HAL
open science

Finite mixture models for sensitivity analysis of thermal hydraulic codes for passive safety systems analysis

Francesco Di Maio, Giancarlo Nicola, Enrico Zio, Yu Yu

To cite this version:

Francesco Di Maio, Giancarlo Nicola, Enrico Zio, Yu Yu. Finite mixture models for sensitivity analysis of thermal hydraulic codes for passive safety systems analysis. 2016. hal-01265884v1

HAL Id: hal-01265884

https://hal.science/hal-01265884v1

Preprint submitted on 1 Feb 2016 (v1), last revised 16 Jul 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FINITE MIXTURE MODELS FOR SENSITIVITY ANALYSIS OF THERMAL HYDRAULIC CODES FOR PASSIVE SAFETY SYSTEMS ANALYSIS

Francesco Di Maio[1], Giancarlo Nicola[1], Enrico Zio[1,2], Yu Yu[3]

[1] *Energy Department, Politecnico di Milano*

Via La Masa 34, 20156 Milano, Italy

francesco.dimaio@polimi.it

[2] *Chair on System Science and Energetic Challenge*

Fondation EDF, Ecole Centrale Paris and Supelec, Paris, France

[3] *School of Nuclear Science and Engineering,*

North China Electric Power University, Beijing, China, 102206

ABSTRACT

For safety analysis of Nuclear Power Plants (NPPs), Best Estimate (BE) Thermal Hydraulic (TH) codes are used to predict system response in normal and accidental conditions. The assessment of the uncertainties of TH codes is a critical issue for system failure probability quantification. In this paper, we consider passive safety systems of advanced NPPs and present a novel approach of Sensitivity Analysis (SA). The approach is based on Finite Mixture Models (FMMs) to approximate the probability density function (i.e., the uncertainty) of the output of the passive safety system TH code with a limited number of simulations. We propose a novel Sensitivity Analysis (SA) method for keeping the computational cost low: an Expectation Maximization (EM) algorithm is used to calculate the saliency of the TH code input variables for identifying those that most affect the system functional failure. The novel approach is compared with a standard variance decomposition method on a case study considering a Passive Containment Cooling System (PCCS) of an Advanced Pressurized reactor AP1000.

1 INTRODUCTION

Passive systems seem to offer several advantages in terms of safety and reliability [Nayak et al. 2009] [Zio et al. 2010]. For this reason, they are considered in the design of innovative and advanced nuclear reactor concepts.

To quantify the reliability of passive safety systems the concept of *functional failure* has been introduced [Cummins et al. 2003] [Pagani et al. 2005] [Burgazzi et al 2007] [Zio et al. 2009]: when counter-forces (e.g., friction) have magnitude comparable to the driving ones (e.g., gravity, natural circulation), the passive systems may fail performing the intended functions even if i) safety margins are met, ii) no hardware failures occur [Burgazzi 2004] [Marques et al. 2005] [Burgazzi 2007b] [Zio et al. 2009].

Thermal Hydraulics (TH) codes are used to predict the physical behavior of the system in nominal and accidental conditions. Conservative TH codes have traditionally been employed to verify that safety limits could be respected with large safety margins [Zio et al. 2010]. Best Estimate (BE) codes have been introduced more recently to provide more realistic results and avoid over-conservatism [Zio et al. 2010] [10 CFR 50.46]; their use requires the identification and quantification of the uncertainties in the code outputs coming from simplifications, approximations, round-off-errors, numerical techniques and uncertainties in the input variable values [Pourgol-Mohammad 2009]. The quantification of the uncertainties in the output can be demanding in terms of computational cost, because it requires a large number of simulations of the BE-code [de Crécy et al. 2008; Di Maio et al. 2014a].

To tackle this challenge, various approaches of Uncertainty Analysis (UA) have been developed, e.g., Code Scaling, Applicability, and Uncertainty (CSAU) [Boyack et al. 1990] [Wilson et al. 1990] [Wulf et al. 1990], Automated Statistical Treatment of Uncertainty Method (ASTRUM), Integrated Methodology for Thermal-Hydraulics Uncertainty Analysis (IMTHUA) [Pourgol-Mohammad 2009], which assume that the uncertainty in the input variables follows a statistical distribution: N input sets are sampled and fed to the BE code and the corresponding N output values are calculated, which reflect the variability of the input variables onto the output. A combination of Order Statistics (OS) [Guba et al. 2003] [Zio et al. 2008] and Artificial Neural Networks (ANN) has been proposed to speed up computation (substituting the TH code with a simpler and faster surrogate) [Secchi et al. 2008]. However, this latter approach does not allow to completely characterize the pdf of the output variable (but only some percentiles), precluding the possibility of: i) obtaining a precise estimate of the safety limit; ii) performing Sensitivity Analysis (SA) at no extra computational cost [Langewisch 2010] [Hong et al. 2011].

Among the SA techniques, it is possible to identify three families: Local, Regional and Global [Saltelli et al. 2000]. The local approach to SA consists in evaluating the effect on the output of small variations around fixed values in the input parameters. Typically, local methods involve the calculation of partial derivatives of the output with respect to the inputs at local fixed points on which the analysis is focused. The local sensitivity indexes obtained, provide information that is valid only locally. Regional SA aims at calculating the sensitivity of the model to partial ranges of the inputs distributions [Pengfei 2014]. Global SA allows to measure the contribution of an input to the variability of the output over the entire range of both the input and the output. To do this, the approach focuses directly on the output and its uncertainty distribution, with no reference to any particular value of the input parameters (unlike local approaches). Global SA is most indicated when models are non-linear and non-monotone, as in these cases local and regional SA cannot provide general results. Compared to local and regional SA methods, global SA methods offer higher capabilities, but these are paid by high computational costs. Examples of global methods are Response Surface Methodology (RSM), Fourier Amplitude Sensitivity Test (FAST) and the variance decomposition method [Helton 1993] [McKay 1996] [Saltelli et al. 2000] [Cadini et al. 2007] [Yu et al. 2010]. RSM consists in approximating the model by a simple and faster mathematical model from a database of computations [Devictor et al. 2005]; with FAST, the model can be expanded into a Fourier series and the Fourier coefficients and frequencies can be used to estimate mean and variance of the model, and the partial variance of individual input parameters of the model [Fang et al. 2003]; variance decomposition is a general and solid method for global SA and has the advantage of not introducing any hypothesis on the model, although it has high computational costs [Carlos et al. 2013].

In this paper, we focus on global SA methods based on the pdf of the output variable and propose a novel alternative to the existing methods. Among these, Polynomial Chaos Expansion (PCE) methods have been used to reconstruct the pdf of the output variable and for SA, with less runs than variance decomposition-based methods [Sudret 2008] [Eaton et al. 2010] [Gilli et al., 2012]. However, in many cases the output variable follows a multimodal distribution for which PCE is unsuitable because the order of the expansion necessary for accurately reconstructing the pdf becomes large and the computational cost too [Nouy, 2010]. We overcome the problem of multimodal distributions by resorting to Finite Mixture Models (FMMs) [McLachlan et al. 2000], which provide, by application of an Expectation Maximization algorithm (EM), a natural “clustering” of the TH code output (e.g., subdividing the data in groups of large safety margin, low safety margin, failure) based on the models composing the mixture. Such models can be used for Sensitivity Analysis, aiming at identifying the most relevant input variables affecting the output

uncertainty, as we shall see in the following. More specifically, in this paper, Gaussian FMMs are used to reproduce the pdf of the TH code output and its natural clustering is originally exploited for SA (Figure 1). The advantages of this approach are i) the possibility to obtain the analytical pdf of the model output and ii) a computational cost for SA significantly lower than classical global methods.

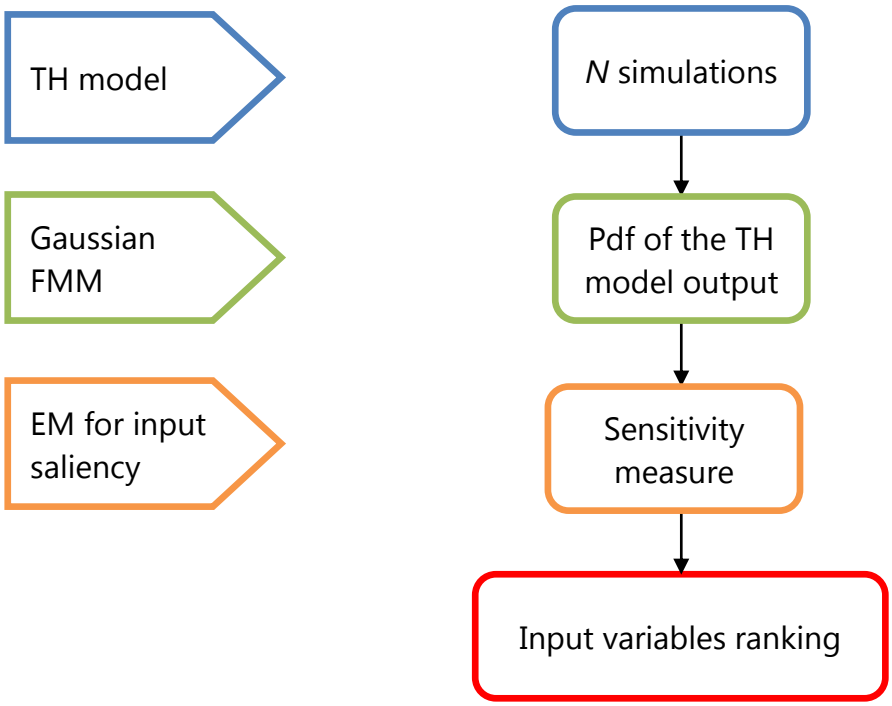

Figure 1 Flowchart of the proposed framework for SA

The paper is organized as follows. In Section 2, FMM are presented and the input saliency approach for global SA is presented. In Section 3, the TH code and the relative case study are illustrated. In Section 4, the experimental results are reported along with the comparison to variance decomposition-based SA results. Section 5 draws some conclusions.

