

ECOLE DOCTORALE
DOCTORAL SCHOOL

PROGRAMME DOCTORALE EN PHYSIQUE
DOCTORAL PROGRAM IN PHYSICS



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Plan de recherche / *Research Plan*

Bayesian Uncertainty Quantification of
Physical Models in Thermal-Hydraulics System Codes

Damar Canggih Wicaksono
Prof. Dr. Andreas Pautz
Omar Zerkak

1 Introduction (cadre général) / *Introduction:*

During the early days of reactor safety analysis, prediction using model of power plant behavior during transient, off-normal or accident conditions, is approached with high-degree of conservatism. Conservatism calls for the most penalizing modeling assumptions to ensure conservative results, which is far below their regulatory limits. This approach, though less realistic, is justified by limited modeling capabilities as well as limited knowledge of physical processes. However, it was later found that there are conditions where conservative assumptions do not necessarily lead to conservative (or even physical) predictions.

As an example of this contradiction, consider the following situation in the analysis of Loss-of-Coolant-Accident (LOCA) as described in [?]. Assuming less interfacial shear between liquid and gas phase of water reducing mist flow during LOCA is conservative in the sense that less heat will be transferred to the coolant flow in the upper region of the core. This penalizes the fuel temperature prediction. But this assumption will also imply that time to refill the core is shorter as there are more liquid retained in the cooling system. Furthermore, with less shear, there is less resistance in injecting emergency coolant into the core (condition known as counter-current flow limitation). Both of these are clearly non-conservative and contradicting the first conservative assumption.

Therefore, a more accurate prediction of two-phase flow transient behavior was deemed necessary for safety analysis of nuclear power plant under accident conditions. As opposed to conservative analysis, the term for this approach was coined *best-estimate* analysis. Such an analysis calls for physically sound thermal-hydraulics models with realistic modeling assumptions which at the same time are also backed by experimental data. With the idea of having more realistic prediction in mind, a best-estimate thermal-hydraulics (TH) system code was developed. The code is designed to be a comprehensive tool capable simulating realistically wide range of transients foreseen in nuclear power plant operation. It was developed using the current best understanding of flow processes expected to happen during the transients.

1.1 Thermal-Hydraulics Models and System Codes

As mentioned, thermal-hydraulics system codes was designed to simulate system behavior of a power plant during off-normal or accident conditions. The typical structure of system code is illustrated in Fig. ???. As can be seen, system code constitutes of several building blocks that can be used to modeled and simulate wide range of system and conditions. The core of system code is the balance equations describing the dynamic state of the fluid. To close the system¹, it has to be complemented by 2 additional set of constitutive equations. First, the equation of state describing the thermodynamic relation between state variables of a given fluid. And second, the closure relationships describing the interaction between phases and each phase to the boundary wall in terms of mass, momentum, and energy.

The state-of-the-art model widely implemented in TH system codes to describe the dynamics of fluid flow in nuclear power plant is based on the two-fluid model [?]. This model separately treats the transport phenomena of the two phases of fluid (gas and liquid) resulting in 6 balance equations (mass, momentum, and energy for each phase). The model can capture phenomena where thermal and mechanical non-equilibrium conditions exist between the two phases, conditions which gives more realistic picture in wide range of nuclear power plant transients.

The validity of the two-fluid model widely implemented in thermal-hydraulics system codes relies on the proper modeling of the transfer terms between phases and each phase to the boundary wall (see Fig. ??). The transfer terms include interfacial drag, wall drag, interfacial heat transfer, and wall heat

¹*i.e.*, to have the same number of equations as the number of unknowns

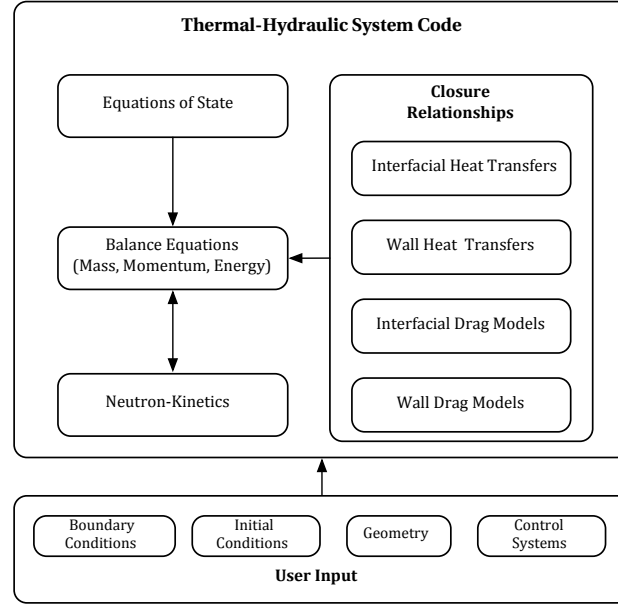


Figure 1: Structure of a system code

transfer. In principle, based on different two-phase flow pattern, different phase distribution as well as different interfacial structure can be observed. As such, each of those transfer terms takes different form with different parameters depending on the pattern of the two-phase flow and they represent different physical processes take place at each pattern.

These physical models, or the so-called *closure laws or relationships* mentioned previously, close the set of balance equations for mass, momentum, and energy of the two phases. They are to be derived experimentally, but proved to be a difficult effort [?, ?] due to various reasons ranging from lack of knowledge to limitation in measurement instruments. Simplifying assumptions and at times extrapolations have to be made simply because of limited available data. Ultimately, the set of models and correlations implemented in the codes is a major of source of uncertainty².

Before continuing our discussion on uncertainty analysis of code predictions, it will be worthwhile to define some terminologies to avoid later confusion. We will refer to the following figure to describe system code in more generic way (*i.e.*, an I/O model). They were adapted from [?] to fit our discussion on system code and its physical model.

