database for uncertainty quantification.

5.4. Step 12: Combination of model input uncertainties if several quantifications are performed

This step is considered if several quantifications are performed. This is not always the case in practice since a first aggregation is achieved in Step 10. The main difference between Steps 10 and 12 is that the aggregation is applied on the information coming from Elements 2-3 in Step 10, i.e. input of the inverse method whereas it is related to quantified model input uncertainties in Step 12, i.e. output of the inverse method.

In practice, the different quantification results can be associated with different studies and the analyst might intend to take them into account. He/she can also be interested in performing several quantifications to clearly evaluate the impact of the information associated with a given SRQ or to a group of experiments (same scale). Moreover, as will become clear in the rest of the section, synthesising the output information is not restricted to building a summary of all information provided by different input uncertainty quantifications. It also allows performing further analysis, such as exhibiting the agreement or disagreement between the results. In case of strong conflict between quantified model input uncertainties, an iteration of the SAPIUM steps-elements is required. This last situation is hardly detected in the aggregation performed in Step 10.

It is common to distinguish three main kinds of synthesis behaviour:

- *Conjunctive*: equivalent to taking the intersection, conjunctive synthesis assumes the reliability of all quantifications, and allows exhibiting the conflict among the different results. It produces precise but potentially unreliable results in case of strong conflict.

- *Disjunctive*: equivalent to taking the union, disjunctive synthesis makes the conservative assumption that at least one quantification is reliable. It produces in general imprecise but reliable results.
- *Arithmetic* (weighted) mean: assumes independence between quantification results, and produces a result between disjunction and conjunction.

These three kinds of behaviour can then be followed through the application of fusion operators, whose construction depends on the input uncertainty modelling.

In the probabilistic framework, the most common approach is based on a weighted average (Cooke, 1991) to combine probability distributions. There exists a second approach, constructed in the Bayesian framework (Dubois et al., 2016; Genest and Zidek, 1986), which requires specifying prior information. There is no real counterpart to set intersections and unions in the probability field; therefore, conjunction and disjunction are difficult to define, formally, if *N* uncertainty quantifications lead to *N* cumulative distribution functions $F_1,...,F_N$ for a given input parameter; and if to each distribution is associated a weight w_i (this weight can be used to take into account the confidence of the analyst in each quantified input uncertainty) such that their sum is one, the arithmetic weighted mean reads

$$F_{mean}(t) = \sum_{i=1}^{N} w_i F_i(t) \qquad (5.13)$$

When the input uncertainty is summarised for each parameter by an interval, an alternative approach is to move to the possibility framework and to model the information provided by each quantification by a triangular possibility distribution:

$$\pi(t) = \begin{cases} \dfrac{t - LUB}{RV - LUB} & if\ t \in [LUB, RV] \\ \dfrac{t - UUB}{RV - UUB} & if\ t \in [RV, UUB] \end{cases} \qquad (5.14)$$

Where *[LUB,UUB]* denotes the uncertainty interval and *RV* is the parameter reference value.

If $\pi_1 ... \pi_N$ are the triangular functions constructed from the information provided by the N quantifications, three fusion operators (Dubois and Prade, 2001) can be constructed in the possibility framework. They are written:

- conjunction:

$$\pi_{\cap}(t) = min_{i=1,...,n}(\pi_i(t))$$

- disjunction:

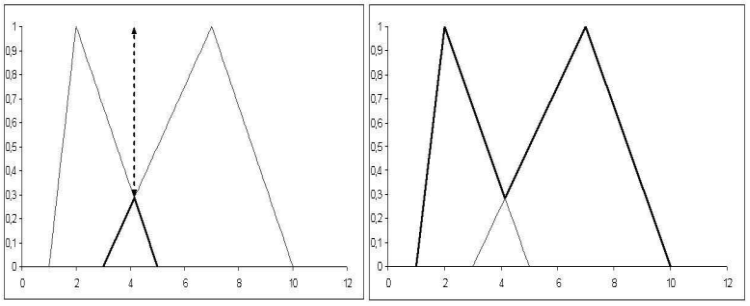
$$\pi_{\cup}(t) = max_{i=1,...,n}(\pi_i(t))$$

- arithmetic mean:

$$\pi_{mean}(t) = \frac{1}{N}\sum_{i=1}^{N}\pi_i(t)$$

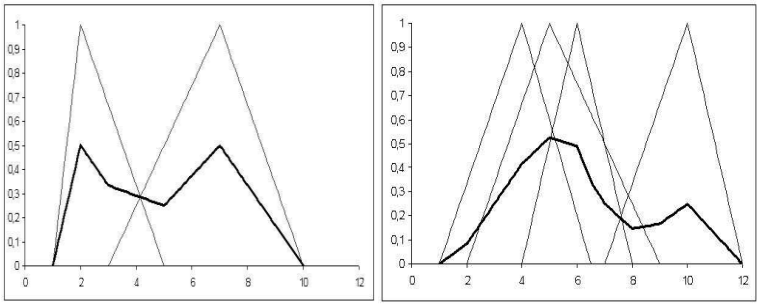
These three operators are illustrated in Figure 5.3 and Figure 5.4 (left) in the case of two quantification results for a given parameter.

Figure 5.3. Fusion of the information provided by two quantifications



Note: left, conjunctive operator; right, disjunctive one. The thin line stands for the possibility model representing each information whereas the thick one is the aggregated result. The dashed arrow represents the disagreement indicator.

Figure 5.4. Mean aggregation of the information provided by different quantifications



Note: left, two quantifications; right, an example with four quantifications. The thin line stands for the possibility model representing each information. The thick one is the aggregated result.

The conjunctive operator makes it possible to quantify the conflict among quantification results. More precisely, for each input parameter, a disagreement indicator can be defined as:

$DI^{\cap} = 1 - max_t \pi_{\cap}(t)$	(5.15)

The mean operator, by construction, tends to average the information. It is similar to the mean operator within the probabilistic framework. It is important to keep in mind that this strategy does not take into account agreement or disagreement between quantifications in

the result. Consider, for example, four quantifications giving the information depicted by Figure 5.4, right. Obviously, the intersection of the four triangles is an empty set due to a conflict between some quantification results. However, the analysis with the mean operator can be carried out and leads to an aggregated model exhibiting a reference value around five, which is outside the uncertainty interval provided by one quantification. Therefore, in order to perform a reliable fusion, it is recommended to combine several fusion operators to get a quantitative insight on the compatibility of the quantification results before combining them by averaging.

An example of application of the different fusion operators previously described can be found in (Destercke and Chojnacki, 2008) in the framework of the BEMUSE project.

5.5. Step 13: Confirmation of an input quantification: running the forward problem

Confirmation comes after input uncertainty quantification. It requires running a forward problem on the same experiment used for the quantification. More precisely, it includes the two following main steps:

- The propagation of the quantified input uncertainties through the simulation model and the estimation of the uncertainties associated with a set of SRQs.
- The comparison between SRQ uncertainties and experimental results using confirmation/validation indicators.

So, experimental data and simulation model calculations are used. Both need to be selected taking into account decisions already made when running the inverse method.

From a technical point of view, confirmation is similar to validation, which is described in Chapter 6, except that the experiments are those used for the quantification.

The first step exploits the set of input uncertainties (e.g. probability density functions) associated to some input parameters and estimated by running an inverse method. Such parameters, usually a limited number of them depending on the methodology, are considered as input uncertain parameters for running the usual forward case. In the subsequent runs, responses are calculated and their associated uncertainty bands are derived. This process is quite standard in nuclear applications. Refer to Chapter 6 (Section 6.3) for an overview of available methodologies to perform it.

The second step requires characterising and quantifying the agreement between simulation and experimental results. A straightforward strategy is to verify whether each experimental data is enveloped by the corresponding SRQ uncertainty band. In the confirmation step established by the CIRCÉ methodology, a high number of simulation model runs is carried out for the propagation step, generally 100 but sometimes higher (e.g. 1000) and order statistics are used to derive quantities such as the 2.5% and 97.5% quantiles associated to each SRQ. Then, a check is performed to see if 2.5% of the experimental values are smaller (resp. larger) than the 2.5% (resp. 97.5%) quantiles. The variation intervals defined by these two quantiles envelop 95% of the experimental responses, which explains the name of "envelop" calculations used for this check. This type of confirmation indicator as well as more complex indicators providing a deeper qualitative insight on the results are fully described in Section 6.4 of Chapter 6. The evaluation of the agreement between simulation and experimental results also requires the definition of a scale of acceptability that depends on the type of indicator. Refer to Section 6.5 for a short discussion on this topic.

A confirmation step has already been performed in the PREMIUM project on the FEBA experiment. It was found that the confirmation results mainly depend on:

- the specification of the problem (choice of SRQs) and the choice of input parameters;
- the quality of the nominal calculation: a good nominal calculation makes it easier for the uncertainty bands to envelop the experimental data;
- the input uncertainty quantification method.

All these topics are connected to the so-called user effect and are covered by the SAPIUM elements (Chapters 2, 4 and 5 respectively).

5.6. Recommendations and open issues

5.6.1. Recommendations

In modelling and simulation, inverse methods provide efficient tools to handle input uncertainty quantification. A first recommendation is to use, as much as possible, separate effects tests (SETs), where a single phenomenon is investigated and the uncertainty associated to the model of the phenomenon can be usually related to a singular measurement. In this case, the input model uncertainty can be estimated using a simple mathematical method through the comparison of the test simulation with the singular measurement from the SET (for a representative result). The quotient of the experimental result to the calculated one (whenever the latter is not zero) is determined. Then, the probability distribution of the quotients can be estimated. The same approach can be applied for quantification of model uncertainties using key model parameters. But, in this case, the relation between the input model uncertainty represented by the key parameter and model output compared with the measurement is implicit, and the determination of the correction factor requires an iteration procedure. In the past this was the most frequently and successfully used method of quantification.

If there is more than one phenomenon (and model) of importance involved in the available experiment, the application of advanced methods is necessary.

Inverse methods can also be used to calibrate a model (simultaneously or not with quantification). The experience in the PREMIUM project (NEA, 2016; NEA, 2017) indicates that the IUQ should be performed without recalibration. This is especially advisable in case of a lack of adequacy of the available experimental database. Therefore, recalibration is not coherent with the definition of best-estimate code and it is not recommended in SAPIUM. More investigations are required if the analyst intends to integrate such a step in the quantification.

A combination of the information from different experiments or databases should be performed before applying an inverse method. It is clear that a connection to the adequacy analysis performed in Chapter 3 should be established to define the contributions of each experiment to the result. Probabilistic inverse methods use the likelihood function in their inferences, and one possibility of taking into account different degrees of importance is to construct a weighted likelihood. The least-square technique, used in the solution of inverse problems, can also assign different weights in the construction of the sum of squares to be minimised. Sensitivity analysis to the weights assigned could be a useful tool.

The assumptions associated with the mathematical methods to solve the inverse problem should be clearly taken into account to evaluate the impact of the analyst's choices on the results.

Depending on the problem to solve, users may choose between statistical (frequentist or Bayesian) methods, or approaches based on forward propagation and DoE. Bayesian methods produce, in general, regularised solutions to the IP, while frequentist methods may produce non-regularised solutions. This fact can be considered when choosing between the two types of methods. The two first categories offer a generic framework that allows taking into account noise and bias in the problem.

Differences between frequentist and Bayesian methods have been explained throughout this chapter. Both of them allow the introduction of a priori information. In Bayesian methods, such information is introduced in the form of the prior distribution of the input, and the methods calculate the posterior. In frequentist methods, the a priori information is introduced as an assumption on the parametric family for the probability distribution of the input, and the methods estimate the parameters of the distribution. A major difference between the two types of methods is that Bayesian methods produce regularised (i.e. well-posed) solutions to the inverse problem, while frequentist methods correspond to non-regularised solutions.

More generally, the choice of a method to quantify input uncertainties depends on the problem under study. The main characteristics of the method to take into account are solidity (rigorous theoretical basis), flexibility, transparency/reproducibility (well documented), reduction of user effect, relevance of the assumptions and tractability.

Most methods described in this chapter are constructed in the probabilistic framework. The estimation of a probability distribution for each uncertain input parameter is not always affordable in practice due to an incomplete state of knowledge or difficulties validating the whole set of underlying assumptions associated with each method. In this case, an alternative uncertainty model could be used. For instance, the third category of methods (which is more empirical but is rooted in fewer assumptions than in the others) can be applied to derive uncertainty intervals.

The influence of additional uncertainty sources (e.g. related to the numerical approximation and especially to the choice of the nodalisation) is of prime importance to derive reliable input uncertainties.

5.6.2. Open issues

Techniques to assign different weights to experimental databases and to solve inverse problems using such weights should be studied further.

Concerning the choice of a mathematical model to represent uncertainties in agreement with the state of knowledge, although several works have been already proposed to treat the direct problem (see Section 6.3.2 of Chapter 6), the adaptation of frequentist or Bayesian inverse methods to alternative theory remains a challenge for nuclear applications. In particular, further investigations are required to tackle the problem of the treatment of epistemic uncertainty by alternative theories in the framework of IUQ.

Exercises of comparison and benchmarks of inverse methodologies (as PREMIUM) must be encouraged. User effect should be an important issue in these exercises. Inverse methods aim to replace engineering judgement in model uncertainty quantification and other IPs.

Nevertheless, inverse methods also have a user effect. This fact was evident in the results of the PREMIUM benchmark, and was one of the reasons for initiating SAPIUM activity.

It is important to remember that inverse problems, in general, do not have unique solutions, and additional information must be added in order to select one of the possible solutions. Users can differ in the ways they introduce given prior information to the IP, thus obtaining different results.

A main objective of the SAPIUM activity is definitely to provide the inverse methods user with guidelines aimed at minimising the user effect.

References

Bachoc, F., G. Bois, J. Garnier and J.-M. Martinez (2014), "Calibration and improved prediction of computer models by universal Kriging", in *Nuclear Science and Engineering*, 176(1), pp. 81-97.

Barbillon, P., G. Celeux, A. Grimaud, Y. Lefebvre and E. De Rocquigny (2011). "Non linear methods for inverse statistical problems". In *Computational Statistics & Data Analysis*, 55, pp. 132-142.

Bates, D.M, and Watts, D.G. (1988), *Nonlinear Regression Analysis and Its Applications*, J. Wiley & Sons.

Box, G.E.P and Draper, N.R. (1987), *Empirical model-building and response surfaces*, Wiley Series in Probability and Mathematical Statistics.

Brynjarsdóttir, J. and O'Hagan, A. (2014), "Learning about physical parameters: the importance of model discrepancy", in *Inverse Problems* 30 (2014) 114007.

Capote, R., D.L. Smith, A. Trkov and M. Meghzifene (2012), "A New Formulation of the Unified Monte Carlo Approach (UMC-B) and Cross-Section Evaluation for the Dosimetry Reaction ^{55}Mn(n,γ)^{56}Mn", in *Journal of ASTM International* 9, JAI104115, available online at www.astm.org.

Castro, E., C. Ahnert, O. Buss, N. García-Herranz, A. Hoefer and D. Porsch (2016), "Improving PWR core simulations by Monte Carlo uncertainty analysis and Bayesian inference", *arXiv*: 1601.08036v1 [nucl-th].

Celeux, G., A. Grimaud, Y. Lefèbvre and E. De Rocquigny (2010), "Identifying intrinsic variability in multivariate systems through linearized inverse methods", in *Inverse Problems in Science and Engineering*, 18, pp. 401-415.

Chadwick, M.B., P. Talou and T. Kawano (2005), "Reducing Uncertainty in Nuclear Data", in *Los Alamos Science*, No 29.

Chavent, G. (2009), *Nonlinear Least Squares for Inverse Problems Theoretical Foundations and Step-by-Step Guide for Applications*, Springer.

Cooke, R. (1991), *Experts in uncertainty*. Oxford, UK: Oxford University Press.

Cressie, N.A. (1993), *Statistics for Spatial Data*, Wiley Series in Probability and Mathematical Statistics.

Damblin, G and Gaillard, P (2018), "A Bayesian framework for quantifying the uncertainty of physical models integrated into thermal-hydraulic computer codes", in *Proc. of BEPU international conference*, Lucca, Italy.

Dasht, M. and Stuart, A.M. (2017), "The Bayesian Approach to Inverse Problems", chapter 10 of book R. Ghanem et al (eds), in *Handbook of Uncertainty Quantification*, Springer.

de Crécy, A. (1997), "Circé: a tool for calculating the uncertainties of the constitutive relationships of CATHARE 2", in *Proc. of the 8th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics (NURETH-8).*

de Crécy, A. and Bazin, P. (2004), "Quantification of the Uncertainties of the Physical Models of CATHARE", in *Proc. of BE2004*, Washington D. C., United States.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm", in *Journal of the Royal Statistical Society*, Series B (Methodological), 39(1), pp.1-38.

Destercke, S. and Chojnacki, E. (2008), "Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes", in *Nuclear Engineering and Design*, 238, pp. 2484-2493.

Dubois, D. and Prade, H. (2001), *Possibility theory in information fusion*, in G. D. Riccia, H. Lenz, and R. Kruse (Eds.), *Data fusion and Perception, Vol. CISM Courses and Lectures N 431, 53–76*. Berlin: Springer Verlag.

Dubois, D., W. Liu, J. Ma and H. Prade (2016), "The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks", in *Information fusion*, 32, pp. 12-39.

Ferson, S., V. Kreinovick, L. Ginzburg and F. Sentz (2003), *Constructing probability boxes and Dempster-Shafer structures*, SAND2002-4015, Sandia National Laboratories.

Freixa, J., E. De Alfonso and F. Reventós (2016), "Testing methodologies for quantifying physical models uncertainties. A comparative exercise using CIRCÉ and IPREM (FFTBM)", in *Nuclear Engineering and Design*, 305, pp. 653-665.

Fu, S., G. Celeux, N. Bousquet and M. Couplet (2015). "Bayesian inference for inverse problems occurring in uncertainty analysis", in *International Journal for Uncertainty Quantification*, 5(1), pp. 73-98.

Fugleberg, E.N. (2016), *Uncertainty Quantification and Calibration in Nuclear Safety Codes Using Gaussian Process Active Learning*, Master Thesis, Massachusetts Institute of Technology.

Genest, C. and Zidek, J.V. (1986), "Combining probability distributions: a critique and an annoted bibliography", in *Statistical Science*, 1(1), pp. 114-148.

Ghosh, J.K., M. Delampady and T. Samanta (2006), *An introduction to Bayesian analysis: Theory and Methods*, Springer.

Hadamard, J. (1902), "Sur les problèmes aux dérivées partielles et leur signification physique", in *Princeton University Bulletin*, 13, pp. 49-52.

He, Q. (2019), "Model validation based on probability boxes under mixed uncertainties", in *Advances in Mechanical Engineering*, 11(5), pp. 1-9.

Heo, J. and Kim, K.D. (2015), "PAPIRUS, a parallel computing framework for sensitivity analysis, uncertainty propagation, and estimation of parameter distribution", in *Nuclear Engineering and Design*, 292, pp. 237-247.

Higdon, D., M. Kennedy, J.C. Cavendisch, J.A. Cafeo and R.D. Ryne (2005), "Combining field data and computer simulations for calibration and prediction", in *SIAM Journal on Scientific Computing*, 26(2), pp. 448-466.

Hoefer, A., O. Buss, M. Hennebach, M. Schmid and D. Porsch (2014), "MOCABA: general Monte Carlo-Bayes procedure for improved predictions of integral functions of nuclear data", *arXiv:* 1411.3172v1.

Hoefer, A., O. Buss and M. Schmid (2018), "Applications of Multivariate Normal Bayesian models in Nuclear Engineering", in *Proc. of the ANS Best Estimate Plus Uncertainty International Conference*, Lucca, Italy.

Idier, J. (Ed) (2008), *Bayesian Approach to Inverse Problems*, ISTE – Wiley.

Kaipio, J. and Somersalo, E. (2007), "Statistical inverse problems: Discretization, model reduction and inverse crimes", in *Journal of Computational and Applied Mathematics* 198, pp. 493-504.

Kass, R.E. and Wasserman, L.W. (1996), "The selection of priors distributions by formal rules", in *Journal of the American Statistical Association*, 91(435), pp.1343-1370.

Kennedy, M. and O'Hagan, A. (2001), "Bayesian calibration of computer models (with discussion)", in *J.R. Statist. Soc. B*, 63, pp. 425-464.

Kuhn, E. (2003), *Maximum likelihood estimation in non linear inverse problems*. PhD Thesis, Paris Orsay.

Kullback, S. and Leibler, R.A. (1951), "On Information and Sufficiency", in *Annals of Mathematical Statistics*, 22 (1), pp. 79-86. MR 39968, DOI:10.1214/aoms/1177729694.

Liu, C. and Rubin, D.B. (1994), "The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence", in *Biometrika*, 81(4), pp. 633-648, DOI: 10.1093/biomet/81.4.633.

McKay, M.D., R.J. Beckman and W.J. Conover (1979), "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code", in *Technometrics*, 21, pp. 239-245.

Marqués de Sá, J.P., L.M.A. Silva, J.M.F. Santos and L.A. Alexandre (2013), *Minimum Error Entropy Classification*. Studies in Computational Intelligence 420, Springer-Verlag Berlin Heidelberg, DOI: 10.1007/978-3-642-29029-9.

Nagel, J.B. (2017), *Bayesian techniques for inverse uncertainty quantification*. PhD Thesis, ETH Zürich, DISS. ETH NO. 24082.

Nguyen, T.N. and Downar, T.J. (2017), "Surrogate-based multi-experiment calibration of the BISON fission gas behaviour model", in *Nuclear Engineering and Design*, 320, pp. 409-417.

NEA (2017), "Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark, Final Report", NEA/CSNI/R(2016)18, OECD Publishing, Paris, www.oecd-nea.org/nsd/docs/2016/csni-r2016-18.pdf

NEA (2016), "PREMIUM report: Methodologies and data review", NEA/CSNI/R(2016)9, OECD Publishing, Paris, www.oecd-nea.org/nsd/docs/2016/csni-r2016-9.pdf

NEA (2011), "BEMUSE Phase VI Report. Status Report on the area, classification of the methods, conclusions and recommendations", NEA/CSNI/R(2011)4, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19020.

NEA (1998), "Report on the Uncertainty Methods Study", Vol.1, NEA/CSNI/R(97)/35, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_16220.

Pastore, G., L.P. Swiler, J.D. Hales, S.R. Novascone, D.M. Perez, B.M. Perez, B.W. Spencer, L. Luzzi, P. Van Uffelen and R.L. Williamson (2015), "Uncertainty and sensitivity analysis of fission gas behaviour in engineering-scale fuel modeling", in *Journal of Nuclear Materials*, 456, pp. 398-408.

Perrin, F., B. Sudret, G. Blatman and M. Pendola (2007). "Use of polynomial chaos expansions and maximum likelihood estimation for probabilistic inverse problems", in *Proc of the 18e Congrès Français de Mécanique*.

Petruzzi, A. (2008), *Development and application of Methodologies for Sensitivity Analysis and Uncertainty Evaluation of the Results of the Best Estimate System Codes applied in Nuclear Technology*, PhD Thesis, Università di Pisa.

Picheny, V., D. Ginsbourger, O. Roustant, R.T. Haftka and H.-H. Kim (2010), "Adaptive designs of experiments for accurate approximation of a target region", in *Journal of Mechanical Design*, 132(7).

Rachdi, N. (2011). *Apprentissage Statistique et Computer Experiments - Approche quantitative du risque et des incertitudes en modélisation*. PhD Thesis, University of Toulouse (in French).

Rochman, D., E. Bauge, A. Vasiliev and H. Ferroukhi (2017), "Correlation $\overline{\nu_p} - \sigma - \chi$ in the fast neutron range via integral information", in *EP Journal of Nuclear Science and Technology*. 3, 14. https://doi.org/10.1051/epjn/2017009.

Rochman, D., E. Bauge, A. Vasiliev, H. Ferroukhi, S. Pelloni, A.J. Koning and J.Ch. Sublet (2018), "Monte Carlo nuclear data adjustment via integral information", in the *European. Phyicsal Journal Plus,* 133, 537.

Shrestha, R. and Kozlowski, T. (2016), "Inverse uncertainty quantification of input model parameters for thermal-hydraulics simulations using expectation–maximization under Bayesian framework", in *Journal of Applied Statistics*, 43(6), pp. 1011-1026.

Skorek, T. (2017), "Input uncertainties in uncertainty analyses of system codes: Quantification of physical model uncertainties on the basis of CET (combined effect tests)", in *Nuclear Engineering and Design*, 321, pp. 301-317.

Skorek, T. (2004), "Determination of Input Uncertainties of Uncertainty and Sensitivity Analyses", *Proc. Of PSAM 7 – ESREL'04,* (Eds. C. Spitzer, U. Schmocker, V.N. Dang), Vol.4, pp. 1803-1808, Berlin, Springer.

Somersalo, E. (2004), *Statistical Inverse Problems, Model Reduction and Inverse Crimes*, Firenze.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.

Wu, X. (2017), *Metamodel- Based Inverse Uncertainty Quantification of Nuclear Reactor Simulators under the Bayesian Framework*, PhD Dissertation, University of Illinois at Urbana-Champaign.

Wu, X. and Kozlowski, T. (2017), "Inverse uncertainty quantification of reactor simulations under the Bayesian framework using surrogate models constructed by polynomial chaos expansion", in *Nuclear Engineering and Design,* 313, pp. 29-52.

Wu, X., T. Mui, G. Hu, H. Meidani and T. Kozlowski (2017), "Inverse uncertainty quantification of TRACE physical model parameters using sparse gird stochastic collocation surrogate model", in *Nuclear Engineering and Design,* 319, pp. 185-200.

Wu, X., T. Kozlowski, H. Meidani and K. Shirvan (2018a), "Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian Process, Part 1: Theory", in *Nuclear Engineering and Design,* 335, pp. 339-355.

Wu, X., T. Kozlowski, H. Meidani and K. Shirvan (2018b), "Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian Process", in *Nuclear Engineering and Design,* 335, pp. 417-431.

Wu, X., T. Kozlowski and K. Shirvan (2018c), "Inverse Uncertainty Quantification Using the modular Bayesian approach in the Presence of Model Discrepancy", in *Proc. of BEPU2018,* Lucca, Italy.

Wu, X., T. Kozlowski and H. Meidani (2018d), "Kriging-based inverse uncertainty quantification of nuclear fuel performance code BISON fission gas release model using time series measurement data", in *Reliability Engineering and System Safety,* 169 pp. 422-436.

Wu, X., K. Shirvan and T. Kozlowski (2019), "Demonstration of the Relationship between Sensitivity and Identifiability for Inverse Uncertainty Quantification", in *Journal of Computational Physics*, 396, pp. 12-30.

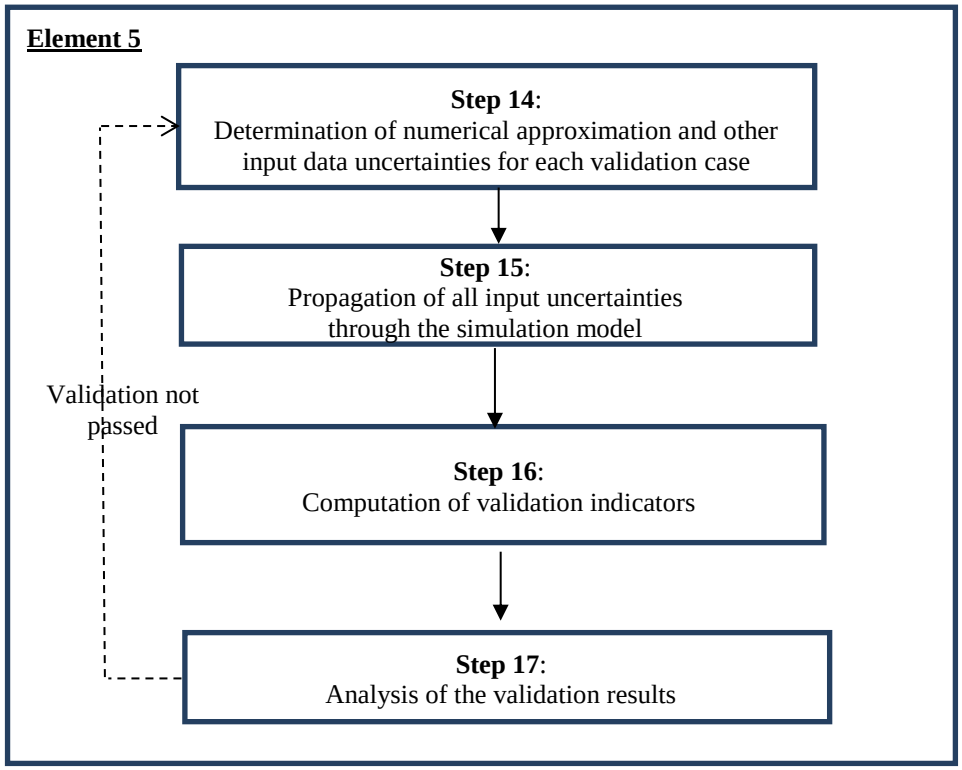
Yurko, J.P. (2014), *Uncertainty Quantification in Safety Codes Using a Bayesian Approach with Data from Separate and Integral Effect Tests*, PhD Thesis, Massachusetts Institute of Technology.

6. Model input uncertainty validation (Element 5)

6.1. Introduction

This chapter is devoted to the last element of the SAPIUM approach, which is performed following the four steps displayed by Figure 6.1.

Figure 6.1. Steps of Element 5



In the case of input uncertainties, the validation cannot be done in the input space since the comparison of the results with experimental data is not possible. It is performed in the SRQ space after input uncertainty propagation through the simulation model. Therefore, as suggested by Step 14, it is mandatory to combine the quantified model input uncertainties with other sources of uncertainty (e.g. model form and numerical approximation) because the validation result is also affected by them. However, since the SAPIUM activity is focused on model input uncertainty, the quantification of the other types of uncertainty are assumed to have been performed in another part of the VVUQ process and their contributions on the total uncertainty is not studied in this document.

It is important to keep in mind that the four previous steps are restricted to the experimental domain (also called validation domain in SAPIUM). For a full validation process for some industrial applications (see AIAA Guide [AIAA, 1998] for example), these steps are not enough for a reactor study in the nuclear field. In this case, following the VVUQ formal

procedure (Oberkampf and Barone, 2006; Ferson et al., 2008) recalled in Chapter 1, the validation process should be followed by:

- A prediction that exploits the comparison in the experimental domain and includes additional uncertainty estimation resulting from interpolation and extrapolation beyond the existing database to satisfy the intended use.
- An acceptability check to evaluate if the input uncertainties are suitable for the intended use in the application domain.

The first point is related to the predictive capability assessment that is discussed in Chapters 7 and 8. The acceptability check depends on the acceptability criteria for IUQ validation. In particular, it involves the choice of acceptability thresholds that are related to the problem of interest and to the risk tolerance of the decision maker. Therefore, this process cannot be formalised and is not addressed precisely in this methodological document. However, this chapter includes a short discussion on the analysis of the validation results and its integration in a loop approach, steps that are compulsory before focusing on the predictive capability.

The focus in this chapter is therefore on the technical treatment of Steps 15, 16 and 17.

6.2. Validation experiments

The validation process first requires the availability of a set of experiments that will be used in the comparison. Two situations can be encountered in practice.

When the experimental database constructed in Element 2 (Chapter 3) is large, validation experiments should correspond to experiments, which are not used for the uncertainty quantification (Chapter 5).

When the number of experiments is not sufficient to split the database, it is recommended to perform a leave-one-out cross-validation (Wackernagel, 1998). This process consists in iterating for each experiment of the database the following steps:

- remove one experiment from the database;
- quantify the model input uncertainties (Chapter 5) using the remaining experiments;
- validate on the removed experiment.

It is important to keep in mind that removing one experiment can artificially create a situation where extrapolation is required, i.e. the removed experiment is outside the convex hull of the remaining ones, or where the remaining subset of experiments has a low adequacy. It is therefore advisable to keep the same validation domain and to perform the cross-validation on adequate subsets of experiments.

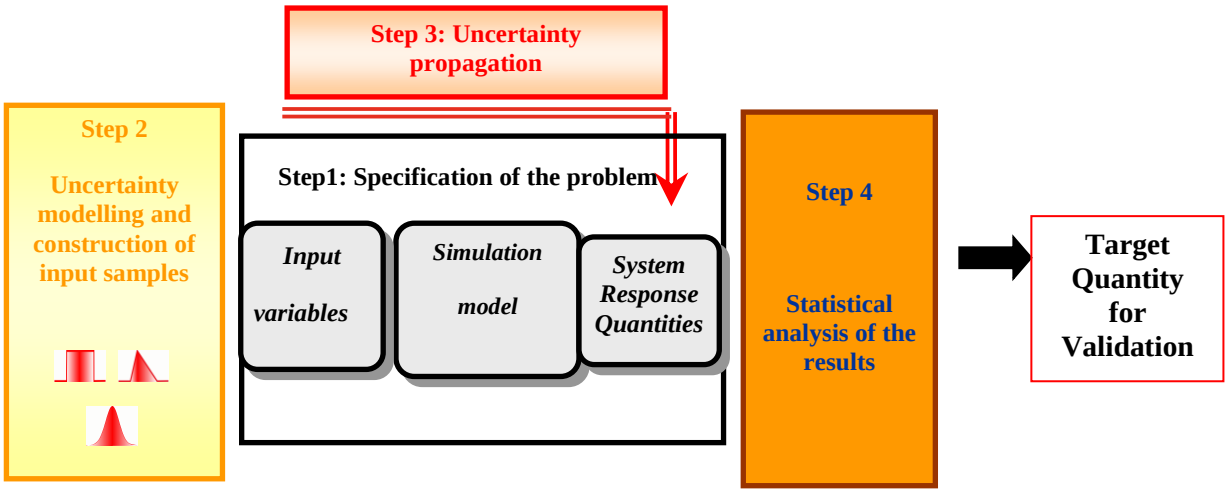
There is a main difference between the two previous situations. Concerning the first one, most of the validation process (except the iteration quantification/validation in a loop-approach) can be performed independently of Element 4 provided input uncertainties have been quantified. The outcome is either the acceptability of the quantified input uncertainties or not. On the contrary, in the second situation, the user needs to go through the previous SAPIUM elements to quantify the model input uncertainties for each subset of experiments considered during the cross-validation. The straightforward outcome is therefore the "acceptability" of the choices adopted in the SAPIUM approach and consequently the acceptability of the input uncertainties integrating model input uncertainties quantified from all the experiments of the database.

However, for both situations the starting point is the experimental database obtained in Element 2 from available traditional experiments. This type of experiment was usually performed to improve the understanding of physical phenomena or the models implemented in the computer code. Therefore, they could not be fully appropriate for validation – refer to the V&V literature (Oberkampf and Trucano, 2007) for guidelines to design new experiments providing high-quality data.

6.3. Step 15: Uncertainty analysis based on input propagation

Among the available uncertainty analysis methods, the input uncertainty propagation approach is widely used in industrial applications. In this type of method, the input uncertainties are propagated to the simulation model output uncertainties taking into account the code calculations with sampled input data coming from a previous mathematical modelling of input uncertainties (see Figure 6.2).

Figure 6.2. The four steps of an input uncertainty propagation method



In the first step, all relevant SRQs for input uncertainties validation have to be listed. Obviously, they have to be taken in the set of SRQs identified in Element 1 of the SAPIUM approach (Chapter 2). The next section provides a review of several approaches to handle Steps 2 to 4.

6.3.1. Review on input uncertainty propagation methods

6.3.1.1. Focus on Monte Carlo-based approach

The input uncertainty propagation method combining probabilistic modelling and Monte Carlo simulations is very popular in nuclear safety (NEA, 2011). In this method, the simulation model is treated as a "black box", and the input uncertainties are propagated to the SRQ uncertainties via several simulation model runs. The construction of this type of method is briefly recalled in the sequel.

Uncertainty modelling and construction of input samples (Step 2)

The uncertainty of each uncertain input parameter is quantified by a probability density function (PDF). For model input parameter, this function is given by the model input uncertainty quantification performed in Element 4 (Chapter 5) when the method allows

deriving a PDF. In other cases (e.g. if the method leads to an uncertainty interval), it is important to select an appropriate one according to the information on uncertainty. If only lower and upper uncertainty bounds (LUB and UUB) are available, a uniform distribution on [LUB,UUB] can be used. If a reference value within the interval needs to be taken into account, histogram or triangular distributions are recommended.

Moreover, if the dependence between uncertain input parameters is known and judged to be potentially important, it needs to be quantified. Correlation coefficients or copulas are, for example, available for dependence quantification. The input sample is finally constructed by drawing input parameter values simultaneously and randomly (a simple random sampling approach is recommended if an order statistics method is used for determining the 95%/95% estimates of the SRQs) from the probability distributions and dependence chosen in the modelling.

Uncertainty propagation through the simulation model (Step 3)

The simulation model is run repeatedly, each time using different values from the input sample. The results of a Monte Carlo simulation lead to a sample of the same size for each SRQ.

Statistical analysis of the results (Step 4)

The SRQ sample is used to get any typical statistics of interest. In the framework of validation, the choice of these statistics should be driven by the Target Quantity for Validation (TQV). Recall in the sequel the most classical statistics that will be used in Section 3 to define the TQV.

In thermal-hydraulic applications, quantiles associated with a chosen level of confidence are usually key quantities (e.g. 95%-quantile) since they can be exploited to check the respect of safety criteria. They can also be used to define an SRQ uncertainty interval (also called tolerance interval). For sake of generality, this type of interval is always denoted [LUB,UUB]. In many cases, with a "direct" Monte Carlo approach, due to the computational cost, the number of runs of the simulation model is not substantial enough (e.g. > 1000) to precisely estimate the quantiles of interest. A popular way to get information on quantiles with limited samples is to use order statistics (Conover, 1999).

The principle of order statistics is to derive results from the ranked values of a sample. If $(Y^1,\ldots,Y^N)$ denotes a sample of any random variable, Y, and $(Y^{(1)},\ldots,Y^{(N)})$ the corresponding ranked one, order statistics first provides an estimation of the quantile of interest since the α-quantile can be estimated by $Y^{(\alpha N)}$. Moreover, it turns out that the cumulative distribution function (CDF) of $Y^{(k)}$, $F_Y(Y^{(k)})$, follows the Beta law $\beta(k,N-k+1)$, which does not depend on the distribution of Y. This key result allows quantifying the probability that any ranked value is smaller than any quantile by the following formula:

$$P(Y^{(k)} <= Y\alpha) = F_{\beta(k,N-k+1)}(\alpha)$$

where $F_{\beta(k,N-k+1)}$ denotes the CDF of the Beta law $\beta(k,N-k+1)$.

The previous equation can then be used to derive:

1. Lower and upper bounds of a quantile of interest (α), given the sample size N and the confidence level δ that controls the probability that $Y^{(k)} <= Y\alpha$. It requires solving the equation $F_{\beta(k,N-k+1)}(\alpha)=\delta$ or $F_{\beta(k,N-k+1)}(\alpha)=1-\delta$.
2. The minimal sample size (and therefore the minimal number of simulation model runs) to perform in order to obtain an upper bound of a given quantile with a given

confidence level. It leads to the so-called Wilk's formula (Wilks, 1941) (and by extension to the Wilk's method also called GRS-type approach (Hofer et al., 1985)):

$$N = \ln(1-\delta)/\ln(\alpha)$$

For a quantile of 95% at a confidence level of 95% (i.e. $\alpha=\delta = 95\%$), the minimal sample size is N=59. This has been adopted in an USNRC approved best-estimate methodology for large-break LOCA (Martin and O'Dell, 2005).

In "Statistical aspects of best estimate method-I" (Guba et al., 2003), a more general form of the Wilk's formula for multiple output parameters is provided in the following equation:

$$\alpha = \sum_{j=0}^{N-p} \frac{N!}{(N-j)!\,j!} \delta^j (1-\delta)^{N-j}$$

Where p is the number of output parameters. By substituting $\alpha=\delta = 95\%$, and p = 3, the number of computer runs, N is found to equal 124. This has been adopted for three SRQs (PCT, ECR and CWO) in an USNRC approved best-estimate methodology for large-break LOCA (Frepoli, 2008).

The use of order statistics does not require any assumption on the type of the distribution of the random variable associated to each SRQ. In addition, the number of code runs is independent of the number of the selected input uncertain parameters, but only depending on the (tolerance limit) quantile and on the desired confidence level. Moreover, this method is very simple to implement and relies on actual simulation model results without fitted response surfaces or other approximations like goodness of fit tests. All these advantages make it widely used for licensing applications to nuclear safety analyses (Martin and Nutt, 2011).

However, it has been shown that this method based on minimal sample size may lead to rather conservative results and variability (e.g. outliers). One way to improve the method is to increase the number of code calculations, and take higher ranks as the estimators for the given probability content (quantile) and confidence level, or to use the "direct" Monte Carlo method (NEA, 2011; Frepoli and Iyengar, 2011; Shockling, 2015; Lee et al., 2014; Zhang et al., 2016). However, this may be limited by the requested large calculation efforts in case of complex coupled code systems.

6.3.1.2. An overview of probabilistic methods for the estimation of quantiles

This section gives some information about alternatives to "direct" Monte Carlo (cf. previous section) to establish a more exhaustive panorama of the available methods for the estimation of a quantile of a certain SRQ. Indeed, empirical quantiles converge very slowly towards the true ones in the case of simple random sampling (crude Monte Carlo) and the Wilks' method provides generally very conservative estimates when the number of simulations is relatively small (which is an issue for validation). It is worth noting that, contrary to the "direct" Monte Carlo approach, the methods mentioned hereafter generally suffer from the curse of dimensionality, and are therefore not suitable for an overly large number of uncertain input parameters (several dozens, typically).

