Ensemble-based sensitivity analysis of a Best Estimate Thermal Hydraulics model: Application to a Passive Containment Cooling System of an AP1000 Nuclear Power Plant

Francesco Di Maio [a,*], Giancarlo Nicola [a], Enrico Zio [a,b,c], Yu Yu [d]

[a] Energy Department, Politecnico di Milano, Via Ponzio 34/3, 20133 Milano, Italy
[b] European Foundation for New Energy – Electricite de France Ecole Centrale, Paris, France
[c] Supelec, Paris, France
[d] School of Nuclear Science and Engineering, North China Electric Power University, Beijing 102206, China

1. Introduction

Thermal Hydraulics (TH) codes are used to predict the response of the systems in nominal and accidental conditions. Conservative TH codes lead to the definition of safety limits that can be respected with large safety margins (Zio et al., 2010). Best Estimate (BE) codes provide more realistic results, thus avoiding over-conservatism (Zio et al., 2010; 10 CFR 50.46); however, it is necessary to identify and quantify the uncertainties affecting the code outputs due to simplifications, approximations, round-off-errors, numerical techniques and variability in the input parameters values (Pourgol-Mohammad, 2009).

Many approaches have been proposed for this 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) and Integrated Methodology for Thermal-Hydraulics Uncertainty Analysis (IMTHUA) (Glaeser et al., 1994), which assume statistical distributions for the input variables, from which N input values sets are sampled and fed to the BE code (Pourgol-Mohammad, 2009) and the corresponding N outputs are calculated. 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 the computations (Secchi et al., 2008), which however allows determining only some percentiles and not the whole distribution, and does not provide insights on the sensitivity to input variability (Langewisch, 2010; Hong et al., 2011).

For the latter point, sensitivity analysis (SA) methods exist of three types: local, regional and global (Saltelli et al., 2000). When considering TH codes and complex models, global SA is most indicated because capable of dealing with non-linear and non-monotone models, whereas local and regional SA do not give an exhaustive representation of the variability of the model (Saltelli et al., 2008). In fact, local approaches consist in evaluating the effect on the output of small variations around fixed values in the input parameters, by calculating partial derivatives of the output with respect to the inputs around the local fixed values on which the analysis is focused. Regional SA methods aim at calculating the sensitivity of the model to partial ranges of the inputs

---

* Corresponding author. Tel.: +39 0223996372.
  E-mail address: francesco.dimaio@polimi.it (F. Di Maio).

distributions, thus allowing to calculate the uncertainty reduction achievable if the inputs parameters distributions ranges were reduced. Various techniques have been proposed for regional SA like Contribution to Sample Mean (CSM) plot (Bolado-Lavin et al., 2009), Contribution to Sample Variance (CSV) plot (Bolado-Lavin et al., 2012), Variance Ratio Functions (Pengfei, 2014). However, these methods encounter difficulties when the model mean varies significantly with respect to the reduced ranges (Pengfei, 2014). Compared to local and regional SA methods, global SA methods offer higher capabilities but at a higher computational cost. Examples of global SA methods are Response Surface Methodology (RSM), Fourier Amplitude Sensitivity Test (FAST), Delta and Variance Decomposition Method (Helton, 1993; Saltelli et al., 2000; Borgonovo, 2007; Cadini et al., 2007; Yu et al., 2010). RSM amounts to approximating the original model outcomes (i.e., responses) by a simple and fast empirical global model that fits a database of computations and covers the as large as possible variability of the original model (Devictor et al., 2005); with FAST, the model is expanded into a Fourier series whose coefficients and frequencies are used to estimate the mean and variance of the model and the partial variances of individual input parameters of the model (Fang et al., 2003). Delta is a moment independent uncertainty indicator that looks at the influence of input uncertainty on the entire output distribution from given data (including simulation input–output data). Variance decomposition is the most used method for global SA and has the advantage that it does not introduce any hypothesis on the model, but has a high computational cost (Carlos et al., 2013).

To avoid a large number of TH code runs, in this work we propose an innovative framework of analysis whose most "reduced" flowchart is shown in Fig. 1. The idea is to directly rely on the information available in the multimodal pdf of the output variable for performing global SA of a TH code. First, a limited number N of simulations of the TH code are performed and a Finite Mixture Model (FMM) is used to reconstruct the pdf of the output variable. It is worth pointing out that (i) the output distribution is linked to the input distributions that must, then, be selected accurately for building the FMM approximating the pdf of the output variable, and (ii) the FMM construction entails learning the structure of the pdf of the output variable, that is an essential step that is explained in detail in Section 3. The natural clustering made by the FMM on the TH code output (McLachlan et al., 2000; Di Maio et al., 2014b) (i.e., a cluster corresponds to each Gaussian model $f_k(y|\theta_k)$ of the mixture, as we shall see in Section 3) is exploited to develop an ensemble of three SA methods, input saliency (Law et al., 2004), Hellinger distance (Diaconis et al., 1982; Gibbs et al., 2002) and Kullback–Leibler divergence (Diaconis et al.,

1982; Gibbs et al., 2002), for ranking the input variables most affecting the output uncertainty. The ensemble strategy allows combining the output of the three individual methods (that perform more or less well depending on the data at hand) to generate reliable rankings. The advantage is the possibility to overcome possible misjudgments of the individual methods. The innovative idea of using an ensemble of methods for sensitivity analysis, will be shown to be particularly useful when the number of TH code simulations is reduced to keep low the computational cost. Due to the limited quantity of data in this situation, in fact, possible misleading rankings can arise from individual methods, whereas the diversity of the methods integrated in the ensemble allows overcoming the problem.

The proposed framework is generally applicable to active and passive safety systems. However, its application is here challenged by the fact that the reliability analysis of passive safety systems used in advanced Nuclear Power Plants (NPPs) must consider the uncertainties affecting passive systems operation, under scarce or null operating experience (Cummins et al., 2003; Pagani et al., 2005; Burgazzi, 2007; Nayak et al., 2009; Zio et al., 2009). When counter-forces (e.g., friction) have magnitude comparable to the driving ones (e.g., gravity, natural circulation), physical phenomena may fail performing the intended function even if (i) safety margins are met, (ii) no hardware failures occur (Burgazzi, 2004; Marques et al., 2005; Burgazzi, 2007b; Zio et al., 2009). It is, therefore, necessary to tailor the analysis framework on the peculiar characteristics of passive safety systems, for a SA accurate and robust in spite of the typically scarce TH code simulations (and operating experience) and large uncertainties affecting the systems behavior.

The successful application of the developed framework is illustrated, tailored to the sensitivity analysis of a TH code that simulates the behaviour of the Passive Containment Cooling System (PCCS) of an Advanced Pressurized reactor AP1000 during a Loss Of Coolant Accident (LOCA). The combination of the three sensitivity methods is shown to make the results more robust, with no additional computational costs (no more TH code runs are required for SA).