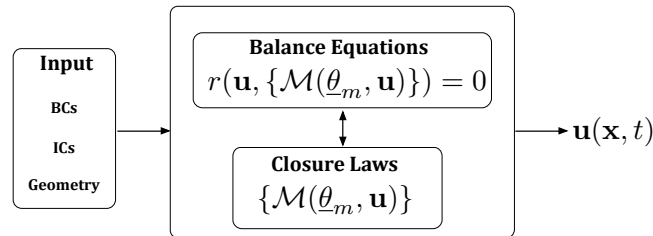


Figure 2: Simplified illustration of system code as an Input/Output model

From the figure, it can be seen that *input* defines specific problem of interest. It includes spec-

²Uncertainty here is defined as a state of limited knowledge, that is of *epistemic* nature

ifications for geometrical configuration, choice of material and fluid involves, as well as initial and boundary conditions. It may also includes specific numerical solution settings. As the input are often parametrized, we will refer to them as *input parameters*, θ_i . Defining the input (as far as user is concerned) completely defines the problem and the code will solve balance equation which output the dynamic state of $\mathbf{u}(\mathbf{x}, t)$ (fluid pressure, temperature, etc.). The balance equations are closed with additional set of closure laws $\{\mathcal{M}(\theta_m, \mathbf{u})\}$. These closure laws in turn are parametrized with a set of parameters θ_m and also a function of the state variables \mathbf{u} . We will refer to θ_m as the *physical model parameters*.

1.2 Statistical Uncertainty Analysis

As explained, best-estimate analysis uses more realistic modeling assumptions for analyzing transient behavior of nuclear power plant. It attempts as realistically as possible to describe the behavior of physical processes occur in power plant transient. And yet, even the best available understanding of the physical processes is still limited. Understanding of complex phenomena might not yet adequate and data support for some processes can be very limited. Simplifying assumptions, approximations and expert judgments to some degree are unavoidable and still required to have a complete analysis.

Hence, best-estimate analysis has to be complemented with uncertainty analysis. The ultimate goal of uncertainty analysis is to associate code prediction with its uncertainty. These combined quantities are then compared with certain regulatory safety limits (*e.g.*, Peak Cladding Temperature, PCT) to check whether the limits still fall outside the uncertainty band of the code prediction.

There are several known sources of uncertainty that render $\mathbf{u}(\mathbf{x}, t)$ prediction uncertain. The following are the sources of primary interest in the current research:

1. Initial and boundary conditions are not perfectly known for a given problem. Additionally, other constituents of input (material properties) might not also not known exactly.
2. The balance equations, originally in partial differential equations, are discretized and solved with finite precision.
3. The closure laws are approximation. They derived from limited experimental data, and in some cases, in different experimental conditions as compared to the purpose of the model.
4. The value model parameters θ_m are often not known *a priori* and requires calibration to fit observed data.

In statistical uncertainty analysis, the input and physical model parameters are modeled as random variables (Θ_i and Θ_m , respectively) with probability distributions. Then using Monte Carlo technique, samples are generated from their respective distribution. These samples are used to run the code multiple times. Finally, statistics of the code output (which is now also a random variable, $\mathcal{U}(\mathbf{x}, t)$) are summarized to obtain uncertainty measure of the prediction. In other words, the uncertainties in the input and physical model are *propagated* forward into the code to quantify the uncertainty of the predictions as shown in Fig. ??.

As can already be concluded, representing the input and physical model parameters as random variables is the key in statistical uncertainty analysis of system code. And indeed, it is the main challenge in making robust statistical analysis of code predictions. The input variables (such as initial and boundary conditions) are usually measured but with finite precision. Thus their uncertainties can be associated with measurement uncertainties.

The physical model parameters, however, are fundamentally different. The physical models we refer to are usually represented either in the form of correlation or mechanistic model. The parameters

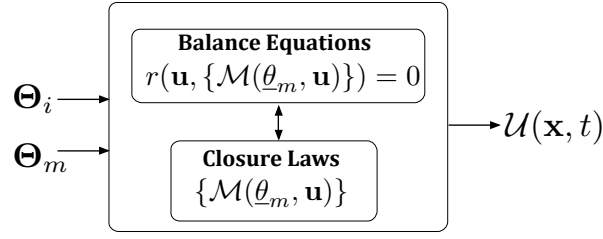


Figure 3: Uncertainty quantification (propagation)

associated with these models are derived from experimental data. They can either represent physically meaningful quantities (reaction rate coefficient) or not (tuning parameter). In either way, there are uncertainties associated with these parameters especially as the condition of their intended use in making prediction can be very different of their derivation (laboratory, controlled condition).

The physical model parameters often cannot be measured directly. To estimate their values, basic experiments are carried out and model or correlation is fitted. Then the optimal value of the estimated parameter is implemented in the code. As the code is used to simulate multiple phenomena at a given period, additional experiments are carried out in more integrated test facility. And, based on the data obtained, the models can be validated and calibrated.

To obtain the uncertainty estimate of the physical model parameters, the problem can be posed as inverse problem. In this setting, given set of experimental data $\{\mathbf{D}\}$ with a known initial and boundary conditions θ_i , the task is to infer the value of the physical model parameters. It is important to acknowledge various sources of uncertainty previously mentioned. Experimental data and initial and boundary conditions though measured are only known to some degree. The models are also only approximation of the real processes. There are uncertainties associated with them.

