



日本原子力研究開発機構機関リポジトリ
Japan Atomic Energy Agency Institutional Repository

Title	An Integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents
Author(s)	Zheng X., Ito Hiroto, Tamaki Hitoshi, Maruyama Yu
Citation	Journal of Nuclear Science and Technology, 53(3), p.333-344
Text Version	Author's Post-print
URL	https://jopss.jaea.go.jp/search/servlet/search?5050372
DOI	https://doi.org/10.1080/00223131.2015.1044262
Right	This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Nuclear Science and Technology on March 2016, available online: http://www.tandfonline.com/10.1080/00223131.2015.1044262 .



ARTICLE

**An integrated approach to source term uncertainty and sensitivity analysis
for nuclear reactor severe accidents**

Xiaoyu Zheng^{a*}, Hiroto Itoh^a, Hitoshi Tamaki^a and Yu Maruyama^a

^aJapan Atomic Energy Agency, 2-4, Shirakata Shirane, Tokai-mura, Naka-gun,

Ibaraki-ken, 319-1195, Japan

Large-scale computer programs simulate severe accident phenomena and often have a moderate-to-large number of models and input variables. Analytical solutions to uncertainty distributions of interested source terms are impractical, and influential inputs on outputs are hard to discover. Runs of such integral codes for complex severe accidents, are generally time-consuming. This article presents an integrated approach to uncertainty and sensitivity analysis for nuclear reactor severe accident source terms, with an example which simulates an accident sequence similar to that occurred at Unit 2 of the Fukushima Daiichi nuclear power plant using an integral code, MELCOR. Monte Carlo based uncertainty analysis has been elaborated to investigate released fractions of representative radionuclides, Cs and CsI. In order to estimate sensitivity of inputs, which have a substantial influence on the core melt progression and the transportation process of radionuclides, a variance decomposition method is applied. Stochastic process, specifically a Dirichlet process, is applied to construct a surrogate model in sensitivity analysis as a substitute of the code. The surrogate model is cross-validated by comparing with corresponding results of MELCOR. The analysis with the simpler model avoids laborious computational cost/load so that importance measures for input factors are obtained successfully.

* Corresponding author. Email: zheng.xiaoyu@jaea.go.jp

Keywords: Fukushima Daiichi NPP accident; source term; uncertainty analysis; sensitivity analysis; variance decomposition; Bayesian nonparametric modeling; Dirichlet process

1. Introduction

When fission products released from a degraded core, they are transformed, both physically and chemically, prior to becoming the “ingredients” of the source term. The definition of “source term” is provided as: the quantity, time, history, and chemical and physical form of radionuclides released to the environment, or present in the containment atmosphere, during the course of a severe accident [1]. The analysis of source term is crucial, both in severe accident analysis and in probabilistic risk assessment of nuclear reactors, since it is an essential quantity for the evaluation of environmental and health-related effects. Precise estimating of the source term can provide more reliable assessment of radiological consequences. Numerical simulation using integral codes is a general way to estimate the source term when large-scale experiments are hard to conduct and relevant data are scarce.

There exist uncertainties in the numerical results, caused by undetermined inputs, model structures or even programming errors. To simplify the present analysis, only parametric uncertainty is considered. Influential input variables are required to be identified to determine uncertainty sources. However, when such codes are rather complex, it is practically impossible to obtain analytical solutions. Let us review the definition of uncertainty and sensitivity analysis at first, and then go on to the proposed methods. Specifically, the uncertainty analysis deals with assessing the uncertainty in model predictions, which may result from imprecisely known input variables, alternative model assumptions and unknown randomness. Monte Carlo based approximation is a currently feasible way to estimate the uncertainty range of source term during a severe accident and then to find out influential variables. Such approaches have been widely applied to reveal the uncertainty in the results originated from variations in the inputs [2]. Sensitivity analysis is to study how the uncertainty in the output can be apportioned to different source in the inputs. The intent of sensitivity analysis is to identify key variables whose uncertainty affects most the output [3].

In order to investigate the uncertainty in severe accident source term and the sensitivity of

corresponding key input variables, an integrated approach has been established at Japan Atomic Energy Agency (JAEA). Methods of uncertainty analysis are elaborated, which consists of four steps: 1) factor screening to reduce the number of inputs, 2) random sampling of inputs using Latin Hypercube sampling (LHS) combined rank correlation, 3) numerical computation of a representative severe accident via integral severe accident codes, for example, THALES-2 developed by JAEA [4][5] and MELCOR by Sandia National Laboratories [6], and 4) verification of uncertainty distributions through Kullback-Leibler (KL) divergence.

Generally, a great number of code runs are required to achieve a stable result of global sensitivity analysis using indices such as Sobol' sensitivity index (a variance decomposition method) [7][8]. The required number of code runs is unacceptable for complex severe accident codes. In order to solve this problem, the sensitivity analysis of the proposed approach consists of three steps: 1) the construction of a surrogate model to execute the function of an integral code, 2) the cross-validation of the surrogate model, and 3) global sensitivity analysis using the surrogate model.

The simpler surrogate model is constructed using Bayesian nonparametric methods. Specifically, Dirichlet processes, one widely-known type of Bayesian nonparametric models, are used to construct the stochastic surrogate model. Since such models are not parametric, we do not have to worry about the fitness of the model to the data (e.g. a linear model is fitted to strongly non-linear data) or poor predictability as a result of overfitting (a deterministic high-order model can fit data well but shows poor predictive ability) [9][10]. Compared with traditional Bayesian models, Dirichlet process is an infinite mixture of probability functions so that complex and multimodal data can be well modelled. The model is rather flexible and the degrees of freedom can be optimized while data are analyzed. The degrees of freedom here refer to the number of parameters in a numerical or statistical model. Moreover, the complexity of one Dirichlet process model, the same as other Bayesian nonparametric models,

will grow as more data are observed [11]. The predictability of the surrogate model is cross-validated by comparing with the results of actual codes. As a result, first order effect and total effect of Sobol' sensitivity indices are calculated and relative importance ranking of all inputs are obtained accordingly. This proposed approach for source term uncertainty and sensitivity analysis shows great advantage and reliability by using probabilistic methods. The prototype of severe accident scenario analyzed is the accident similar to that occurred at Unit 2 of the Fukushima Daiichi Nuclear Power Plant (NPP), which is used as an example for the demonstration of the present approach to source term uncertainty and sensitivity analysis.

2. Framework of source term uncertainty and sensitivity analysis

Uncertainty and sensitivity analyses are always run in tandem. Preceding uncertainty analysis quantitatively reveals the variation in the estimation of source terms during a severe accident in this article. Following sensitivity analysis focuses on what are the uncertainty sources. The detailed design of the presented integrated approach is shown in **Figure 1**. This process of uncertainty and sensitivity analysis is established for the evaluation of severe accident source term. Corresponding methods are summarized in this section and the detailed implementation and analysis with an example are illustrated in the next section.

