Nuclear Safety
NEA/CSNI/R(2016)18
August 2017
www.oecd-nea.org

Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark

Final Report

BETTER POLICIES FOR BETTER LIVES

NUCLEAR ENERGY AGENCY

Unclassified **NEA/CSNI/R(2016)18**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **24-Aug-2017**

English text only

NUCLEAR ENERGY AGENCY
COMMITTEE ON THE SAFETY OF NUCLEAR INSTALLATIONS

NEA/CSNI/R(2016)18
Unclassified

Cancels & replaces the same document of 16 June 2017

Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark

Final Report

JT03418050

Complete document available on OLIS in its original format

English text only

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 35 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, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, 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 31 countries: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, the Netherlands, Norway, Poland, Portugal, Korea, Russia, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, 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, 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 NEA Committee on the Safety of Nuclear Installations (CSNI) is an international committee made up of senior scientists and engineers with broad responsibilities for safety technology and research programmes, as well as representatives from regulatory authorities. It was created in 1973 to develop and co-ordinate the activities of the NEA concerning the technical aspects of the design, construction and operation of nuclear installations insofar as they affect the safety of such installations.

The committee's purpose is to foster international co-operation in nuclear safety among NEA member countries. The main tasks of the CSNI are to exchange technical information and to promote collaboration between research, development, engineering and regulatory organisations; to review operating experience and the state of knowledge on selected topics of nuclear safety technology and safety assessment; to initiate and conduct programmes to overcome discrepancies, develop improvements and reach consensus on technical issues; and to promote the co-ordination of work that serves to maintain competence in nuclear safety matters, including the establishment of joint undertakings.

The priority of the CSNI is on the safety of nuclear installations and the design and construction of new reactors and installations. For advanced reactor designs, the committee provides a forum for improving safety-related knowledge and a vehicle for joint research.

In implementing its programme, the CSNI establishes co-operative mechanisms with the NEA Committee on Nuclear Regulatory Activities (CNRA), which is responsible for issues concerning the regulation, licensing and inspection of nuclear installations with regard to safety. It also co-operates with other NEA Standing Technical Committees, as well as with key international organisations such as the International Atomic Energy Agency (IAEA), on matters of common interest.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Coordination, compilation and additional writing:

R. Mendizábal, Consejo de Seguridad Nuclear (Spain)

E. de Alfonso, J. Freixa and F. Reventós, Technical University of Catalonia (Spain)

The present report compiles and summarizes the work performed in the PREMIUM benchmark and, as such, must be regarded as a collective achievement. The coordinators wish to thank the effort and collaboration of all the participants in PREMIUM. The PREMIUM Phase I, II, III and IV coordinators deserve special mention where the reports from these phases have been the source for Chapter 3. Tomasz Skorek (GRS, Germany), as Phase III coordinator, has rewritten large parts of section 3.3. Section 5.3 has been largely contributed by Jean Baccou (IRSN, France), Jinzhao Zhang (Tractebel, Belgium), and Estelle Nouy and Philippe Emonot (CEA, France).”

TABLE OF PARTICIPANTS IN PREMIUM BENCHMARK

User	Country	Code	Method	II	III	IV
BelV	Belgium	CATHARE2 V25_2 mod8.1	CIRCÉ	yes	yes	yes
CEA	France	CATHARE2 V25_2 mod8.1	CIRCÉ	yes	yes	yes
CVRez	Czech Republic	RELAP5 mod3.3	CIRCÉ		yes	yes
GRS	Germany	ATHLET 2.2B	Own method	yes	yes	yes
IRSN	France	CATHARE2 V25_2 mod8.1	DIPE	yes	yes	yes
KAERI-1	Korea	MARS-KS1.3-COBRA-TF	CIRCÉ	yes	yes	yes
KAERI-2	Korea	MARS-KS1.3-COBRA-TF	MCDA	yes	yes	yes
KINS	Korea	MARS-KS-0003 PREMIUM version	CIRCÉ	yes	yes	yes
KIT	Germany	TRACE Version 5 patch3	FFTBM	yes	yes	*
NRI	Czech Republic	ATHLET 2.1A	-	yes		
OKBM-1	Russian Federation	KORSAR/BR	CIRCÉ		yes	yes
OKBM-2	Russian Federation	RELAP/SCDAPSIM/mod3.4	CIRCÉ	yes	yes	yes
PSI	Switzerland	TRACE V5.0P3-UQ	Own method			yes
SJTU	China	RELAP5/SCADPSIM/mod3.4	FFTBM		yes	yes
TRACTEBEL	Belgium	RELAP5 mod3.3	IUQ	yes	yes	yes
UNIPI	Italy	RELAP5 mod3.3 patch 3	FFTBM	yes	yes	yes
UPC & CSN	Spain	RELAP5 mod3.3 patch 4	CIRCÉ	yes	yes	yes
VTT	Finland	APROS 5.11.2	CIRCÉ(bias)+FFTBM (range)	yes	yes	yes

Columns II, III and IV indicate the participation in the respective Phase of PREMIUM

* KIT had an incomplete participation in Phase IV

LIST OF PARTICIPATING AUTHORS IN PREMIUM

Institution	Country	Name of authors
Bel V	Belgium	Christine Sarrette
CEA	France	Agnès de Crécy, Philippe Emonot, Estelle Nouy
CVRez	Czech Republic	Milos Kyncl, Rostislav Pernica, Vit Kopecek
CSN	Spain	Rafael Mendizábal
GRS	Germany	Tomasz Skorek
IRSN	France	Jean Baccou, Fabrice Fouet, Pierre Probst
KAERI	Korea	Bub-Dong Chung, Tran Tranh Tram
KINS	Korea	Deog-Yeon Oh
OKBM	Russian Federation	Alexey Gusev, Alexander Falkov, Yuri Shvetsov
PSI	Switzerland	Damar Wicaksono, Omar Zerkak, Andreas Pautz
SJTU	China	Dong Li, Xiaojing Liu
TRACTEBEL Engineering	Belgium	Jinzhao Zhang
UNIPI	Italia	Andriy Kovtonyuk, Alessandro Petruzzi
UPC	Spain	Elsa de Alfonso, Francesc Reventós, Jordi Freixa
VTT	Finland	Torsti Alku, Joona Kurki

EXECUTIVE SUMMARY

Introduction

Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) is a benchmark endorsed by the Working Group on the Analysis and Management of Accidents (WGAMA) of the Nuclear Energy Agency (NEA) Committee for the Safety of Nuclear Installations (CSNI). It is devoted to methods for quantification of the uncertainty of the physical models contained in thermo hydraulic (TH) system codes used in nuclear safety, and their application to a specific thermo-hydraulic scenario. It is addressed to model uncertainties quantification on the basis of 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 uncertainty quantification (UQ).

The benchmark application is 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 have participated in PREMIUM benchmark. A majority of them were involved in all the phases of the benchmark. Some organisations were partially involved, but decided not to participate in one or more phases. Some participants provided more than one contribution.

The participants quantified the model uncertainties using results of selected reflood tests in the FEBA facility, and applying different methods. The resulting uncertainties for model parameters were confirmed 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 uncertainty propagation exercises were analysed with two different procedures: a qualitative one established by CEA and a quantitative one developed by IRSN.

Background

PREMIUM is the follow-up of a series of projects promoted by NEA/CSNI and focused on the issue of best estimate plus uncertainty (BEPU) analysis of thermo-hydraulic accident scenarios in nuclear plants. The series started with uncertainty methodology study (UMS) and continued with best estimate methods uncertainty and sensitivity evaluation (BEMUSE).

The quantification of the uncertainty associated to predictive models is an important issue in the field of BEPU analyses. A majority of present BEPU methodologies are based on the propagation of uncertainties from inputs to outputs of predictive models (implemented in computational codes). For this reason, the various input uncertainties must be analysed and quantified.

The issue tackled in PREMIUM can be widely described as a solution to the inverse problem in modelling and simulation, where model input parameters are estimated (with uncertainty) from comparison of model output magnitudes with experimental data. The solution of the inverse problem can include the processes of model calibration, model uncertainty quantification (UQ), or both. PREMIUM is focused on the quantification of model uncertainty. In some cases model parameters are physical magnitudes, and the inverse methods allow their estimation. In other cases, they are simply adjustable parameters (e.g. multipliers) with no physical meaning, and they are used to calibrate the model and/or quantify its uncertainty.

Development

PREMIUM has been organised in 5 consecutive phases.

In ***Phase I*** (co-ordinated by UPC and CSN), the participants presented and described their methods of model UQ. Two methods were offered to the participants (including assistance for their application): CIRCÉ developed by CEA, and FFTBM, developed by University of Pisa. Several participants used these methods, and others used their own quantification methods. The features of the two series of reflood experiments FEBA and PERICLES were also described in this Phase.

In ***Phase II*** (co-ordinated by the University of Pisa), participants identified influential code input parameters, from the point of view of reflooding, and made a preliminary quantification of their variation range. Phase II co-ordinators proposed a methodology for the identification of influential parameters, based on a set of quantitative criteria, and presented a preliminary list of possibly influential parameters on reflooding. Some participants modified the proposed criteria, or established their own criteria.

Each participant obtained a list of influential model parameters to be quantified, assigning ranges of variation to the parameters, based on sensitivity calculations of test 216 of FEBA, and using as responses cladding temperatures and quench front propagation. The influential model parameters obtained by the participants were typically wall and interfacial heat transfer coefficients, interfacial friction coefficient, heat transfer enhancement at the quench front and droplet diameter.

In ***Phase III*** (co-ordinated by GRS), the uncertainty of influential input parameters (identified in Phase II) was quantified. Participants obtained uncertainties in the form of ranges or probability distributions. The uncertainties obtained depended strongly on the quantification method. Additionally, they depended on:

- The responses used on quantification
- The set of input parameters being quantified
- The TH code, and the specific model being used

The dependency on the quantification method was stronger than for the TH code. The results exhibited a significant user effect and a large variability and discrepancy among participants. In some cases, extremely small uncertainty ranges were found for model parameters, which are physically non-realistic (below the attainable accuracy of experimental data). They were obtained by participants who performed a so-called recalibration of the code, additionally to the UQ.

In ***Phase IV*** (co-ordinated by CEA and IRSN), uncertainties calculated for model parameters in Phase III were propagated through the code simulation 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 uncertainties.

For FEBA tests, most of participants obtained, by propagation, uncertainty bands enveloping the experimental values. The width of the bands varied a lot among the participants. The uncertainty bands were influenced by the responses used in the quantification and by the selected input parameters; and there was a significant influence of the quantification method.

For PERICLES tests, the results were less satisfactory. Considering all the contributions, the fraction of experimental values which were enveloped by uncertainty bands was clearly lower than in FEBA, and far from the expected value. Therefore, the direct extrapolation to PERICLES of the model uncertainties calculated in FEBA gave poor results. This outcome could be explained in terms of the significant differences between the two facilities and the analysed tests.

The quality of the nominal calculation was found important in the final results of the propagation calculations. A good nominal calculation facilitated the enveloping of experimental data by the calculated uncertainty bands.

The results were very dependent on the parameter uncertainties produced in Phase III, and therefore were strongly related to the quantification method, rather than to the TH code.

Quantification methods may have the option of performing calibration, as well as quantification, of the models. Applying such option means a recalibration of the code. Some participants in Phase IV compared results with and without the recalibration option, concluding that results of the extrapolation were improved when recalibration was not performed. This outcome is coherent with the “best estimate” qualification of the TH system codes.

As in Phase III, a user effect was observed in Phase IV results. It was found that participants using the same method and the same version of the same system code obtained uncertainty bands which were significantly different. Differences in the input deck (influencing the nominal calculations), and in the choice of input parameters and responses, could account for the discrepancies.

In ***Phase V*** (co-ordinated by CSN and UPC), the main conclusions and lessons learnt of the benchmark have been drawn, lines of future work have been proposed, and the final report of PREMIUM (the present report) has been compiled and written.

Conclusions

PREMIUM benchmark has been a valuable exercise on methods of UQ of physical computational models, and their application to the models involved in the reflooding prediction. Different methods and thermo-hydraulic codes have been used along the benchmark. Results have been very dependent on the quantification method, rather than on the code. Furthermore, the results of quantification have shown a strong dependence on topics such as:

- The set of selected responses used in the quantification
- The set of selected parameters to be quantified
- The selected database for quantification
- The quantified models and their numerical implementation, which, in general depend on the TH code being used.

There is still a lack of clear “best practice” guidelines on these topics. Indeed, participants in PREMIUM took miscellaneous decisions about them according to their own experience or procedures. It is concluded that the quantification methods used in PREMIUM showed a strong user effect. As a final outcome, the results of quantified uncertainties in PREMIUM showed a large variability and discrepancy among participants.

PREMIUM has been useful as a test bed for inverse UQ methods. Some of the used methods have been developed or improved in the course of the participation in the benchmark. Nevertheless, the application conducted in PREMIUM has not allowed a deep and general assessment of the quantification methods, because it has been 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 development of guidelines for evaluation of model input uncertainty. For methods having the option of performing calibration in addition to the UQ, better results were obtained when such recalibration was omitted.

The propagation of the quantified model uncertainties to FEBA tests has given better results than the analogous exercise for PERICLES tests, in the sense that the calculated uncertainty bands for responses envelop 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 model outputs, based on the quantification of two features: informativeness (depending on the width of the band), and calibration (depending on the closeness of predictions to experimental values).

Lessons learnt and recommendations

Methods for UQ of the physical models in system TH codes must be further studied and developed, so that their different performances can be understood. For methods having the option of performing calibration additionally to UQ, such option is not recommended, because the recalibration seems incoherent with the "best estimate" qualification of system codes. Anyway, the use of "calibrated calculations" as reference cases should deserve further study.

A very important point is how to choose the database for development, verification and validation, and uncertainty quantification (VVUQ) of a physical model. The database will define the range of validity of the quantified uncertainties. A compromise must be found between a specific and a generic standpoint. The database should be specific, in relation to the foreseen application of the quantified uncertainties. In other words, if the model uncertainty is needed for calculating a specific scenario in a plant, the database should include experiments related to the scenario. But the database should be generic enough, so that the quantified model uncertainties are applicable to a wide spectrum of simulations.

Quantification methods are intended for intermediate experiments (IT). Some parameters may be quantified on grounds of separate effects tests (SET). In such case, it is important to have guidelines about how to proceed: using only the SET data for quantification, or combining them with IT data.

The selection of parameters, responses and experimental database are fundamental parts of quantification methods. Guidelines and procedures should be established for such processes. Otherwise, the methods would have a strong user effect. Quantification methods are tools to reduce the engineering judgement, but they cannot eliminate it.

The quantified uncertainty obtained for a specific parameter strongly depends on the total set of parameters being simultaneously quantified. This means that quantified uncertainties are attributes of the total set of parameters, rather than intrinsic properties of individual parameters. The set of quantified parameters must include the most influential ones on the responses; otherwise the resulting uncertainty may be completely misleading. As a conclusion, extrapolation of quantified uncertainties (i.e. application to forward calculations outside the range of validity) may lead to erroneous results.

The experimental uncertainties can be influential on the parameter quantification and therefore should be carefully examined.

Quantification methods may be applied to initial conditions, boundary conditions, material properties, and other input magnitudes having full physical meaning, when there is no other source of information about their uncertainty.

Suggestions for future work

Future work in the issue of BEPU analyses should be focused on the development of a systematic approach, including common "best practice" guidelines on:

- Quantification methods
- Selection of the outputs (responses) and the experimental database used for input quantification
- Selection of input parameters to be quantified
- Code modelling and numerical implementation

Key elements of this approach should be:

- Definition of the objectives of the evaluation, selection of NPP and scenario, and of the important outputs. Use of a PIRT to determine the most important phenomena of the scenario.
- Construction of an adequate (sufficient and representative) experimental database (including SETs, IETS and ITs) for input UQ and validation, which will determine the capability of the method to extrapolate its results to real situations. A ranking of the experiments could be performed using multi-criteria decision analysis methods.
- Choice of a frozen version of the code, derived from an assessment of the applicability of the code to the important phenomena and the experiments. Nodalization strategy and model option selection should keep consistency between the experimental facility and the NPP. Uncertain model input parameters should be finally identified.
- Assignation of input uncertainties, and choice of a method for quantification of model parameters from experimental knowledge. The discrepancy between code predictions and experimental results must be quantified. An appropriate uncertainty modelling for each uncertain input (intervals, possibility or probability distributions,...) is required, taking into account the real state of knowledge and reducing as much as possible extra assumptions. Procedures are needed for combining information of multiple experiments, and for combining input uncertainties.
- Propagation of the input uncertainties through the computer code: the input sampling procedure should be specified, as well as the quantities derived from the output sample being used for validation. A validation metrics should be defined. A consensus on the mathematical definition of "acceptable uncertainty bands" is needed.

Very well-known BEPU frameworks as CSAU or evaluation model development and assessment process (EMDAP) should be taken into account in the task. According to the open issues identified in the PREMIUM benchmark, three interacting axes for further developments are:

- Comparison of different strategies to construct experimental databases for evaluation of TH code model input uncertainty.
- Study on how to deal with several experiments in the input UQ as well as in the input uncertainty validation.

- Preparation of a "good practice document" for a systematic approach, reflecting the two previous points and based on experience from industry, advances in research and development and lessons learnt from BEMUSE and PREMIUM projects.

LIST OF ACRONYMS

AA	Average amplitude
ATHLET	Analysis of thermal-hydraulics of leaks and transients
BAF	Bottom of active fuel
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
CATHARE	Code for analysis of thermalhydraulics during an accident of reactor and safety evaluation
CC	Co-ordination committee
CCFL	Counter current flow limiting
CDF	Cumulative distribution function
CEA	Commissariat à l'Énergie Atomique et aux Énergies Alternatives
CHF	Critical heat flux
CIAU	Code with the capability for internal assessment of uncertainty
CIRCÉ	Calcul des Incertitudes Relatives aux Corrélations Élémentaires
CSAU	Code scaling, applicability and uncertainty evaluation methodology
CSN	Consejo de Seguridad Nuclear
CSNI	Committee on the Safety of Nuclear Installations
DAKOTA	Design Analysis Kit for Optimization and Terascale Applications
DBA	Design-basis accident
DIMNP	Dipartimento di Ingegneria Meccanica Nucleare e della Produzione
DIPE	Determination of Input Parameters Empirical Properties
DNB	Departure from Nucleate Boling
DP	Pressure drop
DSA	Deterministic safety analysis
ECC	Emergency core cooling
ECCS	Emergency Core Cooling System

EMDAP	Evaluation Model Development and Assessment Process
ENUSA	Empresa Nacional del Uranio, SA.
FFT	Fast Fourier Transform
FFTBM	Fast Fourier Transformation Based Method
GRNSPG	Gruppo di Ricerca Nucleare San Piero a Grado
GRS	Gesellschaft für Anlangen – und Reaktorsicherheit mbH
HF	Heat flux
HPIS	High pressure injection system
HT	Heat transfer
HTC	Heat transfer coefficient
IAEA	International Atomic Energy Agency
IBP	Input basic parameter
ICP	Input coefficient parameter
ID	Identification
IET	Integral effects tests
IGP	Input global parameter
INEL	Idaho National Engineering Laboratory
IP	Input parameter
IPSN	Institut de Protection et de Sûreté Nucléaire (former name of IRSN)
IRSN	Institut de radioprotection et de sûreté nucléaire
ITF	Integral test facility
IT	Intermediate test
IUQ	Inverse uncertainty quantification
KAERI	Korea Atomic Energy Research Institute
KINS	Korean Institute of Nuclear Safety
LBLOCA	Large break loss-of-coolant accident
LOCA	Loss-of-coolant accident
LPIS	Low pressure injection system
LSTF	Large scale test facility
LWR	Light water reactor
MCDA	Model calibration through data assimilation
MCMC	Markov Chain Monte Carlo
MFBT	Minimum film boiling temperature
NA	Not available
NEA	Nuclear Energy Agency

NPP	Nuclear power plant
OECD	Organisation for Economic Co-operation and Development
PCT	Peak cladding temperature, peak clad temperature
PDF	Probability density function
PIRT	Phenomena identification and ranking table
PREMIUM	Post-BEMUSE reflood model input uncertainty methods
PSA	Probabilistic safety analysis
PWR	Pressurised water reactor
QF	Quench front
RELAP	Reactor excursion and leak analysis programme
SCDAP	Severe core damage analysis package
SET	Separate effects test
SETF	Separate effect test facility
SBLOCA	Small break loss-of-coolant accident
SRS	Simple random sampling
TGTHSB	CSNI Task Group on Thermal-Hydraulic System Behaviour
TH	Thermalhydraulic
UMAE	Uncertainty method based on accuracy extrapolation
UMS	Uncertainty methodology study
UNIPI	Università di Pisa
UPC	Universitat Politècnica de Catalunya
UQ	Uncertainty quantification
VVUQ	Verification, validation and uncertainty quantification
WGAMA	Working Group on Analysis and Management of Accidents

1. INTRODUCTION

Post-BEMUSE Reflood Models Input Uncertainty Methods (PREMIUM) is a benchmark endorsed by the WGAMA Group of NEA/CSNI. It is devoted to methods for quantification of the uncertainty of the physical models contained in thermo hydraulic (TH) system codes, which are important tools in the simulation of accident scenarios in nuclear reactors. The benchmark application is 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.

The realistic analyses of accident scenarios in nuclear reactors are also termed best estimate plus uncertainty (BEPU) analyses. They are based on the use of mechanistic, best estimate system TH codes, supplemented with uncertainty assessments. One of the most important uncertainties that must be taken into account in BEPU analyses is the one derived from the imperfect (i.e. approximate) nature of the physical models and correlations contained in the code. This uncertainty must be calculated through the comparison of code predictions and real data, and the application of an adequate method. The methods of model UQ solve a specific type of the so-called “inverse problem”, in the realm of modelling and simulation. PREMIUM is focused on this type of methods and thus implies a step forward in the development of BEPU methodologies.

PREMIUM has been organised in five consecutive phases. Phase I has been preparatory of the benchmark. The central activities of PREMIUM have been developed in Phases II, III and IV. Finally, Phase V is devoted to the compilation and summary of the results, the drawing of conclusions, lessons learnt and suggestions for future activities, and the composition of the final report.

1.1 Background

Modelling and simulation are essential elements in the performance of safety analyses for nuclear power plants. The deterministic safety analysis (DSA) of accidents, a fundamental tool for the design of safety-related structures, systems and components in NPPs, has traditionally been performed with conservative models and assumptions. But progress in the understanding and modelling of the accident phenomenology, mainly derived from extensive research programmes, has allowed the development of realistic models, implemented in the so-called best estimate codes. The emergence of these advanced tools and, in addition, of techniques for the analysis of prediction uncertainty opened wide the door to realistic analysis, also termed BEPU.

In the United States, realistic LOCA analyses were admitted for the first time in 1988 [1], and the first BEPU methodology, sponsored by the USNRC and named CSAU, was released in 1989 [2]. Since then, extensive works on BEPU tools have been conducted, and a significant number of BEPU methodologies have been developed.

The Committee on the Safety of Nuclear Installations (CSNI) of the OECD Nuclear Energy Agency (NEA) has been promoting since long activities related to realistic calculations in the realm of nuclear safety. In relation with the validation of codes, two important activities have been sponsored by the CSNI:

- The so-called International Standard Problems (ISPs), which address the TH behaviour by comparing predictions of experiments performed with different codes.

- A compilation of validation matrices for LWR [3, 4], covering integral and separate effects tests, following an exhaustive work of phenomena identification and correlation with tests in experimental facilities.

Three major studies and benchmarks have been promoted by CSNI. They are, in chronological order, UMS, BEMUSE and PREMIUM.

In 1989, a State of the Art Report on "Thermohydraulics of Emergency Core Cooling in Light Water Reactors" was released by CSNI [5]. One of the recommendations of this report was the development, test and comparison of uncertainty methods for TH simulations. In 1994, the CSNI Task Group on Thermal Hydraulic System Behaviour (TGTHSB) organised a workshop, with the goal of discussing different methods of uncertainty assessment. Eight methods were presented to the workshop, and it was concluded the necessity of a step-by-step comparison of the methods. The result was the uncertainty methods study (UMS), approved by CSNI in December 1994. The study was performed from May 1995 through June 1997 [6].

UMS was the first exercise of study, application and comparison of uncertainty methodologies in the field of reactor TH and under the auspices of CSNI. The basic objectives were:

- Comparing different methods of uncertainty analysis, step by step, in the application to a common problem;
- Comparing the uncertainties predicted for specified output magnitudes, among the different methods and with measured values.
- Obtaining conclusions for decision takers (e.g. regulators)

Five organisations took part in the study, namely:

- The University of Pisa (Italy), using the codes RELAP5/MOD2 and CATHARE 2; and their uncertainty method base on accuracy extrapolation (UMAE).
- AEA Technology (UK), using the code RELAP5/MOD3.2 and their own uncertainty method based on input uncertainty propagation and bounding analyses.
- Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) (Germany), using the code ATHLET Mod 1.1, and their own method based on input uncertainty propagation and Wilks' nonparametric procedure.
- Institut de Protection et de Sûreté Nucléaire (IPSN), using the code CATHARE 2, and their own method based on input uncertainty propagation and Wilks' nonparametric procedure.

Empresa Nacional de Uranio, SA (ENUSA) (Spain), using the code RELAP5/MOD3.2 and their own method based on input uncertainty propagation and Wilks' nonparametric procedure.

The experiment used to test the methodologies in UMS was LSTF SB-CL-8, a 5% cold leg small break LOCA conducted in the ROSA IV Large Scale Test facility (LSTF), in Japan. This experiment was very well documented, because it was chosen by CSNI as ISP n. 26.

Five UMS workshops were held between May 1995 and April 1997. The participants had to calculate and assign uncertainty to 3 responses as functions of time (pressuriser pressure, primary circuit mass inventory and surface temperature for a specific rod of the facility), and to 4 scalar responses (first and second peak cladding temperature –PCT-, time of maximum PCT and minimum core pressure difference at a specific location).

The NEA Working Group on the Analysis and Management of Accidents (WGAMA) promoted a follow-up study of UMS, a benchmark activity called BEMUSE ("Best Estimate Methods, Uncertainty and Sensitivity Evaluation"). The objectives of the programme were:

- To evaluate the practicability, quality and reliability of best-estimate methods including uncertainty evaluations in applications relevant to nuclear reactor safety;
- To develop common understanding;
- To promote/facilitate their use by the regulatory bodies and the industry.

The structure of BEMUSE was similar to UMS, in the sense that the participants were different organisations, each one working with one or more specific system TH codes and an uncertainty analysis method, and applying them on two analyses: the post-test analysis of the LOFT L2-5 test (which simulated a large break LOCA), and the best estimate analysis of the LBLOCA in the Zion plant (US).

BEMUSE was organised in 6 phases:

I. Presentation of the uncertainty analysis methodologies used by the participants.

II. Analysis of the ISP 13 (post-test analysis of the LOFT L2-5 test).

III. Uncertainty and sensitivity evaluation of the L2-5 test calculations: first conclusions on the methods and suggestions for improvement.

IV. BE analysis of the LBLOCA in the Zion plant.

V. Sensitivity studies and uncertainty evaluation for the analysis of LBLOCA in Zion plant (with or without the improvements from Phase III).

VI. Synthesis of the results, conclusions and recommendations.

Phases I, II and III were performed from September 2003 through May 2006. Phases IV and V were performed from August 2006 through April 2008. The final report was released in March 2011 [7].

The main conclusions of BEMUSE programme are summarised in [7]. The activity compared the applications of two main types of uncertainty methods:

- Statistical methods based on a probabilistic modelling of input uncertainty, which propagate the uncertainty from inputs to outputs of the calculations. Those used in BEMUSE are based on Wilks' procedure for constructing nonparametric tolerance regions. GRS was the first organisation to use such procedure [49], and so the term "GRS-type methods" is sometimes used. Since their emergence, these methods have enjoyed of a high level of acceptance, and have been adopted worldwide.
- UMAE and its extension code with capability of internal assessment of uncertainty (CIAU), developed by the University of Pisa.

The BEMUSE activity showed pros and cons of these two methods. A number of worthy conclusions and lessons learnt were derived in the activity. Two basic conclusions were:

- Prerequisites for any uncertainty study are a computer code which is well qualified and suitable to calculate the scenario under investigation, an adequate uncertainty method and a qualified user, following quality assurance procedures.

- The quality of base case (reference) calculation, and the selection and quantification of important uncertain input parameters (IP), are essential for an adequate uncertainty analysis of outputs. The selection of important IP (using techniques of sensitivity analysis) minimises the task of assigning input uncertainty.

Specific conclusions about statistical methods were:

- The inclusion of all significant input uncertainties is essential for obtaining reliable results of output uncertainty.
- Simple random sampling is needed for application of Wilks' method.
- If the tolerance limit calculated by Wilks' method approaches the acceptance regulatory limit, the number of code runs should be increased, because this fact reduces the dispersion of tolerance limits.
- If, additionally to uncertainty analyses, reliable sensitivity analyses have to be performed, the number of calculations should be significantly larger than the number of uncertain input parameter.

The final report [7] remarked that "the methods used in this activity are considered to be mature for application, including licensing processes". In the conclusions and recommendations section of [7], it is emphasised that "More effort, specific procedures and judgement should be focused on the determination of input uncertainties". And the following remark was added: "…the method used to select and quantify computer code model uncertainties and to compare their effects on the uncertainty of the results could be studied in a future common international investigation using different computer codes. That may be performed based on the same experiments. Approaches can be tested to derive these uncertainties by comparing calculation results and experimental data".

This remark was a first step towards a new project, the PREMIUM benchmark. A brainstorming meeting was held in Barcelona in April 2011, involving members of CEA, CSN, GRS, IRSN, UNIPI and UPC, who later set up the Co-ordination Committee of the benchmark. A first co-ordination meeting of PREMIUM was organised in Pisa in October 2011. The kick-off meeting of the benchmark took place in Paris, 20-21 February 2012.

1.2 Objectives and Phases of PREMIUM

The PREMIUM benchmark is structured in five consecutive phases:

- Phase I: Introduction and methodology review. Methods for quantification of model uncertainty are presented.
- Phase II: Identification and initial quantification of important input uncertainties for reflood prediction. It has been performed on the basis of the results of test 216 of series I in reflood FEBA experiments.
- Phase III: Quantification of physical model uncertainties identified in Phase II, using results of 6 FEBA tests of the series I (including test 216).

- Phase IV: Confirmation / validation of the results of Phase III, through the propagation of input uncertainties to reflood tests. Firstly, the consistency of quantified uncertainty is checked by making the propagation to the 6 FEBA tests used in the quantification. Secondly, the same input uncertainties are propagated to the prediction of another reflood experiment, PERICLES. This exercise is performed blindly. In the two exercises, the uncertainty bands obtained by propagation are compared to the measured responses of the experiments.
- Phase V: compilation and comprehensive analysis of the results, extracting conclusions, lessons learnt and possible suggestions for future work. The final product of this phase is the present report.

A detailed description of the results obtained and conclusions reached in Phases I to IV is contained in the corresponding phase reports [8-11]. A summary of this information is given in Section 3 of the present report.

2. ELEMENTS OF PREMIUM

2.1 BEPU Methodologies

Safety analyses of nuclear power plants are based on calculations. Specifically, deterministic safety analyses (DSA) make use of simulations by computational models (also named codes) of selected accident scenarios, termed design-basis accidents (DBAs). The regulatory authorities of different countries often impose acceptance criteria on the results of the aforementioned calculations. The final goal of the analysis is assessing the fulfilment of the regulatory acceptance criteria.

Traditionally, the DSA methodologies have been conservative, i.e. based on pessimistic models and assumptions. This was a proper framework in the former stages of nuclear safety, when the knowledge of the plant behaviour during hypothetical accidents was still poor. But the development of experimental programmes, together with advances in theoretical and computational fields, produced better predictive models and opened the door to the possibility of realistic simulations. The new framework was that of BEPU methodologies, based on realistic models and uncertainty assessment of the results. In the transition from conservative to BEPU framework, hybrid methodologies were developed, based on the use of BE codes together with conservative assumptions in selected models and input values. Such methodologies, producing bounding values of the output, are still used in DSA [50].

As mentioned in Section 1.1, the USNRC accepted realistic methods in 1988, in the realm of LOCA-ECCS analyses, and sponsored the development of the “first” BEPU methodology, named Code Scaling, Applicability and Uncertainty Evaluation Methodology (CSAU) [2]. CSAU was a landmark in the BEPU developments, regarded as the basic framework for constructing a methodology. Practically all the BEPU methodologies developed after 1989 follow the basic CSAU scheme. This success encouraged NRC to generalise this framework, formerly devised for LOCA analyses, to all transient and accident methods. The result was the evaluation model development and assessment process (EMDAP), described in the Regulatory Guide 1.203 [16].

The shift from conservative to BEPU methodologies does not solely imply the change of predictive models. The fulfilment of regulatory acceptance criteria also changes. The strict fulfilment required for conservative methods becomes a fulfilment “with a high enough degree of certainty” for BEPU ones.

The great majority of current BEPU methodologies make a probabilistic representation of uncertainty, in the sense that uncertain magnitudes are assimilated to random variables, and the uncertainty of a magnitude is described by its probability distribution. In some cases, other representations of epistemic uncertainty are used, on the grounds of interval estimation, possibility theory, evidence theory, etc. [51, 52, 53]. BEPU analyses consider uncertainties of input parameters (IP) and models, and propagate them into the results.

Presently, most BEPU methodologies are based on the probabilistic representation of uncertainty, and on the propagation of uncertainty, from inputs to outputs, through the model or code. The inputs comprise initial and boundary conditions, material properties and also model parameters (i.e. parameters included in the formulation of models). The first stage of the methodologies is the assignation of probability distributions to the uncertain inputs. The propagation is mainly performed by means of Monte Carlo techniques, using random samples of the uncertain inputs for running the code, so that random samples of the code output are obtained.

For conservative methodologies, the regulatory acceptance criteria were simple restrictions, basically in the form of upper or lower limits, imposed on the values of safety magnitudes. But BEPU methodologies produce uncertain magnitudes, and thus the fulfilment of regulatory acceptance criteria is required with a high level of certainty. Specifically, for methodologies based on crude Monte Carlo (i.e. simple random sampling of the inputs), the fulfilment is required with a high level of tolerance (meaning high levels of probability and statistical confidence). In a sense, the regulatory acceptance criteria of DSA become probabilistic in the BEPU realm.

For proving the fulfilment of criteria, several statistical techniques may be used. One of the most obvious is based on the construction of tolerance limits for the safety magnitudes. A very well-known technique for constructing nonparametric tolerance intervals is Wilks' method, which makes use of order statistics. The German GRS has been pioneer [49] in the use of Wilks' method in the BEPU realm. A great majority of current BEPU methodologies follow this method justifying the term "GRS-type methodologies" [44]. For obtaining a given tolerance level, a minimum size of the Monte Carlo sample is needed, and this gives a measure of the computational effort required in the uncertainty analysis. The so-called Wilks' formula relates the sample size with the tolerance level and order statistics used. The big success of these methodologies is partly explained by the fact that the computational effort needed in the uncertainty analysis does not depend on the number of uncertain inputs. Anyway, uncertain inputs must be quantified (i.e. assigned a probability distribution), and this is not an easy task. So, even for GRS-type methodologies it is advisable to limit the set of uncertain parameters to those conveying most of the uncertainty. For this reason, methodologies must include sensitivity or importance analysis methods, suitable for the selection of influential inputs.

Other possibility for conducting a BEPU analyses is to take advantage of the large volume of data obtained from integral experiments performed in test facilities, in the field of the thermo-hydraulic phenomena of accident scenarios in nuclear reactors. The University of Pisa has developed a methodology of this type, named Code with the Capability of Internal Assessment of Uncertainty (CIAU), and formerly named UMAE [45]. It is a methodology based on the accuracy of calculations for integral experiments, obtained by comparison with the recorded experimental values, and the extrapolation of such accuracy to nuclear plants. Time and quantity accuracies are considered. A large experimental database is needed. In a sense, it can be stated that this methodology propagates uncertainty from the outputs, rather than from the inputs.

All the above mentioned methodologies perform the uncertainty analyses with the original codes. An alternative procedure is the replacement of the code by a surrogate model (also termed meta-model), which is a simple and generally local approximation to the code. This replacement is only for uncertainty propagation purposes. The surrogate model is computationally cost-free, and therefore it can be exhaustively studied (e.g. through random sampling), and the probability distribution of their outputs is obtained practically with zero error. This type of methodology may be useful when the original code is very expensive to run and there is a serious limitation on the size of the Monte Carlo analysis. When a surrogate model is used, the limiting step (in terms of computational effort) is not the uncertainty propagation, but rather the construction of the meta-model. Chronologically, the first methods of uncertainty analysis in the nuclear realm were based on surrogate models. In the former application of CSAU methodology (a large break LOCA in a PWR), polynomial response surfaces were used in the estimation of probability distributions for peak cladding temperatures. As previously stated, most present methodologies propagate the uncertainty directly through the original code.

2.2 Methods for model uncertainty quantification

In science and engineering, the knowledge of a physical system and of the laws governing its behaviour allows predicting the values of physical magnitudes linked to the system. This is the so-called "modelling problem", "simulation problem" or "forward problem" [17]. On the other hand, the "inverse problem" is focused on using the actual result of some measurements to infer the values of other type of magnitudes: the parameters that characterise the system.

In deterministic physics, the forward problem has a unique solution. On the contrary, inverse problems usually have multiple solutions (in general, an infinite number). Therefore, for solving them it is necessary to use any a priori information on the model parameters.

This very general statement about forward and inverse problem may be adapted to the BEPU calculations in nuclear safety. In the BEPU realm, the forward problem consists of modelling a given system (e.g. nuclear reactor) and running simulations with the models (codes). Uncertainties associated to the model and its IP are propagated through the calculation, and the final product of the simulation is calculated values of a set of safety magnitudes (typically those involved in regulatory acceptance criteria) with their uncertainties.

The inverse problem has to do with a "backpropagation" of uncertainty. If the measured value of safety magnitudes is known for a real specific scenario, the uncertainty of some inputs may be inferred. Of course, the uncertainty of many IP is *a priori* known, deduced from measurement or control devices, manufacturing specifications, etc. However, there is a special category of inputs which are the real object of these inverse calculations: the **model parameters**. They are quantities involved in the formulation of mathematical models, different from the conventional ones (i.e. the so-called independent variables).

Sometimes, model parameters are physical magnitudes or constants, having a value that is known prior to the solution of the inverse problem. In other cases, their value is unknown, and they are estimated through the solution of the inverse problem [19].

Another type of model parameters encompass adjustable coefficients with no (or little) physical meaning (e.g. multipliers, adders, etc.). These *empirical* parameters are introduced to account for the imperfection of the model, and are used with two main purposes:

- calibration of the model;
- quantification of the model uncertainty (due to model imperfection), as uncertainty assigned to the parameters.

They are often very influential on the model output.

Model calibration is the process of adjusting model empirical parameters in order to adapt the model predictions to a set of real experimental data. The outcome of the calibration depends on the data base being used. Calibration may be regarded as an optimisation procedure, aimed at a minimisation of model bias (discrepancy between the model predictions and the real or true values).

Uncertainty quantification of the model is the estimation of the uncertainty due to the imperfection of the model. The discrepancies between model predictions and real data are used to estimate the uncertainty of the empirical parameters.

Focusing on empirical model parameters, the inverse problem is the estimation of those parameters from a set of real data of the model output. In the most general setting, the estimation results in both calibration and UQ of the model. Indeed, often in the calibration process parameters are estimated with uncertainty, and the model uncertainty is quantified.

There are cases, however, where only calibration is performed: the parameters are adjusted to deterministic values, without any UQ. In other cases, conversely, the estimation only results in UQ, without any calibration of the model.