In probabilistic setting, a natural way to make an inference of parameters based on observed data is through Bayes' theorem [?],

$$p(\theta_m|\{D\}, \theta_i, I) = \frac{\pi(\{D\}|\theta_m, \theta_i, I) \cdot p_0(\theta_m, \theta_i|I)}{p(\{D\}, \theta_i)}$$

Where the left-hand side signifies the posterior probability of the model parameters conditioned on the data, input parameters, and additional background information I . The right hand-side constitutes of π the likelihood (probability of observing data given θ_m), the prior p_o (the degree or state of knowledge we have on the parameter values), and in the denominator, evidence (a normalizing constant to be a valid probability measure). The uncertainty quantification in *inverse* setting can be seen in Fig. ?? to contrast it with Fig. ?. We will revisit this discussion in the subsequent section.

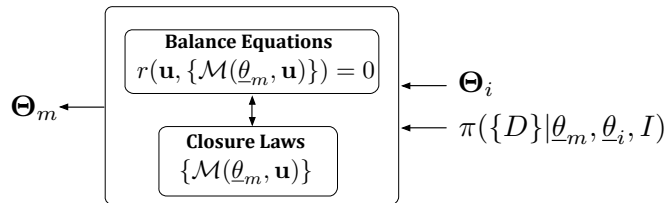


Figure 4: Uncertainty quantification (inverse)

1.3 State of the Research

The importance of characterizing the uncertainty in the physical models parameters is acknowledged by the Working Group on the Analysis and Management of Accidents (WGAMA) of the OECD/NEA³. This led to the PREMIUM⁴ project. Its main goal is to report the state-of-the-art of the available methodologies to quantify the uncertainty in the physical models parameters. The following will briefly describe the PREMIUM project which provides a context for this research as well as some of the available methodologies proposed by different organizations.

1.3.1 OECD/NEA PREMIUM Benchmark

The PREMIUM benchmark is an activity launched by the OECD/NEA in 2012 with the aim to advance the methods for quantifying the physical models parameters uncertainties in thermal-hydraulics system codes. It is a continuation of the previous project BEMUSE⁵, which concentrated on the propagation of input uncertainties in large scale simulation (Large Break - LOCA). The main finding of the BEMUSE project can be found in [?]. The emphasis in the current project is placed on the derivation of the model parameters uncertainty and their validation.

The scope of the project is limited to the physical models in system code to simulate bottom reflood phenomena. Experimental data from two separate effect test facilities (SETF) was made available for the purpose of uncertainty quantification of the model parameters as well as of validation. For the model parameters uncertainty quantification, the data from the FEBA reflood facility [?] is used. The derived uncertainties will then be propagated and compared with the experimental data from another reflood facility (PERICLES). Thus the main goal of the project follows the approach of statistical uncertainty analysis explained previously.

The term “input” defined in the benchmark requires further clarification. The input here is not necessarily coming from usual user specification which is strictly speaking involves mainly system geometry, initial condition, boundary condition specifications. However, the parameters associated with reflood model do become input for the purpose of uncertainty propagation analysis as there are uncertainties associated with these parameters. Following our original convention, the “input” in the benchmark refers to both *model input parameters and model physical parameters*.

1.3.2 GRS Methodology

GRS methodology was developed by GRS⁶ in Germany [?] and for review see [?]. It is based on the statistical technique to propagate the uncertainties in the input and to obtain estimate of reliability measure of the output. The considered sources of uncertainties are comprehensive, ranging from initial conditions, boundary conditions, code models (if an alternative model is available), to the settings for numerical solutions.

GRS adopts subjective view of input and physical model uncertainties which measure state of knowledge. The elicitation procedure to formalize expert judgments and knowledge into probability measure (values between 0 to 1) and distribution was carried out. This procedure of relying on expert judgments by no means simple or non-systematic. In the original paper [?], it was describe that three meetings were organized between the technical experts and statistician. The role of the statistician was to assist the elicitation process. Related to PREMIUM project where uncertainty of input is of interest, GRS adopts the same procedure.

³Organization of Economic Cooperation and Development / Nuclear Energy Agency

⁴Post-BEMUSE Reflood Models Input Uncertainty Methods

⁵Best-Estimate Methods - Uncertainty and Sensitivity Evaluation

⁶Gesellschaft für Anlagen- und Reaktorsicherheit

1.3.3 FFTBM Methodology

FFTBM (Fast Fourier Transform Based Method) is a methodology developed by University of Pisa as part of their integrated uncertainty methodology UMAE (Uncertainty Method Based on the Accuracy Extrapolation) [?, ?]. In FFTBM, the uncertainties of parameters are also treated probabilistically. The distributions of the important parameters are assumed to be uniform and independent. However, the bounds of the distribution are inferred using experimental data and analysis of prediction accuracy.

As the usual quantity of interest both in experimental data or code prediction takes the form of time series, Fast Fourier Transform is applied. The accuracy of code prediction is then treated in the frequency domain. The error or discrepancy (the difference between experiment and calculation) in the amplitude in the frequency domain is compared to the experimental amplitude in the same domain. The ratio between the two is defined as the Average Amplitude (AA).

In the subsequent analysis, sensitivity analysis is first carried out to determine the important parameters that the AA is sensitive to. Then by defining criterion for maximum allowable deviation of AA from the reference case, the range of each parameters can be inferred. As the distributions of these important parameters are assumed to be uniform and independent, the range inferred defines the bounds for this distribution.

1.3.4 CIRCÉ

CIRCÉ (Calcul des Incertitudes Relatives aux Corrélations Élémentaires) is a tool/methodology developed by CEA⁷ in France [?]. The methodology was developed to quantify the uncertainty of the closure laws in CATHARE2 code although applicable to other codes. The main difference of CIRCÉ to the aforementioned 2 is it treats the problem of quantifying parametric uncertainty in system code as an inverse problem in line with the previous discussion.

CIRCÉ operates on two main hypotheses. First, the physical model parameters are assumed to be independent and each is normally distributed. Second, the relationship between the physical model parameters and code prediction is linear. These two hypotheses simplify posterior calculation of the model parameters uncertainty.