2 FINITE MIXTURE MODELS FOR SENSITIVITY ANALYSIS

Let y denote the output of a TH model m, viz:

$$y = m(x_1, x_2, \dots, x_l, \dots, x_D) \qquad l = 1, \dots, D \tag{1}$$

where x_l is the l-th input variable. The random output variable y follows a finite mixture density $f(y)$ with K models if:

$$f(y) = \sum_{k=1}^{K} \pi_k f_k(y|\theta_k) \tag{2}$$

where $f_k(y|\theta_k)$ are K different probability density functions, θ_k is the set of parameters of the k-th model of the mixture and π_k are the mixing probabilities that satisfy:

$$\sum\nolimits_k \pi_k = 1 \qquad with\ \forall\, k, \pi_k \geq 0 \tag{3}$$

In particular, if $f_k(y|\theta_k)$ is Gaussian, then:

$$f_k(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(y-\mu_k)^2}{2\sigma_k^2}} \tag{4}$$

where $\theta_k = (\mu_k, \sigma_k)$ are the mean and the standard deviation of the k-th gaussian mixture model, respectively.

For simplicity of illustration, and without loss of generality, let us consider a mixture of two Gaussians:

$$f(y) = \pi_1 f_1(y|\theta_1) + \pi_2 f_2(y|\theta_2) \tag{5}$$

Expectation Maximization (EM) algorithm [Dempster et al. 1977] [McLachlan et al. 2000] can be used to fit $f(y)$ to N available data $y = \{y_1, \dots, y_N\}$, $i = 1, \dots, N$ and identify its parameters $\theta = (\theta_1, \theta_2)$ and $\pi = (\pi_1, \pi_2)$. To do that, we resort to two classification variables z_{1i}, z_{2i} i.e., $(z_{1i} + z_{2i} = 1)$ that assign one among the two models to a data point y_i:

$$z_{1i} = \begin{cases} 1 \; if \; y_i \; follows \;\; f_1(y|\theta_1) \\ 0 \; if \; y_i \; follows \;\; f_2(y|\theta_2) \end{cases} \qquad z_{2i} = \begin{cases} 0 \; if \; y_i \; follows \;\; f_1(y|\theta_1) \\ 1 \; if \; y_i \; follows \;\; f_2(y|\theta_2) \end{cases} \tag{6}$$

with $\theta = (\theta_1, \theta_2)$.

For each *i*-th datum, the conditional probabilities in Eqs. (7) and (8) hold:

$$P(y_i|z_{1i}, z_{2i}, \theta) = f_1^{z_{1i}}(y_i) f_2^{z_{2i}}(y_i) \tag{7}$$

$$P(z_{1i}, z_{2i}|\theta) = \pi_1^{z_{1i}}(1-\pi_1)^{z_{2i}} \tag{8}$$

Substituting Eqs. (7) and (8) into Eq. (9):

$$P(y_i, z_{1i}, z_{2i}|\theta) = P(y_i|z_{1i}, z_{2i}, \theta) P(z_{1i}, z_{2i}|\theta) \tag{9}$$

and taking its logarithm,

$$\log\big(P(y_i, z_{1i}, z_{2i}|\theta)\big) = z_{1i}\log\big(f_1(y_i)\big) + z_{1i}\log(\pi_1) + z_{2i}\log\big(f_2(y_i)\big) + z_{2i}\log(1-\pi_1) \tag{10}$$

The likelihood function for all the *N* data can be written as:

$$\begin{aligned} &L(y, z|\theta) = \log\big(P(y, z|\theta)\big) = \\ &\sum_{i=1}^{N} z_{1i}\log\big(f_1(y_i)\big) + z_{1i}\log(\pi_1) + z_{2i}\log\big(f_2(y_i)\big) + z_{2i}\log(1-\pi_1) \end{aligned} \tag{11}$$

The maximum likelihood of $L(y, z|\theta)$ cannot be found analytically; this is why we resort to an Expectation Maximization (EM) algorithm for the identification of the model parameters $\theta(\theta_1, \theta_2)$ and $\pi(\pi_1, \pi_2)$, with an initial random estimation of $z, z^{(1)}$:

$$\mu_1^{(1)} = \frac{\sum_{i=1}^{N} z_{1i}^{(1)} y_i}{\sum_{i=1}^{N} z_{1i}^{(1)}}; \qquad \mu_2^{(1)} = \frac{\sum_{i=1}^{N} z_{2i}^{(1)} y_i}{\sum_{i=1}^{N} z_{2i}^{(1)}} \tag{12}$$

$$\sigma_1^{2(1)} = \frac{\sum_{i=1}^{N} z_{1i}^{(1)} (y_i - \mu_1^{(1)})^2}{\sum_{i=1}^{N} z_{1i}^{(1)}}; \qquad \sigma_2^{2(1)} = \frac{\sum_{i=1}^{N} z_{2i}^{(1)} (y_i - \mu_2^{(1)})^2}{\sum_{i=1}^{N} z_{2i}^{(1)}} \tag{13}$$

$$\pi_1^{(1)} = \frac{\sum_{i=1}^{N} z_{1i}^{(1)}}{n}; \qquad \pi_2^{(1)} = \frac{\sum_{i=1}^{N} z_{2i}^{(1)}}{n} = 1 - \pi_1^{(1)} \tag{14}$$

The expectation step follows by application of Bayes rule [McLachlan 2008]:

$$z_{1i}^{(j)} = P\left(z_{1i} = 1 \middle| \theta^{(j-1)}, y_i\right) = \frac{\pi_1^{(j-1)} f_1(y_i,\ \theta^{(j-1)})}{\pi_1^{(j-1)} f_1(y_i,\ \theta^{(j-1)}) + (1 - \pi_2^{(j-1)}) f_2(y_i,\ \theta^{(j-1)})} \tag{15}$$

$$z_{2i}^{(j)} = P\left(z_{2i} = 1 \middle| \theta^{(j-1)}, y_i\right) = \frac{(1 - \pi_2^{(j-1)}) f_2\left(y_i,\ \theta^{(j-1)}\right)}{\pi_1^{(j-1)} f_1(y_i,\ \theta^{(j-1)}) + (1 - \pi_2^{(j-1)}) f_2(y_i,\ \theta^{(j-1)})} \tag{16}$$

The maximization step follows the expectation step: θ^j and π^j are updated to find, for any *j*-th step, the optimum. This continues until the stopping criterion is reached (i.e. μ and σ do not change in two successive iterations) [Figueiredo et al. 2002]:

$$\mu_1^{(j)} = \frac{\sum_{i=1}^{N} z_{1i}^{(j)} y_i}{\sum_{i=1}^{N} z_{1i}^{(j)}}; \qquad \mu_2^{(j)} = \frac{\sum_{i=1}^{N} z_{2i}^{(j)} y_i}{\sum_{i=1}^{N} z_{2i}^{(j)}} \tag{17}$$

$$\sigma_1^{2(j)} = \frac{\sum_{i=1}^{N} z_{1i}^{(j)} (y_i - \mu_1^{(j)})^2}{\sum_{i=1}^{N} z_{1i}^{(j)}}; \qquad \sigma_2^{2(j)} = \frac{\sum_{i=1}^{N} z_{2i}^{(j)} (y_i - \mu_2^{(j)})^2}{\sum_{i=1}^{N} z_{2i}^{(j)}} \tag{18}$$

$$\pi_1^{(j)} = \frac{\sum_{i=1}^{N} z_{1i}^{(j)}}{n}; \qquad \pi_2^{(j)} = \frac{\sum_{i=1}^{N} z_{2i}^{(j)}}{n} = 1 - \pi_1^{(j)} \tag{19}$$

Once the parameters $\theta(\theta_1, \theta_2)$ and $\pi(\pi_1, \pi_2)$ of the mixture models are known, the best approximation of the pdf of the output of the TH model is completely characterized with a small number *N* of TH code simulations. In addition, "natural" clusters corresponding to each Gaussian model $f_k(y|\theta_k)$ of the mixture are defined: some may be representative of normal conditions, whereas others of accidental conditions, allowing for a direct calculation of the probability of exceeding a certain safety limit (i.e., of functional failure). In fact, the area beneath the model

representation of the cluster of accidental conditions can be directly quantified (being the area below a weighted Gaussian distribution as in Eq. (5)) and used as functional failure probability.

For global sensitivity analysis, the FMM of Eq. (2) can be rewritten as a function of the *D* input variables of the TH model, if we assume input variables independence:

$$f(y) = \sum_{k=1}^{K} \pi_k f_k(y|\theta_k) = \sum_{k=1}^{K} \pi_k m\left(\prod_{l=1}^{D} f(x_l|\theta_{kl})\right) \tag{20}$$

where *m* is the TH model function and $f(x_l|\theta_{kl})$ is the pdf of the *l*-th input in the *k*-th cluster. The *l*-th input does not affect the output if its distribution is independent from the cluster, i.e., it follows a common density among all the clusters, denoted by $q(x_l|\lambda_l)$ [Pudil et al. 1995] [Vaithyanathan et al. 1999]. In Eq. (20), $f(x_l|\theta_{kl})$ can be decomposed in a distribution accounting for the contribution of the *l*-th input in the *k*-th cluster $f(x_l|\theta_{kl})$ and in the common distribution $q(x_l|\lambda_l)$, with weights ρ_l, obtaining:

$$f(y|\theta) = \sum_{k=1}^{K} \pi_k m\left(\prod_{l=1}^{D} \rho_l f(x_l|\theta_{kl}) + (1-\rho_l) q(x_l|\lambda_l)\right) \tag{21}$$

We consider the *saliency* ρ_l as the importance of the *l*-th input in affecting the output *y*. In fact, if ρ_l is large it means that the input variable distribution varies significantly from one cluster to another and, thus, the input is important in determining the variability of the output; otherwise, if ρ_l is small the inputs follow the common distribution in any cluster and, thus, the input is not relevant in shaping the distribution of the output. For example, Figure 2 shows the FMM decomposition of $f(y)$ in case of two input variables x_1 and x_2: x_1 contributes in shaping the model output $f(y)$ with $f(x_1|\mu_{11}, \sigma_{11})$ and $f(x_1|\mu_{21}, \sigma_{21})$, whereas x_2 only follows its common distribution $q(x_2|\lambda_2)$.