Inverse methods are parameter estimation methods, and so many of them are statistical. But there are also non-statistical methods, based on optimisation (for instance, on minimisation of measures of the discrepancies between model predictions and real data). Statistical methods may follow the Bayesian framework or the frequentist one.

Model validation is the process of demonstrating that a model (once developed) produces sufficiently accurate predictions. Validation involves running the model and comparing the predictions to observed data not used in the development of the model [18]. If too large discrepancies are obtained, the model cannot be validated. In some cases, this fact may lead to a recalibration, namely a readjustment of model parameters from the comparison to a new database.

The concept of validation extends to model uncertainty. Once model parameter uncertainties are obtained, they must be confirmed and validated. The propagation of these uncertainties through the model, in a forward calculation, should produce uncertainties in the model output compatible with the real data used in the quantification. This operation may be termed the *confirmation* of the quantified uncertainty. Similarly, the *validation* of this uncertainty is based on the propagation to model output and comparison with real data not used in the development and quantification. In some cases, the validation may use data outside the domain where the model has been developed and/or quantified, and so it may be regarded as a check of the capacity of extrapolation.

Calibration, UQ and validation are fundamental stages in the process of model development. When a developed model has to be applied in a specific analysis, it may be necessary to redo the UQ, using a database more adequate to the application. In such case, validation should also be redone. Recalibration of the model (i.e. calibration using the new database) in not strictly needed, but it is sometimes additionally performed.

The uncertainties of model parameters, as obtained from the inverse quantification, have a main source: the imperfection of the models, which causes the discrepancy between predicted and real (measured) output values. There are additional sources, usually of secondary importance: the measurement uncertainty of the outputs, and the finite size of the database used in the quantification.

In some cases, the UQ of a model parameter is a very simple task, when there is only a single output being influenced by the parameter. In such cases, separate effects tests (if they exist) are very suitable, and the uncertainty is quantified from a direct comparison of model output and experimental results.

In other cases, the model parameters have influence on multiple outputs, and cannot be quantified in a separate fashion. Intermediate tests (IT), rather than SET, must be used in the task. The inverse methods considered in PREMIUM are intended for this kind of situations.

PREMIUM benchmark is centred on methods for the solution of the inverse problem in the BEPU realm, and, specifically, on methods of UQ of model parameters. Information on the methods used in PREMIUM has been compiled from [20] and other sources. They are thoroughly described in Phase I report [8] and their appendices. In the sequel, a brief description of each method will be provided.

CIRCÉ ("Calcul des Incertitudes Relatives aux Corrélations Élementaires") is a method developed by CEA [21], and has been extensively applied to the physical models of the CATHARE 2 code [46, 47]. CIRCÉ is based on the principle of maximum likelihood, and quantifies basic parameters, which fulfil two main assumptions: they follow a normal distribution, and keep a linear relation with the code responses. The basic parameters can be transformed logarithmically, so that CIRCÉ can also quantify parameters following a log-normal distribution.

CIRCÉ has three main inputs: the differences between the real and predicted value of the responses, the partial derivatives of the predicted responses with respect to each parameter, and the experimental uncertainties of the responses. From these magnitudes, CIRCÉ estimates the median value and the standard deviation of the basic parameters. The E-M ("expectation-maximisation") algorithm, based on the maximum likelihood principle and Bayes' theorem, is used in the estimation.

In practice, for a small number of output parameters and a relatively small experimental database, CIRCÉ is suitable for quantifying 1-3 parameters, and the application to a higher number is not recommended. In the current version covariances are neglected; output and IP should be selected carefully, so that they could be considered as independent parameters.

CIRCÉ has the option of performing model recalibration, additionally to the UQ.

MCDA ("Model Calibration through Data Assimilation") has been developed and applied by KAERI [22]. Note that this method was used only as an alternative method to CIRCÉ.

MCDA integrates experimental data and computational results, for the purpose of updating the parameters of the computational models based on Bayesian statistics. The prior probability distribution of the parameters is updated with the information from experimental data to calculate the posterior distribution. The mathematical approach used to obtain the calibrated parameter distribution (called the *a posteriori* distribution of the parameters) depends on the linearity of the system.

For a linear system, a deterministic approach, based upon a first order truncated Taylor series for the responses is used. The parameters and observables 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.

FFTBM ("Fast Fourier Transformation Based Method") is a method developed by the San Piero a Grado Nuclear Research Group of University of Pisa, for the estimation of variation ranges of input uncertain parameters [23]. It is based on the use of the fast Fourier transform (FFT), and has been derived from the BEPU methodology named UMAE. The FFT of the experimental signal and of the difference between experimental and calculated trend is obtained. From the amplitude of this FFT, the accuracy of a code calculation is evaluated. A dimensionless Average Amplitude (AA) is used.

The quantification of variation ranges of IP for physical models is achieved by running calculations of a reference case of the model and "sensitivity" cases, obtained from a single-parameter variation. The FFT is applied for calculating the accuracy of calculated responses with respect to experimental data. The values of AA for sensitivity cases are compared to AA for the reference case. The variation range of an input parameter is derived from established criteria for maximum allowed deviation of an AA from the reference one.

DIPE ("Determination of Input Parameters Empirical Properties") is a method developed by IRSN [24, 25]. In a first step ("Experimental Design") the uncertain parameter X is varied in a given region, and responses are calculated and compared to experimental data. In a second step ("Evaluation of the coverage rate"), the fraction of experimental points covered by the predicted response is calculated. From values of this coverage rate for different values of the input parameter, and on the basis of some simple assumptions, an estimate of the CDF ("pseudo-CDF") of X is obtained. If the experimental data are not bounded by the entire predicted responses, the experimental design of X is enlarged. From the pseudo-CDF, uncertainty intervals can be obtained for the parameter X. The aggregation of multiple pseudo-CDFs (e.g. obtained from different experiments) is carried out by averaging individual curves.

Other methods

- **PSI** is developing a new methodology, based on the use of Bayesian inference, to derive the parameters of the probability distribution of parameters, using a representative experimental SET database for bottom reflooding. However, in PREMIUM only the first phase of the methodology has been used. Essentially, *prior* estimates of the uncertainties are defined in the form of a simplified PDF formulation (symmetric uniform or symmetric log-uniform) and centred on the reference parameters value, which allows a PDF characterisation using a single integer value. The width of each PDF is determined via engineering judgement, based on available literature and previous experience. The method can be summarised as follows:

    - **Selection of the relevant parameters**: A set of input and model parameters for the TRACE code and the targeted application (bottom reflood) is determined using engineering judgement and prior experience of the authors.
    - **Preliminary determination of the uncertainty bands**: A prior estimate of the simplified PDF is derived using quantitative information in available literature, and through local sensitivity analysis based on one reference case (216) of the 6 FEBA tests. The initial uncertainty bands are set as large as possible to allow for a gradual reduction during the next step.
    - **Refinement of the uncertainty bands**: The prior estimates of the PDF bands (for physical model parameters only) are iteratively calibrated from UQ results for the 6 available FEBA tests, and using a simple and pragmatic optimisation method.
    - **Optimisation method:** The error model consists of a **visual** inspection of the results obtained (heater rods temperatures and pressure differences). The criterion for selecting the parameter PDF to be modified is based on its **relevance** (one modification of the most influential parameter per iteration). The end of the procedure (sufficiently narrow PDF bands) is determined from the deterioration of the overall **coverage** ratio (also determined visually). Unphysical PDF configurations (too wide PDF bands) of influential parameters are detected from **visual** comparison against FEBA data.

- **GRS** has applied in PREMIUM a method based on evaluation of separate effect tests and "intermediate" tests. A basic idea is that, even when the analysed models are complex and intermediate experiments are needed, separate effects tests should be applied in order to quantify as many model parameters as possible.

    For these models for which SET exist, the quantification is performed on the basis of those tests. If, in the frame of the intermediate experiments, some models can be related to singular measurements, the models are quantified on the basis of these measurements in a similar way as for the SETs (e.g. in the case of FEBA it was pressure drop measurement in the test section and relative velocity model). For the remaining parameters related to global phenomena, influenced by few models, an iterative procedure is used. With all quantified uncertainties (including those quantified on the basis of SETs and/or measurements), an uncertainty analysis is performed.

The global phenomena that are not adequately simulated are identified. Influential parameters on the uncertainty bands are identified, and their ranges are varied and uncertainty analysis is repeated until the uncertainty bands envelop the experimental data.
In the course of iteration procedure the uncertainty ranges can be optimised (i.e. minimised). In the current version of the methodology this optimisation is based on expert knowledge and visual comparison of uncertainty limits and experimental data. This iteration procedure is similar to the one applied by TRACTEBEL.

- **TRACTEBEL** has developed a sampling-based inverse uncertainty quantification (IUQ) approach. It is based on DAKOTA tool, which has an uncertainty quantification (UQ) functionality based on random sampling applicable to the propagation of input uncertainties to output uncertainties, given a prefixed tolerance level. The DAKOTA team has developed an approach of "calibration (or optimisation) under uncertainty" to find a statistical characterisation of IP such that, propagated through the model, they match statistics of output responses. There is a combination of a UQ method with a deterministic calibration method. The UQ analysis is nested within a calibration loop, and therefore can be computationally demanding.
  The IUQ approach has 4 major steps:

  1) Definition of variation ranges and distributions of key model input uncertainty.
  2) Sampling model inputs, obtaining decks and running the code.
  3) Checking if upper and lower bounds of the calculated output parameters cover the selected experimental data (including uncertainties from experiments and scaling).
  4) If the coverage is not satisfactory, adaptation of the ranges and distributions of key model input uncertainty.

  The steps 2-4 are repeated until a reasonable coverage is found.

2.3 Thermo-hydraulic system codes and reflood models

2.3.1 Thermo-hydraulic aspects of reflood

Reflood is one of the most important thermo-hydraulic processes occurring in a large break loss-of-coolant accident (LBLOCA) scenario in PWR plants. The LBLOCA scenario involves a very rapid depressurisation with significant emptying of the primary system and core uncovery taking place within only tens of seconds. When the primary system pressure falls below the injection pressure of the various ECC systems, borated coolant enters the primary system and flows through the available paths to refill the lower-plenum and then to reflood and finally recover the core.

The reflood phase is one of the classical "phenomenological windows" of the LBLOCA scenario, and begins as soon as the ECC water reaches the hot fuel rods at the bottom of the core. The ECC water moves upwards, gradually quenching (i.e. rewetting) the cladding of the fuel rods, ending their heat up. A quench front (QF) is formed on the fuel rods and large amounts of steam are generated by the energy released from the rods at a high temperature. This steam produces a back-pressure opposing the driving head of coolant in the annulus and thereby slowing or even reversing the water level rise in the core. Thus, reflooding of the core proceeds with level oscillations (strong at the beginning, moderate later) occurring in both the core and downcomer.

The rewetting produces the re-establishment of nucleate boiling or annular-flow boiling, and the large heat transfer coefficients associated to them.

The process previously described is termed "bottom-up quenching". Water entering the core quenches the bottom of the fuel rods, bringing the cladding temperature down to saturation, while the rest is driven upwards through the core as a mixture of steam and entrained droplets. The quench front (QF) is the spatial zone marking the transition from wet-wall to dry wall heat transfer regimes. During the bottom-up reflood, it slowly propagates upwards. The mixture provides some cooling at upper core elevations, where the maximum cladding temperatures prior to the final quench are reached. De-entrainment of liquid may occur at the upper core tie plate and on the structures in the upper plenum. Liquid films appear on these structures. Droplets may be entrained by the steam flow to the hot legs or may fall downwards and back into the core. These droplets and films can lead to the formation of a pool of water in the upper plenum and/or a QF which propagates downwards into the core (“top-down quenching”).

As a result of the high temperatures attained by the clad before the emergency coolant arrives, water does not initially wet the hot clad surface. Rewetting or quenching of hot surfaces occurs when the coolant re-establishes contact with the dry surface. The temperature of the fuel pellets and fuel clad is reduced by heat conduction and convection to the coolant. As the coolant rises in a hot channel or in the overheated nuclear fuel rod bundle, complex heat transfer and two-phase flow phenomena take place and also the succession of regimes moves gradually up the rod bundle channels. The hot surfaces along the channels experiences in turn free- or forced-convection cooling by steam, dispersed flow film boiling, inverted annular film boiling, transition boiling, nucleate boiling and finally single-phase convection to the liquid. Almost all the heat transfer modes are encountered during reflooding and quenching phase. In the boiling curve, QF propagation means a movement from the film boiling heat transfer mode through transition boiling to the nucleate boiling regime.

2.3.2 Relevance to nuclear safety

The rewetting characteristics of the overheated core after a large break LOCA was one of the most active research topics in the 1970s and still has a significant influence on acceptance criteria in licensing and PSA safety analyses. The main interest is related to the maximum temperature in the core, but this turn-over temperature is determined by the liquid dispersed flow well before quenching. Depending on the amount of water available the cooldown takes place earlier or later.

The large temperature gradient in the cladding gives rise to a mechanical stress on the cladding and it may result in fuel damage and radioactivity leakages. The rapid temperature drop is also associated with strong steam generation and this may have an effect on system characteristics including:

- liquid entrainment rate;
- counter current flow limitation in the upper tie plate;
- steam binding in the steam generator;
- three-dimensional distribution of flow in the core.

Relevant responses for reflood are:

- Rod surface temperature Tclad;
- Time when rewet starts trew (i.e. time when abrupt change occurs in the rod surface temperature).

Figure 2.3.1 and 2.3.2 present typical time trends for the rod surface temperature and QF elevation, respectively.

Figure 2.3.1: Example of rod surface temperature time trend

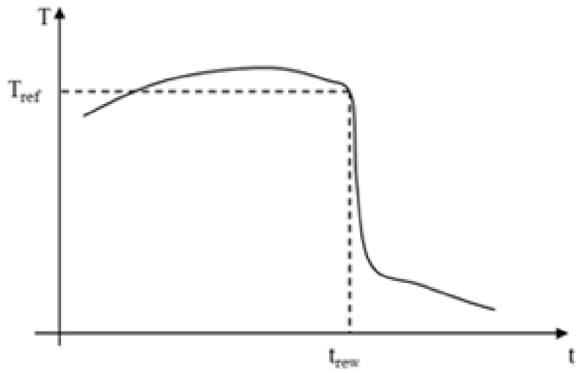
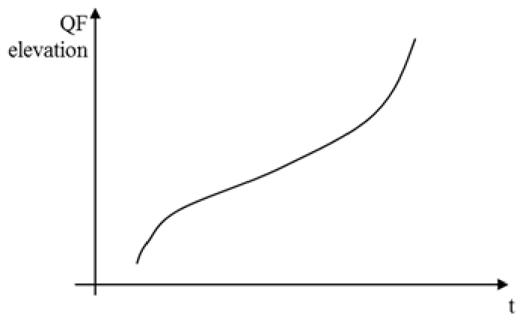


Figure 2.3.2: Example of QF propagation time trend



2.3.3 Reflood models in TH system codes

Methodologies of LOCA analysis must contain adequate models for prediction of the reflood. Specifically, realistic (BEPU) LOCA methodologies make use of best estimate codes able to mechanistically model the phenomena involved in the reflood. Furthermore, the uncertainty of the models used in the reflood prediction must be quantified, so that it can be used in the uncertainty analysis of the LOCA scenario.

The main parameters, driving the QF propagation are primary pressure, ECC flow rate, core void fraction, cladding temperature and minimum film boiling (Leidenfrost) temperature. Regarding the entrainment phenomenon, which plays a significant role during the reflood, presence of restrictions (e.g. spacer grids) or orifices (responsible of steam acceleration), mostly characterised by free area, should be accounted for to estimate the maximum steam velocity.

Advanced system TH codes contain realistic predictive models of reflood. PREMIUM exercises have been carried out using different systems codes, more precisely: CATHARE, APROS, ATHLET, TRACE, RELAP5 and MARS. Each code is using a specific model to deal with reflood or, more exactly, with wall-to-fluid heat transfer and QF propagation.

Usually codes treat the selection of right correlation by considering the different zones of a boiling curve and the QF location.

Libraries of heat transfer correlations are available in each code to allow the right selection. The selection is usually performed by means of evaluating parameters like void fraction, wall temperature, minimum film boiling temperature, wall heat flux and the critical heat flux.

Generally there is some interaction between the treatment of the reflood heat structure and the wall-to-fluid heat transfer. The fact that meshing of the reflood heat structure is refined in the surroundings of the QF has an impact on the way the boiling curve is traced, and as a result on the selection of the different correlations. Also, correlations change when the quench is near to the top of the reflood heat structure.

More detail can be found in Phase II report in which the participants provide a description of the codes that have been used. Next a brief summary for each code is given.

The CATHARE code [26] considers the classical 3 zones of the boiling curve: zone A (wet-wall heat transfer), zone B (Transition boiling) and zone C (post dry out heat transfer). No choice among several correlations is proposed to the user, in order to reduce the user effect. Additionally some physical models are modified in case of reflood (Gas-or vapour-convection, nucleate boiling or film boiling) and some are suppressed (transition boiling). Two heat exchange modes are added for reflood: 2-D conduction wall-to-fluid heat transfer and an evaporation flux added to the interface-wall heat transfer.

In APROS code [27], the wall-to-fluid heat is calculated considering the following heat transfer zones: wetted wall zone (where heat flux is calculated using a forced convection or a nucleate boiling correlation), dry wall zone (where heat flux is calculated using the highest value of the following: film boiling correlation, forced-convection correlation, or natural convection correlation) and the transition zone (where heat flux is interpolated between the critical heat flux and the heat flux calculated for the dry zone).

In the code ATHLET [28], the quench front model determines the current QF position. It defines the boundary between dry position (film boiling or forced convection to steam) and wet position (transition boiling, nucleate boiling). The progression of the QF is calculated on the basis of analytical correlations. The heat transfer in the control volume with QF is calculated on the basis of length weighted average of heat transfer coefficients on rewetted and dry side of the QF. For nucleate boiling conditions the Chen correlation is applied. For film boiling conditions the heat transfer correlation can be chosen between three programmed correlations: Modified Dougall-Rohsenow, Groeneveld 5.9 or Condie-Bengston IV.

In TRACE code [29] the post-CHF heat transfer regime termed dispersed flow film boiling is the most used one. The calculation of the heat transfer coefficient requires the calculation of the single-phase HTC and a two-phase flow correction coefficient.

The RELAP5 reflood heat transfer model [30] has been designed specifically for the reflood process which normally occurs at low flow and low pressure. Besides adding an axial heat conduction model in the heat structures, changes occur in transition and film boiling heat transfer coefficients, both with and without the hydraulic bundle flag activated, when reflood is active. A modified Weismann correlation replaces the Chen transition boiling correlation. The film boiling HTC to liquid uses the maximum of the film coefficient and a Forslund-Rohsenow coefficient. Also, radiation to droplets is added and interfacial heat transfer and interfacial drag are modified when reflood is active.

In MARS code [31], the heat transfer package consists of a library of heat transfer correlations and a selection logic algorithm. Together these produce a continuous boiling curve that is used to determine the phasic heat fluxes. An iterative procedure is used to find the wall temperature at which the heat flux from the Chen nucleate boiling correlation is equal to the critical heat flux. The minimum film boiling temperature is specified as the larger of either equation or that given by Henry’s modification of the Berenson correlation. Heat transfer in the film boiling region is assumed to result from one of two mechanisms: Dispersed flow film boiling or inverted annular film boiling. Heat transfer due to droplets striking the wall is evaluated using the Forslund-Rohsenow equation. The spacer grid heat transfer model originally developed for BART code has been adapted for COBRA-TF.

2.4 Reflood experiments: FEBA and PERICLES Experiments

2.4.1 Experimental facilities investigating the phenomena

As described in Section 2.3, the deep knowledge of reflooding phenomena is very important with a view to the performance and evaluation of LOCA-ECCS analyses. For this reason, a significant number of experimental facilities have been designed and devoted to the study of reflooding transients. Table 2.4.1 provides a list of some separate effect test facilities [4] investigating reflood-related phenomena. Table 2.4.2 provides the list of integral test facilities [3] simulating large break LOCA including reflood phase which are suitable for code assessment.

Table 2.4.1: List of SETF investigating reflood

Facility	Notes	Pressure (MPa)	Inlet mass flow or velocity	Heat flux (W/cm^2)
REWET-II	Triangular array	0.1-1.0	0-15 kg/m^{2}s	20
PERICLES rectangular	Rod bundle	0.2-0.4	0-5 kg/m^{2}s	30-90
PERICLES cylindrical	Rod bundle	0.2-0.4	1-19 kg/m^{2}s	60
TPTF JAERI	Core heat transfer, PWR and BWR bundle	0.5-12	≤ 120 kg/m^{2}s	≤ 20
SCTF JAERI	Large scale	0.6	-	10 MW
CCTF JAERI	Large scale, system	0.6	-	10 MW
GÖTA BWR ECC	Spray cooling, bundle	0.1-2.0	0.045-2.20 kg/s	150-350 kW
ACHILLES reflood loop	PWR bundle	0.3	0.04 m/s	220 kW
NEPTUN-I and –II	Bundle	0.1-0.4	0.015-0.15 m/s	80-140 kW
BWR-FLECHT/GE	Spray cooling, bundle	0.1	0.015-0.15 m/s	10-390
LTSF blowdown/INEL	Single rod, bundle	7	0.4-6.0 m/s	-
RBHT	Rod bundle	0.138-0.414	0.025-0.15 m/s	1.32-2.31 kW/m
FEBA/SEFLEX	Rod bundle	0.2 0.8	0.03-0.1 m/s	200 kW
PDHT-HP	Rod bundle	0.2-2.01	48-1010 kg/m^{2}s	14-360kW/m^2
ERSEC	Single tube	0.3	52 kg/m^2	6.2 kW
THETIS	7x7 bundle	Up to 4 MPa	Max 0.75 kg/s	800 kW
FLECHT-SEASET	10x10 rod bundle	0.17-0.41 MPa	0.01-0.04 m/s	1.7-3.1 kW/m
ROSCO	4x4 rod bundle	0.39		

Table 2.4.2: List of ITF simulating reflood phase of a LBLOCA

Facility name	Facility scale
CCTF	1:25
LOFT	1:50
BETHSY	1:100
PKL	1:145
SEMISCALE	1:1600
ROSA-III	1:424
PMK	1:2070
LSTF	1:48
FIST	1:624
PIPER1	1:2200

Reflooding experiments appear to be suitable for a benchmark application like PREMIUM, because:

- Data for some reflooding tests are available.
- Geometry of the test section is quite simple and average experienced user should not have any problem with its correct simulation.
- Reflooding is a very complex process, involving many physical phenomena. But it can be considered as driven by only a few physical parameters.

During a reflooding experiment, the quenching front progression can be followed by cladding temperature measurements located on different axial levels in the fuel rod. Thermocouples both on the inner and outer sides of the cladding have been located to provide useful data. Due to the time constants of the thermocouples there is only limited possibility for measuring the rapid cooling characteristics in the precursory phase and during the final rewetting. The thermal properties (density, specific heat, thermal conductivity) of the filler material of the fuel rod simulator also have a significant effect to the rewetting characteristics.

In addition to the surface temperature measurements it is essential that the two-phase flow parameters are measured with sufficient accuracy. The minimum instrumentation includes:

- system pressure measurements;
- pressure difference measurements for the core water inventory;
- water inventory measurement in the upper plenum;
- inlet flow measurement both for the net flow and for oscillations.

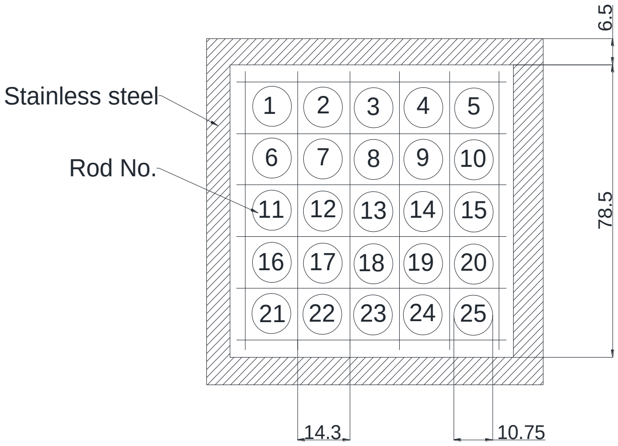
For a more detailed analysis, useful instrumentation includes:

- steam temperature measurement above the swell level;
- droplet size distribution;
- entrained water weighting;
- cladding temperature measurements in different circumferential positions.

2.4.2 The FEBA experimental programme

The FEBA experiment is devoted to the study of the reflood [32]. The experimental data of the six unblocked FEBA tests of the series I were provided to the participants, in order to derive the uncertainty of the physical models influential during reflood. The test section consists of a full length 5×5 rod bundle of electrically heated rods with PWR fuel rod dimensions, surrounded by a housing insulated to reduce heat losses to environment, as shown in Figure 2.4.1.

Figure 2.4.1: FEBA rod bundle – Cross-section view



A cross-sectional view of the FEBA heater rod is provided in Figure 2.4.2.

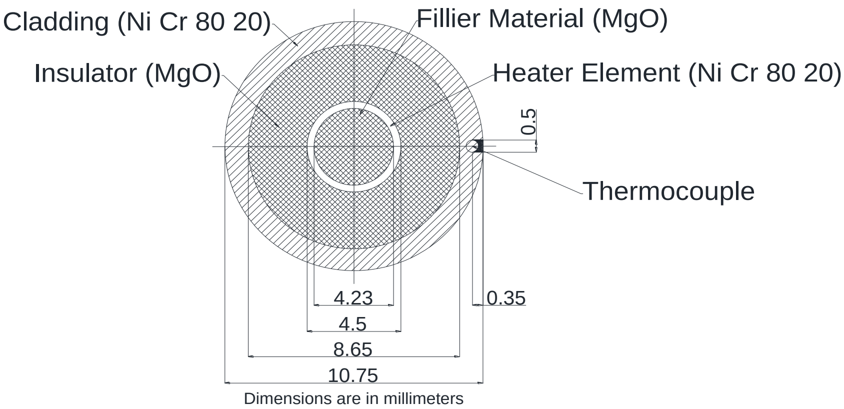
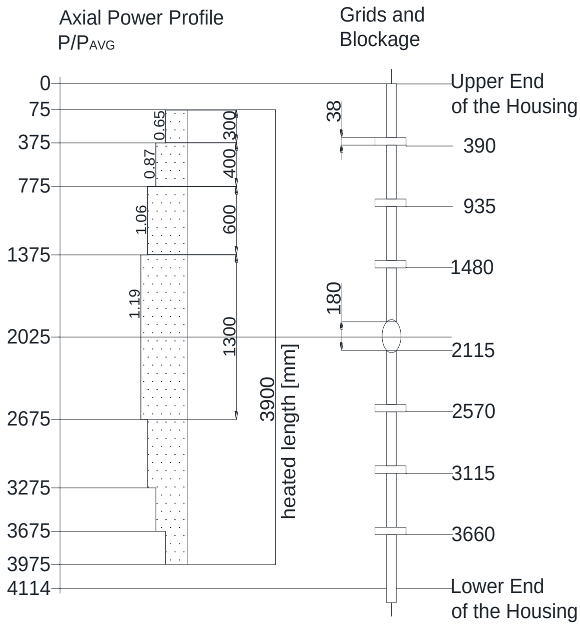


Figure 2.4.2: Cross-section view of the FEBA heater rod

An axial view of the heater rod, and the axial power profile, are shown in Figure 2.4.3

Figure 2.4.3: Axial view of the FEBA heater rod and axial power profile distribution



The power profile is of cosine type approximated by 7 steps of different power density in axial direction. 7 spacer grids are regularly located in the bundle.

Prior to the test run, the fuel rod simulators are heated in stagnant steam to desired initial cladding temperature of roughly 800°C, using a low rod power. In the meantime the test bundle housing is heated passively to the requested initial temperature (roughly 635°C) by radiation from the rods. The aim of choosing a thick wall is to prevent premature quenching of the wall relative to the bundle QF progression.

By starting of the test run, the bundle power is increased to the required level simulating decay heat according to 120% ANS-standard about 40s after reactor shut down. Simultaneously, the cold water supply is activated (its temperature is about 40°C).

The inlet velocity and the pressure are varied for the 6 considered test according to Table 2.4.3.

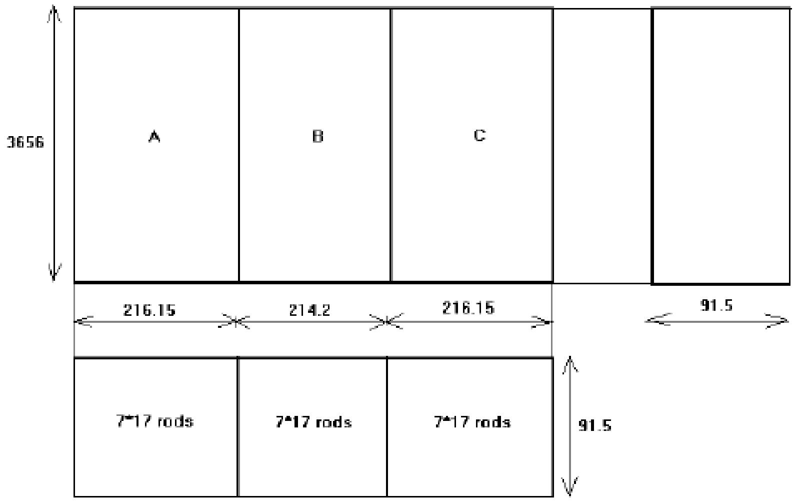
Table 2.4.3: Boundary conditions of the 6 FEBA tests considered for PREMIUM Phase III

Test no.	Inlet water velocity (cm/s)	System pressure (bars)
223	3.8	2.2
216	3.8	4.1
220	3.8	6.2
218	5.8	2.2
214	5.8	4.1
222	5.8	6.2

2.4.3 The PERICLES experimental program

PERICLES has been carried out to investigate 2-D effects which can occur in a PWR core where the rod power is not identical from one assembly to the other ones [33]. The experiment consists of three different assemblies, denoted here by A, B and C (Figure 2.4.4). These assemblies are contained in a vertical cold housing with a rectangular section. Each assembly contains 7×17 = 119 full length heater rods, so that the total number of heater rods is 357.

Figure 2.4.4: The 2-D PERICLES experiment (dimensions indicated in mm)



The rods are heated by two independent electrical power sources, giving the possibility to heat more the central assembly (the ‘hot’ assembly) than the two lateral ones A and C (the ‘cold’ assemblies). The axial power profile is, as for FEBA, of cosine type, with 11 levels.

The experimental procedure is slightly different from that of FEBA, since, at the beginning of the transient, the whole power is immediately switched on until a given initial maximum value of the clad temperature in the hot assembly is reached (generally 600°C). The outer part of the housing is heated, with the aim to maintain its temperature a few degrees above the saturation temperature.

6 tests are considered for PREMIUM benchmark (Table 2.4.5). One is chosen as the reference test, and in the other tests, only one boundary condition is modified with respect to the reference test. The boundary conditions could be:

- nominal heat flux in the hot assembly;
- nominal heat flux in the cold assembly;
- inlet mass velocity entering the bottom of each assembly at the beginning of the reflood stage;
- the sub-cooling of water entering the assemblies;
- the pressure.

Table 2.4.4: Experimental conditions (HA: hot assembly; CA: cold assembly)

Test No	HF_{nom} (HA) W/cm^2	HF_{nom} (CA) W/cm^2	F_{xy}	GO (HA) g/cm^{2}s	GO (CA) g/cm^{2}s	T_{wi} (HA) °C	T_{wi} (CA) °C	DT °C	P (bar)
RE0062	2.93	2.93	1	3.6	3.6	600	600	60	3
RE0064	4.2	2.93	1.435	3.6	3.6	600	475	60	3
RE0069	2.93	2.93	1	3.6	3.6	475	475	60	3
RE0079	4.2	2.93	1.435	3.6	3.6	600	475	90	3
RE0080	4.2	2.93	1.435	5	5	600	475	60	3
RE0086	4.2	2.93	1.435	3.6	3.6	600	475	60	4

The RE0064 test is the reference test. In other tests, only one boundary condition is modified with respect to RE0064. In Table 2.4.4, HF_{nom}(HA) and HF_{nom}(CA) are the nominal heat fluxes in the Hot and Cold Assemblies respectively, F_{xy} is the radial power peaking factor, i.e. the ratio HF_{nom}(HA)/HF_{nom}(CA), GO is the inlet mass velocity entering the bottom of each assembly during the reflood stage, T_{wi} is the initial cladding temperature in the middle of each assembly (at the beginning of the reflood stage). The value for the hot assembly must be reached in the calculation to launch the reflood. The value of T_{wi} in the cold assemblies should also be reached if possible. DT is the sub cooling of water entering the assemblies.

2.5 Methods for the analysis of calculated uncertainty

In Phases III and IV of PREMIUM, the quantified uncertainties obtained for reflood model parameters on FEBA have been validated by propagating them to simulations of FEBA and PERICLES tests, and comparing with the measured data. Methods for analysis of the results of UQ and propagation, including their comparison with real data, are needed to perform these studies.

In Phase IV, two methods of analysis of results have been used by CEA and IRSN, as detailed in [11].

CEA analysis has been qualitative, and is based on answers to 4 questions:

1) Does the uncertainty band envelop the experimental time trend?
2) How is located the nominal calculation with respect to the experiment (under or overestimation).
3) How wide is the uncertainty band?
4) For CIRCÉ users, does the calibrated calculation improve the nominal calculation?

A quantitative methodology developed by IRSN, has been also applied to PREMIUM results. It is based on information fusion [35]. Each uncertainty study performed by each participant can be viewed as an information source. A formal method is defined to combine all the sources, and to estimate agreement or disagreement among them. The methodology is introduced in [36] and [37] and is reformulated in [34] in the framework of the possibility theory [38]. It is briefly recalled when the information on uncertain parameters is summarised by an interval [LUB, UUB] and a reference value RV.

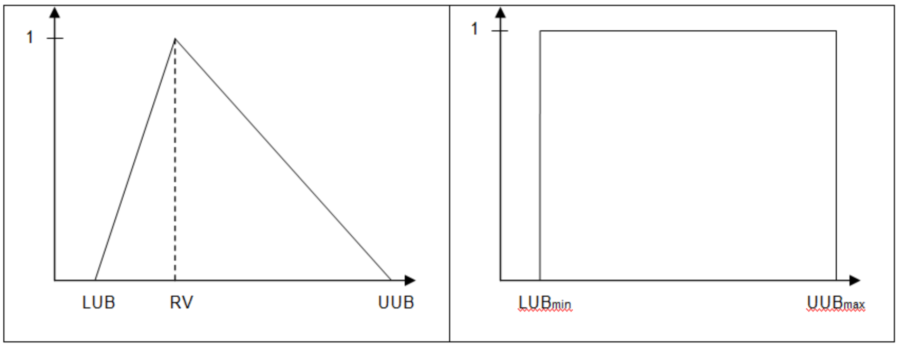
The methodology has three steps:

- INFORMATION MODELLING: a triangular model is associated to each output of interest and each participant. For a source s and an output variable v, an interval [LUB, UUB] and a reference value RV are given and,

$$\pi_{s,v}(t) = \begin{cases} \frac{1}{RV - LUB}(t - LUB) & \text{if} \quad t \in [\text{LUB, RV}] \\ \frac{1}{RV - UUB}(t - UUB) & \text{if} \quad t \in [\text{LUB, RV}] \end{cases} \tag{2.5.1}$$

meaning that the output is modelled as a fuzzy variable.

Figure 2.5.1: Information modelling associated with a source (left) and with complete ignorance (right)



- INFORMATION EVALUATION: for each source s, two scores are calculated, in order to quantify the quality of the information on N output variables $\{v_i\}_{i=1,\dots,N}$

  - **Informativeness**, which measures the precision of the information, and thus its usefulness. The possibilistic model of complete ignorance associated with the output variable is defined as a rectangular fuzzy variable,

$$\pi_{ign,v}(t) = \begin{cases} 1 & \text{if } t \in [LUB_{\min}, UUB_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{2.5.2}$$

where LUB_{min} (resp. UUB_{max}) is the minimum (resp. maximum) of all the values of LUB (resp. UUB) (Figure 2.5.1). An index I(π,s,v) is assigned to each variable and source, as the difference between the area of the rectangle and that of the triangle, divided by the former. Therefore:

$$I(\pi,s,v)=1-\frac{UUB-LUB}{2(UUB_{\max}-LUB_{\min})} \qquad (2.5.3)$$

Index I takes values in the interval [0.5, 1]. Values close to 1 (resp. 0.5) indicate that the uncertainty range is narrow (resp. large).

For N output variables, a global informativeness score is defined as the mean:

$$I(\pi,s)=\frac{1}{N}\sum_{i=1}^{N} I(\pi,s,v_i) \qquad (2.5.4)$$

- **Calibration**, which measures the coherence between information provided by the source and the true (known) value, v* :

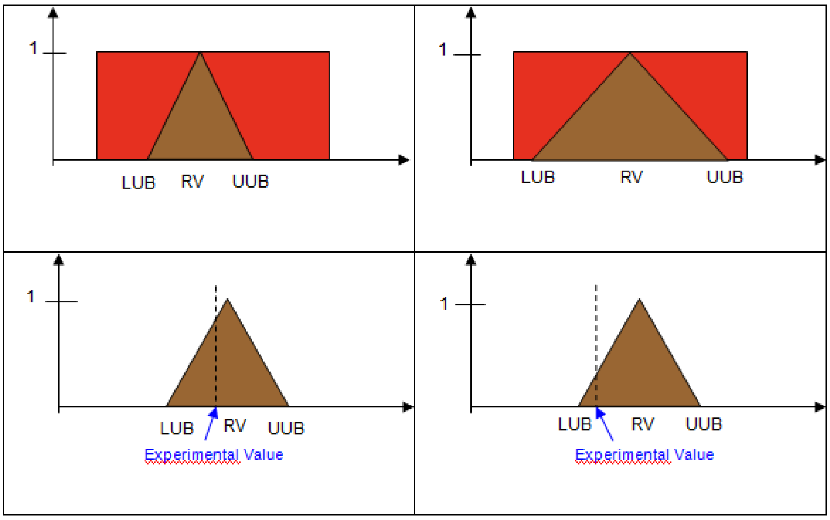
$$Cal(\pi,s,v)=\pi_{s,v}(v^*) \qquad (2.5.5)$$

A value close to 1 (resp. 0) of this score means that the absolute error between the reference value and the true one is small (resp. large). For N output variables, a global calibration score is defined as the mean

$$Cal(\pi,s)=\frac{1}{N}\sum_{i=1}^{N} Cal(\pi,s,v_i) \qquad (2.5.6)$$

Figure 2.5.2 shows the computation of the two defined scores for two sources of information. In the figure, the informativeness and calibration of the second source (on the right) are lower than those of the first source (on the left).

Figure 2.5.2: Computation of the informativeness (top) and calibration (bottom) criteria associated with two sources



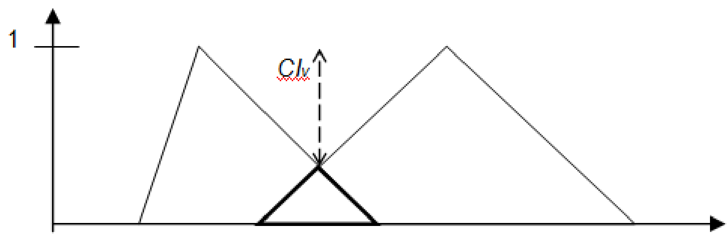
- INFORMATION SYNTHESIS: for each output variable v, the information provided by p sources $\{s_i\}_{i=1,\ldots,p}$ must be aggregated. There are three kinds of aggregation operators in the methodology. For the application to PREMIUM, only the agreement (or disagreement) among participants is studied, and the aggregation is performed by taking the intersection of information. Starting from the p models $\{\pi_{si,v}\}$, i=1,…,p, given by each source, the synthesis result is defined as: $\pi^{\cap}{}_v = \min_{i=1,\ldots,p} \pi_{s_i,v}$ (2.5.7)

  This operator produces precise but potentially unreliable results in case of empty intersection. For the output variable v, the conflict indicator is defined as

$$CI_v = 1 - \max_{t \in IR} \pi^{\cap}{}_v(t) \quad (2.5.8)$$

A value of CI close to 1 (resp. 0) indicates disagreement (resp. agreement) among sources. Figure 2.5.3 shows an example of aggregation between two sources.

