Uncertainty quantification and metamodeling of multi-fidelity CFD computation of a heated fuel assembly

Jean-François Wald, Bertrand Iooss

Uncertainty quantification and metamodeling of multi-fidelity CFD computation of a heated fuel assembly

Jean-François Wald and Bertrand Iooss
EDF R&D, Electricité de France, 6 quai Watier, Chatou, 78401, France

September 11, 2024

Abstract

This study proposes to precisely quantify the uncertainty in a CPU-time costly Computational Fluid Dynamics (CFD) model used to evaluate local temperature field in the situation of blocked fuel assembly in a PWR transfer tube. Several uncertain parameters are identified and a first uncertainty propagation study is conducted on a low fidelity (poorly refined) mesh for CPU cost issues. Then, using the concept of "support points", an algorithm is employed to reduce the size of the initial design of experiments. A high-fidelity model (finer mesh, more CPU-time expensive) is then run on this small-size design of experiments. A metamodel was finally built on those high fidelity results to propagate uncertainties and finely analyze the results. The successful results that are obtained show that metamodeling has potential to overcome the issue of highly costly CPU-time CFD models in the near future.

1 Introduction

In the framework of nuclear safety, simulations tools are used at each step of the process - going from design to safety demonstrations. Those analyses are primarily performed using so called 1D "system" codes (such as RELAP or CATHARE), which are validated against a very large panel of experiments and mainly based on empirical correlations (see, e.g., [1]).

Numerical simulators, often referred to as codes, require numerous input parameters that define the phenomenon being investigated or pertain to its physical and numerical modeling. The information available for many of these parameters is frequently limited or uncertain. These uncertainties primarily arise from gaps in our understanding of the underlying physical processes and the challenges in accurately characterizing the model's input parameters, often due to insufficient experimental data [2]. Additional uncertainties can also stem from the specific accidental scenarios chosen for analysis. As a result, both the

input parameters and the outputs generated by the simulator carry inherent uncertainties. In light of this, it becomes crucial to account for these uncertainties when interpreting the results of computer simulations. The quantification of the variability of the outputs due to the uncertainties in the inputs is calculated via propagation of uncertainties. It is a critical component of safety analysis. In the context of nuclear safety, this approach is termed BEPU ("Best-Estimate Plus Uncertainty") and represents a significant advancement in the rigorous assessment of safety measures [3, 4, 5]. The main idea is to be able to compute low or high-order (e.g. 5% or 95%) quantiles of model output variables of interest.

In the recent decades, the use of 3D Computational Fluid Dynamics (CFD) models to predict the entire local flow field in a nuclear system has considerably increased [6]. Indeed, some phenomena are intrinsically 3D by nature (such as the pressurized thermal shock, steam line break and boron dilution) and the use of 1D system codes might not be accurate enough on such configurations. Moreover, by reducing the modeled part of the governing equations and increasing the resolved one, CFD is expected to reduce conservatism and to give a better physical understanding of safety-relevant phenomena.

The verification and validation ("V&V") process is fundamental to quantify the level of confidence one can have in a CFD code [7] for inputs with no significant uncertainties. Uncertainty Quantification (UQ) (see, e.g., [8]) is however more difficult to set up in the CFD context [9]. When performing UQ on CFD codes, one of the main technical difficulties is the CPU time cost of the simulator evaluation. Indeed, depending on the studied configuration, a single run can take hours or days to return its result. Pointing out that the direct Monte Carlo approach (which includes the Wilks method allowing to compute a conservative quantile estimation [10, 11]) often cannot be applied in this context, this issue has been investigated in the review report [12]. This reference identifies the following methods as solutions:

- The accuracy extrapolation methods [13, 14] which requires experimental data,
- The deterministic sampling approach [15, 16] which does not give access to quantile estimates without strong assumptions on the output variable distribution,
- The use of a metamodel [17] which consists in building a CPU-time inexpensive mathematical function approximating a set of results from simulator runs (coming from a well-chosen sample of input values evaluated with the CFD code).

Therefore, if no experimental data are available, and if we want to compute quantile of an output variable without hypothesis on the output distribution, the metamodels appear as adequate solutions as shown by [18] with the generalized polynomial chaos, [19] with the Gaussian process, and [20] with the so-called optimal statistical estimator.

In this paper, a particular industrial safety issue, coming from the engineering division of EDF (Electricité de France, which is the French electrical

company) and related to the operation of Pressurized Water nuclear Reactor (PWR), is studied from the perspective of CFD calculations. When loading or unloading nuclear fuel in a PWR, fuel assemblies go from the core (containment pool or BR pool) to the spent fuel pool (BK pool). This transit is made possible by a device that allows the fuel assemblies to be switch from a vertical to an horizontal position. It is then placed on a cart that goes from the BR pool to the BK pool in a tube (see Fig. 1). The accidental scenario of interest in such configuration considers that a fuel assembly is blocked in the transfer tube for an undetermined time. The saturation temperature of water in the tube is, in this particular scenario of interest, around 122°C. Thus, ensuring that there is no boiling in the tube leading to the draining of the fuel assembly, is a major issue in such scenario.

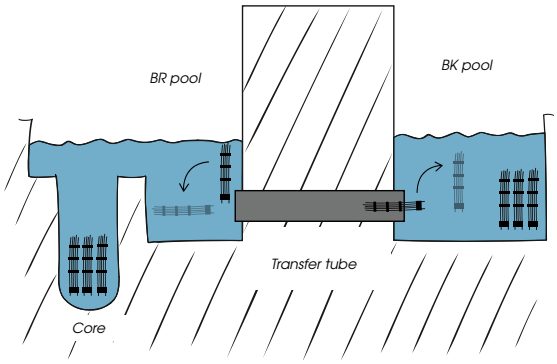

Figure 1: Sketch of the situation studied.

This study proposes to precisely quantify the uncertainty on the temperature in the tube. In particular, it has a threefold objective:

- To quantify the input uncertainties by identifying the uncertain parameters of the study and modeling them with probabilistic distributions,
- To perform an UQ/BEPU study by propagating these uncertainties on a low fidelity CFD mesh (coarse numerical model) to give a first answer to the problem,
- To investigate the possibility to use metamodels to perform uncertainty propagation on a high fidelity mesh (precise numerical model). In this study, we focus on the use of a Gaussian process metamodel [21] which is known to be efficient for simulators up to tens of inputs [22, 23].

Propagating firstly the uncertainties on a mesh that is known to be outside the asymptotic convergence area may appear peculiar as authors seem to already know that the refinement of this mesh is not sufficient to precisely capture the flow field. However, it is the reflect of a day-to-day engineering practice: one build a first mesh, perform some computations and tests on it, before testing

finer meshes to assess solution verification [24]. Therefore, a second step consists in running the code on a high fidelity mesh, but on some locations (in the input space) where the low-fidelity code has been run. Indeed, sampling the two models of different fidelities (framework called "multifidelity code" in the UQ literature [25]) at the same locations would allow to study precisely the difference between their output values (and then to evaluate their convergence). This practice leads to a particular sampling issue that is considered in this paper: How to extract an "optimal" sub-sample of a given size (to run the high-fidelity code) from a first sample (where the low-fidelity code has been run)? In the UQ literature, most of the times, the problem being addressed is the opposite as the user wants to augment a first sample that has been used for the high-fidelity code (see, e.g., the nested designs solution proposed in [26]). Therefore, for our particular purpose, we propose to use the support points technique [27], which is able to solve the problem of sub-sample extraction, as shown in [28].