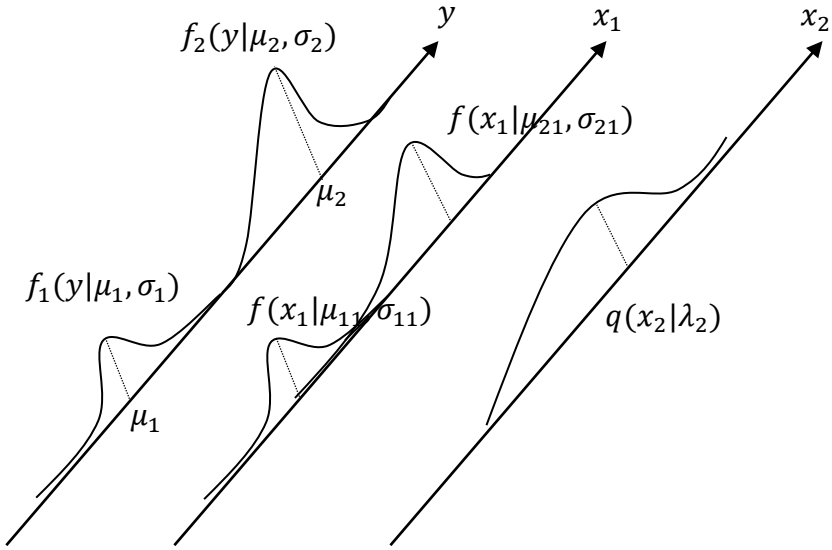

Figure 2 Mixture model decomposition illustrative example

The estimation of the input variable importance ρ_l is a model parameter identification problem that does not admit any closed form analytical solution [Figueiredo et al. 2002]. The problem can again be tackled by the EM algorithm, fitting Eq. (21) to data. In this case, for a FMM with $K = 2$, parameters $\theta(\theta_1, \theta_2)$ and $\pi(\pi_1, \pi_2)$ have already been identified with Eqs. (17), (18) and (19), whereas $\rho(\rho_1, \rho_2, \dots, \rho_l, \dots, \rho_D)$ is initially estimated to $\rho^{(1)}$ and updated at each following *j*-th step as:

$$\rho_l^{(j)} = \frac{\sum_{i,k} u_{ikl}^{(j)}}{N} \tag{22}$$

where

$$u_{ikl}^{(j)} = \frac{a_{ikl}^{(j)}}{(a_{ikl}^{(j)} + b_{ikl}^{(j)})} w_{ik}^{(j)} \tag{23}$$

measures how important the *i*-th datum is in the *k*-th model (cluster), when the *l*-th input is considered,

$$a_{ikl}^{(j)} = P(\, x_{il}|z_{ki} = 1, f(x_{il}|\theta_{kl})) = \rho_l^{(j)} f(x_{il}|\theta_{kl}) \tag{24}$$

is the probability that the *l*-th input of the *i*-th code run belongs to the *k*-th cluster

$$b_{ikl}^{(j)} = P\big(\, x_{il}\big|z_{ki} = 1, q(x_{il}|\lambda_l)\big) = (1 - \rho_l^{(j)}) q(x_{il}|\lambda_l) \tag{25}$$

is the probability that the *l*-th input of the *i*-th code run does not belong to any cluster, and

$$w_{ik}^{(j)} = P(z_{ki} = 1|y_i) = \frac{\pi_k \prod_l (a_{ikl}^{(j)} + b_{ikl}^{(j)})}{\sum_k \pi_k \prod_l (a_{ikl}^{(j)} + b_{ikl}^{(j)})} \tag{26}$$

is the probability that the output of the *i*-th code run belongs to the *k*-th cluster.

It is worth noticing that the term $\sum_{i,k} u_{ikl}$ in Eq. (22) represents the contribution of the *l*-th input to the definition of all *K* models (clusters) when supported by *N* evidences, and thus ρ_l can be considered as a sensitivity index for the *l*-th input.

3 CASE STUDY

The Westinghouse AP1000 is a 1117 MWe (3415 MWth) pressurized water reactor (PWR), with extensive implementation of passive safety systems for reduction of corrective actions in case of accident. The passive safety systems include the passive Residual Heat Removal System (RHRS) and the Passive Containment Cooling System (PCCS). The PCCS cools the containment following an accident, so that the pressure is effectively controlled within the safety limit of a manometric pressure of 0.4 MPa. During an accident, heat is removed from the containment vessel by the continuous, natural circulation of air, supplemented by evaporation of the water that drains by gravity from a tank located on top of the containment shield building by means of three redundant and diverse water drain valves. The steel containment vessel provides the heat transfer surface through which heat is removed from inside the containment and transferred to the atmosphere. In addition, even in case of failure of water drain, air-only cooling is supposed to be capable of maintaining the containment below the failure pressure [Schulz 2006]. Figure 3 shows the PCCS of the AP1000 [Westinghouse Electric Company].

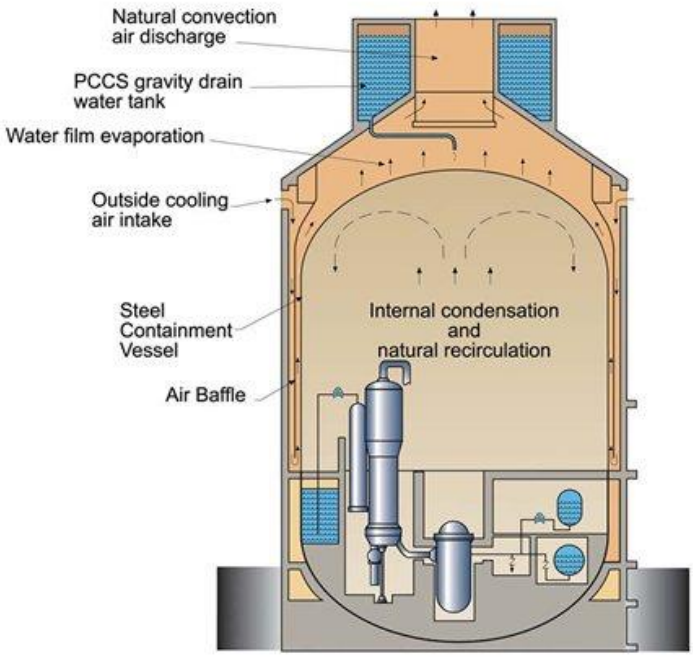

Figure 3 AP1000 Passive Containment Cooling System [Westinghouse Electric Company www.westinghousenuclear.com]

For the quantification of the functional failure of the PCCS of the AP1000, a simplified and lumped-parameter TH model for heat transfer with non-condensed gas has been developed by [Yu et al. 2013; Yu et al. 2013b] and here used for quantifying the response capabilities of the system following a Loss Of Coolant Accident (LOCA), that is one among the design basis accidents for AP1000 reactor design.

A TH code for simulating a LOCA follows the phenomena evolution typically in four phases [Rahim et al. 2011]: 1) blowdown, from the accident initiation (by a double-ended guillotine pipe break in a primary coolant line affecting the normal operation of the reactor at steady-state full power) to the time at which the primary circuit pressure reaches the containment pressure; 2) refill, from the end of the blowdown to the time when the Emergency Core Cooling System (ECCS) refills the vessel lower plenum; 3) reflood, which begins when water starts flooding the core and ends when this is completely quenched; 4) post-reflood, which starts after the core quenching and during which energy is released to the Reactor Coolant System (RCS).

In the post-reflood phase, the steam produced in the RCS is cooled at the internal face of the steel containment vessel and, then, the heat is conducted by the vessel and transferred to the air in the air channels (see Figure 3). Cold air enters the channels through the three rows of air inlets and flows down to the bottom of the channels, where it is heated by the steel vessel up to the air diffuser to the environment.

The TH code is here used to analyze the effects of air temperature and reactor power on the PCCS function at steady state during the post-reflood phase. It is worth pointing out that this TH model has been solely used for the purpose of demonstration of the effectiveness of the proposed UA and SA methods in reducing the computational burden of the analysis. Therefore, the intent is not the demonstration and qualification of the TH model itself, but rather the presentation of a systematic approach for SA that is here verified with respect to this benchmark case study, elsewhere analyzed with other approaches [Di Maio et al. 2014 a; Di Maio et al. 2014c; Yu et al,, 2015]. The input variables used for calculating the PCCS capability of condensing the steam produced are listed in Table 1, together with their distributions chosen based on expert judgment and literature review [Burgazzi 2004], [Zio et al. 2008b], [Zio et al. 2010b]. Three distributions have been used: seasonal, normal and uniform. Seasonal relates to the external air temperature T_{inlet} and pressure P_{air} variability, as inferred by historical data collected by a representative Chinese Automatic Weather Station (CAWS) in different months. Normal distributions, e.g., for the LOCA steam temperature, T_{steam}, are truncated distributions with mean μ and support equal to 4σ where σ is the standard deviation. For uniform distributions, e.g. for the steam mass flow rate G, the interval supports from "Lower value" to "Upper value" are reported.