The use of a surrogate model is a possible way to circumvent the issue of a limited computational budget. A surrogate model is an approximation of the considered simulation model that can be evaluated a much greater number of times than the latter (because its computational cost is very low). Thus, a surrogate model can enable the use of computationally high-demanding methods such as the simple random sampling. A surrogate model has to give an accurate prediction of any numerical simulation result in a

given domain of the uncertain inputs. In the case of a black-box simulation model, such a surrogate model can be obtained by methods of interpolation or of statistical learning (e.g. Regression) from some preceding runs of the simulation model (corresponding to a predetermined or a sequential design of experiments).

A vast statistical literature is available covering many methods and building and validating surrogate models. Some popular approaches are kriging, also known as Gaussian process modelling/regression, polynomial chaos and artificial neural networks, among others.

If the discrepancy between the surrogate model predictions and the simulation results have been estimated as negligible, the "direct" Monte Carlo methods can be applied to the surrogate model in the place of the simulation model. If the discrepancy is not negligible, the surrogate model can be employed nevertheless to outperform the crude Monte Carlo by different techniques (Cannamela et al., 2008): controlled importance sampling, control variate or controlled stratification. Similarly to crude Monte Carlo, asymptotic confidence intervals for the quantile estimates are then available (only the Wilks' method provides an exact confidence interval). Kriging, a particular surrogate model-based method, consists in computing a Bayesian probabilistic posterior distribution which represents the uncertainty about the true (unknown) simulated response, which makes it possible to develop iterative sampling strategies to estimate quantiles (Bayesian optimisation approach); see "Sequential design of experiments for estimating quantiles of black-box functions" (Lapobin-Richard and Picheny, 2017).

Free-surrogate model alternatives to crude Monte Carlo also exist, e.g. importance sampling or multilevel splitting methods (Pastel, 2012). In particular, an asymptotic confidence interval for a multilevel splitting quantile estimate is established in (Walter, 2015).

Table 6.1 summarises the main characteristics of the previously introduced statistical methods for quantile estimation.

Table 6.1. The main characteristics of the statistical methods for quantile estimation

Method	**Possibility to use a coarse metamodel***	**Curse of dimension**	**Confidence interval for the estimator**	**Main references**
Crude Monte Carlo	No	no	asymptotic	-
Wilks	No	no	exact	-
Importance sampling	Yes	no	asymptotic	(Cannamela et al., 2008) (Pastel, 2012)
Control variate	yes	yes	asymptotic	(Cannamela et al., 2008)
Controlled stratification	yes	yes	asymptotic	(Cannamela et al., 2008)
Multilevel splitting	no	no	asymptotic	(Walter, 2015)
Bayesian optimization	yes	yes	no	(Lapobin-Richard and Picheny, 2017)

* whether the use of a surrogate model associated to non-negligible errors is possible or not. In the case of Bayesian optimisation, the surrogate model is improved during the iterations. Surrogate model-based methods suffer from the curse of dimension.

Due to a lack of experience feedback, it does not appear feasible to provide precise guidelines for an efficient application of these methods in the context of system thermal-hydraulic numerical models for nuclear safety.

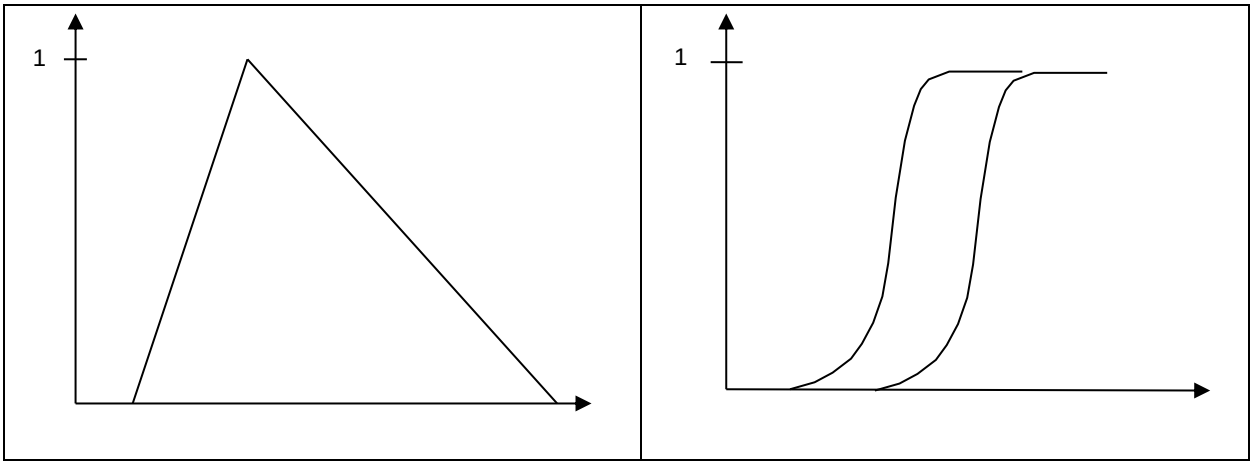
6.3.2. Input uncertainty modelling and uncertainty theory

The previous review is based on a probabilistic modelling of each uncertain input parameter. The probability theory offers an interesting framework to take into account uncertainties since the analysis is based on classical statistical techniques leading to a straightforward estimation of SRQ uncertainty. However, even if there exist several recommendations, the state of knowledge can be, in some situations, not substantial enough to ensure the choice of a unique PDF for each input parameter. Due to this lack of knowledge, the information on uncertainties becomes imprecise and alternative theories (Ferson et al., 2003; Dubois and Prade, 1988; Shafer, 1976) are available in the literature to perform the uncertainty analysis. Refer to *Uncertainty representation and combination: new results with application to nuclear safety issues* (Destercke, 2008) for a general review of existing theories and connections between them.

A straightforward way to model imprecision can first rely on the use of interval. In this case, a lower and upper bound define the uncertainty. However, the information on uncertainties is usually richer and the possibility theory can be exploited to combine a set of intervals. Refer to *Possibility Theory* (Dubois and Prade, 1988) for a full description of the theoretical aspects.

The possibility framework can be interpreted as a generalisation of interval calculation since a possibility distribution can be constructed from a nested set of intervals associated to different degrees of certainty. Figure 6.3 (left) displays an example of triangular possibility when the information on input parameter is reduced to an uncertainty band and a nominal value. Imprecision can be also modelled using probability boxes (or P-box) (Ferson et al., 2003). A P-box is a pair of CDFs ($F^{<}, F^{>}$) satisfying $F^{<} \leq F^{>}$ (Figure 6.3 right). The difference between them represents the imprecision. This kind of modelling also implies a set of intervals, which are translated, and not nested, as in the possibility approach.

Figure 6.3. Two types of input uncertainty modelling



Note: Left, possibility distribution; right, probability box.

In the uncertainty propagation step, classical Monte Carlo techniques need to be adapted to take into account the combination of random (modelled by PDFs) and imprecise variables. It especially requires extending the sampling approach to handle different sets (of values for the probabilistic modelling and of intervals for the possibility or P-box modelling). The uncertainty model after propagation is therefore no more of a random type but a hybrid one. Then, the extraction of the useful information can be performed by post-processing the results. Ferson and Ginzburg (1996) and Baudrit et al. (2006) have proposed for example a summary given in the shape of one or multiple P-boxes to facilitate the interpretation in practice. Optimised sampling strategies have also been introduced in (Chojnacki et al., 2010) in order to reduce the computational cost of the propagation by replacing post-processing by pre-processing.

In most applications, the analysis of the results therefore exploits the collection of P-boxes derived after the propagation. The advantage is that the classical statistical treatment recalled in Section 0 can be used to get the uncertainty related to a SRQ. Since the probability of an event is no more precise, the only difference is that a pair of values is associated to each statistical quantity of interest (such as quantiles), expressing the lack of information.

6.4. Step 16: Comparison between simulation and experimental results using validation indicators

This section is devoted to the description of several validation indicators that can be used to check the agreement between uncertainty results in the output space and the experimental data. The difficulty to derive validation indicators for output uncertainties is essentially due to the different natures of the quantities to compare. The validation of best-estimate simulations is usually based on the comparison between specific values of interest (such as maximum) or time evolution coming from simulations and experiments. In these cases, different mathematical norms can be defined to construct validation metrics. When focusing on uncertainty result validation, it might be necessary to evaluate the discrepancy between an interval and a unique experimental value. A straightforward comparison can consist in checking if the experimental value falls inside the interval. However, with this type of validation indicator, a very large uncertainty band will be considered as acceptable since it is more likely to encompass the experimental value. Therefore, similarly to BE simulation validation, a fully transparent construction of validation indicators should first rely on an appropriate definition of:

- The target quantity of validation (interval, CDF…)
- The important characteristics of the SRQ uncertainty to capture for validation

This involves the use of more complex validation indicators to have a better qualitative insight into the results. Some of them are recalled in the following sections. Annex E provides an illustration of them in the framework of the PREMIUM project.

6.4.1. Target quantity for validation

It is assumed in the sequel that each SRQ is a scalar quantity that has been the most encountered situation in the nuclear studies of interest (NEA, 2011; NEA, 2017) and that its associated uncertainty can be summarised by an interval [LUB,UUB] as well as a reference value RV. Refer to Section 6.4.4 for discussions on the validation of the time-trend of SRQs.

For some validation indicators, it will also be interesting to keep the largest interval $[LUB_{min},UUB_{max}]$ coming from the propagation of the initial model input uncertainties performed in element 3. This largest interval can therefore be split into four disjoined sub-intervals ($[LUB_{min},LUB]$, [LUB, RV], [RV, UUB], $[UUB,UUB_{max}]$). In the probabilistic framework, each bound of the sub-intervals is assumed to be associated for all SRQs to a same quantile that can be estimated by one of the previous methods, for example. The difference between two successive bounds provides the expected percentage (or interquantile) of SRQ realisations that should belong to each sub-interval. For example, if LUB_{min} is associated to the 0%-quantile, LUB to the 5%-quantile, RV to the 50%-quantile, UUB to the 95%-quantile and UUB_{max} to the 100%-quantile, the vector of interquartile is $(p_1,p_2,p_3, p_4)=(0.05,0.45,0.45,0.05)$. More generally, since the computation of the validation indicator recalled in this section is not restricted to four interquantiles, the general notation $(p_1,\ldots,p_B)$ is introduced in the case of B sub-intervals.

6.4.2. Review of some validation indicators

There is significant literature (e.g. (Liu et al., 2011)) on the construction of validation indicators. This section is devoted to recalling classical validation approaches for a fixed SRQ (the question of aggregation of validation results is beyond the scope of this chapter) when the TQV is defined following the previous section. It is assumed that N experimental values are available. They can come, for example, from multiple measurements using the same experimental facility or a unique measurement for multiple experimental facilities.

Most of the indicators are focused on a consistency checking between the information obtained after propagation of input uncertainties and experimental values. This type of indicator is referred to as calibration indicator.

6.4.2.1. A first calibration indicator

The most straightforward indicator consists in checking whether experimental values fall in the uncertainty intervals. If $\{v_i\}_{i=1,\ldots N}$ represents the set of N available experimental values, it leads to the computation of

$$Q_0 = \frac{1}{N}\sum_{i=1}^{N} 1_{\{v_i \in [LUB_i,UUB_i]\}}$$

Where $1_{\{v_i \in [LUB_i,UUB_i]\}}$ is equal to 1 if the experimental value v_i falls in $[LUB_i,UUB_i]$ and 0 otherwise. Since every $[LUB_i, UUB_i]$ is associated to the same interquantile, the discrepancy between Q_0 and the corresponding level of confidence can be used to obtain a quantitative insight on the agreement between the uncertainty results and the experimental information. This type of indicator can also be found in a more general formulation in probabilistic forecasting through the concept of probabilistic calibration (Gneiting et al., 2007).

However, Q_0 does not take into account the position of each experimental value inside the uncertainty interval and does not allow distinguishing between very different situations such as “all experimental values are located close to a bound of the interval” and “all experimental values are uniformly located”. This suggests focusing on the percentage of experimental values falling in each sub-interval and constructing the validation indicator in the framework of hypothesis testing.

6.4.2.2. Hypothesis testing

If r_1*N,..,r_B*N is the number of experimental values falling in each sub-interval associated to the vector (p_1,…,p_B) introduced in Section 6.4.1, a validation indicator can be defined as a measure of the "surprise" of observing r_i*N when p_i is thought to be the right answer. In this case, its construction can be based on hypothesis (or significance) testing which is a well-developed statistical method. It requires formulating a (null) hypothesis that will be tested and that reads for this particular problem "(p_1… pB) defines the true vector of interquantiles". Then, an indicator can be introduced as a measure of the discrepancy between the observed situation and the expected one and a very large value (exceeding a given threshold) will indicate a poor agreement between the uncertainty results and the experimental information. Among classical test, one can mention the χ_2-test that consists in computing the following quantity:

$$Q_1 = \sum_{i=1}^{B} N \frac{(r_i - p_i)^2}{p_i}$$

More formally, if the experimental values are obtained independently, it can be proven that for sufficiently large N, Q_1 follows a χ^2_{B-1} distribution. In other words, the null hypothesis is rejected at a given confidence level when Q_1 is larger than a threshold (associated to the chosen significance level) which is very unlikely to be exceeded for a χ^2 variable. If Q_1 is smaller than the threshold, it does not mean that the hypothesis is accepted but that it cannot be rejected at the chosen confidence level. Table 6.2 provides some examples of thresholds. It is important to keep in mind that the strong underlying assumption when using hypothesis testing is related to the independence between experimental values (e.g. independence between experimental conditions), which does not always correspond to the encountered situations in practice. Therefore, the significance level can be used only in this case to provide some qualitative insight on the validation process.

Table 6.2. Examples of thresholds with respect to the number of sub-intervals and to the significance level

Significance level / **#sub-intervals**	0.1	0.05	0.025	0.01
2	2.706	3.841	5.024	6.635
4	6.251	7.815	9.348	11.345
6	9.236	11.07	12.832	15.086
8	12.017	14.067	16.013	18.475
10	14.684	16.919	19.023	21.666

Following the same idea, another measure of discrepancy can be introduced. It is based on the Kullbach-Leibler divergence (Cooke, 1991) and is written:

$$Q_2 = \sum_{i=1}^{B} r_i * Log\left(\frac{r_i}{p_i}\right)$$

Similarly to the previous indicator, it is known that $2*N*Q_2$ converges to a χ^2_{B-1} distribution as N gets larger.

A normalised indicator can be introduced for both measures of discrepancy exploiting the CDF of the χ^2_{B-1} distribution (see for more details, expression (4) in "Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes" [Destercke and Chojnacki, 2008]).

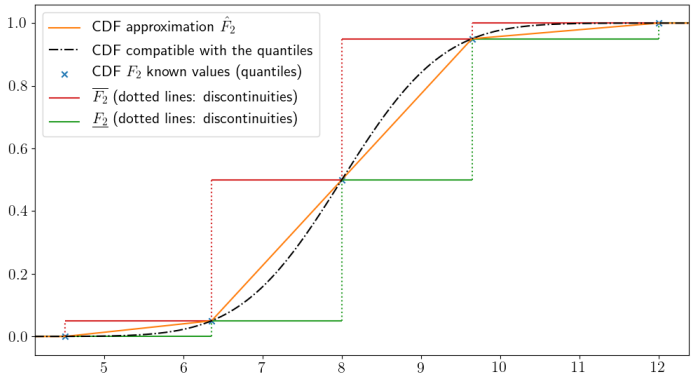
Hypothesis testing approaches can be also constructed in the Bayesian framework (Liu et al., 2011; Kass and Raftery, 1995). In this case, the validation indicator, called the Bayes factor, can be interpreted as the ratio between the likelihood of observing the experimental data under the null hypothesis and the likelihood of observing them under an alternative hypothesis. As mentioned in "Toward a better understanding of model validation metrics" (Liu et al., 2011), its main advantage is that it focusses on hypothesis acceptance rather than hypothesis rejection. However, it is sensitive to the prior knowledge of the alternative hypothesis. Therefore, special attention should be devoted to the modelling choice of the different terms involved in its construction before applying this indicator to practical problems.

By introducing a set of thresholds and under the assumption of independent experimental values, hypothesis testing allows evaluating if the null hypothesis can be rejected or not. Therefore, this type of approach can be interpreted as a consistency check between simulation and experimental results. It does not lead to a quantitative evaluation of the agreement. This last point can be tackled using the following area metric indicator.

6.4.2.2. Area metric indicator

In this section, the N experimental values are assumed to derive from a probability distribution F_1. The distribution of the simulated SRQ due to the initial input uncertainties is denoted by F_2. As in the previous sections, F_2 is assumed unknown except few of its quantiles and the purpose is to check whether these quantiles match with F_1 or not. It is assumed that an interval $[q_0, q_1]$ enclosing the support of F_2, that is all the possible values returned by the simulation model, is given ($F_2(x) = 0$ if $x < q_0$ and $F_2(x) = 1$ if $x \geq q_1$). Figure 6.4 illustrates what is known about F_2 and that many different distributions are compatible with this incomplete information about F_2: any non-decreasing right-continuous function which lies between $F_2^<$ and $F_2^>$ are CDFs compatible with the known quantiles of F_2 (namely the quantiles of order 5, 50 and 95%); in other words, the P-box $[F_2^<, F_2^>]$ encodes all the information about F_2.

Figure 6.4. Illustration of the lack of information about F_2



Even if the experimental probability distribution F_1 is known only through a sample of N data and despite the imprecision on F_2, one idea is to build a validation indicator upon a distance (or even upon a discrepancy) in the space of probability distributions. Such a distance, also known as metric, can be derived from a norm. In particular, the area metric $A(F_1; F_2)$ between two univariate distributions of CDF F_1 and F_2 is defined as the L^1-norm of F_1-F_2:

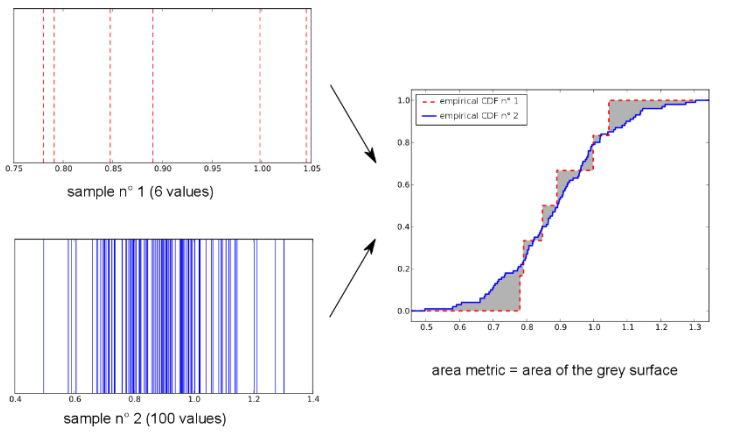
$$A(F_1;\ F_2) = \int_{-\infty}^{\infty} |F_1(x) - F_2(x)| dx.$$

This area metric, proposed in (Ferson et al., 2008) in the context of the validation in scientific computing, has some relevant features:

- $A(F_1; F_2) = 0$ if, and only if, $F_1 = F_2$ (the area metric is a metric in the space of the CDF).
- The area metric is expressed in the same physical unit than the SRQ.
- The preceding definition holds if F_1 or F_2 is discontinuous, for example if there is one exact experimental value, whereas other metrics require the existence of PDF.

It is possible to estimate $A(F_1; F_2)$ by $A(\hat{F}_1; F_2)$, $A(F_1; \hat{F}_2)$ or $A(\hat{F}_1; \hat{F}_2)$ where $\hat{F}_1$ and, respectively, $\hat{F}_2$ are some estimates of F_1 and, respectively, F_2 ; in particular, the empirical CDF of some samples following F_1 and, respectively, F_2 can be used as such estimates. Figure 6.5 illustrates the estimation of an area metric by means of empirical CDF.

Figure 6.5. Area metric of the empirical CDF of two samples



In our context, the idea is to assess the discrepancy between the distribution F_1 of the experimental results and the distribution F_2 of the simulated SRQ by considering the estimate $A(\hat{F}_1; F_2)$ of $A(F_1; F_2)$ where $\hat{F}_1$ is the empirical CDF of the N experimental values. Since F_2 is unknown, one option is to calculate $A(\hat{F}_1; \hat{F}_2)$ where $\hat{F}_2$ is an approximate of F_2, e.g. the continuous piecewise linear interpolation function as displayed on Figure 6.4. A second option is to take into account the imprecision associated to the simulated SRQ and to extend the area metric to this situation (Ferson and Oberkampf, 2009; Lee et al., 2016; He, 2019). In such a framework, an interesting strategy is to compute the minimal and maximal value $A(\hat{F}_1; F)$ for all non-decreasing right-continuous function F inside the P-box $[F_2^<, F_2^>]$ ($F_2^< \leq F \leq F_2^>$). The corresponding interval would be a less accurate but more objective assessment of the discrepancy between F_1 and F_2 than $A(\hat{F}_1; \hat{F}_2)$. It turns out that the minimisation of $A(\hat{F}_1; F)$ with respect to F is simple (the evaluation of this minimal value of the area metric is similar to the approach proposed in "Validation and updating in a large automotive vibro-acoustic model using a P-box in the frequency domain" (Lee et al., 2016), whereas its maximisation appears difficult. Refer to Annex E for further details on this point.

The use of the area metric, as suggested above, raises several objections and questions:

- It does not make it possible to distinguish between precision and trueness (see Section 6.4.3): a great area metric $A(F_1; F_2)$ between two distributions F_1 and F_2, e.g. two Gaussian ones $F_1 \sim N(\mu_1; \sigma_1)$ and $F_2 \sim N(\mu_2; \sigma_2)$, occurs if $|\mu_1-\mu_2|$ is large even if $\sigma_1 = \sigma_2$, but also if $|\sigma_1-\sigma_2|$ is large even if $\mu_1 = \mu_2$.
- Many other non-equivalent metrics are available to compare F_1 and F_2. Contrary to Section 0, the imprecision on F_2 prevents using the area metric to build a statistical test.
- The smart calculation of the supremum of $A(\hat{F}_1; F)$ with respect to F in the P-box $[F_2^<, F_2^>]$ remains to be investigated.

6.4.3. Informativeness and calibration

Calibration indicators are focused on the comparison between simulation and experimental results. In some situations, it can happen that a satisfactory calibration is reached thanks to a wide uncertainty band that is more likely to encompass an experimental value. However,

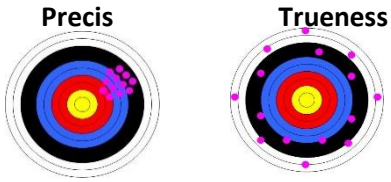
an artificially large uncertainty could be difficult to analyse. The characteristics (and especially the width) of the uncertainty band, independently of the mismatch with experimental values, are therefore important to consider.

Distinguishing concentration of the information associated to the SRQ uncertainty and calibration has already been addressed in other scientific fields. It is the case for example in metrology since the international vocabulary of metrology (JCGM, 2008) introduces the two concepts of precision and trueness (Figure 6.6):

- Measurement precision: closeness of agreement between measured quantity values obtained by replicate measurements on the same or similar objects.
- Measurement trueness: closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value.

They are combined to evaluate measurement accuracy.

Figure 6.6. The two concepts coming from metrology to evaluate measurement accuracy



Note: The pink points represent measurements.

Similar concepts can also be found in forecasting (Gneiting et al., 2007), where important work has been done to define the attributes of "good" probabilistic forecasts (Mason, 2015). Finally, it's important to mention the contributions from expert judgement evaluation (Cooke, 1991; Destercke and Chojnacki, 2008), which introduce the notions of informativeness and calibration. The latter terminology is used in this document. In the framework of input uncertainty validation, informativeness is related to the dispersion of the information obtained after propagation of input uncertainties. The less dispersed the results, the higher the informativeness. Contrary to calibration, this indicator is associated to each SRQ uncertainty. In the case of a unique measurement from multiple experimental facilities, it is computed for each facility and a global informativeness score is usually obtained by averaging.

In the probabilistic approach, informativeness can be studied through the computation of the Kullbach-Leibler divergence:

$$Q_{3,SRQ} = \sum_{i=1}^{B} p_i * Log\left(\frac{p_i}{u_i}\right)$$

Where $(p_1,\ldots,p_B)$ is a vector of interquantiles associated to the SRQ as in Section 6.4.1 and $(u_1,\ldots,u_B)$ is the vector of interquantiles associated to a uniform density defined on the same interval. This last density is commonly used to represent the state of ignorance and Q_3 will therefore measure the improvement of the model built from the estimated input uncertainties compared to a model based on ignorance. From a practical point of view, a large value of this indicator indicates more precise information.

An extension of the informativeness and calibration concepts was developed in (Destercke, 2008) in the framework of the possibility theory (Dubois and Prade, 1988). It has been used to analyse the participants' contributions of the PREMIUM benchmark (NEA, 2017). Starting from a possibility modelling of the SRQ uncertainty result which is for our specific problem a triangular distribution (Figure 6.7, left) defined by:

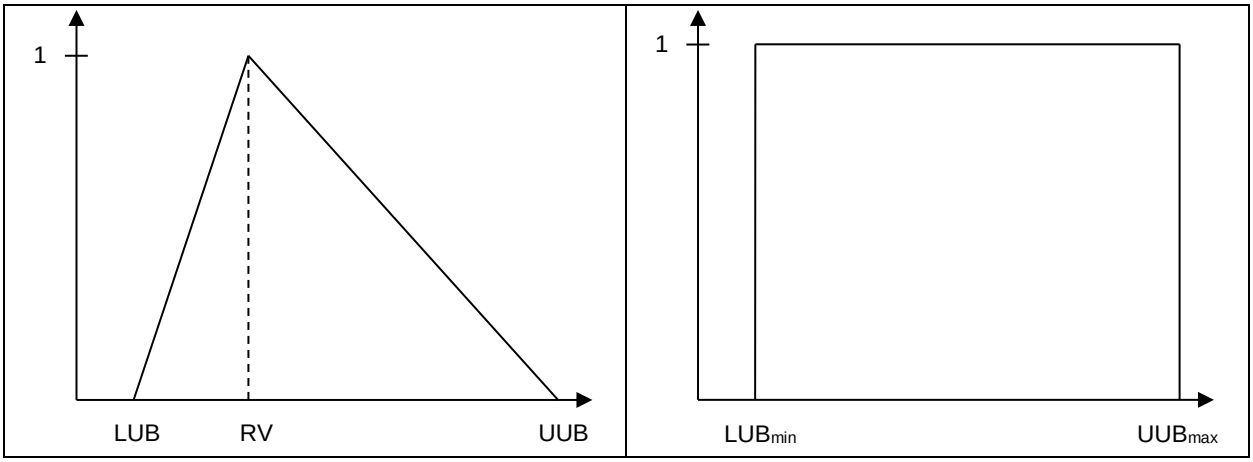
$$\pi_{SRQ}(t) = \begin{cases} \dfrac{t - LUB}{RV - LUB} & if\ t \in [LUB, RV] \\ \dfrac{t - UUB}{RV - UUB} & if\ t \in [RV, UUB] \end{cases}$$

The informativeness (Figure 6.8, left) is related to the ratio between the area of the triangle and the area associated to the uniform possibility defined on [LUB_{min},UUB_{max}] (Figure 6.7, right) and that represents the complete ignorance:

$$Q_{4,SRQ} = 1 - \frac{UUB - LUB}{2(UUB_{max} - LUB_{\min})}$$

A value close to one (resp. 0.5) therefore means that the uncertainty range is narrow (resp. large).

Figure 6.7. Information modelling associated to a SRQ (left) and to complete ignorance (right)

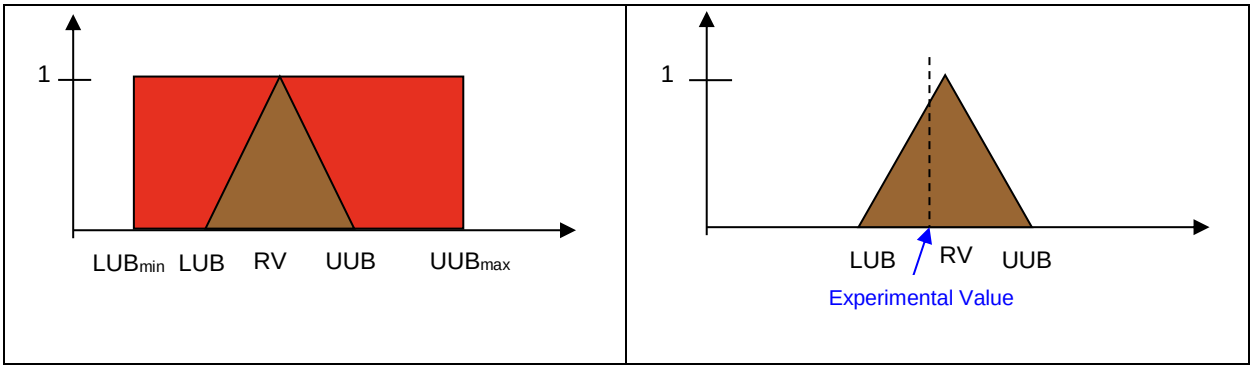


If *v* denotes the experimental value, the calibration (Figure 6.8, right) consists in computing:

$$Q_{5,SRQ,exp} = \pi_{SRQ}(v)$$

A value close to one (resp. 0) therefore means that the absolute error between the reference value and the experimental one is small (resp. large) relative to the uncertainty interval width. More details can be found in "A practical methodology for information fusion in presence of uncertainty: application to the analysis of a nuclear benchmark" (Baccou and Chojnacki, 2014).

Figure 6.8. Computation of the informativeness (left) and calibration (right) indicators associated to a SRQ. $Q_{4,SRQ} = 1 - r$ *.



* r is the ratio between the area in brown and the sum of the area in brown and red, $Q_{5,SRQ,exp}$ is the value on the y-axis corresponding to the intersection of the dashed vertical line and the edge of the triangle

Contrary to the possibility informativeness criterion that has been constructed following the same idea as in the probabilistic framework, the calibration criterion does not have the same meaning in both frameworks. The probabilistic calibration is based on the convergence between the expected SRQ distribution and an experimentally built one. The possibility one does not assume any kind of convergence. It considers, for each SRQ and each experiment, how far the observed value is from the reference one.

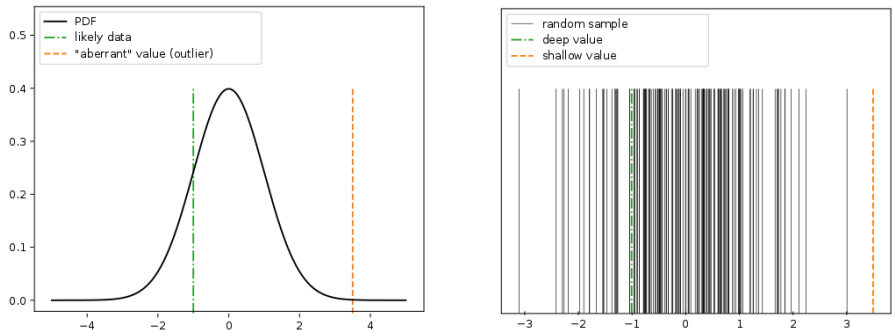
In case of several experimental values (e.g. several experimental facilities), global informativeness and calibration can be obtained by averaging the indicators associated with each single result. Moreover, when possible, it can be interesting to integrate experimental uncertainties in the computation of these two criteria. Refer to "Elicitation, assessment and pooling of expert judgments using possibility theory" (Sandri et al., 1995) and *Uncertainty representation and combination: new results with application to nuclear safety issues* (Destercke, 2008) for an extension of the construction to this situation.

6.4.4. Validation indicators for time-trend SRQs

In this section, the purpose is to evaluate whether an experimental time-trend SRQ is in agreement with a sample of N-1 time-trend simulated SRQ over a time interval [0; T]. The latter are denoted by y_n, for $1 \leq n \leq N-1$, and are regarded as independent and identically distributed realisations of the SRQ in accordance with the input uncertainties; the former is denoted by y_N. The idea is that the validation results should be rejected if y_N appears to be an outlier of the whole sample ($y_1 \ldots y_N$). Such an analysis can be carried out by employing the statistical notion of depth.

First, let us consider a sample of scalar SRQ, as illustrated in Figure 6.9.

Figure 6.9. Illustration of the notion of statistical depth for scalar data

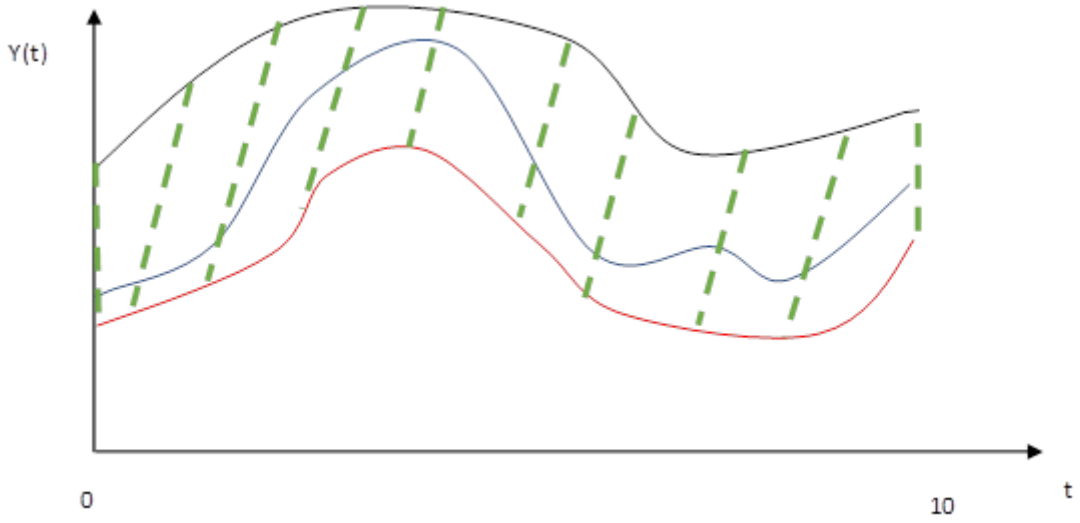


Note: Left: two data are "compared" to a given PDF; Right: the PDF is unknown but a number (150) of values sampled according to it is available; an observation deeply buried in the sample (surrounded by many values) does not seem to be aberrant, whereas an isolated observation (shallow value) is unlikely to come from the distribution.

It illustrates that, given a sample ($y_1 \ldots y_N$), an "aberrant" observation can be identified as a value which is surrounded by relatively few values of the sample: this observation has a small *depth*. The empirical depth of a value y inside the sample can be defined as the proportion of couples (y_i, y_j) with $i \neq j$ and $y_i \leq y_j$ such that $y_i \leq y \leq y_j$. In the right part of Figure 6.9, there are k = 29 sampled values, which are smaller than the observation y = -1: the depth of the latter is, thus, k(N-k) (number of couples bounding y) divided by the number N(N-1)/2 of possible couples, that is about 0.3; whereas the depth of the observation y = 3.5 is zero, while the maximal depth among the sample values is close to 0.5 (which is always true as soon as the size of the sample is large enough). In the case of a scalar SRQ, the notion of depth is connected to the one of quantile: the depth of (empirical or theoretical) quantile of order α is maximised for $\alpha = ½$ (the deepest value is the median), and is a decreasing function of $|1/2 - \alpha|$: a shallow observation corresponds to a quantile of relatively extreme order (α close to 0 or to 1). The notion of depth can be more naturally extended to multivariate data or functional data (e.g. time-trend data) (Zuo and Serfling, 2000) than the notion of quantile.

In particular, the band depth and modified band depth proposed in (López-Pintado and Romo, 2007) can be used to analyse time-trend scalar SRQ data. Their definitions rely on the definition of band: the band of j functions of time ($y_{n1} \ldots y_{nj}$), with j > 1, is the set B (y_{n1}, …, y_{nj}) = {(t, y) : t ∈ [0, T] and $\min_{1 \leq k \leq j} y_{nk}(t) \leq y \leq \max_{1 \leq k \leq j} y_{nk}(t)$}. An example of band is displayed in Figure 6.10.

Figure 6.10. Example of band (dashed line) defined by three functions over [0, T] = [0, 10]



The band depth and modified band depth of a scalar function y: [0, T] → ℝ inside a sample of N functions (y_1, …y_N), with y_n : [0, T]→ ℝ for all 1 ≤ n ≤ N, are defined as follows.

1. Band depth of y among (y_1, …y_N):

$$S_{N,J}(y) = \sum_{j=2}^{J} p_{N,j}(y) \quad with\ p_{N,j}(y) = \binom{N}{j}^{-1} \sum_{1 \le n_1 \le \cdots \le n_j \le N} \mathbb{I}(G(y) \subset B(y_{n_1}, \ldots, y_{n_j}))$$

   where $\mathbb{I}(E)$=1 if E is true, 0 otherwise, and where $p_{N,j}(y)$ is the proportion of bands defined by j functions which enclose the graph G(y) of y, that is G(y) = {(t, y(t)) for all t ∈ [0, T]}. In practice, the depth is computed for J=2 only.

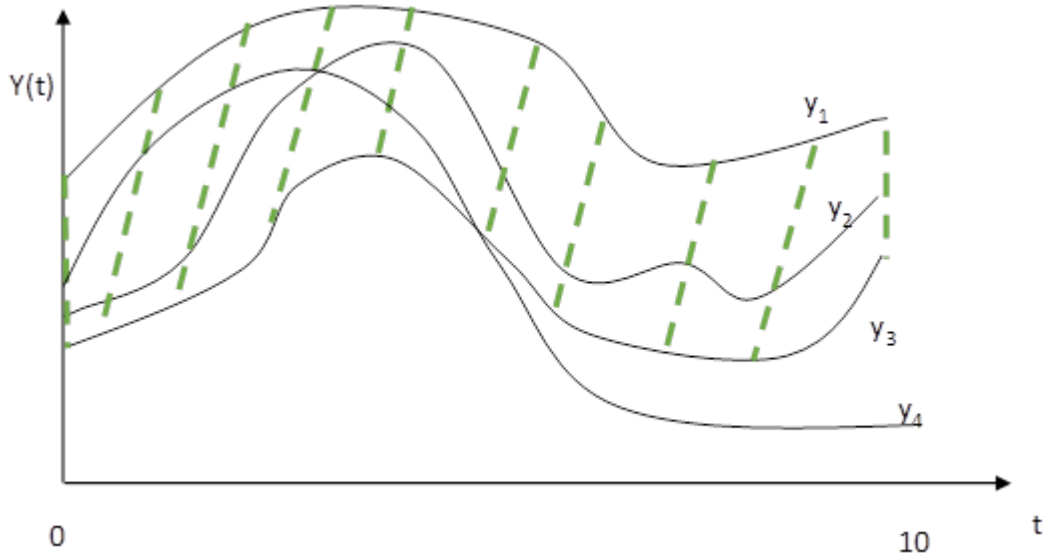
2. The modified band depth of y among (y_1, …, y_N) is derived from the band depth by replacing $\mathbb{I}(G(y) \subset B\left(y_{n_1}, \ldots, y_{n_j}\right)$, which equals either 1 (the graph of y is entirely inside the band) or 0 (there is time t in [0, T] such that (t, y(t)) is out of the band), by the proportion of time during which y is inside the band:

$$GS_{N,J}(y) = \sum_{j=2}^{J} \binom{N}{j}^{-1} \sum_{1 \le n_1 \le \cdots \le n_j \le N} \lambda(\{t \in [0,T] : \min_{1 \le k \le j} y_{n_k}(t) \le y(t) \le \max_{1 \le k \le j} y_{n_k}(t)\}) / T$$

   where λ(.) stands for the Lebesgue measure (just above, it returns the duration during which y is inside the band).

On Figure 6.11, a band B(y_1, y_3) = B(y_1, y_2, y_3) and an observation y_4 are displayed. The contribution to this band to the depth $S_{N,J}(y_4)$ is zero (the graph $G(y_4)$ is not entirely inside B(y_1, y_3)), whereas its contribution to the modified band depth $GS_{N,J}(y_4)$ is 0.4 ($G(y_4)$ is inside the band 40 % of time).

Figure 6.11. Illustration of a band $B(y_1, y_3) = B(y_1, y_2, y_3)$ whose contribution to $S_{N,J}(y_4)$ is different from the one to $GS_{N,J}(y_4)$



In our context of validation, the depth $GS_{N,J}(y_N)$ of the experimental time-trend SRQ and the depths $GS_{N,J}(y_i)$, $1 \leq i \leq N-1$, of the simulations can be computed. The fact that $GS_{N,J}(y_N)$ belongs to the lowest depths, typically the lowest 5%, would be a presumption against the validity of the results.

It is worth noting that an approach other than the evaluation of depth might be carried out to assess whether a time-trend experimental SRQ is unlike in the light of a sample of simulations (therefore might raise a presumption against the validity of the results). The main steps of its construction are described in Annex E. However, this approach is affected by what is called the curse of dimensionality and requires, in general, much more data than the evaluation of depth. The notion of depth may thus appear as a more relevant approach in practice.

6.5. Step 17: Analysis of the validation results and iteration in a loop approach

The analysis is performed in the validation domain. For each SRQ, once the validation indicator has been computed, the evaluation of the agreement between simulation and experimental results requires the definition of a scale of acceptability. This scale obviously depends on the type of validation indicator. If this indicator only consists in checking if each experimental value falls inside each corresponding uncertainty band, it should be based on the percentage of experiments for which this previous agreement check is satisfied. When considering the more complex indicators described in Section 6.4.2, the process is more involved. In the case of hypothesis testing, the indicator is unitless and significance thresholds are available (Table 6.2). However, as mentioned in Section 0, these thresholds can be only used to reject the acceptability and not to accept it. Moreover, they are only valid under an independence assumption between experiments. If the area metric is chosen, since the indicator is expressed in the SRQ physical unit, the scale should be based on expert's judgement. Concerning the informativeness and the possibility calibration, there is not a clear consensus on the choice of significance thresholds.