After the selection of a number of potentially influential uncertain inputs, the uncertainty analysis in **Figure 1** consists of four steps.

- (1) Step 1: factor screening using the elementary effect (EE) method [12][13]. A computational model can be viewed as a function $y = f(\mathbf{X})$ of inputs $\mathbf{X} = (x_1, x_2, \dots, x_N)^T$ that produces an output y . "Factor screening" or "preliminary sensitivity analysis" is to determine which inputs have important effects on an output, by means of low computational cost/load. The efficiency is important since severe accident codes have thousands of input variables and runs of such codes are time-consuming, for example, one simulation via MELCOR would take around

several hours so that the distribution of computational resources is rather important. The elementary effect method, which follows a randomized one-factor-at-a-time design, is efficient with limited number of code runs which is a linear function of the number N of input variables. Important inputs are selected from a large number of candidates.

- (2) Step 2: random sampling using a combined method of LHS and rank correlation. Monte Carlo approaches have been widely applied to severe accident uncertainty analysis recently. By chance, using an original Monte Carlo method may end up with some points clustered closely, resulting in an over-representation of certain regions and under-representation of others. LHS forces the samples to spread out across all portions of the sample space [14]. In a simulation study, a desired correlation matrix is expected to be induced to the sampling scheme, either independence or dependence. One distribution-free approach is to combine the Latin Hypercube sampling strategy with rank correlation [15]. The introduction of correlation will not destroy the structure of LHS and ensure that all marginal distributions of inputs remain intact. Sampling strategies are also applied and evaluated in the risk assessment of other nuclear facilities [16][17].
- (3) Step 3: numerical computation via integral severe accident codes. Integral computer codes model the progression of severe accidents in light water reactor NPP, including a broad spectrum of severe accident phenomena. In this article, representative source terms (cesium and cesium iodide) are estimated based on the setting of accident sequences similar to that occurred at Unit 2 of the Fukushima Daiichi NPP as well as corresponding physical models. Multiple code runs are performed with random samples generated from Step 2.
- (4) Step 4: stability verification of the uncertainty distributions using KL divergence. KL divergence is a non-symmetric measure of the difference between two probability

distributions [18]. The KL divergence is zero when two distributions are identical. The aim of this step is to figure out how many cases of simulation are required to achieve stable uncertainty distributions for source terms. Wilk's formula are used in previous publications to determine tolerance limits, smallest and largest items of sampled outputs in an uncertainty analysis [19][20] whereas shapes of uncertainty distributions are still not clear. We introduce an alternative index of KL divergence to prove that the generated probability distributions can represent uncertainties in source term estimates.

In order to reveal how the uncertainty of outputs can be apportioned according to uncertain input variables, the variance decomposition method (Sobol' sensitivity indices) is used as a measure of importance. However, in order to obtain a converged result of Sobol' importance ranking, a large sample size is required, which is narrated as one main drawback of the method [3]. The simulation with integral severe accident codes is rather time-consuming so that the problem of computational cost gets worse. We propose to construct a stochastic model to replace the code. If the stochastic model can predict outputs as well as the actual code with limited computational cost, the sensitivity analysis can be implemented using the surrogate model. To solve these problems, the sensitivity analysis in **Figure 1** consists of three steps.

- (1) Step 1': the construction of a surrogate model through Dirichlet process, a Bayesian nonparametric model. This technique of stochastic modeling is flexible about how to choose a model at an appropriate level of complexity. The structure of model, the number of clusters in mixture models, grows and adapts to data in a Bayesian way [11]. When more observations of numerical simulation are gained, the predictability of the stochastic model will be strengthened and the model can be used as a substitute executor. The sensitive variables to output uncertainties for integral codes are also sensitive for the surrogate model. A brief review of Dirichlet process mixture

and weight dependent Dirichlet process is provided in Appendices A and B, respectively.

- (2) Step 2': cross-validation. To estimate how accurately the stochastic model will perform in prediction, we use an independent data set to test. The output dataset, results from severe accident codes, is partitioned into two complementary subsets, one is used for the construction of the surrogate model and the other is used for cross-validation. If the prediction of the model is similar to that by actual code, the predictability of the model is cross-validated.
- (3) Step 3': sensitivity analysis via the variance decomposition method (Sobol' sensitivity indices). When the "simpler" model is validated as reliable, all datasets are used to construct a complete surrogate model, by which the global sensitivity analysis is performed and important input factors can be identified.

Based on this integrated approach, let us analyze the source term uncertainties of a postulated accident sequence similar to that occurred at Unit 2 of the Fukushima Daiichi NPP, and then quantitatively identify importance of inputs.

3. An example of uncertainty analysis

The example is provided for easy-understanding of the present approach and a more detailed uncertainty analysis refers to previous studies by authors [13][21], including the setting of severe accident scenarios and plant conditions.

3.1. Severe accident scenarios and analysis conditions

The accident sequence is similar to the accident at Unit 2 of the Fukushima Daiichi NPP. As a demonstration of the integrated approach, MELCOR 1.8.5 is used as the simulation tool. Other important supplementary materials can be found in reference [22]. The release fraction of initial inventories of radionuclides in core, calculated based on core inventory data for the

Peach Bottom Unit 2 [23] using ORIGEN 2 code, is used to survey the source term.

3.2. Factor screening

At the first step of uncertainty analysis, twenty seven uncertainty input variables are selected based on engineering judgment and then screened via preliminary sensitivity analysis using the elementary effect method. Nine inputs are from MELCOR COR package which affects the in-vessel melt progression and the other nineteen inputs are from RN package which affects the transportation of radioactive materials. The mathematical definition of the elementary effect method can be written as follows.

$$EE_i(\mathbf{X}) = \frac{f(x_1, \mathbf{L}, x_{i-1}, x_i + \Delta_i, x_{i+1}, \mathbf{L}, x_{27}) - f(x_1, \mathbf{L}, x_{i-1}, x_i, x_{i+1}, \mathbf{L}, x_{27})}{\Delta_i} \quad (1)$$

Here, f can be viewed as a representation of a code execution to produce a unique output based on specified values of inputs. $\mathbf{X} = (x_1, \mathbf{L}, x_{27})^T$ is the vector of input variables and specified values of \mathbf{X} for each evaluation are selected from several discretized candidates, respectively. Every time one input is selected to change with a Δ_i so that the influence of the input on output can be represented by the EE_i . Since the values of other inputs will affect the EE_i , several rounds of calculations in a randomized manner are performed. Two measures can be used to evaluate the importance of each input: mean (μ_i^*) of absolute values of all EE_i and the standard deviation (σ_i) of all EE_i . Two measures explain 1) whether the influence of an input on the output is great, or; 2) whether the influence of an input on the output is highly dependent on the values of other inputs or not. Since MELCOR is a deterministic code, the consideration of noise in outputs is not required.