Table 1 List of the input variables of the TH code

	Input variable	Description	Unit	Type of distribution	Lower value	Upper value
1	G	Steady state LOCA mass flow rate	kg/s	*uniform*	6	11
2	T_{inlet}	External air temperature	°C	*seasonal*	2	39
3	P_{air}	Inlet air pressure	MPa	*seasonal*	0.0984	0.1011

	Input variable	Description	Unit	Type of distribution	Mean value, μ	Standard Deviation, σ (% of μ)	μ-4σ	μ+4σ
4	T_{steam}	LOCA steam temperature	°C	*normal*	250	5	200	300
5	P_{steam}	LOCA steam pressure	MPa	*normal*	0.1	5	0.1	0.12
6	$\rho_{primary}$	Nominal condition water density in primary circuit	kg/m^3	*normal*	666.7	2	613.36	720.04
7	$P_{primary}$	Nominal condition pressure of primary circuit	MPa	*normal*	15.5	2	14.26	16.74
8	V	Containment volume	m^3	*normal*	58333	1	55999.7	60666.3
9	t	Containment wall thickness	m	*normal*	0.04455	0.5	0.0437	0.0454

10	D	Containment diameter	m	*normal*	39.62	0.5	38.83	40.41
11	H	Containment height	m	*normal*	34.12	0.5	33.44	34.80
12	W	Width of air baffle inlet	m	*normal*	0.92	0.5	0.90	0.94
13	H_1	Height of air baffle downcomer	m	*normal*	38.11	0.5	37.35	38.87
14	H_2	Height of air baffle riser	m	*normal*	59.89	0.5	58.69	61.09
15	D_3	Diameter of air outlet	m	*normal*	9.75	0.5	9.56	9.95
16	H_3	Height of air outlet	m	*normal*	6	0.5	5.88	6.12
17	D_4	Uphead diameter	m	*normal*	39.62	0.5	38.83	40.41
18	H_4	Height of uphead	m	*normal*	11.47	0.5	11.24	11.70
19	d	Diffusive coefficient	m^2/s	*normal*	2.55E-05	20	5.10E-06	4.59E-05
20	λ	Containment thermal conductivity	W/(m K)	*normal*	54	5	43.20	64.80

	Input variable	Description	Unit	Type of distribution	Lower value	Upper value
21	K	Air channel rugosity	-	*uniform*	0.00285	0.00315
22	f_1	Corner friction factor	-	*seasonal*	0.475	0.525
23	f_2	Inlet friction factor	-	*seasonal*	0.9025	0.9975
24	f_3	Riser friction factor	-	*seasonal*	0.1425	0.1575
25	f_4	Outlet friction factor	-	*seasonal*	0.1425	0.1575
26	f_5	Downcomer friction factor	-	*seasonal*	0.1425	0.1575

4 RESULTS

4.1 *Uncertainty propagation*

Figure 4 shows the histogram of the output variable of steady state containment pressure, $P_{containment}$, obtained with 15600 TH simulations of different LOCA scenarios characterized by different values sampled for the D=26 input variables from the distributions of Table 1. This histogram is

representative of the true distribution $f(P_{containment})$ and is taken as reference for the reconstruction by FMM. The calculation time is 4313 s on an Intel Core2Duo P7550.

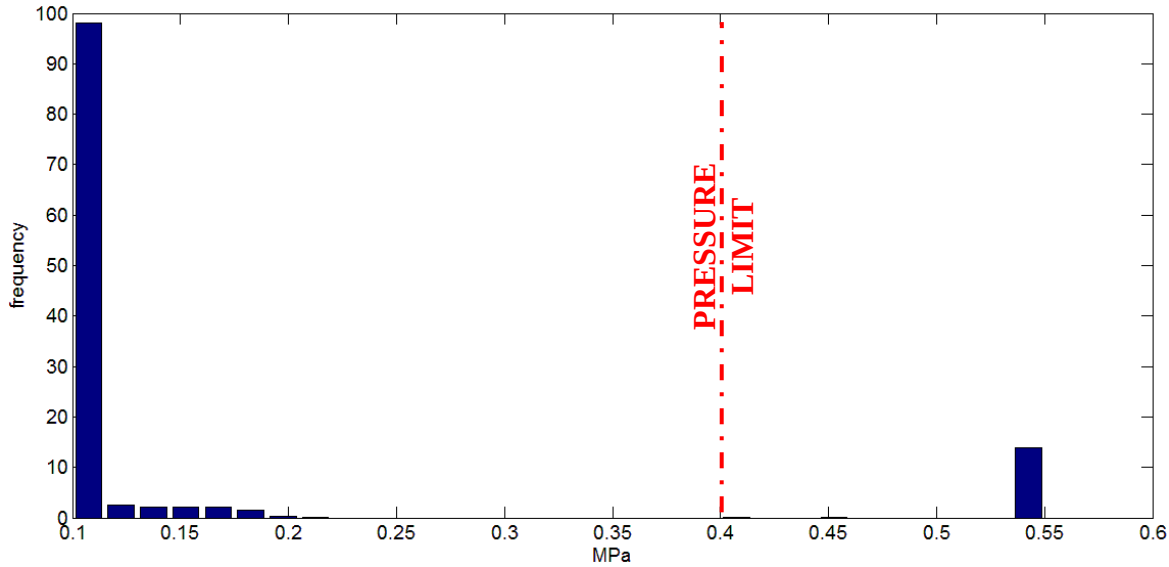

Figure 4 Histogram of the model output variable (*N*=15600 code runs)

The multinomial distribution $f(P_{containment})$ is reconstructed using the EM algorithm of Section 2, with three Gaussian distributions $f_k(\theta_k), k = 1,2,3,$ on the basis of the data obtained by different sets of LOCA simulations, differing in the number of runs. Figure 5, Figure 6 and Figure 7 show results for *N*=15600, 910 and 156 runs of the TH code, respectively (with simulation times 4313 s, 258 s, 47 s including FMM calculation times on an Intel Core2Duo P7550, respectively). It is worth mentioning that the number *N* should be systematically chosen to guarantee satisfying a designated probability with a given confidence level. Actually, in this work *N* has been chosen to compare the proposed approach with the variance decomposition method that is shown to be satisfactory only with *N*=15600. The parameters of the mixture models found for different sample sizes *N* are reported in Table 2. As it is possible to see, the same three clusters are clearly identified in all the three cases (*N*=15600, *N*=910 and *N*=156) with similar mean values μ, standard deviations σ and weights probabilities π. The first two Gaussians ($\mu_1 = 0.1$ and $\mu_2 = 0.15$) are almost exclusively (except for the tails) below the safety limit of 0.4 MPa, while the third Gaussian ($\mu_3 = 0.55$) is almost entirely exceeding the safety limit. The FMM accurately reconstructs the pdf of $P_{containment}$ throughout its interval support of variability, although for large *N* it differs somewhat from the histogram of Figure 4 where the tails of $f_1(y), f_2(y), f_3(y)$ overlap, i.e., around 0.11 MPa. However, since Figure 7 confirms the accuracy of the FMM for small *N* and the analytical pdf $f(P_{containment})$ calculated with the mixture model allows us retrieving the correct clusters, we can define *K*=3 as the optimal number of mixture models to be used for FMM reconstruction.

Table 2 Parameters of the gaussian finite mixture models distributions computed with the EM algorithm for different numbers of code runs

Sample size	Probabilities (π_1, π_2, π_3)	Means (μ_1, μ_2, μ_3)	Standard deviations $(\sigma_1, \sigma_2, \sigma_3)$
15600	(0.81, 0.08, 0.11)	(0.1004, 0.1541, 0.5500)	(0.0023, 0.0236, 0.0000)
910	(0.81, 0.08, 0.10)	(0.1004, 0.1504, 0.5500)	(0.0020, 0.0226, 0.0000)
156	(0.78, 0.08, 0.13)	(0.1001, 0.1496, 0.5500)	(0.0085, 0.0235, 0.0000)

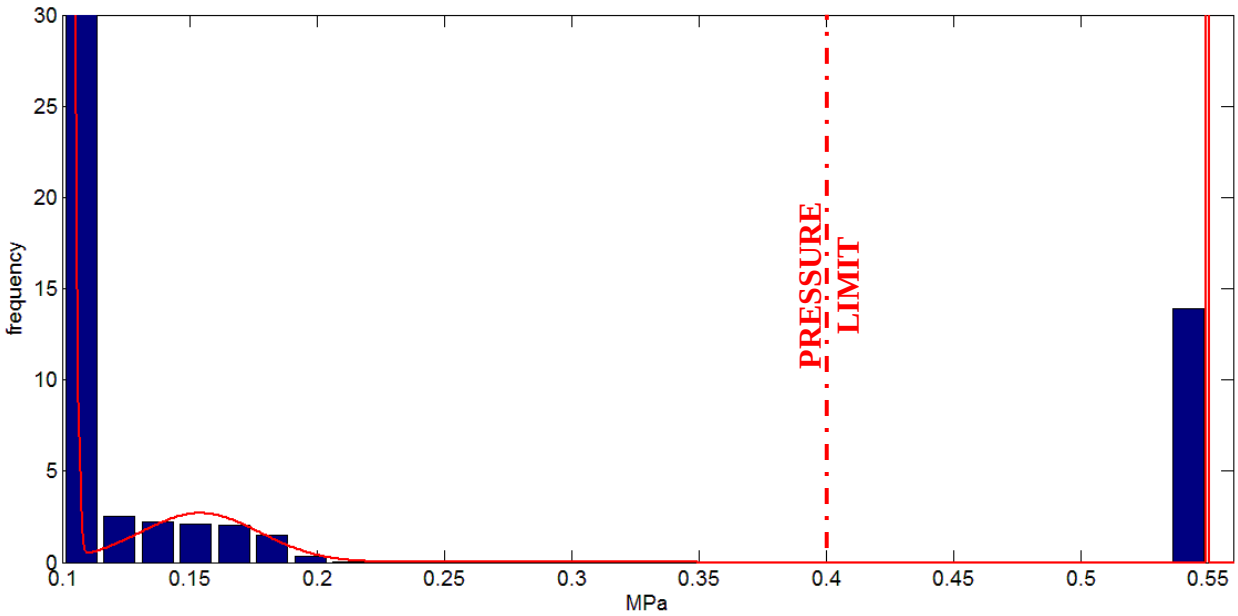

Figure 5 Histogram of the model output values and mixture model reconstruction with *N*=15600 code runs