The paper is organized as follows. In Section 2, the case study and the relative TH code are illustrated. In Section 3, the bases of FMM are presented along with the ensemble of sensitivity methods, i.e., the input saliency, Hellinger distance and Kullback–Leibler divergence. In Section 4, the experimental results are reported. Section 5 draws some conclusions.

2. 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 $P_{containment}$ is effectively controlled within the safety limit 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). Fig. 2 shows the PCCS of the AP1000 [Westinghouse Electric Company].
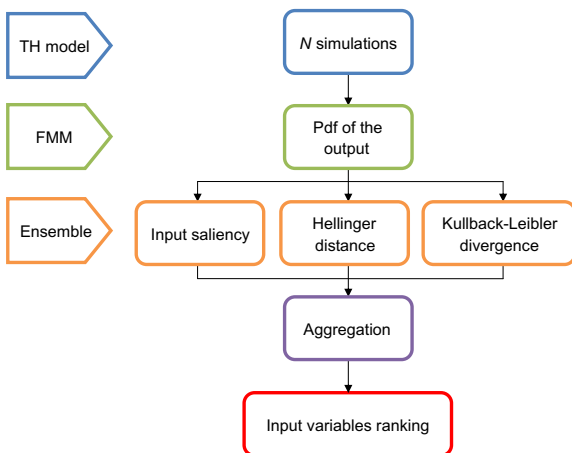

Fig. 1. Flowchart of the proposed framework for SA.

For the quantification of the functional failure of the PCCS of the AP1000 following a LOCA, a TH model that simulates the post-reflood phase dynamics of the stratified heat transfer process with non-condensed heat has been developed.

In general terms, a LOCA evolution typically entails four phases (Rahim et al., 2011): (1) blowdown, from the accident initiation (by a double-ended guillotine pipe break in a primary coolant line determining the leakage of a mass flow rate G, whose initial vapor mass M_{vapor} is at temperature T_{steam}, and 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 (of volume V and diameter D) and, then, the heat is conducted by the vessel and transferred (among the wall thickness t_w with coefficients α and C_p) to the air in the air channels (see Fig. 2). Cold air enters (at temperature T_{inlet} and pressure P_{inlet}) the channels (with speed u_{air}) through the three rows of air inlets and flows down to the bottom of the channels (whose heights are Z_1 and Z_2, and flowing area is A), where it is heated by the steel vessel up to the air diffuser to the environment.

The TH code simulates the dynamics of the heat transfer processes in the post-reflood phase with a stratified dome of the PCCS vessel (Yu et al., 2013). The output variable is the pressure value of $P_{containment}$ after 1000 s from the initiation of the LOCA. The $D = 51$ input variables are listed in the Appendix A together with their distributions (hereafter, also called common densities) chosen from expert judgment and literature review (Burgazzi, 2004; Zio et al., 2008b; Zio et al., 2010b). Three families of distributions have been used: seasonal, normal and uniform. Seasonal relates to the external air temperature T_{inlet} and pressure P_{inlet} 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 friction factors, the supports from "Lower value" to "Upper value" are reported.
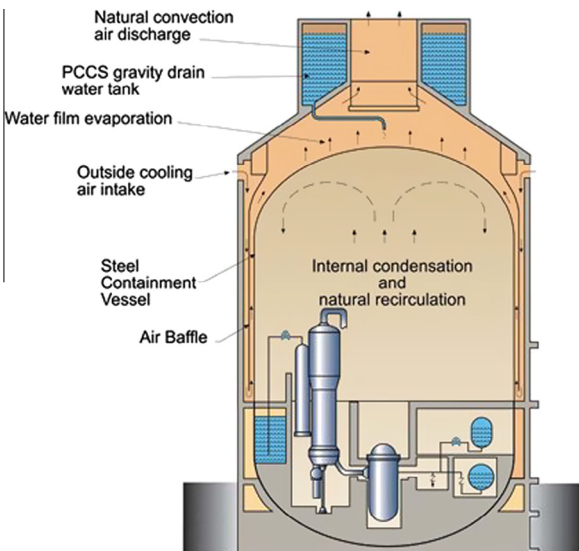

Fig. 2. AP1000 Passive Containment Cooling System [Westinghouse Electric Company].

3. Uncertainty and ensemble based sensitivity analysis

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

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

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:

$$\sum_k \pi_k = 1 \quad \text{with } \forall\, k, \pi_k \geqslant 0 \tag{3}$$

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:

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

The Expectation Maximization (EM) algorithm (Dempster et al. 1977) (McLachlan et al., 2000) can be used to fit $f(y)$ to N available data $y = \{y_1, \ldots, y_N\}$, $i = 1, \ldots, N$ 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:

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

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,

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

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

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

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 $z, z^{(1)}$:

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

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

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

The expectation step follows by application of Bayes rule (McLachlan, 2008):

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

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

The maximization step follows the expectation step, to update θ^j and π^j for any j-th step until the optimum is reached (Figueiredo et al., 2002):

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

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

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

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). These clusters will be exploited for SA within an ensemble of the three methods of input saliency, Hellinger distance, Kullback–Leibler divergence, whose individual outcomes will be aggregated for identifying the input variables most affecting the output uncertainty (Fig. 1).

3.1. Input saliency

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 x 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 its common density $q(x_l|\lambda_l)$ among all the clusters (e.g., its original distribution as in Table 7 in the Appendix A, from which the N set of input decks for building the FMM are sampled) (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:

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

The *saliency* ρ_l is 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, Fig. 3 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 $q(x_2|\lambda_2)$.

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)–(19), whereas $\rho(\rho_1, \rho_2, \ldots, \rho_l, \ldots, \rho_D)$ 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

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

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

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

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

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

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
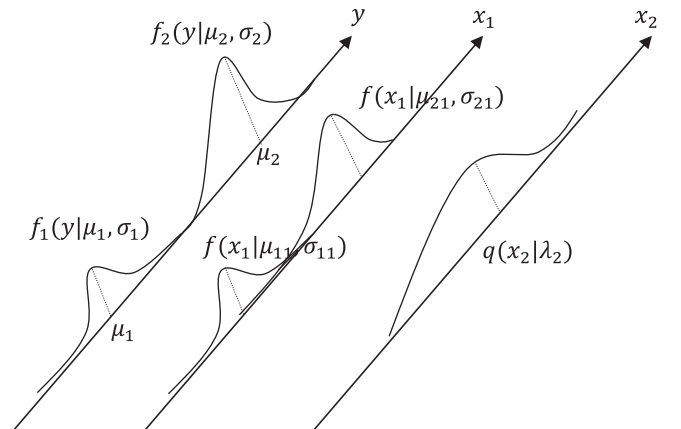

Fig. 3. Mixture model decomposition illustrative example.