Each round of preliminary sensitivity analysis includes 28 (27 for each input and one base case) code runs. Totally, six rounds are performed. Δ_i is the discretized value as 20% or -20% of the corresponding default value. Inputs of the base case are selected from the discretized spaces: $\{d - 2\Delta, d - \Delta, d, d + \Delta, d + 2\Delta\}$, where d is the default value. The

interested source terms are the release fraction of Cs and CsI.

The result of factor screening analysis is shown in **Figure 2** and totally ten input variables are identified as important ones for the uncertainty analysis. Larger mean values of elementary effects indicate inputs with an important “overall” influence on the output. Larger standard deviations indicate influence of inputs is non-linear and highly dependent on the values of others, involved in interaction effects. The definition and assigned distributions for all ten important inputs are listed in **Table 1**. The preliminary rankings of inputs for Cs and CsI show a slight difference, which is possibly caused by the transportation and chemisorption processes of iodine, so some radionuclide transportation related inputs affect more significantly the release of CsI. Among top ten input factors, six are related to the in-vessel melt progression modeling and the other four are related to the radionuclide transportation modeling.

3.3. Random sampling and numerical computations

The coupled method of LHS and rank correlation, also called Iman-Conover restricted pairing technique [17], is applied in the random sampling process. A desired rank correlation matrix among all inputs is determined beforehand and rank correlation coefficient between two correlated inputs is shown in **Table 1**. The method can be simply described that the independently sampled input data via LHS is re-arranged according to the desired matrix. The sampling with correlation brings the joint distribution of inputs closer to the true desired joint distribution. The specified samples would be fundamentally biased when attained under the assumption of independence.

Multiple evaluations using MELCOR are executed. Sixty LHS intervals are partitioned for the first round of uncertainty analysis and then 20 LHS partitions for each round in the further analysis till the obtained uncertainty distribution is stable enough to represent the interested source terms. Totally, two hundred code runs are performed and the stability

analysis of uncertainty distributions will be introduced in the following sub-section.

3.4. Verification of uncertainty distributions

Wilks formula is generally used to determine how many code runs are required to obtain limit values, e.g. the 95th percentile value with a confidence level of 95% is obtained by selecting number of code runs as 59 for the consideration of the one-sided tolerance limit. The upper limit can be decided with a relatively few number of code runs whereas it is still unknown whether the corresponding uncertainty distribution is stable. In order to find an appropriate number of code runs, we introduce the Kullback-Leibler divergence as an alternative index for the number of evaluations. The KL divergence is a measure of the difference between two distributions P and Q . When the KL divergence is small enough, the uncertainty distribution of more code runs can be considered as acceptably reasonable.

The continuous and discretized forms of KL divergence can be written as follows.

$$D_{KL}(P\|Q) = \int_{-\infty}^{+\infty} p(x) \ln \frac{p(x)}{q(x)} dx \stackrel{\text{discretized}}{=} \sum_i P(i) \ln \frac{P(i)}{Q(i)} \quad (2)$$

Here, p and q denote the probability density functions of P and Q . It can be observed from Equation (2) that the KL divergence is a non-symmetric measure. Typically, P represents the “true distribution” and Q an approximate of the “true distribution”. Because there is no “true distribution” available in the severe accident analysis, we use the density function of more code runs as P and that of less code runs as Q . The method of Dirichlet process mixtures is applied to obtain probability density functions of representative source terms based on multiple outputs. The detailed mathematical modeling process is provided in Appendix A. Probability density functions of different number of code runs are shown in **Figure 3**, where the release fraction of CsI corresponding to the initial inventory is taken as an example. When the number of cases increases, two adjacent curves become more similar in shape. It also shows an obvious difference in the interval between 3% and 4%

where the probability densities of less code runs are overestimated. The change of KL divergence for both Cs and CsI is shown in **Figure 4**. The change of KL divergence between two neighbored probability density functions become less when more code runs are performed so that the stability of uncertainty distributions can be confirmed. When 200 runs are executed, the KL divergences between two neighbored distributions reduce to 0.4 and 0.27, for Cs and CsI, respectively. The determination of an acceptable threshold of KL divergence is subjective and when both reduce to be less than 1, we think that the uncertainty distributions converge.

Probability density functions of representative source terms (Cs and CsI) are provided in **Figure 5**. Two curves describe the probability density of Cs and CsI. Two set of points describe the actual MELCOR results and points are jittered to avoid overlapping. The probability density functions show multiple modes so that mixture models are required to fit the data, as introduced in Appendix A. Overall release fraction of Cs is less than that of CsI. Credible intervals that cover 95% highest probability density region are: Cs [9.77E-05, 1.58E-02] and CsI [3.37E-04, 4.09E-02]. The failure status of pressure vessel is a key factor that results in the overall shape of source terms distributions shown as two-modal.

Meaning of representative indices in uncertainty analysis are summarized as: by using EE method, mean values show the importance of input factors and standard deviations show interaction among inputs; the change of KL divergence with the increase of code runs represents the convergence speed of uncertainty distributions; the probability density functions of source terms quantitatively show the uncertainty range of estimates.

4. An example of sensitivity analysis

A more detailed demonstration of the sensitivity analysis, with a simpler example of three input variables, can be found in a previous study by authors [24]. The example in the article is to prove the practicability of the method for a higher-dimensional problem, ten-dimensional surrogate model presently.

4.1. A review of Sobol' sensitivity indices

As we already know uncertainty distributions of source terms (Cs and CsI), it is of further interest to know which factor is the main source of the output uncertainty. Sensitivity analysis is a method to solve the problem. Let us review the calculation procedure of Sobol' sensitivity indices, which is based on the decomposition of output uncertainty. The output variance $V(y)$, one moment of a distribution which represents how far numbers spread out, can be decomposed by conditioning with respect to input variables.

$$V(y) = V(E(y|x_i)) + E(V(y|x_i)) \quad (3)$$

Here, x_i is an arbitrary input variable. The definition of first order effect of the element x_i can be written as

$$S_i = \frac{V(E(y|x_i))}{V(y)} \quad (4)$$

Here, V and E represent variance and expectation, respectively. S_i is the first order effect of input x_i on output y . The total effect of the element x_i can be written as follows, which include all terms of any order that include x_i .

$$V(y) = V(E(y|\mathbf{X}_{\sim i})) + E(V(y|\mathbf{X}_{\sim i})) \quad (5)$$

$$S_{Ti} = \frac{V(y) - V(E(y|\mathbf{X}_{\sim i}))}{V(y)} = \frac{E(V(y|\mathbf{X}_{\sim i}))}{V(y)} = S_1 + S_{12} + S_{13} + S_{123} + \dots \quad (6)$$

Here, $\mathbf{X}_{\sim i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N)^T$ is the vector of inputs with the exclusion of x_i . Higher order indices can be written as

$$S_{ij} = \frac{V(E(y|x_i, x_j))}{V(y)} = S_i + S_j + S_{ij} \quad (1 \leq i < j \leq N) \quad (7)$$

Based on these sensitivity indices, the importance of all input variables can be quantitatively evaluated and ranked. However, in order to obtain converged results of Sobol' sensitivity analysis, thousands of code-runs are required. Thus, we propose an approach to

construct a surrogate model for the sensitivity analysis, so the model should generate approximate outputs compared with the original simulation tool.