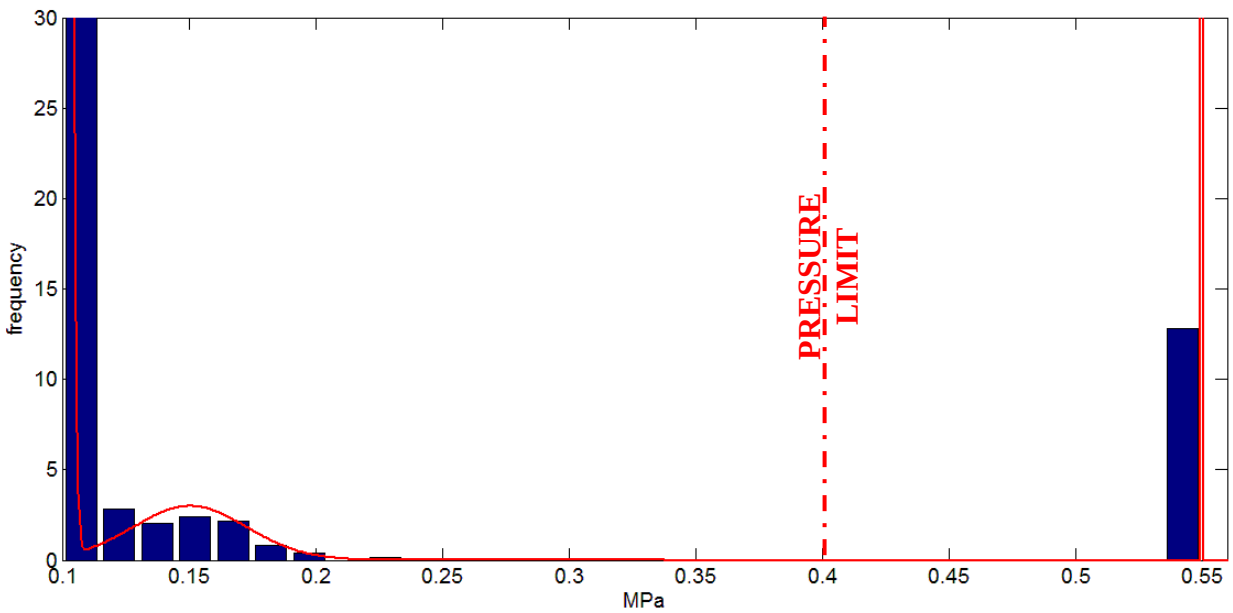

Figure 6 Histogram of the model output values and mixture model reconstruction with *N*=910 code runs

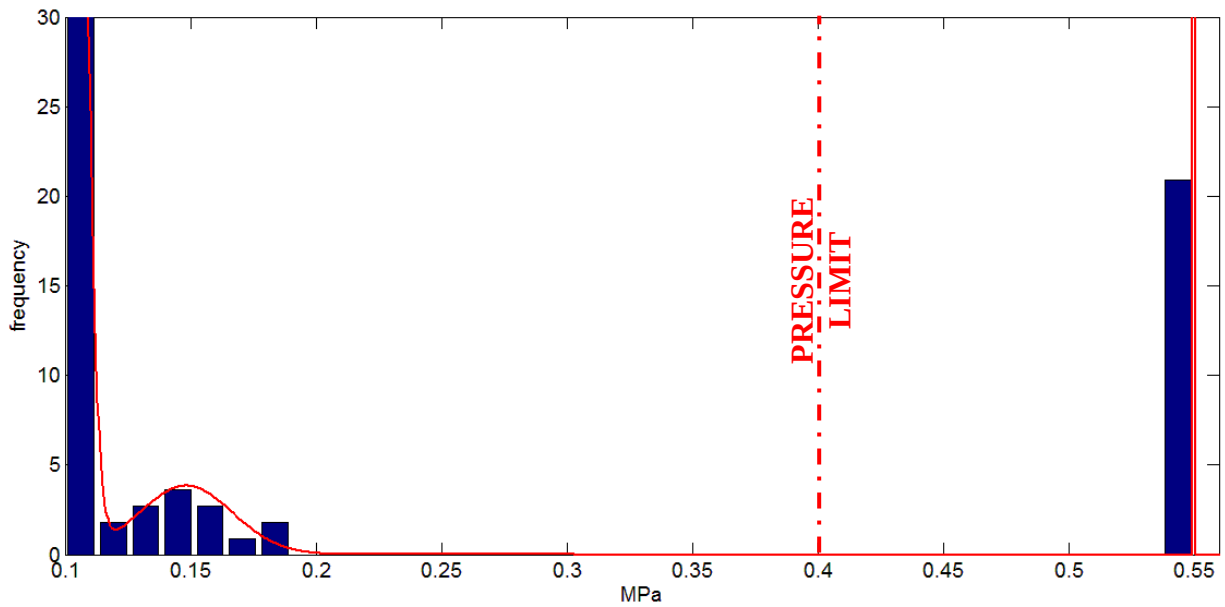

Figure 7 Histogram of the model output values and mixture model reconstruction with *N*=156 code runs

4.2 *Sensitivity Analysis*

In Figure 8, the SA results obtained in [Di Maio et al. 2014b] by a variance decomposition method applied to the TH model of the PCCS of the AP1000 and performed with N=15600 TH code runs (simulation time 4326 s on an Intel Core2Duo P7550) of LOCA scenarios are reported. The sensitivity indexes η^2 of G and T_{inlet} are clearly predominant and those of the other inputs variables are negligible; thus, G and T_{inlet} are by far the most important inputs for the PCCS functional failure. This result aligns with engineering expectations: in fact, not only G and T_{inlet} are directly linked to the energy entering (G) and leaving (T_{inlet}) the PCCS, but they have also, by far, the largest uncertainties as reported in Table 1. The other inputs have low uncertainties and their effects on the output are modest even when sampled at maximum or minimum values of their ranges.

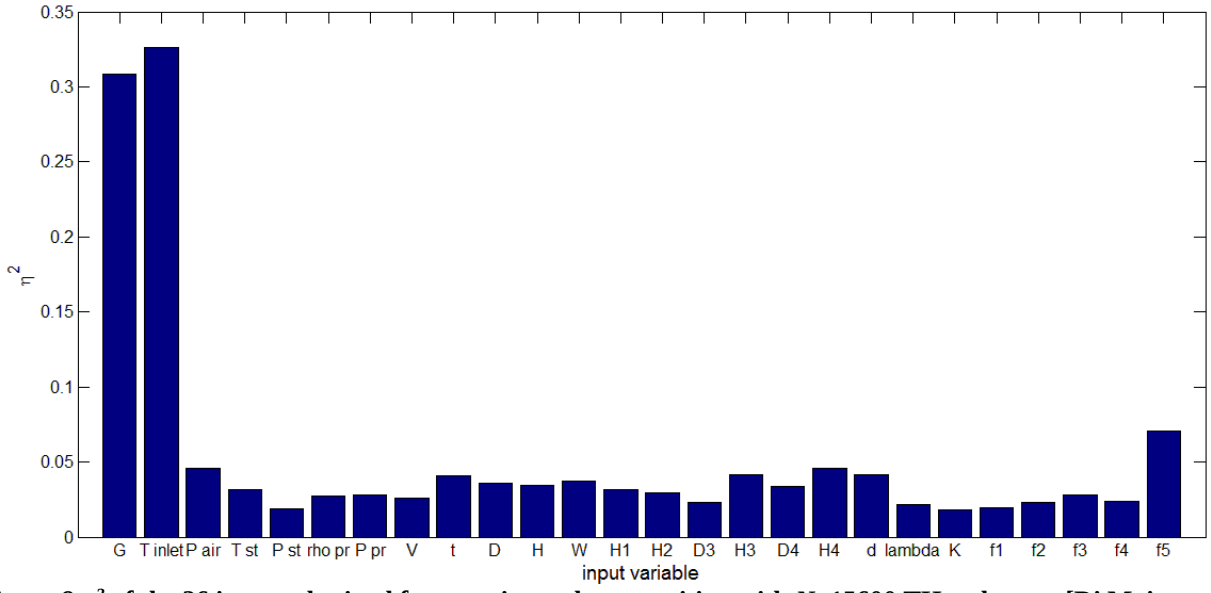

Figure 8 η^2 of the 26 inputs obtained from variance decomposition with N=15600 TH code runs [Di Maio et al. 2014b]

The differences among the η^2 values of the other input variables, (excluding G and T_{inlet}), are non significant as they are affected by large fluctuations depending on the number N of simulations (the larger N, the smaller the fluctuations). In summary, the insights from these results are i) G and T_{inlet} are the most important inputs, ii) the other input variables are not significantly influencing the output, iii) relative input ranking is solid and reliable for G and T_{inlet} but not for the other inputs, iv) large number N of TH code runs is needed and v) the computation of the values of η^2 is burdensome.

In Figure 9, the results of the SA performed via the FMM method as explained in Section 2 are shown: again, G and T_{inlet} are identified as the two most important input variables whereas the saliency of the remaining input variables is negligible. The computation time is 4348 s on an Intel Core2Duo P7550.

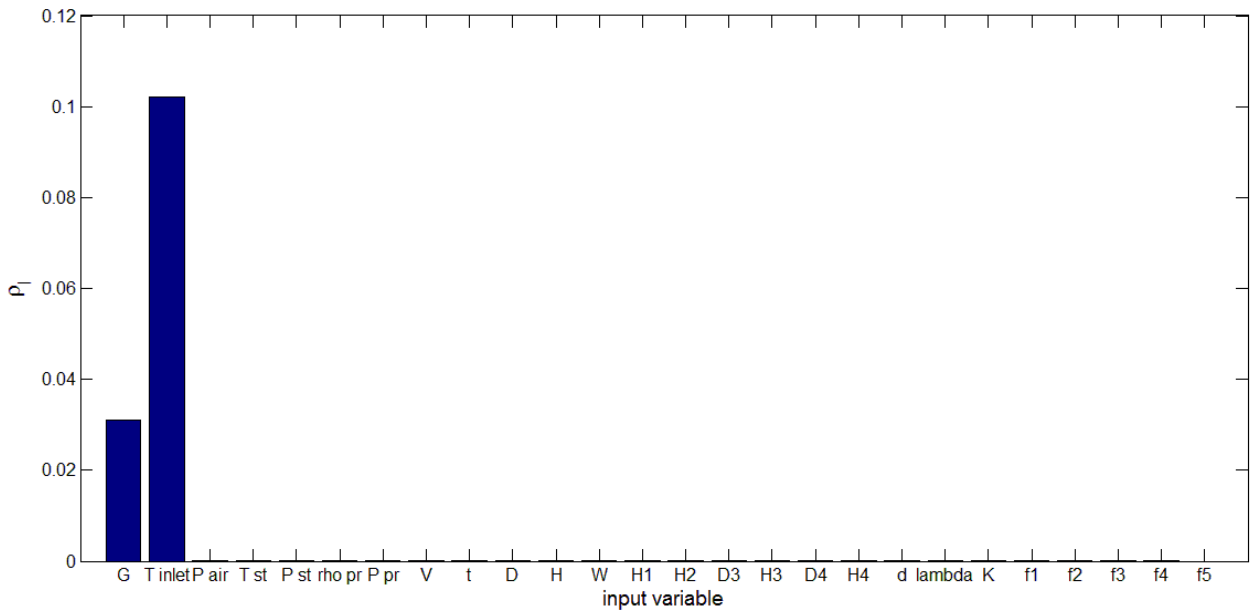

Figure 9 Input saliency obtained with *N*=15600 TH code runs

In Figure 10, the η^2 values obtained by variance decomposition SA are reported when N = 910 TH code runs (simulation time 254 s) are used: the method is not capable of recognizing the importance of G and T_{inlet}. On the other hand, in Figure 11 input saliencies estimated by the FMM method are shown: the importance of G and T_{inlet} is still clearly identified and the other input variables are again reconfirmed as non-influent for the quantification of the final pressure $P_{containment}$ at the end of the LOCA event. Furthermore, the values of the saliencies ρ_G and $\rho_{T_{inlet}}$ are very similar to the ones obtained with N=15600, showing the robustness of method to low numbers of simulation runs.