Figure 2.5.3: Two sources of information (thin line) and their aggregation (thick line)
The dashed arrow represents the conflict indicator



3 PHASES OF PREMIUM

3.1 Phase I: Introduction and Methodology Review

The objective of Phase I was to specify the definition of the benchmark and present the available methodologies of model UQ to be used in the benchmark [8, 12]. It was co-ordinated by UPC and CSN.

Phase I report [8] contains

- A description of the general objectives of the PREMIUM benchmark.
- A brief description of the objectives and co-ordinating institutions for the 5 phases of PREMIUM.
- The list of 16 institutions that decided to participate in the benchmark, grouped in 15 work teams (CSN and UPC participate as a single team), together with the system thermo hydraulic code and the quantification method used by each participant (see Table of participants in PREMIUM benchmark, in the present report). Several national regulatory authorities are involved in PREMIUM.
- A number of appendices containing detailed description of the quantification methods used by the participants.

In [8] the participants are categorised according to the type of quantification method:

I. Participants having at their disposal methodologies for quantification of uncertainties of the physical models: CEA, UNIPI.

II. Participants willing to become users of the available methodologies: BelV, SJTU, CVRez, VTT, KIT, KINS, OKBM and UPC.

III. Participants willing to use an expert judgement based method improved with methods of fitting of data: PSI.

IV. Participants willing to develop and use their own method in parallel with PREMIUM participation: VTT, TRACTEBEL, IRSN, GRS, KAERI.

Several participants are in more than one category.

3.2 Phase II: Identification of influential input parameters.

The goal of Phase II is the identification, by each participant in the benchmark, of the physical models included in their codes which are influential in the reflooding scenario; and the selection of related uncertain parameters and quantification of their uncertainty (in terms of range of variation, or probability distribution), based on sensitivity studies [9, 13]. Data from FEBA/SEFLEX experiment have been used in the task. As a result, the participants have reported the following information:

- Identification of influential phenomena;
- Identification of the associated physical models and parameters, depending on the TH code;
- Quantification of the uncertainty of parameters, by means of sensitivity calculations.

The co-ordinator of Phase II has been the University of Pisa. In the specifications report for Phase II [9], the co-ordinators made a series of suggestions to the participants.

Concerning the input parameters (IP), the co-ordinator proposed some definitions:

- Input Global Parameter (IGP): an input parameter associated with a physical model (e.g. heat transfer coefficient).
- Input Basic Parameter (IBP): an input parameter that can be a boundary or initial condition, a geometric parameter, a material property or a discretisation parameter.
- Input Coefficient Parameter (ICP): a single coefficient inside a correlation.

A procedure for identification and selection of influential IP was proposed to the participants, with 6 steps:

- Step 1: an initial list of IP is set up, by using the knowledge of the related phenomena and the code models. Engineering judgement should be applied in order to screen out non-influential parameters.
- Step 2: best estimate values of the selected parameters are chosen, and they are used to run a code reference calculation, producing the main responses.
- Step 3: a set of quantitative criteria for selection of influential IP is established.
- Step 4: performance of sensitivity code runs, corresponding to the criteria defined in Step 3. The procedure for choosing the value of the parameters may be an "experimental design" (e.g. one-at-a-time variation) or a sampling (e.g. simple random sampling or Latin hypercube sampling). For each sensitivity run, main responses are analysed.
- Step 5: the criteria of IP selection are applied. If a participant decides to keep a parameter that does not meet the criteria, a reasonable justification must be provided.
- Step 6: obtain the list of influential IP, with their variation range and/or probability distribution.

A sample list of parameters, potentially influential on reflood phenomena, was provided by the co-ordinator in the specifications report [13].

The influential IP identification has been performed based on the experimental test 216 of FEBA facility. Geometrical properties, boundary conditions and measured data were provided to benchmark participants

A set of criteria for selection of influential IP has been proposed by the co-ordinator in the specifications report [13], so that an IP is considered as influential whenever its extreme value in the range of variation causes the following change in the two main reflood responses:

- Criterion 1: change in rod surface temperature (in absolute value) higher or equal than 50 K.
- Criterion 2: relative change in rewet time (in absolute value) higher or equal than 10%.

An additional confirmation criterion is:

- Criterion 3: the variation in elevation of the QF versus time is at least 10%

After the IP has been selected, three additional criteria are applied, in order to ensure the validity of its uncertainty:

- Criterion 4: the variation of an IP should not cause drastic changes in rod surface temperature time trends (e.g. sudden deviations or oscillations), which may be caused by phenomenology different to reflood, or by physical or numerical instabilities.
- Criterion 5: the range of variation shall be consistent with the level of knowledge on the IP.
- Criterion 6: if a preliminary uncertainty evaluation is available, the single IP should not be responsible of the overall uncertainty of the responses.

13 organisations were involved in Phase II, using 8 different codes (see Table of Participants in PREMIUM Benchmark). All of them were 1D system TH codes, except one, COBRA-TF which is a sub-channel module of the system code MARS-KS.

Most of participants modelled the test section of FEBA as a single vertical channel and a single heat rod/heat structure. The sole user of TRACE was KIT, and applied a CHAN component, representing a 5×5 bundle. KAERI modelled 1/8 of the bundle with COBRA-TF. All participants included the model of the test section housing. About the spacer grids, some participants represented them by a flow area reduction and activated special models for HT enhancement; others simply applied form loss coefficients.

The number of axial nodes used by the participants ranged from 20 to 78 (Table 3.2.1). This number does not take into account the possible refinement performed by the codes in the vicinity of the QF, as a part of their reflood models.

Table 3.2.1: Characteristics of adopted nodalizations

Participant	Code	Number of axial nodes	Max linear heat rate, W/cm
GRS	ATHLET 2.2B	23	N/A
NRI	ATHLET 2.1A	66	24.4
Bel V	CATHARE 2 V2.5	46	24.4
CEA	CATHARE 2 V2.5	40	24.4
IRSN	CATHARE 2 V2.5	39	24.4
KAERI	MARS-KS1.3(COBRA-TF Module)	26	24.4
KINS	MARS-KS-003	39	23.8
OKBM	RELAP/SCDAPSIM Mod3.4	39	24.4
TRACTEBEL	RELAP5 Mod3.3	51	N/A
UNIPI	RELAP5 Mod3.3 patch3	20	24.4
UPC	RELAP5 Mod3.3 patch4	25	24.3
KIT	TRACE Version 5 patch 3	43	N/A
VTT	APROS 5.11.02	78	24.4

Base calculations were performed by the participants. The results for cladding temperatures at bottom, 2/3 height and top of the active part and QF propagation are respectively presented in Figures 3.2.1 to 3.2.4.

Figure 3.2.1 FEBA: Cladding temperature at BAF in base case

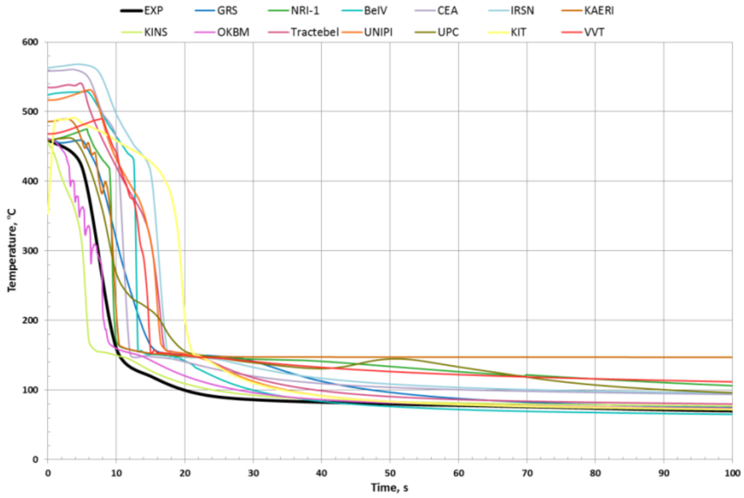


Figure 3.2.2 FEBA: Cladding temperature at 2/3 height in base case

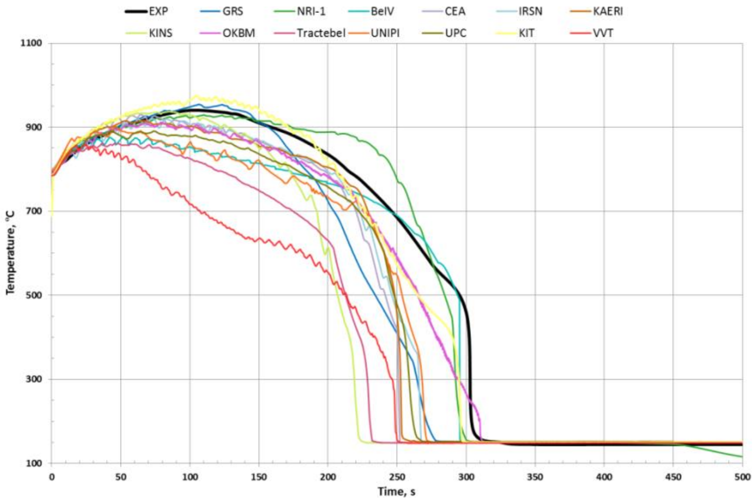


Figure 3.2.3 FEBA: Cladding temperature at TAF in base case

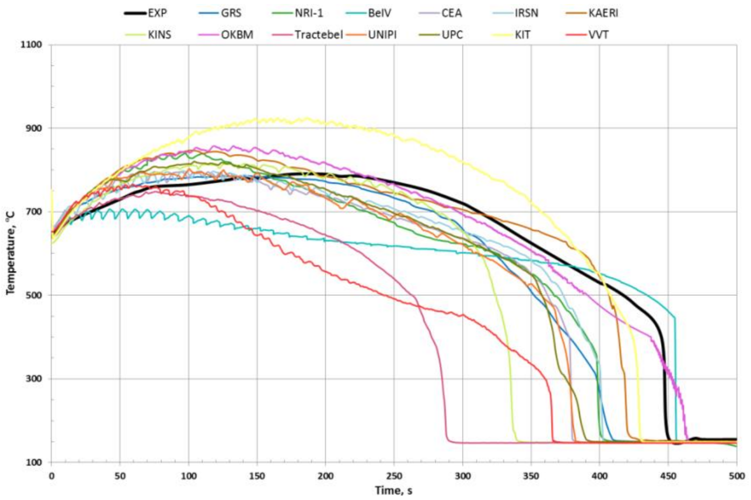
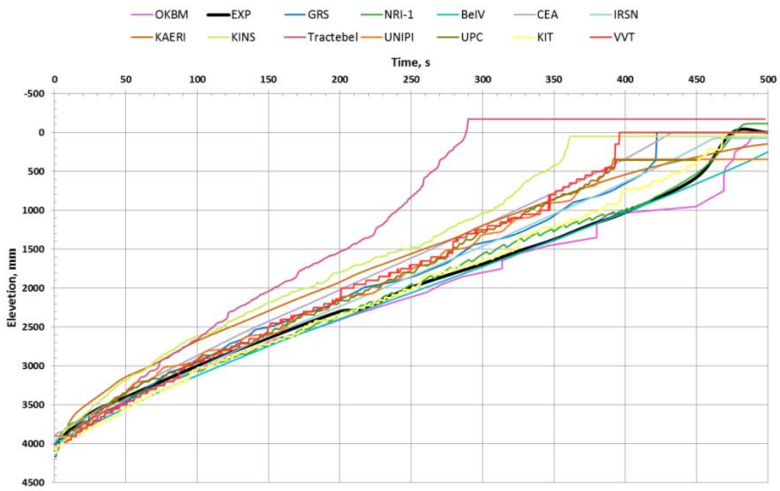


Figure 3.2.4 FEBA: Quench front propagation in base case



Figures 3.2.5 and 3.2.6, and Table 3.2.2 summarise the results from the participants.

Figure 3.2.5 PCT comparison in base case

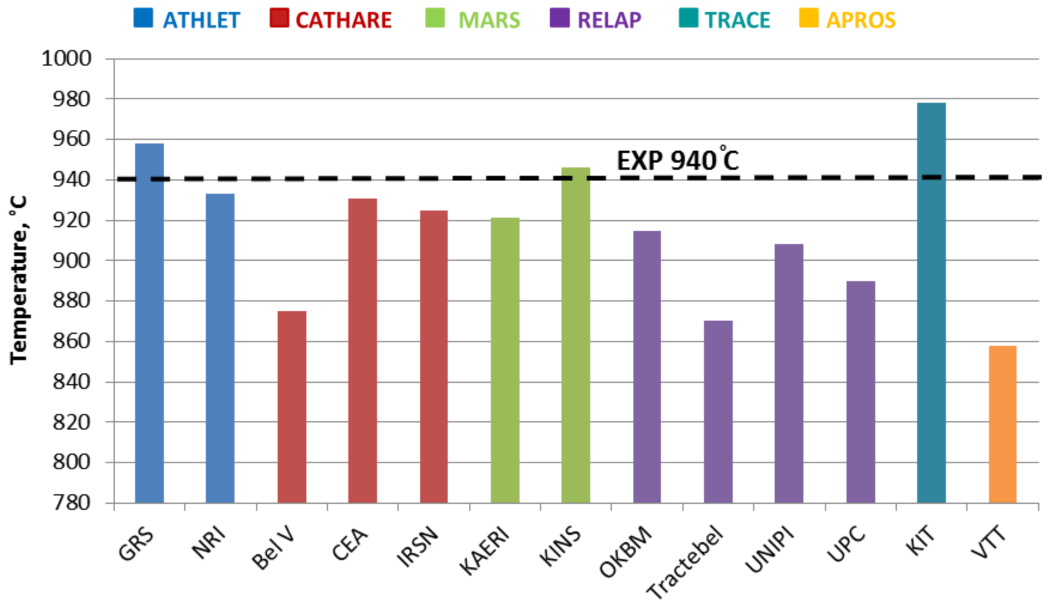


Figure 3.2.6 Bundle quench time comparison in base case

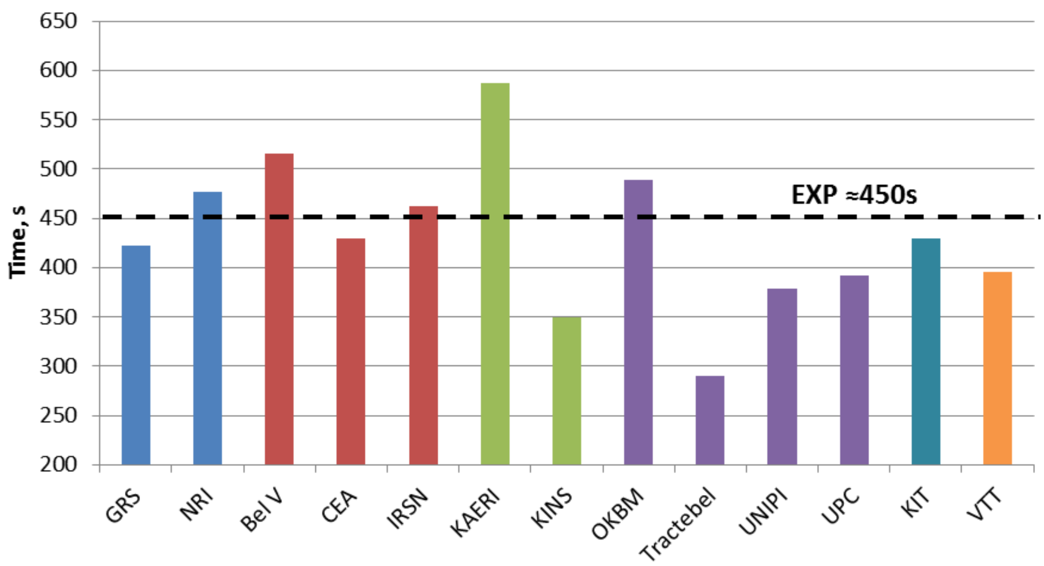


Table 3.2.2 Summary of base case calculations

Participant	Code	PCT, °C	Bundle quenched, s
GRS	ATHLET 2.2B	958	422
NRI	ATHLET 2.1A	933	477
Bel V	CATHARE 2 V2.5	877	516
CEA	CATHARE 2 V2.5	931	429
IRSN	CATHARE 2 V2.5	925	462
KAERI	MARS-KS1.3(COBRA-TF Module)	921	587
KINS	MARS-KS-003	946	350
OKBM	RELAP/SCDAPSIM Mod3.4	915	489
TRACTEBEL	RELAP5 Mod3.3	870	290
UNIPI	RELAP5 Mod3.3 patch3	908	378
UPC	RELAP5 Mod3.3 patch4	890	392
KIT	TRACE Version 5 patch 3	878	430
VTT	APROS 5.11.02	858	396

The base case calculations show spread in predicted cladding temperature and QF propagation, with respect to measured data. All participants except one predict a too fast QF progression. There are some discrepancies in the modelled initial conditions. Some participants simulated the heat up phase, while others initialised the models in conditions of beginning of the transient. The time trends calculated for cladding temperatures by most of participants show oscillatory behaviour (probably having a numerical origin).

Most of participants obtained satisfactory values for peak cladding temperature and bundle rewet time. RELAP and APROS users generally under predicted the PCT, while the sole TRACE user obtained the maximum over prediction. The predicted bundle quench times show a significant spread of ±30% with respect to the measured value.

Anyway, the calculated results reproduce qualitatively the experimental trends, so that they were considered as a suitable basis for performing the sensitivity analysis.

After the base case calculation, each participant compiled an initial list of influential IP. An example list was provided in the specifications report [13]. Each participant considered about 20 parameters, except VTT and KIT, who initially considered 40 and 56 parameters respectively (Figure 3.2.7). In total 72 various IP were taken into account, categorised as IBP, IGP and ICP (Table 3.2.3).

Figure 3.2.7 Initially considered input parameters

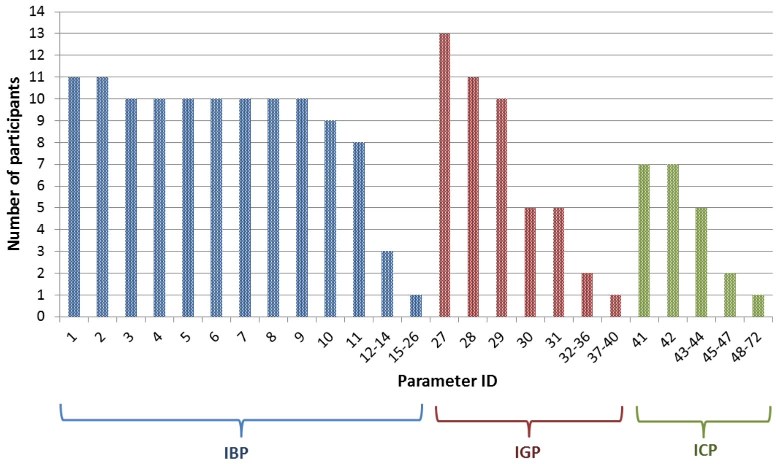


Table 3.2.3: Input parameters initially considered by majority

ID	Parameter
	Input Basic Parameters
1	Inlet liquid temperature
2	Power/power density
3	Pressure
4	Inlet liquid mass flow/flux/velocity
5	Thermal conductivity of heater
6	Heat capacity of heater
7	Thermal conductivity of insulation
8	Heat capacity of insulation
9	Spacer Form loss coefficients
10	Initial wall temperatures
11	Hydraulic diameter
12-26	...
	Input Global Parameters
27	Wall heat transfer
28	Interfacial friction
29	Interphase heat transfer
30	Wall friction
31	Heat transfer (enhancement) at the quench front
32-40	...
	Input Coefficient Parameters
41	Droplet diameter
42	Droplet critical Weber number
43-72	…

The next step was to perform the sensitivity studies and to select the most influential IP. According to the criteria used in the selection, the participants are classified in three groups (Table 3.2.4):

- Those who applied the set of criteria proposed in specifications.
- Those who applied the set of criteria proposed in specifications, but modified the quantitative thresholds.
- Those who applied their own criteria

These different criteria could be one of the causes for the variety of IP ranges identified by the participants, aside from the differences in the physical models contained in the different codes. This variety of ranges makes difficult a meaningful comparison of Phase II results.

Table 3.2.4: Adopted criteria for selection of influential IP

Participant	Code	Criteria
GRS	ATHLET 2.2B	own
NRI	ATHLET 2.1A	own
Bel V	CATHARE 2 V2.5	as in Spec
CEA	CATHARE 2 V2.5	modified Spec
IRSN	CATHARE 2 V2.5	as in Spec
KAERI	MARS-KS1.3 (COBRA-TF Module)	modified Spec
KINS	MARS-KS-003	modified Spec
OKBM	RELAP/SCDAPSIM Mod3.4	modified Spec
TRACTEBEL	RELAP5 Mod3.3	own
UNIPI	RELAP5 Mod3.3 patch3	as in Spec
UPC	RELAP5 Mod3.3 patch4	as in Spec
KIT	TRACE Version 5 patch 3	own
VTT	APROS 5.11.02	as in Spec

Figures 3.2.8 and 3.2.9 show the number of parameters identified as influential

Figure 3.2.8 Selection of influential IP by participants

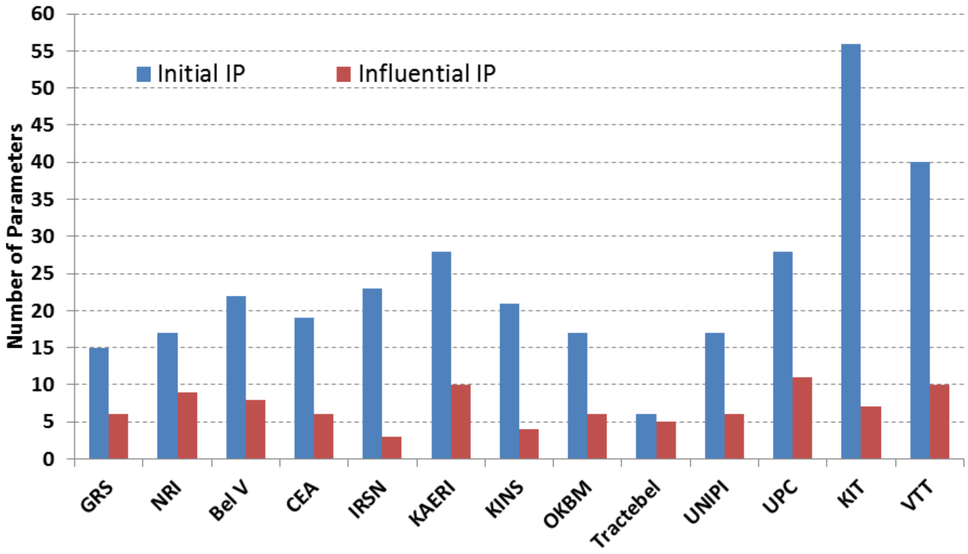
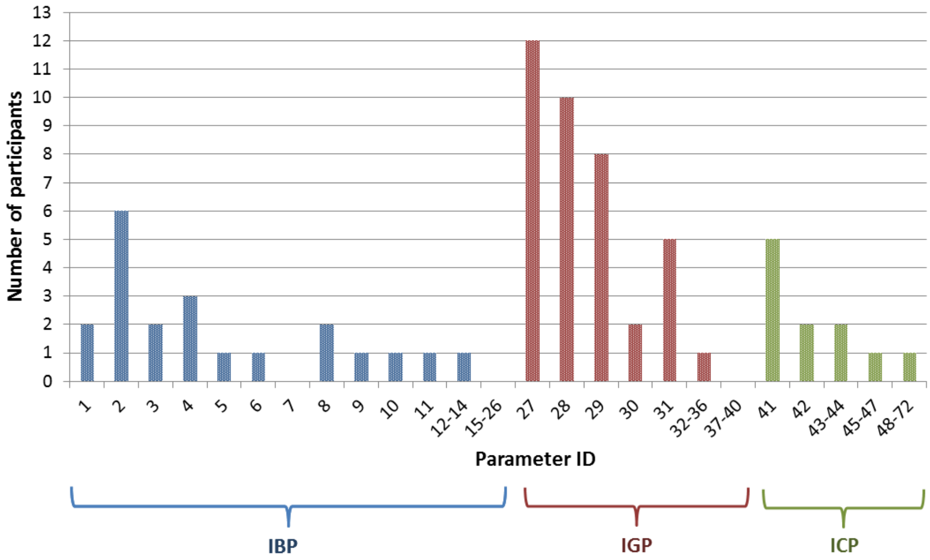


Figure 3.2.9 Selected influential input parameters



Out of the total set of 72 IP, initially considered by all the participants, only 6 were identified as influential by at least 5 participants:

- Bundle power (density).
- Wall heat transfer coefficient: some participants separated the HTC for liquid and for vapour. Some codes showed very low sensitivity of the time of rewet with respect to this parameter.
- Interfacial friction coefficient: some participants distinguished between the friction coefficient for bubbles and droplets, and for dispersed vapour. This parameter influences the cladding temperature in an indirect way, through the void fraction in front of the fuel rods.
- Interfacial HTC.

- Droplet diameter.
- Heat transfer enhancement at the QF: the actual parameters identified are code specific, and may have different influence on calculation results.
- They are highlighted in Table 3.2.5. The applied ranges of variation are shown in Figures 3.2.10 to 3.2.14

Figure 3.2.10 Power variation range (multiplier)

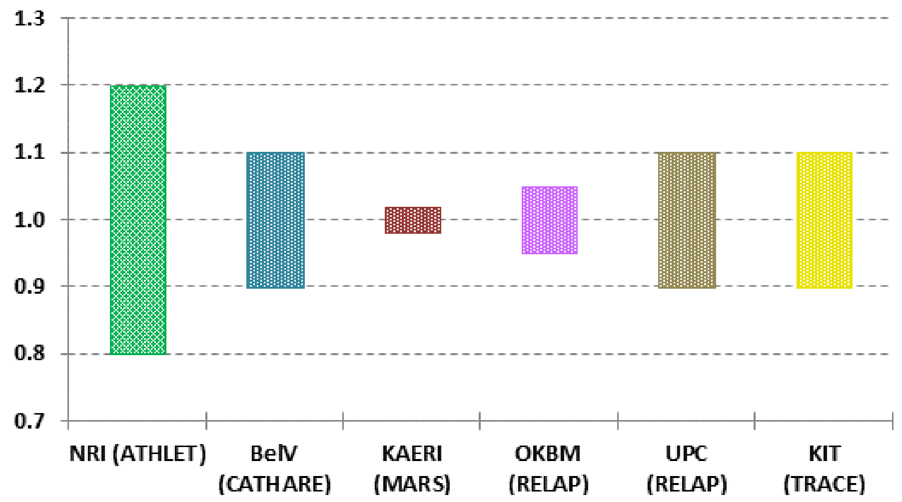


Figure 3.2.11 Wall HTC variation range (multiplier)

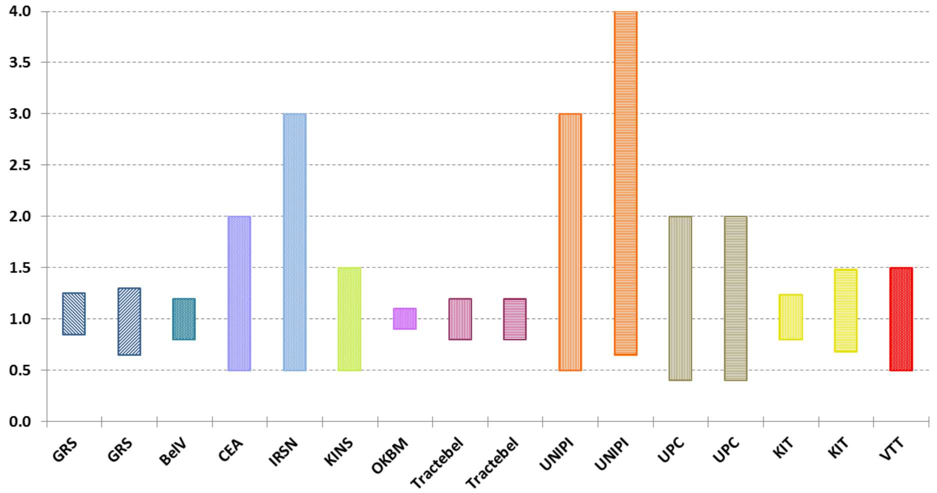


Figure 3.2.12 Interfacial friction coefficient variation range (multiplier)

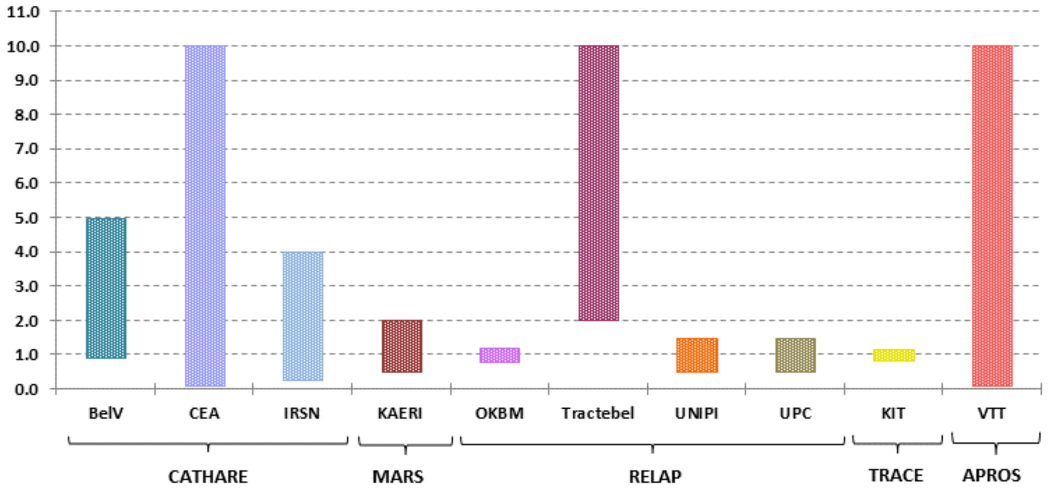


Figure 3.2.13 Interfacial HTC variation range (multiplier)

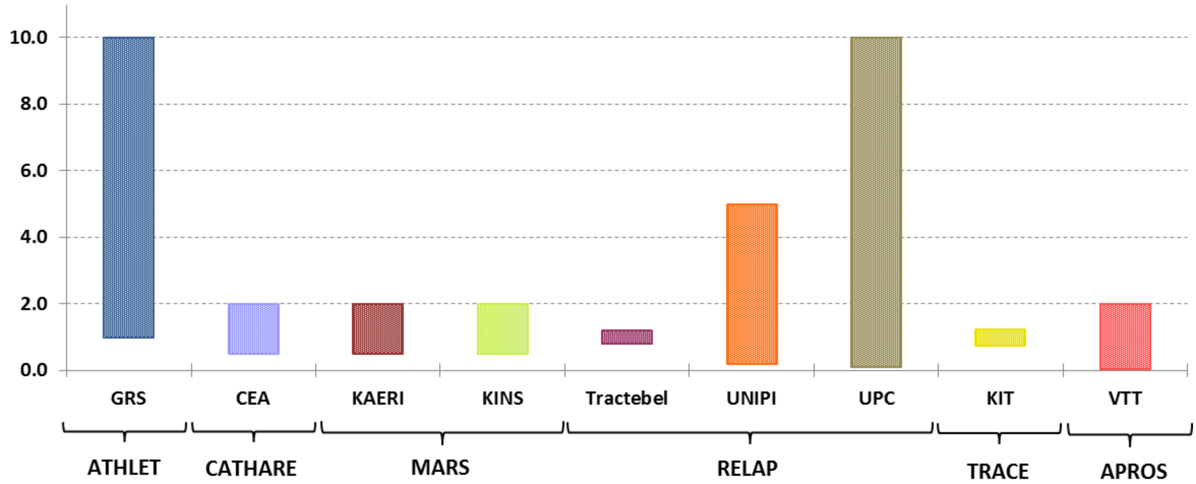
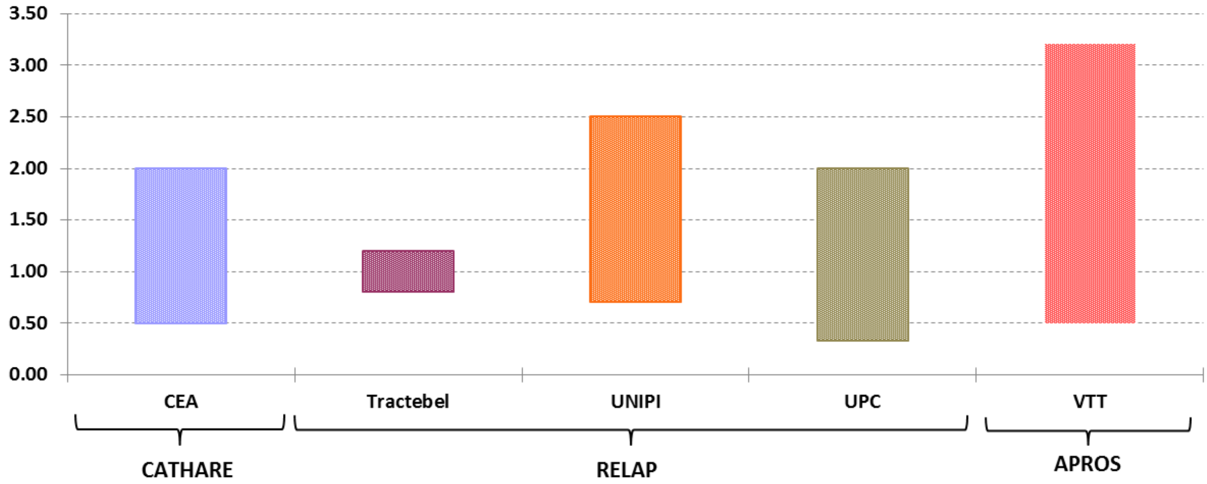


Figure 3.2.14 Droplet diameter variation range (multiplier)



Finally, each participant selected the set of parameters to be quantified at Phase III, as shown in Table 3.2.5.

Table 3.2.5: Parameters identified as influential by majority

ID	Parameter
	Input Basic Parameters
1	Inlet liquid temperature
2	**Power/power density**
3	Pressure
4	Inlet liquid mass flow/flux/velocity
5	Thermal conductivity of heater
6	Heat capacity of heater
7	Thermal conductivity of insulation
8	Heat capacity of insulation
9	Spacer Form loss coefficients
10	Initial wall temperatures
11	Hydraulic diameter
12-26	...
	Input Global Parameters
27	**Wall heat transfer**
28	**Interfacial friction**
29	**Interphase heat transfer**
30	Wall friction
31	**Heat transfer (enhancement) at the quench front**
30-40	...
	Input Coefficient Parameters
41	**Droplet diameter**
42	Droplet critical Weber number
43-73	…

Some participants decided to exclude those parameters whose uncertainty is obtained from experimental data and provided by the co-ordinators (e.g. bundle power). Some participants discarded identified influential parameters (e.g. droplet diameter) because they are part of the correlation for an already considered IGP (existing relation between “coefficient parameter” and “global parameter”). Figures 3.2.15 and 3.2.16 show the statistics of the selected IP.

Figure 3.2.15: IP to be used in Phase III by participants

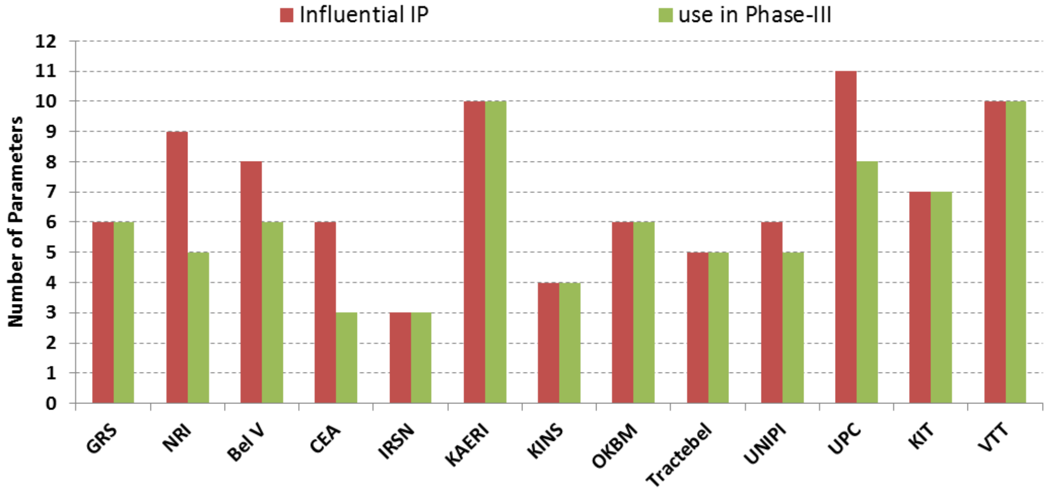
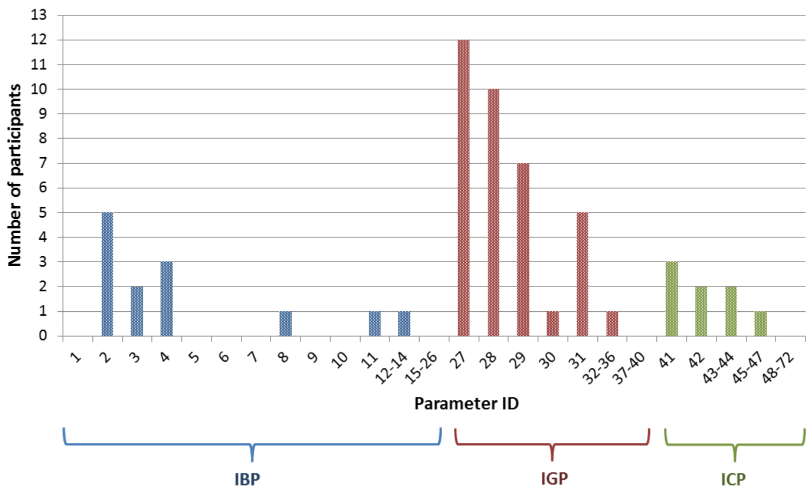


Figure 3.2.16: Input parameters to be used in Phase III



The sensitivity of responses to changes in different parameters, especially in the extremes of the variation range, depends on the type of input parameter and the code used. For some parameters, the behaviour can be very different. The reason may be that the dependence is complex, because many physical models are involved.

The behaviour of the variation of the responses at the extremes of IP range greatly depends on the type of input parameter and on the code used. Mainly, two different types of behaviour can be characterised:

- Qualitative (but not quantitative) agreement among different codes: power, wall HTC and interphase HTC.
- Contrary behaviour (sign change) for different codes, and even for different selected models within the same code: interphase friction coefficient. This shows that the effect of the parameters on cladding temperature is quite complex, involving a lot of physical models.

3.3 Phase III: Quantification of model uncertainties on the basis of FEBA/SEFLEX reflood experiments.

The goal of Phase III is the quantification of model uncertainties on the basis of so-called "intermediate" experiments [10, 14]. Such procedure is addressed in particular to uncertainties quantification of models related to phenomena for which separate effect tests do not exist, as it is the case for heat transfer enhancement at the QF during the reflood phase of a large break LOCA.

The co-ordinator of Phase III has been GRS (Germany). Altogether 14 organisations have participated in Phase III supplying 16 contributions. The participants used different thermal-hydraulic system codes. As it can be seen in Table of Participants in PREMIUM Benchmark two codes have been used by several participants. RELAP5 MOD3.3 has been used in 5 contributions, and CATHARE2 V2.5 has been used by 3 participants.