4.2. The construction of a surrogate model

The estimation of source term can be abstractly represented as a function.

$$y = f(\mathbf{X}) = g(\mathbf{X}; \boldsymbol{\theta}) \quad (8)$$

Here, f is the computational tool that produces an output y based on inputs \mathbf{X} . $g(\mathbf{X}; \boldsymbol{\theta})$ is an approximate model that is built using statistical methods, so $g(\mathbf{X}; \boldsymbol{\theta})$ plays the same role that the actual tool plays, but with far cheaper computational cost. $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)^T$ is the vector of parameters used to build the mathematical model, e.g. all coefficients $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_N, \varepsilon)^T$ in a linear model $y = \beta_0 + \beta_1 x_1 + \dots + \beta_N x_N + \varepsilon$, or $\boldsymbol{\theta} = (\mu, \sigma^2)^T$ in a standard Gaussian model $y: N(\mu, \sigma^2)$, etc.

Based on Bayesian inference, the posterior distribution of parameters ($\boldsymbol{\theta}$) in the stochastic model can be written in an integrated Bayesian form.

$$p(\boldsymbol{\theta} | \mathbf{X}, y) = \frac{p(\boldsymbol{\theta}, \mathbf{X}, y)}{p(y)} = \frac{p(\boldsymbol{\theta}) p(y | \mathbf{X}, \boldsymbol{\theta})}{\int_{\mathbf{x}, \boldsymbol{\theta}} p(\boldsymbol{\theta}) p(y | \mathbf{X}, \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{X}} \quad (9)$$

The predictive posterior distribution of new interested output (\hat{y}) given corresponding input set ($\hat{\mathbf{X}}$) can be written accordingly.

$$\begin{aligned} p(\hat{y} | \mathbf{X}, y, \hat{\mathbf{X}}) &= \int_{\boldsymbol{\theta}} p(\hat{y} | \boldsymbol{\theta}, \mathbf{X}, y, \hat{\mathbf{X}}) d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} p(\hat{y} | \mathbf{X}, y, \hat{\mathbf{X}}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{X}, y) d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} p(\hat{y} | \hat{\mathbf{X}}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{X}, y) d\boldsymbol{\theta} \end{aligned} \quad (10)$$

As the source term uncertainty results shown in **Figure 5**, outputs show nonlinear properties and cannot be represented by any standard stochastic model, e.g. normal distribution, etc. Models of higher-order polynomials are available but too deterministic in the form and show properties of overfitting and poor predictability, for example, the predictions and real value are entirely distinct. We introduce a method using a mixture of some standard

distributions but without the determination of freedom of degree beforehand. The mixture model is one of Bayesian nonparametric model and is called Dirichlet process mixture, which is introduced in Appendix A and applied to density estimation of MELCOR results. More introduction can be found in references [10][11], classic reference on Dirichlet process [25][26], Dirichlet process mixtures [27], and dependent Dirichlet model [28]. A doctoral dissertation also provides easy-understanding introduction of Dirichlet process related methods [29].

When outputs are predicted for new inputs, a weight dependent Dirichlet process model is applied, as introduced in Appendix B. A package of R language, DPpackage, is used in the prediction process in the present research [30] and corresponding introduction of weight dependent Dirichlet process can be found in references [31][32]. The degree of freedom (the number of mixture components: K) and corresponding parameters of mixture components (e.g. the mean vector and covariance matrix, for a mixture of normal) are optimized in a Bayesian way, as introduced in Equations (9) and (10). Using the weight dependent Dirichlet process model, we constructed a flexible and well-fitted surrogate model with reliable predictability.

4.3. Cross-validation of the surrogate model

In order to prove the predictability of the surrogate model, a cross-validation is performed. Based on the randomly-selected 190 cases from data during the uncertainty analysis, a stochastic model has been built. Inputs of the other ten cases are used to compare. If the predicted output of the ten cases shows agreement with the actual code outputs, the model can be proved as acceptably accurate. The results of cross-validation are shown in **Figure 6**. Black curves are predicted released fraction of CsI based on certain inputs via the surrogate model and black dots are the MELCOR results based on the same inputs. The code outputs appear nearly around the mode of predicted distributions. It proved that the stochastic model is

reliable to be used as an alternative tool for source term analysis. It should be mentioned that the prediction of low-dimensional inputs (with fewer variables) is of less uncertainty and higher accuracy than that of high-dimensional problems (with more variables).

4.4. Global sensitivity analysis using Sobol' sensitivity indices

According to the introduction of Sobol' sensitivity analysis, variance of outputs and variance of conditional expectations are used to estimate the importance of all input variable. Based on results of MELCOR, the variance of CsI can be calculated: $V(y) = 1.577E - 4$.

Based on Equation (10), a continuous posterior predictive distribution $p(y|X, y, X^c)$ with one specified input X^c can be obtained. In order to apply the prediction into sensitivity analysis, a point estimate of the unobserved quantity is required. In Bayesian statistics, a maximum a posterior (MAP) estimate is to use the mode of the predictive posterior distribution as the point estimate of the most probable result. It can be written in a mathematical form.

$$\hat{y}_{MAP} = \arg \max_{y} p(y|X, y, X^c) \quad (11)$$

Thousands of input-output evaluations are performed to obtain Sobol' sensitivity measures for input variables, and the computational cost is rather lower than that of MELCOR computations. The results are provided in **Table 2**. The global sensitivity analysis of all ten inputs provides a reference importance measure. Estimation error exists for two reasons. The first is the prediction error existed with the use of a surrogate stochastic model. The surrogate model provides a practically available way to the sensitivity analysis but introduce the approximated error. The second is the point estimate of the posterior predictive distribution. The MAP estimation is a simplification of the results and innate uncertainties exist for all probability distributions. The sensitivity analysis provides a reference ranking from the perspective of substantial influence of inputs on outputs. The most important input variable is

the TZXMX that is one key parameter to influence the failure mechanism of fuel rod claddings. The second and third are DELDIF and TRDFAI, related to the aerosol dynamic process and the fuel failure models, respectively. Without the aid of stochastic tools, importance measure with global sensitivity analysis of complex code would be a difficult task.

5. Conclusions

An integrated approach to source term uncertainty and sensitivity analysis has been developed at JAEA for the risk assessment of nuclear reactor severe accidents. The full-scope approach improves the previous method for source term uncertainty analysis. The introduction of a surrogate model solves the impracticality of global sensitivity analysis for source term.

This integrated approach consists of two main procedures: uncertainty analysis and sensitivity analysis. There are four steps of actions in uncertainty analysis: factor screening using an individually randomized one-factor-at-one-time method; random sampling using Latin Hypercube sampling with the consideration of correlation among inputs; numerical computation via integral severe accident codes; stability analysis using Kullback-Leibler divergences between uncertainty distributions to attain the required number of code runs. Three steps of actions are involved into the global sensitivity analysis: the construction of a surrogate model based on weight dependent Dirichlet process, a nonparametric Bayesian methodology; cross-validation of the model with new inputs; sensitivity analysis using variance decomposition method (Sobol' sensitivity indices).