Section 2 details the CFD modeling of the problem, as well as the description of the uncertain model inputs. Section 3 presents the UQ exercise on the low-fidelity model. Section 4 describes the method used to perform a small number of high-fidelity model simulations and to evaluate the error between low-fidelity and high-fidelity outputs. Section 5 provides the final results of the UQ on the high-fidelity model by the way of metamodeling. A general conclusion stands in Section 6.

2 CFD modeling

All the CFD computations are performed using code_saturne[1], an open-source finite volume CFD solver developed at EDF R&D [29].

2.1 Description

Let consider, in a PWR, a fuel assembly blocked in the tube at a position t_x in the transfer tube open on the BR and BK pools at both ends ($t_x = 0$ being the center of the tube). The residual thermal power contained in the fuel assembly is noted P_d. No mass flow is imposed in the tube (it is the worst scenario): in this situation the fuel assembly cooling is only ensured by natural convection. Hence, a fluid circulation is established between the hot fuel assembly and the BR and BK pools, both maintained at a temperature T_{BR} and T_{BK} respectively.

2.2 Computational domain

The fluid volume consists in the full transfer tube and two arbitrary fluid volumes representing the BR and BK pools (see Fig. 2, left). Inside the transfer tube, an explicit representation of the element it contains is made as represented in Fig. 2 (right). Some hypotheses, sensible with the initial phenomena identification of the study, are made in order to simplify the meshing process:

[1] `https://www.code-saturne.org`

- The top and bottom of fuel assemblies are removed;
- The thermal conduction is neglected in the fuel assemblies (no meshing of the solid part);
- The containment part and the grids are treated as thin wall (which is coherent with their real slenderness);
- The mixing vanes are not represented on the mixing grids.

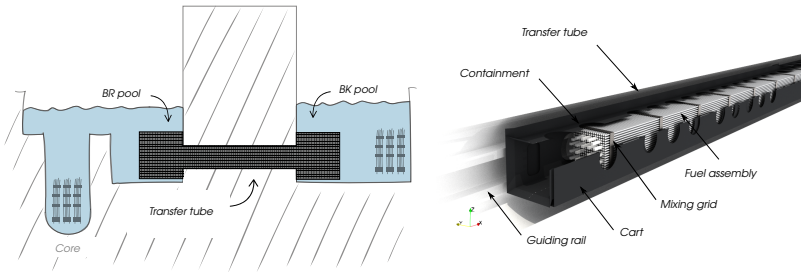

Figure 2: Sketch of the CFD fluid domain. Left: full domain. Right: elements represented in the transfer tube.

Free software Salome platform, with free meshing algorithms, was employed. Three mesh refinements were generated as illustrated in Figs. 3 and 4. These different meshes are noted R_1, R_2 and R_3 from the coarsest to the finest. The coarse R_1 mesh consists in about 8×10^6 cells and allows fast computations with little guarantees of precision and will be called "low fidelity". The finest mesh R_3 consists in 256×10^6 of cells and will be noted hereafter as "high fidelity". Mesh R_2, standing in between, consists in about 45×10^6 cells.

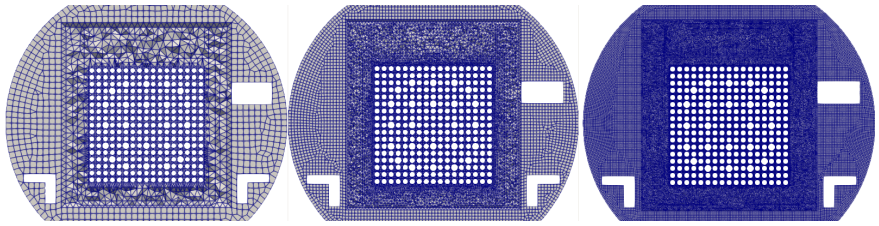

Figure 3: Cut of the CFD mesh in the transfer tube employed in this study. The vicinity of the fuel assembly and the outer regions of the tube are meshed with a grid invariant in the axis of the tube direction. The cart is meshed with isotropic unstructured cells (mainly tetrahedra). The whole mesh is fully conformal (from left to right : R_1, R_2 and R_3).

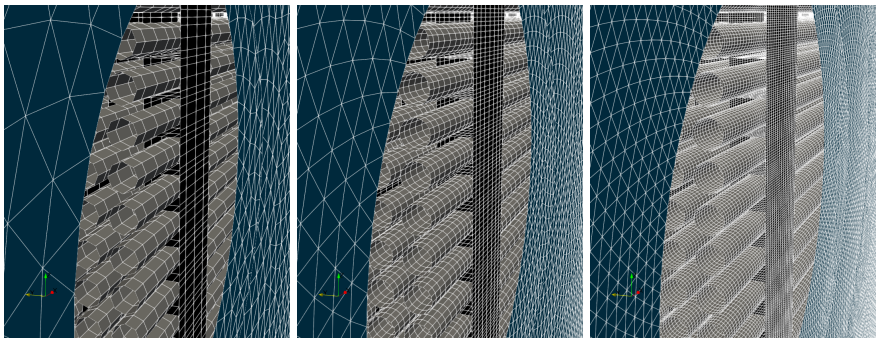

Figure 4: Meshing of the transfer cart, the tube and the grids (from left to right : R_1, R_2 and R_3)

2.3 Computational setup

The physical properties of water are dependent on the temperature and those laws are taken from IAPWS database (`www.iapws.org`). Constant temperature at T_{BR} and T_{BK} in the pools is ensured by a volumetric sink term in the heat equation at the top of each pool (in the last 10% of the pool height, all the energy produced by the fuel assembly is absorbed in this sink term). This models a steady state where all the thermal power dissipated in the fuel assembly has to be absorbed in the pools.

As we are not interested in the transitory part of the scenario, a steady management of the flow is chosen. This ensure a local CFL number of one everywhere in the domain, enabling a faster convergence towards the steady state solution.

Regarding turbulence, a Reynolds Average Navier Stokes (RANS) strategy is adopted. The k-ω SST model will be used. Despite its well known poor ability to predict natural convection flows, this model is able to deal with all type of mesh refinement at walls and show very good numerical stability, which made him the best candidates to perform hundreds of computations. For the sake of precision, a careful sensitivity to turbulence models was performed afterwards by the authors. This study exhibits a very small dependence on the turbulence model as most of the flow is *quasi* laminar where the head loss is the highest. This is coherent with the very low Reynolds number in the fuel assembly that can be estimated at approximately 400.

2.4 Uncertain input parameters

We propose to divide the uncertain parameter of this study in two categories: physical parameters and geometrical parameters.

2.4.1 Physical parameters

As already mentioned, the residual thermal power P_d, as well as the temperature of the BR and BK pool, T_{BR} and T_{BK}, are uncertain parameters. One also want to take into account the uncertainty in the axial distribution of the thermal power P_d. Indeed, noting $f(x)$ (W/m^2) the thermal flux along the fuel assembly, it should always verify

$$P_d = \iint_{S_{AC}} f(x)dx, \tag{1}$$

with S_{AC} the total heated surface of the fuel assembly.