Linear model with a Gaussian distribution will yield normal distribution in the posterior. Joint normal distribution in turn is fully characterized by the mean and standard deviation. The means of model parameters and their standard deviation are calculated using Expectation-Maximization (EM) algorithm. Linearity hypothesis only requires first order derivative of the code prediction to each of the parameters.

⁷Commissariat à l'énergie atomique et aux énergies alternatives

2 Objectifs / Objectives:

2.1 Statement of the Problem

The development of closure laws for reflooding described in [?] showed the difficulties and the amount of assumptions used. In principle, system code development is an effort to consolidate correlations and mechanistic models, to create phenomenological-based code that can provide the best-estimate results. This consolidated effort results in a code that can simulate wide range of transients foreseen in nuclear power plant operation in a best-estimate way. Alas, to come up with a consistent set of closure laws is a great challenge for code developers.

The closure laws required to close the two-fluid model pose particularly difficult challenges [?]. For instance, to have correlation of heat transfer between the wall and the fluid, temperature data from each of the constituents are needed (the wall, the liquid phase, and the gas phase). But measuring temperature of the individual phases in arbitrary interfacial topology has its own technical difficulties to the extent that no such data exists or is available to be implemented in the closure laws. Additionally, the experiments to obtain hydrodynamic closure laws (*e.g.*, interfacial friction factor, wall friction factor, etc.) were generally carried out in adiabatic condition. This excludes the coupling of any heat transfer phenomena between the phases and the wall in such correlation.

Furthermore, during the development of a code, programming considerations also came into the picture. For robustness, simplification is often required and continuity is enforced. Transitional regime between two flow regimes for which experimental data does not exist is modeled to be the averaged of the two bounding regimes. In the end, the phenomenological models implemented in a code are specific to the code. Different code development, which used different assumptions and experimental database, comes up with other set of closure laws with their own parametrization (see for instance [?] for TRAC and [?] for CATHARE). Several authors have expressed their concerns about the uncertainty stemming from the closure laws ([?, ?, ?]).

As an example of the point given above, consider that in the TRACE code, after some derivations, the interfacial drag coefficient closure law in the inverted slug flow regime $C_{i,IS}$ is given by,

$$C_{i,IS} = \Psi_{SET} \times \frac{1}{24} \frac{\rho_g}{La} \frac{(1 - \alpha)}{\alpha^{1.8}} ; \Psi_{SET} = 0.75 \quad (1)$$

There are several remarks about the closure law given above. First, the second term in the right-hand side was derived from experimental data but not directly in the form above. In the inverted slug regime, saturated liquid core breaks up into ligaments. These ligaments are *assumed* to take form as prolate ellipsoid. The drag coefficient of distorted droplet from experimental data is then *assumed*. Then to take into account the multi-particle effect, the coefficient is multiplied by the void fraction α raised to the power of 1.8 (this was taken from experimental data of inertial regime). Lastly, the first term, $\Psi_{SET} = 0.75$ was put *to match, to calibrate against* the experimental data from the FLECHT-SEASET reflood facility. This first term, although clearly non-physical, is an important tuning parameter of the model nonetheless. Its uncertainty should be considered in uncertainty analysis, especially when reflood is expected to occur. Yet, no statement of the uncertainty is given. Several other such terms exist [?].

From the above discussion, it is clear that models in thermal-hydraulics system code are flawed. Various experimental programs were carried out to gain better understanding of important phenomena, and to validate (and, as noted above, to calibrate) the models. Series of the experiments, carried out in Separate Effect Test Facilities (SETFs) were aimed to reproduce limited part of the transient in a selected component following a postulated scenario. For example, in the case of reflooding, several facilities existed and data were accumulated (FEBA, PERICLES, etc.). But, there has not been an orchestrated effort to incorporate the accumulated data into the calibration process of the physical models, in systematic way, while acknowledging the uncertainties of other sources in the process.

2.2 Objective

The overall objective of the research is to quantify the uncertainty of physical models parameters implemented in system code in the form of closure laws. These models are parametrized either with physical parameters or tuning parameters. Usual practice by the code developer is to calibrate them against some an experimental without any statement of their uncertainties. To conform with the framework of statistical uncertainty propagation, uncertainty of these parameters are required in the form of probability distribution. The derivation of these distributions will be based on probabilistic modeling and available experimental data. There are three specific aims in the current research in accordance to the overall objective.

First is to investigate whether the current physical model in system codes can be identified with the available experimental data from test facilities. Global sensitivity analysis methodology can be used to assist in identifying which parameters can be calibrated using the available data. Different facilities might have different type of data. And although the information content might not be the same for the different type, it might be worthwhile to consider each one of them.

Second is to calibrate the physical model parameters against various relevant experimental data. The word *to calibrate* carries a disparaging interpretation related to *to tweak*. However, using Bayesian framework explained below, the aim of calibration is extended to preserve the uncertainty of the parameter estimation. Furthermore, in this framework, various sources of uncertainties can be modeled consistently using probability distributions, including the model discrepancy. At the end, the parameters of interest will be in the form of distributions conditioned on the data. These distributions are the main ingredients in the aforementioned statistical uncertainty analysis (propagation).

Third is to consolidate the findings in different but similar facilities. Indeed, for the same kind of transient, specificity exists in different facilities. Some might have more type of data than the others. Some might have been tested in different operating conditions. Parameter estimation of the same parameter derived from different facility might have different distribution. Yet, the physical models of phenomena implemented in system code is supposed to be general in a sense that the value of their parameters and their uncertainties can be applicable to simulate similar phenomena in different settings. On the other hand, the consolidation effort might also reveal that the models as they are currently implemented simply cannot be informed by relatively large amount of experimental data from various facilities that the model supposed to simulate.