In practice, depending on the type of experimental database (Element 2 of the SAPIUM approach), several validations can be performed, corresponding for example to different components of the reactor, different scales, different SRQs. It is then important to evaluate the overall validation performance by taking into account the information coming from each validation. This implies defining aggregation strategies. A straightforward solution to construct an aggregation operator is to equally consider all the results. This could be improved by weighting each contribution according to its relevance for the intended use.

Keeping in mind that the objective is the application to nuclear power plants, the analysis of the results provided by the computation of the validation indicators is not sufficient. Adapting the works of (Hemez et al., 2010) or (Oberkampf et al., 2007) to the IUQ framework, this analysis in the validation domain should be combined with the adequacy of the experimental database. It leads to the evaluation of the maturity of the IUQ for the intended use. This evaluation should be carried out in a loop approach. If the maturity is proved to not be sufficient, an iteration step is performed. It means the maturity is improved according to findings during quantification and validation and the whole procedure is repeated. It is important to keep in mind that a lack of maturity can be due to different choices adopted in each SAPIUM element. One reason could be the lack of adequacy of the experimental database. In this case, it is preferable to consider integral experiments in the quantification (and not only in the validation) and to iteratively revise (Heo et al., 2018) the quantified model input uncertainties from these new data.

6.6. Recommendations and open issues

6.6.1. Recommendations

The validation should be performed for SRQs that have been taken into account in the quantification step (Element 4 of the SAPIUM approach).

If the experimental database is large enough, validation experiments should correspond to experiments that are not used for the uncertainty quantification (Chapter 5). When the number of experiments is not sufficient to split the database, it is recommended to perform a leave-one-out cross-validation while avoiding removing experiments that would lead to extrapolation of the results and to a strong reduction of the adequacy of the remaining database.

It is important to control (or evaluate) the impact of the methodological assumptions on the validation result. These assumptions are first related to the input uncertainty propagation method and especially to the input uncertainty modelling that should integrate the state of knowledge on uncertainties (e.g. aleatory/epistemic uncertainties). They also concern the construction of some validation indicators that relies on strong assumptions on experiments (e.g. independence of experimental conditions) that are not always satisfied in practice. This should be taken into account when the acceptability of uncertainty results is checked for the intended use by comparison to acceptability thresholds. Another option could be to enforce the connections between experimentalists and developers to design new validation experiments following specific requirements (Oberkampf and Trucano, 2007).

Special attention should be devoted to the choice of the TQV. In this document, this quantity is defined through quantiles, which has been the most encountered situation in previous OECD projects and several adapted validation indicators have been recalled. However, the construction of this type of indicator can be extended to handle CDFs rather than quantiles.

It is recommended to clearly state the important characteristics of the SRQ uncertainty to capture for validation before starting the study. Most of the indicators are focused on a consistency check between simulation and experiment. However, it is advisable to take into account extra features such as the concentration of the information provided by the uncertainty analysis by combining informativeness and calibration type indicators.

The analysis of the results provided by the computation of the validation indicators is not sufficient if the objective is application to nuclear power plants. It is recommended to construct a maturity model combining the acceptability of the validation results in the validation domain with the adequacy of the experimental database. The evaluation of the predictive maturity should be carried out in a loop approach. If the maturity is proved as not sufficient, an iteration step should be performed including a new quantification and validation. If the lack of maturity can be explained by a lack of adequacy of the experimental database, it is preferable to extend the database by considering integral experiments for the quantification and to iteratively revise the quantified model input uncertainties from these new data.

6.6.2. Open issues

Besides the question of predictive capability that is discussed in Chapter 7, several issues remain to be tackled to ensure the validation of the input uncertainties.

Cross-validation is an interesting alternative in the case of a small amount of experiments. However, this process is more involved since it requires performing several quantifications/validations on different subsets of experiments. Moreover, removing one experiment might lead to a strong reduction of adequacy. Therefore, the practicability of such a strategy should be more deeply studied to drive BEPU users in their analysis.

There exists a large literature on validation indicators associated to scalar SRQs. However, further work is required in the case of functional ones (e.g. time-trend). It is not straightforward, for example, to extend the concepts of calibration and informativeness to this situation.

Significant effort should be also devoted to the analysis of the validation results with the objective of extrapolation to the application domain. Besides the definition of a scale of acceptability, it involves the construction of a reliable model for the evaluation of the predictive maturity of the whole process. A general guideline is provided in (Oberkampf et al., 2007) in the VVUQ framework as well as a scale of maturity but adaptation to IUQ and combination with adequacy criteria described in Chapter 3 remain open questions. These challenging problems also imply being able to aggregate several validation results corresponding, for example, to different components of the reactor, different scales or different SRQs.

Finally, a major issue is related to the presence of other uncertainty sources (numerical approximation error and model form) in the validation process. The contributions of this document are restricted to the quantification of model input uncertainties. However, all the types of uncertainty sources will affect the validation, since it operates in the SRQ space. A poor validation that would require iterating the SAPIUM process can therefore be due to other uncertainty sources than model input uncertainties. To avoid any misinterpretation, further work is necessary to understand and evaluate the impact of each category of uncertainty sources on the validation results.

References

AIAA (1998), *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, American Institute of Aeronautics and Astronautics, AIAA-G-077-1998.

Baccou, J. and E. Chojnacki (2014), "A practical methodology for information fusion in presence of uncertainty: application to the analysis of a nuclear benchmark", in *Environment Systems and Decisions* 34(2), DOI: 10.1007/s10669-014-9496-3.

Baudrit, C., D. Dubois and D. Guyonnet (2006), "Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment", in *IEEE Transactions on Fuzzy Systems*, 14, pp. 593-608.

Cannamela, C., J. Garnier and B. Iooss (2008), "Controlled Stratification for Quantile Estimation", in *Annals of Applied Statistics*, 2(4), pp. 1554-1580.

Chojnacki, E., J. Baccou and S. Destercke (2010), "Numerical sensitivity and efficiency in the propagation of epistemic and aleatory uncertainty", in *International Journal of Intelligent Systems*, 39(7), pp. 683-704.

Cooke, R.M. (1991), *Experts in uncertainty*, Oxford University Press.

Conover, W. (1999), *Practical non-parametric statistic*, Wiley, New York.

Destercke, S. (2008), *Uncertainty representation and combination: new results with application to nuclear safety issues*, PhD thesis, University of Toulouse.

Destercke, S. and E. Chojnacki (2008), "Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes", in *Nuclear Engineering and Design*, 238(9), pp. 2484-2493.

Dubois, D. and H. Prade (1988), *Possibility Theory*, New York Plenum Press.

Ferson, S. and L. Ginzburg (1996), "Different methods are needed to propagate ignorance and variability", in *Reliability Engineering System Safety*, 54, pp. 133-144.

Ferson, S., V. Kreinovich, L. Ginzburg, D. Myers and K. Sentz (2003), *Constructing probability boxes and Dempster-Shafer structures*, Technical report, SANDD2002-4015, Sandia National Laboratories.

Ferson, S. et al. (2008), "Model validation and predictive capability for the thermal challenge problem", in *Computer Methods in Applied Mechanics and Engineering*, 197 (29), pp. 2408-2430.

Ferson, S. and W. Oberkampf (2009), "Validation of imprecise probability models", in *International Journal of Reliability and Safety*, DOI: 10.1504/IJRS.2009.026832.

Frepoli, C. (2008), "An Overview of Westinghouse Realistic Large Break LOCA Evaluation Model", in *Science and Technology of Nuclear Installa*tions, Volume 2008, Article ID 498737.

Frepoli, C. and S. Iyengar (2011), "A Comparison of Non-Parametric Tolerance Limits with Linear Combinations of Order Statistics in Safety Analysis", in *Proc. of ICAPP*, Nice, France.

Gneiting, T.; F. Balabdaoui and A. Raftery (2007), "Probabilistic forecasts, calibration and sharpness", in *Journal of the Royal Statistical Society B*, 69 (2), pp. 243-268.

Guba, A., M. Makai and L. Pál (2003), "Statistical aspects of best estimate method-I", in *Reliability Engineering and System Safety*, 80 (2003), pp. 217-232.

He, Q. (2019), "Model validation based on Probability Boxes under mixed uncertainties", in *Advances in Mechanical Engineering*, 11(5), pp. 1-9.

Hemez, F., H.S. Atamturker and C. Unal (2010), "Defining predictive maturity for validated numerical simulations", in *Computers and Structures*, 88, pp. 497-505.

Heo, J., K.D. Kim and S-W. Lee (2018), "Validation and uncertainty quantification for FEBA, FLECHT–SEASET, and PERICLES tests incorporating multi-scaling effects", in *Annals of Nuclear Energy*, 111, pp. 499-508.

Hofer, E., B. Krzykacz, J. Erhardt, F. Fischer, M.J. Crick and G.N. Kelly (1985), "Uncertainty and Sensitivity Analysis of Accident Consequence Sub-models", in *Proc. of ANS/ENS International Topical Meeting on Probabilistic Safety Methods and Applications*, San Francisco, United States.

JCGM 200:212 (2008), *International vocabulary of metrology – Basic and general concepts and associated terms* (VIM).

Kass, R.E. and A.E. Raftery (1995), "Bayes factors", in *Journal of the American Statistical Association*, 90(430), pp. 773-795.

Labopin-Richard, T and V. Picheny (2017), "Sequential design of experiments for estimating quantiles of black-box functions", in *Statistica Sinica*, 28 (2018), pp. 853-877.

Lee, S.W., B.D. Chung, Y-S. Bang and S.W. Bae (2014), "Analysis of uncertainty quantification method by comparing Monte-Carlo method and Wilks' formula", in *Nuclear Engineering and Technology*, 46(4), pp. 481-488.

Lee, D., N.H. Kim and H-S. Kim (2016), "Validation and updating in a large automotive vibro-acoustic model using a P-box in the frequency domain", in *Structural and Multidisciplinay Optimization*, 54, pp. 1485-1508, DOI: 10.1007/s00158-016-1427-0.

Liu, Y., W. Chen, P. Arendt and H-Z. Huang (2011), "Toward a better understanding of model validation metrics", in *Journal of Mechanical Design*, *133*(7), [071005], DOI: 10.1115/1.4004223133.

López-Pintado, S. and J. Romo (2007), "Depth-based inference for functional data", in *Computational Statistics and Data Analysis*, 51, pp. 4957-4968.

Martin, R.P. and L.D. O'Dell (2005), "AREVA's realistic large break LOCA analysis methodology", in *Nuclear Engineering and Design*, 235(16), pp. 1713-1725.

Martin, R.P. and W.T. Nutt (2011), "Perspectives on the application of order-statistics in best-estimate plus uncertainty nuclear safety analysis", in *Nuclear Engineering and Design*, 241, pp. 274-284.

Mason, S.J. (2015), *Guidance on verification of operational seasonal climate forecasts*, World Meteorological Organization, Commission for Climatology XIV Expert Team on CLIPS Operations, Verification, and Application Service.

Oberkampf, W.L. and T.G. Trucano (2007), *Verification and validation benchmarks*, SANDIA report, SAND2007-0853.

Oberkampf, W.L., M. Pilch and T.G. Trucano (2007), *Predictive capability maturity model for computational modeling and simulation*, SANDIA report, SAND2007-5948.

Oberkampf, W.L. and M.F. Barone (2006), "Measures of agreement between computation and experiment: validation metrics", in *Journal of Computational Physics*, 217, pp. 5-36.

NEA (2017), "Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark: Final Report", NEA/CSNI/R(2016)18, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19752.

NEA (2011), "BEMUSE Phase VI Report, Status report on the area, classification of the methods, conclusions and recommendations", NEA/CSNI/R(2011)4, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19020.

Pastel, R. (2012), *Estimation of rare event probabilities and extreme quantiles, Applications in the aerospace domain*, PhD, European University of Bretagne.

Sandri, S.A., D. Dubois and H.W. Kalfsbeek (1995), "Elicitation, assessment and pooling of expert judgments using possibility theory", in *IEEE Transactions on Fuzzy Systems*, 3(3), pp. 313-335.

Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton, University Press.

Shockling, M. (2015), "Non-Parametric Order Statistics: Providing Assurance of Nuclear Safety", in *Proc. of NURETH-16*, pp. 2477-2490, Chicago, United States.

Sun, Y. and Genton, M.G. (2011), "Functional boxplots", in *Journal of Computational and Graphical Statistics*, 20(2), pp. 316-334.

Wackernagel, H. (1998), *Multivariate Geostatistics*, Springer.

Walter, C. (2015), "Moving particles: A parallel optimal Multilevel Splitting method with application in quantiles estimation and meta-model based algorithms", in *Structural Safety*, 55, pp. 10-25.

Wilks, S.S. (1941), "Determination of sample sizes for setting tolerance limits", in *Annals of Mathematical Statistics*, 12, pp. 91-96.

Zhang, H., R. Szilard, L. Zou and H. Zhao (2016), "Comparisons of Wilks' and Monte Carlo Methods in Response to the 10CFR50.46(c) Proposed Rulemaking", in *Proc. of NUTHOS-11*, Gyeongju, Korea.

Zuo, Y. and Serfling, R. (2000), "General notions of statistical depth function", in *Annals of Statistics*, 28(2), pp. 461-482.

7. Scaling issues

7.1. Introduction

The design and operation of nuclear power plants has become highly dependent upon the thermal-hydraulic code, and it is common to use the analysis code specially developed for safety evaluation. In the development of the theory and model for the system code, a wide variety of data from experimental facilities were assessed. Most system codes normally reflect the increased knowledge and new simulation requirements from both large and small-scale experiments. Therefore, the code manual includes the scale-related information described in the equations and models and suggests the guidelines for nuclear power plant modelling. The assessment manual of system code provides the description of discrepancy between the calculated and experimental data. However, even though the thermal-hydraulic code was essentially developed from the first principles, it should be noted that the scaling issues are important since the majority of constitutive relations are semi-empirical correlations calibrated on the base of selected experimental data. That is, it has to be assessed whether the code physical models are applicable for both large-scale and small-scale experiments. The scaling methodology using system codes has been the subject of many studies (e.g. Ransom et al., 1998; Martinez-Quiroga et al., 2014). In the simulation of nuclear power plants, most experimental data are obtained from small-scale experiments. Therefore, it is essential to consider the scaling effect in the system analysis.

Scaling has been widely used in engineering and science for many years, and the methodology for scaling analysis has evolved from a combination of the dimensionless groups of variables to deriving the scale parameters from the complex physical models. However, some limitations and restrictions were found in this process. Since the assessment and validation of the simulation model are essential for the code development, the well-scaled experiments are needed to get the experimental data. In principle, the tests should be conducted after scaling analysis and the design and fabrication of an experiment. Sometimes, scaling effects are evaluated after finishing the tests, if necessary. However, for both cases, it may be impossible to verify the scaling law without a fully scaled experiment for corresponding postulated scenario. In addition, the different scaling distortions can be caused even in the integral effect test facility because of the thermal-hydraulic interaction in the complex geometry.

The scaling issue is a complex process of conversion data and information obtained in the course of experimental investigations performed at the test facilities and related numerical analyses to the full-scale nuclear installations. This is a difficult process in the thermal-hydraulic analyses, particularly difficult in the case of transients when two-phase (or multi-phase) flow occurs. In the past, considerable efforts were devoted to solving the scaling problems but it is still a subject of actual research in the field of nuclear reactor safety and design. A comprehensive review of scaling applications in the system codes has been performed in the frame of NEA/CSNI activities, and published as a state-of-the-art report (NEA, 2017).

The importance of the scaling issue was recognised by USNRC when proposing the code scaling, applicability and uncertainty (CSAU) methodology (Boyack et al., 1989a). An agreeable solution for the scaling issue is mandatory within the framework of the BEPU (Best-estimate plus uncertainty) approach where realistic computational tools are applied to support the licensing of a nuclear power plant.

7.2. Scaling in uncertainty analyses – general considerations

Scaling concerns appear in different phases and activities related to input uncertainty evaluation. As a general statement, one could establish that input uncertainty analysis has to follow the common guidelines of scaling activities and to do so the CSNI report, "A state-of the art report on scaling in system thermal-hydraulics applications to nuclear safety and design" (NEA, 2017), is today the best comprehensive review of scaling applications of system codes.

More in detail and intending to be consistent with the structure of the present report, scaling concerns appear at both the quantification and validation phases.

At the quantification phase, it is possible to distinguish among: experiment selection, quantification itself and confirmation. Major scaling concerns appear in the activity of experiment selection. It is usually recommended to select tests carried out at a scale as close as possible to that of the future use of it. This commonly means great scale close to the nuclear power plant scale. At this level of selecting experiments, scaling considerations will be needed and calculations defining and limiting the effect of the scale may be necessary. Different methodologies outline guidelines for this step. There are also some other recommendations not connected with the scaling issue. It is worth mentioning data availability, data quality and knowledge of the facility. Calling "quantification itself" the activity related to the strict application of an inverse method to derive mean values, probability distributions of uncertain input parameters, all considerations are done at test scale, and no additional scaling concerns appear. In a similar way, confirmation activity, which is a direct calculation at test scale, also does not involve any additional scaling consideration.

At the validation phase, things are different due to two important reasons. On the one hand, usually the validating comparisons are performed using experimental data of other facilities probably at a different scale. On the other hand, the goal of the activity is to qualify the use of the derived parameter and its distribution for future calculations at the nuclear power plant scale. At this point scaling considerations will be needed and calculations maybe necessary. The existing methodologies also address this issue.

7.3. Types of scaling

Consideration of scaling effects is one of the major problems in the quantification of input uncertainties. The scaling effects are of importance for geometrical scaling as well as scaling according to extension of thermal-hydraulic parameters.

Qualitative and quantitative changes of physical phenomena and related uncertainty differences according to variation of the thermal-hydraulics parameters are usually extensively investigated and their treatment belongs to the standard procedure of model uncertainties quantification. Scaling regarding thermal-hydraulic parameters uses not only extrapolation of experimental results but interpolation or approximation of experimental data. It is particularly the case when experiments are not performed with the original fluids. It can be done for instance according to the simplicity of the experimental devices but sometimes also to preserve interesting phenomena in scaled-down geometry.

Geometry scaling is usually related to the scale-up of data and information from small-scale to large-scale applications, i.e. extrapolation from small to large geometry. It is always the case in integral test facilities. For cost reasons, the models of nuclear installations are scaled down to smaller dimensions. The downscaling is performed according to differed scaling

methods with the aim of preserving important features of the system to be able to simulate interesting phenomena in a realistic way.

The major scaling methods most used in nuclear reactor safety are (NEA, 2017):

- linear scaling;
- power to volume scaling;
- three-level scaling;
- hierarchical 2-tiered scaling (H2TS);
- power to mass scaling;
- modified linear scaling;
- fractional scaling analysis (FSA);
- dynamic system scaling.

Some of these scaling methods, e.g. linear scaling, are used only to reduce the size and definition of adequate thermal-hydraulic parameters for construction of test facility. Some of them, for instance H2TS, are integral methods of analysis involving also numerical simulations. They will be described in the following chapter related to uncertainties quantification.

Scaling methods are essential tools for the simulation of thermal-hydraulic phenomena in nuclear reactor safety. However, the basis of the evaluation and extrapolation to full-scale applications is an adequate experimental database.

7.4. Experimental database and predictive capability assessment

The basis of model uncertainties quantification is the establishment of a suitable experimental database (see Chapter 3). The optimal situation is when there are separate effects tests (SETs) available for the whole spectrum of scales where the model will be applied. The experiments considered by evaluation of scale effect underlay the same general requirements regarding representativeness related to investigate phenomena and geometry and accuracy of experimental measurements, applied for selection of tests for experimental database as described in the Chapter 3.

7.4.1. Experiments for thermal-hydraulic scaling

Concerning scaling according to thermal-hydraulic parameters like pressure or temperature, frequently the experimental data cover the whole range of the parameter of interest. Usually, the experiments cover a large field of application, or there are different experiments which cover other ranges of the thermal-hydraulic parameters. A practical problem is that for some ranges of parameters there are many experiments and for other parts of the application field only few. For instance, it is the case with extreme pressure ranges. There are only a few experiments for very high and very low (below atmospheric) pressure, whereas for other pressure ranges, which were considered as most interesting or useful in the past, there are many. This may lead to problems finding enough experiments for quantification and validation. Moreover, it would be necessary to consider the number of available tests by weighting the experiments, not to overestimate some part of the application field and underestimate others. The question of information weighting in the framework of input quantification is discussed in Section 5.1.

7.4.2. Experiments for geometry scaling

Much more difficult is the problem of geometry scaling. It is mainly a problem of up scaling models and simulations for large geometry. The majority of the experiments were conducted in scaled-down facilities. Experimental investigations of important and interesting phenomena were performed usually at small and medium scale experimental facilities. There are plenty of separate effects tests for small scale but there is only a very limited number of experiments for large scale. For some phenomena, there are no large-scale experiments at all.

Even worse is the situation in the case of integral tests. There are practically no full-scale integral test facilities and there are only very few medium-size test facilities like LOFT or LSTF facilities. For some cases, measured data that were collected for some transient situations in the real power plants can be used for validation of the quantified model uncertainties.

It is worth noting that an important part of constitutive models in the thermal-hydraulic codes are developed on the base of small (or medium) size experiments, e.g. for pipe geometry, the frequently used diameters were one or two in. (2.54 cm or 5.08 cm) pipes. There are experiments in larger scale, which are sometimes called large-scale experiments. However, they are mostly in the range of ten cm, sometimes 15 cm, so they can be considered medium rather than large-scale experiments. A few test facilities, such as MHYRESA (35 cm), have a pipe diameter above 20-25 cm. The information on important experiments relevant for validation of numerical models and codes is collected in the OECD NEA TIETHYS on-line database.

The only extensive experimental programme in original reactor geometry seems to be the UPTF/TRAM experimental programme (Weiss et al., 1986). The programme is already finished, and the data are proprietary, so the available information is limited. In the face of a shortage of large-scale experiments, it is not possible to construct a complete experimental database for all phenomena for large-scale geometry. This deficiency has to be compensated by extensive procedures by quantification and validation of input model uncertainties and to some extend by conducting uncertainty analyses themselves. In this context, taking into account counterpart and similar tests is of importance. Their consideration gives information on if, when, and how the phenomena change according to different boundary conditions and different facilities. Even the tests that do not cover the full-scale give some information, if performed scaling is successful.

7.5. Effect of scale on model uncertainties quantification

By quantification of model uncertainties, the geometrical scaling appears to be the major problem.

A typical and simple procedure of model uncertainties deriving (Chapter 4), valid also by consideration of scale effects by uncertainty analysis, is:

- Quantification of model uncertainties (Chapter 5) based on available separate effects tests and as far as necessary combined effect tests for possibly the whole spectrum of expected applications.
- Validation of the quantified model uncertainties (Chapter 6) using integral tests (and eventually combined effect tests).

However, a lack of large-scale experiments requires extrapolation of data, in particular estimation and extrapolation of model uncertainties from small scales to large scale.

In the up scaling, physical models of the thermal-hydraulic codes are of great weight. Since the balance (governing) equations are of general applicability, the closure relations depends if the predictive capability of the code are satisfying for the intended application.

7.5.1. Consideration of thermal-hydraulic scaling effects by model uncertainties quantification

Usually, there is a sufficient number of experiments dealing with different values of thermal-hydraulic parameters. The investigated differences are mainly due to different pressures, but also different temperatures and void fractions are subjects of investigation. Variation of uncertainties according to changing phenomena and models is considered to construct experimental database by subdivision of the experiments into regions for which differences in occurring phenomenon and model uncertainties can be expected. The differences are considered by quantification defining different uncertainty ranges for each sub-region. Usually, there is a need neither for extrapolation nor for interpolation by determination of model uncertainties. Even if there are not enough SETs to cover the whole range of thermal-hydraulic parameter, the correctness of scaling effects consideration can be checked by validation step. The integral tests simulate usually all the transients passing all the interesting regions of thermal-hydraulic parameters.

7.5.2. Quantification of model uncertainties in input uncertainties propagation method

In the course of BEPU analyses, some practical approaches for handling of scaling issues were developed and applied. The main concern of all scaling approaches is extrapolation of the data/results from small to large scales. The usual way is the quantification of model uncertainties based on an existing experimental basis and the estimation of accuracy of the extrapolation to the full-scale applications. This task is an extension of the validation process called predictive capability. The purely mechanistic constitutive models or terms in the governing equations (like momentum term or gravitation term) are expected to work in the same way and be valid for small as well as for large-scale applications.

The phenomena which are described by semi-empirical models or correlations can differ for large and small-scale geometries and therefore can have different prediction accuracy for different scales. For this reason, scalability of the models needs to be proved.

During the code development, application of the code (models) for the large application field (different geometric and thermal-hydraulic scales) is met by developing mechanistic models with minimum of empirical constants and correlations based on dimensionless groups. Semi-empirical correlations using dimensionless groups related to the phenomena

of interest are expected to be able to predict the modelled physical phenomena for different applications and scales, including extrapolation to large scales, in a correct way.

Usually, semi-empirical models are developed on the base of small-scale experiments, but the application field involves large-scale reactors. Quantification of the model uncertainties for large scales requires experiments on large-scale facilities. Large-scale facilities like UPTF (Weiss et al., 1986) are unique. The scaling effect in the quantification can be investigated comparing the model prediction obtained for small and medium scale experiments. The trends observed by comparison small-scale vs. medium-scale experiments can be extrapolated for large scales. It means the dependency of the quantified model uncertainty on the geometry scale, if any exists and can be evaluated, can be extrapolated proportionally for large scales. This approach reflects the situation where model (mainly constitutive equations) predictions depend on the geometry scale. However, this procedure is not generally valid. It can happen that dependency of model predictions on the geometry scale changes; they are not proportional in the whole range considered geometries. For instance, dependent on predicted drift velocities, CCFL velocities are proportional to hydraulic diameters for small and medium diameters but are independent on it above a certain bounding diameter. Such behaviour was observed and modelled in the code ATHLET (ATHLET, 2004). In this case, it has an impact on the estimation of the uncertainty range of quantified uncertainty.

Quantification of model uncertainties for different scales can involve different uncertainty ranges for different scales, as, for instance, practised by GRS (Skorek, 2009). Since models are developed mainly on the base of available small-scale experiments and the evidence of large-scale experiments for quantification is very limited, or sometimes there are no adequate experiments available, the uncertainty ranges for large scale applications are larger than for small-scale applications. This is at first due to fact that semi-empirical models developed on the base of small-scale experiment are calibrated just for small-scale geometries. Additionally, incompleteness/weakness of large-scale experimental database accounts for an increase of quantified uncertainties to compensate the incomplete information.

The scale-up effects are to be considered by the development of a qualified input data set for each reference (best estimate) calculation and by selection and quantification of uncertain input parameters. In particular, differences in uncertainties of physical models according to their application to different scale objects have to be taken into account by model uncertainty quantification. In the uncertainty analyses performed by GRS, differences in model uncertainties by application to small-scale test facilities and to large-scale test facilities or nuclear power plants are usually considered. Different uncertainty ranges mainly express these differences. Mostly, such differences have been identified for the closure relations of the conservation equations. For instance, for the vertical annulus geometry (geometry of the down-comer) in large-scale facilities, a much wider variation range for interfacial friction for ATHLET applications has been established than in small-scale facilities. In analyses of the LB LOCA accident at the Zion Nuclear Power Plant, the variation range of interfacial friction in annulus geometry was 0.05 – 3.0 (Skorek, 2009). A similar range was applied for analyses at the middle scale annulus in the LOFT test facility. For small-scale annulus, much narrower uncertainty ranges are applied. For instance, for LOCA analyses at LSTF test facility the applied uncertainty range was 0.33 – 3.0 (Skorek et al., 2011). The dimensions of the annulus in the LSTF facility are much smaller than the dimensions of the annulus in a real reactor such as the Zion or even in the LOFT facility. The aim of the extension of the variation range is reduction of the interfacial friction in the ATHLET. It takes into account that the 1-D interfacial friction correlation for the annulus

geometry in the ATHLET code was developed based on small-scale experiments. The relatively low value for the large-scale geometry results from different flow behaviour in large-scale and small-scale facilities. In particular, CCFL is much less restrictive for a large-scale annulus than for a small-scale annulus. In addition, for other interfacial friction correlations like interfacial friction in bundle geometry or in pipe geometry, different ranges of variation are applied for small and large scales.

Another possibility to consider scaling effects is selection of different correlations according to their field of application. If there is such an option in the physical model of the thermal-hydraulic code, different correlations/constitutive equations may be applied for small and large-scale facilities according to the recommendation in the code documentation. Such recommendations result from code development and validation and as such express the state of knowledge concerning also the scaling effect.

The variation of the uncertainty range is the main way of scale-up effect consideration in uncertainty analyses. Since the quantification of model uncertainties takes place by comparison with experimental data, an appropriate selection of the adequate experiments is important. The preferable model uncertainties quantification is comparison of code predictions with experimental data from separate effects tests. The experimental data selected for quantification have to be representative for the considered application. In particular, they have to reflect the scale of the analysed facility for which the uncertainty analysis is to be done.

In this context, an approach with consideration of all possible input uncertainties seems to be advantageous in comparison with approaches which require a limitation of the number of input uncertainties. The rankings of parameters required for selection and limitation of number of input uncertainties are based mainly on experts' knowledge. Experts carry out the identification and limitation of parameters based on their experience. Since there are more smaller facilities than large-scale facilities, the knowledge of phenomena in small-scale facilities is much better than in the large-scale facilities. As result, it may happen that some important parameter for the large-scale facility would be not identified, simply because it never appeared as influential in small-scale facilities. It is particularly important when in the approach only a limited number of the most important input uncertainties can be taken into account.

7.5.3. Integral methods for consideration of scale effects by uncertainty analyses

Quantification and validation of model uncertainties is a complicated and difficult task and special methodologies for handling this problem have been developed.

The first formalised approach dealing with the scaling problem by BEPU analyses was the CSAU methodology (described also in Chapter 1) of NRC. The CSAU procedure of code scaling-up evaluation consists of several steps. In the whole procedure of the scale effects consideration, the crucial point appears to be evaluation of the effect of distortion due to scalability of experiments (SETs as well as IETs) on the important processes (step 10-3 in [Boyack et al., 1990]). However, in the methodology it was not explained how this should (or could) be achieved. Therefore, in this formalised approach the probably most important step is dependent on the expertise of the users of the CSAU methodology. Usually, the uncertainty of the upscaling is met with the introduction of conservative biases and enlarged probability distribution functions.

The CSAU methodology evolved into the evaluation model development and assessment process (EMDAP) (USNRC, 2005). An important difference between CSAU and EMDAP

is that scalability of experiments (SETs as well as IETs) and code scalability are separated, and addressed in different elements in EMDAP. Scaling analyses and identification of similarity criteria are to be performed for experiments in the frame of development of assessment database. The assessment of scalability of the models is previewed in two steps within assessment of evaluation model accuracy. At first, the scalability of the closure relations should be assessed. In the next step, the scalability of the integrated calculations and data for distortion, have to be estimated. The following proof of fulfilment of adequacy standards concerning the applicability of the model for plant analyses is the basis for an adequacy decision, if the evaluated model can be applied to power plant event analysis, or a return to the appropriate element of the methodology is necessary.

To evaluate the effect of scale, other methodologies were also developed such as the hierarchical two tired scaling (H2TS) (Zuber et al., 1998) and fractional scaling analysis (FSA) (Zuber et al., 2005).

The H2TS methodology is comprised of four steps:

- system decomposition;
- scale identification;
- top-down system scaling analysis;
- bottom-up process-scaling analysis.

An example of FSA for a LOCA event is described in (Wulff et al., 2005).

D'Auria and Galassi (2010) proposed a "Road Map" methodology using thermal-hydraulic system codes as a tool for practical validation of up-scale analyses of selected scenario(s).

The above-mentioned methodologies, H2TS, FAS and Road Map for evaluation of upscaling analyses of nuclear power plant events, do not follow the principle of uncertainty evaluation of input uncertainties propagation methods. It means determination of input model uncertainties and validation of the quantified uncertainties for each potentially influential model is not performed. These methods try in an integral way to evaluate accuracy of selected up-scaled prediction(s) without identification and quantification of all individual model uncertainties. They can help to proceed by validation of the determined model uncertainties but in fact cannot solve all problems of model uncertainties quantification as required in the input uncertainty propagation method. In case of a lack of suitable large- scale experiments, the uncertainty of important models may relay considerably on the state of knowledge of experts.

In the case of a lack of suitable experiments, the new trend is using CFD simulations as a basis for the model uncertainties quantification of system codes. The CFD simulations can be applied as reference for system code simulations or as a tool to generate simulated experimental data (Lewis et al., 2016). Application of CFD simulations instead of experimental data for uncertainties evaluation of 1-D system code models is a relatively new option but of increasing importance. It is caused by increasing simulation possibilities of CFD codes and increasing computational capabilities of computers. CFD simulations can be applied and are already applied as reference for system codes for evaluation of 2D/3D mixing effects by one-phase flow applications. The two-phase flow applications are still a challenge also for CFD simulation, so the quantification of 1-D model uncertainties for two-phase flow conditions is based on experiments. The use of CFD codes requires their careful validation for the addressed application field, before they can be used as reference for system codes uncertainties quantification.

7.6. Consideration of scaling effects by validation of quantified model uncertainties

The problem of validation of model uncertainties in mechanistic computer codes has been the subject of many analyses. The analysis of possible validation procedures is discussed in Chapter 6. A typical recommendation is addition/consideration of new experiments and improvement of measurements or calibration of the simulation models e.g. (Roy and Oberkampf, 2011). Both are only conditionally applicable. For code users (and not developers) calibration of models in the code is a difficult task which may not even be possible. In the face of a lack of experiments in large-scale it can be impossible to add new data to the experimental database.

7.6.1. Studies and experience regarding validation for large scales

The situation in the case of thermal-hydraulic scaling and geometry scaling is different. The dependency of interesting phenomena on changes in thermal-hydraulic conditions like pressure or temperature are usually known. It is a common practice to investigate the dependency of phenomena on thermal-hydraulic parameters for the full range of possible applications. The geometry of the available experiments is mostly small-scale (or medium-scale) and rather only exceptionally equivalent to full-scale nuclear reactors. Therefore, the problem of the up scaling appears to be the main concern of the uncertainty analysis. Investigations performed in the past showed that the findings obtained based on small-scale experiments are applicable only partially for large scales. In the course of the international programme BEMUSE (Best estimate methods – uncertainty and sensitivity evaluation) scaling issues were investigated, among other topics (NEA, 2011). Within the project, uncertainty and sensitivity analyses of a 2F LB LOCA experiment at LOFT test facility (de Crécy et al., 2008) and hypothetical 2F LB LOCA accident at the Zion Nuclear Power Plant (Perez et al., 2010, 2011) were performed. The Zion Nuclear Power Plant was equipped with a Westinghouse type PWR, which is also simulated by the LOFT test facility in the volumetric scale 1:50.

Users of different thermal-hydraulic codes took part in this project. Almost all of them applied a statistic approach based on Wilks' formula. The comparisons showed that even users of the same code obtained different uncertainty limits. In particular, the compared upper bounds of peak cladding temperature for the LOFT test facility and for the Zion Nuclear Power Plant differ considerably among the users of the same code.

However, not only calculated uncertainty limits differed significantly among participants. Input uncertain parameters identified by different participants as potentially influential are also quite different. Furthermore, users of the same code frequently identified different input parameters as the most influential. Even results of the reference calculations obtained by the use of the same code differ considerably in many cases.

It appears that the large discrepancy even between the results of the same codes had two reasons:

- discrepancy between reference calculations;
- differences between considered input uncertainties.

Input uncertainties were identified as the main reason for the large discrepancy in results. Already by the selection of the potentially influential input uncertainties, two different approaches have been applied:

- setting up a Phenomena Identification and Ranking Table (PIRT) to identify the most influential phenomena with the aim of limiting the number of considered input parameters;
- including all potentially important input parameters, which results in a much larger number of uncertain parameters.

The application of different approaches for the identification of potentially influential uncertain parameters led to a very different number of parameters considered by various participants. The next source of discrepancy related to input uncertainties is the quantification of input uncertainties, in particular the quantification of model uncertainties. Large differences appeared in the quantification of model uncertainties of the same code performed by different users.

Moreover, the users of the same code applied Wilks' formula in different orders. Even the same participant frequently applied Wilks' formula in different orders in the analyses of the LB LOCA at the LOFT test facility and the Zion Nuclear Power Plant.

The differences between the various applications make detailed analyses of the scale-up effects consideration during uncertainty analyses performed within the BEMUSE project very difficult. Consequently, statements on extrapolation of the uncertainty and sensitivity results obtained for test facility to nuclear power plants could be hardly formulated.

Taking all the mentioned difficulties into account, an extension of the BEMUSE project for the ATHLET users had been performed by GRS (Skorek, 2009). In this study, the results of uncertainty and sensitivity analyses of the LOFT test facility and Zion Nuclear Power Plant performed by three different institutions, GRS, NRI and AEKI, using the code ATHLET and GRS software SUSA for uncertainty evaluation were compared. In this study, all three analyses were performed independently. However, the information available for the participating institutions concerning the facilities and the events was the same. The inputs for the thermal-hydraulic code ATHLET were developed using only the information distributed among the participants of the project BEMUSE. In this way it was expected to ensure that the differences between the analyses would be related solely to the application of uncertainty and sensitivity analysis methodology, in particular to determination of input uncertainties.

First, the influence of the reference calculation, input uncertainties quantification and order of the Wilks' formula was investigated. The comparison of the uncertainty results of all three independently performed uncertainty analyses showed that the development of the reference input data and definition of the uncertain input parameters influence the results much more than the applied order of Wilks` formula used for uncertainty bounds determination (Skorek, 2009).

The comparison of the determined uncertainty bounds obtained for the LOFT test facility and Zion Nuclear Power Plant has been performed based on the first and second peak cladding temperatures (PCT), the time of hydro-accumulators injection initialisation and time of complete quenching. The comparison shows that the results of reference runs as well as the width of the uncertainty bands are lower for the LOFT test facility than for the

Zion Nuclear Power Plant with exception of the time of accumulator injection. In particular, the time of complete core quenching is much larger for the Zion Nuclear Power Plant.

Interesting observations could be made by comparing the most influential uncertain input parameters evaluated by GRS for LB LOCA at the LOFT test facility and Zion Nuclear Power Plant. The comparison shows that different parameters have been found important for the LOFT test facility and Zion Nuclear Power Plant. Similar observations could be made by sensitivity analyses performed by NRI. Regarding the NRI analysis, only a few parameters could be identified as important for both LOFT and Zion LB LOCA analyses. It confirms the observation that uncertain parameters related to different phenomena can be important for the accuracy of accident simulation at large and small-scale facilities. The lesson is that the findings obtained by scaled medium size (like LOFT) test facilities are not guaranteed to be valid for nuclear power plants.

7.6.2. Issues to be considered by validation for large scales and assessment of predictive capability

For the sake of nuclear reactor safety, it is necessary to simulate many accident scenarios and predict the thermal-hydraulic phenomena that occur in a nuclear power plant using the well-scaled integral effect facilities. However, it is difficult to conduct the tests for various kinds of scenarios; so, the computer code can be generally used for safety analysis. The system codes have been developed from the first principle of fundamental conservation equations for mass, momentum and energy, while the closure relation at wall and interfacial areas may be affected by scaling of experiment. In addition, there will be some scaling distortion in special component models in system code, such as pump and separator, and flow process models such as critical flow, CCFL and so on, because most models were developed in small-scale experiments.

In the face of a lack of large-scale experiments for quantification of model uncertainties, validation of the determined model uncertainties becomes important. In the case of integral test facilities, the main problem remains the same, a lack of large and full-scale experiments. The procedure of validation is based also on extrapolation of findings obtained by validation for small and medium scale facilities for full-scale facilities. The validation procedure plays an important role in comparing the prediction accuracy of facilities scaled down to different dimensions. Calculation of counterpart tests gives information if the simulation of physical phenomena is correct and independent on scales and experimental conditions.

The numerical scheme for solving the equations in the code is an important factor of scaling distortion. In the verification process by experiment simulations, some numerical techniques such as time step control, smoothing option, etc. are used to obtain the best-estimate calculation, which is not always consistent with those of plant analysis because of different components or sizes of nodes. Such a numerical option may not be proved if there is no full-scale experiment. For example, the scaling can be distorted if the small node size is used in the SET assessment, while the large one is used in the plant analysis for realistic reason. To minimise the effect of nodes, it is recommended that the node size of a plant be the same as that of the SET. However, the scaling distortion may be introduced by using the same node in plant calculation when best noding is used to characterise the special phenomena for SET or IET, which are not occurred in the same way in plant. In the BEPU methodology, code uncertainty for scaling should be evaluated through direct experiment data comparison at different scales, and particularly if predictions are non-conservative, the code should be improved and qualified for nuclear power plants (USNRC, 1989b).

Another aspect of the prediction capability is related to limitation of input uncertainty quantification performed for a particular transient (specific approach contrary to generic approach). It can happen that during uncertainty analysis the range of varied calculations extends the assumed range of parameters considered for development of experimental database and following model uncertainties quantification. In such a case, the frequently recommended best solution is performing an iteration step. Beginning with the extension of the experimental database, quantification and finally validation of model uncertainties have to be performed once more.