As a demonstration, 200 MELCOR code runs are executed for an accident sequence similar to that occurred at Unit 2 of the Fukushima Daiichi NPP. Stable uncertainty distributions of representative source terms are obtained, in non-standard and multi-modal shapes. Highest probability density regions of source terms can provide boundary conditions

for the evaluation of off-site radiological consequences. The relative importance measures of ten inputs are ranked. Important physical models, related to models of fuel and fuel cladding failure and aerosol dynamic process, are identified as of substantial influences on the source term. This integrated approach shows advantages in solving some previously neglected problems in uncertainty and sensitivity analysis of complex severe accident codes.

References

- [1] Sehgal BR. Nuclear safety in light water reactors: severe accident phenomenology. Waltham, Massachusetts, U.S.: Academic Press, Elsevier Inc.; 2012.
- [2] Helton JC. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliab Eng Syst Safe.* 1993; 42: 327-367.
- [3] Homma T, Saltelli A. Importance measure in global sensitivity analysis of nonlinear models. *Reliab Eng Syst Safe.* 1996; 52: 1-17.
- [4] Ishikawa J, Kawaguchi K, Maruyama Y. Analysis for iodine release from unit 3 of Fukushima Dai-ichi nuclear power plant with consideration of water phase iodine chemistry. *J Nucl Sci Technol.* 2014; published online;
- [5] Kajimoto M, Kuramatsu K, Watanabe N, Funasako M, Noguchi T. Development of THALES-2, A computer code for coupled thermal-hydraulics and fission product transport analyses for severe accident at LWRs and its application to analysis of fission product revaporization phenomena. In: *Proceedings of International Topical meeting on Safety of Thermal reactors.* 1991 July 21-25; Oregon, Portland.
- [6] Gauntt RO, Cole RK, Erickson CM, Gido RG, Gasser RD, Rodriguez SB, Young MF. MELCOR computer code manuals: Vol.1 & Vol.2, Rev.2. United States: U.S. Nuclear Regulatory Commission; 2000; NUREG/CR-6119.
- [7] Sobol' I. Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling and Computational Experiments.* 1993; 1:407-414.
- [8] Saltelli A, Tarantola S, Campolongo F, Ratto M. *Sensitivity analysis in practice: a guide to assessing scientific models.* Chichester, West Sussex: John Wiley & Sons Ltd.; 2004.
- [9] Rasmussen CE, Williams CKI. *Gaussian processes for machine learning.* Cambridge, Massachusetts: The MIT Press; 2006.
- [10] Hjort NL, Holmes C, Müller P, Walker SG. *Bayesian nonparametrics.* Cambridge: Cambridge University Press; 2010.

- [11]Gershman SJ, Blei DM. A tutorial on Bayesian nonparametric models. *J Math Pshychol* 2012; 56; 1-12.
- [12]Morris MD. Factorial sampling plans for preliminary computational experiments. *Technometrics*. 1991 May; 33(2): 161-174.
- [13]Zheng X, Itoh H, Tamaki H, Maruyama Y. Estimation of source term uncertainty in a severe accident with correlated variables. In: *Proceedings of the 22th international conference on nuclear engineering (ICONE-22)*. 2014 July 7-11; Prague, Czech Republic.
- [14]Mckay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 1979 May; 21: 239-245.
- [15]Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. *Commun Stat B-Simul*. 1982; 11(3): 311-334.
- [16]Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab Eng Syst Safe*. 2003; 81: 23-69.
- [17]Helton JC, Johnson JD, Sallaberry CJ, Storlie CB. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab Eng Syst Safe*. 2006; 91: 1175-1209.
- [18]Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical recipes: the art of scientific computing*. Cambridge: Cambridge University Press; 2007.
- [19]Wilks SS. Statistical prediction with special references to the problem of tolerance limits. *Annals of Mathematical Statistics*. 1942; 13(4): 400-409.
- [20]Glasser H. GRS method for uncertainty and sensitivity evaluation of code results and applications. *Sci Technol Nucl Ins*. 2008; 2008: ID 798901.
- [21]Zheng X, Itoh H, Tamaki H, Maruyama Y. Source term uncertainty analysis: probabilistic approaches and applications to a BWR severe accident. *Mechanical Engineering Journal*. Forthcoming 2015.
- [22]Hirano M, Yonomoto T, Ishigaki M, Watanabe N, Maruyama Y, Sibamoto Y, Watanabe T,

- Moriyama K. Insights from review and analysis of the Fukushima Dai-ichi accident. *J Nucl Sci Technol.* 2012; 49(1): 1-17.
- [23] Gieseke JA, Cybulskis P, Denning RS, Kuhlman MR, Lee KW. Radionuclide release under specific LWR accident conditions. United States: U.S. Nuclear Regulatory Commission; 1984; BMI-2014.
- [24] Zheng X, Itoh H, Kawaguchi K, Tamaki H, Maruyama Y. Application of Bayesian nonparametric models to the uncertainty and sensitivity analysis of source term in a BWR severe accident. *Reliab Eng Syst Safe.* 2015; 138: 253-262.
- [25] Ferguson TS. A Bayesian analysis of some nonparametric problems. *Ann Stat.* 1973; 1(2): 209-230.
- [26] Sethuraman J. A constructive definition of Dirichlet priors. *Stat Sinica.* 1994; 4: 639-650.
- [27] Antoniak CE. Mixtures of Dirichlet process with applications to Bayesian nonparametric problems. *Ann Stat.* 1974; 2(6): 1152-1174.
- [28] MacEachern SN. Dependent nonparametric processes. In: *ASA proceedings of the section on Bayesian statistical science.* 1999; Virginia, United States.
- [29] Sudderth EB. Graphical models for visual object recognition and tracking [dissertation]. Massachusetts: Massachusetts Institute of Technology; 2006.
- [30] Jara A, Hanson TE, Quintana FA, Müller P, Rosner GL. DPpackage: Bayesian semi- and nonparametric modeling in R. *J stat Softw.* 2011; 40(5).
- [31] Müller P, Erkanli A, West M. Bayesian curve fitting using multivariate normal mixtures. *Biometrika.* 1996; 83(1): 67-79.
- [32] Duson DB, Pillai N, Park J. Bayesian density regression. *J Roy Stat Soc B Met.* 2007; 69(2): 163-183.