The information necessary for PREMIUM has been extracted from the KfK experimental reports by the Phase III co-ordinator, summarised and supplied to the participants: the measured data are available only in the form of plotted curves, so an important part of the task was the digitalisation of such curves.

As can be seen in the Table 3.3.1 the majority of participants applied one of the methods offered in the frame of PREMIUM. Eight participants applied CIRCÉ method and three participants applied FFTBM. One participant (VTT Finland) applied a combination of these two methods. CIRCÉ method has been applied by VTT for determination of biases and FFTBM for the determination of the uncertainty ranges. The remaining four participants applied each one their own method.

The other methods used in this phase are MCDA, DIPE, IUQ and GRS method, described in Section 2.2 and in [8].

Two organisations (KAERI and OKBM) submitted two contributions each. OKBM used the quantification method CIRCÉ with two different codes RELAP and KORSAR. KAERI used for code COBRA two different quantification methods CIRCÉ and MCDA

Phase III has been divided into three steps:

Step 1: <u>Definition of the list of parameters to be considered, and initial quantification</u>. The basis for the definition of parameters to be considered in the Phase III evaluation/quantification step were the results obtained by each participant within the Phase II. The results of the Phase II summarised in [9] enabled comparison with results obtained by other participants, in particular those using the same thermal-hydraulic code. This gave an opportunity for critical review of the own list of selected uncertainties and preliminary quantification ranges on the basis of which the selected parameters have been found as influential. A critical review of the results obtained within the Phase II should lead to improvement of the list of selected parameters which were going to be quantified within the Phase III by each participant

Step 2: <u>Uncertainty quantification for the selected model parameters.</u> Each participant was responsible for choosing an adequate method of model uncertainties quantification. The results of the Phase II sensitivity analyses and experience concerning test run simulation indicate that the experiment and test facility related uncertain IP are secondary comparing to the physical model parameters. Nevertheless, consideration of influence of uncertain parameters different to model parameters in the quantification process was left to individual participant decision.

Step 3: <u>Preliminary check of quantified uncertainty ranges.</u> Uncertainty and sensitivity analysis of the Test run 216 from the Series I of FEBA has been performed. It is the FEBA test most similar to the PERICLES experiment, from the standpoint of boundary conditions. The selection of various calculated quantities for comparison with measured data, and not only cladding temperatures, should give an opportunity to find out if compensating errors take place by the test run simulation. Participants could perform their own uncertainty analyses of farther test runs, should it be found necessary.

3.3.1 Experimental data used for uncertainties quantification

For performing a successful check of model UQ method, it is of importance that other potentially important uncertainties, like uncertainties of spatial modelling can be eliminated. This is the case for relatively simple test facilities, where 1-D approximation is suitable and no particular problems should arise by discretisation and development of an input data set.

In this context FEBA/SEFLEX reflooding experiments appear to be suitable for the benchmark application, for three main reasons:

- Data from some reflooding tests are available.
- Geometry of the test section is quite simple and average experienced user should not have any problem with its correct simulation.
- In reflooding a moderate number of physical phenomena are involved. So, one can expect to identify the reason of differences comparing the results of different calculations (participants) with experimental data.

For the quantification of reflood model uncertainties, all participants have used the tests from Series I of FEBA experiments (Table 3.3.1). Participants were allowed using their own reflood experiments (provided that they are sufficiently validated and made available to other participants). The measured data of FEBA Series II and SEFLEX Series I and II could have been additionally used in the task of quantification. Despite this flexibility neither other reflooding experiment nor tests from Series II were used by the participants in this phase.

Table 3.3.1: Methods and chosen experimental tests and responses in Phase III

Particip.	Code	Method	Tests used	Responses used
CVRez	RELAP	CIRCÉ	FEBA tests: 223, 216, 220, 218, 222	clad temp., quench time, Δp
UPC	RELAP	CIRCÉ	all 6 FEBA tests	clad temp., water carried over, quench time
OKBM-1	RELAP	CIRCÉ	FEBA tests: 223, 216, 220, 218, 214	clad temp., quench time
OKBM-2	KORSAR	CIRCÉ	all 6 FEBA tests	clad temp., quench time
CEA	CATHARE	CIRCÉ	all 6 FEBA tests	clad temp., quench front elevation
BelV	CATHARE	CIRCÉ	all 6 FEBA tests	clad temp., quench time
KINS	MARS-KS	CIRCÉ	all 6 FEBA tests	clad temperature
KAERI-1	COBRA	CIRCÉ	FEBA tests: 214, 216, 218, 223	clad temperature
KAERI-2	COBRA	Data assimilation	FEBA tests: 214, 216, 218, 223	clad temperature
VTT	APROS	FFTM/CIRCÉ	all 6 FEBA tests	clad temp., housing temp., quench time
UniPisa	RELAP	FFTBM	FEBA test 216	clad temp., quench time
SJTU	RELAP	FFTBM	FEBA test 216	clad temp., quench time, Δp
KIT	TRACE	FFTBM	FEBA test 216	clad temp., quench time
IRSN	CATHARE	DIPE	all 6 FEBA tests	clad temp., quench front propagation
GRS	ATHLET	Inverted Unc.	all 6 FEBA tests	Δp, water carried over
			FEBA tests: 216, 223	clad temperature
			Separate effect tests	clad temperature
TRACTEBEL	RELAP	Inverted Unc.	all 6 FEBA tests	clad temperature

A majority of participants used the 6 tests of Series I. Only 3 used solely test 216. They were FFTBM users, and the reason was that the supplied software enabled the application of the method to one test run. VTT made an improvement to the software to apply it to several tests. Thus VTT could use the 6 tests of the series.

Regarding the type of measured responses used for the quantification, all participants considered cladding temperatures, and almost all of them used quench front progression (quench time or QF elevation) too. Some participants considered also pressure drop measurements, and only few used the measured data of water carried over. Only one (VTT) included the measured housing temperature.

Only one participant (GRS) made use of data from separate effects tests, in order to quantify a part of the model uncertainties (two correlations related to wall heat transfer at dry out condition). The goal of this procedure was to benefit of the large database for wall heat transfer correlations in order to accomplish a better quantification of the uncertainty.

3.3.2 Input parameters quantified within Phase III

The input parameters that have been quantified during Phase III are listed in Table 3.3.2. The number of parameters quantified by each participant varied between 2 and 6. They were all model parameters, with an exception. One participant (KIT), in addition to 6 model parameters, made the quantification of uncertainty of the rod bundle power, which in fact is a boundary condition of the experiment, having its own experimental uncertainty.

Table 3.3.2: Uncertain input parameters considered by participants and their probability distributions.

Participant	Code	Number of param.	Wall HTC	Interfacial HTC	Momentum eqn – closure rel.	Method
CVRez	RELAP	2	1 [log-norm]	-	1 [log-norm]	CIRCÉ
UPC	RELAP	3	1 [log-norm]	1 [log-norm]	1 [log-norm]	CIRCÉ
OKBM-1	RELAP	3	2 [log-norm]	1(droplet) [log-norm]	-	CIRCÉ
OKBM-2	KORSAR	2	1 [log-norm]	-	1 [log-norm]	CIRCÉ
CEA	CATHARE	2 + 1(bias)	1 [log-norm] + 1(quen.)	-	1(mist flow) [log-norm]	CIRCÉ
BelV	CATHARE	3	1 + 1(quen.) [log-norm]	-	1 [log-norm]	CIRCÉ
KINS	MARS-KS	2	1 [norm]	1 [norm]	-	CIRCÉ
KAERI-1	COBRA	4	1 + 1(grid) [log-norm]	1(mist flow) [log-norm]	1(mist flow) [log-norm]	CIRCÉ
KAERI-2	COBRA	4	1 + 1(grid) [log-norm]	1(mist flow) [log-norm]	1(mist flow) [log-norm]	Data assimilation
VTT	APROS	6	2 + 1(quen.) [norm, log-norm + log-norm]	1 [log-norm]	1 + 1(wall friction) [log-norm + norm]	FFTBM/CIRCÉ
UniPisa	RELAP	5	2 [uniform]	1 + 1(droplet) [uniform]	1 + 1(droplet) [uniform]	FFTBM
SJTU	RELAP	4	2 [uniform]	1(droplet) [uniform]	1 + 1(droplet) [uniform]	FFTBM
KIT	TRACE	6 + rod power	3 + 1(grid) [uniform]	1 [uniform]	1 [uniform]	FFTBM
IRSN	CATHARE	3	1 + 1(quen.) [histogram]	-	1(mist flow) [histogram]	DIPE
GRS	ATHLET	6	2 + 1(quen.) [uniform]	1 [uniform]	1 + 1(entrainment) [uniform]	Inv. Uncertainty
TRACTEBEL	RELAP	4	1 [uniform]	2 [uniform]	1 [uniform]	Inv. Uncertainty

The number of parameters quantified by RELAP users varies from 2 to 5. On the other hand all users of CATHARE codes considered three parameters. Regarding the quantification method the majority of CIRCÉ users considered two or three parameters. Only KAERI considered 4 parameters. Users of other methods considered generally larger number of parameters.

Taking into account the physical models all participants considered uncertainties related to the wall heat transfer. As it could be expected all participants considered uncertainty of heat transfer correlation at dry out conditions. Almost all participants considered also uncertainties of momentum equation closure relations. The interfacial heat transfer uncertainty has been considered by 11 participants. Usually both the uncertainty of interfacial momentum transfer and interfacial heat transfer were considered. Some participants considered only uncertainty of momentum equation closure relation. Only two participants considered uncertainty of interfacial heat transfer not taking into account interfacial momentum transfer uncertainty.

The following heat transfer related parameters have been considered by participants:

- Film boiling heat transfer coefficient (HTC) – 10 participants.
    - Film boiling HTC total – 2 participants;
    - Film boiling HTC gas/liquid phase separately – 8 participants
- General HTC for dry out condition (above the QF) – 2 participants;
- Global HTC for all heat transfer regimes (applied in the whole bundle) – 2 participants;
- HTC for steam convection – 3 participants;
- Minimum film boiling temperature – 2 participants;
- Heat transfer enhancement at QF – 5 participants;
- Heat transfer enhancement at grid spacers – 3 participants.

The interfacial heat transfer parameters considered by participants are the following:

- Global interphase heat transfer – 4 participants
- Interphase heat transfer for mist flow – 6 participants
- Droplet diameter (key parameter used in evaporation model) – 3 RELAP users

The considered momentum equation closure relation uncertainties are:

- Global interfacial friction – 8 participants;
- Interfacial friction for mist flow – 5 participants;
- Droplet diameter (key parameter used in interfacial drag model) – 2 RELAP users;
- Entrained liquid fraction – 1 participant;
- Wall friction of liquid phase - 1 participant

Although the variety of the considered parameters is limited, even the users of the same code considered frequently different parameters. The code RELAP was used by 6 participants. Only UNIPI and SJTU considered the same 4 parameters out of total number of 5 and 4 parameters considered by them respectively. UNIPI and TRACTEBEL considered the same 3 parameters out of total number 5 and 4 parameters. In the case of CATHARE2 CEA and IRSN considered the same 3 parameters. The third user of CATHARE2 (BelV) considered only 1 parameter in common with CEA and IRSN. On the other hand GRS and VTT using different codes considered three common parameters (parameters related to the same or equivalent models) out of total of six parameters used by each of them.

The type of probability density functions (PDF) applied by the participants is clearly method dependent. As it can be seen in the Table 3.3.3 all the users of the CIRCÉ method and similar data assimilation method applied normal or log-normal probability distribution functions. All other participants applied uniform distributions. The IRSN applied histogram, which is a uniform distribution with 1 as its median (if 1 is not the median, there is a recalibration of the code). The reasons for the selection of the type of distribution are the assumptions making the basis of the methods. In CIRCE formulation normal or log-normal distribution probability distribution functions of the quantified parameters have been assumed. In other methods, the uniform distribution was applied. Even in the IRSN method, where an empirical distribution could be obtained in the course of quantification the histogram distribution has been assumed.

The normal and log-normal distributions determined by CIRCE are not truncated. The truncation of distributions was left to the users' decision. Some users of CIRCE performed this step and truncated the obtained distributions, while others applied the non-truncated distributions. The truncation was performed at 2.5% and 97.5% percentiles of the distribution. The truncation values of the distributions are presented in the tables as "Min" and "Max" values of the variation range. Also those users of CIRCE who did not perform the truncation supplied for comparison the values of 2.5% and 97.5% percentiles of their distribution (also presented in the tables).

Application of non-truncated distributions can lead to generation of extreme values of parameters by sampling. Parameters combinations with such extreme values are extremely improbable in random sampling, but they could result in failure of the code run or production of non-physical code results. Application of truncated distributions prevents generally generation of such extreme elements of the sample. Also by application of truncated distributions there are still differences between normal and uniform distributions. In random sample generation for normal distributions, elements near the mean value of the distribution are preferred. For uniform distributions, the probability of element generation is equal for the whole range of variation. Using of higher order of Wilks' formula can reduce to some extent the differences due to application of different probability distribution functions, particularly those due to application of non-truncated distributions.

3.3.3 Consideration of uncertainties other than physical model uncertainties

PREMIUM is focused on the estimation of uncertainties of model parameters. The uncertainty of other type of IP (initial and boundary conditions, material properties…) should be estimated from other sources (e.g. measurement devices). In the specification of Phase III, information about experimental uncertainties and thermal properties uncertainties in FEBA experiment has been supplied.

The information in Phase III specification concerning the thermal properties of materials used in FEBA test facility was not obtained from experimenters but as a result of Phase III co-ordinator survey of literature and estimations. Therefore, they could be a subject of corrections and modifications by participants. However, taking into account experience from parametric sensitivity study performed within the Phase II of PREMIUM, it seems that in the case of FEBA experiments the uncertainties of material data are considerably less influential than the physical model uncertainties.

Unfortunately, in the description of the FEBA experiment [6, 7, 8, 9] there is only very little information about uncertainties of the experimental data. The cladding and housing temperatures can be expected to be of high accuracy. A typical accuracy of chromel-alumel thermocouple is about ± (0.4% – 0.5%) *Temp. [°C]. For the measured temperature range it is about ±5 °C. In addition it has to be taken into account that the thermocouples measure not exactly the temperature of the cladding surface. The accuracy of the pressure drop measurement was not reported. However, a typical error of pressure transducers is about 1% of measured pressure range by constant temperature. Since the temperature in the FEBA experiments varies strongly along the test section the error can be much higher. The accuracy of the pressure drop measurements in other test facilities with similar bundle configuration were estimated as ±10%. It could be a reasonable estimation also for the FEBA pressure drop measurements.

The uncertainty of the measured mass of water carried over the test section is difficult to estimate. The amount of the water carried over is measured using water collection tank. The mass of water in the tank can be measured quite precisely with estimated accuracy about 1% - 2%. However, the mass collected in the tank may be different compared to the water carried over, particularly shortly after initiation of water carryover phenomenon. At the beginning water carried over the test section can evaporate on the hot surfaces of the upper plenum and some amount can be also accumulated in the upper plenum before water reaches the water collecting tank. A small part of water carried over the test section can be entrained by steam leaving the upper plenum. All these effects lead to underestimation of the measured mass of water carried over. The only possibility for overestimation of the water mass results from inaccuracy of water inventory measurement in the collecting tank. But this seems to be rather small. A reasonable estimation of the measured water mass uncertainty could be the range: (~ +0 kg; -0.5 kg). The measured data of water carried over are available only for the initial part of the transients. The size of the tank was limited to 10 kg, and after filling of the tank no further measurements of the water carried over were possible.

In addition to the measured parameters of the reflooding in the test section the uncertainty of the test boundary and initial conditions could be of importance for quantification of model uncertainties. The inlet velocity, inlet temperature, system pressure and bundle power are constants or slow transients and the measurements should be quite accurate. The measurement error seems to be comparable with digitalisation error of obtaining numerical values on the basis of plotted curve.

The digitalisation error of the bundle power curve is about ±1.5kW – ±2.0 kW. It is about 1.5 – 2.0 % of power, the accuracy in the range of electrical power measurement error.

The inlet water temperature digitalisation error is about ±2°C. It is also in the range of water temperature measurement accuracy. However, during initial period of the test runs cold water filled inlet plenum where walls had much higher temperature. This could result in nonhomogeneous temperature distribution in the inlet plenum and significantly lower accuracy of the estimated inlet temperature. This is a short term phenomenon. After 10 — 30 seconds the inlet temperature stabilised and the inlet temperature measurements are expected to reach ±2°C accuracy.

The system pressure measurements show pressure variation around the value defined as test run parameter. The accuracy of the system pressure measurement is high. But the deviation of the measured system pressure from the defined system pressure can reach ±0.2 bar. It is mainly due to non-stationary character of reflooding and a result of pressure regulators functioning. Considering the measured pressure curve instead of a constant value could improve test run simulation. But in the GRS sensitivity calculation it has been found that even variation of system pressure by 0.2 bar has a small effect on the results in comparison with influence of model uncertainties.

The inlet velocity of single-phase water (flooding velocity in experiment description) can be measured with high accuracy of about ±1% – ±2%. In some test runs during the initial phase of the experiment deviations from the declared test parameters occurs. If the deviation is large, for instance like in the test no. 233 (FEBA Series II), the measured parameters instead of nominal values should be considered by boundary condition modelling.

Since the presented experiment related uncertainties are not obtained from the experimenters, but rather as a result of Phase III co-ordinator (GRS) estimation, an attempt to compare the FEBA experiment uncertainties with those of PERICLES experiment [10] has been undertaken. The estimated uncertainties of FEBA measurements are generally similar or larger than those of PERICLES. Larger are mainly uncertainties of measurements during the initial period of the test runs when, after the stationary phase of test section heating, the flooding is initiated.

Since the experimental uncertainties are only estimations and not uncertainties defined by the experimenters, the participants could correct them and consider them in the quantification process, if they find it necessary.

3.3.4 Results of model uncertainties quantification

The uncertainty ranges found by participants are compared for each type of physical models. Analysing the uncertainty ranges obtained by the participants, it should be kept in mind that, in many cases, the quantified uncertainties are related to different codes and models. Theoretically, direct quantitative comparison should be limited to the users of the same code. One can also remind that for CIRCÉ users, the uncertainties quantified should be regarded as a whole set of values i.e. the value of calibrated parameter and range of its variation is valid only together with the values obtained for other parameter considered in the quantification process, and not alone.

In the tables 3.3.3 to 3.3.6 the uncertainty ranges of model specific uncertainties are compared. In the Table 3.3.3 the uncertainties of wall heat transfer at dry out conditions arranged according to the used method are listed. Majority of participants quantified HTC correlations according to the heat transfer regimes. Only two participants used global HTC as an uncertain input parameter. Two other used general HTC for dry out condition. One participant (KINS) used, instead of heat transfer coefficient, the criterion for minimum film boiling temperature (MFBT) as uncertain parameter. Some participants considered uncertainties of more than one wall heat transfer correlations. Some participants considered code specific parameters like wall heat transfer enhancement at the QF and wall heat transfer enhancement due to the grid spacers. The coefficients related to heat transfer enhancement due to the grid spacers are used by KAERI to express the uncertainty of wall heat transfer in the range of steam convection. During the Phase II analyses KAERI has found multiplication factor for the heat transfer enhancement at grid spacers as more influential than the one for steam convection. With the aim of not increasing too much the number of parameters to be quantified, KAERI has selected only the parameter related to heat transfer enhancement for quantification. Taking it into account, the HTC uncertainty due to grid spacers as applied by KAERI has been considered as comparable with HTC parameters related to steam convection and included in the tables for comparison. KIT also used parameter related to the heat transfer enhancement at grid spacers and considered also the uncertainty of HTC from wall to vapour. In such configuration the grid spacer effect is separated from the HTC for steam convection and not really comparable with uncertainties of wall heat transfer related parameters. Indeed, the uncertainty range obtained by KIT for grid spacer heat transfer enhancement is (0.0 – 1.1), clearly out of the range of typical HTC uncertainties. For this reason the parameter related to the heat transfer enhancement at grid spacers as used by KIT was not included in the comparison.

The parameters related to wall heat transfer enhancement at the QF are very model specific and they are not compared in the tables. Even separate comparison of uncertainties of HTC for the QF is not meaningful, since the parameters are related to different effects. In the case of GRS code ATHLET the heat transfer enhancement at the QF is considered only by QF progression but not by wall temperature determination.

The criterion for MFBT used by KINS was included in all comparisons, since it is the only parameter for wall heat transfer considered by KINS. The minimum and the maximum of the ranges of the wall HTC are presented in the tables in the following convention: {HTC – dry out or film wall to gas correlations} / {HTC – global or film wall to liquid}. In parenthesis are in tables values related to wall dry/wet transition criterion.

The uncertainty ranges of wall heat transfer at dry out conditions are presented graphically in the Figures 3.3.1 and 3.3.2. The green points in the bars represent reference values. For the users who performed model calibration the green points represent the calibrated values. In the case of participants who performed model calibration the reference value was usually 1.0, so it was not necessary to mark it in the figures as additional points.

In the Table 3.3.4 the uncertain parameters related to the interfacial heat transfer model are listed. Majority of participants selected as influential model uncertainty of interfacial heat transfer at the mist (dispersed droplet) flow. This is a typical flow pattern above the QF. Some participants considered as uncertain parameter multiplication factor for global interfacial heat transfer. Parameter used by GRS is not an interfacial heat transfer multiplier but a number of droplets in the evaporation model. However, it can be recalculated on the basis of evaporation correlations used in the code ATHLET as a value equivalent with multipliers used by other participants. The recalculated value of this parameter (given in parenthesis in the Table 3.3.5) is used by graphical comparison of interfacial heat transfer related parameters presented in the Figures 3.3.3 and 3.3.7.

The model uncertainties related to momentum equation closure relations are listed in the Table 3.3.5. The list of considered uncertain parameters is clearly dominated by multiplication factor for global interfacial friction. This parameter was considered by 8 participants. Five participants considered interfacial friction for mist flow. In addition to the interfacial friction some participants considered an additional parameter related to the phase relative velocity. These additional parameters are the droplet diameter in the case of two RELAP users, entrainment rate and wall friction. GRS used as parameter the multiplication factor for phase relative velocity. The uncertainty of phase relative velocity cannot be directly compared with uncertainty of interfacial friction factor. However, on the basis of relative velocity uncertainty an equivalent uncertainty of interfacial friction for the experiment condition could be roughly estimated. This estimated value is compared with uncertainties quantified by other participants in the Figure 3.3.8.

Since almost all considered parameters are multiplication factors the best way to illustrate their ranges is to produce quotient Max (of the range) divided by the Min (of the range). This form of the range presentation is given in the Table 3.3.6 and in the Figures 3.3.5 – 3.3.8.

The uncertainty ranges quantified by the participants differ considerably, even the uncertainties of the same codes. The best basis for comparison gives the uncertainties of HTCs at dry out condition. This parameter has been considered by practically all participants. The comparison of the uncertainty ranges shows that they are quite similar for CATHARE code users and very different for RELAP users. More interesting appears the comparison according to the quantification method. The uncertainty ranges obtained by CIRCÉ users tend to small and moderate ranges. Uncertainty ranges obtained with FFTBM tend to large values. The methods of TRACTEBEL and GRS seem to lead to moderate uncertainty ranges. Very large uncertainty ranges have been obtained by UNIPI and SJTU for film boiling and even larger for forced convection to gas by VTT. Such a large uncertainty range obtained by VTT for practically single-phase heat transfer correlation is astonishing. It is not clear whether this is due to extension of the application of FFTBM to more than one experimental run. The possibility of consideration of more than one test is an obvious advantage. This enables to enlarge the data basis and take into account, for instance, tests considering different thermal-hydraulic conditions. It would enable to consider experiments performed in different test facilities. For quantification of model uncertainties it is of importance to be able to consider many tests. Usually consideration of more experiments by uncertainties quantification leads to extension of the variation range. But it is a typical trend that a wider range of application is usually related with larger uncertainty of simulation. However, the cause of the very large uncertainty obtained by VTT remains unclear (extension of FFTBM, consideration of many tests). As it could be seen in the comparison the FFTBM applied even to only one experiment results sometimes in quantification of large uncertainties. The reason for the large quantified uncertainty may be also the application of combination of CIRCÉ and FFTBM. In the opinion of the VTT the last reason is probably the most important one.

Table 3.3.3: Uncertainty ranges applied for heat transfer at dry out conditions

Particip.	Code	Method	Wall HTC	Min	Max	Ref./calibr. Value
CVRez	RELAP	CIRCÉ	1 (all regimes)	0.433	0.954	0.643 (calibr.)
UPC	RELAP	CIRCÉ	1 (film –liquid)	0.75	1.37	1.01 (calibr.)
OKBM-1	RELAP	CIRCÉ	2 (film-gas/liq)	0.43/1.06	0.97/1.49	0.646/1.272 (calibr.)
OKBM-2	KORSAR	CIRCÉ	1 (film-gas)	0.816	1.34	1.046 (calibr.)
CEA	CATHARE	CIRCÉ	1 (dry out)	0.73	1.44	1.03 (calibr.)
BelV	CATHARE	CIRCÉ	1 (all regimes)	0.692	1.780	1.110 (calibr.)
KINS	MARS-KS	CIRCÉ	1 (MFBT)	(0.493)	(0.891)	0.692 (calibr.)
KAERI-1	COBRA	CIRCÉ	1 (film – liq.)	0.52/0.90	1.02/1.18	0.73/1.03 (calibr.)
KAERI-2	COBRA	Data assimilation	1 (film - liq.)	0.60/0.80	1.26/1.26	0.86/1.01 (calibr.)
VTT	APROS	FFTM/CIRCÉ	2 (steam+ MFBT)	0.05/(0.65)	1.7/(1.7)	0.7/1.2 (calibr.)
UniPisa	RELAP	FFTBM	2 (film-gas/liq)	0.35/0.50	2.8/1.3	1.0/1.0
SJTU	RELAP	FFTBM	2 (film-gas/liq)	0.15/0.77	1.92/1.44	1.0/1.0
KIT	TRACE	FFTBM	2 (steam+film)	0.41/0.36	1.4/1.4	1.0/1.0
IRSN	CATHARE	DIPE	1 (dry out)	0.58	1.56	1.0
GRS	ATHLET	Inverted Unc.	2 (steam+film)	0.85/0.65	1.25/1.3	1.0/1.0
TRACTEBEL	RELAP	IUQ	1 (film-gas)	0.7	1.3	1.0

Table 3.3.4: Uncertainty ranges applied for interfacial heat transfer model

Particip.	Code	Method	IHT correl.	Min	Max	Ref./calibr. Value
CVRez	RELAP	CIRCÉ	-			
UPC	RELAP	CIRCÉ	1 (global)	0.29	2.07	0.77 (calibr.)
OKBM-1	RELAP	CIRCÉ	1(droplet)	0.617	0.844	0.722 (calibr.)
OKBM-2	KORSAR	CIRCÉ	-			
CEA	CATHARE	CIRCÉ	-			
BelV	CATHARE	CIRCÉ	-			
KINS	MARS-KS	CIRCÉ	1 (mist flow)	0.148	1.967	1.058 (calibr.)
KAERI-1	COBRA	CIRCÉ	1 (mist flow)	0.59	1.97	1.07 (calibr.)
KAERI-2	COBRA	Data assimilation	1 (mist flow)	0.77	1.68	1.13 (calibr.)
VTT	APROS	FFTBM +CIRCÉ	1 (global)	0.05	3.5	0.5 (calibr.)
UniPisa	RELAP	FFTBM	1 (global) +1(droplet)	0.2 (0.7mm)	5.0 (2.5mm)	1.0 (1.5mm)
SJTU	RELAP	FFTBM	1(droplet)	0.90mm	2.35mm	1.5mm
KIT	TRACE	FFTBM	1 (global)	0.572/0.000	1.546/2.652	1.0
IRSN	CATHARE	DIPE	-			
GRS	ATHLET	Own method	1 (mist flow)	$(1.0)10^{9}$	$(\sim3.2)10^{10}$	$(1.0)10^{9}$
TRACTEBEL	RELAP	IUQ	1 (global) +1 (mist flow)	0.4 (0.4)	1.1 (1.1)	1.0 (1.0)

Table 3.3.5: Uncertainty ranges applied for momentum equation constitutive equations

Particip.	Code	Method	Interfacial friction	Min	Max	Ref./calibr. Value
CVRez	RELAP	CIRCÉ	1 (below quench)	0.487	0.906	0.664 (calibr.)
UPC	RELAP	CIRCÉ	1 (global)	0.87	1.37	1.09 (calibr.)
OKBM-1	RELAP	CIRCÉ	-			
OKBM-2	KORSAR	CIRCÉ	1 (global)	0.423	2.08	0.938 (calibr.)
CEA	CATHARE	CIRCÉ	1 (mist flow)	0.73	0.77	0.75 (calibr.)
BelV	CATHARE	CIRCÉ	1 (global)	0.076	0.079	0.077 (calibr.)
KINS	MARS-KS	CIRCÉ	-			
KAERI-1	COBRA	CIRCÉ	1 (mist flow)	0.82	0.85	0.83 (calibr.)
KAERI-2	COBRA	Data assimilation	1 (mist flow)	0.49	1.49	0.84 (calibr.)
VTT	APROS	FFTBM /CIRCÉ	1(mist flow) + 1(wall friction)	1.4	6.0	2.0 (calibr.)
UniPisa	RELAP	FFTBM	1 (global) +1(droplet)	0.86 (0.5mm)	1.6 (1.9mm)	1.0 (1.5mm)
SJTU	RELAP	FFTBM	1 (global) +1(droplet)	0.51 (0.90mm)	1.53 (2.35mm)	1.0 (1.5mm)
KIT	TRACE	FFTBM	1 (global)	0.843	1.261	1.0
IRSN	CATHARE	DIPE	1 (mist flow)	0.1	7.3	1.0
GRS	ATHLET	Own method	1 (global) +1(entrain.)	0.64	1.60	1.0 (rel. velocity)
TRACTEBEL	RELAP	IUQ	1 (global)	0.7	3.4	1.0

Table 3.3.6: Quantified uncertainty ranges (as max/min)

Participant	Code	Number of param	Wall HTC	Interfacial HTC	Momentum eqn – closure rel.	Method
CVRez	RELAP	2	/2.20	-	1.86	CIRCÉ
UPC	RELAP	3	/1.83	7.14	1.57	CIRCÉ
OKBM-1	RELAP	3	2.26/1.41	(1.37 – droplet)	-	CIRCÉ
TRACTEBEL	RELAP	4	1.86/	2.75/2.75	4.86	Inv. Uncertain.
UniPisa	RELAP	5	8.0/2.6	18.0(3.80– droplet)	1.86 (3.80 – droplet)	FFTBM
SJTU	RELAP	4	12.8/1.87	(2.61 – droplet)	3.0 (2.61– droplet)	FFTBM
KIT	TRACE	6 +power	3.37/3.91	2.70	1.50	FFTBM
KINS	MARS-KS	2	(1.81)	13.3	-	CIRCÉ
KAERI-1	COBRA	4	/1.31	3.34	1.04	CIRCÉ
KAERI-2	COBRA	4	/1.57	2.18	3.04	Data assimilation
CEA	CATHARE	2 +1(bias)	1.97/	-	1.05	CIRCÉ
IRSN	CATHARE	3	3.0/	-	73.	DIPE
BelV	CATHARE	3	/2.56	-	1.04	CIRCÉ
GRS	ATHLET	6	1.47/2.0	3.2	~6.0 [2.5 – rel. velocity]	Inv. Uncertain.
VTT	APROS	6	34.0/ (2.62)	7.0	4.29	FFTBM +CIRCÉ
OKBM-2	KORSAR	2	1.64/	-	4.92	CIRCÉ

Figure 3.3.1: Quantified uncertainty ranges applied for HTC at dry out conditions: HTC dry out and HTC film wall to gas

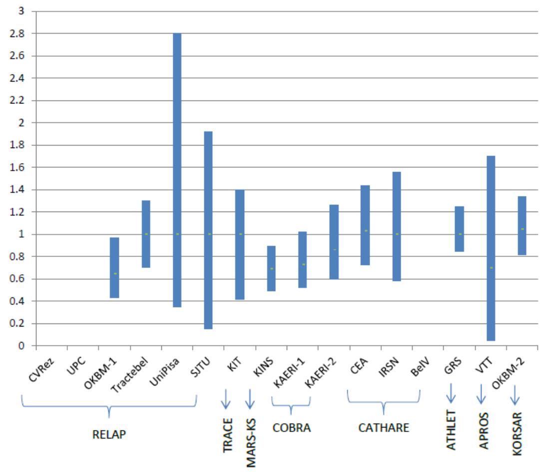


Figure 3.3.2: Quantified uncertainty ranges applied for HTC at dry out conditions: HTC global and HTC film wall to liquid

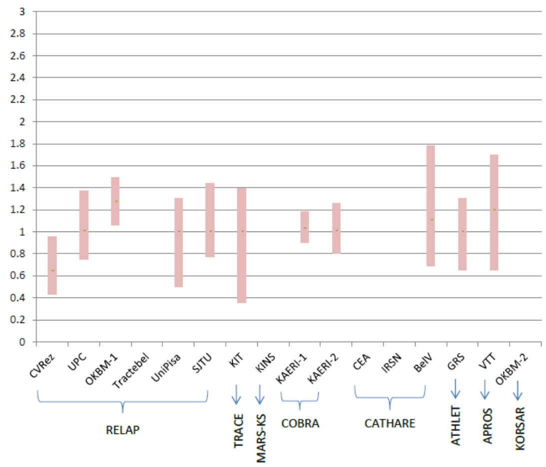


Figure 3.3.3: Quantified uncertainty ranges for interfacial heat transfer

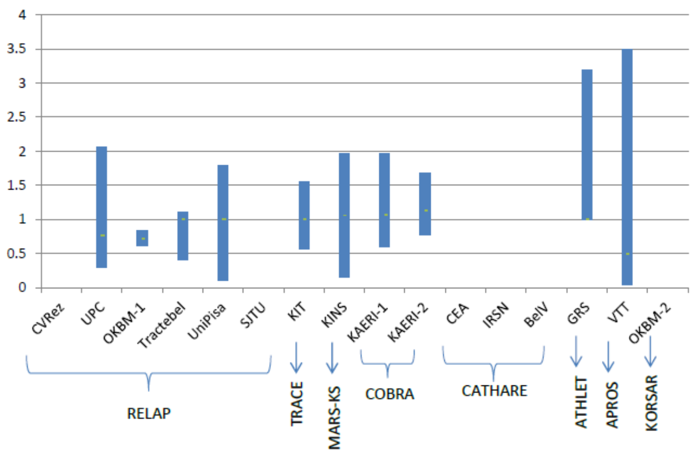


Figure 3.3.4: Quantified uncertainty ranges for interfacial friction

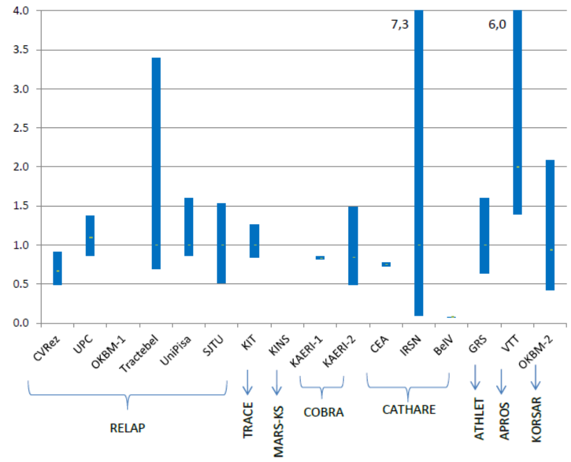


Figure 3.3.5: Uncertainty ranges for wall heat transfer: HTC dry out or HTC film-gas

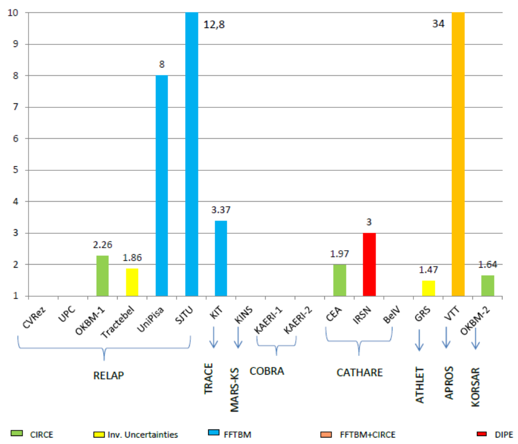


Figure 3.3.6: Uncertainty ranges for wall heat transfer: HTC – global or HTC– film-liquid

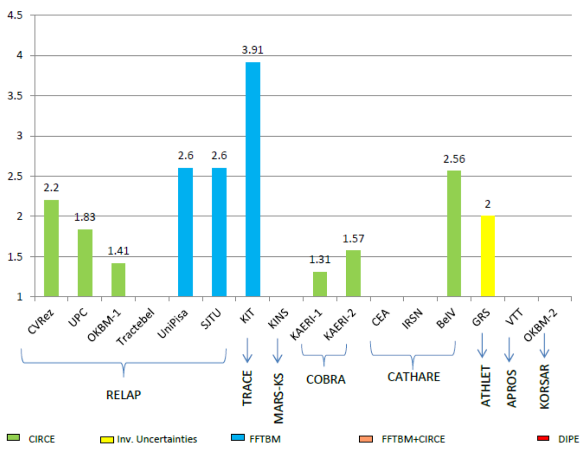


Figure 3.3.7: Uncertainty ranges for interfacial HTC

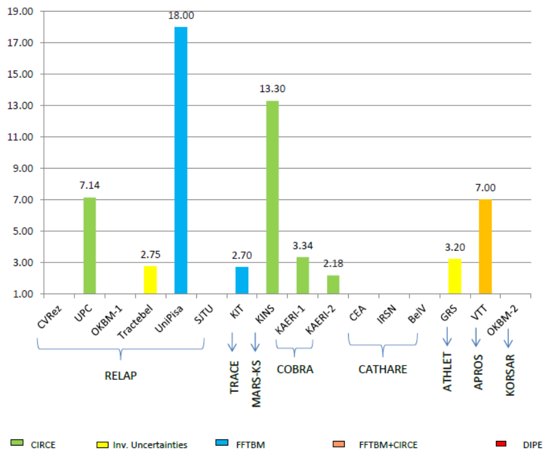
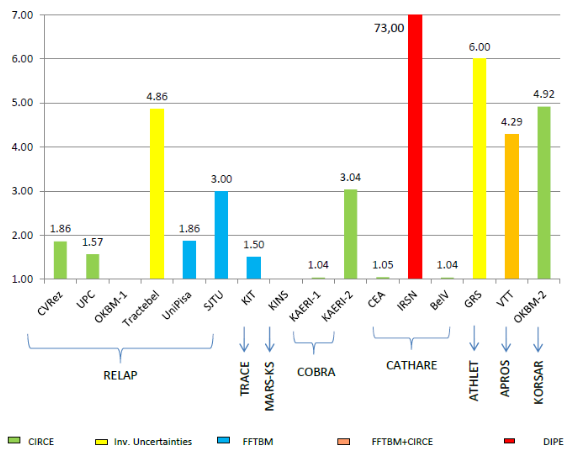


Figure 3.3.8: Uncertainty ranges for interfacial friction



The direct comparison of the lower and upper limits of the quantified uncertainty ranges would be also of interest. For instance, for the HTC at dry out conditions (see Figure 3.3.2) the uncertainty ranges quantified for the RELAP code by CVRez and OKBM-1 are disjoint. The complete uncertainty range quantified by CVRez is below 1.0 and the range quantified by OKBM-1 above 1.0. But CVRez applied in their analyses global multiplier for all the heat transfer regimes, whereas OKBM-1 considered correlation for film boiling heat transfer regime, only. Moreover, OKBM-1 used a special code RELAP version SCDAPSI, what makes the comparison even more difficult. It appears that even for the heat transfer at dry out conditions, a phenomenon considered by all participants, only a qualitative comparison is possible.