(clusters) when supported by N evidences, and thus ρ_l can be considered as a sensitivity index for the l-th input.

3.2. Hellinger distance

The Hellinger distance $H_{lk}(f(x_l|\theta_{kl}), q(x_l|\lambda_l))$ measures the difference between the pdf of the l-th input contributing to shape the k-th cluster $f(x_l|\theta_{kl})$ and its common distribution $q(x_l|\lambda_l)$. It can be defined as (Diaconis et al., 1982; Gibbs et al., 2002):

$$H_{lk}(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) = \left[\frac{1}{2}\int\left(\sqrt{f(x_l|\theta_{kl})} - \sqrt{q(x_l|\lambda_l)}\right)^2 dx\right]^{\frac{1}{2}}$$

$$= \left[1 - \int\left(\sqrt{f(x_l|\theta_{kl})q(x_l|\lambda_l)}\right)^2 dx\right]^{\frac{1}{2}} \quad (27)$$

which satisfies the inequality $0 \le H(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) \le 1$.

In the particular case of $f(x_l|\theta_{kl}) \sim N(\mu_1, \sigma_1^2)$ and $q(x_l|\lambda_l) \sim N(\mu_2, \sigma_2^2)$, the Hellinger distance can be written as:

$$H_{lk}(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) = \left[1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}}e^{-\frac{1}{4}\frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)}}\right]^{\frac{1}{2}} \quad (28)$$

The quantity HL_l is the importance of the l-th input variable in affecting the output y, defined as:

$$HL_l = \sum_{k=1}^{K} H_{lk}(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) \quad (29)$$

where the l-th input variable is important if HL_l is large (in relative terms).

3.3. Kullback–Leibler divergence

The Kullback–Leibler divergence measures the different information carried by the pdf of the l-th input in the k-th cluster $f(x_l|\theta_{kl})$ and its common distribution $q(x_l|\lambda_l)$ (Diaconis et al., 1982; Gibbs et al., 2002):

$$KL_{lk}(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) = \int_{-\infty}^{+\infty} f(x_l|\theta_{kl}) \log\left(\frac{f(x_l|\theta_{kl})}{q(x_l|\lambda_l)}\right) dx \quad (30)$$

with values in $[0, \infty]$.

A symmetric Kullback–Leibler divergence, can be defined (Kullback et al., 1951):

$$KL_{sym_{lk}}(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) = KL_{sym_{lk}}(q(x_l|\lambda_l), f(x_l|\theta_{kl}))$$

$$= \frac{1}{2}KL_{lk}(f(x_l|\theta_{kl}), q(x_l|\lambda_l))$$

$$+ \frac{1}{2}KL_{lk}(q(x_l|\lambda_l), f(x_l|\theta_{kl})) \quad (31)$$

The quantity KLS_l can be taken to represent the importance of the l-th input variable in affecting the output y:

$$KLS_l = \sum_{k=1}^{K} KL_{sym_{lk}}(f(x_l|\theta_{kl}), q(x_l|\lambda_l)) \quad (32)$$

where the l-th input variable is important if HL_l is large (in relative terms).

3.4. Ensemble

A combination of the outputs of the three SA methods in ensemble can be considered for the evaluation of the sensitivity of the output variable on the different inputs. As anticipated in the Introduction, the resulting ensemble is expected to strengthen the results by agreement among the rankings produced by the single methods, without requiring any additional TH simulations for this. In fact, considering the information provided by the three different methods it is possible to gain more confidence on the ranking, when they agree and overcome singular misjudgments by aggregation, when they disagree. The central issue is to decide how to aggregate the sensitivity ranking outcomes provided by the different methods in the ensemble (Baraldi et al., 2011; Di Maio et al., 2012). In this paper, we adopt and compare two simple strategies of aggregation: majority voting R_{mv} and R_{sum} aggregation (Kukkonen et al., 2007). The former consists in taking the majority voting among the three methods: the ranking orders (each one of length $D = 51$) provided individually by the three methods (i.e., O_ρ, O_{HL} and O_{KLS}) are aggregated by assigning to each ranking position $R_{mv,l}$ the l-th input voted by majority, viz:

$$R_{mv,l} = \arg(\max(i)) \text{ where } \begin{cases} i = 1 & \text{if} \begin{cases} O_{\rho,l} = O_{HL,l} \\ O_{\rho,l} = O_{KLS,l} \\ O_{HL,l} = O_{KLS,l} \\ O_{\rho,l} = O_{HL,l} = O_{KLS,l} \end{cases} \\ i = 0 & \text{otherwise} \end{cases} \quad (33)$$

Thus, the l-th ranking position $R_{mv,l}$ is assigned to the input that has been ranked as l-th by at least 2-ou-of-3 methods; if for one l-th position none of the methods agree, no input is placed in that ranking position. The main advantages of this aggregation are its simplicity and negligible computational burden.

On the other hand, the R_{sum} aggregation (Kukkonen et al., 2007) consists in taking for each l-th input variable the sum of the ranking positions provided by the individual methods and, then, sorting them with respect to $R_{sum,l}$, viz:

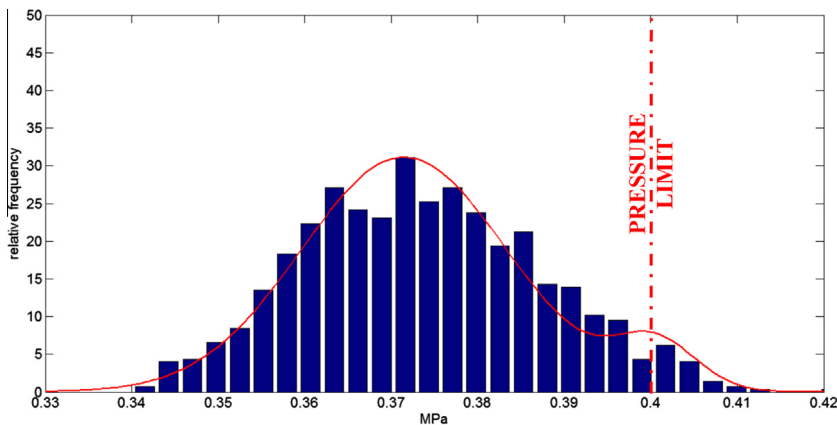$$R_{sum,l} = O_{\rho,l} + O_{HL,l} + O_{KLS,l} \quad (34)$$

Fig. 4. Histogram of the model output and mixture model reconstruction with $N = 1000$ code runs.

4. Results

The SA ensemble approach described in Section 3 is applied to the case study of Section 2. In order to test the effect of different amounts of available information, different numbers of LOCAs are simulated with $N = 1000, 600, 400, 200, 100$, samples of the input variables values drawn from the distributions reported in Appendix A.