Appendix A. A finite mixture model and Dirichlet process mixture

A brief introduction of Dirichlet process and weight dependent Dirichlet process is provided for a quick review. In nonparametric Bayesian statistics, Dirichlet processes are commonly used to learn mixture models whose number of components is not fixed, but instead from data. The Dirichlet process is the infinite form of a finite mixture model, which can be written as

$$p(y) = \sum_{k=1}^K \pi_k p(y|\phi_k) \quad (12)$$

Here, the finite number K is the number of all components, which are also notated as clusters in data analysis. Data from the same cluster share one unique set of parameters ϕ_k . Accordingly, the vector of parameters (θ) in Equations (9) and (10) can be written as $\theta = \{\pi_1, \dots, \pi_K; \phi_1, \dots, \phi_K\}$. Each data point is generated by independently selecting one of K clusters according to a multinomial distribution and then sampling from the chosen cluster's data distributions.

$$\begin{aligned} c_i &: \text{Multi}(\boldsymbol{\pi}), c_i \in \{1, \dots, K\}, \boldsymbol{\pi} = \{\pi_1, \dots, \pi_K\} \\ y_i &: F(\phi_{c_i}) \end{aligned} \quad (13)$$

With the using of mixture models, the stochastic model of any dataset can be a composite of a number of standard distributions, by which complex and multimodal datasets can be modeled. **Figure 7** illustrates the finite mixture model. The multinomial distribution of $\boldsymbol{\pi} = \{\pi_1, \dots, \pi_K\}$ have a conjugate Dirichlet prior with a precision parameter α so the posterior distribution of $\boldsymbol{\pi}$ is still a Dirichlet.

$$\boldsymbol{\pi} : \text{Dir}\left(\frac{\alpha}{K}, \mathbf{K}, \frac{\alpha}{K}\right) \quad (14)$$

Parameters of each cluster are sampled from a hyper-prior distribution with hyper-parameters λ .

$$\phi_k : H(\lambda) \quad (15)$$

As shown in **Figure 8**, mixture models can equivalently be expressed in terms of a discrete distribution (G) on the space (Φ) of cluster parameters.

$$\begin{aligned}
G(\phi) &= \sum_{k=1}^K \pi_k \delta_{\phi_k} \\
\bar{\phi}_i &: G \\
y_i &: F(\bar{\phi}_i)
\end{aligned} \tag{16}$$

The base function of parameters of each cluster is denoted as G_0 instead of $H(\lambda)$. In nonparametric Bayesian setting, K and ϕ_k are not determined beforehand but by data itself. The partitions of data are optimized by Bayesian inference so that the complexity and accuracy of the inferred model will grow as more data is observed. Dirichlet process mixture can be written in an infinite form as

$$\begin{aligned}
p(y) &= \sum_{k=1}^{\infty} \pi_k p(y|\phi_k) \\
G(\phi) &= \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \\
\bar{\phi}_i &: G \\
y_i &: F(\bar{\phi}_i)
\end{aligned} \tag{17}$$

The discrete distribution drawn from a Dirichlet process can be denoted in a simpler way.

$$G : DP(\alpha, G_0) \tag{18}$$

The weight $\{\pi_1, \pi_2, \dots\}$ can be expressed as the stick-breaking construction.

$$\begin{aligned}
\pi_k &= \beta_k \prod_{j=1}^{k-1} (1 - \beta_j), k = 1, 2, \dots, \infty \\
\beta_k &: Beta(1, \alpha)
\end{aligned} \tag{19}$$

Appendix B. Weight dependent Dirichlet process

Weight dependent Dirichlet process is used to predict outputs based on inputs. The mixture model can be expressed in the form that weights and cluster parameters are dependent on inputs or covariates. The influence of covariates on the output can be reflected via linear mean

functions and a set of covariate dependent weights.

$$g(y) = \sum_{k=1}^{\infty} \pi_k(\mathbf{x}) N(y | \gamma_{0k} + \mathbf{x}^T \boldsymbol{\gamma}_k, \sigma_k^2) \quad (20)$$

It can be written in an integral form with multivariate normal distributions.

$$\begin{aligned} (y_i, \mathbf{x}_i)^T &: \int N_D((y_i, \mathbf{x}_i) | \boldsymbol{\mu}, \boldsymbol{\Sigma}) dG(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ G &: DP(\alpha, G_0) \end{aligned} \quad (21)$$

Here, D is the dimension of the vector of complete dataset $(y_i, \mathbf{x}_i)^T$ and in the present research $D=11$. G_0 is the baseline distribution which is assumed as a Normal-Inverted-Whisart (IW) distribution.

$$G_0 \equiv N_D(\boldsymbol{\mu} | \mathbf{m}_1, \kappa_0^{-1} \boldsymbol{\Sigma}) IW_D(\boldsymbol{\Sigma} | \nu_1, \boldsymbol{\Psi}_1) \quad (22)$$

Here, prior parameters are assumed: $\nu_1=12$ and α is a gamma distribution with hyper-prior parameters: $a_0=10$ and $b_0=5$.

$$\alpha : \Gamma(a_0, b_0) \quad (23)$$

The assumptions of other hyper-parameters are listed as

$$\begin{aligned} \mathbf{m}_1 &: N_D(\mathbf{m}_2, \mathbf{S}_2) \\ \kappa_0 &: \Gamma(\tau_1/2, \tau_2/2) \\ \boldsymbol{\Psi}_1 &: IW_D(\nu_2, \boldsymbol{\Psi}_2) \end{aligned} \quad (24)$$

In the present article, \mathbf{m}_1 is the mean vector: $\mathbf{m}_1 = (\overline{y_i}, \overline{\mathbf{x}_i})^T$, $\nu_2=12$, $\tau_1=6.01$ and $\tau_2=2.01$. \mathbf{S}_2 is the covariance matrix of $(y_i, \mathbf{x}_i)^T$ and $\boldsymbol{\Psi}_2$ is the inverse matrix of \mathbf{S}_2 .

The covariate dependent weights and parameters in mean functions are defined as

$$\begin{aligned} \pi_k(\mathbf{x}) &= \frac{\pi_k N_D(\boldsymbol{\mu}_{2k}, \boldsymbol{\Sigma}_{22k})}{\sum_{j=1}^{\infty} \pi_j N_D(\boldsymbol{\mu}_{2j}, \boldsymbol{\Sigma}_{22j})} \\ \gamma_{0k} &= \mu_{1k} - \boldsymbol{\Sigma}_{12k} \boldsymbol{\Sigma}_{22k}^{-1} \boldsymbol{\mu}_{2k} \\ \boldsymbol{\gamma}_k &= \boldsymbol{\Sigma}_{12k} \boldsymbol{\Sigma}_{22k}^{-1} \\ \sigma_k^2 &= \sigma_{11k}^2 - \boldsymbol{\Sigma}_{12k} \boldsymbol{\Sigma}_{22k}^{-1} \boldsymbol{\Sigma}_{21k} \end{aligned} \quad (25)$$

Here, π_k follows the stick-breaking construction as introduced in Equation (19). The standard mean and covariance matrix in Equation (20) for each cluster k can be written as

$$\boldsymbol{\mu}_k = \begin{pmatrix} \mu_{1k} \\ \mu_{2k} \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_k = \begin{pmatrix} \sigma_{11k}^2 & \Sigma_{12k} \\ \Sigma_{21k} & \Sigma_{22k} \end{pmatrix}. \text{ Based on observations } \mathbf{y} \text{ and inputs } \mathbf{x}^T, \text{ the solving}$$

of the Dirichlet process model can be realized by using Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm.