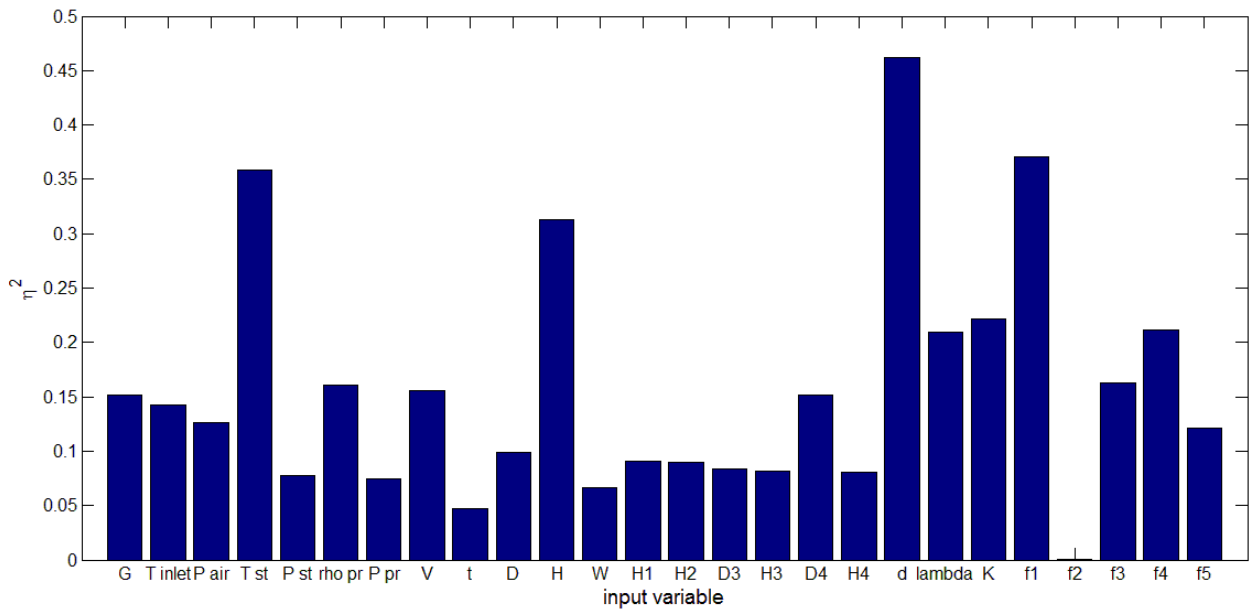

Figure 10 η^2 of the 26 inputs obtained from variance decomposition with *N*=910 TH code runs

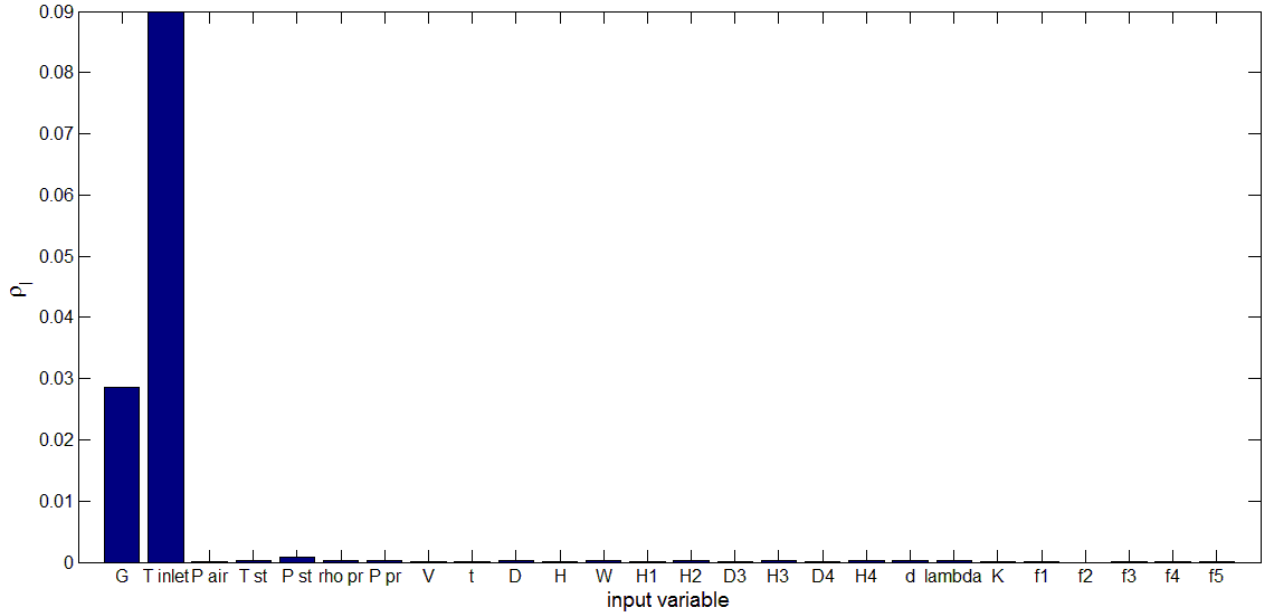

Figure 11 Input saliency obtained with *N*=910 TH code runs

Even with smaller numbers of TH code simulations, the analysis leads to similar conclusions: Figure 12, shows that the variance decomposition method is not capable of reproducing a reliable importance ranking when fed with N=156 code runs (simulation time 46 s); on the other hand, in Figure 13 the saliencies obtained with N=156 still unambiguously show the dominant importance of G and T_{inlet}.

Finally, in Figure 14 the features saliencies obtained with only N=70 runs of the TH code are shown: it is still possible to identify G and T_{inlet} as the two most important inputs, with less than 1% of the number N of simulations needed for variance decomposition to provide reliable results.

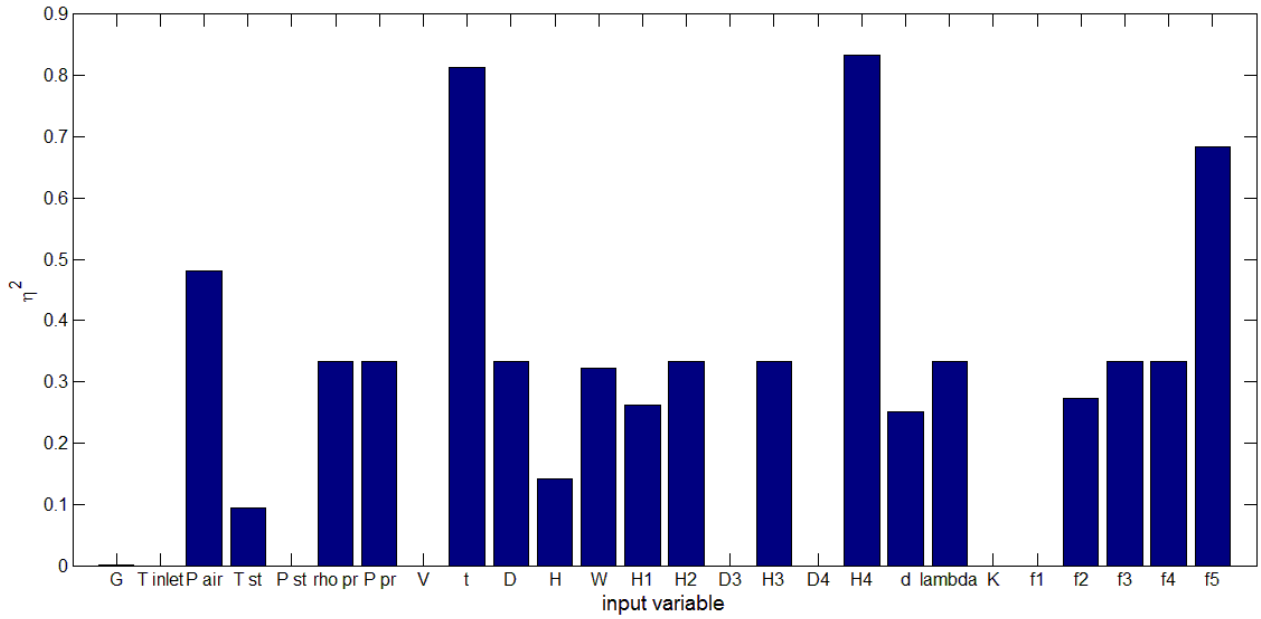

Figure 12 η^2 of the 26 inputs obtained from variance decomposition with *N*=156 TH code runs