Taking a constant heat flux $f = P_d/S_{AC}$ is however not fully representative of the reality. Indeed, a typical spent fuel with high decay heat (encountered in an unplanned stop of the plant) will show a bump around the middle of the fuel assemblies, with decreasing heat flux towards the top and the bottom. The position at which this bump occurs (hereafter noted x_p) as well as its intensity (hereafter noted f_p) is subject to uncertainty. This parameterization is illustrated in Fig. 5, where a typical spent fuel power distribution is compared to the present linear approximation resulting from the introduction of x_p and f_p.

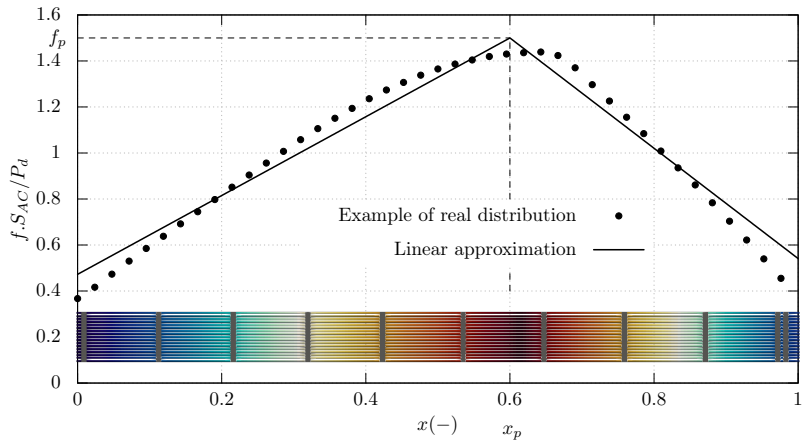

Figure 5: Typical spent fuel dimensionless power distribution compared to the parameterization used in this study (linear approximation with location (x_p) and intensity (f_p).

Table 1 summarizes all the uncertain physical parameters of the study. For sake of simplicity, parameters x_p and f_p are made dimensionless with respect to the length of the fuel assembly L and the constant heat flux value P_d/S_{AC}, respectively.

Parameter	Notation	Unit
BR pool temperature	T_{BR}	°C
BK pool temperature	T_{BK}	°C
Total thermal residual power	P_d	kW
Location of the thermal flux peak	x_p	-
Intensity of the thermal flux peak	f_p	-

Table 1: Uncertain physical parameters of the present study.

2.4.2 Geometrical parameters

The first obvious geometrical parameters is the blocked position of the fuel assembly as it can be locked at any position t_x inside the tube. We also decided to take the represented volume of the pool as an input parameter. Indeed, as we do not explicitly represent the whole pools, an arbitrary cubical volume is constructed at each side of the transfer tube (see Fig. 2, left). We note f_{BR} and f_{BK} the multiplicative coefficients applied to the side length of this cube. Table 2 summarizes all the uncertain geometrical parameters of the study.

Parameter	Notation	Unit
Locked position of the fuel assembly	t_x	m
Size factor for the BR pool representative volume	f_{BR}	-
Size factor for the BK pool representative volume	f_{BK}	-

Table 2: Uncertain geometrical parameters of the present study.

3 Uncertainty propagation on low-fidelity model

The peculiarity of CFD computations is its high requirements in terms of HPC. It is indeed common in the nuclear industry to tackle problems involving meshes with 50 millions to several hundreds millions of cells. The mean return time (the total elapsed time, from the beginning of the computation to the moment the engineer get the results) for those types of computation ranges from a day to several weeks. This makes CFD particularly not suitable, as such, for UQ which requires many evaluations of the same problem. That is why, as a first step, we conducted an UQ on the low fidelity mesh R_1 described previously. This mesh contains 8×10^6 cells and it takes around 8 hours on 144 cores of the EDF Cronos supercomputer (Atos BullSequana X system with 3,400 Intel Xeon Platinum 8260 processors) to reach a fully steady solution.

The design of experiments, the uncertainty propagation as well as the analysis of the results which are described in the next sections were performed using OpenTURNS[2], an open-source software platform developed by EDF R&D,

[2] `https://openturns.github.io/www/`

Airbus, ONERA, Phimeca and IMACS, and dedicated to the treatment of uncertainty [30].

3.1 Probabilistic representation of the problem

The uncertain parameters of the study being identified (seeSection 2.3), one needs a knowledge on their variability.

Physical parameters. The residual thermal power in a fuel assembly has been extensively studied in the nuclear community. A normal distribution of this power, with a mean value of 120 kW and a standard deviation σ of 5 kW, has been considered as a reasonable choice. The BR and BK pools are always maintained at a temperature below 50°C and theoretically never overshoot this temperature. However, for the purpose of this UQ exercise, we adopted a penalizing approach consisting in taking values given by a normal distribution centered on $\mu = 50$°C with a standard deviation σ of 1°C for both the BR and BK pools. Finally, the authors were not able to collect enough data to precisely estimate the variability in the residual thermal flux peak location and intensity. However, some observations allow to consider that it is:

- more likely to have light peak than a very sharp peak of the thermal flux,
- more likely to have a position of the peak in the upper part of the fuel assembly.

Geometrical parameters. We assume that all the geometrical parameters follows uniform probability distributions: there are no preferential values for those parameters.

Summary. Table 3 summarizes all the adopted probabilistic definition for the uncertain parameters of the study, where $\mathcal{N}(\mu, \sigma^2)$ is the normal probability distribution centered on μ with standard deviation σ, $\mathcal{N}(\mu, \sigma^2, a, b)$ is the normal probability distribution truncated on $[a, b]$, and $\mathcal{U}(a, b)$ is the uniform probability distribution on $[a, b]$.

Parameter	Notation	Probability density function
Total thermal residual power	P_d	$\mathcal{N}(120, 5^2)$
Location of the thermal flux peak	x_p	$\mathcal{N}(0.7, 0.2^2, 0.5, 0.9)$
Intensity of the thermal flux peak	f_p	$\mathcal{N}(1.3, 0.2^2, 1.0, 1.7)$
BR pool temperature	T_{BR}	$\mathcal{N}(50, 1)$
BK pool temperature	T_{BK}	$\mathcal{N}(50, 1)$
Locked position of the fuel assembly	t_x	$\mathcal{U}(-0.7, 0.7)$
Size factor for the BR pool representative volume	f_{BR}	$\mathcal{U}(1.0, 2.0)$
Size factor for the BK pool representative volume	f_{BK}	$\mathcal{U}(1.0, 2.0)$

Table 3: Uncertain parameters and their probabilistic models.

3.2 Output variables

As specified in Section 1, the main objective of this study is to ensure that temperature in the transfer tube remains far from the saturation temperature, $T_{max} \ll T_{sat}$. Several output variables may be looked at as for example the maximum temperature in the entire fluid domain. However, the latter parameter is highly sensitive to the wall refinement used in the mesh (the temperature is maximal at the wall around the fuel tube). For the purpose of this exercise, and to reduce this sensitivity, the maximal temperature 25cm above the fuel assembly in the tube is taken as the output variable of this study and is noted T_{max}.

3.3 Design of experiments generation

An optimized Latin Hypercube sampling (see, e.g., [31]), allowing to have a good coverage and space filling properties of the design points, was chosen to generate the numerical experiments involving our 8 parameters. It is noted $\mathcal{P}_{600}$. The number of evaluations of the CFD model was set to $N = 600$. It has been found that it offers the best compromise between the return time and the variability representation of the parameters.