Figure captions

Figure 1. The proposed approach to severe accident source term uncertainty and sensitivity analysis (gray symbols are essential steps)

Figure 2. Estimated elementary effects of 27 input variables for Cs and CsI with the elementary effect method

Figure 3. Probability density functions based on different numbers of code runs by using Bayesian nonparametric density estimation (Dirichlet process mixture of normals)

Figure 4. Kullback-Leibler divergences between two adjacent probability density functions

Figure 5. Probability distributions of representative source terms

Figure 6. Cross-validation of simulated posterior predictive distributions (solid curves) with actual MELCOR results (single black dots)

Figure 7. A finite mixture model

Figure 8. A Gaussian process mixture model

Table 1. Uncertain input variables selected according to the elementary effect method

	Inputs	Package	Description	Distribution
1	TRDFAI	COR	Temperature at which oxidized fuel rods can stand when unoxidized Zr is absent from the cladding (default = 2500.0 K).	Normal distribution ($\mu = 2500.0, \sigma = 250.0$); Range = [2100, 2990] Rank correlation: 0.5 with TZXMX
2	TSPCB	COR	Time constant for the channel-bypass relocation of debris after the failure of canisters (default = 1.0 s).	Uniform distribution; Range = [0.5, 1.5] Rank correlation: -0.25 with TZXMX
3	SC7155 (1)	RN	It is used to determine the decontamination factor for small Stocks number of the impaction process which is modeled by SPARC-90 particle impaction model (default = 1.79182).	Normal distribution ($\mu = 1.79182, \sigma = 1.53755$) Range = (0, $+\infty$)
4	TZXMX	COR	Maximum ZrO ₂ temperature permitted to hold up molten Zr (default = 2400.0 K).	Normal distribution ($\mu = 2400.0, \sigma = 240.0$) Range = [2100, 2990] Rank correlation: 0.5 with TRDFAI; -0.25 with TSPCB; -0.25 with HFRZZR
5	SC7160 (1, 1)	RN	Chemisorption rate coefficient for CsOH on the surface of stainless steel (default = 0.139 m/s).	Normal distribution ($\mu = 0.139, \sigma = 0.0417$) Range = (0, $+\infty$)
6	STICK	RN	Particle sticking coefficient which is one of miscellaneous coefficients used for the aerosol dynamic process (default = 1.0).	Uniform distribution Range = [0.5, 1.0]
7	DELDIF	RN	Diffusion boundary layer thickness which is one of miscellaneous coefficients used for the aerosol dynamic process (default = 1.0×10^{-5})	Uniform distribution Range = [$5.0 \times 10^{-6}, 1.5 \times 10^{-5}$]
8	TPFAIL	COR	Failure temperature of the penetration or the lower head. It is used as one of parameters to model the failure of the vessel lower head and its penetrations (default = 1273.15 K).	Normal distribution ($\mu = 1273.15, \sigma = 127.315$) Range = [1000, 1700]
9	TNSMAX	COR	Temperature above which nonsupporting structure (control blades or rods) will collapse, independent of remaining metal thickness (default = the melting point of steel which is the cladding of control blades or rods, 1700.0 K)	Normal distribution ($\mu = 1700.0, \sigma = 170.0$) Range = [1000, 1870]
10	HFRZZR	COR	Refreezing heat transfer coefficient for Zircaloy which is used in the candling model for the molten Zircaloy. (default = 1000.0 W/m ² ·K)	Normal distribution ($\mu = 100.0, \sigma = 300.0$) Range = (0, $+\infty$) Rank correlation: -0.25 with TZXMX

Table 2. Results of global sensitivity analysis and importance ranking

		V_j	$V_{\sim j}$	First order	Total	Importance
	Inputs	$j=1,2,L,10$	$j=1,2,L,10$	Effect	Effect	ranking
				(MAP)	(MAP)	(Total)
x_1	TRDFAI	5.396E-06	4.333E-05	0.034	0.725	3
x_2	TSPCB	5.337E-06	8.409E-05	0.034	0.467	9
x_3	SC7155(1)	4.930E-06	6.999E-05	0.031	0.556	4
x_4	TZXMX	5.403E-05	3.126E-05	0.343	0.802	1
x_5	SC7160(1,1)	5.705E-06	7.595E-05	0.036	0.518	5
x_6	STICK	1.806E-06	8.112E-05	0.011	0.486	8
x_7	DELDIF	9.533E-06	4.160E-05	0.060	0.736	2
x_8	TPFAIL	5.282E-06	9.429E-05	0.033	0.402	10
x_9	TNSMAX	5.313E-06	7.708E-05	0.034	0.511	6
x_{10}	HFRZZR	1.088E-05	7.803E-05	0.069	0.505	7

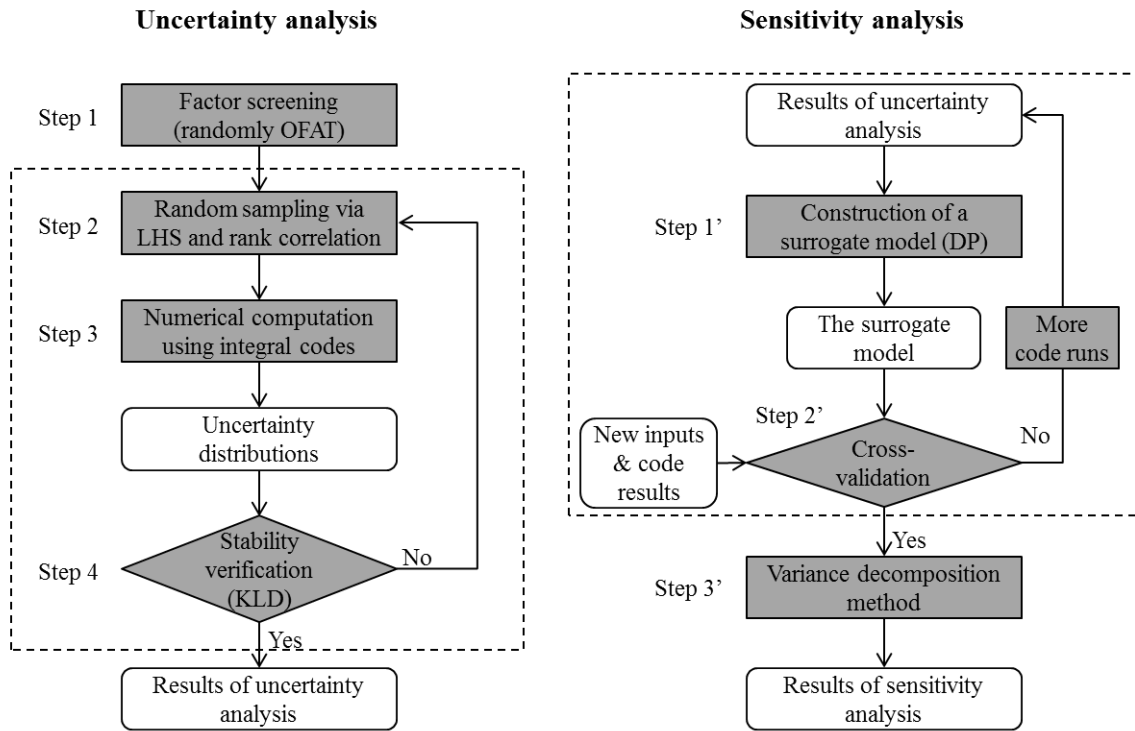


Figure 1. The proposed approach to severe accident source term uncertainty and sensitivity analysis (gray symbols are essential steps)