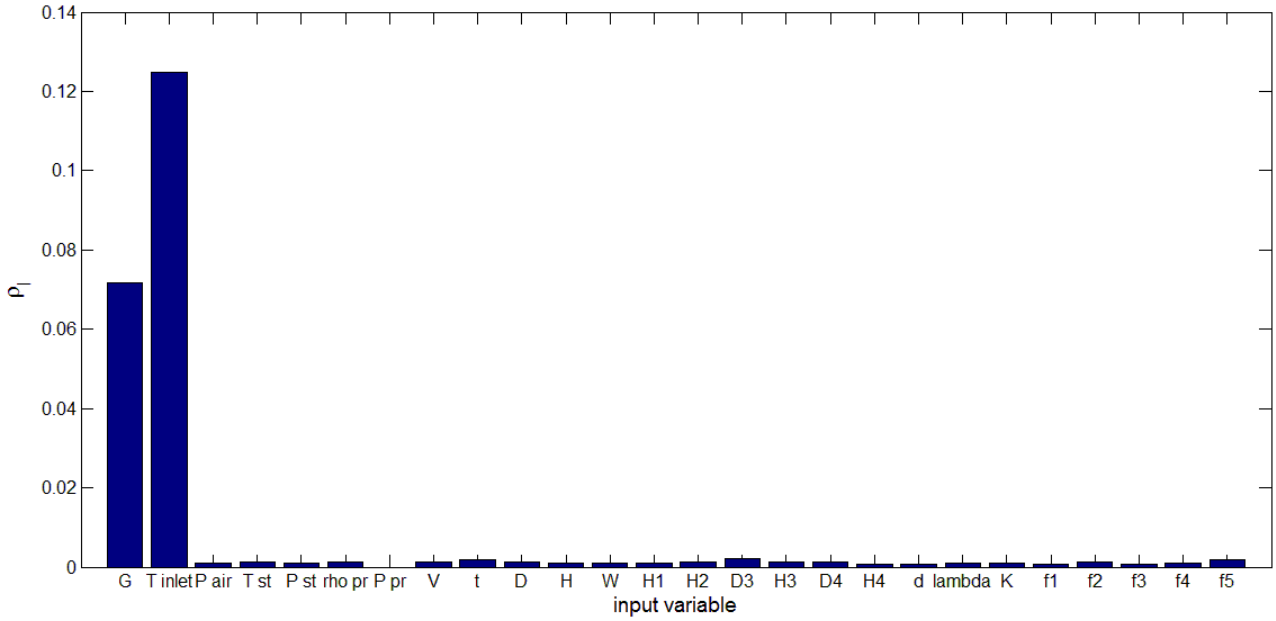

Figure 13 Input saliency obtained with *N*=156 TH code runs

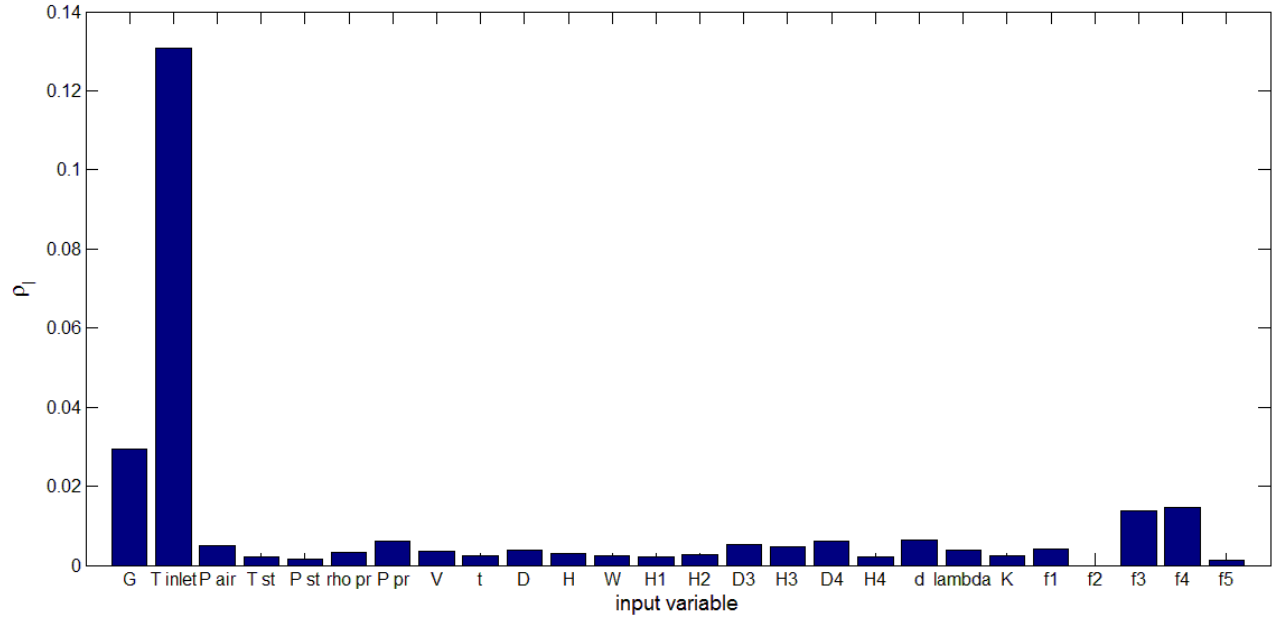

Figure 14 Input saliency obtained with *N*=70 TH code runs

Figure 15 shows the FMM decomposition for T_{inlet}, the most important input variable which is seen to contribute significantly to shaping the output pdf, and *W*, a non contributing one.

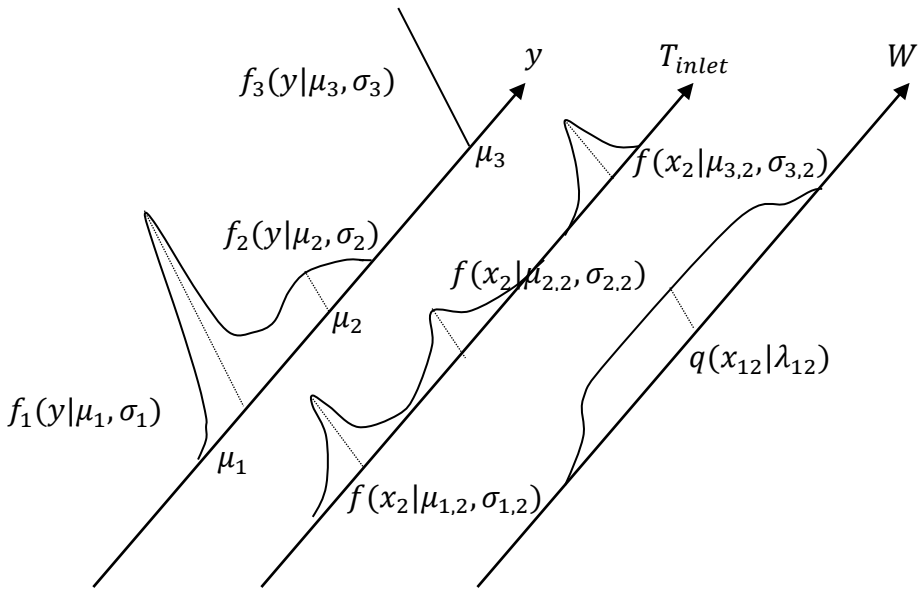

Figure 15 Mixture model decomposition illustrative example

As concluding remarks, we can observe in Table 3 that the FMM method for SA i) has a lower computational cost, providing better results with fewer TH code runs than the variance decomposition method, ii) provides credible results even with reduced number *N* of samples (and, thus, reduced computational cost), and iii) is an effective way to carry out SA and UA within a unified framework, whereby UA provides the input data for SA.

Table 3 Correctness of the ranking produced by the FMM and the variance decomposition methods with different sample sizes *N* for the two most relevant inputs *G* and T_{inlet}

	Ranking of the first two inputs (*G* and T_{inlet})	
***N*=Sample size**	**FMM**	**Variance decomposition**
15600	✓	✓
910	✓	✗
156	✓	✗

5 CONCLUSIONS

In this paper, we have presented an approach for SA for the functional reliability analysis of a passive safety system. The proposed approach has proven capable of reducing the computational cost with respect to variance decomposition-based methods. For SA, we have applied a Gaussian Finite Mixture Model (FMM) to retrieve the analytical pdf of the passive safety system TH code

output with few simulations. An EM algorithm has been innovatively used for retrieving the importance of the input variables of the TH code, making direct use of the "natural" clustering provided by the mixture model. The results obtained on a case study have been compared with those of variance decomposition, with clear demonstration of the capability of the framework of providing satisfactory results with less TH code runs.

Acknowledgements

The participation of Yu Yu to this research is supported by "The National Natural Science Foundation of China" (51206042).

REFERENCES

[10 CFR 50.46] Acceptance criteria for emergency core cooling systems for light-water nuclear power reactors, NRC Regulations.

[Boyack et al. 1990] Boyack, B.E., Catton, I., Duffey, R.B., Griffith, P., Kastma, K.R., Lellouche, G.S., Levy, S.,Rohatgi, U.S., Wilson, G.E., Wulf, W., Zuber, N., 1990. Quantifying reactor safety margins. Part 1: an overview of the code scaling, applicability and uncertainty evaluation methodology. Nucl. Sci. Des. 119, 1–15.

[Burgazzi 2004] Luciano Burgazzi, "Evaluation of uncertainties related to passive systems performance", Nuclear Engineering and Design 230 (2004) pp. 93–106.

[Burgazzi 2007] Burgazzi, L., Addressing the uncertainties related to passive system reliability. Progress in Nuclear Energy 49, 93–102.

[Burgazzi 2007b] Burgazzi, L., Thermal–hydraulic passive system reliability-based design approach. Reliability Engineering and System Safety 92 (9), 1250–1257.

[Cadini et al. 2007] F. Cadini, E. Zio, F. Di Maio, V. Kopustinkas, R. Urbonas, "A Neural-network-based Variance Decomposition Sensitivity Analysis", International Journal of Nuclear Knowledge Management, Vol. 2, No. 3, 2007, pp. 299-312.

[Carlos et al. 2013] S. Carlos, A. Sánchez, D. Ginestar, S. Martorell, "Using finite mixture models in thermal-hydraulics system code uncertainty analysis". Nuclear Engineering and Design, Volume 262, September 2013, Pages 306-318.

[Cummins et al. 2003] "Westinghouse AP1000 Advanced Passive Plant", W.E. Cummins, M.M. Corletti, T.L. Schulz, Westinghouse Electric Company, LLC. Proceedings of ICAPP '03 Cordoba, Spain, May 4-7, 2003 Paper 3235.

[de Crecy et al. 2008] de Crécy, A., Bazin, P., Glaeser, H., Skorek, T., Joucla, J., Probst, P., Fujioka, K., Chung,B.D., Oh, B.D., Kyncl, B.D., Pernica, R., Macek, J., Meca, R., Macian, R., D'Auria, F.,Petruzzi, F., Batet, L., Perez, M., Reventos, F., 2008. Uncertainty and sensitivity analysis of the LOFT L2-5 test: results of the BEMUSE programme. Nucl. Eng. Des.238, 3561–3578.