2.3 Scope

Although the steps taken in this research can be applicable to any system code physical model calibration and validation, the attention will be focused on the models of particular importance during reflooding, the so-called Post-CHF⁸ flow regimes. The reasons for this emphasis are:

1. Reflooding is an important part in the simulation of nuclear power plant transient during Loss-of-Coolant-Accident. Modeling reflooding determines the appropriate representation of the dynamics of heat transfer phenomena during the effort to rewet an uncovered core. Of paramount interest is to estimate the time the rod can be expected to be rewet as well as the maximum temperature reached prior to rewet. Reflood is a transient with highly coupled hydrodynamic-heat transfer effects and it challenges the assumption made on the implemented closure laws. Indeed, several Separate Effect Test (SET) reflood programs existed and were designed to validate reflood models in system code. Unfortunately, no orchestrated effort was done so far to consolidate the generated data in general and into TRACE code in particular.

⁸Critical Heat Flux

2. The models are adequately complex. It is complex that 5 flow regimes are involved in a single phenomena. But as the source of data is from reflooding separate effect test facilities (SETFs), real plant system effect can be excluded and analysis can be concentrated on limited set of models. In fact, as already pointed out, reflood SETFs are designed to validate (and calibrate) reflood models in sytem codes.
3. Additional data from another reflooding facilities are also available, including data obtained from series of experiments at the Neptun facility conducted at the Paul Scherrer Institut during the 1990s, as well as data from experiments at the PERICLES facility conducted by CEA.
4. It is the model considered in PREMIUM benchmark, thus there is possibility to compare the results of this research with the results of the other participants.

Furthermore the thermal-hydraulics system code considered in this thesis is TRACE (TRAC/RELAP Computational Engine) developed by US NRC. The main reason to consider solely this particular code is the fact that TRACE is the thermal-hydraulic code used for the purpose of Swiss nuclear power plant safety analysis within the STARS (Steady-state and Transient Analysis Research for Swiss Reactor) project carried out at Paul Scherrer Institut (PSI) [?].

2.4 Bayesian Framework

System code is a complex tool to simulate complex problem. Usage of partial information, assumptions, subjectivity, and judgments are inevitable and are, in fact, crucial to make initial progress. The key then is to be as transparent as possible about these assumptions, to systematically check them, and in the light of new data, to allow the change in assumptions. These are certainly more constructive than merely being objective from the beginning, rejecting any notion of assumptions or approximations. All of these can be accommodated naturally using Bayesian approach to quantify uncertainty.

Bayesian approach stated here does not correspond to a specific methodology. Instead, it refers to a framework which pose the inverse problem of parameter estimation in terms of probability. Beck [?] refers to this procedure as *stochastic embedding*. Posed in this way, Bayesian inference machinery can be used to infer the parameters of interest given the available data. Furthermore, as the key idea of Bayesian inference is to explore the posterior distribution (instead of finding single set of optimum parameters in the traditional deterministic calibration setting), the uncertainty of the parameters can be characterized. This characterization of parameter uncertainty of the model while simultaneously taking into account other sources of uncertainty is the ultimate goal of this thesis.

The first step to carry out Bayesian analysis is to cast the problem in terms of probability. The source of uncertainty that caused discrepancy (error) in the model prediction will be modeled as a random variable. The framework for error models used in this research is adapted and elaborated from the work of Huard and Mailhot [?] (see Fig. ??). It is assumed that there exists true process, true input, true output, true model parameters. However, in practice data, is finite and distorted, and models are an approximation due to lack of knowledge. That is, our analysis and prediction are distorted by uncertainties.

According to the figure there are three main types of errors due to uncertainty: input, output, and structural errors. They will have to be modeled probabilistically. The following paragraphs will describe each of the errors.

Input error refers to the difference between the true input $\tilde{\mathbf{x}}$ and the measured input \mathbf{x} . The measured input is limited by the imperfection of the measurement device. In the context of separate effect test, the input is the imposed initial and boundary conditions of the experiment. These conditions drive the system response during the test and they are measured with finite precision. Because the test

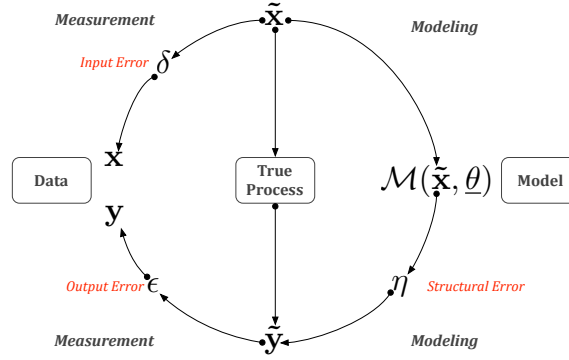


Figure 5: Errors and their relation for Bayesian uncertainty analysis (adapted from [?])

conditions differ even within the same facility, these errors has to be taken into account for robust parameter estimation.

Output error is analogous to the input error described above. It is the difference between the true output $\tilde{\mathbf{y}}$ and the measured output \mathbf{y} . It is the system response for the imposed initial and boundary conditions.

Structural error is defined as the difference between model results and the true output $\tilde{\mathbf{y}}$. Modeling process relates the input to the output via a model parametrized with $\underline{\theta}$. By definition of model, even with a perfect set of parameters (which in principle, are also unknown, uncertain), model will still be inadequate. There are processes which are not taken into account either because of a lack of knowledge, a fallacy in the hypothesis or even an error in the numerical solution. As shown in [?], not taking into account structural error during calibration biases the parameter estimates and reduces the model predictive capabilities.