The distribution (histogram) of the output variable $P_{containment}$ and its FMM (line) obtained using two Gaussian distributions $f_k \theta_k, k = 1, 2$, are shown in Figs. 4–8 for $N = 1000, 600, 400, 200$,

and 100 TH code simulations, respectively. The parameters of the mixture model found for different sizes of N are reported in Table 1. The choice of using two Gaussian distributions for the FMM parameters identification is the result of a trial and error procedure, but automatically optimizing the number K of distributions to be used in the FMM is also possible (Figueiredo et al., 2002). It is worth pointing out that resorting to the FMM method for any application requiring a number K of distributions larger than two is straightforward (Di Maio et al., 2014).

It can be seen that the multinomial pdf of $P_{containment}$, $f(P_{containment})$ is well reconstructed by all the FMMs built with a

Fig. 5. Histogram of the model output and mixture model reconstruction with $N = 600$ code runs.

Fig. 6. Histogram of the model output and mixture model reconstruction with $N = 400$ code runs.
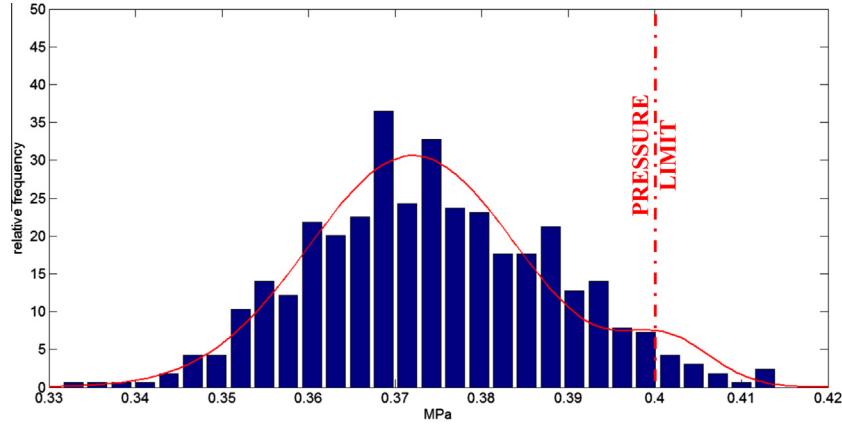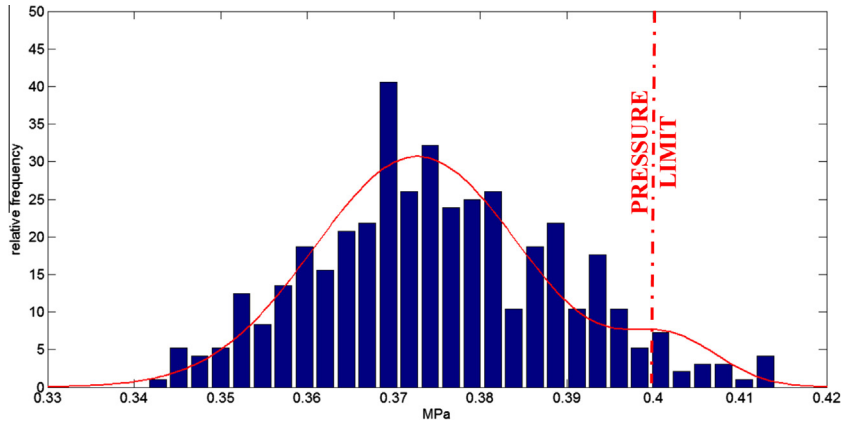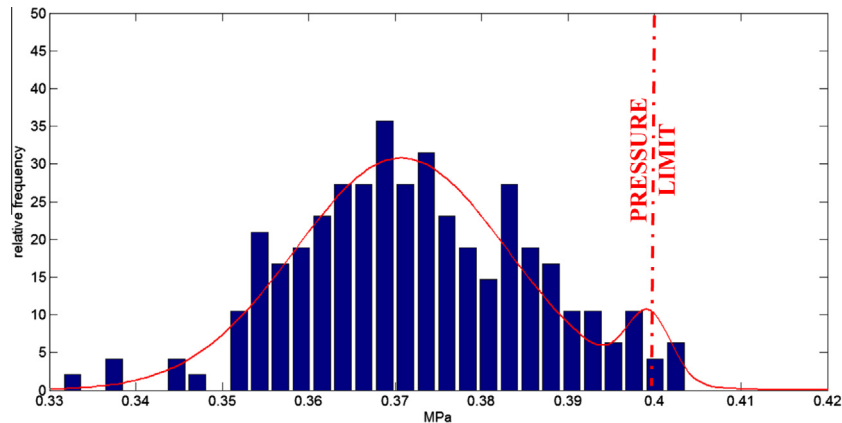

Fig. 7. Histogram of the model output and mixture model reconstruction with $N = 200$ code runs.
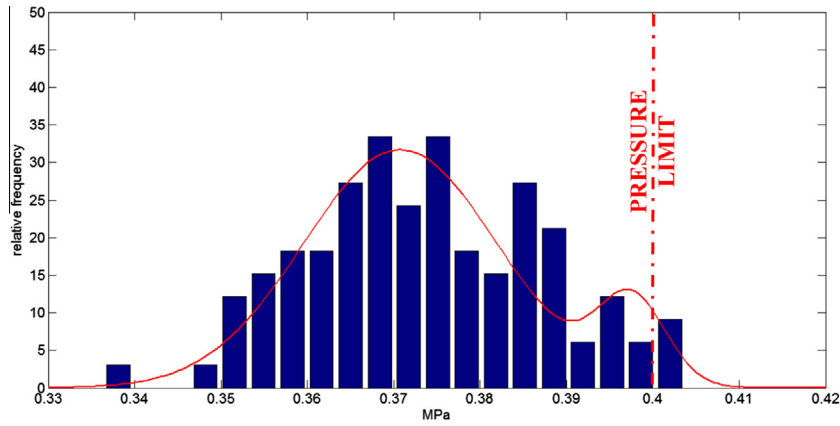
Fig. 8. Histogram of the model output and mixture model reconstruction with $N = 100$ code runs.

Table 1
Parameters of the finite mixture of Gaussian distributions computed with the EM algorithm for different sample sizes.