[Dempster 1977] Dempster, A., Laird, N., Rubin, B., 1977. Maximum likelihood estimation from incomplete data via EM algorithm. J. Roy. Stat. Soc. B 39, 1–38.

[Devictor et al. 2005] Nicolas Devictor, Ricardo Bolado Lavín, "Uncertainty and sensitivity methods in support of PSA level 2", 2005, Proceedings of the Workshop on evaluation of uncertainties in relation to severe accident and level 2 probabilistic safety analysis, Aix-en-Provence, 7-9 November 2005.

[Di Maio et al. 2014a] F. Di Maio, G. Nicola, E. Zio, Y.Yu, Ensemble-based sensitivity analysis of a best estimate thermal hydraulic model of a Passive Containment Cooling System of an AP1000 Nuclear Power Plant, Annals of Nuclear Energy 73 (2014) pp. 200-210

[Di Maio et al. 2014b] F. Di Maio, G. Nicola, E. Zio, Sensitivity Analysis and Failure Damage Domain Identification of the Passive Containment Cooling System of an AP1000 Nuclear Reactor, accepted, PSAM12, Probabilistic Safety Assessment & Management, 22-27 June 2014, Honululu, USA.

[Di Maio et al. 2014c] F. Di Maio, G. Nicola, E. Borgonovo, E. Zio, Ensemble of Invariant Methods for Sensitivity Analysis of a Passive Containment Cooling System of an AP1000 Nuclear Power Plant, ESREL2014.

[Eaton et al. 2010] Eaton, M., Williams, M.M.R., 2010. A probabilistic study of the influence of parameter uncertainty on solutions of the neutron transport equation. Prog. Nucl. Energy52, 580–588.

[Fang et al. 2003] Shoufan Fang, George Z. Gertner, Svetlana Shinkareva, Guangxing Wang, Alan Anderson "Improved generalized Fourier amplitude sensitivity test (FAST) for model assessment",2003, Statistics and Computing Volume 13, Issue 3 , pp 221-226.

[Figueiredo et al. 2002] Figueiredo, M., Jain, A.K., 2002. Unsupervised learning of finite mixture models. IEEE Trans. Pattern Anal. Mach. Intell. 42 (3), 1–16.

[Gilli et al. 2012] Gilli, L., Lathouwers, D., Kloosterman, J.L., van der Hagen, T.H.J.J., 2012. Perform-ing uncertainty analysis of nonlinear point-kinetics/lumped parameters using polynomial chaos expansion. Ann. Nucl. Energy 40, 35–44.

[Guba et al. 2003] Guba, A., Makai, M., Pal, L., 2003. Statistical aspects of best estimate method-I. Reliab. Eng. Syst. Safe. 80, 217–232.

[Helton 1993] J. C. Helton, “Uncertainty and Sensitivity analysis Techniques for Use in Performance Assessment for Radioactive Waste Disposal”, Reliability Engineering and System Safety, 42, 1993, pp. 327-367.

[Hong et al. 2011] Hong, I.S., Oh, D.Y., Kim, I.G., 2011. Generic application of Wilks tolerance limit evaluation approach to nuclear safety. In: Proceedings of the OCDE/CSNI Workshop on Best Estimate Methods and Uncertainty Evaluations, Barcelona, Spain.

[Langewisch 2010] Langewisch, D.R., 2010. Uncertainty and Sensitivity Analysis for Long-running Computer Codes: A Critical Review. Massachusetts Institute of Technology, http://hdl.handle.net/1721.1/58285

[Marques et al. 2005] Marques, M., Pignatel, J.F., Saignes, P., D’Auria, F., Burgazzi, L., Muller, C., Bolado-Lavin, R., Kirchsteiger, C., La Lumia, V., Ivanov, I., 2005. Methodology for the reliability evaluation of a passive system and its integration into a probabilistic safety assessment. Nuclear Engineering and Design 235, 2612–2631.

[McKay 1996] M. D. McKay, “Variance-Based Methods for Assessing Uncertainty Importance in NUREG-1150 Analyses”, LA-UR-96-2695, Los Alamos National Laboratory, 1996, pp. 7-27.

[McLachlan et al. 2000] McLachlan, G., Peel, D., 2000. Finite Mixture Models. John Wiley & Sons Inc., New York.

[McLachlan et al. 2008] McLachlan, G., Krishnan, T., 2008. The EM Algorithm and Extensions. Wiley-Interscience, Hoboken.

[Nayak et al. 2009] A.K. Nayak, M.R Gartia, A. Antony, G. Vinod, R.K. Sinha, “Reliability assessment of passive isolation condenser system of AHWR using APSRA methodology”, 2009, Reliability Engineering and System Safety 94, 1064–1075.

[Nouy 2010] Nouy, A., 2010. Identification of multi-modal random variables through mixtures of polynomial chaos expansions. Comptes Rendus Mecanique 338, 698–703.

[Pagani et al. 2005] Pagani, L., Apostolakis, G.E., Hejzlar, P., 2005. The impact of uncertainties on the performance of passive systems. Nuclear Technology 149, 129–140.

[Pengfei 2014] Pengfei Wei, Zhenzhou Lu, Wenbin Ruan, Jingwen Song, “Regional sensitivity analysis using revised mean and variance ratio functions”. Reliability Engineering and System Safety 121 (2014) 121–135.

[Pourgol-Mohammad 2009] Pourgol-Mohammad, M., 2009. Thermal-hydraulics system codes uncertainty assessment: a review of the methodologies. Ann. Nucl. Energy 36, 1774–1786.

[Pudil et al. 1995] P. Pudil, J. Novovicova´, and J. Kittler, “Feature Selection Based on the Approximation of Class Densities by Finite Mixtures of the Special Type,” Pattern Recognition, vol. 28, no. 9, pp. 1389-1398, 1995.

[Rahim et al. 2011] Farzad Choobdar Rahim, Mohammad Rahgoshay, Seyed Khalil Mousavian, 2011 “A study of large break LOCA in the AP1000 reactor containment”,

[Saltelli et al. 2000] Saltelli A, Chan K, Scott E. Sensitivity analysis. New York: John Wiley & Sons Inc.; 2000.

[Schulz 2006] T.L.Schulz, 2006, “Westinghouse AP1000 advanced passive plant,” Nuclear Engineering and Design, vol. 236,pp 1547-1557.

[Secchi et al. 2008] P. Secchi, E. Zio, F. Di Maio, “Quantifying Uncertainties in the Estimation of Safety Parameters by Using Bootstrapped Artificial Neural Networks”, Annals of Nuclear Energy, Volume 35(12), Pages 2338-2350, doi: 10.1016/j.anucene.2008.07.010, 2008.

[Sudret 2008] Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansion. Reliab. Eng. Syst. Safe. 93, 964–979.

[Vaithyanathan et al. 1999] S. Vaithyanathan and B. Dom, “Generalized Model Selection for Unsupervised Learning in High Dimensions,” Advances in Neural Information Processing Systems 12, pp. 970-976, Cambridge, Mass. MIT Press, 1999.

[Wilson et al. 1990] Wilson, G.E., Boyack, B.E., Catton, I., Duffey, R.B., Griffith, P., Kastma, K.R., Lellouche, G.S., Levy, S., Rohatgi, U.S., Wulf, W., Zuber, N., 1990. Quantifying reactor safety margins. Part 2: characterization of important contributions to uncertainty. Nucl. Sci. Des. 119, 17–31.

[Wulf et al. 1990] Wulf, W., Boyack, B.E., Catton, I., Duffey, R.B., Griffith, P., Kastma, K.R., Lellouche, G.S., Levy, S., Rohatgi, U.S., Wilson, G.E., Zuber, N., 1990. Quantifying reactor safety margins. Part 3: assessment and ranging of parameters. Nucl. Sci. Des.119, 33–65.

[Yu et al. 2010] Y. Yu, T. Liu, J. Tong, J. Zhao, F. Di Maio, E. Zio, A. Zhang, “Multi-Experts Analytic Hierarchy Process for the Sensitivity Analysis of Passive Safety Systems”, Proceedings of the 10th International Probabilistic Safety Assessment & Management Conference, PSAM10, Seattle, June 2010.

[Yu et al. 2013] Yu Yu, Shengfei WANG, Fenglei NIU, “Thermal–hydraulic performance analysis for AP1000 passive containment cooling system”, Proceedings of the 21th International Conference on Nuclear Engineering ICONE21 July 29- August 2, 2013, Chengdu, Sichuan, China.

[Yu et al. 2013b] Yu Yu, Shengfei Wang, Fenglei Niu; “Effect of Air Temperature on Failure Prognosis of Passive Containment Cooling System in AP1000”. Chemical Engineering Transactions Vol. 33, 2013 pp. 835-840

[Yu et al., 2015] Yu, Y., Ma, G., Hao, Z., Niu, F., Zio, E., “Correlation analysis for screening key parameters for passive system reliability analysis”, Annals of Nuclear Energy, 77, pp. 23-29, 2015.

[Zio et al. 2008] E. Zio, and F. Di Maio, “Bootstrap and Order Statistics for Quantifying Thermal-Hydraulic Code Uncertainties in the Estimation of Safety Margins”, Science and Technology of Nuclear Installations, Volume 2008 (2008), Article ID 340164, 9 pages, doi:10.1155/2008/340164.

[Zio et al. 2008] 33. E. Zio, F. Di Maio, S. Martorell and Y. Nebot, “Neural Networks and Order Statistics for Quantifying Nuclear Power Plants Safety Margins”, Safety, Reliability and Risk Analysis, Taylor & Francis Group, London, ISBN 978-0-415-48513-5, proceedings of ESREL 2008 Conference, Valencia, Spain, September 2008.

[Zio et al. 2009] E. Zio and N. Pedroni, 2009, “Estimation of the functional failure probability of a thermal-hydraulic passive system by Subset Simulation”, 2009, Nuclear Engineering and Design , vol. 239, pp580-599.

[Zio et al. 2010] E. Zio, N.Pedroni, “How to effectively compute the reliability of a thermal–hydraulic nuclear passive system”, 2010.

[Zio et al. 2010b] E. Zio, F. Di Maio, J. Tong, “Safety Margins Confidence Estimation for a Passive Residual Heat Removal System”, Reliability Engineering and System Safety, RESS, Vol. 95, pp. 828–836, 2010.