The uncertainty ranges obtained for interfacial heat transfer are very different for RELAP code. No user of CATHARE code has considered it as uncertain parameter. Regarding the method of quantification the uncertainty ranges obtained by participants using CIRCÉ as well as those using FFTBM are quite different. The ranges obtained by GRS and TRACTEBEL can be described as moderate.

The largest differences between uncertainty ranges can be observed for interfacial friction. The uncertainty ranges are very different for all codes. Some regularity can be found by uncertainty ranges arranged according to the quantification method. The participants using CIRCÉ quantified very small uncertainty ranges with exception of OKBM-2. The application of FFTBM lead also to rather small uncertainty ranges. However, two users of FFTBM used, in addition to interfacial friction multiplier droplet diameter, a key parameter available in RELAP input and used in the interfacial friction model. GRS and TRACTEBEL obtained moderate uncertainty ranges. IRSN, using DIPE method, obtained for code CATHARE a very large uncertainty range. It is particularly remarkable that the uncertainty range of interfacial friction obtained by IRSN for the code CATHARE2 was the largest uncertainty range obtained by a participant. Simultaneously, the uncertainty ranges obtained for code CATHARE2 interfacial friction by CEA and BelV (users of CIRCÉ) were the smallest uncertainty ranges obtained by any participant in Phase III. In the case of CEA it is the uncertainty of exactly the same model as that considered by IRSN.

The problem of the origins of these large differences appears to be very complex. CEA and BelV did not consider pressure drops as responses in their quantification process, and this fact could lead to underestimation of the interfacial friction quantification. However, another CIRCÉ user (OKBM-2), who did not considered pressure drops obtained quite reasonable ranges for interfacial friction. IRSN did not consider pressure drops as well. It seems that the main reason for the discrepancy would lie in the different assumptions of both methods concerning the parameters dependency.

The differences between CIRCÉ users are more difficult to explain. But the uncertainties quantified should be regarded as a whole set of values i.e. the value of calibrated parameter and range of its variation is valid only together with the values obtained for other parameter considered in the quantification process and not alone.

3.3.5 Preliminary check of quantified model uncertainties.

The last step in Phase III has been a preliminary check of quantified uncertainty ranges, by performing an uncertainty and sensitivity analysis of the Test run 216 from Series I of FEBA. It is the FEBA test most similar to the PERICLES experiment, from the standpoint of boundary conditions. This preliminary check has been introduced for confirmation of the quantification results.

The participants performed uncertainty analyses with the aim to obtain (95, 95) tolerance intervals for the following results:

- Cladding temperature as a function of time at three different levels: 3315 mm, 2225 mm and 1135 mm.
- Pressure drop along the total channel length and along the middle part of the channel.
- Water carried over the heated channel.
- Quench front propagation.

The results of the analysis were compared with the experimental data.

In general, the experimental data were bounded by the tolerance intervals. Only in the case of two CIRCÉ users and two FFTBM users some experimental data are not bounded by calculated (95, 95) uncertainty limits. It concerns mainly cladding temperatures at time when QF is passing the level of thermocouple location. There are some discrepancies of experimental data and predicted uncertainty intervals at the beginning of the transient. The measured cladding temperature at 3315 mm decreases immediately after the start of reflooding, but the predictions show instead a short period of temperature increase. The discrepancy can be attributed to differences between the reported initial conditions and the real ones during the test.

There are also discrepancies for the pressure drop along the total channel length. The measured value is clearly higher than the predicted ones at the start of the transient. The reason may be a difference between nominal and real initial conditions and/or a measurement error. Since the participants did not use the data measured at the very beginning of the transient for quantification, it can be expected that it did not affect the quantified uncertainties.

For FEBA test 216, sensitivity analyses have been performed. A majority of participants identified as the most influential parameters for cladding temperature those related to the wall heat transfer at dry out conditions. Some participants identified parameters related to interfacial heat transfer as the most influential for the cladding temperature.

A majority of participants identified the interfacial friction factor as the most influential parameter on pressure drop in the test section. GRS and TRACTEBEL identified this factor as the most influential on all investigated output. BelV identified interfacial friction factor as the most influential parameter not only for pressure drop but also for QF propagation and cladding temperature in the upper part of the test section. BelV determined very small variation range of the interfacial friction but a very large bias. Obviously the interfacial friction factor was identified as so important parameter due to the large bias and not due to the variation range, which is one of the smallest ones identified by participants.

For the water carried over, mainly interfacial friction factor and interfacial HTCs were identified as most influential.

CEA was the sole participant who identified parameters related to the special heat transfer model at the QF as the most influential for QF elevation, cladding temperature at the bottom of the test section and water carried over. This results from the large bias found by CEA for this parameter; however, the parameter was not varied explicitly. Other participants considering such parameters have not counted them among influential parameters.

Concerning QF propagation, the majority of participants found the interfacial friction factor as the most influential one. Some found the HTC correlations as the most influential. Only CEA has found that the parameter related to heat transfer enhancement at the QF is the most influential parameter for the QF propagation.

3.4 Phase IV: confirmation / validation of the uncertainties found within Phase III

The goal of Phase IV has been the confirmation and validation of the uncertainties determined during Phase III, by propagating them to the 6 FEBA tests considered in Phase III, and to the 6 tests of the 2D reflood PERICLES experiment [11, 15]. This latter step has been performed blindly.

Phase IV has been co-ordinated by CEA and IRSN. 15 organisations were involved (see Table of Participants). They are the same involved in Phase III, with the addition of PSI. KIT had an incomplete participation in Phase IV, concerning only the uncertainty propagation for the FEBA tests, and excluding the PERICLES tests.

The participants basically considered the same parameters as in Phase III, all of them related to physical models. Nevertheless, two participants suppressed one parameter each, because they were considered as included in another choice. One participant (TRACTEBEL) added 5 IP representing experimental uncertainties of the boundary conditions rather than model parameters (local heat flux, bundle power, inlet water temperature, system pressure and inlet velocity).

PSI (not involved in Phase III) considered a large number of IP, including boundary conditions, material properties and physical models. The uncertainty of these IP is estimated by expert judgement, literature review and confirmatory UQ based on the 6 FEBA tests available through PREMIUM.

The ranges of variation found for the IP in Phase III are never modified in Phase IV.

The type of PDF assigned to the selected inputs was dependent on the method, and included normal, log-normal, uniform, log-uniform and histograms. The influence of the PDF was not investigated by most of the participants. It is an issue out of the scope of PREMIUM benchmark.

In the specification of Phase IV [15] it was required, both for FEBA and PERICLES, 200 code runs and the use of 5th and 195th order statistics, which are point estimators of the 2.5 and 97.5 percentiles, respectively. Additionally, they are, approximately, one-sided tolerance limits with level (95, 95). They define a two-sided interval that has an expected coverage of 0.945.

Some participants did not follow these specifications, and used different number of calculations as well as different order statistics:

- IRSN performed 119 calculations and used the sample extremes. This defines an interval with an expected coverage of 0.983. SJTU performed 93 calculations and used the extremes. This defines a (95, 95) two-sided interval, with an expected coverage of 0.979.

- KINS performed, only for FEBA, 124 calculations and used 3rd OS (i.e. orders 3rd and 122nd). This corresponds to (95, 95) one-sided limits, and defines an interval with an expected coverage of 0.952.

Table 3.4.1: Expected coverage for different choices of interval (sample size/order statistics).

SAMPLE SIZE	INTERVAL (OS)	EXPECTED COVERAGE
200	[5th , 195th]	0.945
119	[1st, 119th]	0.983
93	[1st, 93rd]	0.979
124	[3rd, 122$^{nd]}$	0.952

Table 3.4.1 summarises the features of the aforementioned intervals. The expected coverage of an interval with endpoints in two order statistics $X_{r:n}$ and $X_{s:n}$ (r<s) of a scalar random variable X, obtained from a simple random sample of size n, is

$$\frac{s-r}{n+1} \qquad (3.4.1)$$

The influence of these different choices is assumed to be of secondary importance compared to the effect of the used quantification method and other effects.

Phase IV has been developed in two steps:

- Step 1: uncertainty analysis of the 6 FEBA tests considered in Phase III.
- Step 2: uncertainty analysis of the 6 PERICLES tests. This step is performed blindly.

3.4.1 Uncertainty analysis of FEBA experiments

The analysis of the results has been performed according to two methods (see Section 2.5). The first one is qualitative and carried out by CEA, the second one is quantitative and carried out by IRSN. They use different calculated outputs.

The qualitative analysis performed by CEA is intuitive. To begin with, all the contributions are systematically analysed. More precisely, three issues are addressed for each output:

1) Does the uncertainty band envelop the experimental time trend?
2) How the nominal calculation is located with respect to the experiment (under or overestimation)?
3) How wide is the uncertainty band?

For the CIRCÉ users, a fourth question must be raised:

4) Does the calibrated calculation improve the nominal calculation?

The results of this analysis are afterwards gathered in order to distinguish main trends and try to draw some conclusions.

This method is extensively applied to time trends of clad temperature type:

- Clad temperature time trends at the middle of the bundle (respectively at 2225 mm and 1828 mm for FEBA and PERICLES);
- Clad temperature time trends in the upper part of the bundle (respectively at 1135 mm and 2998 mm for FEBA and PERICLES, knowing that for FEBA the elevation 0 mm corresponds with the top of the bundle, whereas for PERICLES, the elevation 0 mm corresponds with the bottom of the bundle);

Middle pressure drops time trends are also analysed, but less detailed, due to their very oscillatory behaviour.

For the quantitative analysis performed by IRSN, scalar outputs have to be provided by the participants. They are deduced from the time trends used by CEA for its qualitative analysis and are:

- For FEBA: The clad temperatures at the same time for all the time trends of a given test: the time of the maximum experimental clad temperature. These temperatures are considered in the middle and in the upper part of the bundle, i.e. respectively at 2225 mm and 1135 mm.
- For PERICLES: The maximum clad temperatures, even if the time where this maximum value is reached is not the same one for all the time trends. These temperatures are considered in the middle and in the upper part of the bundle, i.e. respectively at 1828 mm and 2998 mm.
- For FEBA and PERICLES: The quench times in the middle and in the upper part of the bundle, i.e. respectively at 1828 mm and 2998 mm.

3.4.1.1 Qualitative analysis by CEA

A synthesis of the CEA analysis is presented in Table 3.4.2. Participants have been ranked in four groups, according to the position of the uncertainty bands with respect to the experimental data of both types: cladding temperatures during the whole transient (not only the peak cladding temperature, PCT) and quench times:

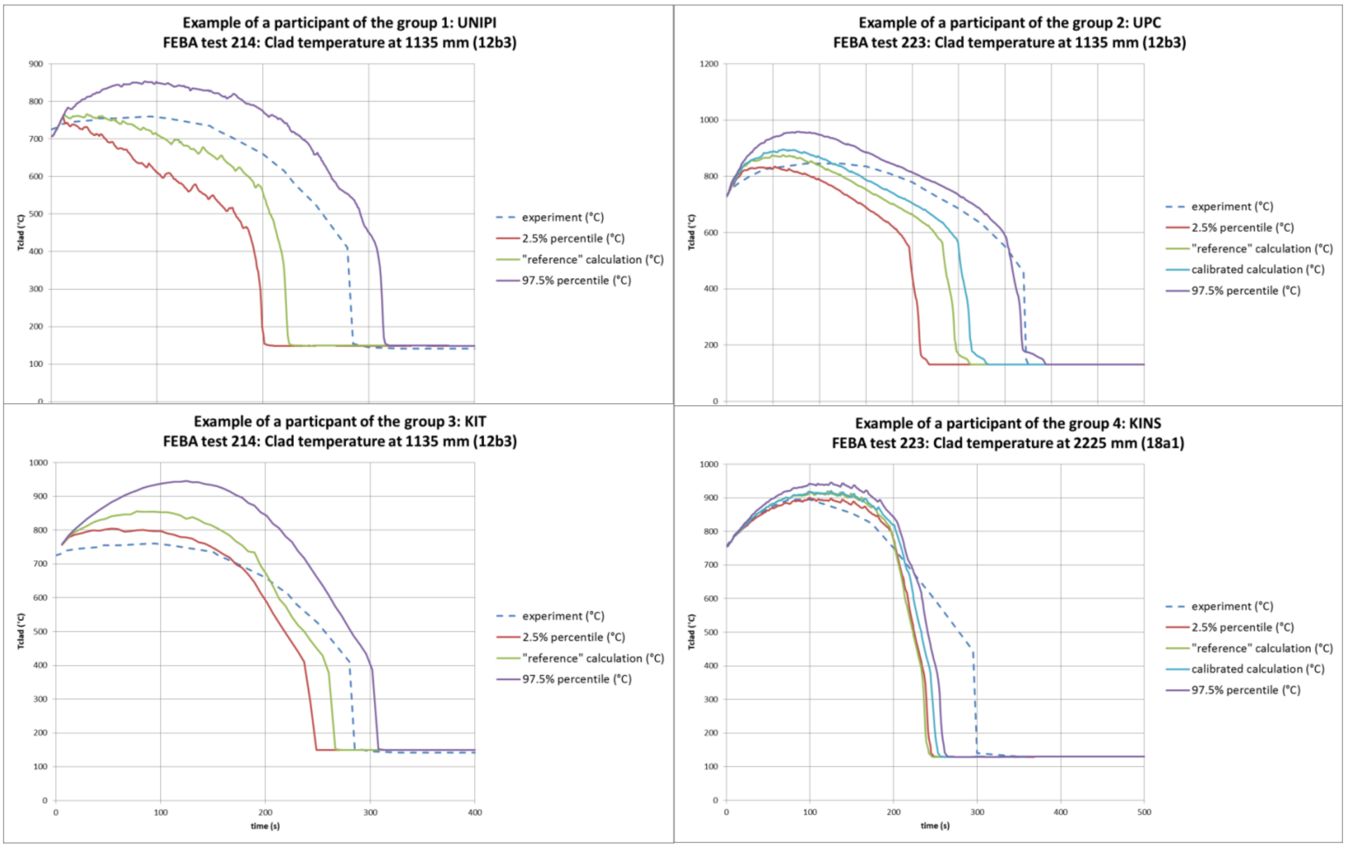
- First group (7 participants): the uncertainty bands envelop the experimental clad temperatures and quench times for all tests during the whole (or almost the whole) transient.
- Second group (5 participants): experimental data are bounded except for some tests just before the quenching.
- Third group (2 participants): experimental data are not enveloped during a rather long interval at the beginning of the transient.
- Fourth group (3 participants): experimental data are systematically not bounded at the end of the transient.

Table 3.4.2: Summary of the FEBA uncertainty results

General result	Participant	Code	Method	Features of the nominal calculation				Width of the uncertainty bands
				Tclad at 2225mm	tque at 2225mm	Tclad at 1135mm	tque at 1135 mm	
Exp. data bounded for all the time trends	IRSN	CATHARE	DIPE			underestimated		very wide
	SJTU	RELAP	FFTBM	underestimated	underestimated	underestimated	underestimated	very wide
	PSI	TRACE	expert judgement		overestimated			very wide
	UNIPI	RELAP	FFTBM	underestimated	underestimated	underestimated	underestimated	wide
	VTT	APROS	FFTBM+CIRCE	underestimated		underestimated		wide
	Tractebel (limit)	RELAP	IUQ	underestimated		underestimated	underestimated	wide
	CVRez (limit)	RELAP	CIRCE	underestimated	underestimated	underestimated	underestimated	medium
Exp. data bounded except for some tests just before the quenching	CEA	CATHARE	CIRCE		underestimated		underestimated	medium
	UPC	RELAP	CIRCE				underestimated	medium
	OKBM-Korsar	KORSAR	CIRCE		underestimated		underestimated	medium
	BelV	CATHARE	CIRCE			underestimated		medium
	GRS	ATHLET	IUQ	overestimated	underestimated	overestimated	underestimated	wide
Exp. data not at all bounded at the beginning	OKBM-Relap	RELAP	CIRCE			overestimated		narrow
	KIT	TRACE	FFTBM			overestimated	underestimated	medium
Exp. data not bounded	KAERI	COBRA	CIRCE		underestimated		underestimated	narrow
	KAERI	COBRA	MCDA		underestimated		underestimated	narrow
	KINS	MARS-KS	CIRCE		underestimated		underestimated	narrow

An example of results for each group is shown in Figure 3.4.1

Figure 3.4.1: Examples of FEBA results, for participants in each group of Table 3.4.3



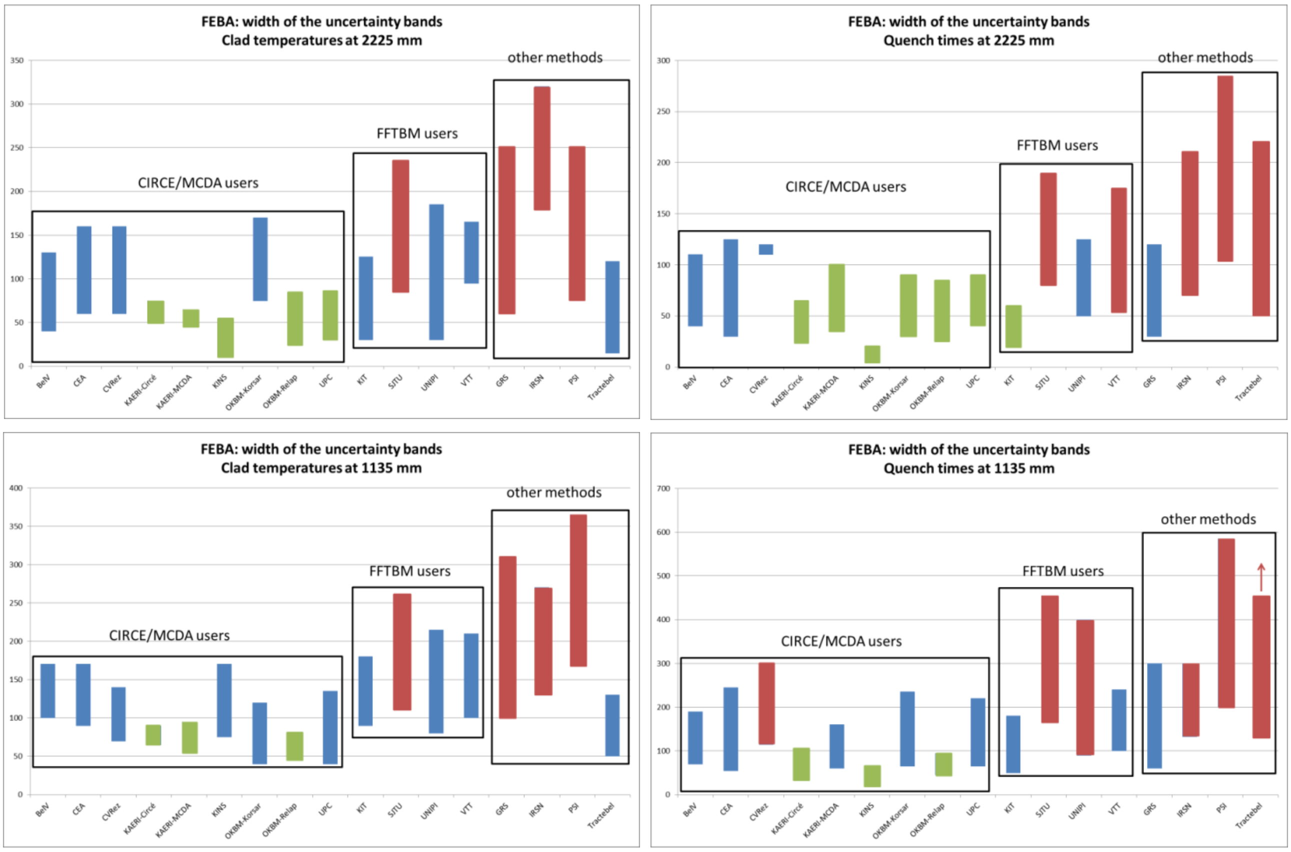
The under or overestimation is indicated for the following scalar outputs:

- Clad temperature and quench time at 2225 mm.
- Clad temperature and quench time at 1135 mm.

The empty boxes correspond to cases where no systematic trend is observed or where the nominal calculations are globally good.

In the last column, the width of the uncertainty bands is indicated, with 4 levels: very wide, wide, medium and narrow. For each of the 4 scalar outputs, 6 uncertainty bands are obtained for the 6 FEBA tests. The global width is expressed as an interval formed by the minimum and maximum of the 6 widths, and is represented in Figure 3.4.2 for the 17 participants and the four scalar outputs of FEBA tests. Colour green, blue and red correspond to narrow, medium and wide uncertainty bands, respectively. Small arrows indicate that the quenching is not reached in the calculations.

Figure 3.4.2: Width of the uncertainty bands obtained for the 4 scalar outputs of the analysed FEBA tests



A main conclusion is that, except for the last group of participants, the envelop calculations are globally successful. It seems logical, since the participants used data from FEBA in the quantification, and the exercise was not blind. Only KAERI and KINS, among CIRCÉ users, considered exclusively the cladding temperature as responses, and excluded the quench times; as a consequence, their results do not envelop the quench times. Moreover, KINS only consider 2 IP, without any parameter related to interfacial friction.

The 7 participants of the first group who enveloped all the data obtained very wide or wide uncertainty bands (except for CVRez). In the opposite side, the participants of the last group who did not envelop the experimental values had narrow bands. The two intermediate groups had mostly medium uncertainty bands. This means that the apparently more efficient or precise results failed to encompass the real results, and, conversely, those that did envelop were the less informative.

Another observation concerns the quality of the nominal calculations. Results may envelop real data even if the nominal calculations are not very satisfactory. It is very evident for the participants of the 1st group, whose bands are very wide.

In some cases, wide bands do not envelop data due to problems in the nominal calculation. The participants of the 3rd group have problems in predicting the adiabatic rise. In the opposite, there are participants who, despite having fairly narrow bands, envelop the data, because they obtain a good nominal calculation. A conclusion is that having a good nominal calculation is a sufficient, but not necessary condition, to envelop the FEBA data.

There is a very clear relationship between the type of quantification method and the width of the uncertainty bands. CIRCÉ and MCDA produced the narrowest bands, while FFTBM produced medium-large bands. The widest bands were produced by the "other methods" (Figure 3.4.2).

A particularity of CIRCÉ is the possibility to estimate the so-called "calibrated value" of the IP, which is the median of their uncertainty distribution, in general different from the nominal value. Theoretically, the calibrated calculation improves the nominal one, by decreasing any systematic under or overestimation. All the CIRCÉ users performed this calculation, except for OKBM-KORSAR. For most of the participants, the calibrated calculation slightly improved the nominal one but not by far.

About the number of IP considered in the quantification process, it ranges from 2 (CVRez, OKBM-KORSAR, KINS) to 8 (TRACTEBEL, who considered 3 inputs for Phase III, and 5 additional parameters for Phase IV). PSI considered the largest number, 26 IP, but used expert judgement in the quantification. The CIRCÉ users consider generally less inputs than FFTBM and other methods. CIRCÉ estimates the range of variation of the IP so that the coverage of the experimental response by the uncertainty bounds is independent of the number of parameters. If the number of inputs is reduced, the ranges of variation increase in order to account for the total uncertainty.

The uncertainty bands for quench times and clad temperatures are wider at 1135 mm than at 2225 mm.

Another feature observed in the results is an influence of boundary conditions on the results. UNIPI noticed that the PCT prediction was better for tests with higher inlet flow rate, while the width of uncertainty bands seemed unchanged. Contrary to this, pressure had no influence on the quality of PCT prediction, but the width of the bands decreased when pressure increases. This behaviour illustrates the problems of extrapolating the uncertainty determined with a set of boundary conditions to a different set.

There are few cases where the experimental scalar outputs are bounded by the results, while the time trends are not bounded in some intervals. Problems in predicting the adiabatic rise may produce this kind of behaviour.

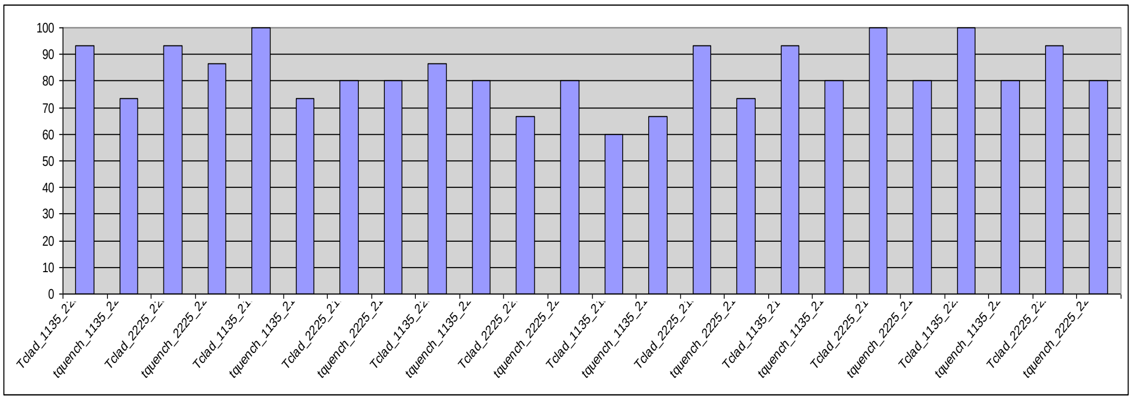
In the quantification of input uncertainties, some participants did not consider the quenching times as responses. The consequence was that, contrary to clad temperatures, quenching times were not enveloped. The opposite was also observed in the 3rd group of Table 3.4.2: the quench times were enveloped, but not the clad temperature. The cause seemed to be a problem in the prediction of the adiabatic rise.

3.4.1.2 Quantitative analysis by IRSN

The results of FEBA calculations were quantitatively analysed with IRSN methodology (Section 2.5), using the IRSN SUNSET software [48]. 15 contributions of 24 scalar outputs (12 related to clad temperatures and 12 to quench times) were collected.

First of all, the percentage of uncertainty bands covering the corresponding experimental value was calculated, taking into account all participants (Figure 3.4.3). The average percentage is 83%, lower than the values of expected coverage for calculated tolerance bands (Table 3.4.1). The frequency of coverage was better for clad temperatures (88%) than for quench times (77%).

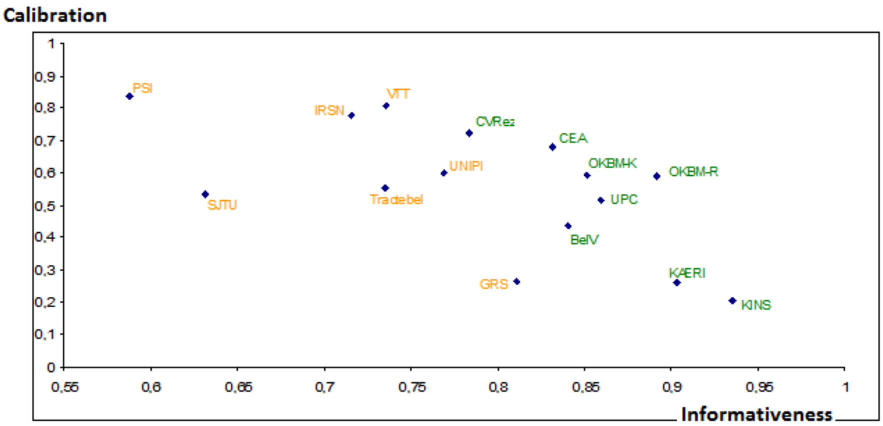
Figure 3.4.3: FEBA: percentage of experimental values falling inside uncertainty bands



The information provided to IRSN for each scalar output was an interval [LUB, UUB] and a reference value RV. However, some participants did not provide the whole information for all outputs. For most of them, it was due to not attained UUB. Therefore, in order to perform the analysis on the same number of outputs, the not attained UUBs were set to the maximum of the UUBs given by participants. The IRSN methodology was applied, and for each participant, two scores were calculated, quantifying the informativeness and the calibration (Figure. 3.4.4). A negative correlation coefficient (-0.6) is found in these data, pointing out that wide uncertainty bands tend to bound the experimental data, giving a good calibration score, but they are not very informative. Conversely, narrow bands are very informative, but tend to fail in enveloping the data. UNIPI score is the closest to the centre of gravity of the cloud.

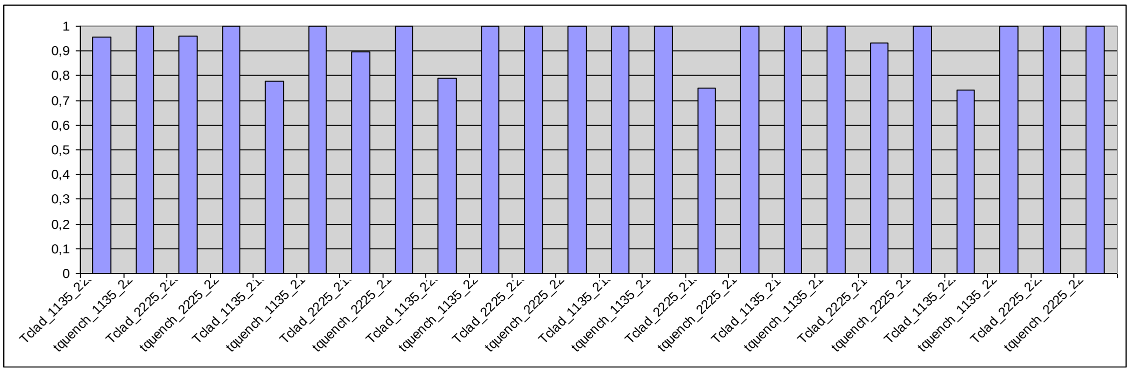
KINS and KAERI, who are CIRCÉ users, provide narrow uncertainty bands and poor results that never encompass the experimental value for all time variables.

Figure 3.4.4: FEBA: informativeness and calibration scores for each participant
Green colour stands for CIRCÉ users



According to the methodology, the conflict indicator was computed for the 24 scalar outputs (Figure 3.4.5). The large values obtained for this indicator points out a strong disagreement in the results among the participants.

Figure 3.4.5: FEBA: Value of the conflict indicator for each output



The effect of the quantification methodology has been studied as well. In Figure 3.4.4, the results obtained by CIRCÉ are plotted in green. It is observed that the results are method dependent. CIRCÉ produces narrow uncertainty bands, so that the informativeness score is generally high but the calibration score tends to be lower compared to other participants.

Figures 3.4.6 and 3.4.7 show the calibration vs. informativeness scores plot separating the clad temperatures and the quench times. The very low calibration score of GRS, KAERI and KINS in Figure 3.4.7 has an easy explanation: these participants did not consider as responses the quench times in their quantification process.

Figure 3.4.6: FEBA: Informativeness and calibration scores for each participant for temperature variables

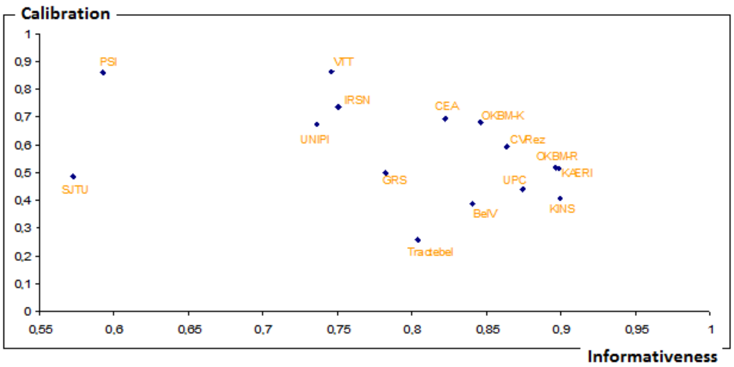


Figure 3.4.7: FEBA: Informativeness and calibration scores for each participant for time variables

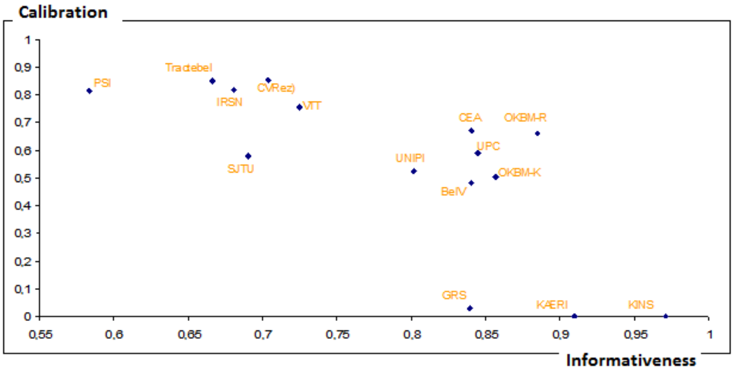
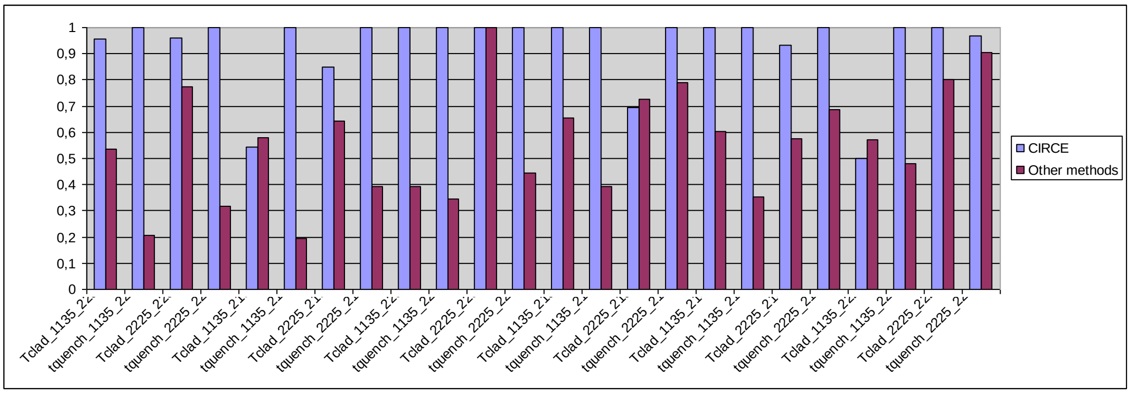


Figure 3.4.8 plots the conflict indicator, separating CIRCÉ from the other methods. There is a higher lack of coherence among the CIRCÉ users. This is explained by the combination of narrow uncertainty bands, and the fact that some participants (KAERI, KINS) never encompass the experimental values.

Figure 3.4.8: FEBA: conflict indicator with respect to the used methodology (CIRCÉ/ Other methods)



3.4.2 Uncertainty analysis of PERICLES experiments

The second part of Phase IV was the propagation of the model uncertainties obtained in Phase III to the results of the PERICLES tests.

The exercise was performed blindly, using the *a priori* information provided in the specifications [15]. As described in Section 2.4, the difference of 2D reflood PERICLES experiment with respect to FEBA in the presence of 2D effects. As shown in Table 3.4.3, the participants used different modelling criteria. In general, a multi-channel model with crossflows was chosen (e.g. RELAP users). Four participants chose a 3-D model. IRSN used CATHARE with the 1D modelling, arguing that this would allow the extrapolation to PERICLES of the uncertainties found for FEBA. Using a 2D or 3D model would add a new set of uncertain models (mainly those governing the crossflows), which cannot be estimated from the FEBA experimental data.

Table 3.4.3: Type of modelling for PERICLES

Code	Participant	Type of modelling		
		multi-channels	3-D	1-D
APROS	VTT	x		
ATHLET	GRS	x		
CATHARE	Bel V		x	
	CEA		x	
	IRSN			x
KORSAR	OKBM	x		
MARS PREMIUM version	KINS	x		
MARS COBRA-TF	KAERI		x	
RELAP	CVRez	x		
	OKBM	x		
	SJTU	x		
	TRACTEBEL	x		
	UNIPI	x		
	UPC	x		
TRACE	PSI		x	

The number of axial nodes in the heated part of the bundle is generally close to the corresponding number for FEBA, except for the users of a 3D model, who assigned a number of axial nodes for PERICLES significantly lower than for FEBA. Contrary, PSI duplicated the number of nodes in moving from FEBA to PERICLES. The majority of participants model also the bottom plate, upper tie plate, upper core plate and the housing. Average values were assigned to thermal conductivity and heat capacity of structural materials.

Unlike for FEBA, the experimental clad temperature profile was not given, except for the central assembly of one of the tests. The majority of participants respected as far as possible the experimental procedure, with the whole power imposed to the fuel rods until the initial maximum clad temperature of the central assembly reaches the specified value, and after that start the injection of cold water.

The theoretical maximum initial clad temperature is slightly higher than the really observed one. The reason is that the measurement of only one thermocouple was considered. This little discrepancy will be kept in memory for the comparison of the experimental clad temperatures with the uncertainty bands. It can explain why these uncertainty bands do not perfectly envelop the experimental data at the beginning of the transient, independently from the input uncertainties.

The quantified IP considered are the same as for FEBA with two exceptions:

- GRS considers 2 additional parameters: relative velocity in cross connection, and bundle total power.
- PSI considers a total of 34 IP for PERICLES (26 for FEBA), with the 8 supplementary parameters being related to an increased number of boundary conditions for outlet pressure, inlet

mass flow and temperature of the coolant, and heater rods power (physical model parameters were the same for PERICLES and FEBA exercises).

The parameters considered by the participants for interfacial friction and wall-to-fluid heat transfer correlation were not always the same for FEBA and PERICLES. Some participants observed high sensitivity of the quality of nominal calculations to these modelling features.

The uncertainty propagation for PERICLES was based on 200 code runs, as required in the specification. Exceptions were IRSN (119 runs) and SJTU (93 runs). All the code runs were successful, except for PSI and TRACTEBEL, who simply replaced the failed code runs by new runs.