The main concern of the predictive capability of uncertainty analyses is geometrical and thermal-hydraulic scaling. It means that the identification and quantification of model uncertainties based on selected/available experiments are valid for large-scale geometry and the range of thermal-hydraulic parameters characteristic for nuclear reactors.

While the predictive capability according to a range of thermal-hydraulic parameters is usually enough and sufficient experimental data exists, predictive capability according to geometry effects remains a difficult issue. The problem of the up scaling in the best-estimate thermal-hydraulic simulations is a central problem in nuclear reactor safety. This is a general problem and affects all activities in this field: experimental work as well as numerical analyses. This topic was a subject of intensive investigations and review of the activities and findings in this field were analysed in the framework of NEA projects.

The report, "A state-of the art report on scaling in system thermal-hydraulics applications to nuclear safety and design" (NEA, 2017), addresses issues regarding uncertainties and scaling in the BEPU approach. In such an approach, it is noticed that the uncertainty method can bound the code errors expected in applications of the code at full-scale (i.e. at the nuclear power plant scale). The report gives discussions on the account of the scaling effect in UQ methods included into three established procedures (CSAU, UMAE-CSAU and GRS methodologies). The CSNI report points out: "*The analyses of small-scale facilities could be used for identifying potentially important uncertain input parameters. However, for each application, a careful identification and selection process must be performed. As it has been found in such studies, […] that* ***very different parameters may be influential for small- and large-scale facilities****."* It is also noticed *"The condition for a correct consideration of scale-up effects is to carry out carefully complete uncertainty and sensitivity analyses for each application. The results of uncertainty analyses for small-scale facilities are important source of information and experience, but cannot be directly transformed to large-scale applications. The most important step to be performed in order to consider the scale-up effects is the identification and quantification of input uncertainties, in particular, model uncertainties for large-scale applications. Since some large-scale separate-effect experiments exist, the quantification can be performed in the best way on the basis of comparison with the experimental data."*

Therefore, predictive capabilities and the general safety of nuclear reactors rely on the quality of extrapolation from existing information to the full-scale application of nuclear reactors. As an ultimate measure of ensuring security of the safety analysis of nuclear power reactors, the application of safety margins compensates for a lack of exact information on the uncertainty of the upscaling to the full-scale nuclear power plant applications in nuclear reactor safety.

7.7. Recommendations and open issues

The condition for a correct consideration of scale-up effects is to carry out carefully complete uncertainty and sensitivity analyses for each application. The most important step of the scale-up effects is the identification and quantification of input uncertainties. It is particularly important but also a difficult task for model uncertainties for large-scale applications. Since some large-scale separate effects experiments exist, the quantification can be performed in the best way based on comparison with available experimental data, as far as possible.

Concerning the quantification of model uncertainties based on SETs, the methodology is mature and widely used. Recently a number of methods for the simultaneous quantification of several uncertainty parameters based on CETs have been developed. They are presented in Chapter 5. Some of them have already been successfully applied to uncertainty quantification. The main problems are the semi-empirical closure relations and their uncertainty prediction for the small and large-scale geometry. The results of uncertainty analyses for small-scale facilities are an important source of information and experience but cannot be directly transformed to large-scale application. It must be remembered that physical model uncertainties for small scales can be, and often are, different than for large scales. In the case of a lack of experimental evidence for large scales, the extrapolation of quantified model uncertainties has to be performed carefully and an increase in the uncertainty ranges has to be taken into account.

An optimal way of obtaining reliable model uncertainties would be evaluation of the physical model uncertainties by the code/model developers. In such an approach, inherent quantification of model uncertainties, called advanced validation (Unal et al., 2011) or extended validation (Skorek, 2018) are considered by code developers. The quantification of physical model uncertainties is conducted in the frame of code development and validation. This approach is already applied and some results can be seen in Annex F. However, complete quantification of all relevant models in the code application field is a difficult issue that requires many resources. Even if there were an obvious interest in performing a systematic evaluation of code model uncertainties, it would still require a lot of work and a long time to complete the process.

Once the model uncertainties have been quantified, the propagation of the input uncertainties through the mechanistic codes enables carrying out the best-estimate plus uncertainty analyses for any transient or accident in the field of the code application, as well as for events for which integral tests do not exist. This capability is a clear advantage of the uncertainty estimation method based on input uncertainties propagation (e.g. using Wilks' formula). However, it requires a proper quantification of input uncertainties and sufficient experimental basis of separate effects tests for model uncertainties quantification.

A clear deficiency of performing uncertainty analyses for reactor scale geometries is the lack of large-scale experiments. CFD calculations can be used in some cases, mainly for single-phase flows, as reference for system code simulations. However, in the case of two-phase flow the accuracy of CFD predictions is still limited. Therefore, additional large-scale experiments are important and would help in the quantification and validation of model uncertainties.

References

ATHLET, (2004), *Models and methods*, GRS Rep. P-1, Vol. 4.

Boyack, B.E. et al. (1990), "Quantifying Reactor Safety Margin, Part1: An overview of the code scaling, applicability and uncertainty evaluation methodology", in *Nuclear Engineering and Design*, 119, pp. 1-15.

D'Auria, F. and G. Galassi (2010), "Scaling in nuclear reactor system thermal hydraulics", in *Nuclear Engineering and Design*, 240, pp. 3267-3293.

de Crécy, A. et al., (2008), "Uncertainty and sensitivity analysis of the LOFT L2-5 test: Results of the BEMUSE programme", in *Nuclear Engineering and Design*, 238, pp. 3561-3578.

Lewis, A., R. Smith, B. Williams and V. Figueroa (2016), "An information theoretic approach to use high-fidelity codes to calibrate low-fidelity codes", in *Journal of Computational Physics*, 324, pp. 24-43.

Martinez-Quiroga, V., F. Reventos and J. Freixa (2014), "Applying UPC Scaling-Up Methodology to the LSTF-PKL Counterpart Test", in *Science and Technology of Nuclear Installations*, 2014, Article ID 292916.

NEA (2017), "A state-of the art report on scaling in system thermal-hydraulics applications to nuclear safety and design", NEA/CSNI/R(2016)14, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19744.

NEA (2011), "BEMUSE Phase VI Report – Status report on the area, classification of the methods, conclusions and recommendations", NEA/CSNI/R(2011)4, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19020.

Perez, M. et al. (2011), "Uncertainty and sensitivity analysis of a LB LOCA in a PWR Nuclear Power Plant: Results of the Phase V of the BEMUSE Programme", in *Nuclear Engineering and Design*, 241, pp. 4206-4222.

Perez, M., et al. (2010), "Main Results of Phase IV BEMUSE Project: Simulation of LB LOCA in an NPP", in *Science and Technology of Nuclear Installations*, 2010, Article ID 219294.

Ransom, V.H., W. Wang and M. Ishii (1998), "Use of an ideal scaled model for scaling evaluation", in *Nuclear Engineering and Design*, 186, pp. 135-148.

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

Skorek, T. (2018), "V&V and Uncertainty Quantification of Code Models", in *Proc. of ANS Best Estimate Plus Uncertainty International Conference BEPU2018*, Paper no 110, Lucca, Italy.

Skorek, T. (2009), "Uncertainty and Sensitivity Analyses of Experiment and NPP Accidents: Large Break LOCA at Cold Leg of Zion Nuclear Power Plant and Comparison with LOFT Test L2-5", in *Proc. of the 13th Intern. Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH13)*, Kanazawa, Japan.

Skorek, T., B. Krzykacz-Hausman and H. Austregesilo (2011), "Investigation of the Uncertainty of Governing Equation Systems in the Thermal-Hydraulic Calculation", in

Proc. of the 14th Int. Topical Meeting on Nuclear Reactor Thermal Hydraulics, NURETH14, Toronto, Canada.

Unal, C., B. Williams, F. Hemez, S.H. Atamturktur and P. McClure (2011), "Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors", in *Nuclear Engineering and Design*, 241, pp. 1813-1833.

USNRC (2005), *Regulatory Guide 1.203, Transient and Accident Analysis Methods*, December 2005.

Boyack, B., R. Duffey, G. Wilson, P. Griffith, G. Lellouche, S. Levy, U. Rohatgi, W. Wulff, W. and N. Zuber (1989a), "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, United States.

USNRC (1989b), *Regulator Guide 1.157, Best-Estimate Calculations of Emergency Core Cooling System Performance*, May 1989.

Weiss, P., M. Sawitzki and F. Winkler (1986), "UPTF, a full-scale PWR loss-of-coolant accident experiment program", in *Atomkernenergie Kerntechnik*, 49; pp. 61-67.

Wulff, W. N. Zuber, U. Rothagi and I. Catton (2005), "Application of Fractional Scaling Analysis (FSA) to loss of Coolant Accidents (LOCA), Part 2: System Level Scaling for System Depressurization", in *Proc. of NURETH-11*, Paper 111, Avignon, France.

Zuber, N. U. Rothagi, W. Wulff and I. Catton (2005), "Application of Fractional Scaling Analysis (FSA) to loss of Coolant Accidents (LOCA), Part 1: Methodology Development", in *Proc. of NURETH-11*, Paper 153, Avignon, France.

Zuber, N. et al. (1998), "An Integrated Structure and Scaling methodology for Severe Accident Technical Issue Resolution: Development of Methodology", in *Nuclear Engineering and Design*, 186, pp. 1-21.

8. Conclusion

8.1. Outcome of the SAPIUM project

This project was devoted to the development of a systematic approach for model input uncertainty quantification based on inverse propagation of the information associated to the discrepancy between simulation results and experimental data, using verified and validated simulation models. The starting point of the work performed in SAPIUM was the available state of knowledge from previous related NEA projects and current practices in regulation, industry and research. The main outcome is a first good practice document that offers a general framework to analyse IUQ and reduces expert judgement.

The SAPIUM approach was split into five elements described in separated chapters and summarised by Figure 5.4.

It clearly emphasised that IUQ should not be reduced to the application of inverse methods, as was the case during the PREMIUM activity. It is a more general process that involves a clear specification of the problem and efficient strategies to construct an adequate experimental database and to combine the information from different experiments. It also requires assessing the simulation model before quantifying input uncertainties.

Finally, the validation of the quantified model input uncertainties has to be taken into account in the whole process in order to check the acceptability of the results for the intended use. The description of the different elements revealed the need to use mathematical tools integrating the physical knowledge in order to be fully rigorous, transparent and reproducible.

8.2. Main progress compared to previous activities

SAPIUM was proposed to provide an approach that minimises the user effect identified in previous projects. In particular, it allows handling the different sources of user effect from the PREMIUM activity. Table 8.1 recalls them and indicates the corresponding SAPIUM element that can be used to handle them.

Two important topics were also discussed during the PREMIUM project and can be treated in the SAPIUM framework. The first is the validation and acceptability of the quantified model input uncertainties. It can be addressed by Element 5 (Chapter 6), which introduces different types of validation indicators and recommends performing a loop-approach based on the computation of the predictive maturity index in order to reach the acceptability of the results. The second is related to scaling issues. The importance of scaling in IUQ is emphasised in Chapter 7 and some good practices are given to progress on the extrapolation of uncertainty results to nuclear power plants.

Figure 8.1. Flowchart of the SAPIUM approach

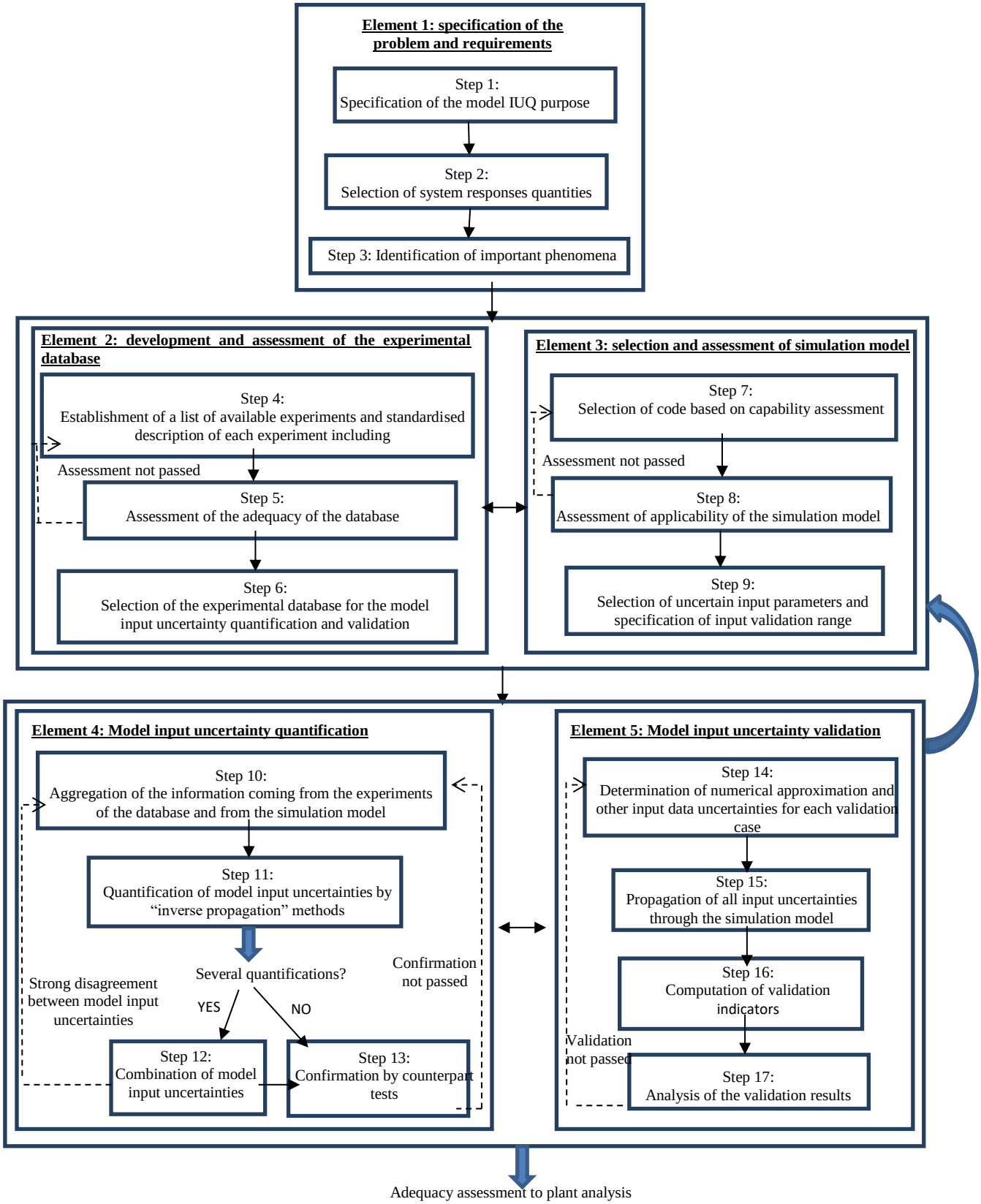


Table 8.1. User effect identified in PREMIUM and SAPIUM elements to handle them

User effect identified from the PREMIUM benchmark	Element of the SAPIUM approach to minimise/reduce the user effect
Selection of the SRQs used for input quantification: Some participants only focused on cladding temperature while others also considered quench times.	Element 1 (Chapter 2) provides recommendations for the problem specification including the selection of SRQs.
Selection of the experimental database: The six available FEBA tests were not taken into account in the quantification step by some of the participants. The lack of adequacy of the FEBA experiment (used for input uncertainty quantification) for a validation on the PERICLES data was also pointed out.	Element 2 (Chapter 3) provides the different tools to establish a standardised list of available experiments and to analyse the adequacy of a given database.
Selection of uncertain input parameters: Some participants did not consider parameters related to interfacial friction. A global heat exchange coefficient multiplier was used rather than several multipliers for each correlation involved in the global heat exchange.	Element 3 (Chapter 4) recalls the different categories of input parameters and a review on sensitivity analysis tools is given as well.
Code modelling and numerical implementation: All participants considered a 1-D modelling for FEBA but 4 participants considered a 3-D modelling for PERICLES. Moreover, three participants have a significantly lower number of meshes in the vertical direction for PERICLES than for FEBA.	Element 3 (Chapter 4) provides the requirements for the code selection and the process to assess the applicability of the simulation model.
Choice of the quantification method: Six methods were used, including different assumptions related to the input uncertainty modelling (interval/PDF, type of PDF, with or without calibration of the reference calculation).	Element 4 (Chapter 5) is mainly devoted to a practical description of available inverse methods including assumptions and tractability for industrial applications. It also provides a comparison between them and recommendations for their use.

8.3. Main recommendations and open issues

The SAPIUM project has led to several recommendations to handle properly each element of the approach. They have fully been described at the end of each chapter. It has also revealed several remaining open issues for a complete application of the proposed IUQ approach. The next section provides a summary of the main recommendations and open issues.

8.3.1. Main recommendations

Element 1 - specification of the problem and requirements

This element is part of the BEPU methodology, and is common to all VVUQ processes. Therefore, the available best practices remain applicable except that the focus is on the model input uncertainty quantification (i.e. IUQ), rather than on the uncertainty analysis of the calculated SRQs of the simulation model.

It is recommended to clearly specify the model IUQ problem for the developed or selected simulation model, according to the transient classification as documented in the FSAR of the targeted nuclear power plant or, in case of new plant design, to specify the application domain for the newly identified transient scenarios. It is recommended to choose the SRQs based on parameters that are directly and accurately measured in the experiments, which are used for verification of the design limits or surrogates in the nuclear power plant accident analysis. It is recommended to group different accident scenarios (e.g. RCS heat-up or cool-down accidents, reactivity-initiated accidents) into a single IUQ problem with common SRQs of interest, and make a generic model input uncertainty quantification and validation for the developed or selected simulation model.

The nuclear power plant accident scenario identification and IUQ definition process can rely heavily on expert opinion and can be subjective. Therefore, iteration of the process, based on experimentation and analysis, is important. It is recommended to use the Phenomena Identification and Ranking Table (PIRT) technique to first identify and rank the physical phenomena, and use the sampling-based global sensitivity analysis (GSA) technique to confirm the PIRT (also called Q-PIRT).

It is important to keep in mind the phenomena ranking could be subject to the limitations of the knowledge of the expert, the experimental databases or simulation models. An iteration with other elements may be necessary if such limitations are identified during the SAPIUM process.

Element 2 - development and assessment of the experimental database

The construction of the experimental database is a key element of the model input uncertainty quantification process since, together with the simulation model, it provides to the mathematical methods of IUQ the information to back-propagate in order to derive input uncertainties. It is recommended to follow a structured and transparent approach to perform this construction.

It first requires collecting all available experiments, including various scales tests that are mandatory for the application to the reactor case. Concerning the question of dependency of the experimental database with respect to the reactor transient, a balanced mixture of specific and generic approach, according to actual application and capabilities, seems to be the best solution.

Then a standardised description of each experiment should be established with a description of each experiment with respect to a fixed set of criteria including the covered phenomena, the geometry, the scaling effect of the experimental facility and the available measurements with the associated uncertainties.

An important issue is the assessment of the adequacy of a database. Two properties of the database contribute to its adequacy: the representativeness of each experiment and the completeness of the database for the intended use. It is recommended to use mathematical tools to quantitatively perform the analysis. The multi-criteria decision making (MCDM) approach can be exploited to evaluate the representativeness. The evaluation of the completeness of the experimental database requires the development of a completeness index that should be restricted to the database and not related to the whole quantification/validation process, as is done classically in VVUQ.

The experimental database should be split in two parts, one for input uncertainty quantification and one for input uncertainty validation. If the number of available

experiments is too limited to perform this split, all available experiments should be considered for the quantification step and the validation step should be adapted.

Element 3 - selection and assessment of the simulation model

The assessment of the applicability of the SM is an iterative process that applies to all experimental tests of the validation database. The possible failure of the process for one test implies the improvement of the SM (either code or nodalisation or both) and the repetition of the assessment of the applicability of an SM for all tests of the validation database.

The validity of a simulation model is defined over the domain of model form, inputs, parameters and responses. This fact effectively limits use of the model to the particular application for which it was validated; use for any other purpose would require the assessment of the applicability of the SM to be performed again.

The SM should not be tuned to a particular data set and the data used to assess the SM should not be deliberately selected to make the SM appear to be more accurate than it truly is.

An important aspect of the simulation model assessment is the nodalisation strategy and model option selection. They should be consistent between the experiment and the nuclear power plant.

In the assessment process, special attention should be devoted to the construction of consistency indicators to evaluate the accuracy between simulation and experiment.

The selection of important uncertain input parameters (including nature of uncertainties e.g. aleatory, epistemic) should be confirmed by sensitivity analysis methods in order to reduce the subjectivity by expert judgement.

Element 4 - model input uncertainty quantification

It is recommended to use, as far as possible, separate effects tests (SETs) where a single phenomenon is investigated and the uncertainty associated with the model of the phenomenon can be usually related to a singular measurement. In this case, the input model uncertainty can be estimated using a simple mathematical method through the comparison of the test simulation with the singular measurement from the SET (for a representative result).

If there is more than one phenomenon (and model) of importance involved in the available experiment, the application of advanced inverse methods is necessary. These methods can also be used to calibrate a model (simultaneously or not with quantification). The experience in PREMIUM indicates that the IUQ should be performed without recalibration.

A special attention should be devoted to the set of pairs of simulated/experimental value used in the back-propagation and that can have different degrees of importance for the intended study. Inverse methods should allow assigning different weights to different pairs in order to avoid under or overweighting the influence of a given one.

The assumptions associated with the inverse problem mathematical methods (type of method, uncertainty modelling to handle aleatory and epistemic uncertainties, etc.) should be clearly taken into account to evaluate the impact of the analyst’s choices on the results. Moreover, depending on the problem to solve, users may choose between probabilistic or non-probabilistic modelling (i.e. based on alternative uncertainty theories). For instance, an inverse problem where the noise has a known probability distribution (e.g. Gaussian) can benefit from probabilistic methods. In presence of incomplete knowledge on input

uncertainties, it is advisable to combine different methods to avoid formulating extra assumptions.

Element 5 - model input uncertainty validation

If the experimental database is large enough, validation experiments should correspond to experiments, which are not used for the uncertainty quantification. When the number of experiments is not sufficient to split the database, the validation should be adapted by performing for example a cross-validation.

It is also important to control (or evaluate) the impact of the methodological assumptions on the validation result. These assumptions are first related to the input uncertainty propagation method and especially to the input uncertainty modelling that should integrate the state of knowledge on uncertainties (e.g. aleatory/epistemic uncertainties). They also concern the construction of validation indicators that can rely on strong assumptions on experiments (e.g. independence of experimental conditions) that are not always satisfied in practice. This should be taken into account when the acceptability of uncertainty results is checked for the intended use by comparison to acceptability thresholds.

Validation requires the selection of a validation indicator. It is recommended to clearly state the choice of the target quantity to validate defining the SRQ uncertainty and the important characteristics of this uncertainty to capture for validation before starting the study. Most of the indicators are focused on a consistency check between simulation and experiment. However, it is advisable to take into account extra features such as the concentration of the information provided by the uncertainty analysis.

The computation of the validation indicators is not sufficient if the objective is application to nuclear power plants. It is recommended to construct a maturity model combining the acceptability of the validation results in the validation domain with the adequacy of the experimental database. The evaluation of the predictive maturity should be carried out in a loop approach. If the lack of maturity can be explained by a lack of adequacy of the experimental database, it is preferable to extend the database by considering integral experiments for the quantification and to iteratively revise the quantified model input uncertainties from these new data.

Scaling issues and predictive capability assessment

The condition for a correct consideration of scale-up effects is to carry out carefully complete uncertainty and sensitivity analyses for each application. The most important step by consideration of the scale-up effects is the identification and quantification of input uncertainties. It is a particularly important but also difficult task for model uncertainties for large-scale applications. Since some large-scale separate effects experiments exist, the quantification can be performed in the best way based on a comparison with available experimental data, as far as possible.

Concerning the quantification of model input uncertainties based on SETs, the methodology is matured and widely used. A number of methods for simultaneous quantification of several model input uncertainties based on combined effect tests (CETs) have been developed. The main problems are the semi-empirical closure relations and their uncertainty prediction for the small- and large-scale geometry. It must be remembered that physical model uncertainties for small scales can be different from for large scales. In the case of a lack of experimental evidence for large scales, the extrapolation of quantified model uncertainties has to be performed carefully and an increase in the uncertainty ranges has to be taken into account.

When lacking large-scale experiments, CFD calculations could be used in some cases, mainly for single-phase flows, as reference for system code simulations. However, in the case of two-phase flow, the accuracy of CFD predictions is still limited. Therefore, additional large-scale experiments are important and would greatly help in the quantification and validation of model uncertainties.

8.3.2. Main open issues

Adequacy of the experimental database

The IUQ process is based on the comparison between simulation and experimental results. Therefore, the quantified model input uncertainties strongly depend on the adequacy of the experimental database. Two main aspects of this topic should be further investigated.

The first is related to the quantitative analysis of a database and to the construction of representativeness and completeness indices. Several generic tools have already been developed but their extension to the framework of IUQ still remains an open issue.

The second concerns the lack of experiments. In the case of a poor validation, the SAPIUM approach provides some tools and recommendations to evaluate if the experimental database should be enlarged (e.g. computation of adequacy indicator, loop approach, cross-validation). However, in practice, it is not always affordable when the number of available experiments is too limited and the few available experiments do not define an adequate experimental database. There is no clear strategy to deal with this last situation and to measure the impact on the acceptability of the quantified model input uncertainties.

Integration of the state of knowledge in the construction of inverse methods

There is a large literature related to the construction of inverse methods. However, the degree of reliability (for the problem under study) of the information provided by each pair experimental/simulation value coming from Elements 2 and 3 of the SAPIUM approach is hardly taken into account. It requires the construction of procedures to combine information with different degrees of importance.

Moreover, most of the available inverse methods are constructed in the probabilistic framework and require the choice of unique PDFs to model uncertainties. In case of incomplete knowledge, the use of alternative uncertainty model could be an interesting strategy. If several work projects have already been proposed to treat the direct problem, the adaptation of probabilistic inverse methods to alternative theory remains a challenge in the framework of nuclear applications. In particular, further investigations are required to tackle the problem of the treatment of epistemic uncertainty by alternative theories for IUQ (possibility, Dempster-Shafer…).

Acceptability of the validation results

Significant efforts should also be devoted to the analysis of the validation results in the validation domain with the objective of extrapolation to the application domain. It first involves the construction of a reliable model for the evaluation of the predictive maturity of the whole process. A general guideline has been provided in the VVUQ framework as well as a scale of maturity, but the adaptation to IUQ and the combination with adequacy indices remain open questions.

The acceptability of the validation results strongly depends on the quantification of other uncertainty sources (numerical approximation error and model form) in the validation process. The contributions of the SAPIUM document are restricted to quantification of

model input uncertainties. However, all the types of uncertainty sources will affect the validation since it operates in the SRQ space. A poor validation that would require iterating the SAPIUM process can therefore be due to other uncertainty sources than model input uncertainties. To avoid any misinterpretation, further work is necessary to understand and evaluate the impact of each category of uncertainty sources on the validation results.

Predictive assessment and extrapolation to the application domain

The main weakness of the predictive capability of application inverted uncertainties quantification in the field of Reactor BEPU analyses is the lack of suitable large-scale experiments. As long as an adequate experimental basis is available and the quantification of the input uncertainties can be proved by validation procedure, the predictive capability for large scales applications is ensured.

However, frequently it is not the case, and there are available only small-scale experiments and the quantified uncertainties needed to be extrapolated for the analyses in the reactor scale. Some efforts were performed to solve this problem by developing integral methods for estimation of scale effects in BEPU analyses, e.g. EMDAP methodology. One possible way, based on scaled integral test, is to extrapolate the behaviour of reactor systems from test facilities to the reactor safety analyses using system codes. Nevertheless, there is no matured extrapolation method that could be recommended without a doubt for such applications.

This is the main issue to be solved. An option for further investigation in this field could be the combination of input uncertainties propagation methods with methods based on output uncertainties extrapolation.

8.4. Further work: demonstration of the SAPIUM approach

The SAPIUM approach can be used to reduce the user effect identified during the PREMIUM activity. However, most of the tools that have been introduced to replace engineering judgement in model input uncertainty quantification have themselves a user effect. It is the case, for example, when a scale of acceptability has to be defined for the validation of the simulation model or of the quantified model input uncertainties, when prior information or thresholds have to be chosen. Moreover, as mentioned in the previous section, several open issues remain to be tackled in order to use this approach for industrial applications.

Therefore, exercises demonstrating the SAPIUM approach are encouraged. These could be the subject of a follow-up project. More precisely, demonstration cases could be conducted, with the objective of verifying:

- the applicability of the best practices;
- that whatever the problematic, the best practices can deal with it;
- the issues identified during PREMIUM can be avoided by following the SAPIUM best practices.

For that, these demonstrations could be split into the following three steps:

- First step: The experimental database contains enough data and there are few influential phenomena. The problem is well posed and seems easy to treat.
- Second step: In these cases, several influential phenomena are involved, but the experimental databases provide enough data of different types. An example of this

case is the critical flow at the break, which is of significant relevance during a LOCA.

- Third step: Several influential phenomena are present, but the experimental database does not contain enough varied data (such as in PREMIUM benchmark).

This progressive approach could help to treat the key issues identified, gradually use the recommendations developed in SAPIUM and give feedback on the proposed approach. This is recommended as a follow-up project to the SAPIUM project, and may lead to a revision of the current SAPIUM document.

ANNEX A - Introduction (Chapter 1)

A.1 Milestones

Table A1. Milestones of the SAPIUM project

Task	Milestone
Start of the project and "kick-off" meeting (organisation of the writing group)	26-27 January 2017
1st meeting of the writing group, discussion on the additional simple demonstration case studies and description of key issues for each chapter	29-30 May 2017
2rd meeting of the writing group, discussions on the first draft of the document	21-22 March 2018
Presentation of the guidelines proposal to WGAMA for discussion	17 September 2018
3th meeting of the writing group, discussions on the complete version of the document	28-30 November 2018
4th meeting of the writing group, discussions on the final version of the document	29-30 April 2019
Submission to WGAMA for comments	August 2019

A.2 List of key issues and analysis of the answers to the questionnaire

Table A.2. The 16 identified key issues associated to the SAPIUM approach

Key elements	Key issues
Element 1: Specification of the problem and requirements	1) Selection of important phenomena 2) Selection of SRQs
Element 2: Development and assessment of the experimental database	3) Assessment of adequacy of an experiment and of completeness of an experimental database 4) Selection of experiments according to adequacy and completeness criteria 5) Evaluation of experimental uncertainties 6) Selection of experimental SRQs per test 7) Splitting of the experimental database in two separated parts, one for input uncertainty quantification and one for input uncertainty validation
Element 3: Selection and assessment of the simulation model	8) Assessment of applicability of the model for simulating all the tests of the experimental database (nodalisation strategy and model options) 9) Selection of uncertain input parameters (including nature of uncertainties e.g. aleatory, epistemic) and specification of input validation range 10) Construction of accuracy indicators (accuracy simulation/experiment) and definition of a scale of accuracy
Element 4: Input uncertainty quantification	11) Aggregation of the information coming from different experiments to be used in the "inverse propagation", if considered experiments and experimental measurements should have the same weight or if criteria for different weights should be defined/applied. 12) "Inverse propagation" and derivation of input uncertainties 13) Combination of input uncertainties if several quantifications are performed
Element 5: Input uncertainty validation	14) Input uncertainty propagation for experiments belonging to the validation database (input uncertainty modelling and sampling methods) 15) Construction and computation of criteria (also called validation indicators) to evaluate the information on input uncertainties 16) Iteration quantification/validation in a loop-approach

In order to obtain a better overview of existing knowledge and experiences in the field of input uncertainty quantification and validation, a questionnaire was sent to a group of organisations that declared their interest in this topic (including the SAPIUM participants and BelV). All the organisations provided a knowledge ranking (low, medium or high) on existing methods or practices to handle the 16 previously identified key issues. The following ranking definition was used:

- Low (L): no available method or well-known practice to handle the key issue.
- Medium (M): available methods or well-known practices but not mature enough for industrial applications.
- High (H): available methods or well-known practices and mature enough for industrial applications.

They were also asked to indicate the top five issues that, according to them, should be addressed during the SAPIUM project. The SAPIUM approach is not reduced to these issues but the answers to the questionnaire helped at the beginning of the project to distinguish between key issues with well-known methods/practices (requiring just a review) from key issues for which extra methodological developments are necessary. The answers on the top five key issues are exploited to identify issues that deserve more attention.

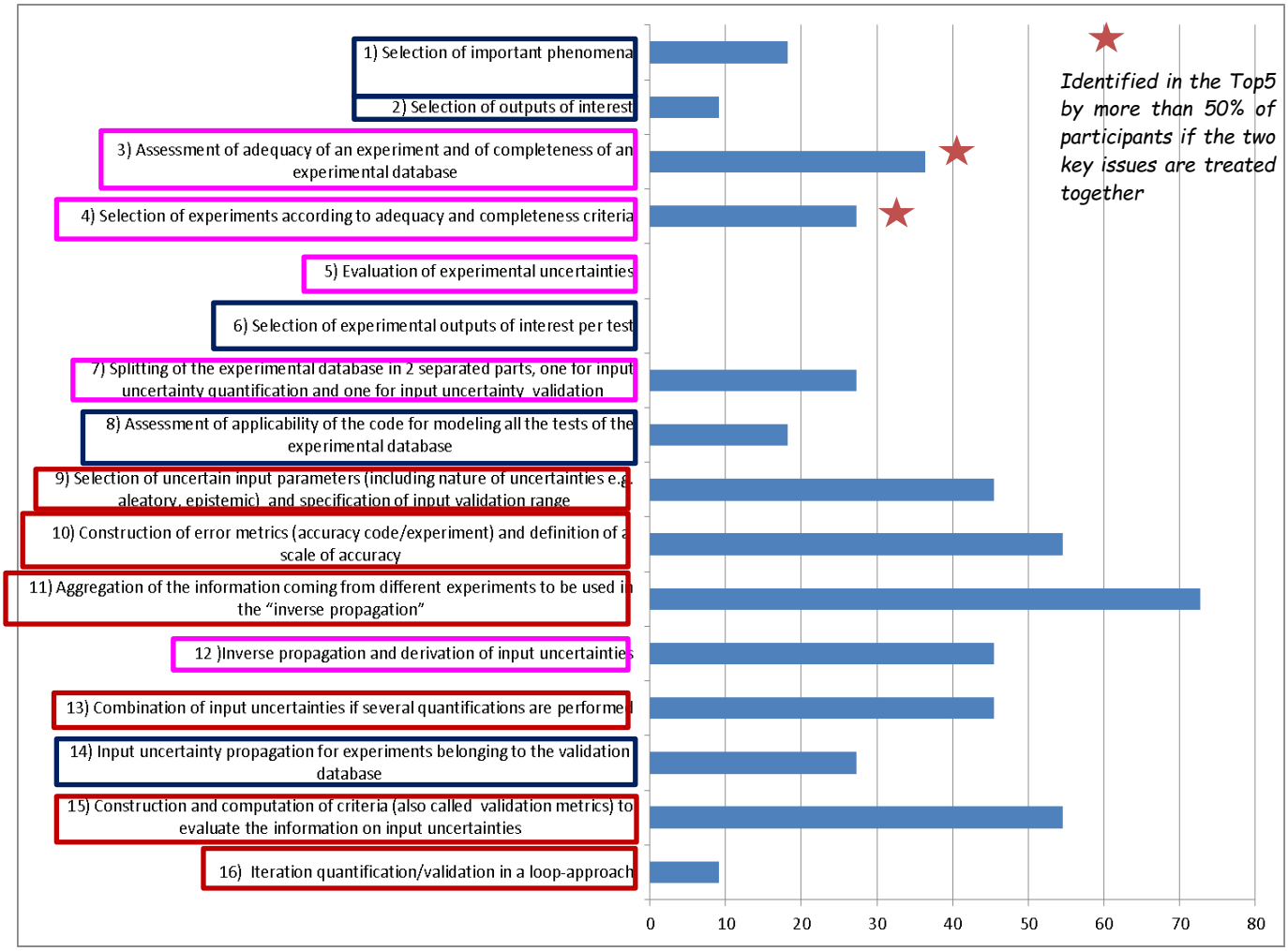
IRSN performed a statistical analysis of the contributions received from each organisation and related to the 16 key issues of Table A.2. Figure A.1 summarises the results.

Three groups of key issues could be identified according to their associated state of knowledge:

- Satisfactory knowledge level (industrial practices are available): key issues related to the first key element of the SAPIUM approach as well as key issues concerning the assessment of applicability of the model for simulating all the tests (Issue 8) and input uncertainty propagation (Issue 14). These two last topics were addressed during the BEMUSE project.
- Existing methods/practices but very few participants consider them as mature for industrial applications: majority of the key issues related to the development and assessment of the experimental database (Element 2).
- Existing methods not mature enough for industrial practices: majority of the key issues related to the selection and the assessment of the simulation model (Element 3) and to input uncertainty quantification and validation (Elements 4 and 5).

These results have been combined in Figure A.1 with the top five key issues provided by each contributor. One could first notice the coherence between the knowledge level ranking on a given key issue and its identification as among the top five issues. More precisely, poor (resp. satisfactory) knowledge leads to high (resp. low) importance to be addressed during SAPIUM. Moreover, the top five key issues identified by the majority of contributors belong to Element 3 (simulation model, Issue 10), Element 4 (quantification, Issue 11) and Element 5 (validation, Issue 15). To a less extent, Issues 3 and 4 related to the construction of the experimental database (Element 2) also appear.

Figure A.1. Combination of knowledge level ranking and identified top five key issues. The x-axis represents the percentage of participants that considers a given key issue among the top five ones.



ANNEX B - Description of integral-effect tests with the LSTF (Chapter 3)

In Chapter 3, the criteria to describe an experiment are shown in Table 3.1. Among several types of experiments, integral-effect tests (IETs) are employed for input uncertainty validation. This Annex gives a description of the IETs with the large-scale test facility (LSTF) (The ROSA-V Group, 2003) as an example of the selected IETs database.

The LSTF of the Japan Atomic Energy Agency (JAEA) is the world's largest integral test facility designed to investigate thermal-hydraulic responses during accidents and transients in a pressurised water reactor (PWR). The LSTF simulates a typical 3423 MW (thermal) Westinghouse-type four-loop PWR with a two-loop system model by full height and 1/48 in volume. Important thermal-hydraulic phenomena are greatly dependent on the accident or transient scenarios.

JAEA selected typical experimental database of IETs from past tests using the LSTF, in view of covering important phenomena that may affect peak cladding temperature (PCT) as a safety-related parameter. These IETs with the LSTF have been useful for the assessment of best-estimate computer codes such as RELAP5, TRACE, CATHARE, and ATHLET.

Seven LSTF experiments chosen by JAEA are described as follows:

1) PWR 5% cold leg small-break LOCA (loss of coolant accident),
2) PWR 1.5% hot leg small-break LOCA with accident management (AM) measures,
3) PWR 1% cold leg small-break LOCA with AM measures,
4) PWR 17% cold leg intermediate-break LOCA,
5) PWR 17% hot leg intermediate-break LOCA,
6) PWR 1% cold leg small-break LOCA without scram,
7) PWR loss-of-feedwater transient without scram.

Test (1) was defined as NEA/CSNI international standard problem (ISP) No.26. Tests (2) and (3) on small-break LOCAs with AM measures were related to counterpart tests with different integral test facilities of LSTF and PKL (Primärkreisläufe Versuchsanlage) in Germany (Umminger et al., 2012). Tests (4) and (5) were classified into intermediate-break LOCAs at cold leg or hot leg. Tests (6) and (7) belonged to anticipated transient without scram (ATWS) focusing on natural circulation under high core power.

Tables B.1 through B.7 indicate the description of each experiment selected by JAEA, referring to Table 3.1 in Chapter 3. For each of the seven LSTF experiments selected by JAEA, publications have some information on the validation of different best-estimate computer codes against the experiment.

Among the seven LSTF experiments chosen by JAEA, Takeda has performed uncertainty evaluations for the four experiments (2), (3), (4), and (5) in terms of the PCT during the small-break LOCAs with the AM measures or the intermediate-break LOCAs (Takeda and Ohtsu, 2017a; Takeda and Ohtsu, 2017b; Takeda and Ohtsu, 2018; Takeda, 2018). For that purpose, first, post-test analysis of the experiment was conducted using the RELAP5 code to assess the code predictive capability. Next, phenomena identification and

ranking table (PIRT) for each component for the small-break LOCA with the AM measure or the intermediate-break LOCA was created in view of the importance of phenomena in determining the PCT. The PIRT was established based on the LSTF test data analysis and the post-test analysis with the RELAP5 code, referring to the PIRTs that had been developed for various kinds of LOCA scenarios (Boyack and Ward, 2000; Griffiths et al., 2014). Finally, uncertainty analysis with the RELAP5 code was carried out to investigate the influences of input uncertain parameters determined by the PIRT on the PCT.