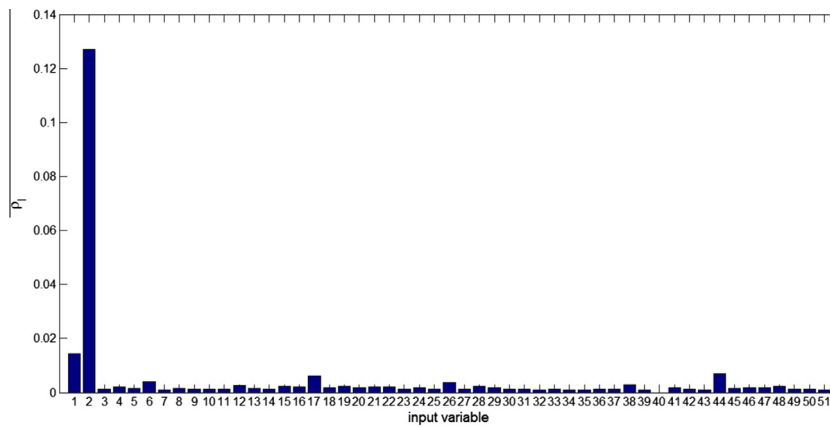| Sample size, N	Probabilities (π_1, π_2)	Means (μ_1, μ_2)	Standard deviations (σ_1, σ_2)
1000	(0.93, 0.07)	(0.3715, 0.4006)	(0.0119, 0.0047)
600	(0.93, 0.07)	(0.3720, 0.4015)	(0.0121, 0.0052)
400	(0.92, 0.08)	(0.3727, 0.4022)	(0.0119, 0.0057)
200	(0.94, 0.06)	(0.3707, 0.3993)	(0.0122, 0.0025)
100	(0.93, 0.07)	(0.3717, 0.3998)	(0.0119, 0.0030)

Fig. 9. Input variables saliencies (Eq. (22)) obtained with $N = 1000$ TH code runs.
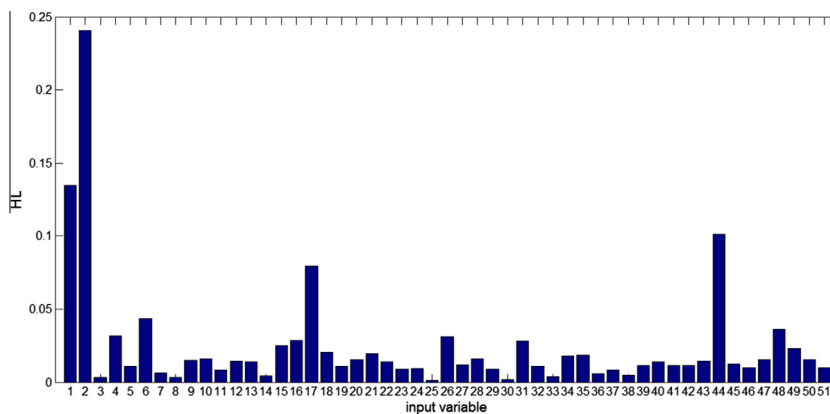

Fig. 10. Hellinger distance (Eq. (29)) obtained with $N = 1000$ TH code runs.
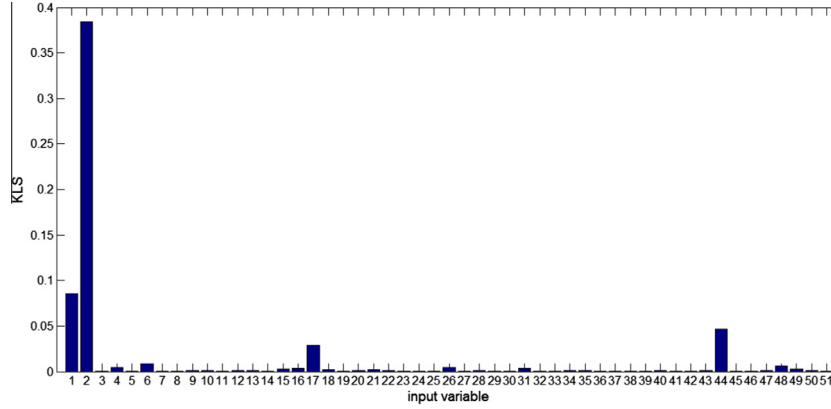
Fig. 11. Kullback–Leibler divergence (Eq. (32)) obtained with $N = 1000$ TH code runs.

Table 2
Input variables ranking obtained with $N = 1000$ TH code runs.

	1°	2°	3°	4°
Saliency, ρ_l (Eq. (22))	T_{inlet} (0.127)	G (0.014)	α_1 (0.007)	D_3 (0.006)
Hellinger distance, HL_l (Eq. (29))	T_{inlet} (0.241)	G (0.135)	α_1 (0.101)	D_3 (0.079)
Kullback–Leibler, KLS_l (Eq. (32))	T_{inlet} (0.384)	G (0.086)	α_1 (0.046)	D_3 (0.029)
Ensemble R_{mv}	T_{inlet}	G	α_1	D_3
Ensemble R_{sum}	T_{inlet}	G	α_1	D_3

number of code runs from $N = 1000$ down to $N = 100$. The case of $N = 100$ brings a considerable time saving since $N = 1000$ TH code runs take 277.3 h on an Intel Core2Duo P7550, whereas $N = 100$ only 28 h.

4.1. Sensitivity analysis

The three SA methods illustrated in Section 3 have been applied exploiting the analytical pdf obtained by the FMMs and the related clustering. In general terms, the methods backtrack the contribution of the inputs variables pdfs into the K clusters in which the analytical output pdf, obtained by the FMM, can be represented. Their individual rankings and that obtained by their ensemble are analyzed in the following, for the different cases of code runs form $N = 1000$ to $N = 100$.

Figs. 9–11 show ρ_l, HL_l and KLS_l for the $L = 51$ input variables, calculated using Eqs. (22), (29), and (32), respectively, on the basis of $f(P_{containment})$ obtained with $N = 1000$ TH code runs. The results of this case are taken as reference, i.e., as correct ranking. The ranking is reported in Table 2 along with the values of the sensitivity indexes (in the brackets). All the three methods and their ensemble (with both aggregation R_{mv} and R_{sum}) identify input variable T_{inlet}

($l = 2$) as the most important one and, in fact, agree on the ranking of the four most important input variables T_{inlet}, G, α_1 and D_3 ($l = 2$, 1, 44, 17, respectively). These input variables play a direct role in the heat transfer process of the PCCS operation. In particular, T_{inlet} (i.e., the temperature of the ultimate heat sink) is related to the capability of the atmosphere to absorb the heat generated, G (i.e., the steam mass flow rate) is related with the energy entering the containment, α_1 (i.e., the conductivity of the containment) has an active role in determining the heat flux leaving the containment, and D_3 (i.e., the containment diameter) determines the internal volume and the heat exchange surface, both critical quantities for internal pressure and heat transfer.

Table 3 shows the results obtained with $N = 400$ TH code runs: the ranking of the first four most important input variables agrees with that of Table 2 obtained with $N = 1000$.

Table 4 shows that when $N = 200$, the three methods all still identify the previous four variables as most important but no longer produce the same ranking order. Only the input saliency method is able to correctly identify the first three most influencing input variables, T_{inlet}, G, and α_1, but misses the fourth which is indicated to be the $l = 42$, H_c, (D_3 is ranked 11th); the other two methods switch the position of D_3 and α_1 in the 3rd and 4th positions.