3.4.2.1 Qualitative analysis by CEA

As in FEBA, CEA performed a qualitative analysis of the PERICLES results. In Table 3.4.4, a synthesis of results is provided, with the same structure than Table 3.4.2 for FEBA, but only considering three groups:

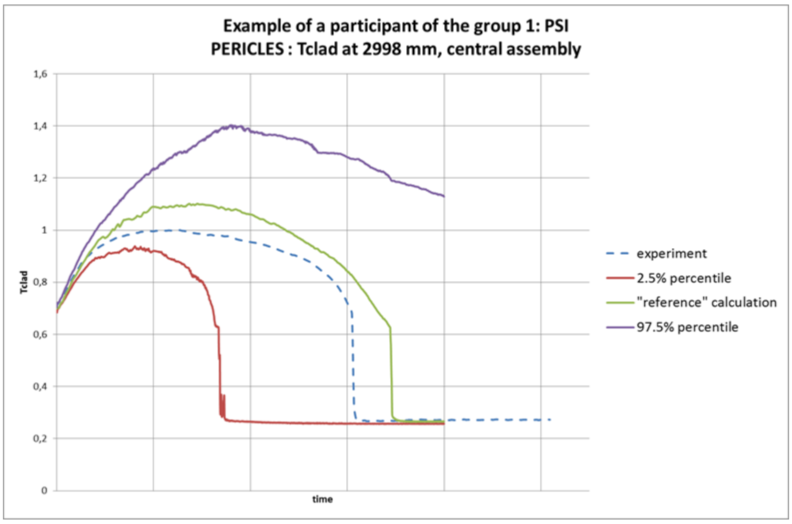
- First group (6 participants): the uncertainty bands satisfactorily envelop the experimental data for all time trends, even if they are not perfect at the beginning. The problem at the beginning is not provoked by the input uncertainties, and is probably due to the lack of precision in initial clad temperature. Uncertainty bands are wide or very wide.
- Second group (3 participants): experimental data are bounded most of the time, except for the end of the transient for some cases. Uncertainty bands are medium.
- Third group (7 participants): experimental data are, in general, not bounded. Uncertainty bands are narrow or medium, with the exception or GRS (very wide)

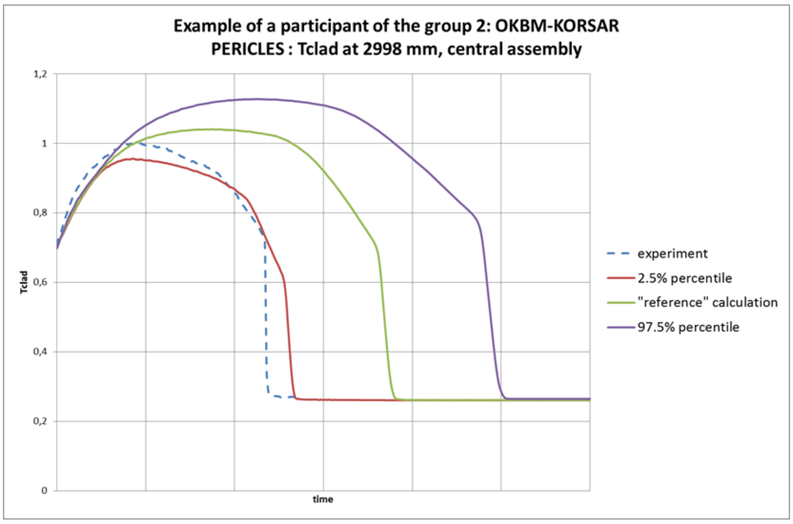
Table 3.4.4: Summary of the PERICLES uncertainty results

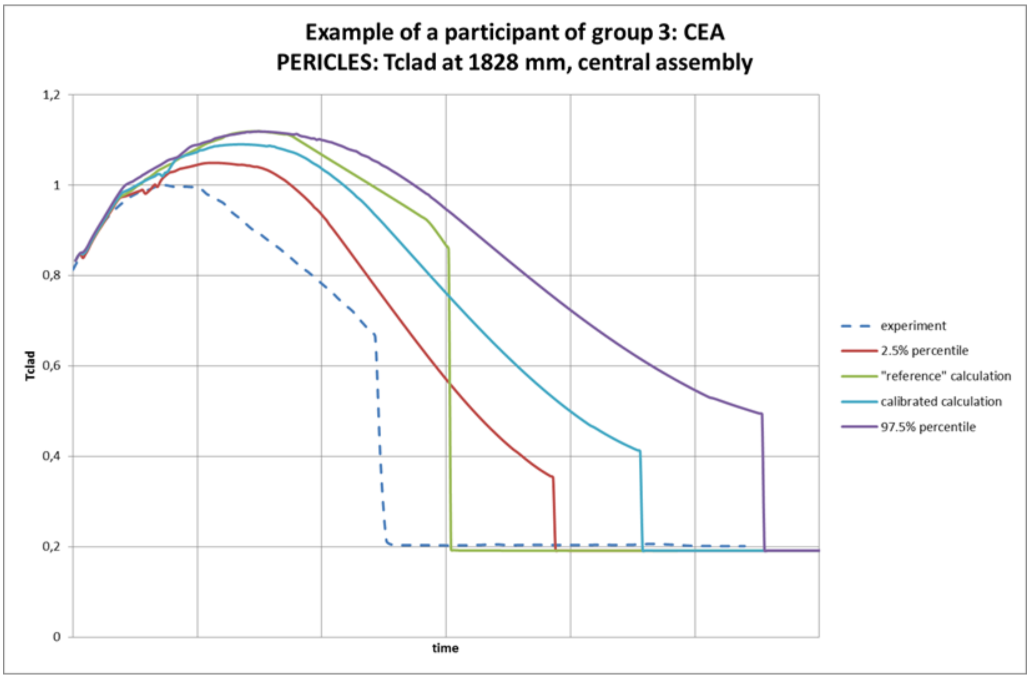
General result	Participant	Code	Method	Features of the nominal calculation				Width of the uncertainty bands
				Tclad at 1828 mm	t_{que} at 1828 mm	Tclad at 2998 mm	t_{que} at 2998 mm	
Exp. data bounded for all the time trends	VTT	APROS	CIRCE+FFTBM	underestimated	overestimated	underestimated	**very overestimated**	very wide
	IRSN	CATHARE	DIPE	overestimated				very wide
	SJTU	RELAP	FFTBM	underestimated	overestimated		**very overestimated**	very wide
	UNIPI	RELAP	FFTBM	underestimated	underestimated		underestimated	wide
	PSI	TRACE	Expert judgement	underestimated				very wide
	Tractebel	RELAP	IUQ	underestimated		underestimated		very wide
Exp. data bounded the majority of time, but not always	OKBM-Korsar	KORSAR	CIRCE		overestimated		**very overestimated**	medium
	UPC	RELAP	CIRCE		underestimated		**very underestimated**	medium
	CVRez	RELAP	CIRCE		underestimated		**very underestimated**	medium
Exp. data not bounded	GRS	ATHLET	IUQ		**very overestimated**	underestimated	**very overestimated**	very wide
	OKBM-Relap	RELAP	CIRCE			**very overestimated**	overestimated	narrow
	KAERI-MCDA	COBRA	MCDA		underestimated	**very overestimated**		medium
	KAERI-Circé	COBRA	CIRCE		underestimated	**very overestimated**		medium
	KINS	MARS-KS	CIRCE		underestimated	overestimated	**very underestimated**	narrow
	CEA	CATHARE	CIRCE	**very overestimated**	overestimated		overestimated	medium
	BelV	CATHARE	CIRCE	**very overestimated**	overestimated		**very overestimated**	narrow

An example of results of each group is presented in Figure 3.4.9. All the results are in a no dimensional form.

Figure 3.4.9: PERICLES: Examples of results of participants in groups of Table 3.4.4

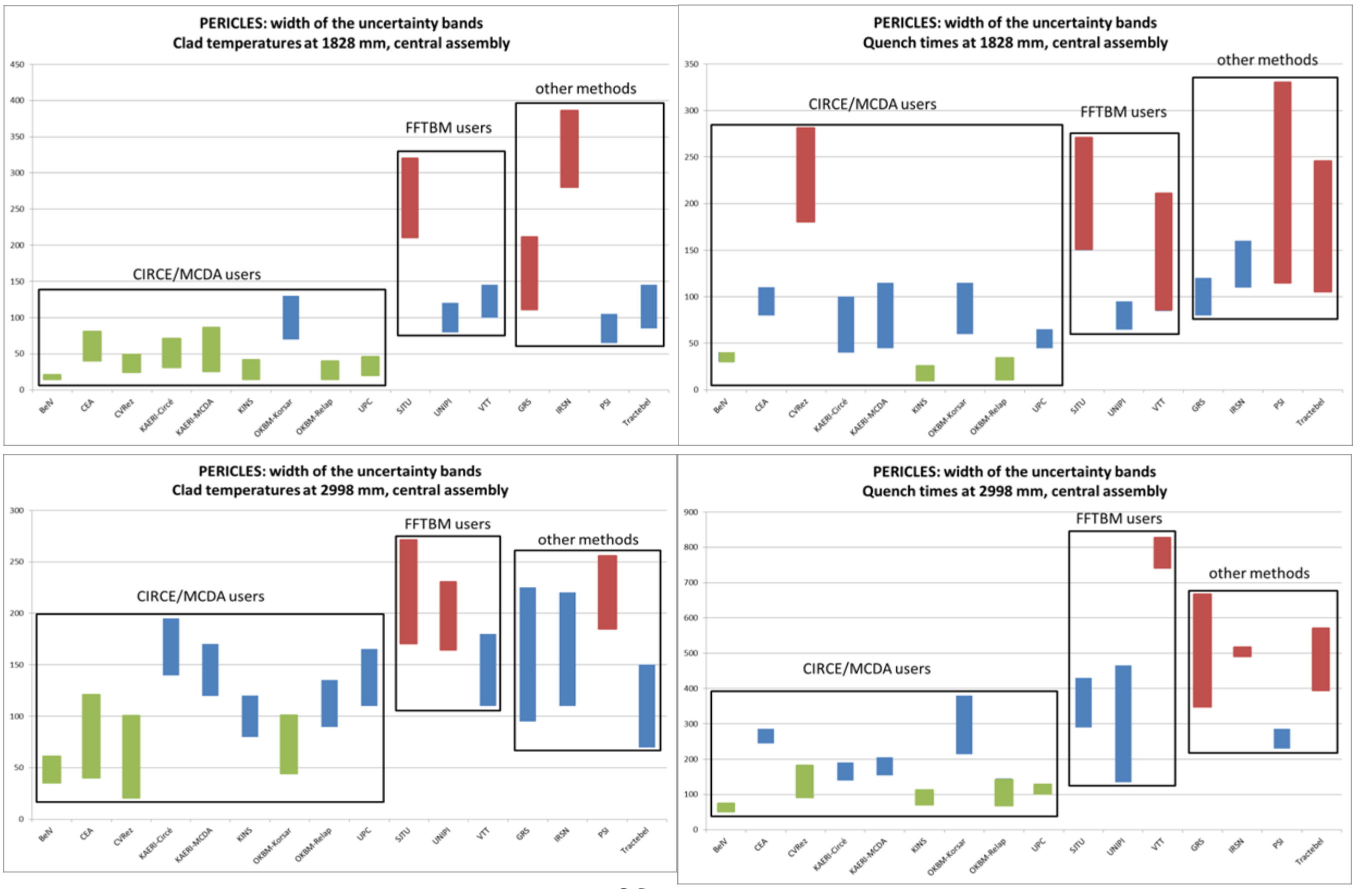






In the three groups, participants have been ranked by decreasing order of quality. As in the FEBA case, the table includes a classification of the systematic under or overestimations of the results, and of the uncertainty bands width. In Figure 3.4.10, the width of the bands (expressed by an interval) is plotted versus the participants for the 4 scalar outputs in the central assembly.

Figure 3.4.10: PERICLES: Width of the uncertainty bands for different participants and methods



More participants fail to envelop the experimental data for PERICLES than for FEBA. The uncertainties were quantified for FEBA, and the PERICLES exercise was blind.

There is a relationship between the quality of nominal calculations and the bounding of experimental data. This relationship seems more decisive for PERICLES than for FEBA, especially when the quality is poor. All participants having unsatisfactory nominal calculations do not have successful envelop calculations, especially when the PCT is poorly predicted. The exceptions are participants with very wide uncertainty bands.

Nominal calculations are globally better in the first group, though in some cases the quench time is overestimated. In the second group, the opposite behaviour is observed, with underestimation of the quench time and thus a fail to envelop the data at the end of the transient. Participants of the 3rd group have poor nominal calculations, particularly for the prediction of PCT.

Participants gave miscellaneous reasons for the poor quality of the nominal calculations. For instance:

- Insufficiencies of the specifications, concerning the description of PERICLES. E.g. the housing and the heat losses through it.
- Deviation of the axial mesh centres from thermocouples position.
- Deficiencies in the modelling of thermo-hydraulic phenomena: small break droplet breakup by spacer grids, wall-to-fluid heat transfer, interfacial friction (a too strong friction, due to the use of a 5 equations model, may lead to overestimating the water carryover), vapour mixing between central and lateral assemblies. The predicted reflood is, in some cases, too rapid and too slow in others.
- Differences between the 1D and 3D modules of CATHARE2.

Figure 3.4.10 shows a clear relation between the type of quantification method and the width of uncertainty bands. As in the FEBA exercise, CIRCÉ and MCDA users have, in general, the narrowest bands. FFTBM and other methods users have wider bands, with similar widths.

In the specific case of CIRCÉ users, no systematic change is observed when the FEBA-calibrated calculation is used.

Other remarks included in [11] are the following:

- There are globally less IP considered in CIRCÉ. But, according to CEA, this fact does not theoretically have any impact on the width of uncertainty bands.
- It is generally more difficult to envelop the clad temperature in the upper part (2998 mm) than in the middle of the bundle (1828 mm). But is some cases the opposite trend is observed. In FEBA, the bands are wider at 2998 mm than at 1828 mm. This seems logical for quench times, but it is also generally found for clad temperature.
- Uncertainty bands are wider for the central assembly than for the lateral ones, at both elevations. It is normal, because clad temperatures and quench times are higher in the central assembly.
- There is a pressure effect: calculated results are degraded for the test at higher pressure, compared to the other PERICLES tests. Uncertainty bands are narrower when pressure increases.
- There are some cases where the 4 scalar outputs are bounded, but not the whole time trends. It is the case of participants having difficulties in the prediction of the adiabatic rise. Participants who

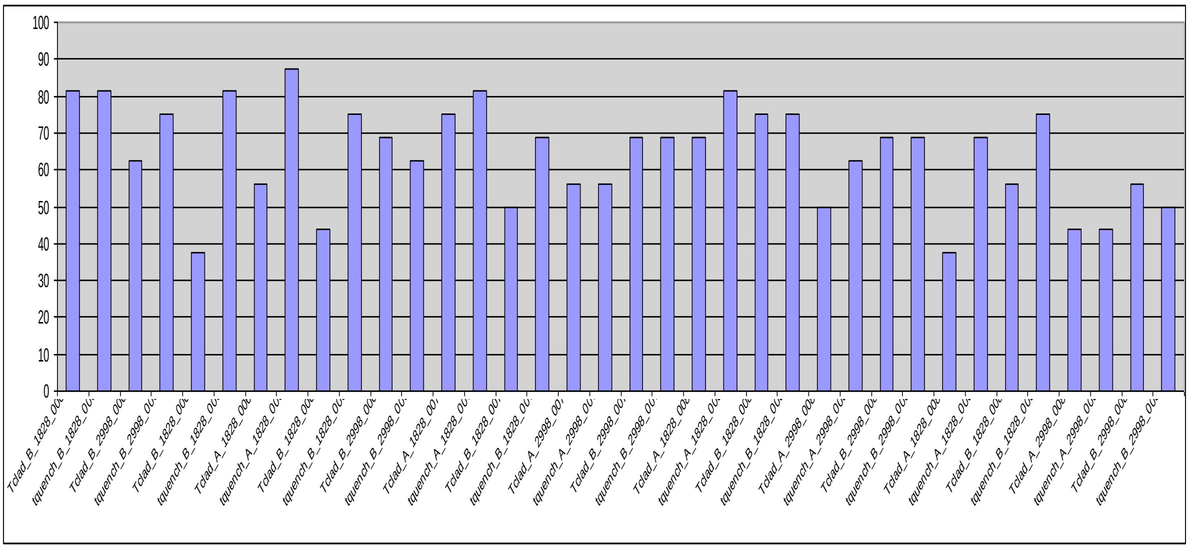
did not consider quench time as responses in Phase III generally envelop the experimental quench times for PERICLES.

3.4.2.2 Quantitative analysis by IRSN

It has been performed with the SUNSET software of IRSN [48]. 16 contributions were collected on 36 scalar outputs (18 related to clad temperature, 18 related to quench time).

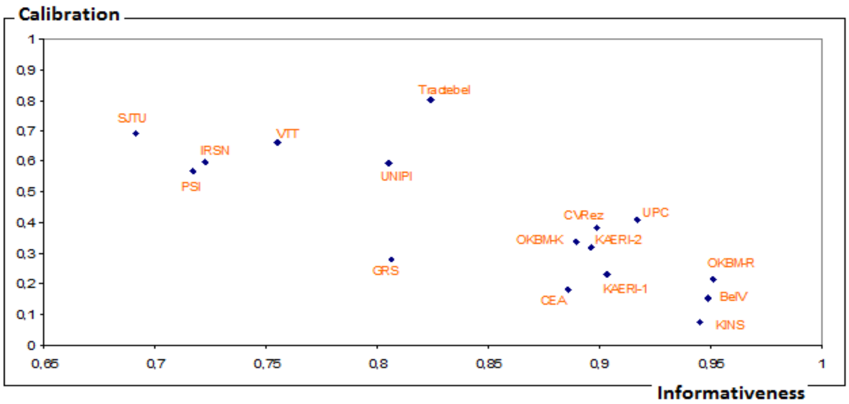
Firstly, the fraction of cases where the experimental value is enveloped by the uncertainty band was calculated and represented in Figure 3.4.11. Percentages were between 37% and 87%, with a mean value of 65%, less satisfactory than in FEBA, and far from the expected values shown in Table 3.4.1. In average, the results are better for quenching time than for cladding temperature (70% and 60%, respectively). In a number of cases the time variables did not attain upper uncertainty bounds, and the experimental value was considered to be enveloped as soon as it was larger than the lower endpoint of the provided interval (LUB).

Figure 3.4.11: PERICLES: percentage of experimental values falling inside uncertainty bands



A total number of 36 responses, for different participants, were considered in the analysis. Figure 3.4.12 shows the plot of calibration vs. informativeness scores for the participants.

Figure 3.4.12: PERICLES: Informativeness and calibration scores for each participant



The values of the two criteria are more correlated (-0.8) than in FEBA. This means that narrow uncertainty bands are not able to envelop most of the PERICLES experimental values, and, therefore, that the extrapolation of uncertainties obtained for FEBA to PERICLES seems inadequate.

The conflict estimator is 1 for all outputs indicating that participants are highly conflicting in their results.

The results are method dependent (Figure 3.4.13). The discrepancy between CIRCÉ and other methods is even more important than in FEBA. CIRCÉ produces narrow uncertainty bands encompassing few experimental values (high informativeness / low calibration). The other methods produce wide bands encompassing a large number of experimental data (low informativeness / high calibration). An analysis of variance reveals the inter-group variance to be ten times larger than the intra-group variance. The conflict indicator is 1 for all outputs in the CIRCÉ subgroup and 0.73 for the other one (Figure 3.4.14). In both cases, it is higher than in FEBA.

Figure 3.4.13: PERICLES: Informativeness and calibration scores for each participant. Green colour is for CIRCÉ users. The blue (resp. brown) solid line connects the participants using RELAP (resp. CATHARE).

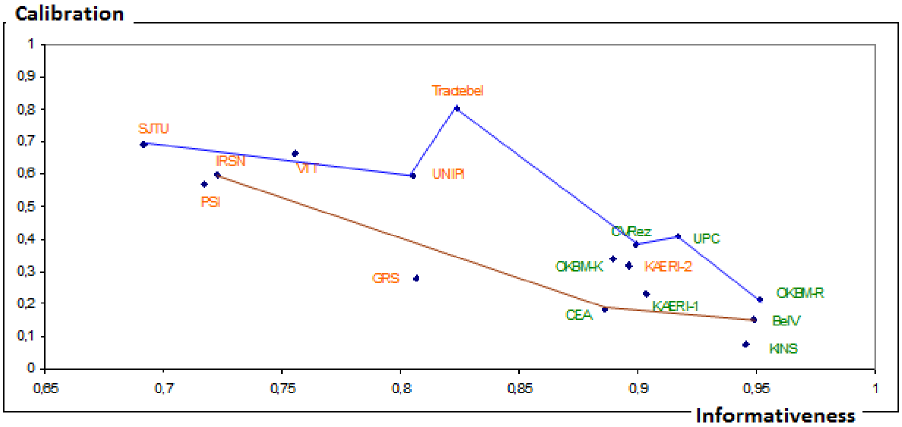


Figure 3.4.14: PERICLES: Conflict indicator with respect to the quantification method

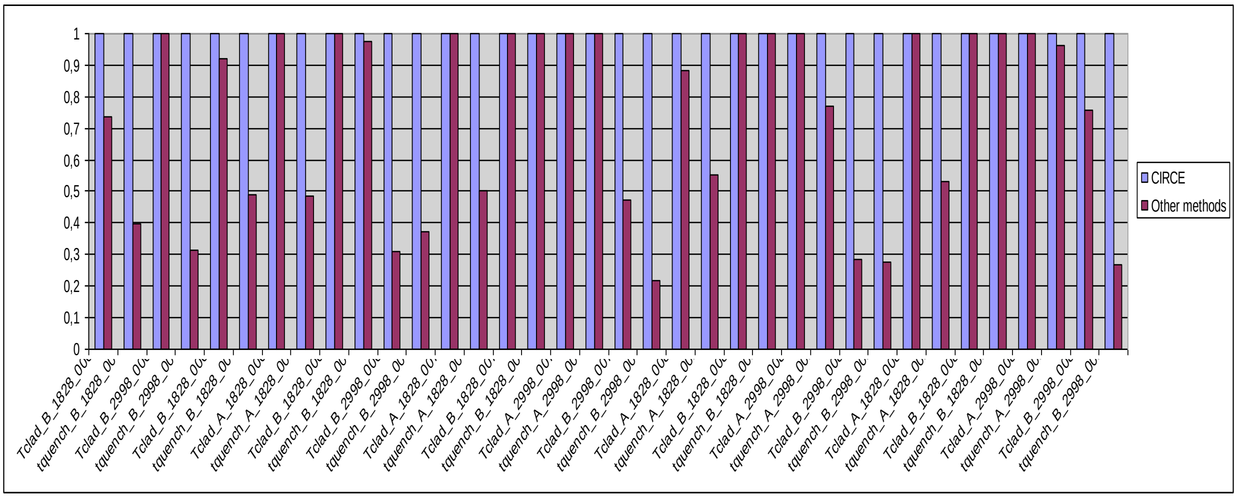


Figure 3.4.13 shows that the results are more dependent on the quantification method than on the code.

Figures 3.4.15 and 3.4.16 plot informativeness and calibration scores, distinguishing temperature and time variables.

Figure 3.4.15: PERICLES: Informativeness and calibration scores for temperature variables

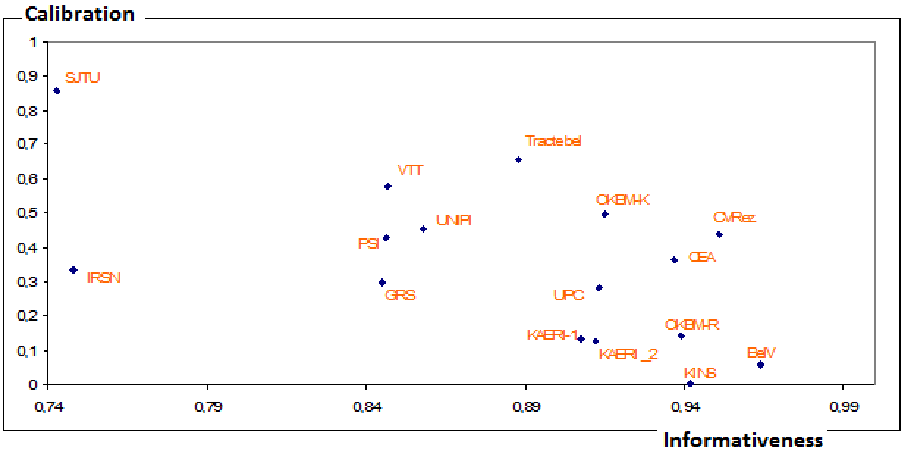


Figure 3.4.16: PERICLES: Informativeness and calibration scores for time variables

3.4.2.3 Joint conclusions for FEBA and PERICLES

When FEBA and PERICLES results are jointly considered, 4 groups can be distinguished (Table 3.4.5):

- 1st group: FEBA and PERICLES data are enveloped by the uncertainty bands.
- 2nd group: FEBA data are rather well bounded, PERICLES data not always enveloped.
- 3rd group: FEBA data enveloped, except close to the quench time or at the beginning. PERICLES experimental data not always enveloped.
- 4th group: FEBA and PERICLES experimental data are not bounded by envelop calculations.

From this ranking it is clear that having enveloping bands for FEBA is a necessary (but not sufficient) condition to have the same for PERICLES.

Table 3.4.5: Summary of the uncertainty results, by considering FEBA and PERICLES jointly

General result	Participant	Code	Method	Width of uncertainty bands (FEBA/PERICLES)
FEBA and PERICLES exp. data well bounded	IRSN	CATHARE	DIPE	very wide/very wide
	PSI	TRACE	expert judgment	very wide/very wide
	SJTU	RELAP	FFTBM	very wide/very wide
	Tractebel	RELAP	Inv. Uncertainty	wide/very wide
	UNIPI	RELAP	FFTBM	wide/ wide
	VTT	APROS	FFTBM+CIRCE	wide/very wide
FEBA roughly bounded, PERICLES not always bounded	CVRez	RELAP	CIRCE	medium/medium
	OKBM-Korsar	KORSAR	CIRCE	medium/medium
	UPC	RELAP	CIRCE	medium/medium
FEBA bounded, except close to the quench time or at the beginning, PERICLES not bounded	BelV	CATHARE	CIRCE	medium/narrow
	CEA	CATHARE	CIRCE	medium/medium
	GRS	ATHLET	Inv. Uncertainty	wide/very wide
	OKBM-Relap	RELAP	CIRCE	narrow/narrow
FEBA and PERICLES not bounded	KAERI-Circé	COBRA	CIRCE	narrow/medium
	KAERI-MCDA	COBRA	MCDA	narrow/medium
	KINS	MARS-KS	CIRCE	narrow/narrow

3.4.3 Axial pressure drops

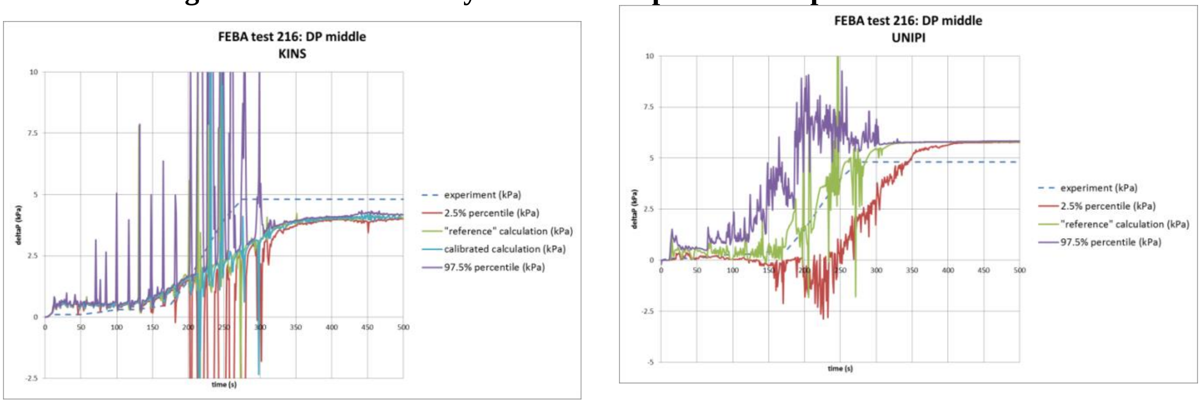
The participants in Phase IV were also requested to provide 2.5 and 97.5 percentiles for the time trends of middle pressure drops. Only three participants included these responses in the quantification of model uncertainty. Interfacial friction is the a priori most influential physical model on pressure drops; it is considered by a majority of participants. CEA considers only the interfacial friction downstream from the QF.

Two parts in the time trends of middle pressure drops are considered: the rapid increase during the passage of the QF, and the end of the transient. The prediction of the first part is strongly related to the prediction of the QF progression. One could expect that, if the quench times are enveloped, this part will be also enveloped.

The situation is different for the end of the transient. For FEBA, in most of the tests the test section is filled up with liquid at the end. For PERICLES, there is a vapour-liquid mixture, mainly controlled by interfacial friction, but only upstream from the QF.

In the FEBA case, a lot of oscillations are observed in the time trends, which may produce an artificial broadening of the uncertainty bands (Figure 3.4.17). It is possible to perform a smoothing of the curves (e.g. with a low frequency of storage of the times). The drawback is that a rigorous comparison of the uncertainty bands width is difficult.

Figure 3.4.17: Oscillatory behaviour in pressure drops calculated for FEBA



For FEBA, the majority of participants envelop experimental data during the passage of the QF, consistently with the results of cladding temperatures. At the end of the transient, the test section is filled up with liquid, so that the pressure drop is roughly the hydrostatic pressure. The participants did not reproduce well this value, probably because the pressure was calculated in different points than the measurements and because of the possible presence of vapour in the calculation. The band widths are very variable among participants (from 0.5 to 4 bars).

For PERICLES, some calculations also present very large oscillations (Figure 3.4.18). The curves are presented in a nondimensional form.

Figure 3.4.18: Oscillatory behaviour in pressure drops calculated for PERICLES

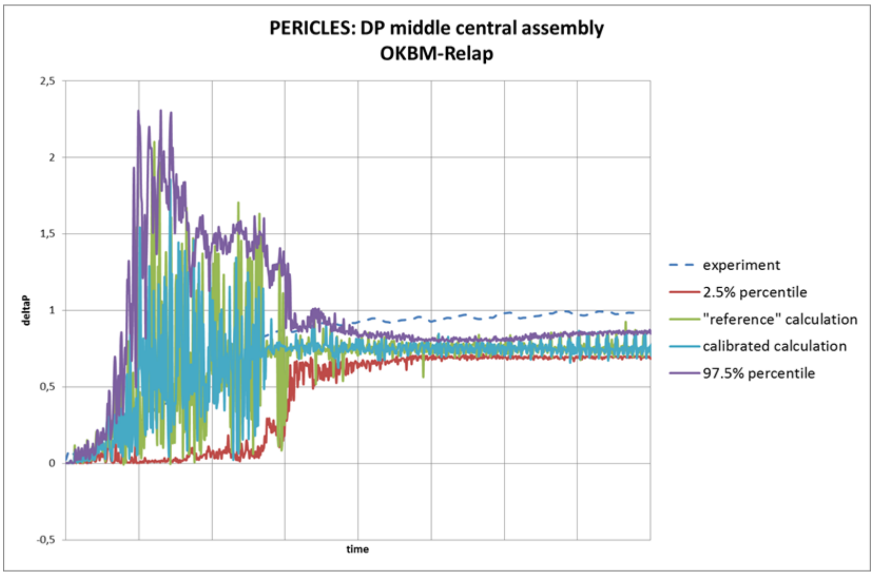


Table 3.4.6 shows the ranking of participants according to their results. Two groups are clearly distinguished: those who bound the data for all time trends, and those who generally do not bound the data. The quality of the nominal calculation is described.

The quality of the nominal calculation is indicated in Table 3.4.6. For some participants it is impossible to see the behaviour of the nominal calculation, due to strong oscillations (Figure 3.4.19). In this case, the feature of the nominal calculation is indicated as being very oscillatory.

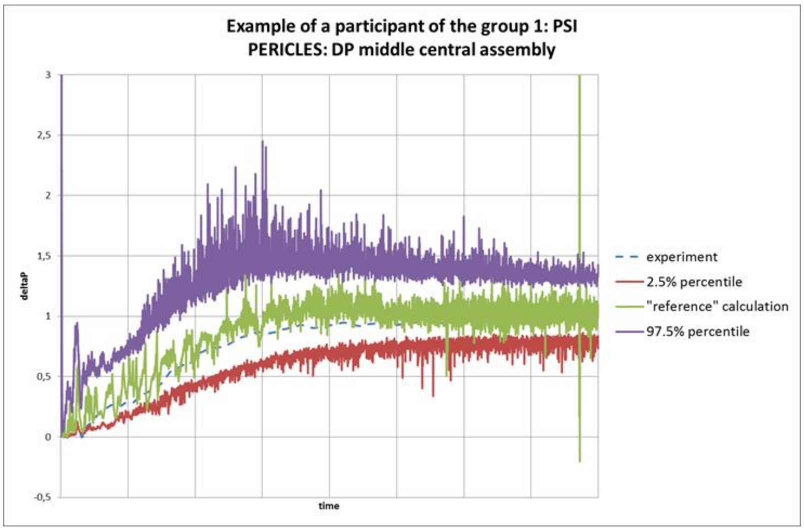
Participants who enveloped the clad temperatures were also successful with pressure drops. Those in the 2nd group had a bad nominal calculation. In some cases the discrepancies between calculation and experiment are too high. In PERICLES, there was a two-phase mixture under the QF at the end of the transient. A bad calculation of void fraction leads to a poor prediction of the pressure drop. The void fractions are controlled by interfacial friction upstream of the QF.

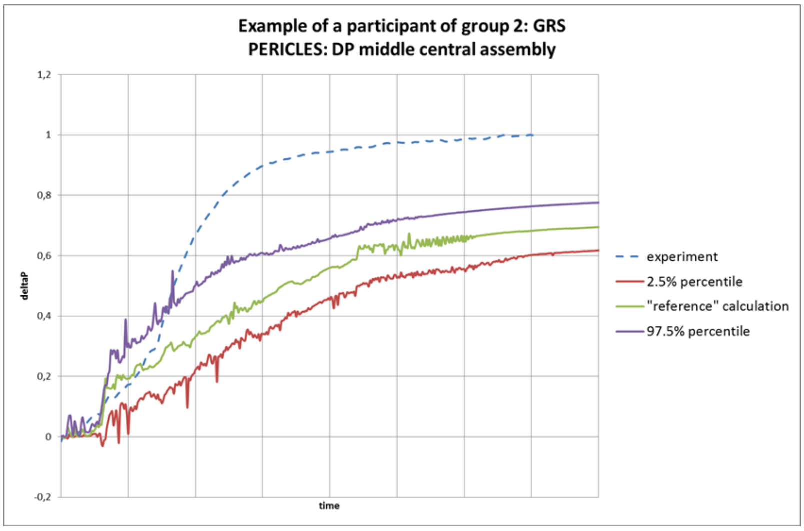
The fact of taking into account the pressure drops as responses during Phase III does not necessarily imply successful envelop calculations. The quality of the nominal calculation and the width of the uncertainty bands seem more determining.

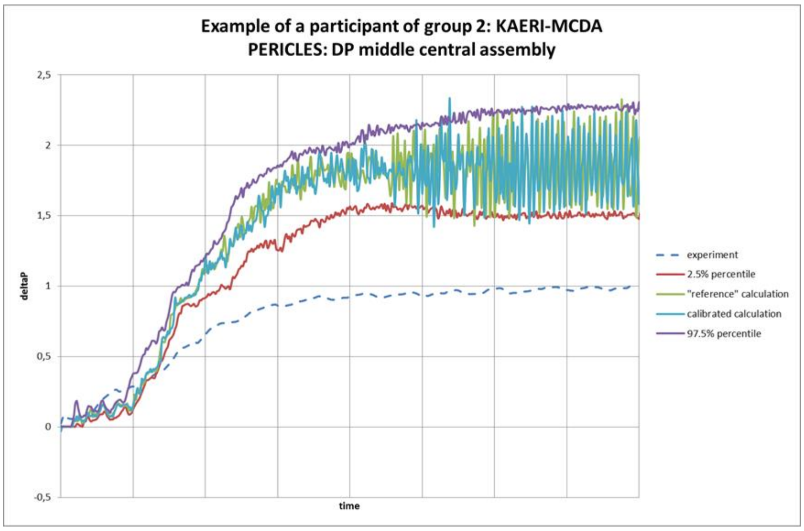
Table 3.4.6: PERICLES: Summary of the uncertainty results for the middle pressure drops.

General result	Participant	Code	Method	Features of the nominal calculation		Width of the uncertainty bands
				during the passage of the quench front	at the end	
Exp. data bounded for all the time trends	CVRez	RELAP	CIRCE	very oscillatory	very oscillatory	medium
	IRSN	CATHARE	DIPE			wide
	PSI	TRACE	Expert judgement	very oscillatory	very oscillatory	wide
	SJTU	RELAP	FFTBM			medium
	Tractebel	RELAP	Inv. uncertainty	very oscillatory	very oscillatory	medium
	UNIPI	RELAP	FFTBM	very oscillatory	very oscillatory	medium
	VTT	APROS	FFTBM+CIRCE			medium
Exp. data not bounded	BelV	CATHARE	CIRCE	underestimated	underestimated	medium
	CEA	CATHARE	CIRCE	underestimated	underestimated	narrow
	GRS	ATHLET	Inv. uncertainty	underestimated	underestimated	medium
	KAERI-MCDA	COBRA	MCDA	overestimated	overestimated	medium
	KAERI-Circé	COBRA	CIRCE	overestimated	overestimated	medium
	KINS	MARS-KS	CIRCE	overestimated	overestimated	narrow
	OKBM-KORSA	KORSAR	CIRCE	underestimated	underestimated	medium
	OKBM-RELAP	RELAP	CIRCE	very oscillatory	very oscillatory	wide
	UPC	RELAP	CIRCE	underestimated	overestimated	medium

Figure 3.4.19: PERICLES: examples of results of participants







4 ANALYSIS OF SELECTED TOPICS

4.1 Comparing methods

PREMIUM is focused on methods of quantification of model uncertainty. The benchmark has included a description of the theoretical basis of the different methods [8], and has allowed a comparison of their results in the application to the reflooding scenario.

4.1.1 Comparing the basics of the methods.

PREMIUM is a benchmark focused on the issue of the UQ of reflooding models in system thermo-hydraulic codes. The problem has been introduced in Section 2.2 of this report. Model UQ is a special case of the more general topic of parameter estimation, which is the object of several scientific disciplines, such as statistical inference and optimisation techniques. In fact, there are both statistical and non-statistical methods for quantifying model uncertainty

Methods used in quantification solve an inverse problem for the model; in general, they can also be used for the calibration of models. In the most general approach, parameter estimation produces calibration and UQ of the model, but methods have options to perform only one of the two operations.

Most methods used in PREMIUM are statistical (CIRCÉ, DIPE, MCDA,…). According to their basic statistical framework, they can be classified as

- **Classical or frequentist**: the parameters are considered as fixed but unknown constants which must be estimated from real and predicted responses. E.g. DIPE.

- **Bayesian:** parameters are considered as random variables. Experts may assign them prior probability distributions. Then, the information on real and predicted responses allows the updating of distributions, by means of Bayes' theorem, producing the posterior distributions. E.g. CIRCÉ, MCDA.

CIRCÉ and MCDA are based on a linearisation of the model (first order Taylor series), and the use of maximum likelihood principle and Bayes' theorem.

There is also a non-statistical, optimisation method involved in PREMIUM (FFTBM). It is based on a measure of discrepancy of predicted and measured responses (the so-called average amplitude, AA). The uncertainty ranges of the model parameters are defined so that they induce a small enough change in AA

4.1.2 Comparing the application of two different methodologies

Two participants in PREMIUM (KAERI and UPC) submitted a double contribution. Each one used two different quantification methods and compared the results, using two different quantification methods. These are very enlightening exercises, because the user effect is expected to be reduced in the comparison.

KAERI has used the CIRCÉ method in a contribution and its own MCDA method in the other, using the same sub-channel code (COBRA-TF module of MARS-KS1.3). CIRCÉ and MCDA are very similar methods, and, accordingly, their results have been completely similar, both for FEBA and PERICLES.

UPC has participated in PREMIUM as a CIRCÉ user. After the completion of Phase IV, UPC decided to perform an additional study, by repeating the UQ performed in Phase III with another of the available methods. In particular the CIRCÉ calculation of FEBA has been repeated with FFTBM.

The sole difference between the application to FEBA case of CIRCÉ and FFTBM methods is that, for CIRCÉ, data from all the 6 tests of FEBA Series I were considered as input, while only the reference test 216 has been considered as input data for FFTBM.

Other aspects, such as response choice and parameter choice, have been preserved as much as possible, i.e. the same types of magnitudes and same heights were considered and the parameters quantified also were the same.

The probability distributions obtained as a result of these two exercises are shown in Table 4.1.1.

Table 4.1.1: Comparing CIRCÉ and FFTBM – Uncertainty of parameters calculated by UPC

	P1: Film boiling heat transfer coefficient: wall-to-liquid		P4: Interfacial friction coefficient: bubbles and droplets		P6: Interphase heat transfer coefficient: global		
Method	Min	Max	Min	Max	Min	Max	Range type
Expert judgement	0.4	2	0.5	1.5	0.1	10	estimated range
CIRCE	0.75	1.37	0.87	1.37	0.29	2.07	95% log-normal law
FFTBM	0.59	1.26	0.82	1.67	0.19	1.60	uniform law

The representation of the different ranges and probability distribution functions in relation with the engineering judgement range for each one of the three considered parameters is shown in Figures 4.1.1 to 4.1.3.