Some examples of the model input uncertainty employed in the uncertainty evaluations are described below. As for critical flow at the break, the break was simulated by utilising a sharp-edge orifice for the small-break LOCA test or a nozzle for the intermediate-break LOCA test, and break discharge coefficient (C_d) was regarded as the relevant uncertain parameter. For the break simulated by the sharp-edge orifice, the model used the Bernoulli incompressible orifice flow equation (Fauske, 1965) for single-phase discharge liquid, and the maximum bounding flow theory (Ardron and Furness, 1976) was applied to two-phase discharge flow. In the base case the values of C_d of 0.61, 0.61, and 0.84 (Sallet, 1984) respectively, were employed for single-phase discharge liquid, two-phase discharge flow, and single-phase discharge steam. For the break simulated by the nozzle, the analysis used the model that has been developed by Ransom and Trapp (1980), and the C_d value was given as 1.0 for the discharge flow in the base case. The uncertain range of the C_d was defined to clarify the effects of differences in the C_d on the PCT. Concerning liquid accumulation in the steam generator (SG) U-tube up flow-side as phenomenon specific to the cold leg intermediate-break LOCA case, slope and intercept of Wallis-type correlation (Wallis, 1969), on counter-current flow limiting (CCFL) at the SG U-tube inlet were selected as the associated uncertain parameters. In the base case the slope and the intercept of the Wallis CCFL correlation were set to 1 and 0.75, respectively, referring to a separate-effect test (SET) with the LSTF focusing on the CCFL at the SG U-tube inlet (Yonomoto et al., 1991). The uncertain ranges of the slope and the intercept of the Wallis CCFL correlation were defined to make clear the effects of differences in the slope and the intercept of the Wallis CCFL correlation on the PCT. Regarding the decay heat of fuel rods, the core decay power was chosen as the related uncertain parameter, and the specified value was given in the base case. The uncertain range of the core decay power was defined based on the measurement uncertainty (The ROSA-V Group, 2003).

Table B.1. Description of LSTF test on PWR 5% cold leg small-break LOCA

Criteria/items	LSTF test denoted as SB-CL-18 (conducted in 1988)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• 5% cold leg small-break LOCA • Total-failure of high-pressure injection system
Range of main parameters	• Primary pressure; max. 15.5 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 740 K
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1064 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Critical flow; steam condensation on coolant of accumulator (ACC) system; coolant injection from ACC system; unbalance in coolant holdup between upflow and downflow sides of steam generator; loop seal clearing; core two-phase mixture level; core heat transfer; core boil-off; core quench
Covered model	Critical flow model; Ransom and Trapp model, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through break
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact NEA Data Bank
Available documents/reports	See research report (Kumamaru et al., 1989)
Publications	See research reports (Kukita et al., 1992; Lee et al., 1993)
Remark	OECD/NEA/CSNI international standard problem (ISP) No.26

Table B.2. Description of LSTF test on PWR 1.5% hot leg small-break LOCA with AM measures

Criteria/items	LSTF test denoted as SB-HL-18 (conducted in 2011)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• 1.5% hot leg small-break LOCA • Total-failure of high-pressure injection (HPI) system • In high-pressure phase test, manually inject coolant from HPI system into vessel upper plenum for core cooling at max. cladding surface temperature of 750 K • In low-pressure phase test, start secondary-side depressurisation by opening relief valves in both steam generators (SGs) at max. core exit temperature of 623 K with auxiliary feedwater injection into SG secondary-side, as accident management measures
Range of main parameters	• Primary pressure; max. 15.5 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 780 K (in high-pressure phase test), max. 822 K (in low-pressure phase test)
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1008 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Critical flow; steam condensation in SG U-tubes; steam condensation on coolant of accumulator (ACC) and low-pressure injection (LPI) systems; steam discharge through SG relief valve; coolant injection from ACC and LPI systems; core two-phase mixture level; core heat transfer; core boil-off; core quench
Covered model	Critical flow model; Ransom and Trapp model, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through break
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact NEA Data Bank
Available documents/reports	See research report (NEA, 2017)
Publications	See research papers (Freixa et al., 2015; Carlos et al., 2016; Takeda, 2018)
Remark	NEA ROSA-2 Project; low-pressure phase test as counterpart test in NEA PKL-2 Project

Table B.3. Description of LSTF test on PWR 1% cold leg small-break LOCA with AM measures

Criteria/items	LSTF test denoted as SB-CL-32 (conducted in 1996)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• 1% cold leg small-break LOCA • Total-failure of high-pressure injection system • Start secondary-side depressurization of both steam generator (SG) to achieve depressurization rate of 200 K/h in primary system 10 min after break with auxiliary feedwater injection into SG secondary-side, as accident management measures
Range of main parameters	• Primary pressure; max. 15.5 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 772 K
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1008 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Critical flow; steam condensation in SG U-tubes; steam condensation on coolant of accumulator (ACC) and low-pressure injection (LPI) systems; steam discharge through SG relief valve; coolant injection from ACC and LPI systems; loop seal clearing; core two-phase mixture level; core heat transfer; core boil-off; core quench
Covered model	Critical flow model; Ransom and Trapp model, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through break
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact Japan Atomic Energy Agency (JAEA)
Available documents/reports	See research report (Takeda, 2014)
Publications	See research paper and report (Takeda and Ohtsu, 2017a; NEA, 2018)
Remark	Approval as counterpart test in NEA PKL-3 Project

Table B.4. Description of LSTF test on PWR 17% cold leg intermediate-break LOCA

Criteria/items	LSTF test denoted as IB-CL-03 (conducted in 2010)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• 17% cold leg intermediate-break LOCA • Actuations of high-pressure injection (HPI), accumulator (ACC), low-pressure injection (LPI) systems in intact loop only • Single-failure of HPI and LPI systems
Range of main parameters	• Primary pressure; max. 15.5 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 978 K
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1008 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square-root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Critical flow; steam condensation on coolant of HPI, ACC and LPI systems; steam discharge through steam generator (SG) relief valve; coolant injection from HPI, ACC and LPI systems; loop seal clearing; liquid accumulation in U-tube upflow-side and upper plenum of SG, and upper plenum due to counter-current flow limiting (CCFL); core two-phase mixture level; core heat transfer; core dryout; core quench
Covered model	• Critical flow model; Ransom and Trapp model, etc. • CCFL model; Wallis-type correlation, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through break
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact NEA Data Bank
Available documents/reports	See research report (NEA, 2017)
Publications	See research papers (Takeda et al., 2012a; Freixa et al., 2013; Takeda and Ohtsu, 2017b)
Remark	NEA ROSA-2 Project experiment

Table B.5. Description of LSTF test on PWR 17% hot leg intermediate-break LOCA

Criteria/items	LSTF test denoted as IB-HL-01 (conducted in 2009)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• 17% hot leg intermediate-break LOCA • Total-failure of high-pressure injection system • Design flow rates of accumulator (ACC) and low-pressure injection (LPI) systems to become 3:1 to cold legs in loops with and without pressurizer
Range of main parameters	• Primary pressure; max. 15.5 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 607 K
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1008 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Critical flow; steam condensation on coolant of ACC and LPI systems; steam discharge through steam generator (SG) relief valve; coolant injection from ACC and LPI systems; loop seal clearing; liquid accumulation in upper plenum due to counter-current flow limiting (CCFL); core two-phase mixture level; core heat transfer; core boil-off; core quench
Covered model	• Critical flow model; Ransom and Trapp model, etc. • CCFL model; Wallis-type co-relation, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through break
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact NEA Data Bank
Available documents/reports	See research report (NEA, 2017)
Publications	See research papers (Freixa et al., 2012; Takeda et al., 2012a; Takeda and Ohtsu, 2018)
Remark	NEA ROSA-2 Project experiment

Table B.6. Description of LSTF test on PWR 1% cold leg small-break LOCA without scram

Criteria/items	LSTF test denoted as SB-CL-38 (conducted in 2006)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• 1% cold leg small-break LOCA without scram • Total-failure of high-pressure injection system
Range of main parameters	• Primary pressure; max. 15.5 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 903 K
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1008 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Critical flow; steam discharge through steam generator (SG) relief valve; natural circulation under high core power; supercritical flow; liquid accumulation in U-tube upflow-side and upper plenum of SG due to counter-current flow limiting (CCFL); core two-phase mixture level; core heat transfer; core boil-off
Covered model	• Critical flow model; Ransom and Trapp model, etc. • CCFL model; Wallis-type co-relation, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through break
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact NEA Data Bank
Available documents/reports	See research report (NEA, 2013)
Publications	See research papers (Takeda et al., 2009; Gallardo et al., 2012a; Martinez et al., 2012a)
Remark	NEA ROSA Project experiment

Table B.7. Description of LSTF test on PWR loss-of-feedwater transient without scram

Criteria/items	LSTF test denoted as TR-LF-13 (conducted in 2007)
Type	IET
Working fluid	Steam/water
Material properties	Electrically-heated rod cladding of Inconel 600; density, specific heat, thermal conductivity
Component and/or reactor	Overall system of PWR
Experimental conditions	• Loss-of-feedwater transient without scram • Total-failure of high-pressure injection system
Range of main parameters	• Primary pressure; max. 16.2 MPa • Core power; max. 10 MW • Cladding surface temperature; max. 873 K
Geometry	• Four primary loops of Westinghouse (WH)-type PWR are represented by two equal-volume loops to simulate two-phase flows. • Full assembly has mostly the same dimensions as those of WH-type 4-loop PWR 17×17 fuel assembly to preserve heat transfer characteristics of core. • Core, 3.66 m in active height, consists of 1008 electrically-heated rods in 24 rod bundles to simulate fuel rod assembly in WH-type 4-loop PWR.
Scale	• Full-height model of WH-type 4-loop PWR • Volumetric scaling ratio of primary loops is 1/48 of WH-type 4-loop PWR. • Time scale of simulated phenomena is one to one to those in WH-type 4-loop PWR. • Flow area in horizontal leg is scaled to conserve ratio of length L to square-root of pipe diameter D; $L/D^{0.5}$ of WH-type 4-loop PWR to better simulate flow regime transitions in primary loops (Froude number basis).
Covered phenomena	Coolant discharge through pressurizer (PZR) power-operated relief valve (PORV); steam discharge through steam generator (SG) relief valve; natural circulation under high core power; liquid holdup in PZR due to counter-current flow limiting (CCFL); liquid level oscillation at SG U-tube due to CCFL; non-uniform flow among SG U-tubes; core two-phase mixture level; core heat transfer; core boil-off; core quench
Covered model	• CCFL model; Wallis-type co-relation, etc.
Validate complete system	• Experimental data are manually qualified through comparison of published ranges and uncertainty values. • Bad trend data among all experimental data are excluded. • Available experimental data are finally obtained.
Available measurements	Pressure; differential pressure; fluid temperature; wall temperature; flow rate; liquid level; fluid density; electric power; pump rotation speed; integrated discharge flow through PZR PORV
Instrumentation	Pressure transducer; differential pressure transducer; thermocouple; flow meter; gamma-ray densitometer; electric power meter; magnetic pickup; level meter
Data access condition	Contact NEA Data Bank
Available documents/reports	See research report (NEA, 2013)
Publications	See research papers (Gallardo et al., 2012b; Martinez et al., 2012b; Takeda et al., 2012b)
Remark	NEA ROSA Project experiment

References

Ardron, K.H. and R.A. Furness (1976), "A study of the critical flow models used in reactor blowdown analysis", in *Nuclear Engineering and Design*, 39, pp. 257–266.

Boyack, B.E. and L.W. Ward (2000), "Validation test matrix for the consolidated TRAC (TRAC-M) code", in *Proc. of International Meeting on Best Estimate Methods in Nuclear Installation Safety Analysis (BE '00)*, Washington DC, United States.

Carlos, S. et al. (2016), "Post-test analysis of the ROSA/LSTF and PKL counterpart test", in *Nuclear Engineering and Design*, 297, pp. 81–94.

Fauske, H.K. (1965), "The discharge of saturated water through tubes", in *AlChE Symposium Series*, 61(59), pp. 210–216.

Freixa, J. et al. (2015), "Modelling guidelines for core exit temperature simulations with system codes", in *Nuclear Engineering and Design*, 286, pp. 116–129.

Freixa, J. et al. (2013), "Post-test thermal-hydraulic analysis of two intermediate LOCA tests at the ROSA facility including uncertainty evaluation", in *Nuclear Engineering and Design*, 264, pp. 153–160.

Freixa, J. et al. (2012), "Thermal-hydraulic analysis of an intermediate LOCA test at the ROSA facility including uncertainty evaluation", in *Nuclear Engineering and Design*, 249, pp. 97–103.

Gallardo, S. et al. (2012a), *Assessment of TRACE 5.0 against ROSA Test 3-1, Cold Leg SBLOCA*, NUREG/IA-0413, USNRC, Washington DC, USA.

Gallardo, S. et al. (2012b), *Assessment of TRACE 5.0 against ROSA Test 3-2, High Power Natural Circulation*, NUREG/IA-0412, USNRC, Washington DC, United States.

Griffiths, M.J. et al. (2014), "Phenomena identification and ranking table for thermal-hydraulic phenomena during a small-break LOCA with loss of high pressure injection", in *Progress in Nuclear Energy*, 73, pp. 51–63.

Kukita, Y. et al. (1992), *OECD/NEA CSNI International Standard Problem No.26 ROSA-IV LSTF Cold-Leg Small-Break LOCA Experiment Comparison Report*, NEA/CSNI/R(91)13.

Kumamaru, H. et al. (1989), *ROSA-IV/LSTF 5% Cold Leg Break LOCA Experiment Run SB-CL-18 Data Report*, JAERI-M 89-027, Japan Atomic Energy Research Institute, Ibaraki, Japan.

Lee, S. et al. (1993), *RELAP5 Assessment Using LSTF Test Data SB-CL-18*, NUREG/IA-0095, USNRC, Washington DC, United States.

Martinez, V. et al. (2012a), *Post-Test Calculation of the ROSA/LSTF Test 3-1 using RELAP5/moD-3.3*, NUREG/IA-0409, USNRC, Washington DC, United States.

Martinez, V. et al. (2012b), *Post-Test Calculation of the ROSA/LSTF Test 3-2 using RELAP5/moD-3.3*, NUREG/IA-0410, USNRC, Washington DC, United States.

NEA (2018), "Summary and Conclusions of the Joint PKL-3-ATLAS Workshop on Analytical Activities", NEA/CSNI/R(2017)10, OECD Publishing Paris, www.oecd-nea.org/jcms/pl_19796.

NEA (2017), "Final Integration Report of the Rig-of-safety Assessment (ROSA-2) Project - 2009–2012", NEA/CSNI/R(2016)10, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19728.

NEA (2013), "Final Integration Report of OECD/NEA ROSA Project 2005–2009", NEA/CSNI/R(2013)1, OECD Publishing, Paris, www.oecd-nea.org/jcms/pl_19244.

Ransom, V.H. and J.A. Trapp (1980), "The RELAP5 chocked flow model and application to a large scale flow test", in *Proc. of the ANS/ASME/NRC International*

Topical Meeting on Nuclear Reactor Thermal-Hydraulics, Saratoga Springs, New York, United States.

Sallet, D.W. (1984), "Thermal hydraulics of valves for nuclear applications", in *Nuclear Science and Engineering*, 88, pp. 220–244.

Takeda, T. (2014), *Data Report of ROSA/LSTF Experiment SB-CL-32 –1% Cold Leg Break LOCA with SG Depressurization and No Gas Inflow–*, JAEA-Data/Code 2014-021, Japan Atomic Energy Agency, Ibaraki, Japan.

Takeda, T. et al. (2012a), "RELAP5 Analyses of OECD/NEA ROSA-2 Project Experiments on Intermediate Break LOCAs at Hot Leg or Cold Leg", in *Journal of Power and Energy Systems*, 6(2), pp. 87–98.

Takeda, T. et al. (2012b), "RELAP5 Analysis of OECD/NEA ROSA Project Experiment Simulating a PWR Loss-of-Feedwater Transient with High-Power Natural Circulation", in *Science and Technology of Nuclear Installations*, 2012, Article ID 957285.

Takeda, T. et al. (2009), "Analysis of the OECD/NEA ROSA Project experiment simulating a PWR small break LOCA with high-power natural circulation", in *Annals of Nuclear Energy*, 36, pp. 386–392.

Takeda, T. and I. Ohtsu (2018), "Uncertainty analysis of ROSA/LSTF test by RELAP5 code and PKL counterpart test concerning PWR hot leg break LOCAs", in *Nuclear Engineering and Technology*, 50, pp. 829–841.

Takeda, T. and I. Ohtsu (2017a), "ROSA/LSTF Test and RELAP5 Analyses on PWR Cold Leg Small-Break LOCA with Accident Management Measure and PKL Counterpart Test", in *Nuclear Engineering and Technology*, 49, pp. 928–940.

Takeda, T. and I. Ohtsu (2017b), "RELAP5 uncertainty evaluation using ROSA/LSTF test data on PWR 17% cold leg intermediate-break LOCA with single-failure ECCS", in *Annals of Nuclear Energy*, 109, pp. 9–21.

Takeda, T. (2018), "ROSA/LSTF test and RELAP5 code analyses on PWR hot leg small-break LOCA with accident management measure based on core exit temperature and PKL counterpart test", in *Annals of Nuclear Energy*, 121, pp. 594–606.

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, Japan Atomic Energy Research Institute, Ibaraki, Japan.

Umminger, K. et al. (2012), "Integral test facility PKL: experimental PWR accident investigation", in *Science and Technology of Nuclear Installations*, 2012, Article ID 891056.

Wallis, G.B. (1969), *One-Dimensional Two-Phase Flow*, McGraw-Hill Book, New York, United States.

Yonomoto, T. et al. (1991), "CCFL characteristics of PWR steam generator U-tubes", in *Proc. of the ANS International Topical Meeting on Safety of Thermal Reactors*, American Nuclear Society, Portland, Pre, United States.

ANNEX C - Selection and assessment of the simulation model (Chapter 4)

C.1 The NEMM Validation Process

A brief summary of each step of the Validation Process, depicted in Figure C.1, is provided hereafter.

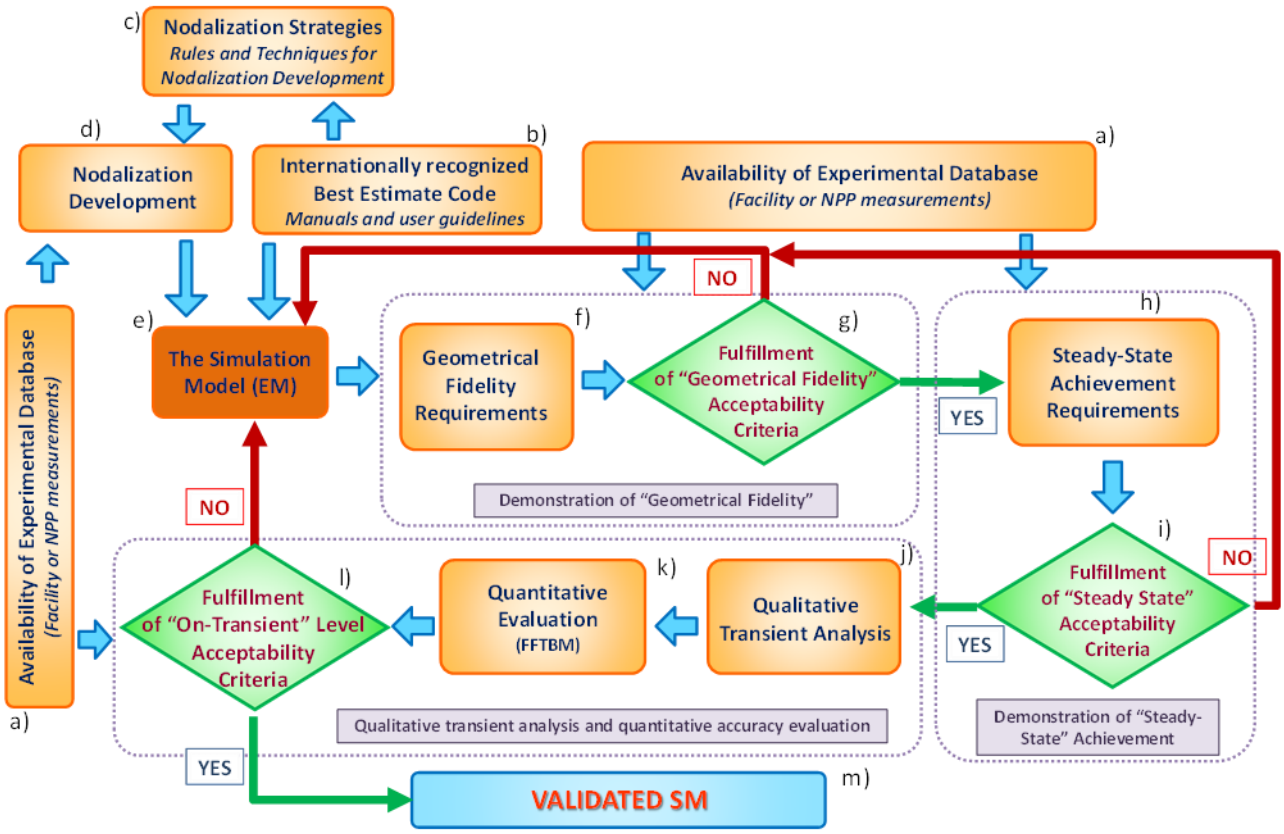
Step "a" (the experimental database)

The availability of a database of experiments for assessing the applicability of the SM is a pre-requisite of the procedure. Criteria to build up the experimental database are part of the procedure and involve mainly aspects related to the representativeness of the experiments with respect to the nuclear power plant, the use of "qualified" experiments in the sense of experiments designed with certain criteria (e.g. power/volume and time preserving features) and the presence of experimental data at different levels of scale to address the scaling issues. This step a) is connected with Element 2 of SAPIUM.

Step "b" (the best estimate code)

An internationally recognised code version must be available and no special deficiencies should have been detected in predicting the phenomena to be considered. The step is also related to the information that is available from the code-user manuals and the guidelines for the use of the code. This type of information takes into account the specific limits and assumptions of the code (specific of the code adopted for the analysis) and provides guidelines about the way to best realise nodalisations. This step b) is connected with Element 3 Step 8 (see Section 4.3 of Chapter 4) of SAPIUM.

Figure C.1. Flow chart of the validation procedure of a SM



From a generic point of view, the following statements should be considered:

- homogeneous nodalisation (i.e. nodalisation developed adopting the same set of criteria and following the same set of guidelines);
- strict observation of the user guidelines;
- standard use of the code (and model) options.

Step "c" (the nodalisation strategies)

Code user experience and code developers' recommendations are useful to set up procedures to be applied for a better nodalisation. These special procedures are related to the specific code adopted for the analysis. For instance, in relation to the use of the RELAP5 code, based on acquired experience achieved through the comparison between experimental and calculated results, the following rules (among many others) are generally adopted by the authors' organisations during the development of the nodalisation:

a) The ratio between the length and the diameter of a node shall be larger than 1.

b) The ratio between the volumes of two adjacent nodes shall be between 0.5 and 2.

c) The ratio between the lengths of two adjacent nodes shall be between 0.5 and 2.

d) To use a standard set of code options.

e) To use more than nine mesh points for simulating the heat structures of the fuel bundles.

f) To adopt special techniques like the "slice" nodalisation to improve the capability of the code to simulate phases of the transient involving with natural circulation phenomena.

However, special cases can be identified where the rules in the above list cannot be respected.

Step "d" (the nodalisation)

The realisation of the nodalisation depends on several aspects like availability of data, code-user capability and experience, and code capability. Data must be qualified and this implies that the database for the realisation of the nodalisation shall derive from:

- Qualified facility (if the analysis is performed for a facility).
- Qualified test design.
- Qualified test data.

In addition, the traceability of each reference shall be maintained in the database. However, three different types of data sources can be distinguished for the preparation of the database:

- Qualified data, which refer the same type of facility or plant under investigation.
- Data derived from similar facility or plant or from other qualified nodalisation for the same type of facility or plant. The use of these data can introduce potential errors and the effect on the calculation results must be carefully evaluated.
- Data assumed by the code user. These types of data constitute assumptions of the code user (based on the experience or on similitude with other similar plants) and their use should be avoided. Any special assumptions adopted by the code user or special solutions in the nodalisation must be recorded and documented.

The SCCRED (Standardised consolidated calculated and reference experimental database) methodology (Petruzzi and D'Auria, 2016) provides a consolidated and systematic approach to develop in a standardised way the relevant source of data that characterise the experimental test and the associated facility as well as the features and the rationales at the basis of the developed SM (see also Element 2 of SAPIUM).

Step "e" (the simulation model (SM))

The simulation model (SM), made by the best estimate code and the developed nodalisation, must reproduce all the relevant parts of the reference facility/plant - including geometrical data and material compositions - and the reproduction of the systems and of the related logics. A SM representing an actual system (facility or nuclear power plant) is validated when:

- It has a geometrical fidelity with the involved system (see steps f-g).
- It reproduces the measured nominal steady state condition of the system (see steps h-i).
- It shows a satisfactory behaviour in time dependent conditions (see steps j-k-l).

Step "f" (the geometrical fidelity requirements)

This step is related to the demonstration that the developed nodalisation has a geometrical fidelity with the involved system (facility or nuclear power plant). Hardware

and modeled geometrical parameters are selected based on the established geometrical fidelity requirements as in the second column of Table C.1. This step is related with the derivation of the values of the geometrical features of the hardware and the estimated numerical values implemented in the nodalisation.

Step "g" (fulfillment of the geometrical fidelity acceptability criteria)

Hardware and modeled geometrical values are compared in order to satisfy the acceptability criteria set-up as in the third column of Table C.1. This validation checking point should be performed by an analyst different from the code user who has developed the nodalisation. The relevant geometrical values (e.g. volume, heat transfer area, elevations, etc.) of the hardware are identified (step "e") and compared with the values implemented into the nodalisation (Step "f").

Table C.1 lists thirteen categories of geometrical parameters to be checked against acceptable criteria. For each category in Table C.1, acceptable errors (AE) have been set-up and listed in the right column. Several parameters can be identified per each category, in relation to the level of detail of the qualification, the user wish to apply (e.g. the category #5 "Non-active structure heat transfer volume" can include only one parameter – the total volume in the whole facility – or as many parameters as the different structures in the facility – core barrel, RPV wall, primary piping etc...). The category #3 "Volume vs height curve (i.e. "local" volume per each circuit)" needs additional explanations: it implies building both for the hardware system and for the modeled one a curve that represents the amount of geometrical volume of each circuit (e.g. primary and secondary circuits) below a certain elevation. The criterion imposes a maximum difference of 5% among the two curves (hardware system and modeled one) at any elevation except for the topmost position, where the Item 1 and 2 in Table C.1 impose a maximum difference of 1% and 2% for the primary and secondary circuit respectively.

The path "from g to e" must be activated if any of the acceptability requirements of the geometrical fidelity is not fulfilled. In this particular case, the nodalisation element of the SM should be improved.

Table C.1. Demonstration of geometrical fidelity acceptable errors

#	CATEGORY OF GEOMETRICAL PARAMETERS	ACCEPTABLE ERROR [(a)]
1	Primary Circuit Volume (Component and Overall)	1%
2	Secondary Circuit Volume (Component and Overall)	2%
3	Primary/Secondary Circuit Volume Vs Elevation Curve	5%
4	Active Structure Volume	0.2%
5	Non-Active Structure Volume (Component and Overall)	10%
6	Primary/Secondary Structure Volume Vs Elevation Curve	10%
7	Active Structure Heat Transfer Area	0.1%
8	Non-Active Structure Heat Transfer Area	10%
9	Flow Area of Components like valves, pumps, orifices	1%
10	Generic Flow Area	10%
11	Component Relative Elevation	0.01 m
12	Primary Circuit Flow Path Length	1%
13	Secondary Circuit Flow Path Length	2%

The % error is defined as the ratio: $100 \cdot |(\text{measured value} \pm \text{measured error}) - \text{calc value}| \,/\, |\text{measured value}|$

The "dimensional error" is the numerator of the above expression.

It shall be noted that the acceptable errors (AEs) in Table C.1 were derived by the developers of the methodology and:

- are based on the engineering judgments derived from the analysis of several integral experiments;
- are applied for all SM developed, independent of the facility type and accident scenario type;
- different developers can derive a different set of AEs based on their engineering judgments derived from the analysis of a consistent set of experiment.

Step "h" (the steady state achievement requirements)

This step is related to the capability of the SM to reproduce the steady state qualified conditions of the system. Experimental and calculated steady state parameters are selected based on the established steady state achievement requirements as in the second column of Table C.2. It shall be considered that experimental measurements are typically available with an error band that must be considered when performing the comparison with the calculated results. No error is made if the calculated value is inside the experimental uncertainty bands. More generally, the error E between a measured Y_E and calculated Y_C values can be calculated by the following formulas where U_E is the measurement uncertainties:

$$\text{if} \quad Y_E - U_E \le Y_C \le Y_E + U_E \quad \rightarrow \quad E = 0 \text{ [19]}$$

$$\text{if} \quad Y_C < Y_E - U_E \quad \rightarrow \quad E = \left(Y_E - U_E - Y_C\right)/Y_E$$

$$\text{if} \quad Y_C > Y_E + U_E \quad \rightarrow \quad E = \left(Y_C - Y_E - U_E\right)/Y_E$$

Measurement uncertainties UE shall take into account different contributions:

o The error due to instrumentation measurement, this is the uncertainty ±2σ which should be provided with each sensors;

o The error due to the time oscillation (or even drift) of the parameters before the transient onset (t = 0 s), this is the value of the two standard deviations connected with the oscillation of the parameter around its average value.

In addition, it shall be carefully considered that SM in SYS-TH are rarely simulating the 3D geometry of the facility, thus an extra error in the evaluation of UE shall be connected with the positions of the sensors, in particular when more than one sensor is close each other and the sensors show different values.

Step "i" (fulfillment of the steady state achievement acceptability criteria)

Experimental and calculated steady state parameters are compared in order to satisfy acceptability criteria set-up as in the third column of Table C.2.

This validation checking-point implies performing a "steady state" calculation. This activity depends on the different code peculiarities. As an example, for the RELAP5 code, the steady state calculation is performed by a "null transient" calculation, which implies adopting the "transient" option without triggering any event (e.g. valve opening or pump

19. If the calculated data point lies within the experimental uncertainty band, there is no possibility to perform a better calculation and thus consistently with the qualification methodology and the "acceptable errors", the associated error is zero.

switching off) that can bring to a transient evolution (i.e. time-dependence) of the thermal-hydraulic parameters.

The relevant thermal-hydraulic parameters of the steady state conditions have to be identified. A thermal-hydraulic parameter is considered as relevant when it is of major relevance to determining the plant behaviour and can be reliably measured. Then, the selected relevant parameters are derived from the results of the steady state calculation (Step "h") for a comparison against experimental parameters (Step "I").

Table C.2 lists sixteen categories of thermal-hydraulic parameters to be checked against acceptable criteria. For each parameter of the categories in Table C.2, acceptable errors (AE) have been set up and listed in the right column. Several parameters can be identified per each category in relation to the level of qualification the user wishes to apply (e.g. the category #7 "fluid temperature" can include the hot leg, the cold leg, the steam generator inlet and out liquid temperatures). The category #6 "local pressure drops" implies building a curve both with the experimental measurements and with the calculated results that represents the pressure drops (ΔP) distribution versus the length of each circuit (e.g. primary and secondary circuits). The criterion imposes a maximum difference of 10% of the difference between the maximum and minimum pressure, in each circuit at any position along the loop.

The values of the thermal-hydraulic parameters listed in Table C.2 to be compared against the experimental measurements are extracted from the end of the steady state calculation, which consists of a null-transient simulation in case the RELAP5 code is used. In this case the convergence of the steady state calculation shall be checked and an additional acceptance criterion is established (acceptance criterion-SS): the inherent drift of each parameter time trend shall be less than 1% over the last one hundred seconds of the steady state calculation.

Table C.2. Steady-state achievement acceptable errors

#	CATEGORY OF THERMAL-HYDRAULIC PARAMETERS [(a)]	ACCEPTABLE ERROR [(b)]
1	Primary/Secondary Circuit Power	1%
2	Primary/Secondary Circuit Power Balance	2%
3	Axial and Radial Power Distribution	1%
4	Heat Losses, PRZ Heaters, Pump Dissipation	10%
5	Absolute Pressure (PRZ, SG, ACC, …)	0.1%
6	Primary/Secondary Circuit Pressure Drops Vs Length Curve	10% [(c)]
7	Fluid Temperature (K)	0.5% [(d)]
8	Primary/Secondary Circuit Fluid Temperature (K) Vs Length Curve	0.5% [(d)]
9	Rod Surface Temperature (K)	10 K
10	Pump Velocity	1%
11	Flow Rates (Primary and Secondary Circuit)	2%
12	Bypass Mass Flow Rates	10%
13	Mass Inventory in Primary Circuit	2% [(e)]
14	Mass Inventory in Secondary Circuit	5% [(e)]
15	Pressurizer Level (Collapsed)	0.05 m [(f)]
16	Secondary Side or Downcomer Level	0.1 m [(f)]

[(a)] With reference to each parameter, the solution must be stable with an inherent drift < 1% / 100 s (acceptance criterion-SS).

[(b)] The % error is defined as the ratio: $100 \cdot |(\text{measured value} \pm \text{measured error}) - \text{calc value}| \,/\, |\text{measured value}|$

The "dimensional error" is the numerator of the above expression.

(c) 10% of the difference between the maximum and minimum pressure in the loop.

(d) The acceptable error shall be consistent with the power error. Fluid Temperature are expressed in Kelvin.

(e) The acceptable error shall be consistent with the other errors (temperatures/densities and levels).

(f) The acceptable error shall be consistent with the other errors (temperatures/densities and pressure drops). Usually in the experiments, the measurements of levels are provided by a correlation with the pressure drops. These correlations are usually not accurate to provide estimation of levels for situations like moving fluids, saturated conditions and thermal stratified conditions.

It shall be noted that the values of the acceptable errors listed in Table C.1 and Table C.2 have been derived based on the engineering judgment of the developers of the methodology. Notwithstanding the subjectivity of the derivation, the following has to be highlighted:

- The engineering judgment has been supported by the analysis of several tens of experiments and associated code predictions. During this process, it was found that when the value of one parameter (in Table C.1 or C.2) was different respect to the reference more than a certain threshold, the code predictions start to differ from the experimental results in a considerable way.
- The values of the acceptable errors are not modified from one application to another; i.e. the engineering judgment has been frozen and does not change with the application.

If one or more than one of the acceptability criteria in Step "i" of the steady-state achievement is not fulfilled (i.e. $E_i > AE_i$ where "I" is a generic parameter) a review of the SM (step "e") must be performed. This process can request more detailed data (Step a), improvement of the nodalisation including user-choices (Step d), different code model-choices (Step b), etc…. The path "from i to e" must be activated until all acceptability criteria in Table C.2 are satisfied (i.e. all $E_i < AE_i$ where "i" is a generic parameter).

Step "j" (the qualitative transient analysis)

This is the first step of the "on transient" level qualification. This activity is necessary to demonstrate the capability of the code and of the developed nodalisation to reproduce the relevant thermal-hydraulic phenomena expected during the transient. This step also makes it possible to verify the correct implementation into the code model of some systems operating only during transient events.

The qualitative transient analysis must be completed before any meaningful attempt to perform the quantitative evaluation and shall address topics like the relevance of the experimental test and facility in respect to the reference nuclear power plant.

The relevant thermal hydraulic phenomena and parameters shall be selected to perform the comparison between calculated and experimental/measured results. The following sub-steps are involved during the qualitative transient analysis:

1. ***Visual observation***: visual comparisons are performed between experimental and calculated relevant parameters time trends.

2. ***Resulting time sequence of events***: the list of the calculated significant events with the corresponding calculated time of occurrences is compared with the experimental events and values.

3. ***Use of the Phenomena specified in the CSNI Validation Matrix (NEA, 1996;NEA, 1994a; NEA, 1994b)***. The relevant phenomena suitable for the code assessment and the relevance of the phenomena in the selected facility and in the selected test can be derived from the CSNI matrix. A judgment can be expressed taking into account the characteristics of the facility, the test peculiarities and the code results.

4. ***Use of the Phenomenological Windows (PhW), Key Phenomena and Relevant Thermal-hydraulic Aspects (RTAs) (D'auria et al., 1995)***. Each test scenario (measured or calculated) shall be divided into Phenomenological Windows (i.e. time spans in which a unique relevant physical process mainly occurs and a limited set of parameters controls the scenario). In each PhW, key phenomena and RTAs must be identified. Key phenomena are attributed to a class of experiments. The lists prepared by the NEA/CSNI are used in the process (NEA, 1996; NEA, 1994a; NEA, 1994b). RTAs are defined as the characterisation of the key phenomena for the specific transient and selected facility and are characterised by numerical values of significant parameters:

- Single valued parameters, SVP (e.g. minimum level in the core).
- Non-dimensional parameters, NDP (e.g. Froude numbering the hot leg at the beginning of reflux condensation).
- Time sequence of events, TSE (e.g. time when dryout occurs).
- Integral parameters, IPA (e.g. integral of break flow rate during subcooled blowdown).
- Derivative parameters, DPA (e.g. derivative of primary and secondary pressure).

Around 20 RTAs, characterised by more than 40 values of significant parameters, must be selected for the qualitative evaluation of a database. Key phenomena and RTAs are used for the following purposes:

- to judge the relevance respect to the scaling and the quality of a test facility (key phenomena);
- to judge the relevance respect to the scaling and the quality of a test design (key phenomena);
- to judge the relevance of an experimental database (key phenomena and RTAs);
- to judge the calculation performance (RTAs);
- to assess the success of a similarity study and of the nodalisation qualification process (RTAs);
- to assess the similarity of different experimental databases (RTAs).

The qualitative transient analysis is finally synthesised by the use of five subjective judgment marks, which are applied to the matrix of phenomena, to the visual observation of the time trends and to the list of RTAs:

- The SM predicts qualitatively and quantitatively the parameter (**Excellent** - the calculation falls within the experimental data uncertainty bands).

- The SM predicts qualitatively, but not quantitatively the parameter (**Reasonable** - the calculation shows only correct behaviour and trends).
- The SM does not predict the parameter, but the reason is understood and predictable (**Minimal** - the calculation does not lie within the experimental data uncertainty bands and does not have correct trends).
- The SM does not predict the parameter and the reason is not understood (**Unqualified** - calculations do not show the correct trend and behaviour, and reasons are unknown and unpredictable).
- Not applicable (-).

Step "k" (the quantitative evaluation)

If the qualitative transient analysis step is acceptable, the degree of consistency between experimental data and predicted results of the SM can be quantified through consistency indicator(s), CI.

Step "l" (fulfillment of "transient" level acceptability criteria)

The fulfillment of the "transient" level acceptability criteria implies that:

- no "unqualified (U) mark" is assigned during the qualitative transient analysis and,
- with reference to the quantitative evaluation, the set of J derived consistency indicators CIj shall be less than fixed thresholds values, TVj:

$$\mathrm{CI_j} \leq \mathrm{TV_j} \quad \text{with } \mathrm{j} = 1, \ldots, \mathrm{J}$$

The path "from "l" to "e" is actuated if one or more than one acceptability criteria (qualitative and/or quantitative) of the transient qualification is not fulfilled. The SM must be improved by changing the nodalisation, adopting different code model-choices or increasing the level of details using new data. It shall be emphasised that every time the SM is modified, a new validation process shall be performed through the loop "from d to e" which implies a new demonstration of geometrical fidelity and steady-state achievement and a new transient qualification.

The validated SM (Item m) can be used to predict nuclear power plant scenario characterised by the same phenomenological windows and key phenomena of the assigned transient.

Step "m" (achievement of a validated SM)

This is the last step of the procedure. The obtained SM consists of a validated nodalisation running on a validated code by a qualified user. The SM can be used to predict plant scenario characterised by the same phenomenological windows and key phenomena of the assigned transient.

It must be pointed out that a modification of the SM (which can be requested for instance to better reproduce the experimental results) requests the application of a new entire assessment of applicability process, i.e. demonstration of geometrical fidelity, demonstration of steady-state achievement and qualitative and quantitative transient analysis.

Quantitative statements about the pedigree of the SM in terms of a) fidelity, respect to the real hardware of the facility, b) achievement of steady state conditions and c) accuracy of both selected single responses, and of the overall prediction are derived for each test of

the experimental database and can be used to judge the assessment of the applicability of the SM for all tests of the experimental database. It shall be clarified that the statement "for all tests" refers to tests of the experimental database that are either performed in the same facility (different boundary and initial conditions) or in very similar facilities (i.e. not characterised by large geometrical distortion) in a such way that the nodalisation, part of SM, is more or less the same.

C.2 The FFTBM Tool

The FFTBM allows for a quantitative judgment of a given simulation of an experimental test by the developed SM. Each set of two curves constituted by a calculated and a measured time trend can be processed by FFTBM. The transformation from time to the frequency domain avoids the dependence of the error from the transient duration. Weight factors are attributed to each time trend to make possible the summing up of the error and the achievement of a unique threshold for accepting a calculation. The quantification of the consistency between experimental and results of the SM must be carried out following demonstration that the SM results are qualitatively acceptable. The same time trends selected for carrying out the qualitative transient analysis (see Item "e" in Figure C.1 of the previous section) shall be utilised as input to the FFTBM.

The quantification of the degree of consistency between a code calculation and an experiment considers the amplitude, in the frequency domain, of the experimental function Fexp(t) and the error function ΔF(t):

$$\Delta F(t) = F_{calc}(t) - F_{exp}(t) \quad \text{(C.1)}$$

To apply the Fast Fourier Transform, a number of values must identify functions, which is a power of two. Thus, if the number of points defining the function in the time domain is N=2m+1, the FFT gives the frequencies:

$$f_n = n/T \quad (n = 0, 1, \dots, 2^m) \qquad \text{(C.2)}$$

In which T is the time duration of the sampled signal.