The treatment of input and output errors belong to experimental data analysis . However, there is no standard approach to the model structural error. The structural error can be expected to show bias (due to incorrect or excluded process) as well as to be correlated in time and space, depending on the nature of problem. Several possibilities are available to model this error [?, ?]. In this context, an additional strength of Bayesian framework is that the plausibility of different error model classes \mathcal{M} can be checked and compared via their evidence as shown in [?].

In general, the above error models can be written as follows in probability density p ,

$$\begin{array}{llll} \text{Input} & \delta & \implies & p(\delta) \iff p(\mathbf{x} | \tilde{\mathbf{x}}) \\ \text{Output} & \epsilon & \implies & p(\epsilon) \iff p(\mathbf{y} | \tilde{\mathbf{y}}) \\ \text{Structure} & \eta & \implies & p(\eta) \iff p(\tilde{\mathbf{y}} | \tilde{\mathbf{x}}, \underline{\theta}, \mathcal{M}) \end{array}$$

Note that the structural error model is conditioned on a specific model class \mathcal{M} chosen and defines its parametrization. To obtain the posterior distribution of the parameters based on their priors and conditioned on the data (both output and input) and specific model class, Bayes' theorem is applied,

$$\begin{aligned} p(\underline{\theta} | \mathbf{y}, \mathbf{x}, \mathcal{M}) &= \int \int p(\underline{\theta}, \tilde{\mathbf{y}}, \tilde{\mathbf{x}} | \mathbf{x}, \mathbf{y}, \mathcal{M}) d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \\ &= \int \int p(\mathbf{x} | \tilde{\mathbf{x}}) \cdot p(\mathbf{y} | \tilde{\mathbf{y}}) \cdot p(\tilde{\mathbf{y}} | \tilde{\mathbf{x}}, \underline{\theta}, \mathcal{M}) \cdot p(\tilde{\mathbf{x}}) \cdot \frac{p(\underline{\theta} | \mathcal{M})}{p(\mathbf{x}, \mathbf{y})} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}} \end{aligned}$$

Where the posterior is conditioned on the observed data and model class \mathcal{M} . The integration is carried out to marginalize the joint distribution. To characterize the posterior distribution using their moments, Monte Carlo simulation is usually performed as the integration becomes intractable. One of the popular approach is to use Markov Chain Monte Carlo (MCMC) simulation to generate sample representative of the posterior.

3 Premiers résultats / *First results*: (1-2 pages)

3.1 Bayesian Calibration of a Turbulence Model

As an illustrative example and to familiarize us with the main idea of the Bayesian framework and its method, simple case of physical model calibration were carried out during the first year. Here the RNG k-epsilon turbulence model (parametrized by 7 parameters and shown below) is calibrated using measurement data from backstep flow experiment [?].

$$\begin{aligned}\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) &= \frac{\partial}{\partial x_j} \left[\left(\mu + \frac{\mu_t}{\sigma_k} \right) \frac{\partial k}{\partial x_j} \right] + P_k - \rho \epsilon \\ \frac{\partial}{\partial t}(\rho \epsilon) + \frac{\partial}{\partial x_i}(\rho \epsilon u_i) &= \frac{\partial}{\partial x_j} \left[\left(\mu + \frac{\mu_t}{\sigma_\epsilon} \right) \frac{\partial \epsilon}{\partial x_j} \right] + C_{1\epsilon} \frac{\epsilon}{k} P_k - C_{2\epsilon}^* \rho \frac{\epsilon^2}{k}\end{aligned}$$

where

$$\begin{aligned}C_{2\epsilon}^* &= C_{2\epsilon} + \frac{C_\mu \eta^3 (1 - \eta/\eta_0)}{1 + \beta \eta^3} \\ \eta &= S k / \epsilon \\ S &= (2 S_{ij} S_{ij})^{1/2}\end{aligned}$$

The 7 parameters from above set of equations are $C_\mu, \sigma_k, \sigma_\epsilon, C_{1\epsilon}, C_{2\epsilon}, \eta_0$, and β . Their initial values are given in Table ???. The turbulence model is implemented in the OpenFOAM framework in which the backstep flow simulation model is part of the OpenFOAM validation case.

The first step in statistical calibration is to develop a stochastic error model that probabilistically relates measured data and simulation results. There are two stochastic error models considered in the present analysis, the model either assumed to be independent or spatially correlated with an exponential form of correlation (with correlation length scale to be estimated). Then Markov Chain Monte Carlo simulation were carried out to characterize the posterior of the parameter conditioned on the data.

The likelihood (error function) is formulated in line with the procedure given in [?]. Here as opposed to approach given by [?], the multiplicative form discrepancy error was assumed such that no-slip boundary condition is retained. The discrepancy error was assumed to be Gaussian with mean= 1 and standard deviation to be estimated.

$$p(\{D\}|\underline{\theta}, M_j) = \frac{1}{\sqrt{((2\pi)^N \times |K|)}} \exp \left[-\frac{(d - y)^T K^{-1} (d - y)}{2} \right] \quad (2)$$

Where d is the measured velocity data, y is the simulation results with a given set of parameters, and K is the covariance matrix which depends on the model structure M_j . For $j = 1$, $K = K_e + K_i$ (measurement and model discrepancy independent variance), while for $j = 2$, $K = K_e + K_c$ (measurement and model discrepancy correlated variance).

This resulted in 8 parameters for M_1 and 9 parameters for M_2 to be calibrated. The priors are set to be uniform and independent with the mean for the turbulence model parameters corresponds to the default values in OpenFOAM and with $\pm 50\%$ upper and lower bound values for each parameter. 1000 samples from the posterior is then generated using Markov Chain Monte Carlo algorithm implemented in Matlab [?]. The priors and the results are tabulated in Table ?? and Table ??, respectively, while the marginal posterior distribution is given in Figure for the 9 parameters model.