Table 3
Input variables ranking obtained with $N = 400$ TH code runs.

	1°	2°	3°	4°
Saliency, ρ_l (Eq. (22))	T_{inlet} (0.081)	G (0.032)	α_1 (0.015)	D_3 (0.012)
Hellinger distance, HL_l (Eq. (29))	T_{inlet} (0.246)	G (0.142)	α_1 (0.118)	D_3 (0.075)
Kullback–Leibler, KLS_l (Eq. (32))	T_{inlet} (0.357)	G (0.097)	α_1 (0.061)	D_3 (0.025)
Ensemble R_{mv}	T_{inlet}	G	α_1	D_3
Ensemble R_{sum}	T_{inlet}	G	α_1	D_3

Table 4
Input variables ranking obtained with $N = 200$ TH code runs.

	1°	2°	3°	4°
Saliency, ρ_l (Eq. (22))	T_{inlet} (0.666)	G (0.104)	α_1 (0.079)	H_c (0.040)
Hellinger distance, HL_l (Eq. (29))	T_{inlet} (0.285)	G (0.108)	D_3 (0.100)	α_1 (0.064)
Kullback–Leibler, KLS_l (Eq. (32))	T_{inlet} (0.764)	G (0.058)	D_3 (0.047)	α_1 (0.018)
Ensemble R_{mv}	T_{inlet}	G	D_3	α_1
Ensemble R_{sum}	T_{inlet}	G	α_1	D_3

The ensemble of methods overcomes the problem: both ensemble aggregated by majority voting and sum of rankings identify all four most important variables and the latter provides the correct ranking: T_{inlet}, G, α_1, D_3.

The rankings produced on the basis of $N = 100$ TH code simulations are shown in Table 5. It is seen that the ensembles remain able to identify the three most important inputs, T_{inlet}, G and D_3 with both aggregation strategies, whereas there is no longer agreement on the fourth ranking position. This results from the fact that the Kullback–Leibler divergence again switches the positions of D_3 and α_1, the Hellinger distance mistakes the 4th ranked input and the input saliency mistakes the 3rd and 4th ranked inputs. In particular, the input saliency method starts misjudging the importance of the input variables when N approaches 200 code runs (in Table 4, H_c is considered as 4th important input, whereas it is not if compared with the ranking order of Table 2). This is due to the fact that the input saliency method is based on the EM algorithm that is sensible to the initialization of the indexes $\rho^{(1)}$ (see Eq. (22)) when N is reduced (Figueiredo et al., 2002). On the other hand, even with small N, it is important to keep the input saliency method in the ensemble because, with respect to the other two methods it is the only one that correctly assigns the right ranking order to the first three most important input variables T_{inlet}, G, and α_1 with $N = 200$ (see Table 4 in comparison to the ranking order of Table 2).

Finally, for further comparison, the variance decomposition-based method with $N = 15,600$ correctly identifies only T_{inlet}, and G as the two most important input variables, and with $N = 910$ it is not able to recognize anyone of the first four most important input variables (see Table 6) (Di Maio et al., 2014b) (for further details on the variance decomposition method, see Appendix B).

As for some concluding remarks drawn from the case study analyzed, it appears that: (i) all three SA methods used are computationally more efficient than variance-based decomposition analysis ($N = 910$ for variance-based decomposition analysis takes ~270 h on an Intel Core2Duo P7550, whereas $N = 200$ takes ~55 h), (ii) the ensembles proposed are effective with small numbers N (e.g. $N = 200$), compensating the errors of the individual methods, (iii) the R_{sum} aggregation strategy as provided the best results, especially with $N = 200$, as shown in Table 4. For the case study here analyzed, our suggestion is, thus, to resort to the ensemble of SA methods and assembling the different ranking orders by the R_{sum} aggregation strategy.

5. Conclusions

In this paper, we have presented a novel framework for performing the sensitivity analysis of passive safety systems TH codes, at reduced computational cost. The framework consists in estimating a Gaussian FMM to retrieve the analytical pdf of the model output with as few simulations as possible and to induce a clustering of the output variable space; then, two ensemble strategies have been applied to aggregate three sensitivity methods, namely input saliency, Hellinger distance and Kullback–Leibler divergence.

We have tested the framework to a long-running TH code that simulates the behaviour of a Passive Containment Cooling System (PCCS) of an Advanced Pressurized reactor AP1000 during a Loss Of Coolant Accident (LOCA).

The results obtained show the capability of the framework in discerning between influent and negligible input variables at a reasonable computational cost, especially when relying on the R_{sum} aggregation strategy.

Appendix A.

Table 7

Appendix B.

For simplicity of illustration, and without loss of generality, let us consider a model m whose output value y depends only on the values x_1 and x_2 of two uncertain input parameters X_1 and X_2, viz:

$$y = m(x_1, x_2) \tag{a1}$$

No hypotheses are made on the structure of the model.

We consider a set of s realizations of the two input parameters drawn from the assigned pdfs $q(x_1)$, $q(x_2)$, respectively:

$$\bar{x}^j = \left[x_1^j, x_2^j\right] \quad j = 1, 2, \ldots s \tag{a2}$$

The model is evaluated for each of the s independently generated vectors $\bar{x}^j$, $j = 1, 2, \ldots, s$, to obtain a corresponding set of output values:

Table 5
Input variables ranking obtained with $N = 100$ TH code runs.

	1°	2°	3°	4°
Saliency, ρ_l	T_{inlet} (0.539)	G (0.172)	Z_1 (0.100)	Z_2 (0.088)
Hellinger distance, HL_l	T_{inlet} (0.289)	G (0.138)	D_3 (0.135)	H_1 (0.102)
Kullback–Leibler, KLS_l	T_{inlet} (0.415)	G (0.079)	D_3 (0.077)	α_1 (0.043)
Ensemble R_{mv}	T_{inlet}	G	D_3	–
Ensemble R_{sum}	T_{inlet}	G	H_1	D_3

Table 6
Variance decomposition ranking for the lumped PCCS-TH code (Di Maio et al., 2014a).

N (code runs)	1°	2°	3°	4°
15,600	T_{inlet}	G	f_5	H_4
910	d	f_1	T_{st}	H

$$y^j = m\left(x_1^j, x_2^j\right) \quad j = 1, 2, \ldots s \qquad (a3)$$

Such set represents an independent random sample of size s of the distribution of the output y. Therefore, the dependence of the value of the output variable (Y) on the value of one of the two input variables, e.g. X_1, can be approximated by the expected value of Y with respect to the other variable X_2, conditioned on X_1 being equal to a given value x_1:

$$Y^*(x_1) = E_{X_2}(Y|x_1) = \int m(x_1, x_2) q_{X_2|X_1}(x_2|x_1) dx_2 \qquad (a4)$$

where $q_{X_2|X_1}(x_2|x_1)$ is the conditional probability density of X_2 given X_1. Note that, since X_1 is fixed at x_1, y^* depends only on the variable X_2.