As stated above, one evaluation of the CFD model has an approximate return time of 8h. Running 20 evaluations in parallel on the EDF supercomputer (i.e. using 2880 processors) leads to a total return time for the 600 evaluations of about 8 days.

3.4 Results

From a general physical point of view, a stratification is established in the tube. Figure 6 gives an illustration of the temperature field in the domain. One can clearly see the hot plumes coming from the fuel assembly and rising towards the top of the tube and then going towards the pool, which in turn will cool the fluid. As stated above, the maximum temperature in the tube above the fuel assembly is noted T_{max}.

Figure 6: Temperature field for one evaluation of the CFD model (in this view, the cart has been cut for the sake of visibility).

Table 4 presents the mean, the standard deviation and the 99% empirical quantile for this output variable after 600 evaluations on the design of experiments previously defined. The 95%/95% quantile, computed with the Wilks formula [10, 11], is also provided. Indeed, with the aim of increasing "safety", a conservative estimate of a quantile is often required. The Wilks method allows

to impose that the obtained estimate does have a certain level of confidence (we take here 95%) to overestimate the true (unkwnown) quantile.

Output	Mean μ	Standard deviation σ	99% quantile	95%/95% quantile
T_{max}	60.08	0.86	62.1	61.68

Table 4: Main statistical quantities of interest (in °C) for T_{max}.

Figure 7 gives the frequency histogram (probability density representation) for this output variable with the main statistical quantities of interest described above. As mentioned in the introduction, the accidental scenario studied here exhibits maximal temperature that are far from the saturation temperature (about 122°C, represented by the red line) and no boiling can occur in the tube.

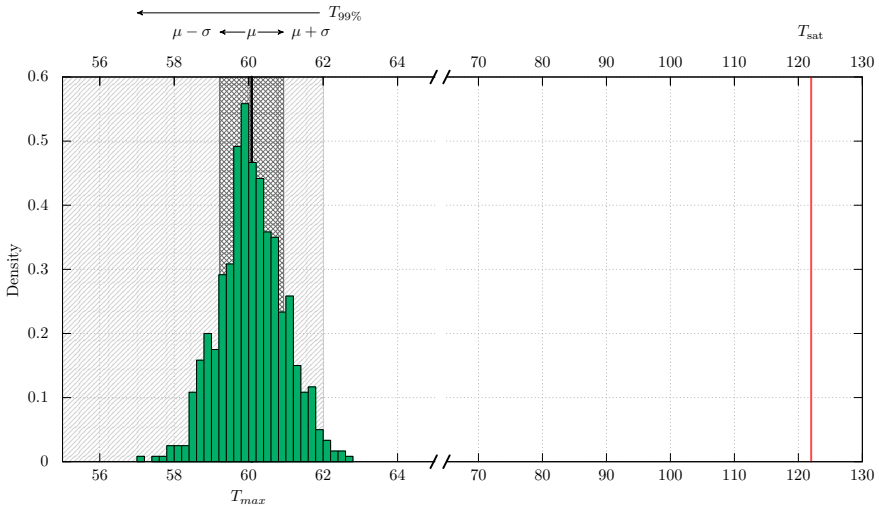

Figure 7: Probability density representation of the output variable T_{max} for the low fidelity model.

4 High fidelity modeling

It is well known in CFD that the choice made for the mesh (quality, refinement, etc.) has a huge impact on the results. As a consequence, the mesh is itself a parameter that can induce a strong variability in the distribution of the maximal temperature described in the previous section. It is however very difficult to include the mesh in the uncertainty propagation: the finest mesh that was generated ("high fidelity" mesh R_3) has 256×10^6 cells and one evaluation of the CFD model has a return time of about 4 days on 2880 processors. Running

the whole $\mathcal{P}_{600}$ design of experiments on such mesh is thus totally unreachable (6 years of computation with actual cluster limitation).

The strategy that is proposed in this study is to reduce the size of $\mathcal{P}_{600}$ to several well chosen points, so that the total number of CFD evaluations on finer meshes has a reasonable return time. Sampling the two codes of different fidelities at the same locations would allow to study precisely the difference between their output values (and then to evaluate their convergence). One important constraint is to ensure that the extracted sub-sample has the same distribution than the initial sample (then the inputs' probability density functions defined in Section 3.1 are kept). Another objective is to have a sub-sample with good space filling properties (see Section 3.3) as the initial sample. Therefore, for this particular purpose, we propose to use the support points technique which can easily be adapted to extract a sub-sample of a given size from a first sample.

4.1 High-fidelity design of experiments via support points

The support points method [27] allows to compact a continuous probability distribution F into a set of representative points (called support points). Among its different uses, it can extract a subsample from a given sample, for example for building a test sample in machine learning [32, 33, 34]. It is highly efficient in terms of computational cost, even for large-size sample N (up to $N = 10^4$) and in high input space dimension d (as large as $d = 500$). Mathematical formulas on support points and technical details on our extraction algorithm (called SPNN), are described in Appendix A.

Figure 8 illustrates in two cases the results of applying the SPNN algorithm, in comparison to applying a purely random extraction (Monte Carlo subsample). Through this example, the representativeness of the SPNN points is clearly illustrated:

- Red circles (given by SPNN algorithm) are located on all the different zones of the initial sample points (black crosses) and are well-spaced each other;
- Blue squares (given by pure random sampling) do not sample all the different zones of the initial sample points (black crosses) and can be close to each other.

In the present study, the SPNN algorithm has been used to generate two sets of points representative of the $\mathcal{P}_{600}$ design of experiments. The number of points of each subset is chosen so that the return time of the CFD model on mesh R_2 and R_3 is reasonable. A subset $\mathcal{P}_{50}$ with 50 points extracted from the initial distribution $\mathcal{P}_{600}$ is created for the evaluation of the model on mesh R_2, and a subset $\mathcal{P}_{10}$ with 10 points is thus created for the evaluation of the model on mesh R_3. As a consequence, those sets of points fulfill the following inclusions:

$$\mathcal{P}_{10} \subset \mathcal{P}_{50} \subset \mathcal{P}_{600}. \tag{2}$$

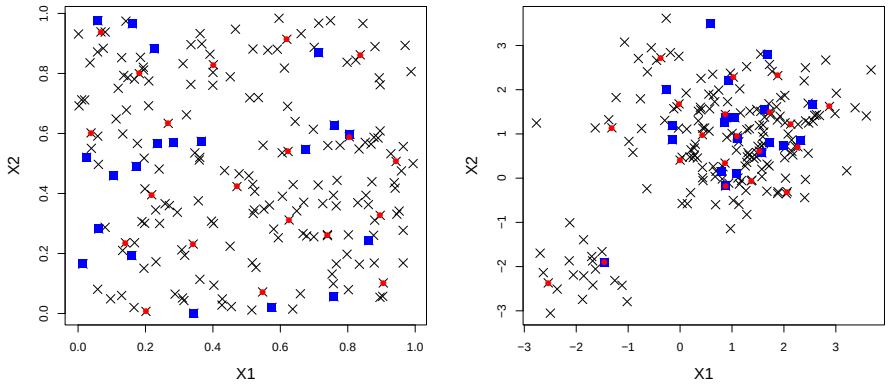

Figure 8: Illustrations of 20-size Monte Carlo samples (blue filled squares) and SPNN samples (red filled circles), both extracted from the same 200-size Monte Carlo samples (black crosses). Left: X_1 and X_2 follows the same uniform probability distribution; Right: X_1 and X_2 follows the same Gaussian mixture model.