In particular, the method introduces the definition of two figures of merits: the average amplitude (AA) given in Eq. (C.3) and the weighted frequency (WF) in Eq. (C.4), which provide a synthesis of the information about the error function in Eq. (C.1):

$$AA = \frac{\sum_{n=0}^{2^m} |\tilde{\Delta}F(f_n)|}{\sum_{n=0}^{2^m} |\tilde{F}_{exp}(f_n)|} \quad \text{(C.3)}$$

$$WF = \frac{\sum_{n=0}^{2^m} |\tilde{\Delta}F(f_n)| \cdot f_n}{\sum_{n=0}^{2^m} |\tilde{\Delta}F(f_n)|} \quad \text{(C.4)}$$

The average amplitude represents the relative magnitude of the discrepancy deriving from the comparison between the addressed calculation and the corresponding experimental trend: the lower is the AA – the better is agreement between the experiment and the calculation. The weighted frequency factor characterises the kind of error, because its value emphasises if the error has more relevance at low or high frequencies. Depending upon the transient, high frequency errors can be more acceptable than low frequency ones. In other terms, better accuracy is achieved by low AA values at high WF values.

Trying to give an overall picture of the degree of consistency between a given calculation and the experiment, it is required to combine the information obtained for the

single parameters into average indexes of performance. This is obtained by defining the following quantities: the global average amplitude AA_{TOT} given in Eq. (C.5) and the global weighted frequency WF_{TOT} in Eq. (C.6):

$$AA_{TOT} = \sum_{i=1}^{N_{var}} (AA)_i \cdot (w_f)_i \qquad (C.5)$$

$$WF_{TOT} = \sum_{i=1}^{N_{var}} (WF)_i (w_f)_i \quad (C.6)$$

with

$$\sum_{i=1}^{N_{var}} (w_f)_i = 1 \qquad (C.7)$$

where Nvar is the number of analysed parameters and (wf)i are weighting factors that take into account the different importance of each parameter from the viewpoint of safety analyses. This introduces some degree of engineering judgment that has been fixed by a proper and unique definition of the weighting factors, necessary to account for the different relevance, from the point of view of safety and reliability of the measurement, of the various addressed parameters.

Briefly, each (wf)i takes into account (Bovalini et al., 1992):

- The "experimental reliability": experimental measures of thermal-hydraulics parameters are characterised by a more or less sensible uncertainty due to:
  - intrinsic characteristics of the instrumentation;
  - assumptions formulated in getting the measurement;
  - un-avoidable discrepancies existing between experimental measures and the code calculated ones (mean values evaluated in cross sections, volume centres, or across junctions, etc.).
- The "safety relevance": particular importance is given to the accuracy quantification of calculations concerned with those parameters (e.g. clad temperature, from which peak cladding temperature (PCT) values are derived) which are relevant for safety and design.
- Last, a further contribution is included in the weighting factors definition; this is a component aiming at accounting for the physical correlations governing most of the thermal-hydraulics parameters. Taking as reference parameter the primary pressure (its measurement can be considered highly reliable), a normalization of the AA values calculated for other parameters with respect to the AA value calculated for the primary side pressure is carried out.

The weighting factor $(wf)_i$ for the generic *i*-th parameter, is defined as:

$$(w_f)_i = \frac{(W_{exp})_i \cdot (W_{saf})_i \cdot (W_{norm})_i}{\sum_{i=1}^{N_{var}} (W_{exp})_i \cdot (W_{saf})_i \cdot (W_{norm})_{ji}} \qquad (C.8)$$

Where the specific factors $(W_{exp})_i$, $(W_{saf})_i$ and $(W_{norm})_i$ are:

- $(W_{exp})_i$ is the contribution related to the experimental reliability.
- $(W_{saf})_i$ is the contribution expressing the safety relevance of the parameter.
- $(W_{norm})_i$ is the component of the normalisation vector with reference to the AA evaluated for the primary side pressure.

It shall be noted that while the figures of merit AA and WF are not affected by the user's choices, the definition of the weighting factors, which is needed to evaluate AA_{tot} and WF_{tot},

introduces a degree of engineering judgment that has been fixed by a proper and unique definition of the weighting factors as discussed in Bovalini et al. (1992). Notwithstanding the subjectivity of the derivation, the following has to be highlighted:

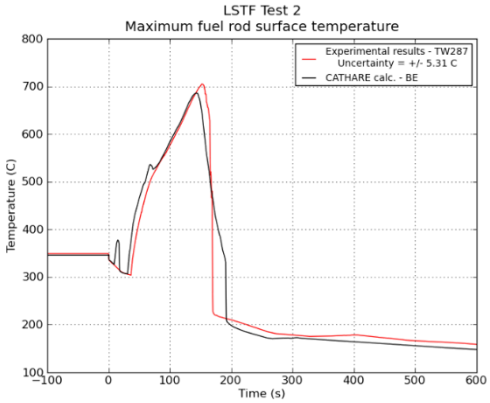
- The engineering judgment has been supported by the analysis of several tens of experiments and associated code predictions.
- The values of the specific factors $(W_{exp})_i$, $(W_{saf})_i$ and $(W_{norm})_i$ are not modified from one application to another; i.e. the engineering judgment has been frozen and does not change with the application independent on the facility type and accident scenario type.
- Different developers of the methodology can derive a different set of specific factors based on their engineering judgments gained from the analysis of a consistent set of experiment.

C.3 Example of a screening sensitivity analysis of a LOCA scenario simulation test-case (Iooss and Marrel, 2019)

A simplified "intermediate break loss of coolant accident" (IBLOCA) scenario is considered that takes into account a double-ended guillotine break with a specific size piping rupture. The numerical model is based on code CATHARE-2 (V2.5_3moD-3.1), which simulates the time evolution of physical quantities during a thermal-hydraulic transient. It models a test carried out on the mock-up "large scale test facility"' (LSTF) in the framework of the NEA/ROSA-2 project, and which is representative of an IBLOCA. This mock-up represents a reduced scale Westinghouse PWR (1/1 ratio in height and 1/48 in volume), with two loops instead of the four loops on the actual reactor and an electric powered heating core (10 MWe). It operates at the same pressure and temperature values as the reference PWR. The simulated accidental transient involves a break on the cold leg and no safety injection on the broken leg. The test under study reproduces a PWR 17% (of cold leg cross-sectional area) cold leg IBLOCA transient with total failure of the auxiliary feedwater, single failure of diesel generators and three systems only available in the intact loop (high pressure injection, accumulator and low pressure injection).

CATHARE-2 is used to simulate this integral effect test. During an IBLOCA, the reactor coolant system minimum mass inventory and the peak cladding temperature (PCT) are obtained shortly after the beginning of the accumulators' injection. Figure C.2 shows the CATHARE-2 prediction, and the experimental values of the maximal cladding temperature (also called maximal heater rod temperature) obtained during the test. This CATHARE-2 modelling of the LSTF allows reproducing the global trends of the different physical phenomena during the transient of the experimental test. In the following, the SRQ will be a single scalar, which is the PCT during the accident transient.

Figure C.2. Experimental values and physical simulation output of the CATHARE-2 model: maximal rod cladding temperature during the transient



The input parameters of the CATHARE-2 code correspond to various system parameters as boundary conditions, some critical flow rates, interfacial friction coefficients, condensation coefficients, heat transfer coefficients, etc. In this study, only uncertainties related to physical parameters are considered, and, no uncertainty on scenario variables (initial state of the reactor before the transient) is taken into account. All uncertain physical models identified in a IBLOCA transient of a nuclear power plant are supposed to apply to the LSTF, except phenomena related to fuel behaviour because of the fuel absence in the LTSF. A physical model uncertainty consists in an additive or multiplicative coefficient associated to a physical model. Finally, d=27 scalar input parameters are considered uncertain and statistically independent of each other. They are then defined by their marginal probability density function (uniform, log-uniform, normal or lognormal). Table C.3 gives more details about these uncertain inputs and their probability density functions (PDF). The nature of these uncertainties appears to be epistemic since they come from a lack of knowledge on the true value of these parameters.

Table C.3 List of the 27 uncertain input parameters and associated physical models in CATHARE-2 code

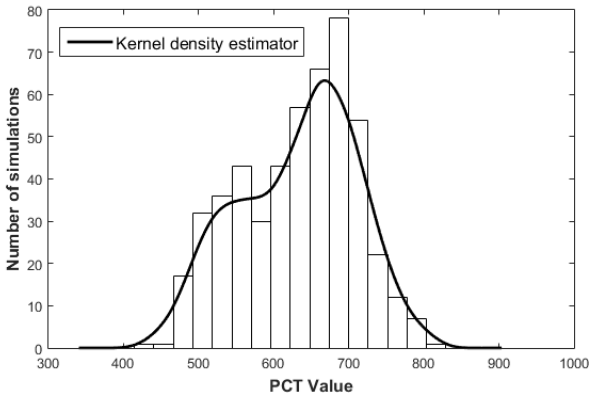
Type of inputs	Inputs	pdf [a]	Physical models
Heat transfer in the core	X_1	N	Departure from nucleate boiling
	X_2	U	Minimum film stable temperature
	X_3	LN	HTC[b] for steam convection
	X_4	LN	Wall-fluid HTC
	X_5	N	HTC for film boiling
Heat transfer in the steam generators (SG) U-tube	X_6	LU	HTC forced wall-steam convection
	X_7	N	Liquid-interface HTC for film condensation
Wall-steam friction in core	X_8	LU	
Interfacial friction	X_9	LN	SG outlet plena and crossover legs together
	X_{10}	LN	Hot legs (horizontal part)
	X_{11}	LN	Bend of the hot legs
	X_{12}	LN	SG inlet plena
	X_{13}	LN	Downcomer
	X_{14}	LN	Core
	X_{15}	LN	Upper plenum
	X_{16}	LN	Lower plenum
	X_{17}	LN	Upper head
Condensation	X_{18}	LN	Downcomer
	X_{19}	U	Cold leg (intact)
	X_{20}	U	Cold leg (broken)
	X_{27}	U	Jet
Break flow	X_{21}	LN	Flashing (undersaturated)
	X_{22}	N	Wall-liquid friction (undersaturated)
	X_{23}	N	Flashing delay (undersaturated)
	X_{24}	LN	Flashing (saturated)
	X_{25}	N	Wall-liquid friction (saturated)
	X_{26}	LN	Global interfacial friction (saturated)

The screening sensitivity analysis process (often called screening) aims at performing the factor fixing setting in order to reduce the complexity (number of uncertain model input parameters) of the problem and to identify the primary influential inputs (PII) on the variability of the SRQ. Once this step is done, a more precise and quantitative sensitivity analysis can be realised and efficient uncertainty propagation techniques can be performed.

In the following, a quantitative screening approach is used. It is based on the kernel-based sensitivity indices called Hilbert-Schmidt independence criterion (HSIC). It allows detecting general complex and non-linear dependence between two random variables.

An initial Monte Carlo sample of CATHARE-2 calculations is required. Then, a Monte Carlo sample of $n = 500$ values for the inputs is defined. After performing the corresponding runs with CATHARE-2, the obtained sample of inputs/outputs constitutes the sample available for the screening step. The histogram of the obtained values for the SRQ, namely the PCT, is given by Figure C.3 (temperature is in °C). A kernel density estimator of the data is also added on the plot to provide an estimator of the probability density function. A bimodality seems to be present in the histogram. It underlines the existence of bifurcation or threshold effects in the code, probably caused by a phenomenon of countercurrent flow limitation between the bend of hot legs and the steam generator inlet plena.

Figure C.3. Histogram of the PCT from the learning sample of *n* = 500 simulations



The bimodality that is observed on the PCT distribution strengthens the use of advanced sensitivity indices (i.e. more general than linear ones or variance-based ones) in the subsequent analysis.

From the estimated HSIC, independence tests can be performed for identifying PII. It shows:

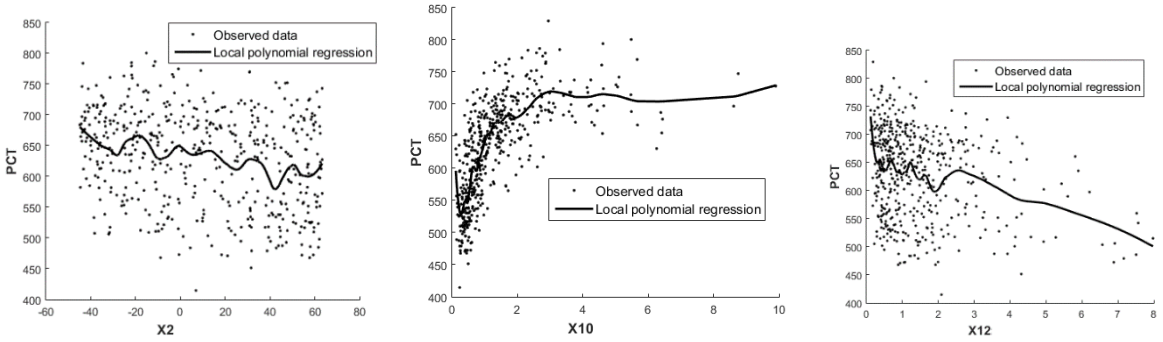
- the large influence of the interfacial friction coefficient in the horizontal part of the hot legs (X_{10});
- followed by the minimum stable film temperature in the core X_2, the interfacial friction coefficient in the SG inlet plena X_{12} and the wall to liquid friction (in under-saturated break flow conditions) in the break line X_{22};
- followed by seven parameters with a lower influence: the interfacial friction coefficients in the upper plenum X_{15}, the down-comer X_{13}, the core X_{14} and the SG outlet plena and crossover legs together X_9, the heat transfer coefficient in the core for film boiling X5, the interfacial friction coefficient of the saturated break flow X_{26} and the condensation coefficient in the jet during the injection X_{27}.

These results clearly underline the predominant influence of the uncertainties on various interfacial friction coefficients.

From the learning sample, some scatterplots of the PCT with respect to some well-chosen inputs (the three most influential ones: X_2, X_{10}, X_{12}) are displayed in Figure C.4. An additional local regression using weighted linear least squares and a first-degree polynomial model (moving average filter) is added on each scatterplot to extract a possible tendency. It is possible to observe that larger values of the interfacial friction coefficient in the horizontal part of the hot legs (X_{10}) lead to larger values of the PCT. This can be explained by the increase of vapour which brings the liquid in the horizontal part of hot legs, leading to a reduction of the liquid water return from the rising part of the U-tubes of the SG to the core (through the hot branches and the upper plenum). Since the amount of liquid water available to the core cooling is reduced, higher PCT are observed. In addition, it has been noticed that there is a threshold effect concerning this input: beyond a value of two, the water non-return effect seems to have been reached, and X_{10} no longer appears to be influential. It was also noted that the minimum stable film temperature in the core (X_2) shows a trend: the more it increases, the lower the PCT. This is explained by the fact that in the film-boiling regime in the core (i.e. when the rods are isolated from the liquid by a film of vapour), X_2 represents

(with a decrease in heat flux) the temperature from which the thermal transfer returns to the nucleate boiling regime. Thus, the larger X_2, the faster the re-wetting of the rods, the faster the cladding temperature excursion is stopped, and thus the lower the PCT.

Figure C.4. Scatterplots with local polynomial regression of PCT according to several inputs, from the learning sample of *n* = 500 simulations



References

Bovalini, R. et al. (1992), *Qualification of the Fast Fourier Transform based methodology for the quantification of thermal-hydraulic system code accuracy*, DCMN NT 194(92), Pisa.

D'auria, F. et al. (1995), "Outline of the Uncertainty Methodology based on Accuracy Extrapolation (UMAE)", in *J. Nuclear Technology*, 109(1), pp. 21-38.

Iooss, B. and A. Marrel (2019), "Advanced methodology for uncertainty propagation in computer experiments with large number of inputs", in *Nuclear Technology*, in press.

NEA (1996), "CSNI Integral Test Facility Validation Matrix for the Assessment of Thermal-Hydraulic Codes for LWR LOCA and Transients", NEA/CSNI/R(96)17, OECD Publishing, Paris.

NEA (1994a), "Separate effects test matrix for thermal-hydraulic code validation. Vol. I Phenomena characterisation and selection of facilities and tests", OECD/GD(94)82, Vol. I, OECD Publishing, Paris.

NEA (1994b), "Separate effects test matrix for thermal-hydraulic code validation. Vol. II: Facility and experiment characteristics", OECD/GD(94)83, Vol. II, OECD Publishing, Paris.

Petruzzi, A. and F. D'auria (2016), "Standardized Consolidated Calculated and Reference Experimental Database (SCCRED): A Supporting Tool for V&V and Uncertainty Evaluation of Best-Estimate System Codes for Licensing Applications", in *Nuclear Science and Engineering*, 182(1), pp. 13-53.

ANNEX D - Model input uncertainty quantification (Chapter 5)

This annex provides more details on inverse problems and methods. A list of significant references on these topics is also given at the end.

D.1 Inverse problems

An introduction to inverse problems

In science and engineering, the term *inverse problem* does not have an undisputed definition. It can be said, in a very general fashion, that inverse problems are aimed at getting information about unknown properties of an object by using indirect and possibly noisy data. The term "indirect" means here that the data do not correspond to the unknown properties that we want to learn about, but from magnitudes that are related to them. Moreover, "noisy" means that data are afflicted by errors, for instance measurement errors.

Sometimes it is said that the object is *recovered* or *reconstructed* from indirect observations, corresponding to magnitudes called responses. In other words, in IPs unknown causes are inferred from known effects

The theory of inverse problems comprises mathematical techniques for obtaining useful information about physical systems based on observations.

IP result from the fact that physical quantities may not be directly accessible for measurement. If we want to learn about them, a possibility is to observe other quantities that are connected with them by physical laws and then try to "invert" such laws. The physical laws have a "natural" direction (from input to output, from cause to effect) which corresponds to the so-called direct or forward problem (DP). The DP can be defined as the "theory" that explains the data observed. In the IP, the response of interest is known and, based on it, one learns about the unknown magnitude. It is important to emphasise this asymmetry between DP and IP: IPs are always more problematic and difficult to solve than DPs.

A typical example of inverse problem is the estimation of a distribution of electrical charges from measurements of the electric field that they produce. The direct problem is the calculation of the electric field induced by a known distribution of electric charges. It is clear that many different distributions of charge may produce the same electric field, and this shows a typical feature of IP. While direct problems have unique solutions, IPs routinely have many (even infinite) solutions.

Two remarks can be made here:

- Even if the quantity of interest can be measured, the errors associated to the measurement process should give rise to an IP. As stated in (Idier, 2008), every data processing chain has an IP associated.

- Instead of *observed* responses it is possible to use *desired* responses. The first case relates to the identification or reconstruction of causes; the second case relates to control or design of causes. Control theory solves inverse problems.

Inverse problems are ubiquitous. Strictly speaking, they are a part of mathematics, but as practical problems, they arise almost in every branch of Science and Engineering. In fact, the idea of *inference* is clearly connected with the inverse problem; "effects" or "consequences" are observed and, from them, their "causes" are inferred or deduced. Science is mainly constructed from inferences; the gathering and interpretation of real data ultimately produces theories and models with the objective of explaining such evidence.

This same Annex includes an overlook of inverse problems in science and engineering.

IP form a very active field of research. Many monographs are devoted to IP. Many scientific journals contain contributions about inverse problems, and specific journals are devoted to the topic (e.g. Inverse Problems, Inverse problems and Imaging, Journal of Inverse and Ill-posed Problems, Inverse Problems in Science and Engineering). As previously stated, for solving an IP it is necessary to have a theory or a predictive model of the data generation, i.e. a model of the object or system under study, and of the process of indirect observation. Such theory is the DP. For this reason, IP are an essential topic in modeling and simulation. Given a predictive model of a specific phenomenology, it is possible to distinguish:

- The forward or direct problem is defined as the calculation of model output from known input parameters. Models are designed and used in order to solve forward problems. Prediction is a forward problem.
- Inverse problems are the estimation of input parameters from known values of a set of model outputs. It is assumed that such model outputs are known because they are measured, or controlled, or postulated magnitudes.

Predictive models transform inputs into outputs. For deterministic models, (those producing always the same output to a given input set), when forward problems have a solution, this solution is unique. However, as previously pointed out, in general IP do not have unique solutions; the model to the same response value can map different assignments of inputs values.

Formally, the deterministic model can be represented as a function or operator F acting on the input x and producing the response y

$y = F(x)$	(D.1)

The unknown input x and the response y can be scalar or multidimensional quantities (finite dimensional IP) or functions (infinite dimensional IP).

The mapping F is sometimes described as the "forward operator", or simply the predictive model. The DP is to determine y from a given x. The IP is to determine values x such that they are transformed by (D.1) in a given value of y.

It is important to notice that y in (D.1) is the response predicted by the model. The real or true value of the response can be written as:

$y_T = F(x) + b(x)$	(D.2)

In (D.2), b(x) is the model bias or model error (sometimes the terms model inadequacy or model discrepancy are used too). The bias is the difference between the real value and the predictive value of the response, and is due to the imperfection, as a predictive tool, of the model. The bias depends on x, but it is known imperfectly (otherwise, it would be used to construct a perfect model). Therefore, it is common to model the bias as an uncertain quantity.

The importance of recognising model discrepancy is emphasised in references (Kennedy and O'Hagan, 2001; Brynjarsdóttir and O'Hagan, 2014), where it is stated that overlooking this term in the analysis of IPs may lead to biased and overconfident estimates of the unknown input parameters.

Expression (D.2) relates the true and the predicted response. Nevertheless, in the IP typically there is an effort to estimate x from measured values of the response, which can be expressed as:

$y_M = y_T + e$	(D.3)

Expression (D.3) states that the measured response is the true response plus the measurement error. Then, the relation between measured and predicted response is found:

$y_M = F(x) + b(x) + e$	(D.4)

Equivalent to:

$y_M = F(x) + \varepsilon$	(D.5)

ε in (D.5) is the difference between measured and predicted response. Usually, it is termed the data noise or observational noise, and is the sum of measurement error and model error. The IP seeks a solution x from the noisy responses in (D.5).

The noise is a very important element in IP. It can be modelled either deterministically or stochastically. In the first case, it is common that the noise "level" is known, but its exact value is unknown. In this case, the IP must be solved for the "worst" noise. Most often, the noise is modelled stochastically, i.e. as a random variable, and then the IP becomes a statistical problem.

When F is a linear function or operator, there is a linear IP, and the solution derives from finding the inverse operator F^{-1}. In a finite dimension, F is represented by a matrix A applied to the vector x to produce the response vector y.

Types of inverse problems

A basic classification of forward and inverse problems is focused on the dimensionality of the data and the unknown, as follows:

- Continuous data – continuous unknown.
- Continuous data – discrete unknown.
- Discrete data – continuous unknown.
- Discrete data – discrete unknown.

Theoretical problems (e.g. Fredholm integral equations) are examples of continuous-continuous IP. In practice, there is a finite number of data, and furthermore the computer implementation is discrete. Therefore, the consideration of data or unknown as continuous is an idealisation, and can be useful as leading to simplified analyses.

Inverse problems can be overdetermined or underdetermined depending if the number of equations is higher or lower than the number of unknowns. An overestimated IP has too much information, so that it may have no solution. However, commonly IP have a lack of information and are underdetermined. In that case, there may be many solutions.

Ill-posedness

One of the distinctive features of inverse problems is ill-posedness. A problem is termed well posed (in the sense of Hadamard) when it fulfills the three following conditions:

I. EXISTENCE: the problem has a solution.

II. UNIQUENESS: the solution is unique.

III. CONTINUITY: the solution depends continuously on the data.

A problem is ill posed when it is not well-posed.

In general, IPs do not have unique solutions; the model into the same response value can map different assignments of inputs values. No uniqueness may be seen because of scarcity of data and information. The general procedure to restore uniqueness is to inject more information to the problem, e.g. prior information about the solution.

Referring to expression (D.1), the existence of solution is ensured when F is a surjection i.e. when every point in the response space has at least a point in the input space such that $y=F(x)$. The solution is unique, when, additionally, F is an injection, i.e. when different inputs are mapped to different responses. So existence and uniqueness are fulfilled when F is a bijection so that the inverse mapping F^{-1} exists. In other words, if F is not invertible, the problem is ill posed.

Condition iii requires F^{-1} to be continuous. This is not a trivial condition, because even if F is continuous, F^{-1} may be discontinuous.

Summarising, it is possible to say that, typically, in inverse problems, F^{-1} does not exist or is not continuous.

Many IPs have unstable or non-robust solutions, meaning that small disturbances on data (due for e.g. to the noise) can produce large changes on the solutions and make the computed solution completely useless. Continuity of the solution as a function of data is a necessary but not sufficient condition for stability or robustness of the solution. The solution of a well-posed IP can be very sensitive with respect to the data (i.e. small changes on data

produce large changes on solution). In this case, the IP and its solution are termed ill-conditioned, and resemble very much an ill-posed problem.

Pseudo-solutions

In order to solve the noisy problem (D.5), it is possible to start by solving the noise-free problem (D.1). If the inverse of F exists, the solution is $x= F^{-1}(y)$. The problem is that the inclusion of the noise may imply that y does not belong to the image of F, making impossible the inversion of F on the data.

When the noise-free IP has no solution, a possibility is to reformulate the problem and look for approximate solutions. The well-known least squares (LS) method adopts the solution, which minimises the Euclidean norm of the difference between the predicted values of y and the data:

$$x_{LS} \equiv \arg min_x \|F(x) - y\|^2 \quad \text{(D.6)}$$

The function of x to minimise sometimes is called cost function:

$$J(x) \equiv \|F(x) - y\|^2 \quad \text{(D.7)}$$

The LS solutions are also called pseudo-solutions. So, the IP (D.1) is replaced by the LS problem (D.6). This change may transform an ill-posed problem into a well-posed problem, and then it is adequate to say that well-posedness has been restored by reformulating the problem.

However, LS problems are not necessarily well-posed. When the LS solution is not unique, a possibility is to select one of the solutions. This amounts to introducing prior information, for instance requiring a solution with minimum norm. Therefore, a new optimisation problem is solved, the minimisation of the norm of x subject to the fact that x is a LS solution. This minimum norm solution is also called the best approximate solution or the generalised solution.

Regularisation

Regularisation is the transformation of an ill-posed or ill-conditioned IP in a well-posed IP, by reformulating the problem. In fact, it has been seen that an IP can be turned well-posed by accepting LS solutions, which are approximate. Nevertheless, there are cases where the LS problem is still ill-posed. Then, regularisation methods may be applied to meet well-posedness.

A major issue in the solution of IP is the construction and analysis of regularisation methods and numerical schemes. Regularisation, generally, speaking, involves the injection of additional prior information about the solution and the noise.

LS solutions are obtained from the minimisation of a functional representing the Euclidean distance between data and model predictions. Regularisation introduces modifications in this minimisation.

Regularisation methods are divided in two broad families: those based on dimensionality control and those based on minimisation of a modified functional. Methods based on dimensionality control minimise the LS functional in a subspace of reduced dimension or by an iterative method (with a limited number of iterations)

The best-known method based of minimisation of a modified functional is Tikhonov regularisation, where the LS functional is modified with the addition of a "penalty term":

$$x_T \equiv \arg min_x \{\|F(x) - y\|^2 + \alpha\|x - x_0\|^2\} \quad \text{(D.8)}$$

x_0 in (D.8) is an "initial guess" of the exact solution of the IP. Solution of (D.8) is a value x close to a guess x_0 and such that F(x) is not far from y. In other words, the regularised solution is a compromise between fidelity to measured data and compatibility with prior information. The value of α (termed regularisation parameter) governs this tradeoff. There are procedures to assign "optimal" values to the regularisation parameter and obtain efficient regularisations.

The Tikhonov method can be generalised to:

$$x_T \equiv \arg min_{x \in C} \{\|F(x) - y\|^2 + \alpha\rho(x)\} \quad \text{(D.9)}$$

Where ρ is the regularisation function and C is a constraint set capturing some of the a priori information on x (i.e. non-negativity).

Inverse problems with or without uncertainty

Forward problems in modelling and simulation can be solved without taking into account uncertainties, implicitly assuming that:

- Predictive models are "perfect" versions of reality.
- Input parameters are perfectly known and have no intrinsic variability.

The resulting model outputs are point values (i.e. they are not uncertain).

Similarly, inverse problems can be solved without uncertainty, producing (as previously shown) true solutions, or pseudo-solutions, or regularised solutions, which are point estimations of the unknown input parameters. It is possible to speak of "deterministic inverse problems".

However, in order to obtain significant outcomes, realistic modelling and simulation must take uncertainties into account. In forward problems, uncertainties introduced by input parameters and by the imperfection of models must be considered and propagated to the model outputs. Similarly, in inverse problems, the different sources of uncertainty must be counted, so that the unknown inputs are estimated with uncertainty.

In modelling and simulation (M&S) inverse problems can be solved with or without uncertainty. When IPs are solved without uncertainty, point estimates of input parameters are obtained, and we can properly speak of "deterministic inverse problems". On the other hand, when the IP is aimed at the estimation with uncertainty of unknown inputs, the different sources of uncertainty must be taken into account.

An example of IP without uncertainty is model calibration, defined in this document (see the Glossary) as the process of adjusting model-selected parameters (termed calibration parameters) in order to adapt the model predictions to a set of experimental data. On the other hand, an example of IP with uncertainty is, obviously, input uncertainty quantification, defined by the Glossary as the estimation of the uncertainty associated with model inputs that does not include model calibration. It is possible, also, to perform simultaneously the calibration and the UQ of the calibration parameters.

The uncertainty introduced by model imperfection has two sources: the uncertainty of model parameters and the imperfectly known model error. Therefore, the quantification of model uncertainty is an important inverse problem, and constitutes the main objective of the present document.

The calibration parameters are a special type of model inputs (i.e. they are a subset of x in our notation). The calibration, or the uncertainty quantification, of the model represented by F, uses a database formed by pairs (x_j, y_j), $j = 1,\ldots,N$, and obtains estimates (with or without uncertainty) of the calibration parameters. These inverse methods are "parameter estimation" methods.

When there are few calibration parameters, the inverse problem is typically overestimated (more data than unknowns did), and commonly has no solution. Then, approximate solutions, obtained via classical techniques as regression and least squares, are adopted.

On the opposite side, when F represents a very complex model with many free parameters, the IP is typically underestimated and has multiple solutions. In that case, the LS solution is not unique, and additional criteria are needed to recover uniqueness (e.g. minimum norm solution).

Dynamical inverse problems. Data assimilation

A special type of inverse problem arises when the predictive model is dynamical, i.e. describing the time evolution of a system. Dynamical models commonly use the concept of *state system*. The state (a concept common in physics, thermodynamics, etc.) can be defined as a set of variables (state variables) fully describing the condition of the system as a function of time. Dynamical models describe the time evolution of the system state via a "state equation". The response is calculated as a function of the state, and the observed response may be noisy.

Data assimilation (DA) is a discipline of applied mathematics that combines mathematical dynamical models of a system with observational data. The model collects the past knowledge of the system, while the observations represent the injection of new information. DA is an inverse method for dynamical models, aiming to estimate the state of the system, initial conditions for a numerical forecast model, model parameters etc.

DA techniques use all the information available within a period, including observational data, prior information and threw dynamical model. The system state is updated by a combination of new observations and prior forecasts. Much of the work in DA is focused on how to weight adequately predictions and observations. The difficulty of DA grows with the dimensionality and the nonlinearity of the problem.

As in other IP, Bayesian methods offer a good framework for DA. The prior information is encapsulated in a prior distribution, which is updated via Bayes theorem with the observational information.

DA is very popular in the geosciences (weather forecasting, ocean forecasting, hydrology, space weather…) but finds application in many other fields (traffic control, image processing, earthquakes, chemical processes in atmosphere, planetary atmospheric circulation, robotics, bio-medicine…). Section 5.2.2.4 of Chapter 5 mentions DA methods in the nuclear safety field.

A simplified approach to data assimilation is the technique known as Kalman Filter (KF), based on the sequential state estimation in linear discrete-time dynamics. It can be proved that, with the assumption of linearity and gaussianity of prior and noise, the KF solves the Bayesian IP.

There are extensions of KF to nonlinear problems (e.g. extended KF, ensemble KF, polynomial chaos expansion KF…). Their results can be related to the Bayesian estimation.

An overlook of Inverse problems in science and engineering

Inverse problems (IP) are intimately linked to scientific inference. For this reason, IP are found in every branch of science and engineering. The catalogue of applications is so large that the present annex will just try to give a summarised overlook (Kabanikhin, 2012; Tanaka and Dulikravich, 1998, 2000; Beilina and Shestopalov, 2013; Bertero and Boccacci, 1998; Biegler et al., 2011).

It is appropriate to emphasise, as a first precaution, that every IP should be described in conjunction with the corresponding forward or direct problem (DP).

Many classifications of IPs can be done, according to different criteria.

A first classification should distinguish between mathematical IP and practical and computational IP.

The first includes generic mathematical techniques, e.g. inverse scattering problems, inverse boundary value problems, inverse initial value problems, inverse eigenvalue problems, mathematical tomography. Theory of IP has impact in many areas of mathematics: algebra, calculus, geometry, differential equations, mathematical physics, functional analysis, computational mathematics, etc.

In (Kabanikhin, 2012) it is stated that, in mathematical physics, a DP usually is aimed to find the field or process (electromagnetic, acoustic, heat transfer…) at any point of a given spatial domain at any time (if the field is not stationary). Thus, the formulation of a DP includes the equations describing the process (typically, integral-differential equations), the definition of the domain, initial and boundary conditions, source terms, etc. Using all this information, the solution of the DP produces the value of the involved magnitudes in given spatial and time points. For these DPs, several types of IP can be derived:

- Retrospective IP, to determine initial conditions.
- Inverse boundary value problem, to determine boundary conditions.
- Extension problem, when the solution of the DP only is known on a part of the boundary of a domain, and the solution must be extended to the interior of the domain.

- Inverse source problem: determine the source term of the equation.
- Coefficient problem: reconstruct the coefficient of the main equation.

In the corresponding inverse problems, the value of the magnitudes is known at some spatial sets and in a certain time set.

The unknowns (i.e. the solutions of the IP) are a part of the aforementioned elements. If they are initial conditions, we have a retrospective IP; if they are boundary conditions, we have an inverse boundary value problem. If the unknown is the source term, we have an inverse source problem. The IP is called a coefficient problem if it is aimed at reconstructing coefficients in the main equations. And so on.

Taking into account the variety of physical phenomena modeled in science and engineering (S&E), the different types of known responses and of unknown parameters, the immense variety of possible inverse problems becomes clear.

According to the branch of S&E involved, there are IPs in:

- Physics (quantum mechanics, optics, acoustics, heat transfer, electrodynamics, spectroscopy, astronomy…).
- Earth and space sciences, including Geophysics (seismic exploration; electrical, magnetic and gravimetric prospection, magnetotelluric sounding…).
- Medicine (X-ray and NMR tomography, ultrasound testing…).
- Ecology (air and water quality control, space monitoring…).
- Aerospace industry.
- Nuclear industry.
- Civil engineering.
- Economics (econometrics, financial mathematics…).
- Structural chemistry.

A special case is statistics. Mathematical statistics may be regarded as an inverse problem with respect to probability theory.

Another possible classification is based on the type of techniques that can be applied to different types of fields:

- Image processing and reconstruction, specially applied in medicine, but also in industrial nondestructive evaluation and process monitoring, astronomy, meteorology. There are many techniques: computerised tomography, positron emission tomography, ultrasound imaging, electrical impedance tomography, optical imaging, diffraction tomography, radar imaging, etc.
- Biomedical engineering.
- Instrumentation.
- Signal processing.
- Heat exchange.
- Mechanics.

Some special types of IPs can be singled out. For instance, many forward problems can be modelled as the **convolution** of two functions.

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(s)g(t-s)ds \quad \text{(D.10)}$$

The inverse problem of recovering one of the functions from data of the convolution function is called *deconvolution problem*. Deconvolution techniques are applied in many scientific and engineering fields. They are widely used in signal processing, considering signals as functions of time, and in image processing and optics. A well-known application is the *deblurring* of images. Under some assumptions, the blurred image is the convolution of the "true" image and a point spread function. Other fields of application include weather forecasting, economics, seismology, optics and radio astronomy. Many measurement devices produce an output signal that is the convolution of the input signal and a transfer function. In the IP, the input signal is reconstructed from the known output.

Fourier transform (FT) is a very important technique in the solution of deconvolution problems. FT of the convolution of two functions is the product of the individual FT.

D.2 Description of CASUALIDAD method

The predictive modelling methodology constitutes a third kind of approach to perform uncertainty analysis, different from propagation of input uncertainties or from propagation of code output accuracies already discussed in previous sections.

The method is based upon powerful mathematical tools to perform sensitivity analysis, and upon the *Data Adjustment/Assimilation* (DAA) methodology by which experimental observations are combined with code predictions and their respective errors to provide an improved estimate of the system state and of the associated uncertainty, considering all input parameters $\boldsymbol{\alpha}$ that affect any prediction, being part either of the code models or of the input deck.

Since in most problems of practical interest, the number of input parameters $\boldsymbol{\alpha}$ exceeds the number of responses **R**, the method requires the most efficient deterministic method for computing local sensitivities **S** for large-scale systems, which is the ASAP. However, as large-scale systems are frequently characterised by critical points like bifurcations, turning points, saddle points, response extrema or cliff-edge effects, local sensitivity tools cannot be properly used, and a deterministic global sensitivity method has to be applied. In this context the global adjoint sensitivity analysis procedure (GASAP) is one of the most advanced tools to deterministically compute the system's critical points **y** in the combined phase-space formed by the parameters, forward state variables, and adjoint variables. Subsequently the local sensitivities of the responses **R** located at critical points **y** are analysed by the ASAP. Once the sensitivity matrix **S** of the responses **R** with respect to the parameters $\boldsymbol{\alpha}$ is available, the moment propagation equation is adopted to obtain the computed covariance matrix $\mathbf{C_R}$ of the responses starting from the covariance matrix $\mathbf{C_\alpha}$ of the system parameters. The technique, by which experimental observations are combined with code predictions and their respective errors, to provide an improved estimate of the system state is known as DAA and it is based on a Bayesian inference process. The idea at the basis of DAA can be made more specific, as follows: the computed results **R,** and the

respective statistical errors $\mathbf{C_R}$ predicted by mathematical models, and based on a 'prior' or "first" guess PDF for the input parameters (i.e. $\mathbf{C_\alpha}$) are combined with proper experimental observations **M** of the states of a system, and associated uncertainty $\mathbf{C_M}$ to generate "adjusted" values for the system parameters ($\boldsymbol{\alpha}^{IE}$, where the suffix IE stays for Improved Estimate values), and the respective input covariance matrix ($\mathbf{C_\alpha}^{IE}$, or "posterior" PDF). From this process, which can be considered as improved estimate analysis of the system's states, the responses $\mathbf{R}^{IE}$ and the respective covariance matrix ($\mathbf{C_R}^{IE}$) are finally derived. A short description of a method based on predictive modelling methodology is given hereafter.

The CASUALIDAD method (Code with the capability of adjoint sensitivity and uncertainty analysis by internal data adjustment and assimilation) has been developed as a fully deterministic method based on advanced mathematical tools for performing internally to the thermal-hydraulic system code the sensitivity and the uncertainty analysis. The availability of a suitable database of experiments (SETF and/or ITF) and related qualified code calculations constitutes a pre-requisite for the development and the following application of the methodology.

The bases of the CASUALIDAD method

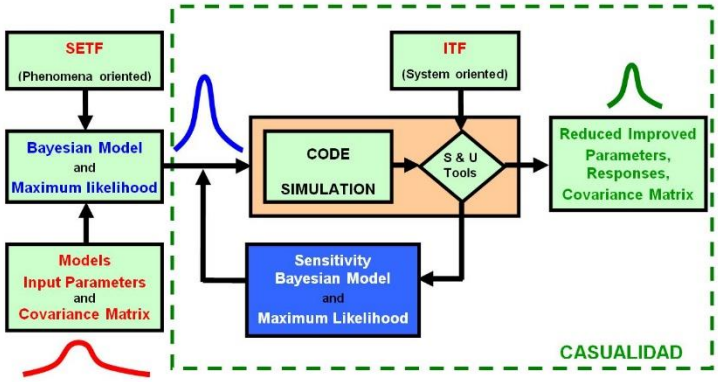
Establishing the range and probability distribution function of parameters is fairly easy for parameters which describe the condition of the plant (such as initial power or geometrical data) or for parameters describing physical data (such as thermal conductivity of UO_2). It is more difficult for parameters relative to the constitutive relationships (such as interfacial friction) because they cannot be directly measured in facilities. To this end, the results from SETF together with their experimental uncertainties are used for establishing or assessing these constitutive relationships. However, the models developed from SETF are applicable to a certain range of parameters, when often the correlations are extrapolated to more extreme conditions in correspondence of which the data is lacking. In order to justify these extrapolations, the ITF data, whose ranges of parameters are much closer to typical conditions, are used. However, the derivation of the PDFs of the input uncertainty parameters is never an easy and objective task and a common or agreed methodology is still not available to the nuclear community.

The fundamental principle of the methodology here proposed can be derived from Figure D.1. The initial uncertainty distribution obtained from SETFs is shown coloured in red and represents the uncertainty in each of the model parameters (multiple model parameter distributions). After the application of the Bayesian Theorem (BT) and of the Principle of the Maximum Likelihood (PML), the uncertainty distributions (in blue), which are prior estimations for the code-simulation, are obtained. The code simulations are further constrained (using BT & PML) against the ITF for considering both the system behaviour and the use of the correlations (developed starting from SETFs) outside their range of validity. At the end of this step, the key parameters, their contributions to the uncertainty of the system responses and the posterior parameter distributions (green curve in Figure D.1) are obtained. It shall be noted that if the extrapolation of the SETF conditions to the ITF conditions were not required and the various physics models employed in the ITF were truly uncoupled, then it is expected that the prior and posterior distributions resulting from the application of the BT & PML are very similar.