To evaluate how the uncertainty in the input propagates to the output of the model, the variance of the distribution of the output variable Y is decomposed as follows:

$$Var[Y] = Var_{X_1}\left[E_{X_2}(Y|X_1)\right] + E_{X_1}\left[Var_{X_2}(Y|X_1)\right] \qquad (a5)$$

where X_1 has been indicated explicitly as subscript of the variance and expectation operators to highlight that these are applied with respect to such variable. The sensitivity relevance of X_1 can be associated to its contribution to the output variance, i.e. the term $Var_{X_1}\left[E_{X_2}(Y|X_1)\right]$ in (a5). Quantitatively, it is then customary to take the following measure as an index of the importance of the variable X_1 with respect to its contribution to the uncertainty in the output Y:

$$\eta_1^2 = \frac{Var_{X_1}\left[E_{X_2}(Y|X_1)\right]}{Var[Y]} \qquad (a6)$$

Table 7
List of input variables and their distributions.

	Input variable	Description	Units	Type of distribution	Mean value, μ	Standard deviation, σ	$\mu-4\sigma$	$\mu+4\sigma$
1	G	Steam mass flow rate	kg/s	Normal	182	5	162	202
4	T_{steam}	Steam temperature	°C	Normal	163	8.15	130.4	195.6
5	t_{w1}	Containment wall thickness	m	Normal	0.052	0.00026	0.051	0.053
6	t_{w2}	Containment wall thickness	m	Normal	0.043	0.000215	0.04214	0.04386
7	t_{w3}	Containment wall thickness	m	Normal	0.043	0.000215	0.04214	0.04386
8	t_{w4}	Containment wall thickness	m	Normal	0.015	0.000075	0.0147	0.0153
9	t_{w5}	Containment wall thickness	m	Normal	0.015	0.000075	0.0147	0.0153
10	D_1	Diameter of uphead	m	Normal	43	0.215	42.14	43.86
11	H_1	Height of uphead	m	Normal	13.51	0.06755	13.24	13.78
12	D_2	Diameter of containment	m	Normal	43	0.215	42.14	43.86
13	H_2	Height level of operative plant layer	m	Normal	0	0.05	−0.2	0.2
14	H_3	Height level of the blade	m	Normal	17.25	0.08625	16.905	17.595
15	H_4	Height level of the containment bottom	m	Normal	−11	0.055	−11.22	−10.78
16	H_5	Containment height	m	Normal	46.58	0.2329	45.65	47.51
17	D_3	Containment diameter	m	Normal	43	0.215	42.14	43.86
18	V	Containment volume	m³	Normal	75000	750	72000	78000
19	A_1	Area in air baffle	m²	Normal	42.28	0.4228	40.5888	43.9712
20	A_2	Area in air baffle downcomer	m²	Normal	131.1	1.311	125.856	136.344
21	Z_1	Height of the downcomer air baffle	m	Normal	32.51	0.16255	31.86	33.16
22	Z_2	Height of the riser air baffle	m	Normal	32.51	0.16255	31.86	33.16
23	A_3	Area in air baffle riser	m²	Normal	125.8	1.258	120.768	130.832
24	A_4	Area in air baffle riser	m²	Normal	34	0.34	32.64	35.36
25	A_5	Area in air baffle riser	m²	Normal	156.97	1.5697	150.6912	163.2488
26	A_6	Area in air baffle downcomer intake	m²	Normal	113	1.13	108.48	117.52
27	A_7	Area at the blade turn	m²	Normal	68.49	0.6849	65.7504	71.2296
28	A_8	Area at diffusers level	m²	Normal	463.1	4.631	444.576	481.624
29	A_9	Area at the inlet of the chimney	m²	Normal	41.53	0.4153	39.8688	43.1912
30	A_{10}	Area at the exhaust of the chimney	m²	Normal	74.82	0.7482	71.8272	77.8128
42	H_c	Height of chimney	m	Normal	8.27	0.04135	8.1046	8.4354
43	ρ_1	Central containment layer density	kg/m³	Normal	7750.476	38.75238	7595.47	7905.49
44	α_1	Central containment layer conductivity	W/(m K)	Normal	51.9	5.19	31.14	72.66
45	Cp_1	Central containment layer heat capacity	J/(kg K)	Normal	447.9876	2.239938	439.03	456.95
46	ρ_2	Covering layer density	kg/m³	Normal	3324.15	16.62075	3257.66	3390.63
47	α_2	Covering layer conductivity	W/(m K)	Normal	0.52246	0.052246	0.31	0.73
48	Cp_2	Covering layer heat capacity	J/(kg K)	Normal	544.284	2.72142	533.39	555.16

	Input variable	Description	Unit	Type of distribution	Lower value	Upper value		
2	T_{inlet}	External air temperature	°C	Seasonal	2	39		
3	P_{inlet}	External air pressure	MPa	Seasonal	0.0984	0.1011		

	Input variable	Description	Unit	Type of distribution	Mean value, μ	Lower value	Upper value	
31	f_1	Air baffle friction factor	–	Uniform	1.15	1.035	1.265	
32	f_2	Air baffle friction factor	–	Uniform	3.74	3.366	4.114	
33	f_3	Air baffle friction factor	–	Uniform	1.6	1.44	1.76	
34	f_4	Air baffle friction factor	–	Uniform	0.5	0.45	0.55	
35	f_5	Air baffle friction factor	–	Uniform	1.13	1.017	1.243	
36	f_6	Air baffle friction factor	–	Uniform	0.5	0.45	0.55	
37	f_7	Air baffle friction factor	–	Uniform	3.9	3.51	4.29	
38	f_8	Air baffle friction factor	–	Uniform	1	0.9	1.1	
39	f_9	Air baffle friction factor	–	Uniform	3.68	3.312	4.048	
40	f_{10}	Air baffle friction factor	–	Uniform	2.76	2.484	3.036	
41	f_{11}	Air baffle friction factor	–	Uniform	1.27	1.143	1.397	
49	T_{water}	Pool cooling water temperature	°C	Uniform	25	1	32	
50	u_{air}	Air baffle intake air speed	m/s	Uniform	2	0.5	4	
51	M_{vapor}	Initial in-containment vapor mass	kg	Uniform	36,600	33,600	39,600	

References

Acceptance criteria for emergency core cooling systems for light-water nuclear power reactors, NRC Regulations.

Baraldi, P., Zio, E., Gola, G., Roverso, D., Hoffmann, M., 2011. Two novel procedures for aggregating randomized model ensemble outcomes for robust signal reconstruction in nuclear power plants monitoring systems. Annals of Nuclear Energy 38, 212–220.