As can be seen, all the parameters are informed by the data. This mean that there are more plausible values of parameters with associated uncertainties than the others which fit the data better. From the above analysis, it is clear that the uniform parameter distribution as it was assume a priori can be improved by the data. The uncertainties of the parameter obtained from the posterior can then be propagated to obtain robust prediction in other quantities (*e.g.*, the wall shear stress). However it was not checked the robustness of the parameter estimation using another data sets.

Table 1: Bounds of Uniform PDFs

	Parameter	Initial	Lower	Upper
RNG k- ϵ	C_μ	0.08450	0.04225	0.12675
	C_1	1.42000	0.71000	2.13000
	C_2	1.68000	0.84000	2.52000
	σ_k	0.71942	0.35971	1.07913
	σ_ϵ	0.71942	0.35971	1.07913
	η_0	4.38000	2.19000	6.57000
	β	0.01200	0.00600	0.01800
M1 (Independent)	σ_{mod}	0.10000	0.00000	0.20000
M2 (Correlated)	σ_{mod}	0.10000	0.00000	0.20000
	$\log(\alpha)$	1.00000	0.00000	2.00000

Table 2: Statistics of MCMC samples of Correlated Model

Parameter	mean	std.deviation	MC Error
C_μ	0.061714	0.0060212	0.0013311
C_1	1.43980	0.026246	0.0047692
C_2	1.8468	0.10874	0.023902
σ_k	0.51179	0.07683	0.017243
σ_ϵ	0.81953	0.01831	0.0039055
η_0	5.4496	0.24451	0.055019
β	0.0095692	0.00049043	9.8941×10^{-5}
σ_{mod}	0.14952	0.0072768	0.0015929
$\log(\alpha)$	0.79485	0.045782	0.0099883

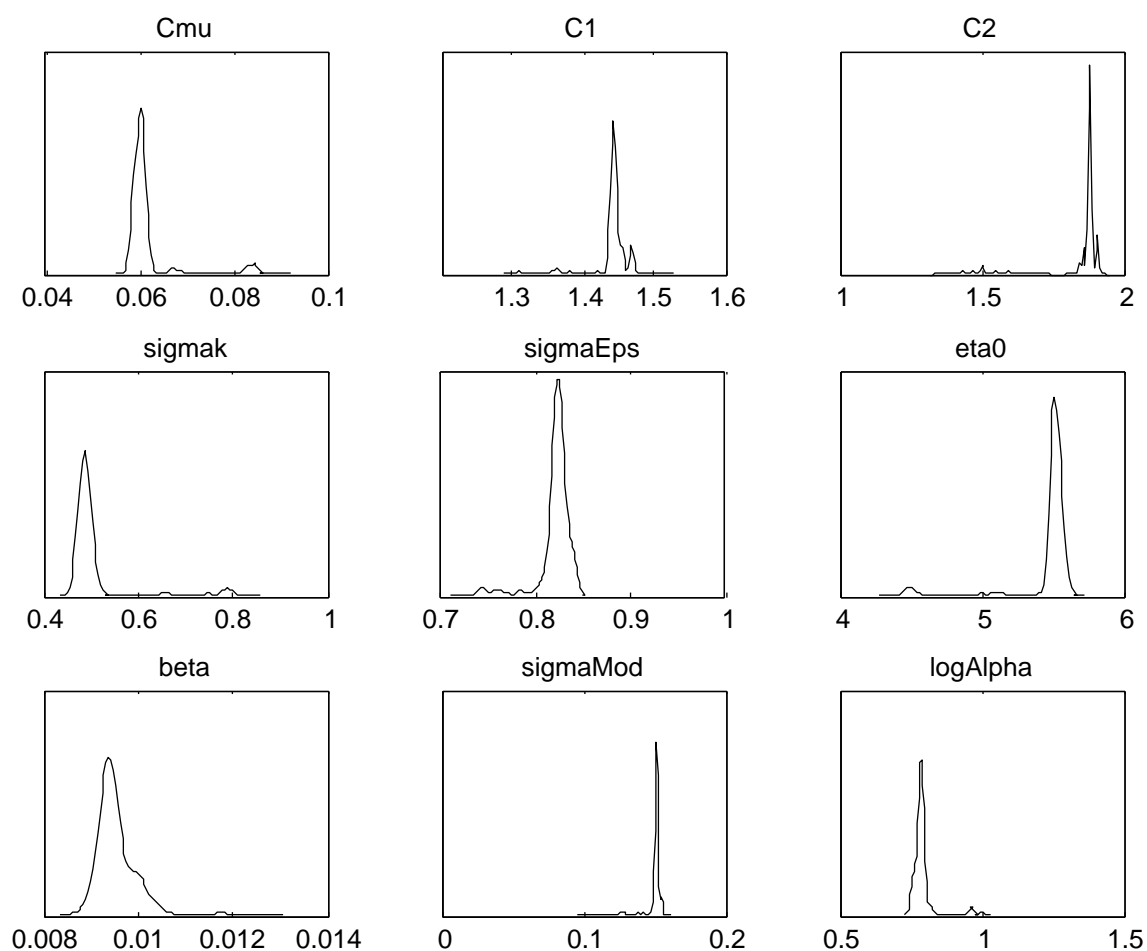


Figure 6: Marginal Posterior PDFs for Model Parameters

This exercise revealed some considerations in applying the Bayesian framework to calibrate TRACE reflood model with the data from reflood experiments:

1. The reflood process is inherently transient in nature, so in addition to measurement on several locations, the data will also be in time series. This need to be considered in deriving error models especially if there is correlation structure both in space and time assumed.
2. Incorporating several type of available data. In some facilities, several other quantities other than cladding temperature were also measured, including for example, pressure drop and void fraction. These call for several goodness-of-fit measures between data and model to be considered at once. Even if the importance of these additional data is less than clad temperature, they nevertheless contain additional information of reflood process.
3. The boundary conditions in reflood experiment are dynamic in nature, the imposed power as well as the injected water temperature is time-dependent. Their uncertainties should also be taken into account.
4. Parameters that need to be calibrated in TRACE reflood model are not part of the user input specification. Some were already hard-coded with values from a calibration with a data from a facility (without any statement of their uncertainties). Some modifications in the source code are required to accommodate sensitivity and uncertainty analysis of TRACE.
5. Aggregating the uncertainty estimation of parameters of reflood model from a facility with another might not be straightforward. Although each reflood experiment in different facility basically experimented a reflood process in a rod bundle, specificities exist and might bias the results from each.