Figure 9 illustrates these different nested design with the points' projection on two inputs.

The return time thus reached a total of 40 days on $\mathcal{P}_{10}$ for the evaluation of the CFD model on mesh R_3. A representation of those subsets for two parameters (T_{BR} and f_{BR}, see Tab. 3) is given on Fig. 9. The left figure illustrates all the points from $\mathcal{P}_{10}$, $\mathcal{P}_{50}$ and $\mathcal{P}_{600}$ and the right shows the good representativity of the $\mathcal{P}_{50}$ subset.

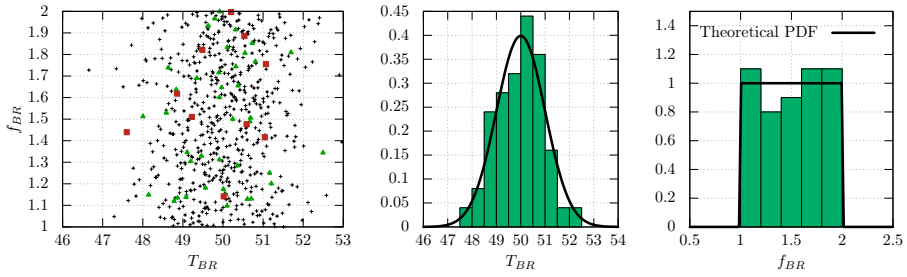

Figure 9: Point projections on the inputs (T_{BR}, f_{BR}) of the three designs of experiments $\mathcal{P}_{10}$ (red squares), $\mathcal{P}_{50}$ (green triangles) and $\mathcal{P}_{600}$ (black crosses) obtained from SPNN algorithm (left). Probability density functions of the two inputs from design $\mathcal{P}_{50}$ (middle and right).

4.2 Error quantification on the low-fidelity CFD model

Figure 10 presents a qualitative view of the CFD model for one set of uncertain parameters on all three meshes. One can notice differences, especially the location of the stratification or the intensity of the thermal plume above the fuel assembly. Figure 11 presents a more quantitative view of those differences as it draws a temperature profile above the fuel assembly (from the beginning of the cart to the end of the cart). It clearly appears that the hot peaks in the temperature are much higher on mesh R_2 and R_3 than the ones observed on the low-fidelity mesh R_1. Moreover, there are less differences between mesh R_2 and R_3.

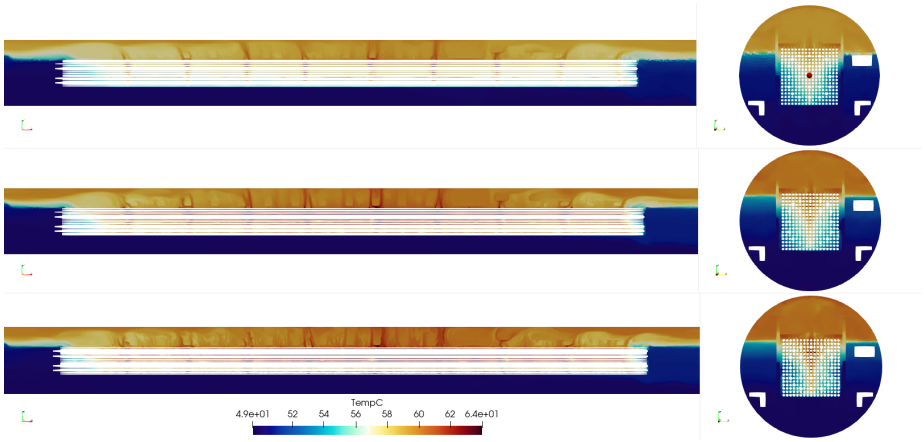

Figure 10: Transversal cut of the instantaneous temperature field for mesh R_1 (top), R_2 (middle) and R_3 (bottom).

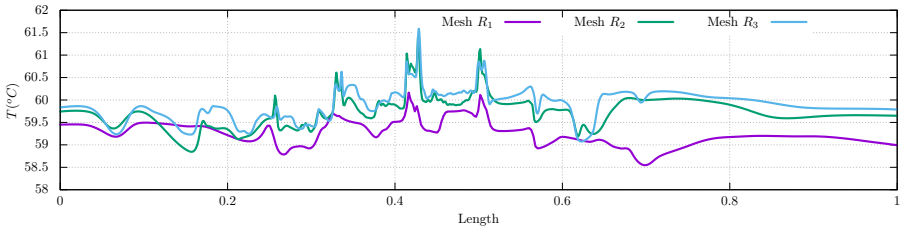

Figure 11: Temperature profile above the fuel assembly in the transfer tube.

To properly quantify this error, let us consider $\mathcal{P}$ the smallest design of experiments between meshes R_i and R_j. We introduce the two following standard metrics:

- **The bias on the temperature** between two meshes R_i and R_j, defined

as follows:

$$\mathcal{B}_T^{ij}(X) = T^{R_i}(X) - T^{R_j}(X), \quad X \in \mathcal{P}. \tag{3}$$

For example, we write $\mathcal{B}_{T_{max}}^{21}(X)$ the bias on T_{max} between meshes R_2 and R_1. We then write $\mathbb{E}\left(\mathcal{B}_T^{ij}(X)\right)$ the mean bias value for all the points of $\mathcal{P}$.

- **The Q^2 criterion on the temperature** between two meshes R_i and R_j, defined as follows:

$$Q_T^{2,ij}(X) = 1 - \frac{1}{N}\frac{\sum_{k=1}^{N}\left(T^{R_i}(X^{(k)}) - T^{R_j}(X^{(k)})\right)^2}{\mathrm{Var}(T^{R_i}(X))}, \quad X \in \mathcal{P}. \tag{4}$$

where the variance at the numerator is chosen to be computed on the finest mesh. It quantifies the loss of information between two meshes and their variability. The more the criterion is close to 1, the less information is lost going from the coarsest to the finest mesh.

Table 5 provides the results of these metrics for the maximum temperature T_{max}.The conclusions of those observations are threefold:

- The bias decreases as the mesh is refined. Going from 1.1°C between mesh R_1 and R_3 to 0.42°C between mesh R_2 and R_3;
- The variability of T_{max} decreases as the mesh is progressively refined;
- The loss of information between two meshes is minimized when going from mesh R_2 to R_3.

Meshing	$\mathbb{E}\left(\mathcal{B}_{T_{max}}(X)\right)$	$\sigma\left(\mathcal{B}_{T_{max}}(X)\right)$	$Q^2_{T_{max}}$
R_1/R_2	0.936	0.296	0.902
R_1/R_3	1.01	0.208	0.927
$\mathbf{R_2/R_3}$	**0.129**	**0.124**	**0.974**

Table 5: Metrics between all meshes for the output variable T_{max}.