The framework for which the CASUALIDAD method has been developed deals with the dashed area in Figure D.1, or, in other terms, the methodology is system oriented, having as its main objective the improvements of the estimations of the system output responses

and related covariance matrix (i.e. uncertainty) through the reduction and improvement of the input parameter values and covariance. Key elements for the methodology are the following:

Figure D.1. The basic idea and framework of the CASUALIDAD method



1. Availability of a frozen qualified and internationally recognised thermal-hydraulic code.
2. Availability of a suitable large database of ITF containing different transient scenarios and different scales of facility (e.g. the set of tests belonging to).
3. Use of a robust and rigorous procedure (i.e. set of acceptability criteria to be satisfied) for the qualification and acceptance of code calculations (Petruzzi and D'Auria, 2005).
4. Availability of adequate (i.e. exact and efficient from CPU time point of view) methods for performing the local and global sensitivity analysis. The powerful tool based on ASAP is implemented for the derivation of the sensitivity matrix **S** containing the local derivatives of any response R_n (obtained by the code) with respect to any parameter α_i ($\partial R_n / \partial \alpha_i$). For large-scale systems, in which the number of system parameters and/or parameter variations to be considered exceeds the number of responses of interest, the ASAP is, by far, the most advantageous method to employ. The global sensitivity analysis is performed through the implementation of the GASAP method aimed at determining all of the system's critical points and subsequently analyse them locally by ASAP. Thus, the strong non-linearities characterising the nuclear power plant system can be efficiently considered during the sensitivity analysis and taken into account for the following step dealing with the uncertainty evaluation.
5. Implementation of the DAA method based on the Bayesian theorem and the principle of the maximum likelihood for updating the "a priori" PDF of the input parameters **α** and responses **R** with the available experiments **M** ("likelihood observations", see also Element 2 above) for getting the "posterior" improved estimation (IE) of the input parameters, responses and related covariance matrixes.
6. Use of the concept of status approach for grouping together the "posterior" improved estimations of the input parameters, responses and related covariance

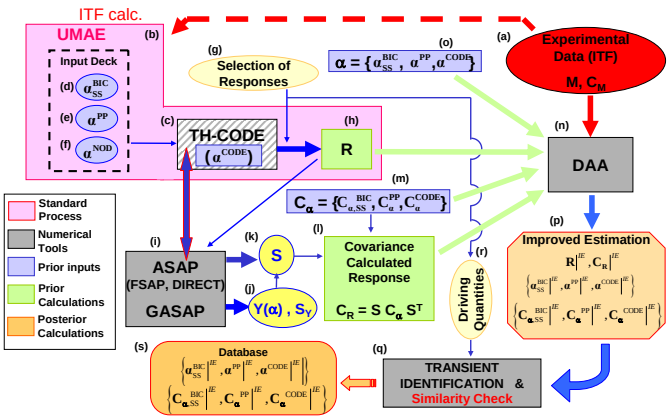
matrices derived from similar transients, i.e. transients selecting the same path in the phase space of the selected driving quantities. A database of improved estimations is then generated, and the appropriate information stored inside (i.e. the one in the phase space selected by the nuclear power plant calculation) can be used during the application process of the methodology.

In addition to the above, the following assumptions are used in the methodology, including criteria and steps embedded in the methodology which itself makes it possible to detect when the assumptions are violated (and thus when the methodology cannot be applied) or demonstrate their well-founded basis:

1. Phenomena and transient scenarios in larger scale facilities shall be close enough to plant conditions (see also similar discussion in (Petruzzi and D'Auria, 2005) in Section D.1 above). This assumption also supports Element 6 in the list above.
2. Given that the discrepancy between measured and calculated responses (and not the absolute values) is used for evaluating the uncertainty, this difference is randomly dispersed around the zero-value independent of the volume scaling factors of the facilities. Moreover, the methodology provides a consistent indicator χ^2 that quantifies the degree of consistency between calculation and experiment and allows neglecting the contributions of the related discrepancies to the uncertainty estimation when the consistency is poor (i.e. χ^2 far from unity).
3. The influence of user and nodalisation upon the uncertainty of the response is minimised in the methodology by the use of robust qualification procedures (see Element 3 of the previous list). The consistency indicator χ^2 contributes to minimise this influence discharging the cases where the consistency between calculation and experiment is poor.

The flow chart of the CASUALIDAD method is depicted in Figure D.2 below.

Figure D.2. Flow-chart of CASUALIDAD method - development process

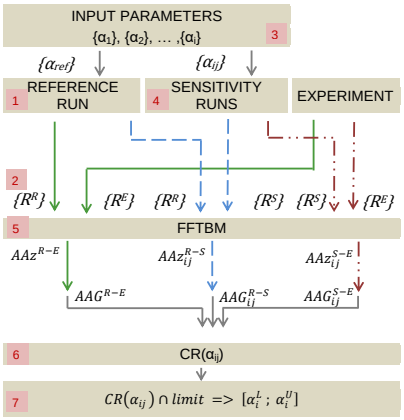


D.3 Description of IPREM method

The feature of FFTBM (Ambrosini et al, 1990) is to provide a quantitative evaluation of the accuracy of a time-dependent code output parameter with respect to experimental. The methodology has been defined as input parameter range evaluation methodology (IPREM)

(Kovtonyuk et al., 2015) and it is characterised by the following steps outlined on the flowchart of Figure D.3.

Figure D.3. Flowchart of IPREM.



The quantification of the variation ranges of the input uncertain parameters, for the selected physical models, is achieved through running the calculations of a "reference case" of a physical model and "sensitivity" cases, constituted by a single-parameter variation, application of the FFTBM for quantification of the accuracy of calculated responses respect to experimental data and further comparison of differences between average amplitude (AA) values obtained from sensitivity cases and an AA of the reference case.

Once the model has been set up, the analyst performs the reference calculation (Step 1 on Figure D.3) of experimental test and performs the qualitative check of obtained results to ensure that relevant thermal-hydraulic phenomena observed in the experiment are predicted by the code.

The relevant thermal-hydraulic parameters that describe the phenomena of interest should be selected as responses {RR} (Step 2 on Figure D.3), e.g. cladding temperature and quench front propagation are the representative code output parameters for the reflood phenomenon. Reliable experimental measurements {RE} must be available for the responses of interest.

The list of studied input parameters $\{\alpha_i\}$ is established at Step 3. A preliminary sensitivity analysis may be applied to identify those parameters that are influential to the selected responses. For each i-th input parameter of interest, a number of j calculations must be performed (Step 4 on Figure D.3), by varying only the i-th parameter. As a result, the set of calculated selected responses {RijS} is available for each j-th run. Two applications of FFTBM analyses must be performed for each j-th sensitivity run of each i-th input parameter:

To perform FFTBM for the "Sensitivity calculation – Experiment data" pair ($\{R_{ij}^S\}$ vs $\{R^E\}$);

To perform FFTBM for the "Sensitivity calculation – Reference calculation" pair ($\{R_{ij}^S\}$ vs $\{R^R\}$).

As a result, the analyst obtains two sets of Average Amplitudes for each j-th sensitivity run of each i-th input parameter:

AAz_{ij}^{S-R} That quantifies the "deviation" of the sensitivity run from reference case;

AAz_{ij}^{S-E} That quantifies the "accuracy" of the sensitivity run with respect to experimental data;

Where z is the consecutive number of a response.

At this point, a AAz_{ij} value is available for each selected response for each j-th sensitivity run of each i-th input parameter. The following step is performed in order to produce a single figure of merit that makes it possible to:

- quantify the sensitivity of entire simulation model to the input parameter variation;
- quantify the accuracy of entire simulation model performance in each sensitivity run with respect to experimental data.

This is achieved by calculating the Global AA (AAG) for each of two sets of AAzij derived from j-th sensitivity run of i-th input parameter (equation (D.11)):

$$\mathrm{AAG} = \sum_{z} \mathrm{w_z AA_z} \qquad \text{(D.11)}$$

Where w_z is the weighting factor assigned for each type of response. The weighting factors are determined by the type of thermal-hydraulic parameter selected as a response. A proposed set of weighting factors has been developed for "intermediate" experimental tests based on weighting factors used in original FFTBM procedure for evaluation of code calculation accuracy. After the weights are selected, and assigned to the AAz_{ij} of corresponding responses, they are normalised by Equation (D.12):

$$w_z = \frac{W_z}{\sum_{i=1}^{Nz} W_i} \qquad \text{(D.12)}$$

Once the AAG values are calculated for each j-th value of input parameter αi, the criterion quantity $CR(\alpha_{ij})$ is defined by equation (D.13):

$$\mathrm{CR}(\alpha_{ij}) = \frac{\mathrm{AAG^{S-E}}(\alpha_{ij}) + \mathrm{AAG^{S-R}}(\alpha_{ij}) - \mathrm{AAG^{R-E}}}{\left(1 - \mathrm{AAG^{S-E}}(\alpha_{ij})\right)} \qquad \text{(D.13)}$$

In equation (D.13) the following constituents are included:

AAG^{S-E}+AAG^{S-R} is a measure of the total "deviation" of a sensitivity calculation from both the reference calculation and the experimental data;

AAG^{R-E} is calculated from the comparison of reference calculation and experimental data;

(1-AAG^{S-E}) "slows" the increase of CR if the change of an input parameter leads to improvement of results with respect to experiment.

As a last step, the variation ranges of each parameter α are quantified by applying the limiting value (threshold) to CR(α). The lower and upper bounds of α are defined by equation (D.14):

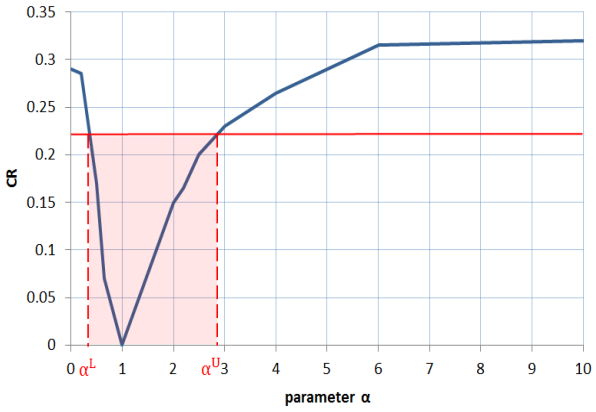
$$[\alpha^{L};\ \alpha^{U}] = \mathrm{CR}(\alpha) < \mathrm{limit} \qquad \text{(D.14)}$$

The limit value must be set once and consistently applied for all the analyses performed. In IPREM a value of limit=0.22 has been adopted. The value of 0.22 is based on the consideration of the maximum allowed "deviation" of responses (at extremes of the range of input parameters) of 10% (in the IPREM metrics) in the hypothetical case when reference calculation exactly matches the experimental data ($AAG^{S-E} = AAG^{S-R}$), see Equation (D.15):

$$\mathrm{CR} = \frac{2 \cdot \mathrm{AAG}}{(1 - \mathrm{AAG})} \leq 0.22 \quad \Leftrightarrow \quad \mathrm{AAG} \leq 0.1 \qquad \text{(D.15)}$$

A typical trend of CR(α) obtained for one input parameter is showed on the Figure D.4.

Figure D.4. Sample trend of CR quantity



The IPREM methodology proved to allow evaluation of input parameter uncertainty from a single "intermediate" experimental test. However, the use of a number of other experimental tests is required to validate the obtained ranges [α_L ; α_U]. This methodology does not depend on the applied thermal-hydraulic system code, as well as on the type of investigated input parameter and analysed responses, since the procedure involves only post-processing of calculation results. In principle, it also does not require code modification for assessment of an input parameter uncertainty given access to a parameter of interest (e.g. CATHARE code and other system thermal-hydraulic and severe accident codes like MELCOR). However, the proposed methodology is rather based on engineering considerations and previous experience from the application of FFTBM than on statistical methods. It does not take into account or provide as a result the Probability Density Function for each input parameter (adoption of uniform or likewise distributions is therefore suggested). The IPREM reduces the use of engineering judgment in terms that the proper procedure, mathematical apparatus and corresponding criteria are clearly defined. However,

the choice of analysed code responses may affect the resulting ranges of input parameters, and in some particular cases the CR(α) may not reach the limit value.

- **References**

Alquier, P. et al. (Eds) (2009), *Inverse Problems and High-Dimensional Estimation*, Stats in the Château Summer School, Springer.

Ambrosini, W. et al. (1990), "Evaluation of Accuracy of Thermal-hydraulic Codes Calculations". In *Energia Nucleare*, 7.

Aster, C. et al. (2005), *Parameter Estimation and Inverse Problems*, Elsevier Academic Press.

Bal, G. (2012), *Introduction to Inverse Problems*, Columbia University, New York.

Banks, H.T. et al. (2014), *Modeling and Inverse Problems in the Presence of Uncertainty*, monographs and research notes in mathematics, CRC Press.

Beilina, L. and Y.V. Shestopalov (Eds) (2013), *Inverse Problems and Large-Scale Computations*, Springer.

Bertero, M. and P. Boccacci (1998), *Introduction to Inverse Problems in Imaging*, IOP Publishing Ltd.

Biegler, L. et al. (Eds.) (2011), *Large-Scale Inverse Problems and Quantification of Uncertainty*, J. Wiley & Sons.

Bovalini, R. and F. D'Auria (1993), "Scaling of the accuracy of Relap5/moD-2 Code", in *J. of Nuclear Engineering and Design*, 139, pp. 187-204.

Bryan, K. (2011), *An Introduction to Inverse Problems*, www.rose-hulman.edu/~bryan/invprobs/inversefin1.pdf

Brynjarsdóttir, J. and A. O'Hagan (2014), "Learning about physical parameters: the importance of model discrepancy", in *Inverse Problems*, 30(11), pp. 114007.

Burger, M. (2008), *Inverse Problems. Lecture Notes*, University of Münster.

Cacuci, D.G. (2006), *On the Use of Adjoint Operators for Global Optimization and Data Assimilation*, Lectures Notes, Workshop on Predictability, Observations, and Uncertainties in Geosciences, Tallahassee, United States.

Cacuci, D.G. (2003), *Sensitivity and Uncertainty Analysis*, Theory, Vol. I, Chapman & Hall/CRC, Boca Raton, Florida, United States, ISBN 1-58488-115-1.

Cacuci, D.G. et al. (2014), *Computational Methods for Data Evaluation and Assimilation*, CRC Press.

Cacuci, D.G. et al. (2005), *Sensitivity and Uncertainty Analysis, Application to Large Scale Systems*, Vol. II, Chapman & Hall/CRC, Boca Raton (Fl, US), ISBN-10 1-58488-116-X.

Cacuci, D.G. and M. Ionescu-Bujor (2010), *Sensitivity and Uncertainty Analysis, Data Assimilation and Predictive Best-Estimate Model Calibration*, Chapter 17 in Vol.3, pp 1913 – 2051, *Handbook of Nuclear Engineering*, D. G. Cacuci, Editor, ISBN: 978-0-387-98150-5, Springer New York / Berlin.

Cacuci, D.G. and M. Ionescu-Bujor (2010), "Model Calibration and Best-Estimate Prediction through Experimental Data Assimilation: I. Mathematical Framework", in *Nuclear Science and Engineering*, 165, pp. 18-44.

Evensen, G. (2009), *Data Assimilation. The Ensemble Kalman Filter*, second edition, Springer.

Fox, C. et al. (2010), *An Introduction To Inverse Problems*, Course notes for ELEC 404.

Grewal, M.S. and A.P. Andrews (2001), *Kalman Filtering: Theory and Practice Using MATLAB*, Second Edition, J. Wiley & Sons.

Groetsch, C.W. (1999), *Inverse Problems: Activities for Undergraduates*, The Mathematical Association of America.

Groetsch, C.W. (1993), *Inverse problems in the Mathematical Sciences*, Vieweg.

Guzzi, R. (2016), *Data Assimilation: Mathematical Concepts and Instructive Examples*, Springer.

Hohage, T. (2002), *Lecture notes on Inverse Problems*, University of Göttingen.

Idier, J. (Ed) (2008), *Bayesian Approach to Inverse Problems*, ISTE – Wiley.

Ito, K. and B. Jin (2015), *Inverse problems: Tikhonov Theory and Algorithms*, World Scientific Publishing Co. Pte. Ltd.

Kabanikhin, S.I. (2012), *Inverse and Ill-Posed Problems. Theory and Applications*, De Gruyter.

Kennedy, M.C. and A. O'Hagan (2001), "Bayesian calibration of computer models", in *J.R. Statist. Soc. B*, 63(3), pp. 425-64.

Kirsch, A. (2011), *An Introduction to the Mathematical Theory of Inverse Problems*, 2nd edition, Springer.

Kovtonyuk, A. et al. (2015), "Development and Assessment of a Method for Evaluating Uncertainty of Input Parameters", in *Proc. of NURETH16*, Chicago, USA.

Lahoz, W. et al. (Eds) (2010), *Data Assimilation. Making Sense of Observations*, Springer.

Lasanen, Y. (2014), *Introduction to inverse problems (Spring 2014)*, 802360A.

Law, K. et al. (2015), *Data Assimilation. A Mathematical Introduction*, Springer

Maréchal, P. (2016), *Introduction to Inverse Problems of Fourier Synthesis, Lecture notes*, Université Paul Sabatier. https://perso.math.univ-toulouse.fr/marechal/files/2016/06/LN-FourierSynthesis.pdf.

Petruzzi, A. (2008), *Development and Application of Methodologies for Sensitivity Analysis and Uncertainty Evaluation of the Results of the Best Estimate System Codes Applied in Nuclear Technology*, PhD Thesis, University of Pisa.

Petruzzi, A. and F. D'Auria (2014), "Uncertainties in Predictions by System Thermal-Hydraulic Codes: the CASUALIDAD Method", *Proc. of 10th International Topical meeting on Nuclear Thermal-Hydraulics, Operation and Safety (NUTHOS-10)*, Okinawa, Japan.

Petruzzi, A. and F. D'Auria (2005), "Development, Qualification and Use of the CIAU Method for Internal Assessment of Uncertainty in System Codes", in *Proc of 11th*

International Topical Meeting on Nuclear Reactor Thermal-Hydraulics (NURETH-11), Avignon, France.

Petruzzi, A. et al. (2010), "Model Calibration and Best-Estimate Prediction through Experimental Data Assimilation: II. Application to a Blowdown Benchmark Experiment", in *Nuclear Science and Engineering*, 165, pp. 45-100.

Ramm, A.G. (2005), *Inverse Problems. Mathematical and analytical techniques with applications to Engineering*, Springer.

Sun, N. and A. Sun (2015), *Model Calibration and Parameter Estimation For Environmental and Water Resource*, Springer.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.

Tikhonov, A.N. and V.Y. Arsenin (1977), *Solutions of ill-posed problems*, V. H. Winston & Sons.

Tanaka, M. and G.S. Dulikravich (Eds) (2000), *Inverse Problems in Engineering Mechanics II*, Elsevier.

Tanaka, M. and G.S. Dulikravich (Eds) (1998), *Inverse Problems in Engineering Mechanics*, Elsevier.

Van Leeuwen, P.J. et al. (2015), *Nonlinear Data Assimilation*, Springer.

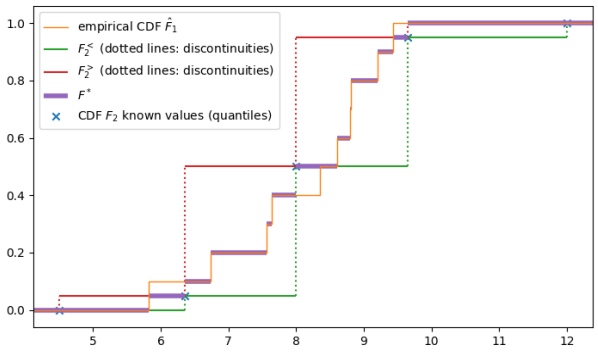
Vogel, C.R. (2002), *Computational Methods for Inverse Problems*, SIAM.

ANNEX E - Model input uncertainty validation (Chapter 6)

E.1 Minimisation and maximisation of the area metric A($\hat{F}_1$; F)

It can be shown that the minimum of A($\hat{F}_1$; F) is reached for F = F* where F* is defined as follows. Let $x_0 < \ldots < x_N$ be the ranked values of q_0, of q_1, of the known quantiles, and of the realisations of F_1 together, hence $\hat{F}_1$, $F_2^<$ and $F_2^>$ take some constant values, respectively $\hat{F}_1(i)$, $F_2^<$ (i) and $F_2^>$ (i), over every $[x_i; x_{i+1}[$, then $F^*(x) = \arg\min |F^i - \hat{F}_1(i)|$ for $F_2^<$ (i) $\leq F^i \leq F_2^>$ (i) (that is $F^*(x) = F_2^<$ (i) if $\hat{F}_1(i) < F_2^<$ (i), $F_2^>$ (i) if $\hat{F}_1(i) > F_2^>$ (i), $\hat{F}_1(i)$ otherwise) for all x in $[x_i; x_{i+1}[$ (F^* is indeed a non-decreasing right-continuous function). Figure E.1 illustrates this. Besides, A($\hat{F}_1$; F) can be upper-bounded with respect to F by A($\hat{F}_1$; G) with $G(x) = \arg\max |F^i - \hat{F}_1(i)|$ for $F_2^<$ (i) $\leq F^i \leq F_2^>$ (i) over $[x_i; x_{i+1}[$, however G is generally not non-decreasing and this bound may be over-pessimistic.

Figure E.1. CDF F* minimising A($\hat{F}_1$; F) with respect to F in the P-box [$F_2^<$, $F_2^>$]



E.2 Alternative approach to depth evaluation for the validation of time-trend results

This approach consists in the following steps:

(S1) Finding a low-dimensional representation of the time-trend SRQ on the basis of the data by means of a Principal Component Analysis or of another technique of dimension reduction; any time trend y^k is then represented by a vector $\alpha^k = (\alpha_1^k, \ldots, \alpha_D^k)$ of low dimension D: there are some functions Φ and Ψ such that $\alpha^k = \Phi(y^k)$ and $y^k \approx \Psi(\alpha^k)$.

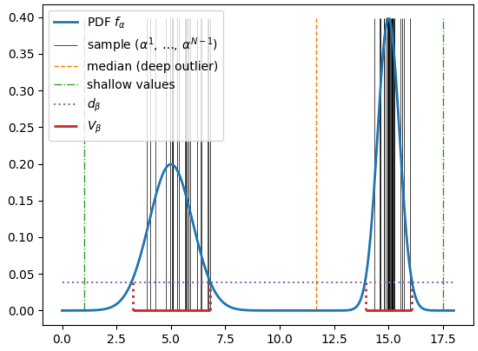
(S2) Estimating the distribution from which $(\alpha^k, \ldots, \alpha^N)$ is supposed to be sampled, e.g. by a multivariate Gaussian mixture model (see (Nanty et al., 2017)) or by another technique.

(S3) Let α be a random vector ruled by this distribution, given a level of confidence β (for instance 95 %), compute the set $V_\beta \subset \mathbb{R}^D$ of minimum volume such that $P(\alpha \in V_\beta) = \beta$. This set is delimited by a level set of the PDF f_α of α: there is $d_\beta > 0$ such that $V_\beta = \{x \in \mathbb{R}^D : f_\alpha(x) \geq d_\beta\}$ (http://openturns.github.io).

The experimental SRQ y_N would then appear unlike if $\alpha^N = \Phi(y^N) \notin V_\beta$, that is if $f_\alpha(x) < d_\beta$ (presumption of invalidity). Despite step 1, step 2 may still suffer from the curse of dimensionality (which comes to a lack of data).

Figure E.2 illustrates steps 2 and 3, and the benefits that could be expected by adopting this approach rather than the one based on the notion of depth.

Figure E.2. Example of deep outlier (median)



Let us assume that the data are not some functions of time, but some scalars denoted by α^k instead of y_k, or that the dimension reduction (Step 1) is so efficient that D = 1. The sample (α^1, …, α^N) of N scalars is displayed together with its underlying PDF f_α. Because of the bimodality of f_α, the depth cannot be used to detect all the outliers: if the values in green (shallow values) have a low depth, any value around the median should be considered as an outlier despite its high depth. On the contrary, the estimation of V_β (that is d_β) enables the identification of such a value as unlike, thus as an outlier.

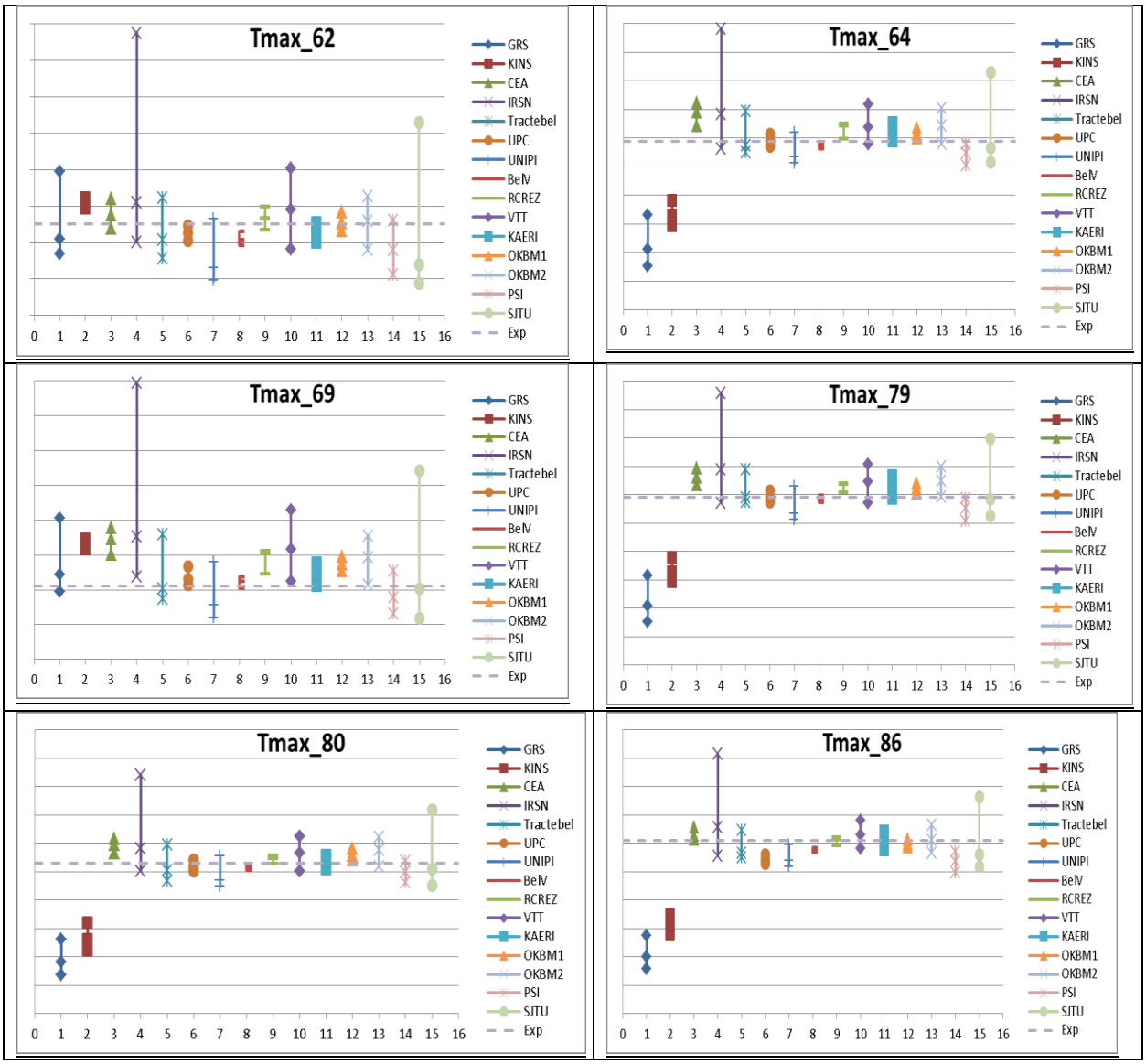
Nevertheless, the accurate estimation of the set V_β is generally difficult and requires very much data.

E.3 Illustration of some validation indicators

This part is devoted to an illustration of some validation indicators described in Chapter 6. More precisely, Q_0, Q_1, Q_4 and Q_5 are applied to analyse the uncertainty results associated to the PERICLES data and performed in the PREMIUM benchmark. All details concerning this experiment can be found in the final PREMIUM report. By construction, each validation indicator is focused on different important characteristics of the SRQ uncertainty to capture and the objective is here to clearly show the impact on validation results.

The PERICLES SRQs considered in this study are reduced to maximum clad temperatures for a given assemble (B) in the middle part of the bundle (1825 mm). Each PREMIUM participant was asked to perform a probabilistic uncertainty analysis (as described in Section 6.3.1.1) for each PERICLES tests with the input model uncertainties quantified in a previous phase. Uncertainty results after propagation are summarised by an interval [LUB,UUB] where LUB and UUB are respectively associated to an estimation by order statistics of the 2.5%-percentile and of the 97.5%-percentile. Figures E.3 displays the participants' results as well as the experimental value for the six considered PERICLES tests.

Figure E.3. Uncertainty results provided by each participant for the 6 PERICLES tests. Three types of information are given for each contribution: LUB, Reference calculation and UUB.



As can be seen in Figure E.3, there is a large dispersion between participants' results. In the sequel it is investigated how Q_0, Q_1, Q_4 and Q_5 can capture such dispersion. According to the uncertainty analysis specification, the following choices are adopted:

Q₀: for each participant, it is computed as: $Q_0 = \frac{1}{6}\sum_{i=1}^{6} 1_{\{v_i \in [LUB_i, UUB_i]\}}$

where v_i, LUB_i and UUB_i correspond respectively to the experimental value, the lower and the upper bounds for each SRQ. Since *LUB* and *UUB* are associated to estimations of the 2.5% and 97.5%-percentiles, we also introduce:

$$\widetilde{Q_0} = |0.95 - Q_0|$$

Q_1: as mentioned in Chapter 6, this indicator can take into account a larger interval [LUB_{min},UUB_{max}] coming from the propagation of initial input uncertainties performed in Element 3. Since this information was not available in PREMIUM, these two bounds are computed as follows: for each SRQ, if q_{min} and q_{max} are the minimal and maximal values of the lower and upper bounds given by all participants

$$LUB_{min} = 0.98 * q_{min}$$

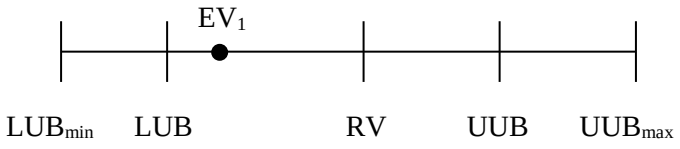
$$UUB_{max} = 1.02 * q_{max}$$

It then requires defining the interquartile associated to the 4 disjoined sub-intervals [LUB_{min},LUB], [LUB, RV], [RV, UUB], [UUB,UUB_{max}]. Since reference values are often associated to the median of the distribution, we take: (p_1,p_2,p_3, p_4)=(0.025,0.475,0.475,0.025). Therefore, the null hypothesis to test reads for this particular problem "(0.025,0.475,0.475,0.025) defines the true vector of interquantiles". The validation indicator is:

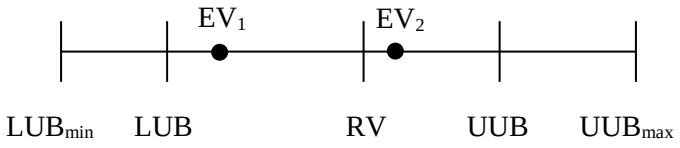
$$Q_1 = \sum_{i=1}^{4} 6\frac{(r_i - p_i)^2}{p_i}$$

Where (r_1,r_2,r_3, r_4) is the vector of observed interquartiles deduced from each participant's contribution. Figure E.4 provides an illustration of its computation.

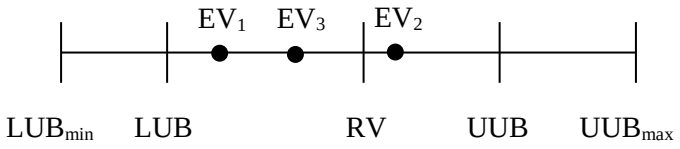
Figure E.4. Example of evaluation of (r_1,r_2,r_3, r_4) from the number of experimental values (EV_i) falling in each sub-intervals



(r_1,r_2,r_3, r_4)=(0,1,0,0)



(r_1,r_2,r_3, r_4)=(0,0.5,0.5,0)



(r_1,r_2,r_3, r_4)=(0,0.667,0.333,0)

Q_4, Q_5 : their expression is given in Section 6.4.3. LUB_{min} and UUB_{max} are the same as before. These types of indicator are computed for each participant and each SRQ. Two global indicators by participant are obtained by averaging over the six SRQs.

Figures E.5 and E.6 provide the comparison between the four previous indicators.

Figure E.5. Comparison between calibration indicators

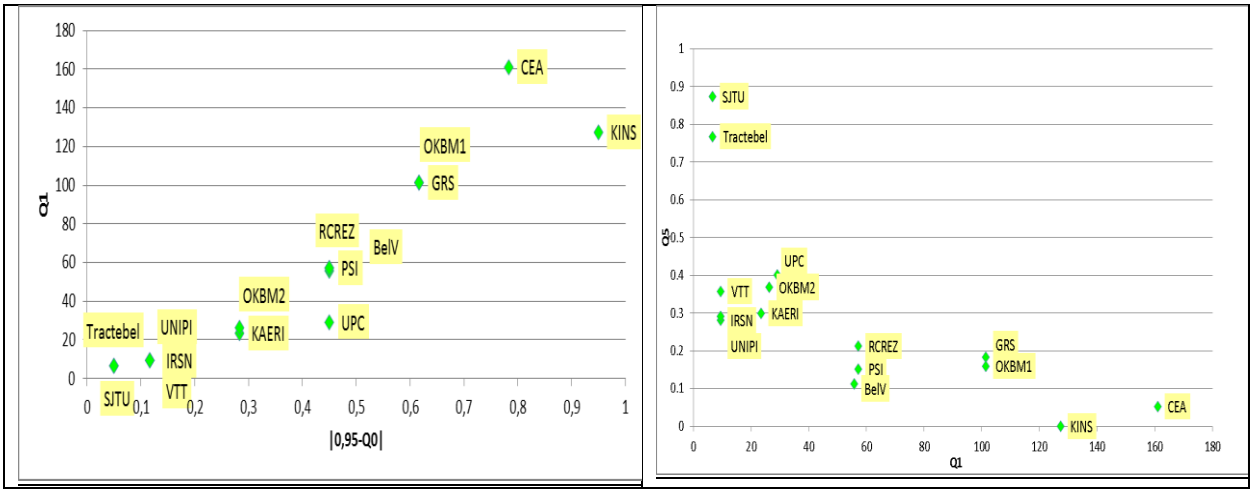
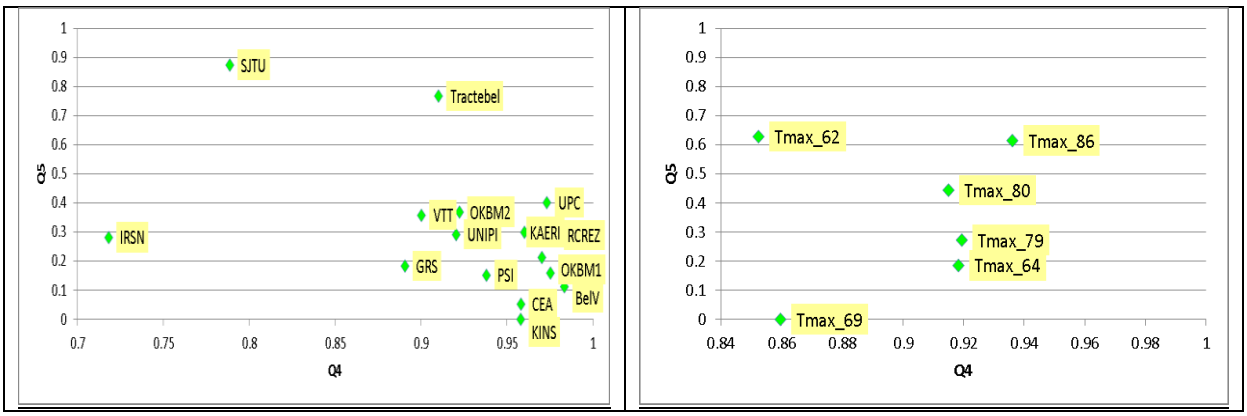


Figure E.6. Informativeness (Q_4) and calibration (Q_5)indicators. Left, for each participant, right, for each SRQ associated to a given participant (VTT).



From Figure E.5, left, it appears that $\widetilde{Q_0}$ and Q_1 lead to the same ranking of the participants' results except for CEA and KINS. This is because, contrarily to Q_0, Q_1 takes into account the number of experimental values falling in each sub-interval associated to the vector of interquartiles. Therefore, since KINS results are better balanced around the reference calculation (1/3 of the experimental values are smaller than the reference values and 2/3 are larger) than CEA ones (all experimental values are smaller than the reference values), the corresponding Q_1 is smaller. It is also important to notice that some participants such as Tractebel and SJTU have the same value for both $\widetilde{Q_0}$ and Q_1. However, according to Figure E.3, the uncertainty intervals provided by SJTU are larger than the Tractebel ones. In this case, the uncertainty interval width has therefore no impact on the validation.

Figure E.5, right, exhibits the same trend[20] in the validation results provided by Q_1 and Q_5. However, Q_5 allows better separating the participants' contributions. The two indicators capture different properties of the uncertainty results. Q_5 is focused on the position of the experimental value within the uncertainty interval and on the discrepancy between experimental and reference values whereas Q_1 evaluates how balanced experimental values are around reference calculations. As a result, Q_5 allows separating two participants even if the experimental values fall inside the same sub-interval, see for example Tractebel and SJTU. For these two participants, it is also interesting to notice that according to this indicator, SJTU is better calibrated than Tractebel even if the discrepancy between experimental and reference values is smaller for the latter for a large majority of SRQs. This is due to the construction of Q_5 that focusses on the discrepancy but relatively to the uncertainty interval width. Therefore, for the same discrepancy, a wide interval might lead to a higher calibration.

It appears that uncertainty interval width is an important property to capture. This is achieved by computing Q_4. This indicator focusses on the uncertainty interval width relatively to the width obtained from a previous quantification. Therefore, it can be used to evaluate the reduction of the uncertainty interval width. Figure E.6, left, exhibits a negative linear trend meaning that high/low informativeness (i.e. narrow/wide uncertainty interval width) tends to lead to low/high calibration.

Since Q_4 and Q_5 can be computed for each SRQ, they can be plot for a given participant. An example is displayed on Figure E.6, right. This figure provides a synthetic graphical representation of the validation results that can be used to identify the SRQs requiring further analysis and that might involve new input uncertainty quantification.

Table E.1 summarises the main characteristics of the four previous validation indicators.

Table E.1. Summary of the main characteristics of the validation indicators applied in this annex

Indicator		Main characteristics
Calibration (focus on the agreement between uncertainty results and experimental values)	Q_0	Information on uncertainty: LUB, UUB. Check if the experimental value falls inside the uncertainty interval.
	Q_1 $(p_1,p_2,p_3,p_4)=(0.025,0.475,0.475,0.025)$.	Information on uncertainty: LUB_{min}, LUB, RV, UUB, UUB_{max}. Quantitative insight on the spreading of the experimental values within the uncertainty interval taking into account the sub-intervals associated to the vector of interquartiles. Check how well balanced the experimental values are around the reference values.
	Q_5	Information on uncertainty: LUB, RV, UUB. Focus on the position of the experimental value within the uncertainty interval, evaluate the discrepancy between experimental and reference values.
Informativeness (focus on uncertainty interval width)	Q_4	Information on uncertainty: LUB_{min}, LUB, RV, UUB, UUB_{max}. Focus on the relative uncertainty interval width (evaluate the interval width reduction).

20. It is recalled that « validated » results correspond to low, resp. high, values of Q_1, resp. Q_5.

References

Nanty, S., C. Helbert, A. Marrel, N. Perot and C. Prieur (2017), "Uncertainty quantification for functional dependent random variables", in *Computational Statistics*, 32(2), pp. 559-583, Springer Verlag, https://dx.doi.org/10.1007/s00180-016-0676-0

http://openturns.github.io/openturns/latest/user_manual/_generated/openturns.Distribution.html#openturns.Distribution.computeMinimumVolumeLevelSet

ANNEX F - Examples of scale dependence of model input uncertainties for the ATHLET code (Chapter 7)

No.	Parameter	Parameter explanation	Component/ Geometry	Ranges		Reference	Scale
				min	max		
1	OIBUN	Correction factor for interfacial shear in non-dispersed vertical bundle flow	Core	0.01	2.5	1.0	Large scale
2	OIANU	Correction factor for interfacial shear in non-dispersed vertical down-comer flow	Annular geometry – down-comer	0.05	3.0	1.0	Large scale
1	OIBUN	Correction factor for interfacial shear in non-dispersed vertical bundle flow	Core	0.15	2.5	1.0	small scale
2	OIANU	Correction factor for interfacial shear in non-dispersed vertical down-comer flow	Annular geometry – down-comer	0.15	3.0	1.0	small scale
3	OHWFC	Correction factor for single phase forced convection to water (Dittus-Boelter)	All heat slabs	0.85	1.15	1.0	Uniform
4	OHVFC	Correction factor for single phase forced convection to steam (Mc Eligot)	All heat slabs	0.85	1.25	1.0	Uniform

ANNEX G - Illustration of the MCDA method for the input uncertainty quantification of the critical flow model

The purpose of this appendix is to give an example of input uncertainty quantification of physical models in thermal-hydraulic code using SET and applying the MCDA (model calibration through data assimilation) method described in Chapter 5. This work therefore mainly addresses the different steps developed in Element 4 of the SAPIUM approach even if it also requires going through Elements 1 to 3. These last elements are partially addressed and the conclusion emphasises the important issues to tackle in order to improve the results. This annex remains an illustration and the quantified input uncertainties should not be considered as certified to be used in nuclear power plant studies.