Bolado-Lavin, R., Castaings, W., Tarantola, S., 2009. Contribution to the sample mean plot for graphical and numerical sensitivity analysis. Reliability Engineering and System Safety 94, 1041–1049.

Tarantola, S., Kopustinskas, V., Bolado-Lavin, R., Kaliatka, A., Uspuras, E., Vaisnoras, M., 2012. Sensitivity analysis using contribution to sample variance plot: application to a water hammer model. Reliability Engineering and System Safety 99, 62–73.

Borgonovo, E., 2007. A new uncertainty importance measure. Reliability Engineering and System Safety 92, 771–784.

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. Nuclear Engineering and Design 119, 1–15.

Burgazzi, Luciano, 2004. Evaluation of uncertainties related to passive systems performance. Nuclear Engineering and Design 230, 93–106.

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

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

Cadini, F., Zio, E., Di Maio, F., Kopustinkas, V., Urbonas, R., 2007. A neural-network-based variance decomposition sensitivity analysis. International Journal of Nuclear Knowledge Management 2 (3), 299–312.

Carlos, S., Sánchez, A., Ginestar, D., Martorell, S., 2013. Using finite mixture models in thermal-hydraulics system code uncertainty analysis. Nuclear Engineering and Design 262, 306–318.

Cummins, W.E., Corletti, M.M., Schulz, T.L., 2003. Westinghouse AP1000 Advanced Passive Plant. In: Proceedings of ICAPP '03 Cordoba, Spain, May 4–7, 2003 Paper 3235. Westinghouse Electric Company, LLC.

Nicolas Devictor, Ricardo Bolado Lavín, 2005. Uncertainty and sensitivity methods in support of PSA level 2. In: 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, F., Hu, J., Tse, Peter W., Pecht, M., Tsui, K., Zio, E., 2012. Ensemble-approaches for clustering health status of oil sand pumps. Expert Systems with Applications 39, 4847–4859.

F. Di Maio, G. Nicola, E. Zio, 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, F., Nicola, G., Zio, E., Yu, Y., 2014b. Finite mixture models for sensitivity analysis of thermal hydraulic codes for passive safety systems safety analysis. Nuclear Engineering and Design.

Diaconis, Persi, Zabell, Sandy L., 1982. Updating subjective probability. Journal of the American Statistical Association 77 (380), 822–830.

Fang, Shoufan, Gertner, George Z., Shinkareva, Svetlana, Wang, Guangxing, Anderson, Alan, 2003. Improved generalized Fourier amplitude sensitivity test (FAST) for model assessment. Statistics and Computing 13 (3), 221–226.

Figueiredo, M., Jain, A.K., 2002. Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (3), 1–16.

Gibbs, Alison L., Edward Su, Francis, 2002. On choosing and bounding probability metrics. International Statistical Review 70 (3), 419–435.

Glaeser, H., Hofer, E., Kloos, M., Skorek, T., 1994. Uncertainty and sensitivity analysis of a post-experiment calculation in thermal hydraulics. Reliability Engineering and System Safety 45 (1/2), 19–33.

Guba, A., Makai, M., Pal, L., 2003. Statistical aspects of best estimate method-I. Reliability Engineering and System Safety 80, 217–232.

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering and System Safety 42, 327–367.

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.

Kukkonen, Saku, Lampinen, Jouni, 2007. Ranking-dominance and many-objective optimization. Evolutionary Computation, 3983–3990.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79–86. http://dx.doi.org/10.1214/aoms/1177729694.

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.

Law, Martin H.C., Figueiredo, Mario A.T., Jain, Anil K., 2004. Simultaneous feature selection and clustering using mixture models. IEEE Transactions on Patterns Analysis and Machine Intelligence 26 (9), 1154–1166.

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.

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

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

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

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

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

Pudil, P., Novovicová, J., Kittler, J., 1995. Feature selection based on the approximation of class densities by finite mixtures of the special type. Pattern Recognition 28 (9), 1389–1398.

Farzad Choobdar Rahim, Mohammad Rahgoshay, Seyed Khalil Mousavian, 2011. A study of large break LOCA in the AP1000 reactor containment.

Saltelli, A., Chan, K., Scott, E., 2000. Sensitivity Analysis. John Wiley & Sons Inc., New York.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al., 2008. Global Sensitivity Analysis: The Primer. John Wiley & Sons Ltd., Chichester.

Schulz, T.L., 2006. Westinghouse AP1000 advanced passive plant. Nuclear Engineering and Design 236, 1547–1557.

Secchi, P., Zio, E., Di Maio, F., 2008. Quantifying uncertainties in the estimation of safety parameters by using bootstrapped artificial neural networks. Annals of Nuclear Energy 35 (12), 2338–2350. http://dx.doi.org/10.1016/j.anucene.2008.07.010.

Vaithyanathan, S., Dom, B., 1999. Generalized model selection for unsupervised learning in high dimensions. In: Advances in Neural Information Processing Systems. MIT Press, Cambridge, Mass., pp. 970–976.

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. Nuclear Engineering and Design 119, 17–31.

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. Nuclear Engineering and Design 119, 33–65.

Yu,Y., Liu, T., Tong, J., Zhao, J., Di Maio, F., Zio, E., Zhang, A., 2010. Multi-experts analytic hierarchy process for the sensitivity analysis of passive safety systems. In: Proceedings of the 10th International Probabilistic Safety Assessment & Management Conference, PSAM10, Seattle, June.

Yu Yu, Shengfei Wang, Fenglei Niu, 2013. Thermal–hydraulic performance analysis for AP1000 passive containment cooling system. In: Proceedings of the 21th International Conference on Nuclear Engineering ICONE21 July 29-August 2, Chengdu, Sichuan, China.

Zio, E., Di Maio, F., 2008. Bootstrap and order statistics for quantifying thermal-hydraulic code uncertainties in the estimation of safety margins. Science and Technology of Nuclear Installations 2008, 9. http://dx.doi.org/10.1155/2008/340164.

Zio, E., Di Maio, F., Martorell, S., Nebot, Y., 2008, Neural networks and order statistics for quantifying nuclear power plants safety margins. In: Proceedings, European Safety & Reliability Conference (ESREL), Valencia, Spain.

Zio, E., Pedroni, N., 2009. Estimation of the functional failure probability of a thermal-hydraulic passive system by Subset Simulation. Nuclear Engineering and Design 239, 580–599.

Zio, E., Pedroni, N., 2010. How to effectively compute the reliability of a thermal–hydraulic nuclear passive system.

Zio, E., Di Maio, F., Tong, J., 2010b. Safety margins confidence estimation for a passive residual heat removal system. Reliability Engineering and System Safety, RESS 95, 828–836.