This emphasizes that the mesh R_1 is not suitable to properly quantify the absolute value of T_{max}. However, the loss of information between mesh R_2 to R_3 is relatively small compared to going from mesh R_1 to R_2. From here, we can state that the low-fidelity mesh R_1 is not refined enough and the mesh R_2 is sufficiently refined to obtain precise results as it has small difference with the high-fidelity mesh R_3. This approach is a rather state-of-the-art procedure in CFD engineering to analyze the sensitivity of a model to the mesh and to use a fine enough mesh.

5 Uncertainty quantification on high fidelity modeling

Given the latter conclusion, how is possible to quantify uncertainty on a higher resolution model? To answer this question, the only available data are the evaluations of the CFD model on the $\mathcal{P}_{50}$ design of experiments. Those evaluations will be used to create a metamodel, which in turn will be used as a cheap CPU-time emulator of our model to propagate uncertainties on the $\mathcal{P}_{600}$ design of experiments.

5.1 Metamodel construction

The UQ of CPU-time expensive computer models can be done by approximating the computer model by a CPU-time inexpensive mathematical function called "surrogate model" or "metamodel". A metamodel can be based on polynomials, splines, Gaussian processes, random forests, neural networks, etc. [17], in fact on any machine learning techniques. Built from a set of computer code simulations, they must be as representative as possible of the code output values in the domain of variation of the uncertain parameters while having good prediction capabilities.

It was chosen in the present study to built a metamodel using a Gaussian process model (see Appendix B) to estimate T_{max}. This metamodel was designed using OpenTURNS software [30] and validated against 15 points (noted $\mathcal{T}_{15}$). Those 15 points were randomly chose among the 50 remaining points of a 100 points subplan which includes $\mathcal{P}_{50}$. To quantify the validation of the metamodel it is common to draw, for those validation points, the results of the metamodel against the values of the original model in a single plot as illustrated in Fig. 12. The Q^2 criterion (percentage of the output variable variance explained by the metamodel in prediction, see [33]) value reaches 90.79% for those data. This result is rather suitable for a subsequent uncertainty propagation study, as shown in [19]. This approach is a first work of metamodel-based UQ applied on a CFD codes. A more powerful approach, that will be applied in future studies, would be to build confidence-intervals around the quantile estimates using Gaussian process conditional simulations [19].

5.2 Uncertainty propagation using the metamodel

Given that we are now able to perform an evaluation of the metamodel at an extremely low CPU cost, it is possible to perform any evaluation with the precision of the high-fidelity CFD model. Table 6 compares, for T_{max}, the previous statistics resulting from the low-fidelity CFD model (coarse mesh R_1, see Tab. 4) to the ones coming from the high-fidelity metamodel. Both were evaluated on the $\mathcal{P}_{600}$ design of experiments in order to provide a precise comparison between the results. Figure 13 illustrates the results by giving the frequency histogram (probability density representation) for this output variable.

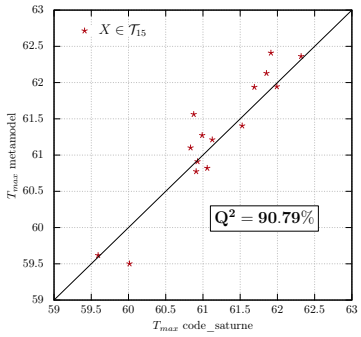

Figure 12: Metamodel validation for T_{max}.

Output	Mean μ	Standard deviation σ	99% quantile	95%/95% quantile
CFD model (evaluated on R_1)	60.08	0.86	62.1	61.68
Metamodel (built on R_2, evaluated on R_1)	61.01	0.93	63.04	62.67

Table 6: Comparison of the main statistical quantities of interest (in °C) for T_{max}.

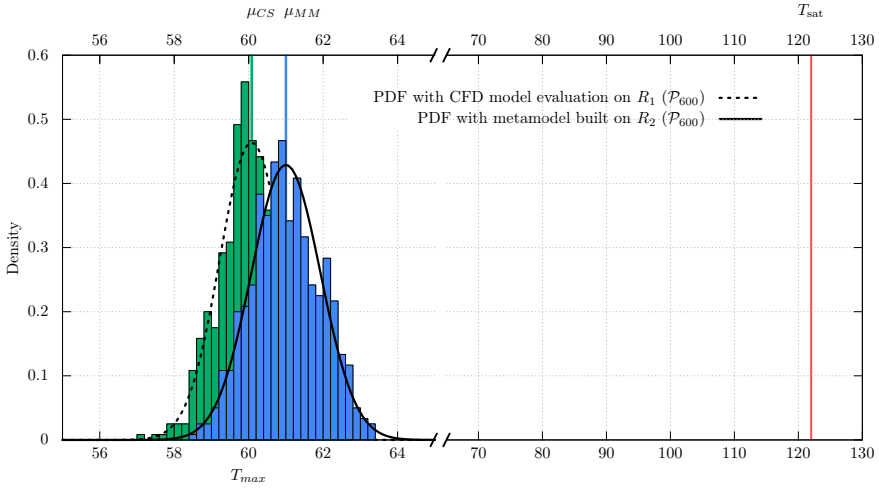

Figure 13: Probability density representation (green for the low-fidelity model and blue for the high-fidelity one) of the output variable T_{max}.

From Tab. 6 and Fig. 13, a bias of approximately one degree between the low fidelity data and the results of the high-fidelity metamodel is clearly

observed. This is consistent with the observations made in Section 4.2. It tends to prove that it is possible to capture the high fidelity resolution of mesh R_2 using a metamodel evaluated on the original design of experiments $\mathcal{P}_{600}$.

6 Conclusion

This paper studies the situation of a blocked fuel assembly in a PWR transfer tube. In order to refine results obtained with previous CFD computations, model inputs' uncertainties are propagated through the CFD model with a low-fidelity mesh. This mesh was the only one that ensures a reasonable return time. This first UQ exercise leads to the conclusion that temperature in the tube remains far from the defined saturation threshold.

To perform a global convergence study and quantify the error made in this first part, the initial design of experiments has been reduced using the SPPN algorithm. The CFD model was then evaluated on those reduced subsets. It emphasizes a constant bias between the low-fidelity mesh and the higher resolution meshes, proving that it is not possible to have good confidence in the first uncertainty propagation study.

In nuclear industrial studies, despite recent progresses both on hardware (CPU and GPU supercomputer get more powerful each year) and software (fully operational coupling between code_saturne and OpenTURNS), CFD is a domain where UQ remains highly challenging (mainly due to time cost issues), and then cannot be considered as a standard practice for industrial CFD applications. In this paper, it has been shown that metamodeling has potential to overcome this main issue in the near future. Indeed, results in building CPU-time inexpensive mathematical models (based on CFD computations) to quickly propagate uncertainty are very encouraging. Regarding the present use case, a metamodel was built on a high fidelity mesh and model inputs' uncertainties were propagated quickly on all points of the initial design of experiments with success.

Despite this difficulty, attempts to use UQ with CFD become more common in the literature. For example, a recent work [35] used physic-informed machine learning to quantify the closure relations of a physical model. Also recently, Wening et al [36], combined Reynolds Avergage Navier Stokes, Large Eddy Simulation and Direct Numerical Simulation to produce a mutli-fidelity model (built with Polynomial Chaos Expansion) and perform UQ on a buoyant flow in the framework of nuclear safety.