G.1 Introduction

Best-estimate (BE) calculation is more broadly used in nuclear industries and licensing process to reduce the significant conservatism for evaluating loss-of-coolant-accident (LOCA). A key feature of BE evaluation requires quantifying the uncertainty of the calculations. The BE methodology, KINS-realistic evaluation model (KINS-REM) (Kim et al., 2005), was developed for independent audit calculations in 1991, and has been improved with code accuracy and statistical methods. It utilises the thermal hydraulic code MARS-KS (KINS, 2016) and treats uncertainty analysis in the nonparametric statistical method based on the third Wilks' formula. In the BE methodology, it is very important to determine the uncertainty input distribution before the uncertainty evaluation is conducted. This includes the uncertainty of physical models and correlation, plant operational parameters, material properties and so forth. The input uncertainty quantification process has been often performed mainly by subjective expert judgment, or the uncertainty input distributions were obtained from manual documents of computer code. In this respect, methods that are more mathematical are needed to reasonably determine the uncertainty ranges. Recently, the statistical methods in input uncertainty quantification process have been used to reasonably determine the input uncertainty ranges of physical models instead of expert judgment. In this study, the MCDA (model calibration through data assimilation) method (Heo and Kim, 2015) was used to quantify the distribution of influential input uncertainties in the critical flow model of MARS-KS thermal-hydraulic code. There are two MCDA methods; one for linear systems and one for non-linear systems. The deterministic MCDA method for linear system assumes that the system behaviour is linear to the uncertain input parameters, while the probabilistic MCDA method for non-linear systems is appropriate to treat non-linear behaviour for complex systems. Before input uncertainty quantification by MCDA method, it is recommended that a linearity test be conducted. These methodologies are implemented in the statistical data analysis toolkit PAPIRUS, which was developed by KAERI in Korea (Heo and Kim, 2015).

This annex is illustrates input uncertainty quantification using the MCDA method. Therefore, it is mainly restricted to Element 4 of the SAPIUM approach. The other SAPIUM elements are partially addressed in order to apply the MCDA method.

In addition, the quantified input uncertainties are applied for the large-scale IET experiment LOFT L2-5 in order to evaluate the effect of extrapolation, contributing to the illustration of the discussions provided in Chapter 7 on scaling.

G.2. Specification and selection of experiments (Elements 1 and 2)

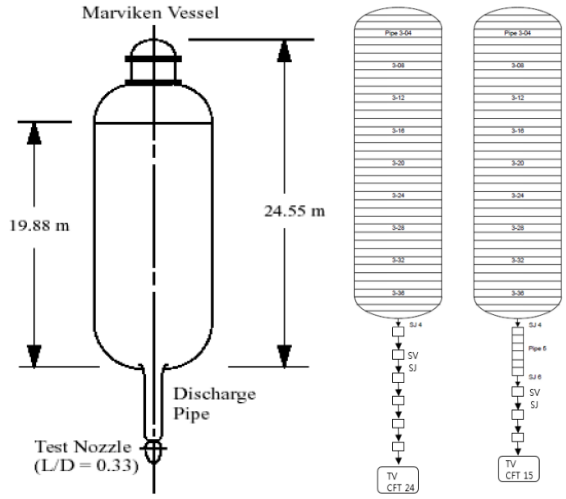
The aim of the Marviken critical flow test (CFT) was to determine the critical mass flow rate of a two-phase mixture of steam and hot water. In input uncertainty quantification, the internal pressure and mass flow are considered as system response quantities (SRQs). Marviken test equipment consisted of four major components: a pressure vessel, a discharge pipe, a test nozzle, and a rupture disc assembly, as shown in Figure G.1. The vessel had an inside diameter of 5.22 m and was 24.55 m high as measured from the vessel bottom to the top of the top-cupola. The net available internal volume was 420 m^3 (USNRC, 2012).

Marviken test 24, a full-scale critical flow test, was selected to check out and evaluate the MARS-KS choked flow model. Because of the short nozzle design (L/D = 0.33) and the long duration (about 20 seconds) of sub-cooling at the break, the test is particularly well-suited for validating the MARS-KS sub-cooled choking model. The MARS nodalisation is also shown in Figure G.1. The vessel was represented by 39 volumes and subdivided from the top. The discharge pipe was modelled by six single volumes. The third and fifth junctions of the discharge pipe were modelled using the abrupt area change option, while the rest were modelled with the smooth area change option. Table G.1 shows that Marviken test 15 has some different initial conditions and long nozzle (L/D = 3.6). The MARS-KS computational model consisted of two pipe components, five single junctions, and time dependent volume as shown in Figure G.1. Therefore, two Marviken tests, 15 and 24, are chosen to quantify input uncertainty ranges in MCDA method.

Table G.1. Initial conditions for Marviken tests

Item/Test	CFT 15	CFT 24
Steam Dome Pressure, MPa	5.04	4.96
Saturation Temperature, ℃	264	263
Normal Sub-cooling at Vessel Bottom, ℃	31	33
Nozzle inlet Initial Temperature, ℃	177	27
Steam Water Inventory, Mg	327	330
Initial Water Level, m	19.93	19.88
Nozzle Length, mm	1809	169
Nozzle Diameter, mm	500	500
Length to Diameter Ratio, L/D	3.6	0.3

Figure G.1. Marviken vessel schematic and MARS nodalisation



G.3. Selection of code and simulation model including uncertainty input parameters (Element 3)

The MARS-KS code has been developed by KAERI for a multi-dimensional and multi-purpose realistic thermal-hydraulic system analysis of light water reactor transients. Choking is defined as the condition wherein the mass flow rate becomes independent of the downstream conditions. This occurs when the fluid velocity equals or exceeds the propagation velocity. The Henry-Fauske critical flow model is the default choked flow model in MARS-KS code (KINS, 2016; KAERI, 2009).

In this code, the final expression for the critical value of the mass flux is:

$$G_c^2 = [\frac{x_o V_v}{\eta P} + (V_v - V_{l,0})[\frac{(1-x_o)N}{s_{v,eq} - s_{l,eq}} \frac{ds_{l,eq}}{dP} - \frac{x_o C_{p,v}(1/\eta - 1/\gamma)}{P_t(s_{v,o} - s_{l,o})}]]_t^{-1} \qquad \text{(G.1)}$$

Where G_c: mass flux, V: velocity, x: quality, s: entropy, C_p: specific heat, P: pressure, subscript o: static condition, c: critical condition, t: throat, eq: equilibrium.

If the thermal nonequilibrium factor, N, is taken to be unity, the prediction of equation (G.1) is close to that of the homogeneous equilibrium model, and if it equals zero the solution is approximately the homogeneous frozen model. Therefore, the quantity N attempts to correlate the partial phase change occurring at the throat. The nonequilibrium factor in terms of the equilibrium quality at the throat is

$$N = \frac{x_{eq}}{C_{ne}} \qquad \text{(G.2)}$$

The implementation of the Henry-Fauske critical flow model in MARS-KS provides for two adjustable coefficients, the traditional discharge coefficient and a thermal nonequilibrium constant, to provide the analyst with the means to better characterise the break. The nonequilibrium constant retains the Henry-Fauske value of 0.14 as a default but can be overridden by the user through input. The thermal non-equilibrium constant is only the region near the saturation line that is greatly affected by this parameter, whereas the discharge coefficient is applied uniformly over all conditions.

First, the critical mass flux is modified by the user specified value of the discharge coefficient. Only one value of the discharge coefficient can be specified for each junction. Then, this single value is applied to the critical mass flux regardless of whether the upstream conditions are subcooled liquid, two-phase, or single-phase vapor. If the mass flux predicted by the critical flow model is less than that resulting from the normal solution of the momentum equations, then the junction is considered to be choked.

Therefore, the non-equilibrium default constant 0.14 and discharge coefficient can be treated as uncertainty input parameters.

G.4. Inverse uncertainty quantification (Element 4)

G.4.1 Chi-squared linearity test

A linearity test is important in evaluating the degree of linearity of the target scenario. A mathematical approach for data assimilation and input uncertainty quantification depends on whether the system behaviour is linear to the input parameters. If the system behaves linearly, the results can be linearised by the first-order Taylor series expansion. Thus if the input parameter distributions are Gaussian and the system responds linearly over the range of the input parameter values, then the calculation result distributions are Gaussian as well. In order to determine the linearity of the system behaviour, a random sampling was employed to develop distributions of the simulation results assuming Gaussian distributions for the parameters. The Chi-squared test statistics are defined as:

$$x^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \qquad \text{(G.3)}$$

Where O_i and E_i are the observed and the expected frequencies for bin i, respectively.

Chi-squared linearity test was performed for Marviken 15 and 24 tests by MARS-KS code with 100 samples of the parameters, respectively.

For Marviken 15 test, the Chi-square values as a function of time (60 seconds) for the pressure are shown in Figure G.2. The result shows the non-Gaussian distribution of the pressure within ten seconds giving a large Chi-square value. In addition, large values of Chi-square for break flow indicate the nonlinearity in early ten seconds and last ten seconds. For Marviken 24 test, the nonlinearity for pressure appears at the early few seconds, while for break flow between 40 and 50 seconds as shown in Figure G.3.

However, because most data during transients has the Chi-squared values below 20, the system behaviour is assumed to be linear with Gaussian distribution. So, for uncertainty quantification in critical flow models, the MCDA for deterministic approaches may be used in this study together with a probabilistic approach.

Figure G.2. Linearity test result for Marviken test 15

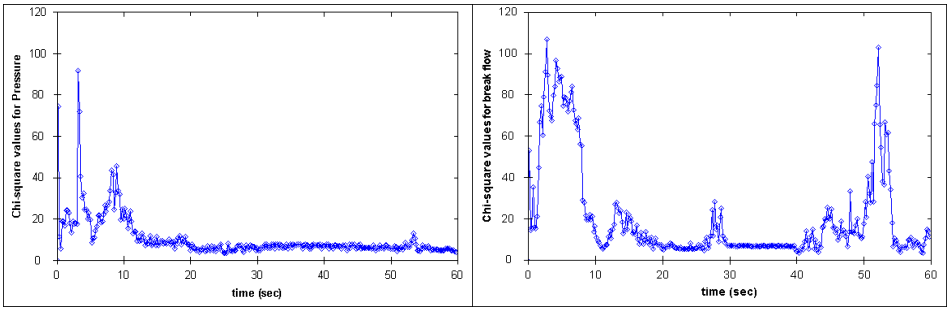
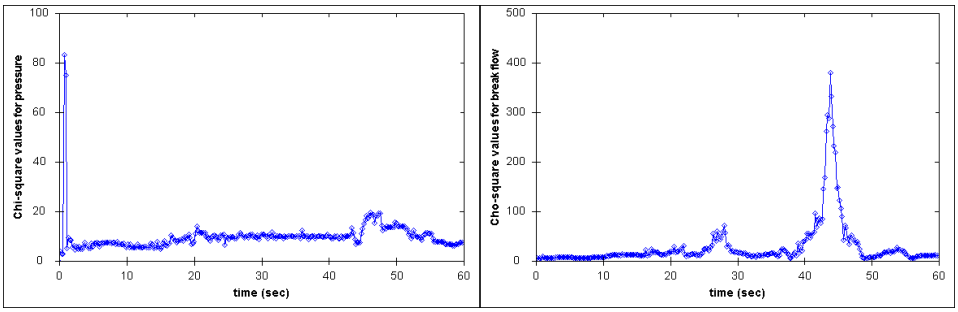


Figure G.3. Linearity test result for Marviken test 24



G.4.2 MCDA method

Inaccuracy in the prediction of physical phenomena can arise from multiple sources of input uncertainties, including physical models, initial and boundary conditions, numeric, etc. For input uncertainty quantification, there are two MCDA methods; one for linear systems and the other for non-linear systems. MCDA integrates experimental data and computational results, for updating the input parameters of the simulation models based on Bayesian statistics. The prior probability distribution of the input parameters is updated with the information from experimental data to calculate the posterior distribution. Bayes' theorem is stated mathematically in the following equation:

$$f(x|\mathrm{Y})=\frac{f(Y|x)f(x)}{f(Y)} \qquad \text{(G.4)}$$

Where *f(x)* is prior probability distribution and $f(x|Y)$ is posterior probability distribution.

Added knowledge contained in the posterior probability density function leads to a reduction of uncertainties in both input parameters and SRQs.

Data assimilation seeks model values α to fit best simulated and measured SRQs based upon Bayes' theorem (Bayes, 1763);

$$\rho(\mathbf{p}\mid\mathbf{r}_m)=c\cdot\exp\left[-\frac{1}{2}\left(\{\mathbf{r}_m-\mathbf{r}\}^{\mathrm{T}}\mathbf{C}_m^{-1}\{\mathbf{r}_m-\mathbf{r}\}+\alpha^2\{\mathbf{p}-\mathbf{p}_0\}^{\mathrm{T}}\mathbf{C}_p^{-1}\{\mathbf{p}-\mathbf{p}_0\}\right)\right] \qquad \text{(G.5)}$$

where p denotes the input parameter vector, r_m the measurement data vector, r the simulation result vector, C_m the measurement error covariance matrix, p_0 the nominal value

vector of the input parameter, C_p the input parameter covariance matrix, α the regularisation parameter, and c the normalisation constant.

For a linear system, a deterministic approach based upon a first-order truncated Taylor series for the responses is used. The mathematical approach used to obtain the quantified input parameter distribution (called the a posteriori distribution of the input parameters) depends on the linearity of the system. In addition, the parameters and observable uncertainties are assumed to follow a normal distribution. For nonlinear relation of responses and parameters, a sampling approach is employed to estimate the posterior distributions of the parameters. This is conducted using the Markov Chain Monte Carlo (MCMC) simulation.

First, the deterministic MCDA for the linear system is studied. To quantify the distribution of the input parameters, users should enter the parameter information including their prior distribution and initial mean value and standard deviation. Therefore, the sensitivity analysis was performed on tests, the number of data, prior distribution, perturbation, etc. In addition, Table G.2 summarises the mean and standard deviation of each input uncertainty parameter depending on the code input, respectively.

Case 1 uses two Marviken tests during 60 seconds for uncertainty quantification, Also, based on the previous study (Bang et al., 2017), the prior distributions are assumed to be normal distribution with mean 1.0 and standard deviation 0.025 for discharge coefficient, and normal distribution with mean 1.0 and standard deviation 0.05 for nonequilibrium constant, respectively. Case 2 has the same inputs except Marviken test 15. Case 6 represents the same as Case 1, but the standard deviation is doubled with the same mean values and perturbation to the nominal values is 0.90 for uncertainty parameters. Case 8 uses the same input except the standard deviation of prior distribution as shown in Table G.2. Case 10 is not the MCDA calculated result, but the arbitrary standard deviation with the mean of case 6 in order to use for envelope calculation. As shown in Figure G.4, the estimated mean values of discharge coefficient and non-equilibrium constant are almost the same for most cases except Case 8, while the standard deviations of both input uncertainty parameters are a little different. The main difference was caused by the prior distribution determined by expert judgement. So, the final value of uncertainty can be decided by envelope calculation in Section 4.3.

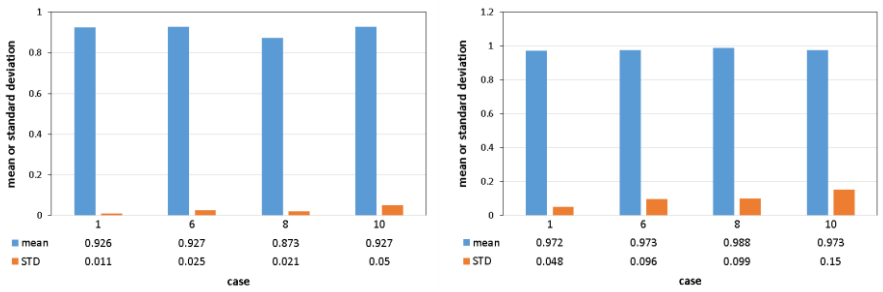
Even though the statistical methods actually aid to decide the input uncertainty for individual parameters, it is inevitable to include the user effect in selecting the data set in both methods because the final uncertainties deeply depend on the simulated results corresponding to the experimental data, prior probability, etc.

Table G.2. Sensitivity analysis and results for deterministic MCDA method

Case	CFT	No. of Data	Prior Dist. for Cd, NE	Perturbation	Cd (m, σ)*	NE (m, σ)*
1	15, 24	300, 300	N(1, 0.025), N(1, 0.05)	0.95	(0.926, 0.011)	(0.972, 0.048)
2	24	300	N(1, 0.025), N(1, 0.05)	0.95	(0.930, 0.011)	(0.972, 0.048)
3	24	200	N(1, 0.025), N(1, 0.05)	0.95	(1.037, 0.003)	(0.457, 0.027)
4	15, 24	300, 300	N(1, 0.025), N(1, 0.05)	1.10	(0.912, 0.007)	(0.908, 0.044)
5	15, 24	150, 190	N(1, 0.025), N(1, 0.05)	0.95	(0.939, 0.005)	(0.829, 0.039)
6	15, 24	300, 300	N(1, 0.05), N(1, 0.10)	0.90	(0.927, 0.025)	(0.973, 0.096)
7	15, 24	300, 300	N(1, 0.05), N(1, 0.10)	0.90	(0.958, 0.017) * nom: 1.0	(0.976, 0.094) * nom.: 0.14
8	15, 24	300, 300	N(1, 0.10), N(1, 0.10)	0.90	(0.873, 0.021)	(0.988, 0.099)
9	15, 24	250, 250	N(1, 0.10), N(1, 0.10)	0.90	(0.952, 0.049)	(0.994, 0.098)
10	Assumed (mean: No.6, σ: assumed)				(0.927, 0.050)	(0.973, 0.150)
11	15, 24	250, 250	N(1, 0.10), N(1, 0.10)	0.95	(0.963, 0.046)	(0.996, 0.098)
12	15, 24	250, 250	N(1, 0.05), N(1, 0.10)	0.95	(0.965, 0.025)	(0.985, 0.094)

* These represent the mean & standard deviation to the multiplier for Cd & NE

Figure G.4. Uncertainty ranges by deterministic MCDA method



Second, about 2 000 simulations were performed using MARS-KS code to predict the thermal hydraulic behaviours such as tank pressure and break flow. The probabilistic MCDA method is used assuming the nonlinear system behaviour, and Table G.3 shows the estimated mean and standard deviation for discharge coefficient and non-equilibrium factor. Those obtained values are somewhat different from those of the deterministic MCDA method. However, the mean and standard deviation for discharge coefficient are similar to Case 8, while the mean for a non-equilibrium factor is generally higher than the deterministic MCDA and the standard deviation has smaller values.

Table G.3. Uncertainty ranges by probabilistic MCDA method

Uncertainty parameter	Mean	Standard deviation
Discharge coefficient	0.8592	0.0245
Non-equilibrium factor	1.0904	0.0499

G.4.3 Confirmation Calculation

The confirmation is performed by the Marviken envelope calculation for the experimental data to evaluate the quality of the uncertainty information how much the experimental data were covered within the lower and upper bound. The uncertainty ranges for two uncertainty input parameters determined in both deterministic and probabilistic MCDA methods were used for envelope calculation, and 124 input files randomly sampled with their uncertainty distributions were generated. The internal pressure and mass flow are considered as SRQs as like the uncertainty quantification.

For the Marviken 15 test, the calculated pressures at the top of tank with uncertainty band of three deterministic cases (1, 6, and 8) are shown in Figures G.5 to G.7. Near the initial time of transient, the MARS-KS code does not predict the pressure rising in the single phase due to the narrow bound, while during the two-phase region, the calculated pressures are well matched with experimental data despite the narrow band. For the break flow, the upper and lower band were increased for about ten seconds as uncertainty ranges become wider, but did not completely cover the experimental data. In addition, in the middle of transient, the code over-predicted the break flow with narrow band. And, with the uncertainty values obtained from probabilistic MCDA, the pressure behaviour near the early time of transient was not well enveloped as shown in Figure G.8, but well predicted in the middle and late time in spite of a narrow band. The break flow until about 20 seconds was well enveloped except in the early five seconds. However, after that, the break flow was over-predicted with the two-phase mixture.

For Marviken 24, the pressure behaviour was almost the same as for the Marviken 15 test, even with a much wider band. For Cases 6 and 8, the break flows until 20 seconds were well enveloped within the lower and upper bounds, while in the middle of the transient with two-phase mixture, the code predicted higher break flow since it is difficult to reach an equilibrium state between liquid and vapour phases due to the short nozzle. The enveloped results were improved, getting uncertainty ranges that are wider, as shown in Figures G.9 to G.11. In the probabilistic MCDA method, the envelope calculation result was not improved during the period of the two-phase mixture in Figure G.12.

Figure G.5. Envelop calculation for Marviken 15 (Case 1)

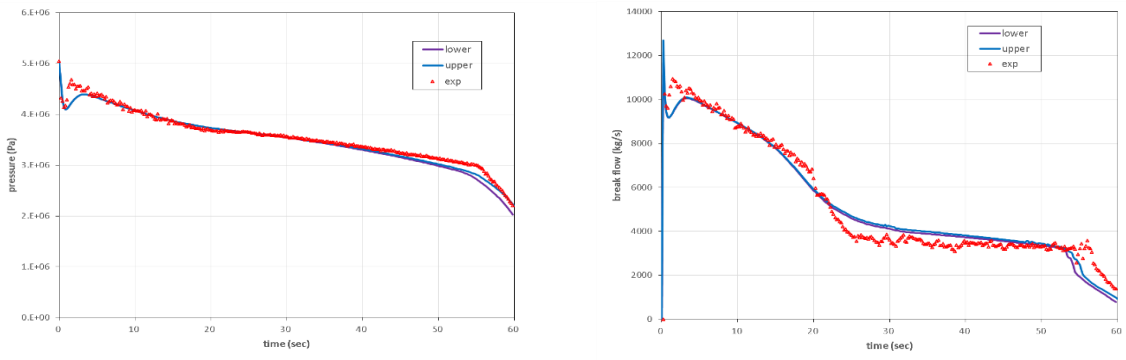


Figure G.6. Envelop calculation for Marviken 15 (case 6)

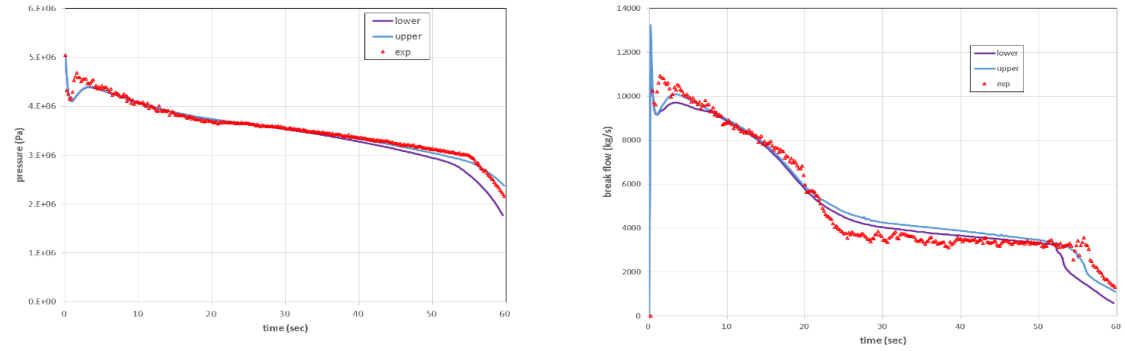


Figure G.7. Envelop calculation for Marviken 15 (case 8)

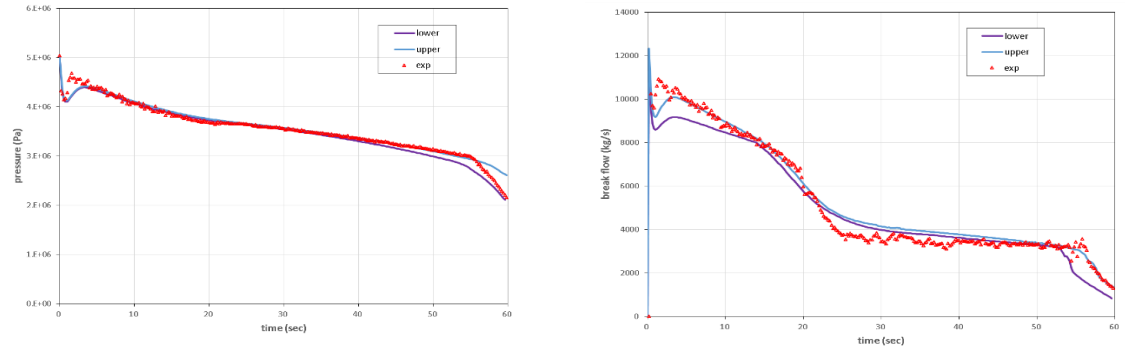


Figure G.8. Envelop calculation for Marviken 15 (probabilistic)

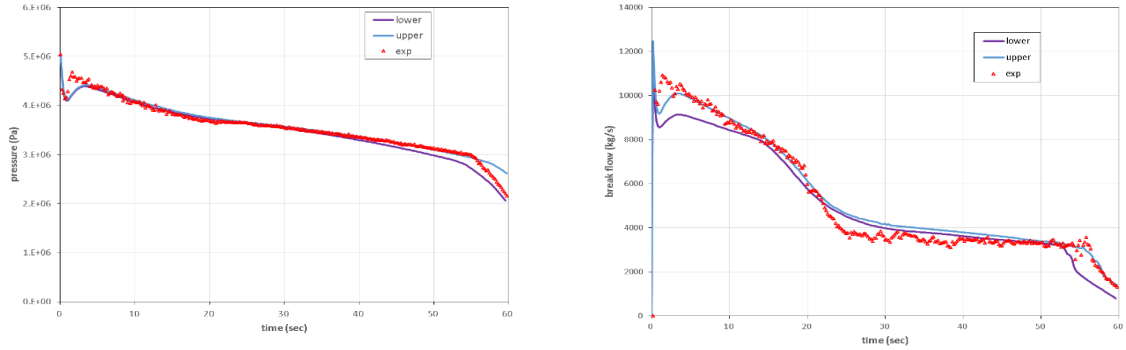


Figure G.9. Envelop calculation for Marviken 24 (case 1)

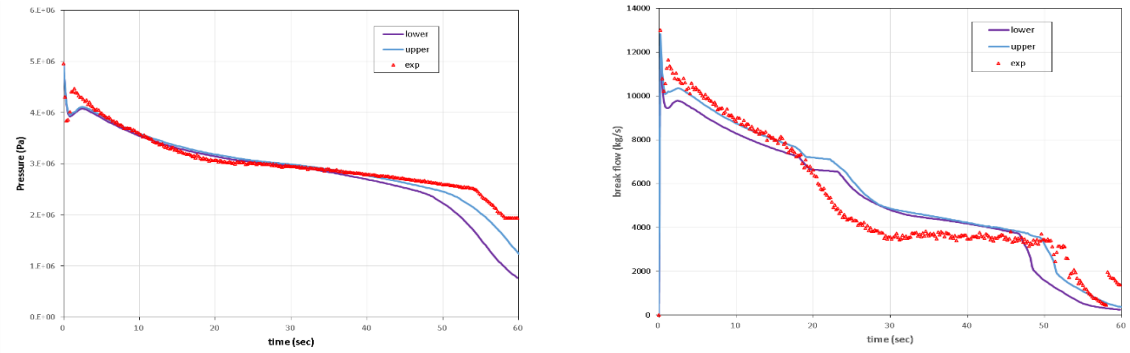


Figure G.10. Envelop calculation for Marviken 24 (case 6)

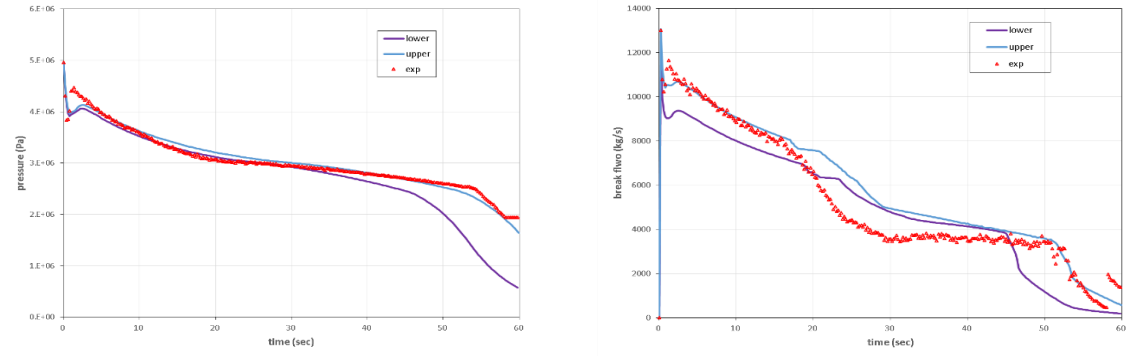


Figure G.11. Envelop calculation for Marviken 24 (case 8)

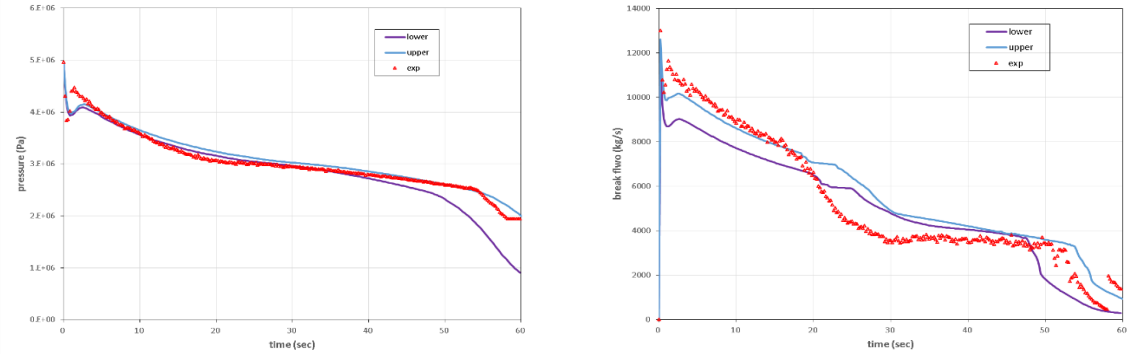
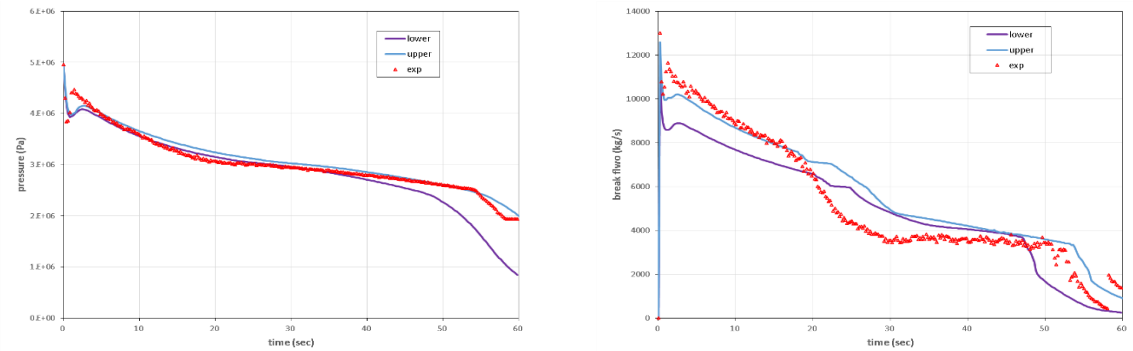


Figure G.12. Envelop calculation for Marviken 24 (probabilistic)



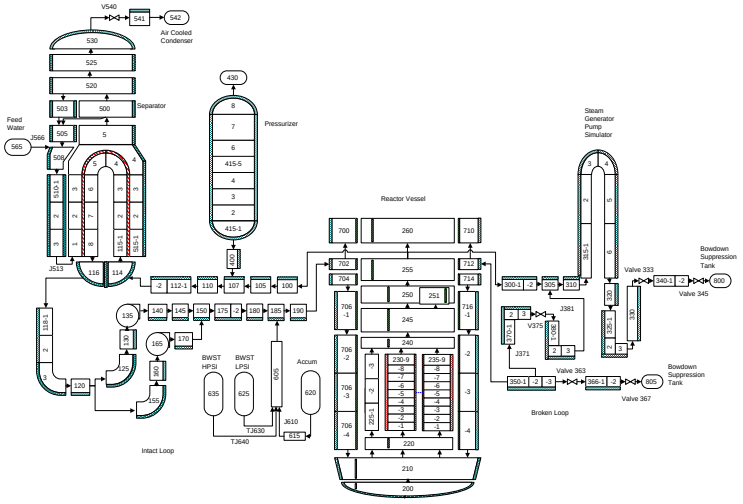
Some reasons for the uncovered region are the nonlinearity of the system behaviour and the deficiency of the critical flow model. In addition, two input uncertainty parameters in the Henry-Fauske model may not be sufficient to describe the critical flow behaviour. These two problems can be tackled in Element 3 of the SAPIUM approach. Considering the envelope calculation results, the uncertainty ranges for two input parameters determined in Case 6 are used as final.

G.5. Evaluation of Scaling Effect for LOFT L2-5 Experiment (related to Chapter 7)

The obtained uncertainty ranges for critical flow models were applied to uncertainty evaluations for IET such as LOFT L2-5 following an input uncertainty propagation method as described in Chapter 6. This is to check whether the statistically determined uncertainty range for a small-scale facility to test the specific thermal-hydraulic phenomena can be extrapolated to scaled-up experiments. Influences for two input uncertainty parameters with respect to the specific SRQs were evaluated from the results of 124 sampled runs performed for uncertainty analysis.

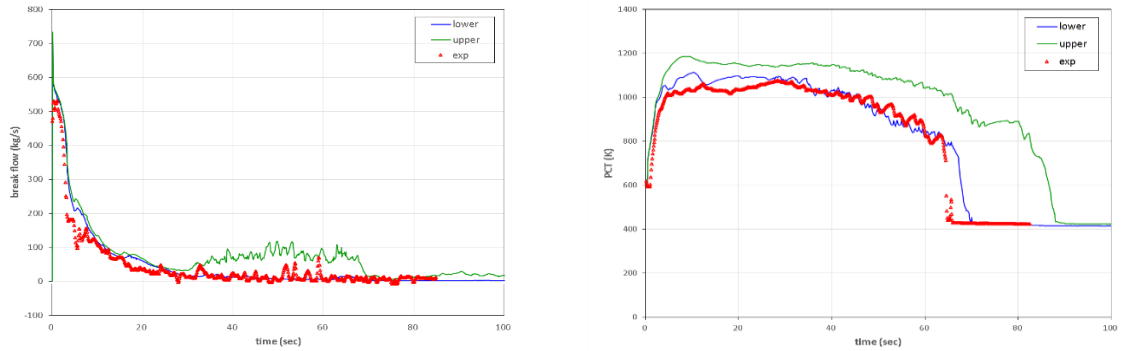
The LOFT facility has been designed to simulate the major components and the system responses of a commercial PWR during a LOCA (USNRC, 1982). For the performance of Experiment L2-5, the LOFT facility was configured to simulate a double-ended 200% cold leg break. Figure G.13 shows the nodalisation for MARS-KS code.

Figure G.13. Nodalisation for LOFT L2-5 test



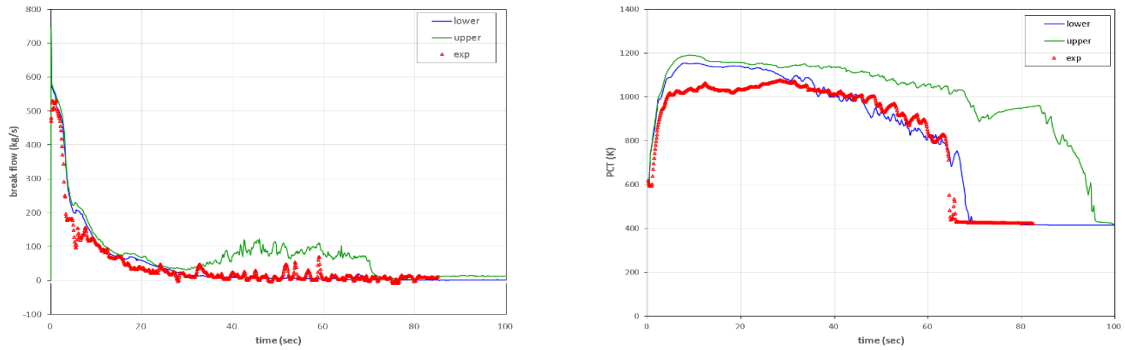
Two input uncertainty distributions for Case 6 were applied to uncertainty calculation for LOFT L2-5. In general, the calculated break flow at the cold-leg side was higher than the experimental data and did not well envelope the experimental data at the early 20 seconds with a very narrow band, as shown in Figure G.14. However, after 30 seconds, the break flow becomes a two-phase mixture and is enveloped by the wider band. Due to the higher break flow, the peak cladding temperature (PCT) was highly predicted in the early time of transient and did not envelope the experimental data in Figure G.14. With only two uncertain input parameters in the critical model, the PCT cannot be enveloped with upper and lower bounds since even best-estimate calculation results by the MARS-KS code showed a higher prediction for PCT. However, if considering the most influential uncertainty parameters, PCT would be enveloped despite over-prediction in the MARS-KS code.

Figure G.14. Confirmation calculation for LOFT L2-5 (Case 6)



With the input uncertainty ranges determined from the probabilistic MCDA, the uncertainty evaluation for LOFT L2-5 was performed. During LBLOCA, Figure G.15 shows a similar trend for break flow as in Case 6. It is judged that the effect of uncertainty variation for the critical model is not large enough to change the break flow substantially. However, the lower bound for PCT with narrower band was higher than that of Case 6. It means that the uncertainties of the Henry-Fauske model do not have a deep impact on the break flow, but the PCT trend during LBLOCA can be influenced a little by the complex thermal-hydraulic behaviour in the reactor coolant system.

Figure G.15. Confirmation calculation for LOFT L2-5 (probabilistic)



G.6. Results and conclusions

This annex is devoted to illustrating the input uncertainty quantification using the deterministic and probabilistic MCDA method in the critical flow model in the MARS-KS code (Element 4 of the SAPIUM approach).

Following Chapter 5, this illustration includes the description and the application of the MCDA method as well as a confirmation step of the quantified input uncertainties. In the deterministic MCDA method, a sensitivity analysis was performed on the kind of test, the number of data, the parameters of the normal prior distribution, and the perturbation to nominal values. The result showed little difference for the input uncertainty results in terms of mean and standard deviation of the quantified normal distribution. The difference is more important when moving from deterministic to probabilistic MCDA, which assumes nonlinear system behaviour. As emphasised in Chapter 5, this type of analysis is therefore of prime importance for a clear evaluation of the user effect and to measure the impact on the uncertainty results of the assumptions associated with each method. The reduction of the user effect is one of reasons why the SAPIUM project was initiated to provide "good practice guidelines".

To evaluate the quality of the input uncertainty, the quantified uncertainties were confirmed by the envelope calculation for the Marviken experimental data itself. Some reasons for uncovered regions may be the nonlinearity of system behaviour and a deficiency of the critical flow model. In addition, the number of uncertainty parameters in the physical model may not be sufficient to describe the complex thermal-hydraulic phenomena. Therefore, it is important to improve the physical model in the code for better accuracy and to find other uncertain input parameters that are important in the model in accordance with Element 3.

The validation process in Element 5 was not performed in this Annex, and more quantitative validation can be possible following Element 5. Additionally, the application calculation for the LOFT L2-5 experiment was conducted with the quantified input uncertainties in the Marviken tests. This work was to evaluate the effect of extrapolation on the large scale IET experiments, which contributed to illustrating the scaling issues on model uncertainty quantification treated in Chapter 7 of the SAPIUM report. Furthermore, the application to nuclear power plants should be more carefully considered through the assessment of the scaling adequacy.

References

Bayes, T. (1763), "An essay towards solving a problem in the doctrine of chance", in *Philosophical Transactions of the Royal Society*, 53, pp. 370-418.

Bang, Y.S. et al. (2017), *Development of Safety Evaluation Technology for New Design Features and New Safety Criteria*, KINS/RR-1696 (in Korean).

Heo, J. and Kim, K.D. (2015), "PAPIRUS, a parallel computing framework for sensitivity analysis, uncertainty propagation, and estimation of parameter distribution", in *Nuclear Engineering and Design*, 92 (2015), pp.237-247.

KAERI (2009), *MARS Code Manual Volume V: Models and Correlations*, KAERI/TR-3872/2009.

Kim, I.G. et al. (2005), *Improvement of the ECCS Best Estimate Methodology and Assessment of LOFT L2-5 Experiment*, KINS/RR-279 (in Korean).

KINS (2016), *MARS-KS Code Manual Volume II: Input Requirements*, KINS/RR-1282 Rev.1.

USNRC (2012), *Assessment of Two-Phase Critical Flow Models Performance in RELAP5 and TRACE Against Marviken Critical Flow Tests*, NUREG/IA-0401.

USNRC (1982), *Experiment Data Report for LOFT Large Break Loss-Of-Coolant Experiment L2-5*, NUREG/CR-2826.