Figure 4.1.1: Probability distribution for film boiling HTC wall-to-liquid

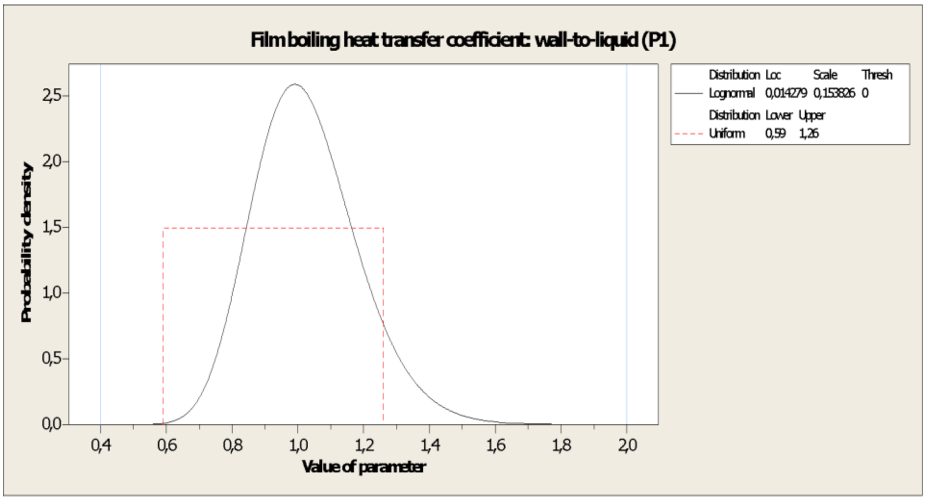


Figure 4.1.2: Probability distribution for interfacial friction coefficient

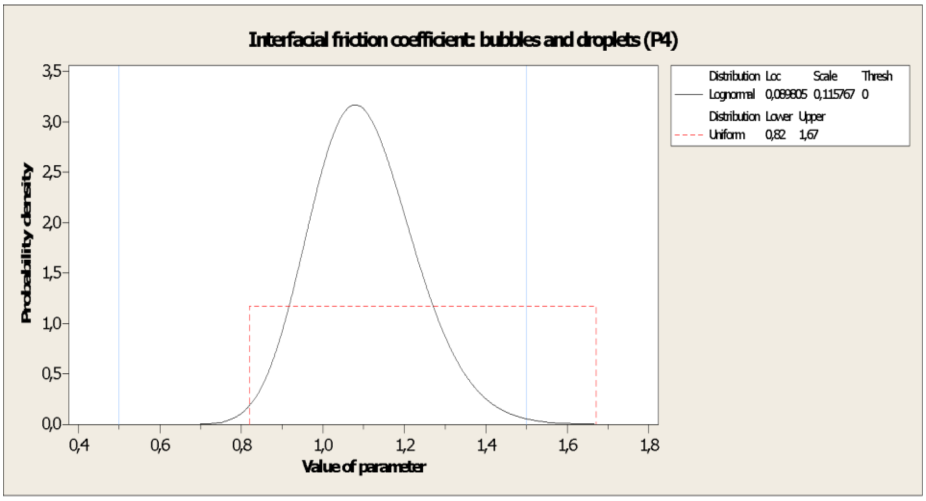
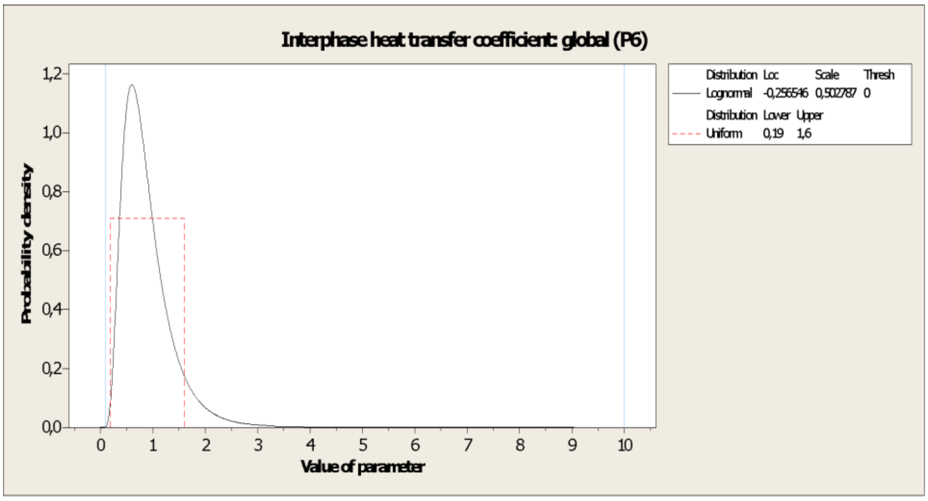


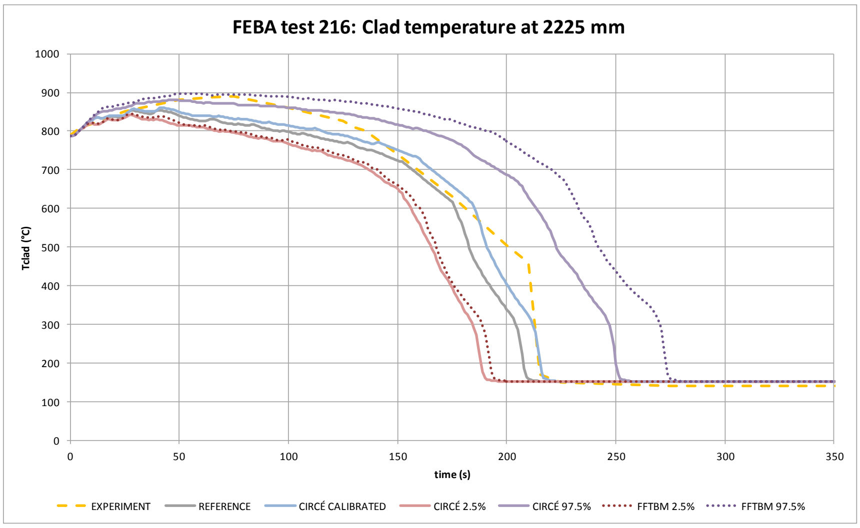
Figure 4.1.3: Probability distribution for interfacial HTC



From these figures, a similarity between the results of CIRCÉ and FFTBM can be clearly observed for all the parameters, and the ranges found by these two methods are in all cases narrower that the engineering judgement approach.

Figure 4.1.4 represents the envelop results compared for FEBA test 216. The magnitude represented is the cladding temperature at height 2225 mm (measured by the thermocouple which is closest to the middle of the bundle).

Figure 4.1.4: Comparing CIRCÉ and FFTBM results for FEBA 216 clad temperatures



The bands obtained with CIRCÉ are narrower than those obtained with FFTBM, especially because the upper band is farther away from the reference calculation (actually the lower band is very close to the one obtained with CIRCÉ). This was expected due to the fact that, even if the ranges obtained from the two methodologies are similar, truncated log-normal laws give less probability to extreme values in the range, compared with uniform laws (with the same range) obtained with FFTBM. Another possible explanation is the higher upper range of interfacial friction in the case of FFTBM.

The results were, as performed during Phase IV, extrapolated to PERICLES facility. In this case, a similar observation can be made. FFTBM bands are broader and give higher values, enveloping better the experimental results, but with less accuracy (fig 4.1.5 and 4.1.6)

Figure 4.1.5: Comparing CIRCÉ and FFTBM results for PERICLES RE0064 clad temperatures (1828 mm, lateral assemblies)

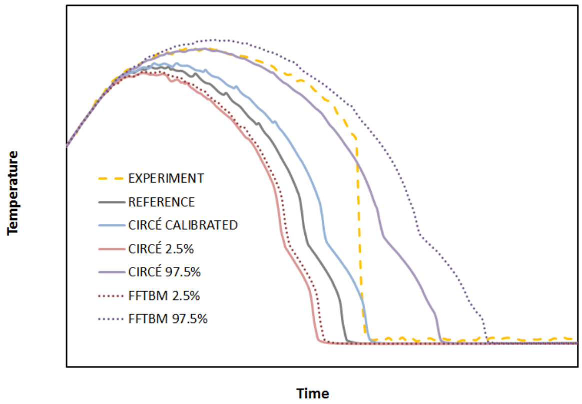
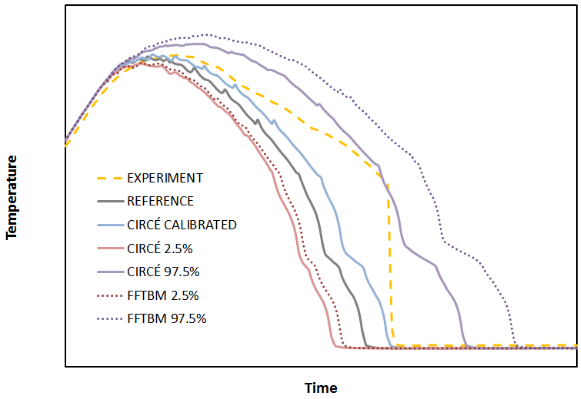


Figure 4.1.6: Comparing CIRCÉ and FFTBM results for PERICLES RE0064 clad temperatures (1828 mm, central assembly)



The analysis of the results of the study is in good agreement with conclusions of Phase IV. The CIRCÉ method results in medium or narrow bands in the enveloping calculation, which are more informative but do not encompass all the experimental measurements, while FFTBM method results in wide or very wide bands which are less informative but do envelop more experimental results.

4.2 User effect

PREMIUM has shown that the quantification methods of model parameter uncertainty have a significant user effect. Their application needs the use of engineering judgement, in addition to the well-known user effect issue when using system codes.

The existence of user effect can be detected in the different stages of the application of a quantification method. The following are important features of the quantification which depend on the judgement and experience of the user:

1) Choice of the responses (outputs) on which the quantification is based
2) Choice of the model parameters to be quantified
3) Selection of the database of responses (experimental measurements) used in the quantification
4) Choice of IP in the modelling of experiments. This includes the selection of specific process models, coefficients and nasalisation approaches

4.2.1 User effect in the selection of model parameters, responses and quantification database.

The selection of model parameters to be quantified is very important. The model quantification can be viewed as a backpropagation of uncertainties, from the responses (outputs) to a specific set of inputs. In a sense, it may be said that the back-propagated uncertainty is apportioned among the selected inputs. The uncertainty quantified for a given parameter depends, in general, on the set of selected parameters, and should not be considered as an “intrinsic” property of the parameter. The IP should be mutually independent. The “grain” of the parameters is also a difficult point: one may deal, for instance, with a global HTC or with all the coefficients involved in all the sub-models being used in the flow map.

The choice of responses to be used in the quantification is also important. The very low calibration score of some participants shown in Figure 3.4.7 (Section 3.4.1.2) is due to the fact that they did not consider as responses the quench times in their quantification process.

Responses must be clearly dependent on the selected IP. But application of inverse methods may require the use of responses that are not mutually dependent. E.g. the user guidelines of CIRCÉ state that the selected responses must be as independent as possible [21]. Two dependent responses are considered by the method as only one response with a double weight.

On the other hand, responses should be used so that the effects of various parameters could be discriminated (e.g. not selecting the pressure drop is an error, as one can no longer distinguish the effects of interfacial friction and HTC).

Selection of the data base for quantification may have a large influence on the results. The uncertainty obtained for model parameters depends on the selected database.

Experimental databases are made up of separate effects tests, integral effects tests and the so-called intermediate tests (IT). Data from SET are useful for the quantification of “simple” models, e.g. including a single model parameter. For more complex models, effects are difficult to separate, and IT must be used in quantification. Quantification methods (e.g. those used in PREMIUM) are intended for the use of IT data. Some model parameters may be quantified with SET data; others must be quantified with IT data.

An interesting point is how specific or general should be the quantification database. Every physical model has a range of application, meaning a region of the input space where the model is deemed to produce adequate results. The model is developed, validated and quantified inside its range, and should not be applied outside it.

It is clear that the process of development, validation and quantification of a model must be somehow guided by the foreseen application of the model. But whereas model development and validation are always dealing with the full range of foreseen applications (physical models implemented in codes have as large applicability as possible, meaning that they can be applied to a wide variety of scenarios), the quantification of the model should be focused on the region of the input space where the model will work during the specific application. This means that data from experiments resembling the desired scenario should be privileged in the quantification database.

On the other hand, it is desirable that physical models implemented in codes have as large applicability as possible, meaning that they can be applied to a wide variety of scenarios. A compromise must be found, so that the selected database is guided by the intended application without losing generality. The decision on this trade-off will be based on engineering judgement, and the quantified uncertainty will depend on the selected database.

Deciding whether the UQ should be determined on the full database or on a specific subset, and choosing the level of specificity, is not an easy question, and usually involves a compromise between the cost and the desired sharpness of the UQ process. In some cases a larger database may enlarge the physical domain of the model, and produce wider uncertainty bounds. In other cases, the larger database will only have the effect of increasing the sample size, thus reducing the statistical uncertainty and producing narrower uncertainty bands.

4.2.2 User effect in the calculation with the codes.

One of the most relevant findings of BEMUSE was the importance of the user effect in the application of BEPU methodologies. Within the PREMIUM project several participants performed a similar work using different codes and methodologies. Therefore, it is interesting for this synthesis to draw some conclusions about both user and code effect. Table 4.2.1 lists all participants of the PREMIUM project along with the codes and methodologies employed. It is similar to Table 3.1.1, but is more explicit on the code versions used.

This section is devoted to analysing user effect in the main results of the PREMIUM project.

Table 4.2.1: List of participants, codes and methods

User	Country	Code	Method	II	III	IV
BelV	Belgium	CATHARE2 V25_2 mod8.1	CIRCÉ	yes	yes	yes
CEA	France	CATHARE2 V25_2 mod8.1	CIRCÉ	yes	yes	yes
CVRez	Czech Republic	RELAP5 mod3.3	CIRCÉ		yes	yes
GRS	Germany	ATHLET 2.2B	Own method	yes	yes	yes
IRSN	France	CATHARE2 V25_2 mod8.1	DIPE	yes	yes	yes
KAERI-1	Korea	MARS-KS1.3-COBRA-TF	CIRCÉ	yes	yes	yes
KAERI-2	Korea	MARS-KS1.3-COBRA-TF	MCDA	yes	yes	yes
KINS	Korea	MARS-KS-0003 PREMIUM version	CIRCÉ	yes	yes	yes
KIT	Germany	TRACE Version 5 patch3	FFTBM	yes	yes	*
NRI	Czech Republic	ATHLET 2.1A	-	yes		
OKBM-1	Russian Federation	KORSAR/BR	CIRCÉ		yes	yes
OKBM-2	Russian Federation	RELAP/SCDAPSIM/mod3.4	CIRCÉ	yes	yes	yes
PSI	Switzerland	TRACE V5.0P3-UQ	Own method			yes
SJTU	China	RELAP5/SCADPSIM/mod3.4	FFTBM		yes	yes
TRACTEBEL	Belgium	RELAP5 mod3.3	IUQ	yes	yes	yes
UNIPI	Italy	RELAP5 mod3.3 patch 3	FFTBM	yes	yes	yes
UPC	Spain	RELAP5 mod3.3 patch 4	CIRCÉ	yes	yes	yes
VTT	Finland	APROS 5.11.2	CIRCÉ(bias)+FFTBM(range)	yes	yes	yes

* KIT had an incomplete participation in Phase IV (as described in Section 3.4).

FEBA calculations

The first step for the participants of PREMIUM benchmark was to build input data for Test 216 performed at FEBA reflood facility. Test 216 has been selected as FEBA test with conditions mostly similar to

PERICLES experiment. This test belongs to FEBA experiment Series I ("Base line tests with undisturbed bundle geometry with 7 grid spacers"), the only series taken into account.

Most of participants adopted a nodalization approach representing the test section of FEBA with a single vertical channel and a single heater rod/heat structure. A specific CHAN component of TRACE code was used by KIT, which actually simulates a 5x5 bundle. KAERI, in its turn, modelled 1/8 of the bundle with a sub-channel COBRA-TF module of MARS-KS code.

Different approaches were adopted by participants for modelling the spacer grids: some organisations actually reduced the flow area at the location of the grids and activated special models for heat transfer enhancement; others took into account the grids only by applying form loss coefficients at the corresponding elevations.

The number of axial nodes in the different nodalizations, representing the test section, ranges from 20 to 78 (Table 3.2.2 and Figure 4.2.1). It should be mentioned that the provided number of axial nodes does not take into account the possible refinement, as it can be the case in the vicinity of the QF to calculate the axial conduction (whenever performed by a code).

Figure 4.2.1: number of nodes adopted by the participants

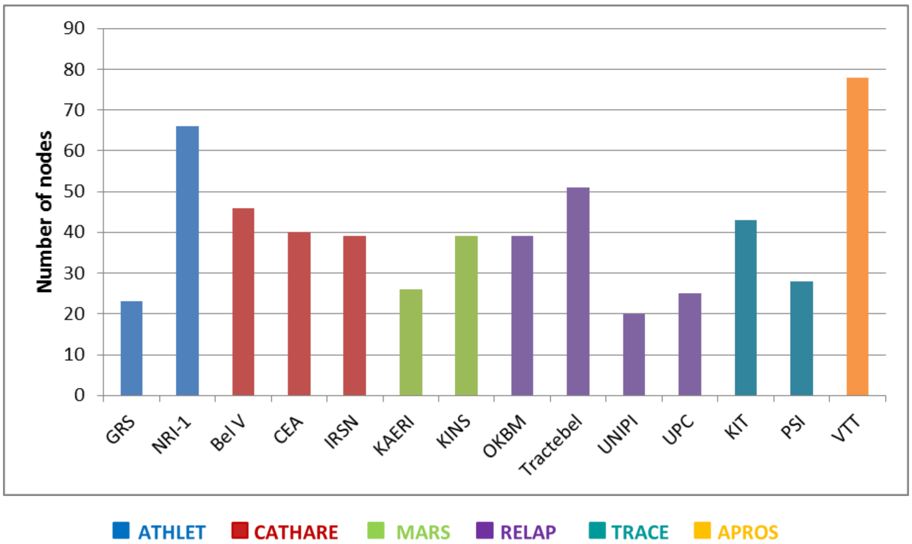


Figure 3.2.5 represents the scalar values found by all the participants of base case PCT grouped by code, the black dashed line shows the experimental measurement. The results show that there is still a significant user effect. On the other hand, code effect is minimal or might be shadowed by the user effect. The same can be observed for the quench time scalar values (Figure 3.2.6) [9].

Figure 3.2.3 displays the time trends of all participants of the magnitude of interest "Cladding temperature at Top of Active Fuel (TAF)" of the base case. This figure confirms that there is a strong user effect and that no conclusions can be drawn for the code effect. For instance, one can consider the CATHARE2 results, it can be observed that the results obtained by CEA and IRSN present very similar time trends, however it is visible that the BelV calculation is considerably far away from those two.

The other large group of participants is the RELAP5 users, here again one can see significant differences. In this case, TRACTEBEL results are different from UNIPI and UPC results, which are quite close but also show different oscillations.

A very general conclusion can be drawn observing this behaviour. Thermal-hydraulic codes are very complex tools that can give different answers depending on how the question is postulated. Even a quite simple and unambiguous problem of a single bundle in a reflood situation has enough degrees of freedom to show a considerable spread of results, depending on how the model is built by the user.

PERICLES calculations

In Phase IV a base case model for the PERICLES experiment facility was built. The simulation of the base case was carried out blindly, although the participants had information of boil-off experiment in order to qualify the created input data.

The specificity of the 2D reflood PERICLES experiment with respect to FEBA is the presence of 2-D effects. As shown in Table 3.4.4, the modelling of these 2-D effects is different according to the code[1] used and also depending on the participant choices. In the most general case, a multi-channel modelling with crossflows is chosen, among others by the RELAP users. Depending on the participants, two or three channels are modelled. A 3-D modelling is chosen by 4 participants, their code offering this possibility. The case of IRSN with CATHARE2 is apart since this participant uses a 1-D modelling whereas CATHARE2 has a 3-D module: IRSN chooses this option to have a similar modelling for both FEBA and PERICLES, so that, according to them, the uncertainties found with FEBA can be used for PERICLES.

The number of axial meshes in the heated part of the bundle is generally close to the corresponding number of meshes in FEBA, except for the users of a 3-D model and, among the users of a multi-channel modelling. The users of a 3-D model, BelV, CEA and KAERI consider a significantly lower number of axial meshes for PERICLES than for FEBA (e.g. for BelV: 46 axial meshes for FEBA and 11 axial meshes for PERICLES, i.e. the number of levels of the axial power profile), whereas for PSI, it is the opposite: 28 axial meshes for FEBA and 65 for PERICLES. TRACTEBEL considers too a lower number of axial meshes for PERICLES (15) than for FEBA (51).

4.2.3 User and code effect in the model uncertainty quantification

As described in the previous section, there are sources of user effect in methods of model parameter UQ. In fact, results of Phases III and IV of PREMIUM reveal that the quantified uncertainty of the model parameters, and the uncertainty propagated to the results of FEBA and PERICLES depend primarily on the quantification method, rather than on the system code used. In a sense, this is an astonishing result, because the physical models have a specific implementation in each system code.

The quantified uncertainties in Phase III were propagated, during Phase IV, to the 6 Tests of FEBA Series I, and other 6 Tests of PERICLES facility in order to confirm and validate the results against experimental data.

The dependence on quantification method is thoroughly discussed in Section 3.4 of this report. The main findings may be summarised as follows:

- The width of propagated uncertainty bands depend basically on the quantification methods. For CIRCÉ results, bands are narrow or medium. For FFTBM and other methods, bands are wide or very wide. In general, the wider the band, the higher the ability of bounding real values.

1. KIT, which used TRACE for FEBA, did not participate in Phase IV, and consequently is not quoted.

- The quality of the base case calculation is important; the closer the calculation to the real data is, the higher the ability of bounding real bands by the uncertainty bands will be.

As described in Section 3.4 and shown in Table 3.4.2, the results of the uncertainty propagation to FEBA depended on the method rather than on the code. Table 3.4.2 distinguishes four groups of participants according to the quality of the bounding of the experimental data.

One can note that there are RELAP users in the first three groups, in the same way there are CATHARE users in the 1st and the 2nd group and both TRACE users belong to the 1st group and the 3rd group. The 4th group is apart since the corresponding participants, KAERI and KINS are alone to use their code.

A similar observation can be made from PERICLES results. IRSN analysis points out that RELAP and CATHARE users include participants with both low and high informativeness and calibration scores (see blue and brown lines respectively in Figure 3.4.12).

Finally, one can quote also a user effect for the quantification of the uncertainties. For example, CEA and BelV apply the same method of UQ, CIRCÉ, to the same version of CATHARE2 and their envelop calculations are significantly different. It is due firstly to a different nominal calculation, coming from a different input data deck, but also to a different choice of the IP and, to a less extent a different selection of experimental responses used for quantification. The same observation can be made for CVRez, OKBM-RELAP and UPC, all of them being RELAP and CIRCÉ users. FFTBM users are also concerned: SJTU and UNIPI, which are FFTBM users, use both of them RELAP and have different uncertainty bands.

4.3 Other effects: Nominal calculation versus calibrated calculation in CIRCÉ

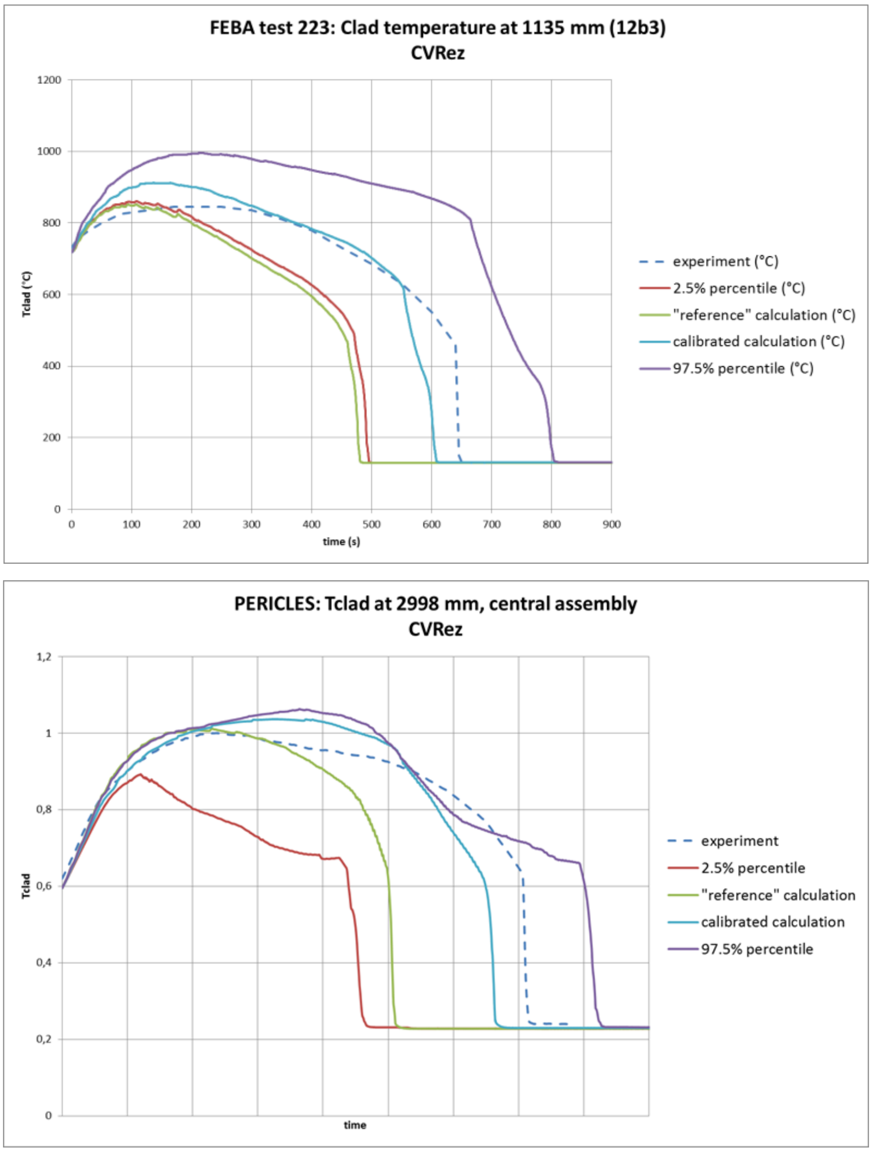
In Phase IV [11], a majority of CIRCÉ users obtained envelop calculations which were globally satisfactory for FEBA, but not for PERICLES. Focusing on the behaviour of the nominal calculation, it is observed that most of participants predicted a too rapid QF progression for FEBA. On the contrary, a majority of participants predicted a too slow QF progression for PERICLES.

CIRCÉ users estimate median values for their IP, different from the nominal values. The median values produce what is termed “calibrated calculation”. Indeed, this corresponds to a “recalibration” of the system code. Theoretically, this action corrects systematic deviations of the nominal calculation with respect to the experiment used for the quantification. This effect can be especially observed on the prediction of the QF progression.

In PREMIUM, the calibrated value of the IP, found for FEBA, is applied to PERICLES. But the prediction of the QF progression is often different for FEBA and PERICLES. In fact, the calibrated calculation may degrade the prediction of the QF for PERICLES.

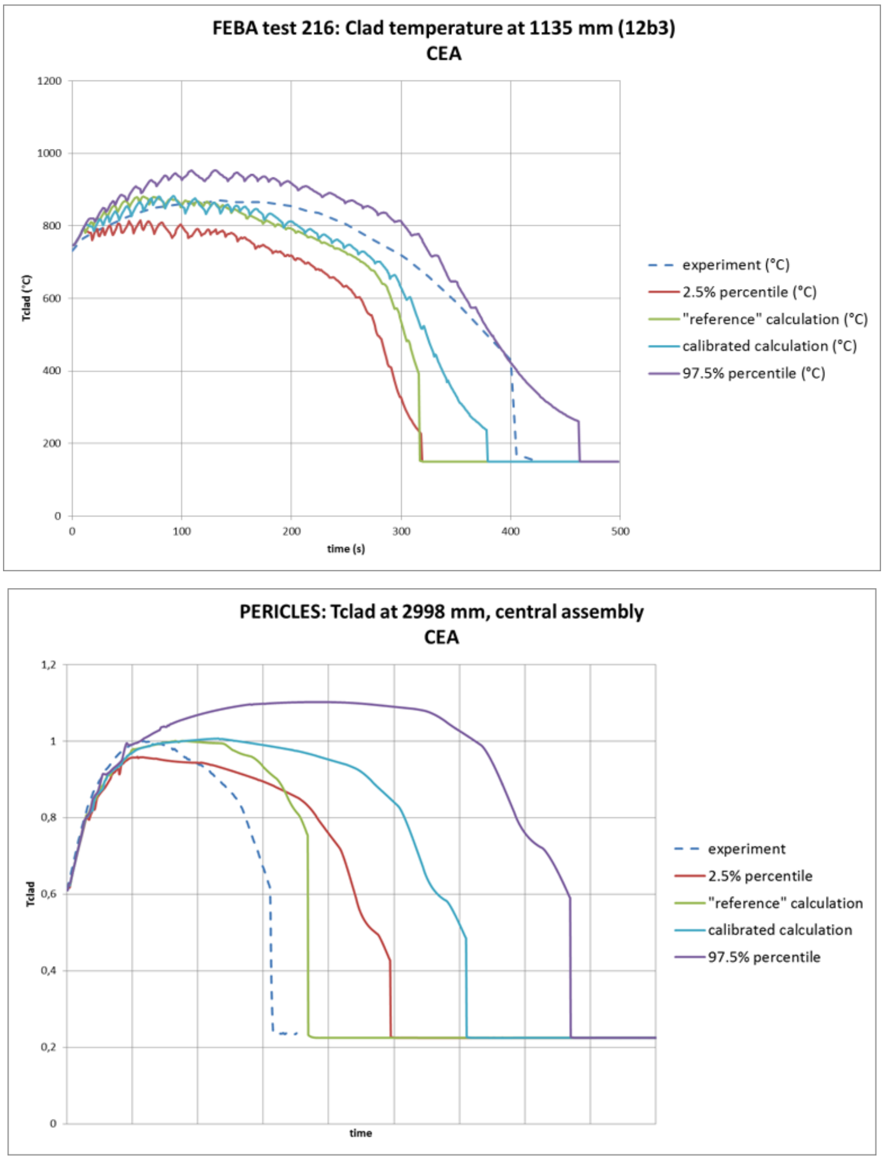
An example of improvement of recalibrated calculation with respect to nominal calculation, for FEBA and PERICLES, is shown in Figure 4.3.1.

Figure 4.3.1: Example of a favourable case for the calibrated calculation for PERICLES compared to FEBA



On the other hand, in Figure 4.3.2 a case is shown where the calibrated calculation predicts a later quench than the nominal one. The effect is that the calibrated calculation improves the nominal one for FEBA but degrades it for PERICLES when extrapolating results from one to the other.

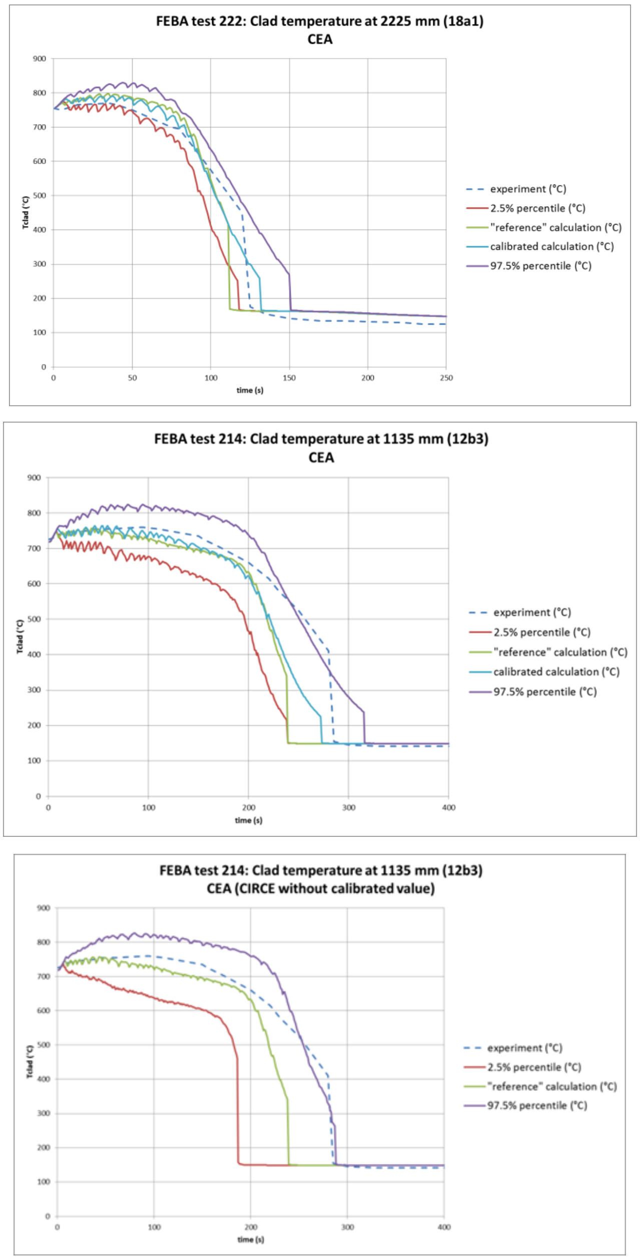
Figure 4.3.2: Example of an unfavourable case for the calibrated calculation for PERICLES compared with FEBA.



For participants of the third group the recalibration degrades the nominal calculation. This fact, associated with the medium/narrow condition of the uncertainty bands, explains why the envelop calculations of these participants are not successful.

For the reasons outlined, CEA recommends the suppression of the recalibrated value estimation, especially when the database used for recalibration is small and does not cover the physical conditions in which the model will be used. In fact, CEA has repeated the quantification of the physical models with the FEBA tests, with the same parameters and responses, but without estimation of a calibrated value. Once estimated the uncertainty for the IP, the propagation has been repeated, firstly for FEBA. Figure 4.3.3 shows the comparison between two clad temperatures, calculated with and without estimation of the calibrated value.

Figure 4.3.3: Comparison of cladding temperatures for FEBA, with and without estimation of recalibrated value

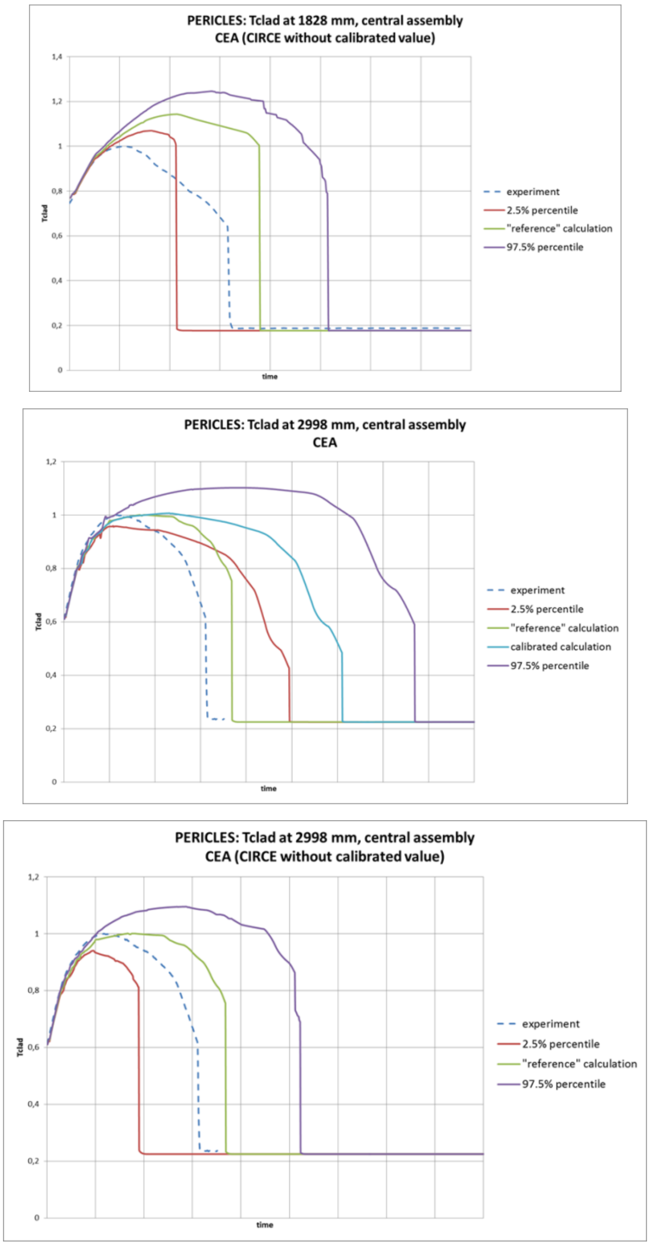


It is apparent that the resulting uncertainty bands are a little bit wider but are still reasonable when the recalibrated value is suppressed. The bands envelop the real values (except at the very end of transient at 1135 mm).

The input uncertainties have been also propagated for PERICLES (Figure 4.3.4). It has been observed that the PERICLES uncertainty bands improve by far, and become a little wider, without recalibration.

Another participant (OKBM) did not perform the recalibrated calculation, and obtained good envelop calculations for PERICLES, better than those of the majority of the CIRCÉ users.

Figure 4.3.4: Comparison of cladding temperatures for PERICLES, with and without estimation of calibrated value



4.4 Extrapolating results

One of the main questions raised in connection with the PREMIUM activity has been the capacity of extrapolating the quantified uncertainties of IP. The quantification methods may use a specific experimental database in the process, but the same parameters can be involved in other tests, facilities or scenarios. The level of applicability of quantified uncertainty to other cases must then be addressed.

From FEBA to PERICLES

PREMIUM Phase IV has been an attempt to extrapolate results found in the quantification performed in Phase III.

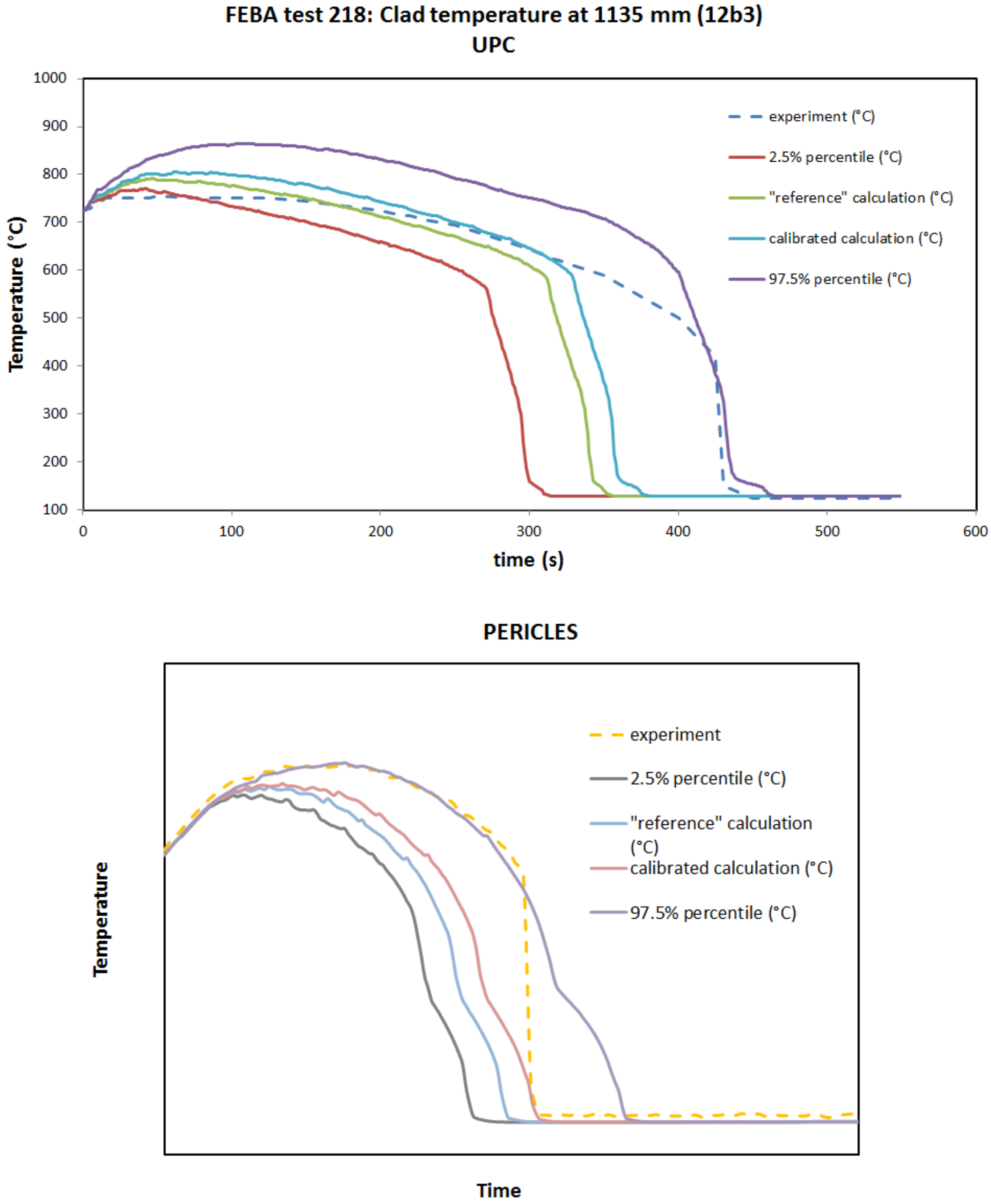
This quantification used the data from a very simple facility, FEBA, with only one bundle subject to a reflood situation. The validation step of the next phase proposed to apply the uncertainties found with FEBA to PERICLES, which is a slightly more complex facility. It was very similar to FEBA but had 3 bundles with different power rate (therefore the ability to experiment with 2-D effects in the core). Therefore, the validation exercise was in fact also a validation of the extrapolation of quantified model uncertainties.