For sampling the two levels CFD models and solving the "optimal" subsample extraction (from a given sample) issue, this paper has shown that using the support points method is adequate. However, recent works [33, 28] have shown the efficiency of the kernel herding technique [37], which generalizes the mathematical concept of support points. Kernel herding is available in a module[3] of the OpenTURNS UQ platform [30], and then will be tested in future studies. Another avenue for improvement would be to use co-kriging models

[3] `https://efekhari27.github.io/otkerneldesign/master/index.html`

(see, e.g., [25]) to build the metamodel on the high-fidelity level code runs, by using all the information on the different low-fidelities code runs. This solution is known to be more efficient than just metamodeling each fidelity levels independently to each other.

Appendix A: Support points and SPNN algorithm

The support points approach [27] consists in compacting a continuous probability distribution F into a set of representative points, called support points. The construction of the support points is based on the optimization of the energy distance which is a particular case of the Maximum Mean Discrepancy criterion. This criterion provides a distance between F and a uniform distribution (via a kernel metric) and can be used with a relative good computational efficiency in high dimension. Let denote $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The discrete distribution of N_v support points $\mathbf{x}^{N_v} = (\mathbf{x}^{(i)})_{i=1\ldots N_v}$ is denoted F_{N_v} and the energy distance between F and F_{N_v} writes:

$$d_E^2(F, F_{N_v}) = \frac{2}{N_v} \sum_{i=1}^{N_v} \mathbb{E}\|\mathbf{x}^{(i)} - \zeta\| - \frac{1}{N_v^2} \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} \mathbb{E}\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\| - \mathbb{E}\|\zeta - \zeta'\| \quad (5)$$

with $\zeta, \zeta' \sim F$ and by using the Euclidean norm. The energy distance is always non-negative and equals zero if the two distributions are the same. The support points $(\xi^{(i)})_{i=1\ldots N_v}$ are then defined by minimizing $d_E^2(F, F_{N_v})$. Finding the support points corresponds to solving an optimization problem where F is empirically known by the sample points. To solve it, the objective function is approximated by a Monte Carlo estimate giving

$$(\xi^{(i)})_{i=1\ldots N_v} = \arg\min_{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(N_v)}} \left(\frac{2}{N_v n} \sum_{i=1}^{N_v} \sum_{k=1}^{n} \|\mathbf{x}^{(i)} - \mathbf{x}'^{(k)}\| - \frac{1}{{N_v}^2} \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} \mathbb{E}\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\| \right) \quad (6)$$

where $(\mathbf{x}'^{(k)})_{k=1\ldots n}$ is the n-size sample from F. This cost function can be written as a difference of convex functions in $\mathbf{x}^{N_v}$ and then can be minimized thanks to efficient optimization algorithms [27]. The examples given by [27] clearly show that support points distribution are more uniform than the ones of Monte Carlo and quasi-Monte Carlo samples [17].

In this procedure, the selected points are not extracted from the dataset but are the "best" points representative of the full dataset distribution. Therefore, an additional step is required in order to find the N_v representative points inside the dataset. For each support point, the nearest dataset point is selected. The whole algorithm is called SPNN ("support points nearest neighbor"). This algorithm is highly efficient in terms of computational cost, even for large-size test sample N (up to $N = 10^4$) and in high input space dimension d (as large as $d = 500$).

Appendix B: Basics on the Gaussian process metamodel

Gaussian process modeling [21], also called kriging model, treats a scalar output variable of a computer code $G(\mathbf{x})$ ($\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$) as a realization of a random function $Y(\mathbf{x})$, including a regression part and a centered stochastic process:

$$Y(\mathbf{x}) = h(\mathbf{x}) + Z(\mathbf{x}). \tag{7}$$

We consider n evaluations of the computer code at different inputs' values; these n points are called the experimental design and are denoted as $\boldsymbol{X}_s = (x^{(1)}, \ldots, x^{(n)})$. The output values are denoted as $Y_s = (y^{(1)}, \ldots, y^{(n)})$ with $y^{(i)} = G(\mathbf{x}^{(i)}) \; \forall \; i = 1 \ldots n$. The deterministic function $h(x)$ provides the mean approximation of the computer code. We can use for example a one-degree polynomial model:

$$h(x) = \beta_0 + \sum_{j=1}^{d} \beta_j x_j, \tag{8}$$

where $\beta = [\beta_0, \ldots, \beta_d]^t$ is the regression parameter vector. The stochastic part $Z(x)$ is a Gaussian centered stationary process fully characterized by its covariance function: $\mathrm{Cov}(Z(x), Z(u)) = \sigma^2 K_\theta(x - u)$, where σ^2 denotes the variance of Z, K_θ is the correlation function and $\theta \in \mathbb{R}^d$ is the vector of correlation hyperparameters. This structure allows to provide interpolation and spatial correlation properties. Several parametric families of correlation functions can be chosen [21, 38].

If a new point $\mathbf{x}^* = (x_1^*, \ldots, x_d^*) \in \mathbb{R}^d$ is considered, we obtain the predictor and variance formulas for the scalar output $Y(\mathbf{x}^*)$:

$$\mathbb{E}[Y(x^*)|Y_s] = h(x^*) + k(x^*)^t \boldsymbol{\Sigma}_s^{-1} (Y_s - h(\boldsymbol{X}_s)) \; , \tag{9}$$

$$MSE(x^*) = \mathrm{Var}[Y(x^*)|Y_s] = \sigma^2 - k(x^*)^t \boldsymbol{\Sigma}_s^{-1} k(x^*) \; , \tag{10}$$

with

$$\begin{array}{rcl} k(x^*) & = & [\mathrm{Cov}(y^{(1)}, Y(x^*)), \ldots, \mathrm{Cov}(y^{(n)}, Y(x^*))]^t \\ & = & \sigma^2 [K_\theta(x^{(1)}, x^*), \ldots, K_\theta(x^{(n)}, x^*))]^t \end{array} \tag{11}$$

and the covariance matrix

$$\boldsymbol{\Sigma}_s = \sigma^2 \left(K_\theta \left(x^{(i)} - x^{(j)} \right) \right)_{i=1\ldots n, j=1\ldots n} \; . \tag{12}$$

The conditional mean (Eq. (9)) is used as a predictor. The variance formula (Eq. (10)) corresponds to the mean squared error (MSE) of this predictor and is also known as the kriging variance. This analytical formula for MSE gives a local indicator of the prediction accuracy. More generally, Gaussian process model defines a Gaussian distribution for the output variable at any arbitrary new point. Regression and correlation parameters β, σ and θ have to be estimated (see the recent review of [38] describing the various optimization algorithms that can be used).

Acknowledgments

The authors acknowledge Romain Camy for its initial work on the subject, its accurate advise along this study and its comments on a first version of the paper. Many thanks also to Nicolas Merigoux for his support and whose project hosted this work.

References

[1] T. Skorek, A. de Crécy, A. Kovtonyuk, A. Petruzzi, R. Mendizábal, E. de Alfonso, F. Reventós, J. Freixa, C. Sarrette, M. Kyncl, R. Pernica, J. Baccou, F. Fouet, P. Probst, B-D. Chung, T.T. Tram, D-Y. Oh, A. Gusev, A. Falkov, Y. Shvestov, D. Li, X. Liu, J. Zhang, T. Alku, J. Kurki, W. Jäger, V. Sánchez, D. Wicaksono, O. Zerkak, and A. Pautz. Quantification of the uncertainty of the physical models in the system thermal-hydraulic codes - PREMIUM benchmark. *Nuclear Engineering and Design*, 354:110199, 2019.