4 Planification future (plan de travail) / *Future planning (scheme of work)*:

The proposed research plan, with emphasized placed on bottom reflood model and TRACE code, is expected to involve the following phases:

1. First, comprehensive review of TRACE closure laws related to reflood phenomena. Most are part of the post-CHF closure package, which includes interfacial drag, wall drag, interfacial heat transfer, and wall heat transfer models or correlations. The goal is to obtain understanding how reflood is conceptualized in TRACE, and along with it how the models are parametrized. This will result in list of parameters that is perceived to be important in reflood simulation.
2. Second, as these parameters are often not part of user specification and hard-coded in TRACE, modification of TRACE is required for it to accomodate sensitivity and uncertainty analyses. The goal here is to externalize important parameters hard-coded in TRACE, thus make them available for user to modify for each run.
3. Third, based on FEBA separate effect test facility, global sensitivity analysis of assumed important parameters (including the boundary and initial conditions) is carried out for the available type of measured data. In Bayesian framework, sensitivity analysis is an important for the purpose of system identification. In essence, if a measured value is insensitive to a (perceived) important parameter, then the uncertainty of this parameter cannot be inferred from this particular data. Excluding them prior to posterior sampling will save computational time or even reveal modeling error.
4. Fourth, based on several error models that can be assumed, posterior sampling is carried out to obtain the uncertainty in the parameters. The posterior sampling is performed via Markov Chain Monte Carlo (MCMC) illustrated in the example above. Further analysis for selecting (or averaging) the most plausible error model is done in this phase.
5. Next is the validation phase where the uncertainties obtained will be propagated in “forecasting” setting, using different data sets than the one used to derive the uncertainty of the model parameter.
6. Sixth, the uncertainties of parameters obtained from two different facilities need to be consistently combined while taking into account the facilities specificity and different range of experimental conditions. This phase involves a repetition of phase 3-5 but based on different facility (*e.g.*, PERICLES).
7. Finally, the results of combined uncertainties estimates need to be validated yet against another facility (*e.g.*, NEPTUN). The goal here is to investigate if indeed experimental data from different reflood SETFs can be consolidated into the current implementation of reflood model in TRACE to increase its predictive capability in simulating reflood process.

Basically for a given of reflood facility, the routine procedure would be: sensitivity analysis - uncertainty (backward) analysis for parametric calibration - uncertainty (forward) analysis for validation. The analyses at first are carried out separately for each reflood facilities. At the latter stage, however, the results will be consolidated where the uncertainty information of reflood physical model based on the data from various facilities are combined together in a consistent way, minimizing the specificity of the facilities and/or the experimental conditions.

Based on the phases identified above, a preliminary schedule for the 3-year dissertation work is proposed as follows:

1. Task 1 (3 months)

- List the important parameters in TRACE that are perceived to be important for reflood simulation in SETF.
- Externalize these parameters such that it can be specified in an input deck for the purpose of sensitivity and uncertainty analyses

2. Task 2 (4 months)

- Test the modified TRACE for global sensitivity analysis.
- Finalize the development of TRACE model for FEBA test facility.
- Carried out global uncertainty analysis of FEBA test facility.

3. Task 3 (10 months)

- Develop various plausible error models for input, output, and structural errors applicable for reflood model calibration. The nature of available data needs to be taken into account, that they are time series of measured values taken from different location. Possibility to incorporate information from different type of data (*e.g.*, temperature and void fraction) need also to be considered.
- Develop or adapt available MCMC sampling algorithm that is efficient, can be parallelized, with informative diagnostics.
- Carried out MCMC sampling based on each of the plausible structure of error models to characterize the uncertainty in the parameters conditioned on the data.
- Select the more plausible model based on the evaluation of the evidence of these models.
- If the reflood simulation proved to be too computationally demanding, develop a surrogate models to assist the calibration procedure.

4. Task 4 (1 month). Validation of the obtain uncertainty estimates by propagating them for another data set within the same facility (FEBA).

5. Task 5 (10 months). In principle, repeat most of the previous four tasks for other reflood facilities (*e.g.*, PERICLES and ACHILLES) to obtain the uncertainties in the important parameters.

6. Task 3 (6 months): Consolidation and Validation

- Develop procedure to consolidate parameter estimation of reflood models based on two/three different reflood facilities.
- Validation of the consolidated uncertainty estimates of the parameters by propagating them in the simulation of reflood using data set from another facility (*e.g.*, NEPTUN).

7. Thesis write-up (3 months):

5 Divers / *Miscellaneous*

5.1 Eventuelles publications / *Publications*

5.2 Présentation(s) orale(s) / *Oral presentation(s)*

5.3 Cours suivis / *Courses attended*

6 Références / *References*

7 Signatures

Candidat(e)/ *Candidate*

Directeur de thèse/ *Thesis director*

Co-directeur de thèse/ *Thesis co-director*
(Officiellement habilité / *Officially authorized*)

Date du rapport/ *Date of report*

Directeur du programme/ *Program director*

Signature:

Date:

A retourner à : EDPY – Station 3 – 1015 Lausanne
To be returned to : EDPY– Station 3 – 1015 Lausanne