Phase IV results reveal that even if the facilities and studied tests are similar, significant differences in conditions can make the extrapolation of the quantified model uncertainties very questionable.

The first obvious observation is that more participants fail to envelop the experimental data for PERICLES than for FEBA. It may be partly due to the blind feature of the part of the benchmark devoted to PERICLES. But other reasons may explain this fact. Aside from this, the trends already observed for FEBA can be checked. See Figure 4.4.1 which represents UPC participation.

The results of the uncertainty analysis of PERICLES are less satisfactory than for FEBA, with a fraction of experimental values falling inside the uncertainty band very far from the values in Table 3.4.2. One reason to explain this low capability of the used methodology for extrapolation to PERICLES is the lack of representativity of the FEBA database. Indeed, the thermo-hydraulic conditions of PERICLES and FEBA are different. Even if the pressure, the velocity of the injected water and the generated power of FEBA and PERICLES are rather similar, it is not the case of the initial clad temperatures just before the start of the reflood, which in higher in FEBA than in PERICLES tests. More energy has to be released in the FEBA tests, resulting in longer transients. Another difference between FEBA and PERICLES is the temperature of the injected water, which is lower for FEBA tests. The influence of the thermo-hydraulic conditions on the results can be already observed for the high pressure PERICLES test. Another difference is that PERICLES is at higher scale than FEBA with the presence of 2-D effects in the form of crossflows. This particularity has been pointed out by some RELAP users (OKBM-RELAP, UNIPI and UPC). RELAP users may choose between two different wall-to-fluid heat transfer correlations depending on whether crossflows are considered or not. Thus the correlation used for FEBA and PERICLES may be different.

Figure 4.4.1: Extrapolation of the uncertainties found in FEBA to PERICLES test for UPC participation



General remarks about extrapolation of results

As discussed in previous sections of this report, the uncertainties that quantification methods assign to model parameters depend on a number of features including user and code effects, and assessment base for quantification of model input uncertainty. An important conclusion is that the quantification depends on the set of parameters analysed. The uncertainty, back-propagated from the real and predicted responses, is apportioned between the quantified parameters taking into account the information provided by the quantification database. So, the quantified uncertainty cannot be considered as an intrinsic property of the parameter. This fact implies that a direct application of the quantified uncertainties to different conditions that are not covered by the quantification database may produce completely misleading results.

As stated in [19], the adjustable parameters used for the calibration of a model generally are not physically measurable magnitudes, and have little or no physical meaning outside of the model. In this setting, the confidence in extrapolating the model decreases significantly.

4.5 Topics on computation

4.5.1 Computational effort

Table 4.5.1 shows estimates of the CPU time spent by the participants in the completion of Phases II, III and IV.

These estimates have been provided by the participants to the Phase V co-ordinators [20]. Some participants decided to provide the CPU time spent per code run, rather than total time estimates. Some provided very detailed estimates; others made very rough ones.

On the other hand some participants provided additional information:

- computer used;
- time spent in other tasks: preparation of input decks, writing of scripts, pre and post-processing of data, etc.

Several participants remarked the fact that computational time was not an issue for PREMIUM, being negligible in comparison with the time of “human work”, devoted to the analysis of results, study and development of methods, construction of input decks, etc.

Table 4.5.1: CPU time spent by the participants in Phases II, III and IV

Participant	Computer used	CPU time		
		Phase II	Phase III	Phase IV
GRS	_	4 hr (including input deck development)	20 hr	100 hr (FEBA) 180 hr (PERICLES)
SJTU	Laptop with 2 cores (Ph II and III) Desktop with 8 cores (Ph IV)	8 hr 20 min	8 hr 15 min	366 hr
CEA	_	1 min per FEBA CATHARE run 3 min per PERICLES CATHARE run FEBA envelop calculation: one morning or an afternoon PERICLES envelop calculation: one day		
CVrez	-	90% of CPU time spent in Phases II and III 10% of CPU time spent in Phase IV		
OKBM	-	2 hr (RELAP/SCDAPSIM/MOD3.4) 5 hr (KORSAR/BR)	8 hr (RELAP/SCDAPSIM/MOD3.4) 18 hr (KORSAR/BR)	67 hr (RELAP/SCDAPSIM/MOD3.4) 150 hr (KORSAR/BR)
KINS	PC with Intel Core i7 CPU, 2.93 GHz	11191 sec	12494 sec	671100 sec
KAERI	Intel Core™ i5, based on single CPU. Operating syst: 32 bit Window-7	40600 sec	1034160 sec	18627000 sec
TRACTEBEL		10-40 hr	10-40 hr	10-40 hr
Bel V		14 sec per CATHARE run	15-20 sec per CATHARE run Few minutes per CIRCÉ run 20-130 min per 200 runs of FEBA 90 min per 153 runs of FEBA 216	60-100 min per 200 runs of FEBA 280-330 min per 200 runs of PERICLES
IRSN		1 hr	40 hr	24 hr
PSI		53 hr: 7 hr for reference model development & verification + 46 hr for selection of uncertainties	105 hr	2064 hr: 105 hr for FEBA uncertainty quantification + 215 hr for PERICLES Model Development + 1744 hr for PERICLES Uncertainty Quantification
UPC				
Univ. of PISA				

4.5.2 Treatment of failed runs

In Phase IV, the propagation of uncertainty is based on the extraction of a simple random sample from the uncertain inputs, and the running of the code for such input samples. In the process, code runs may fail. Indeed, only 2 participants had failed runs. They simply replaced the failed code runs by new runs, in order to complete the prefixed total sample size.

Some controversy may arouse about this procedure. The propagation of uncertainties in Phase IV is made by a crude Monte Carlo procedure [39], based on simple random sampling (SRS) of the uncertain inputs to the calculations. If failed runs are replaced by new, successful ones, do we still have a simple random sample of outputs?

In some cases, the run failure could be due to the performance of the code outside the validity range of their models. In such case, the uncertainty ranges of the IP, as well as the dependencies among them, should be revised.

Some code failures are due to the fact that conditions would produce very extreme values of the response. For instance, in reflooding experiments, some conditions may produce very high clad temperatures and the code failure. It is argued that the failure of the code is an evidence of a very extreme response. Following this idea, a procedure could be to keep the failed run in the sample and estimate the value of the response. This means a sort of extrapolation in time of the response.

5 CONCLUSIONS, LESSONS LEARNT AND SUGGESTIONS FOR FUTURE WORK.

5.1 Conclusions

PREMIUM is a benchmark devoted to methods of UQ of physical model parameters, and their application to the models involved in reflooding scenarios simulation.

A number of organisations have participated in PREMIUM benchmark. A majority of them were involved in the three computational phases (II, III and IV). Some organisations were partially involved, because skipped one or more phases. Some participants provided more than one contribution, using different TH codes or different quantification methodologies.

<u>PHASE I</u>

Inverse methods of UQ have been used in PREMIUM.

Two methods (CIRCÉ and FFTBM) were offered to the participants. Some participants used their own methods. All methods (except for FFTBM) can be considered as statistical, using either a Bayesian or a frequentist framework.

All the quantification methods have been adequately presented and documented.

<u>PHASE II</u>

Participants in Phase II identified influential code IP, from the point of view of reflooding, and made a preliminary quantification of their variation range. The identification was performed on the basis of sensitivity analyses of experimental test 216 of FEBA facility.

The co-ordinator of Phase II proposed a methodology for the identification of influential parameters, based on a set of quantitative criteria, and presented a preliminary list of possibly influential parameters on reflooding. Some participants modified the proposed criteria, or established their own criteria. The use of this type of methodologies at this phase of the uncertainty analysis reduces the use of engineering judgement, although does not eliminate it completely, and provides a structured approach to further optimisation of the computation and analytical resources.

Base case results (cladding temperatures and QF propagation) showed dispersion among the participants. Most of them predicted a too fast quench progression. Temperature time trends showed oscillatory behaviour which may have numerical origin.

Only 6 IP were identified as influential by more than 4 participants. One was the bundle power (a boundary condition of the calculation), and 5 were model parameters (wall and interfacial HTCs, interfacial friction coefficient, heat transfer enhancement at the QF and droplet diameter).

Several participants discarded some identified influential parameters, considering that their effect was included in other parameters, or because they are not model parameters (e.g. bundle power).

Participants assigned ranges of variation to the considered parameters, based on sensitivity calculations for test 216 of FEBA and using as responses cladding temperatures and QF propagation.

The variations of responses at the extremes of the calculated ranges were checked. For some parameters (power, wall and interfacial HTC), there was a qualitative agreement among participants. For other parameters (droplet diameter, interfacial friction coefficient), the sign of response variations changed for different codes, even for different models within the same code. This is probably due to a complex dependence of the responses on the parameters.

PHASE III

In Phase III, the uncertainty of influential IP (identified in Phase II) was quantified. Participants in Phase III obtained uncertainties in the form of ranges or probability distributions. The model parameters being quantified are related, with few exceptions, to wall heat transfer, interfacial heat transfer and interfacial friction.

The uncertainties obtained depend on a number of features:

- the responses used on quantification;
- the set of IP being quantified;
- the TH code, and the specific model being used.

The three of them have a significant important user effect.

The results show a stronger dependence on the quantification method than on the TH code.

The results exhibit a large variability and discrepancy among participants. In some cases, extremely small uncertainty ranges have been found for models parameters (e.g. interfacial friction) which are physically non-realistic (they are below the attainable accuracy of experimental data). They have been obtained by CIRCÉ users, who performed the recalibration and evaluated the uncertainty ranges for it, instead of the nominal calculation. This was particularly problematic, if only cladding temperature and quenching times were used for recalibration and quantification.

PHASE IV

In Phase IV, uncertainties calculated for model parameters in Phase III are propagated through the code simulation for the selected tests of FEBA and PERICLES experiments.

For FEBA tests, the exercise was aimed at the confirmation of the quantified uncertainties. Most of participants obtained uncertainty bands enveloping the experimental values. The width of the bands varies a lot among the participants, so that a strong lack of coherence of the results has been concluded from the analysis with IRSN method.

The obtained uncertainty bands are very influenced by the responses used in the quantification and by the selected IP.

For PERICLES tests, the exercise can be viewed as a validation of the quantified uncertainties, and the results are less satisfactory than those for FEBA. Considering all the contributions, the fraction of experimental values which are enveloped by uncertainty bands is clearly lower than in FEBA, and very far from the expected value inferred from the random sample size and the order statistics used in the construction of the band. This means that a direct extrapolation of the model uncertainties calculated in FEBA to PERICLES gives poor results. There are significant differences between PERICLES and FEBA tests. PERICLES is larger and has 2-D effects (crossflows). Furthermore, TH conditions in the tests are similar but also different in the two facilities.

The quality of the nominal calculation is important in the final results of the propagation calculations. A good nominal calculation facilitates that the uncertainty bands envelop the experimental data.

The results are strongly related to the quantification method, and are almost independent of the TH code used. CIRCÉ produces narrower uncertainty bands than the other methods used in PREMIUM. CIRCÉ users tend to envelop the experimental data for FEBA, while for PERICLES they generally fail to envelop. MCDA method gives similar results.

The rest of methods, including FFTBM, produce wide or very wide uncertainty bands, that tend to envelop the experimental data, and thus give more coherent results (according to IRSN methodology for analysis of calculated uncertainty).

CEA repeated the calculations of CIRCÉ without the estimation of a recalibrated value. The results for PERICLES improved significantly. CEA believes that working without a recalibrated value can be a possible solution to improve CIRCÉ results in PREMIUM. When considering simultaneously different experiments, the code is well calibrated and there is no more systematic over or underestimation, there is no need of a recalibrated calculation. That corresponds with the idea of a “best estimate” code.

A user effect is observed in the quantification of uncertainties. Participants using the same method and the same version of the same system code obtain uncertainty bounds which are significantly different. This may be due to differences in the input deck, which produce different nominal calculations, and also to different choice of the IP and the responses.

GENERAL CONCLUSIONS

PREMIUM benchmark has been a valuable exercise on methods for UQ of physical computational models, and the application to the models involved in the reflooding prediction.

Different methods and thermo-hydraulic codes have been used along PREMIUM. Results have been very dependent on the quantification method, rather than on the code. Good evidence is the fact that the uncertainty bands produced by participants using CIRCÉ and MCDA are significantly narrower than those using the remaining quantification methods.

Furthermore, the results of quantification have shown dependence on topics such as:

- the selected responses used in the quantification;
- the selected parameters to be quantified;
- the selected database for quantification;
- the code modelling and the numerical implementation;
- the quantified models, which, in general depend on the TH code being used.

There is still a lack of clear guidelines on these topics. Indeed, participants in PREMIUM took miscellaneous decisions about them. It is concluded that the quantification methods used in PREMIUM showed a strong user effect.

As a final outcome, the results of quantified uncertainties in PREMIUM showed a large variability and discrepancy among participants

PREMIUM has been useful as a test bed for inverse quantification methods as CIRCÉ, FFTBM, MCDA, DIPE, etc. Some of these methods have been developed or improved for the participation in PREMIUM. The benchmark has revealed the necessity of further development of inverse methods, and the development of guidelines for evaluation of model input uncertainty.

The propagation of the quantified model uncertainties to FEBA tests has given better results than the analogous exercise for PERICLES tests, in the sense that the calculated uncertainty bands for responses 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. It is a methodology useful in validation exercises, where responses calculated with uncertainty are compared to real values. The methodology captures the basic criteria:

- the narrower the bands, the more informative they are from the predictive standpoint.
- the closer the reference value is to the experimental one, the better is the calibration of the models.

The two criteria are negatively correlated. The IRSN method has supplemented the qualitative analysis of results performed by CEA.

For methods having the option of calculating a calibrated value of the responses (e.g. CIRCÉ), better results are obtained when such calibration is omitted. This conclusion is coherent with the "best estimate" qualification of the TH codes being used, which would not need any recalibration.

5.2 Lessons learnt and recommendations

<u>Quantification methods</u>

Methods for UQ of the physical models in system TH codes must be further studied and developed, so that their different performances can be understood. Issues that should be tackled are:

- statistical versus non-statistical methods;
- Bayesian versus frequentist statistical methods;
- procedures for modelling uncertainty;
- choice and influence of the parametric probability distribution assumed for the parameters (in statistical methods);.
- validation criteria for quantified uncertainties.

For methods having the option of performing calibration additionally to UQ, such option is not recommended, because this recalibration seems incoherent with the best estimate qualification of system codes. Anyway, the use of "calibrated calculations" as reference cases would deserve further study.

Quantification database

A very important point is how to choose the database for development, validation and quantification of a physical model. The database will define the range of validity of the quantified uncertainties.

A compromise must be found between a specific and a generic standpoint. The database should be specific, in the sense that must be related to the foreseen application on the quantified uncertainties. In other words, if the model uncertainty is needed for calculating a specific scenario in a plant, the database should include experiments related to the scenario. But the database must be generic enough, so that the quantified model uncertainties are applicable to a wide spectrum of simulations.

Quantification methods are intended for intermediate experiments. Some parameters may be quantified on grounds of separate effects tests (SET). In such case, it is important to have guidelines about how to proceed: using for quantification only the SET data, or combining them with intermediate tests (IT) data.

User effect

Complex physical models may have a considerable number of physical parameters and produce a large number of responses. The results of the quantification of model parameters uncertainty are very dependent on the selected parameters to be quantified and the selected responses to be used in the process.

The selection of parameters, responses and database are fundamental parts of quantification methods. Guidelines and procedures should be established for such processes. Other aspects of the UQ process that deserve development of guidelines are

- scaling issue;
- assessment of the applicability of codes;
- modelling and simulation of experimental tests;
- validation of quantified uncertainties

Without these type of guidelines and procedures, the methods will have a strong user effect. Quantification methods are tools to reduce the engineering judgement, but they cannot eliminate it.

Quantification and extrapolation

In many instances, model parameters are adjustable coefficients used for fitting the models to real data, and have null or little physical meaning. These fitted models may have poor ability of extrapolation outside the range of development and validation.

The quantified uncertainty obtained for a specific parameter strongly depends of the total set of simultaneously quantified parameters. This means that quantified uncertainties are attributes of the total set of parameters, rather than *intrinsic* properties of individual parameters. Extrapolation of quantified uncertainties (i.e. application to forward calculations outside the range of validity) may lead to erroneous results.

The set of quantified parameters must include the most influential ones on the responses; otherwise the resulting uncertainty may be completely misleading. On the other hand, it is advisable to include in the quantification all potentially important model parameters, not only the most influential ones, because in other applications the set of dominant parameters may be different.

Method and code effect

Results of model quantification seem to be more dependent on the quantification method than on the TH code used. Large differences have been observed in the quantified uncertainty, depending on the method used.

In a computational code, models are organised in a hierarchical structure, so that an individual model generally encompasses several sub-models or correlations. This structure has to be taken into account in the quantification of model parameters, and guidelines are needed for that matter. For instance, the different results obtained by quantifying different sub-models or by quantifying the complete model via a global multiplier should be analysed.

Model input quantification

The experimental uncertainties should be carefully examined, because they can be influential on the quantification.

Quantification methods should not be applied to initial conditions, boundary conditions, material properties, and other magnitudes having full physical meaning, unless there is no other source of information about their uncertainty.

5.3 Suggestion for future work

A main conclusion of the PREMIUM activity is that methods for UQ of physical models should be further developed and studied. Specifically, guidelines and procedures must be established for the development and application of these methods.

The development of this type of guidelines should be a basic part of a systematic approach to the UQ issue, where future work should be focused, as described in the sequel [40].

The analysis of PREMIUM Phases III and IV has shown a large dispersion of participants' results. Moreover, the results were not satisfactory when moving from FEBA to PERICLES. One reason could be the lack of common consensus and practices in the used process and method, more specifically related to:

- The selection of the outputs of interest (responses) used for input quantification: it appeared that some participants only focused on cladding temperature while others also considered quench times.
- The selection of IP whose uncertainties should be evaluated: it appeared for example that some participants did not consider parameters related to interfacial friction or that a global heat exchange coefficient multiplier was used rather than several multipliers for each correlation involved in the global heat exchange.
- The selection of the experimental database: the 6 available FEBA tests were not taken into account in the quantification step by some of the participants. The lack of representativeness of the FEBA experiment (used for input UQ) for PERICLES (used for input uncertainty validation) was also pointed out.
- The code modelling and the numerical implementation: all the participants considered a 1-D modelling for FEBA but 4 participants considered a 3-D modelling for PERICLES. Moreover, 3 participants have a significantly lower number of meshes in the vertical direction for PERICLES than for FEBA.
- The quantification methods: 6 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).

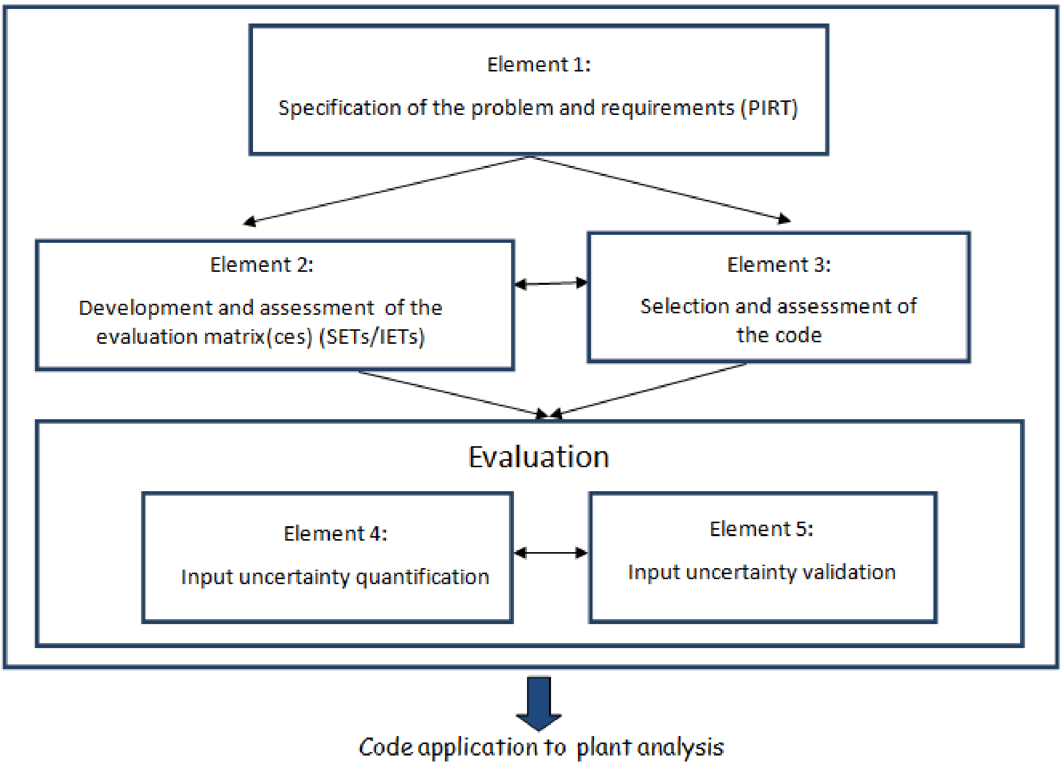
Therefore, it should be valuable to investigate in further developments if a systematic approach devoted to model input uncertainty evaluation (i.e. quantification and validation) can improve the reliability of the analysis and ensure the extrapolation of its results to the NPP case.

From a methodological point of view, a systematic approach has already the advantage of providing a common and generic framework to facilitate discussions between participants and applications to several industrial problems.

Such a type of approach is already widely spread in industrial applications where it is used by experts and engineers for the choice and calibration of physical models in computer codes, for Validation and Verification of computer codes [19] and also for evaluation of uncertainties associated with code calculations (e.g. CSAU [2]) . Moreover, Regulatory bodies integrate code and method development and assessment process in their regulatory guides (e.g. EMDAP [16]). Finally, the development of new procedures for the treatment of model uncertainty remains an active research field of interest [41, 42]. The proposed further works should therefore exploit the current state of knowledge to develop a systematic approach for TH code model input uncertainty evaluation.

In [40] 5 key elements of this approach are pointed out. They are displayed in Figure 5.3.1 and fully detailed in the sequel.

Figure 5.3.1: Key elements of the systematic approach



Element 1 allows to share/check a common understanding between participants on the problem to analyse. It includes the definition of the objectives of the evaluation (e.g. quantify and validate the uncertainties of the reflooding heat transfer models for application to plant analysis), the selection of an NPP and a scenario as well as the code outputs of interest and the important physical phenomena thanks to a PIRT.

Element 2 is related to the construction of an adequate (sufficient and representative) experimental database for input UQ and validation that will control the capability of the method to extrapolate its results to real situations. It should be based on available SETs and IETs but can also require extra experiments if necessary. The completeness/sufficiency of the database (e.g. every important phenomenon should be covered) has to be carefully checked as well as its adequacy. At the end of this step, a ranking between experiments within the database could be performed using multi-criteria decision analysis methods [43].

The feasibility of constructing standardised databases for key physical models (similar to those existing for critical heat flux) in the scenarios under study should be considered.

One important issue of this element concerns the question of dependency of the experimental database with respect to the reactor transient. More precisely, if the database is dependent on the reactor transient, this transient is divided into several "parts" within which the uncertainty of a given model should be evaluated. These parts are defined according to several possibilities that may be combined:

- Different components of the reactor (e.g. vessel, downcomer, cold leg, etc.).
- Different period of times (e.g. depressurisation, refilling, reflood, safety injection, for a LBLOCA).
- Several macro-phenomena (e.g. reflood, condensation in the cold leg, break flow, etc.) or
- Thermal-hydraulic conditions (power, pressure, flow rate, quality, etc.).
- Uncertainty quantification should be performed for each "part" and may be validated both parts by parts and globally on IET

If the database is designed independently from the reactor transient, the uncertainty of any physical model depends only on the thermal hydraulic conditions and one solution could be to quantify the uncertainties using all SET and to validate them on every transient and every reactor (all IET).

A consensus on the type of strategy and the way to perform them should be reached at this step of the systematic approach.

Element 3 is related to the code and leads to a frozen version that will be used in the next elements of the approach. It includes the assessment of the applicability of the code for modelling the identified important phenomena as well as for modelling the considered SETs/IETs. A special attention has also to be devoted to nodalization strategy and model option selection that should be consistent between the experimental facility and similar components in the nuclear power plant. Once the code version is frozen, uncertain model IP are finally identified.

Element 4 consists in inferring, from the experimental knowledge, the information related to 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 input uncertainty validation). It then requires the selection of a set of differences between code calculation and experimental value. To do that, a special attention should be devoted to the quantification of the discrepancy Code/Experiment that could depend on the outputs of interest (scalar, time trend). Finally, the inference can be performed. Besides the choice of the input UQ method (FFT, Monte Carlo, Bayesian, …), it requires an appropriate uncertainty modelling for each uncertain input (interval, possibility distribution, probability distribution…) that should take into account the real state of knowledge (nature of uncertainty and available information) and reduce as much as possible extra assumptions. Key questions of this element are also related to the strategy to follow in presence of several experiments (a single quantification per experiment or rather a unique quantification for all experiments considered together) as well as in case of several quantifications (how to combine input uncertainties, keeping in mind that several options exist). It seems that some recommendations should be done on this point.

Element 5 is based on the propagation of the input uncertainties obtained in Element 4 through the computer code. It can be included in an iterative process with Element 4. It exploits the remaining subset of the experimental database identified in Element 2 and not used in Element 3. The propagation first implies the selection of an uncertainty model for each uncertain input (interval, possibility, probability distribution, etc.) that can be different from the uncertainty modelling associated with Element 4. Moreover, the input sampling procedure should be specified as well as the quantities of interest derived for the output sample that will be used for validation (e.g. percentiles in the probabilistic framework). Finally, a key point of this step is the definition and computation of validation metrics. It requires to reach a consensus on the definition of “validated uncertainty bands” (i.e. which important properties an uncertainty band has to satisfy to be accepted) and to introduce relevant criteria that mathematically translate this definition.

It should be noted that Elements 1-3 are common to any BEPU methodology based on CSAU or EMDAP, focusing on the application of a quantified (fully verified and validated, with model input uncertainties quantified and validated) code for accident analysis. The good practices from those industrial development and applications will be taken in this framework.

The PREMIUM benchmark has been devoted to Elements 4 and 5. The sources of discrepancy between participants recalled at the beginning of this section are included in Elements 1, 2, 3 and 4 respectively. Therefore, a “top-down” systematic approach can be seen as a way to share a common understanding about "good practices” for input uncertainty evaluation.

According to the open issues identified in the PREMIUM benchmark, in [40] 3 interacting axis for further developments have been identified. They concern:

- The comparison of different strategies to construct experimental database in order to orientate engineers in the process of evaluation of TH code model input uncertainty.
- The study on how to deal with several experiments in the input UQ as well as in the input uncertainty validation.
- The preparation of a “good practice document” for a systematic approach shared by all participants and based on:
  - experience from industry;
  - advances in the research and development;
  - lessons learnt from BEMUSE and PREMIUM;
  - strategy and method exhibited by previous axes.

All these further developments could be performed in task group(s) organised within WGAMA as a follow-up of PREMIUM.

REFERENCES

/1/ Code of Federal Regulations, Title 10, Section 50.46. ***"Acceptance criteria for emergency core cooling systems for light-water nuclear power reactors"***. 1989

/2/ Technical Program Group. ***"Quantifying Reactor Safety Margins: Application of Code Scaling, Applicability and Uncertainty Evaluation Methodology to a Large Break Loss of Coolant Accident",*** Technical Report NUREG/CR-5249 EG&G Idaho, Inc. 1989

/3/ OECD/Nuclear Energy Agency/Committee on the Safety of Nuclear Installations: **"*CSNI Integral Test Facility Validation Matrix for the Assessment of Thermal-Hydraulic Codes for LWR LOCA and Transients"***. NEA/CSNI/R(96)17, OCDE/GD(97)12, July 1996.

/4/ OECD/Nuclear Energy Agency/Committee on the Safety of Nuclear Installations: **"*Separate Effects Test Matrix for Thermal-Hydraulic Codes Validation"***. NEA/CSNI/R(93)14, OECD/GD(94)82, 83, September 1993.

/5/ OECD/Nuclear Energy Agency/Committee on the Safety of Nuclear Installations: **"*Thermohydraulics of Emergency Core Cooling in Light Water Reactors"***. CSNI Report no. 161. October, 1989

/6/ OECD/Nuclear Energy Agency/Committee on the Safety of Nuclear Installations: **"*Report on the Uncertainty Methods Study"***. NEA/CSNI/R(97)35, June 1998.

/7/ OECD/Nuclear Energy Agency/Committee on the Safety of Nuclear Installations: ***"BEMUSE Phase VI Report: status report on the area, classification of the methods, conclusions and recommendations".*** NEA/CSNI/R(2011)4, March 2011.

/8/ F. Reventós, E. de Alfonso and R. Mendizábal, ***"OECD/NEA PREMIUM Benchmark. Phase I: Methodologies and data review".*** Universitat Politécnica de Catalunya & Consejo de Seguridad Nuclear.

/9/ A. Kovtonyuk, A. Petruzzi and F. D'Auria, ***"Post-BEMUSE Reflood Model Input Uncertainty Methods (PREMIUM) Benchmark. Phase II: Identification of influential parameters".*** OECD/Nuclear Energy Agency/Committee on the Safety of Nuclear Installations. NEA/CSNI/R(2014)14, February 2015.

/10/ T. Skorek, ***"OECD/NEA PREMIUM Benchmark. Phase III: Quantification of model uncertainties on the basis of FEBA/SEFLEX reflood experiments".*** GRS Garching, Germany, September 2014.

/11/ A. de Crécy and J. Baccou, ***"OECD/NEA PREMIUM Benchmark. Phase IV: Confirmation / validation of the uncertainties found within Phase III". orrelation for boiling Heat transfer to Saturated Fluids in convective Flow,*** 1-EC CEA & IRSN (France), December 2014.

/12/ E. de Alfonso, F. Reventós, R. Mendizábal, ***"Specifications Phase I General requirements and description specification of methodologies and experimental data to be used in PREMIUM benchmark",*** Universitat Politècnica de Catalunya. Grup d' Estudis Termohidràulics. December 2011.

/13/ A. Petruzzi, A. Kovtonyuk, F. D'Auria, ***"Specifications for Phase-II of the PREMIUM benchmark"***, PREMIUM-1/01(11) Rev.1, Pisa, Italy, March 2012.

/14/ T. Skorek, ***"PREMIUM Benchmark Phase III Specification"***. Gesellschaft für Anlagen – und Reaktorsicherheit (GRS) mbH. October 2012.

/15/ A. de Crécy, P. Bazin, J. Baccou, ***"PREMIUM: specifications foer the phase IV using the 2-D reflooding PERICLES tests",*** CEA and IRSN, October 2013.

/16/ US Nuclear Regulatory Commission, ***"Transient and Accident Analysis Methods"***, Regulatory Guide 1.203 (2005). Revision 0. December, 2005

/17/ A. Tarantola, ***"Inverse Problem Theory and Methods for Model Parameter Estimation"***. SIAM, 2005.

/18/ J.G. Arnold *et al*, ***"SWAT: Model Use, Calibration and Validation"*** Transactions of the ASABE. Vol. 55(4): 1491-1508, 2012.

/19/ W.L. Oberkampf & C.J. Roy. ***"Verification and Validation in Scientific Computing"***, Cambridge University Press, 2010.

/20/ R. Mendizábal, E. de Alfonso, J. Freixa, F. Reventós, ***"Questionnaire for PREMIUM participants"***. E-mail submitted on March 20th, 2015.

/21/ A. de Crécy, ***CIRCÉ: A methodology to quantify the uncertainty of the physical models of a code"***, CEA, STMF/LGLS, March 2012.

/22/ J. Heo et al, ***"Implementation of Data Assimilation Methodology for physical Model Uncertainty Evaluation using post-CHF Experimental Data"***. Nuclear Engineering and Technology, Vol. 46, No. 5, October 2014.

/23/ A. Kovtonyuk, A. Petruzzi & F. D'Auria, ***"Methodology for the Quantification of the Input Uncertain Parameters by the Use of the FFTBM. "PREMIUM" Benchmark".*** Universitá di Pisa**,** PREMIUM-1/02(12) Rev. 0, February, 2012.

/24/ J. Joucla & P. Probst, ***"DIPE: Determination of input parameters uncertainties. Methodology applied to CATHARE V2.5_1 Thermalhydraulics Code".*** 15th International Conference on Nuclear Engineering (ICONE15), Nagoya, Japan, April, 2007.

/25/ J. Joucla & P. Probst, ***"DIPE: Determination of input parameters uncertainties. Methodology applied to CATHARE V2.5_1"***. Journal of Power and Energy Systems, 2008.

/26/ J. Bartak, D. Bestion, T. Haapalehto, ***"The top-down reflooding model in the Cathare code"***, Nuclear. Engineering. & Design,149 (1994) 141-152.

/27/ Silvernnoinen et al., **"The APROS Software for Process Simulation and Model Development"** VTT publications, 1989

/28/ H. Austregesilo et al., "**ATHLET Mod2.2 Cycle A Models and Methods**", GRS-P-1, Vol.4, GRS, July 2009

/29/ USNRC, ***"TRACE V5.0, Theory Manual: Field Equations, Solution Methods and Physical Models"***, Division of Risk Assessment and Special Projects, Office of NRR.

/30/ *Information system Laboratories.* **"Relap5/Mod3.3 Code Manual. Volume IV: Models And Correlations"**, *October 2010.*

/31/ B. D. Chung et al., "**MARS Code Manual Volume V: Models and Correlations**". KAERI/TR—3872/2009 February 2010.

/32/ T. Skorek, ***"Description of FEBA Test Facility and FEBA/SEFLEX Experimental Program"***. OECD/NEA PREMIUM benchmark, GRS, Germany, October 2012.

/33/ R. Deruaz et al, ***"Study of two-dimensional effects in the core of a light water reactor during the ECCS phase following a Loss Of Coolant Accident"***, Centre d'Etudes Nucléaires de Grenoble, February, 1985.

/34/ J. Baccou, E. Chojnacki, ***"A practical methodology for information fusion in presence of uncertainty: application to the analysis of a nuclear benchmark"***, Environ Syst Decis 34(2): doi 10.1007/s10669-014-9496-3, 2014.

/35/ R.M. Cooke, ***"Experts in uncertainty"***, *Oxford University Press, 1991.*

/36/ S. Destercke, ***"Uncertainty representation and combination: new results with application to nuclear safety issues"***, PhD Thesis, University of Toulouse, 2008.

/37/ S. Destercke, E. Chojnacki, ***"Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes"***, Nuclear Engineering and Design 238(9):2484–2493, 2008.

/38/ D. Dubois, H. Prade, ***"Possibility theory"***, New York Plenum Press, 1988.

/39/ M.H. Kalos & P.A. Whitlock, ***"Monte Carlo Methods".*** Second edition, Wiley-VCH, 2008.

/40/ (Contributed by IRSN, TRACTEBEL, CEA), ***"Proposal for further development. Submitted to Phase V coordinator"***. June, 2015.

/41/ M. Pourgol-Mohamad, A. Mosleh, M. Modarres, ***"Structured treatment of model uncertainty in complex thermal-hydraulics codes: Technical challenges, prospective and characterisation"***, Nuclear Engineering and Design, 241, 285-295, 2011.

/42/ C. Unal, B. Williams, F. Hemez, S.H. Atamturktur, P. McClure, ***"Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors"***, Nuclear Engineering and Design, 241, 1813-1833, 2011.

/43/ V. Belton, T.J. Stewart, ***"Multiple Criteria Decision Analysis : An Integrated Approach"***, Kluwer Academic Publishers, 2002.

/44/ H. Glaeser, E. Hofer, M. Kloos, T. Skorek, ***"Uncertainty and Sensitivity Analysis of a Post-Experiment Calculation in Thermals Hydraulics"***, *Reliability Engineering and System Safety, Vol. 45, pp. 19-83, 1994.*

/45/ A. Kovtonyuk, A. Petruzzi, F. D'Auria: ***"Proposal of Methodology for Assessment of Uncertainty of System Codes 'Input Parameters'"***, 7th Workshop on OECD/NEA Uncertainty Analysis Methodology Program, OECD/NEA headquarters, Issy-les-Moulineaux, France, 2013.

/46/ A. de Crécy, ***"Determination of the Uncertainties of the Constitutive Relationships of the CATHARE 2 Code"***, M&C 2001, Salt Lake City, Utah, USA, September 2001.

/47/ A. de Crécy, P. Bazin, ***"Quantification of the Uncertainties of the Physical models of CATHARE"***, BE2004, Washington D.C., USA, November 2004.

/48/ E. Chojnacki, A. Ounsy, ***"Description of the IPSN method for the uncertainty and sensitivity analysis and the associated software: SUNSET"***, Proc. ASME/JSME, ICONE, Louisiana, USA 3:545–550, 1996.

/49/ E. Hofer et al: ***"Uncertainty and Sensitivity Analysis of Accident Consequence Submodels"***, Proc of ANS/ENS International Topical Meeting on Probabilistic Safety Methods and Applications, San Francisco 1985.

/50/ International Atomic Energy Agency, ***"Best Estimate Safety Analysis for Nuclear Power Plants: Uncertainty Evaluation"***, Safety Reports Series, No. 52, Vienna 2008.

/51/ C. Baudrit, D. Guyonnet & D. Dubois, ***"Joint propagation and exploitation of probabilistic and possibilistic information ir risk assessment"***, IEEE Trans. Fuzzy Systems, 14:593.608, 2006.

/52/ E. Chojnacki, J. Baccou & S. Destercke, ***"Numerical sensitivity and efficiency in the treatment of epistemic and aleatory uncertainty".*** International Journal of Intelligent Systems, 39(7): 683-704, 2010.

/53/ J.C. Helton et al, ***"Representation of analysis results involving aleatory and epistemic uncertainty",*** International Journal of genneral Systems, 39(6): 605-646, 2010.