[2] J. Baccou, J. Zhang, P. Fillion, G. Damblin, A. Petruzzi, R. Mendizábal, F. Reventós, T. Skorek, M. Couplet, B. Iooss, D-Y. Oh, and T. Takeda. Development of good practice guidance for quantification of thermal-hydraulic code model input uncertainty. *Nuclear Engineering and Design*, 354:110173, 2019.

[3] G.E. Wilson. Historical insights in the development of Best estimate Plus Uncertainty safety analysis. *Annals of Nuclear Energy*, 52:2–9, 2013.

[4] V. Mousseau and B.J. Williams. Uncertainty quantification in a regulatory environment. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncertainty Quantification*, pages 1613–1648. Springer, 2017.

[5] B. Iooss, V. Vergès, and V. Larget. BEPU robustness analysis via perturbed-law based sensitivity indices. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236:855–865, 2022.

[6] CSNI. CSNI Technical Opinion Paper No. 20 - Use of Computational Fluid Dynamics for Nuclear Safety. Technical Report NEA No. 7615, OECD/NEA, 2024.

[7] W.L. Oberkampf and T.G. Trucano. Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*, 38(3):209–272, 2002.

[8] R.C. Smith. *Uncertainty quantification*. SIAM, 2014.

[9] G.E. Karniadakis. Quantifying uncertainty in CFD. *Journal of Fluids Engineering*, 124:2–3, 2002.

[10] S.S. Wilks. Determination of sample sizes for setting tolerance limits. *Annals of Mathematical Statistics*, 12:91–96, 1941.

[11] J.P. Hessling and J. Uhlmann. Robustness of Wilks' conservative estimate of confidence intervals. *International Journal for Uncertainty Quantification*, 5:569–583, 2015.

[12] CSNI. Review of uncertainty methods for Computational Fluid Dynamics application to nuclear reactor thermal hydraulics. Technical Report NEA/CSNI/R(2016)4, OECD/NEA, 2016.

[13] F. D'Auria, N. Debrecin, and G.M. Galassi. Outline of the uncertainty methodology based on accuracy extrapolation (UMAE). *Nuclear Technology*, 109:21–38, 1995.

[14] A. Petruzzi and F. D'Auria. Approaches, relevant topics, and internal method for uncertainty evaluation in predictions of thermal-hydraulic system codes. *Science and Technology of Nuclear Installations*, 2008(Article ID 325071, 17 pages, DOI:10.1155/2008/325071), 2008.

[15] J.P. Hessling. Deterministic sampling for propagating model covariance. *SIAM/ASA Journal on Uncertainty Quantification*, 1:297–378, 2013.

[16] A.C. Rakhimov, D.C. Visser, and E.M.J. Komen. Uncertainty Quantification method for CFD applied to the turbulent mixing of two water layers - II: Deterministic Sampling for input uncertainty. *Nuclear Engineering and Design*, 348:146–158, 2019.

[17] K-T. Fang, R. Li, and A. Sudjianto. *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, 2006.

[18] J. Ko, D. Lucor, and J. garnier. Multi-layer growth response to inflow forcing with random phase shift. In *Proceedings of the ASME 2010 3rd Joint US-European Fluids Engineering Summer Meeting and 8th International Conference on Nanochannels, Microchannels, and Minichannels FEDSM-ICNMM2010*, Montreal, canada, august 2010. ASME.

[19] B. Iooss and A. Marrel. Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. *Nuclear Technology*, 205:1588–1606, 2019.

[20] Andrej Prošek, Boštjan Končar, and Matjaž Leskovar. Bepu method applied to cfd simulation of mixing flows. *Nuclear Technology*, 205(12):1661–1674, 2019.

[21] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[22] A. Marrel, B. Iooss, and V. Chabridon. The ICSCREAM methodology: Identification of penalizing configurations in computer experiments using screening and metamodel – Applications in thermal-hydraulics. *Nuclear Science and Engineering*, 196:301–321, 2022.

[23] A. Marrel and B. Iooss. Probabilistic surrogate modeling by Gaussian process: A new estimation algorithm for more robust prediction. *Reliability Engineering & System Safety*, 247:110120, 2024.

[24] ASME. Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer. Technical Report Standard No. V&V20-2009, American Society of Mechnical Engineers, New-York, 2009.

[25] Loic Le Gratiet. Bayesian analysis of hierarchical multifidelity codes. *SIAM/ASA Journal on Uncertainty Quantification*, 1:244–269, 2013.

[26] P.Z.G. Qian, M. Ai, and C.F.J. Wu. Construction of nested space-filling designs. *Annals of Statistics*, 37:3616–3643, 2009.

[27] S. Mak and V.R. Joseph. Support points. *The Annals of Statistics*, 46:2562–2592, 2018.

[28] E. Fekhari, V. Chabridon, J. Muré, and B. Iooss. Given-data probabilistic fatigue assessment for offshore wind turbines using bayesian quadrature. *Data Centric Engineering*, 5:e5, 2024.

[29] F. Archambeau, N. Méchitoua, and M. Sakiz. Code Saturne: A Finite Volume Code for the computation of turbulent incompressible flows - Industrial Applications. *International Journal on Finite Volumes*, 1(1), 2004.

[30] M. Baudin, A. Dutfoy, B. Iooss, and A-L. Popelin. Open TURNS: An industrial software for uncertainty quantification in simulation. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Springer Handbook on Uncertainty Quantification*, pages 2001–2038. Springer, 2017.

[31] G. Damblin, M. Couplet, and B. Iooss. Numerical studies of space filling designs: Optimization of Latin hypercube samples and subprojection properties. *Journal of Simulation*, 7:276–289, 2013.

[32] B. Iooss. Sample selection from a given dataset to validate machine learning models. In *Proceedings of 50th Meeting of the Italian Statistical Society (SIS2021)*, pages 88–93, Pisa, Italy, june 2021.

[33] E. Fekhari, B. Iooss, J. Muré, L. Pronzato, and J. Rendas. Model predictivity assessment: incremental test-set selection and accuracy evaluation. In N. Salvati, C. Perna, S. Marchetti, and R. Chambers, editors, *Studies in Theoretical and Applied Statistics, SIS 2021, Pisa, Italy, June 21-25*, pages 315–347. Springer, 2023.

[34] L. Pronzato and M-J. Rendas. Validation of machine learning prediction models. *New England Journal of Statistics in Data Science*, 1:394–414, 2023.

[35] Yang Liu, Nam Dinh, Xiaodong Sun, and Rui Hu. Uncertainty quantification for multiphase computational fluid dynamics closure relations with a physics-informed bayesian approach. *Nuclear Technology*, 209(12):2002–2015, 2023.

[36] Philipp J. Wenig, Stephan Kelm, and Markus Klein. Cfd uncertainty quantification using stochastic spectral methods—exemplary application to a buoyancy-driven mixing process. *Nuclear Engineering and Design*, 409:112317, 2023.

[37] Y. Chen, M. Welling, and A. Smola. Super-samples from kernel herding. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 109 – 116. AUAI Press, 2010.

[38] A. Marrel and B. Iooss. Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation. *Reliability Engineering & System Safety*, 247:110094, 2024.