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents

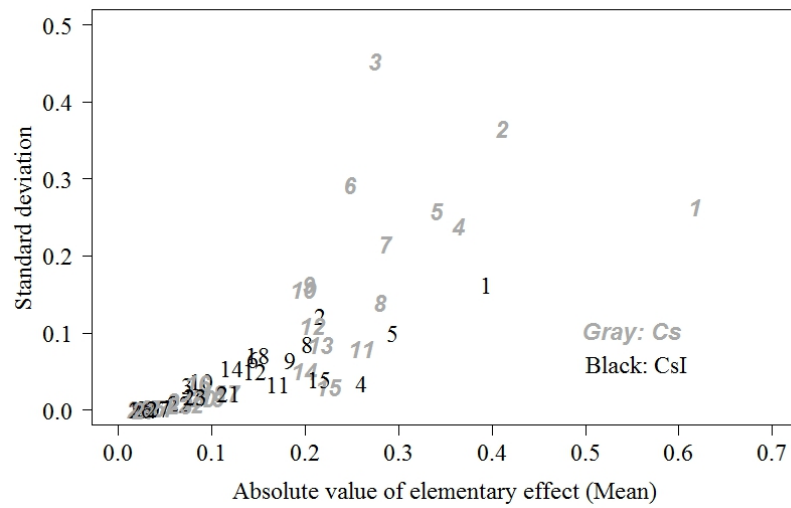
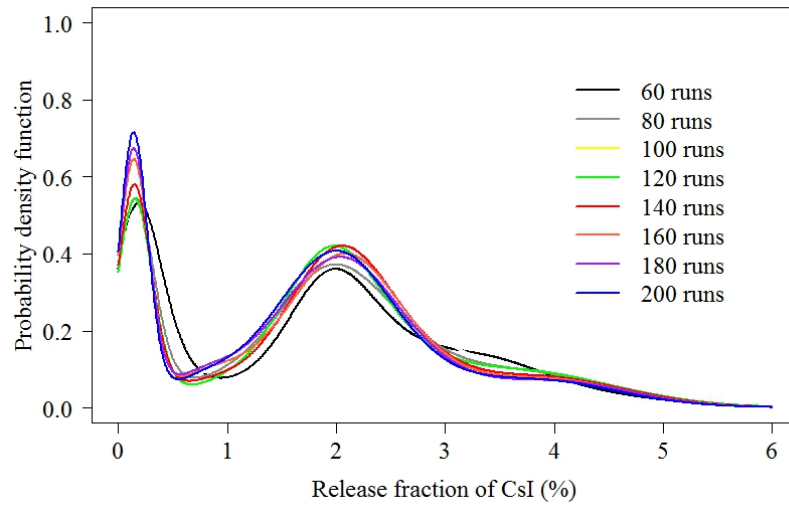


Figure 2. Estimated elementary effects of 27 input variables for Cs and CsI with the elementary effect method

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents



**Color
Print**

Figure 3. Probability density functions based on different numbers of code runs by using Bayesian nonparametric density estimation (Dirichlet process mixture of normals)

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents

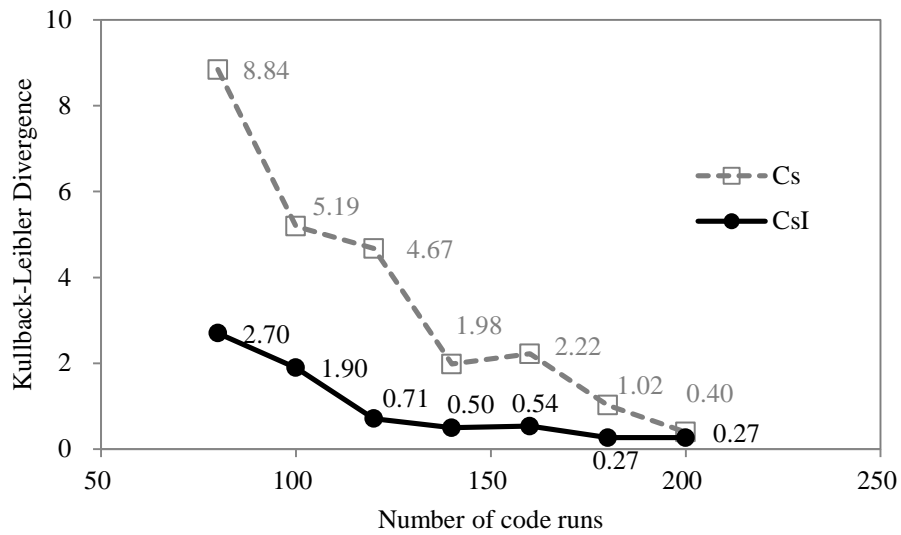


Figure 4. Kullback-Leibler divergences between two adjacent probability density functions

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents

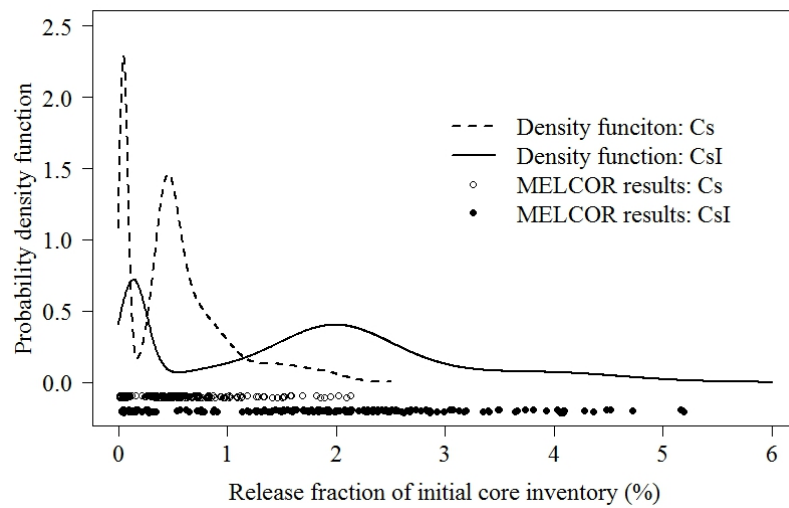


Figure 5. Probability distributions of representative source terms

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents

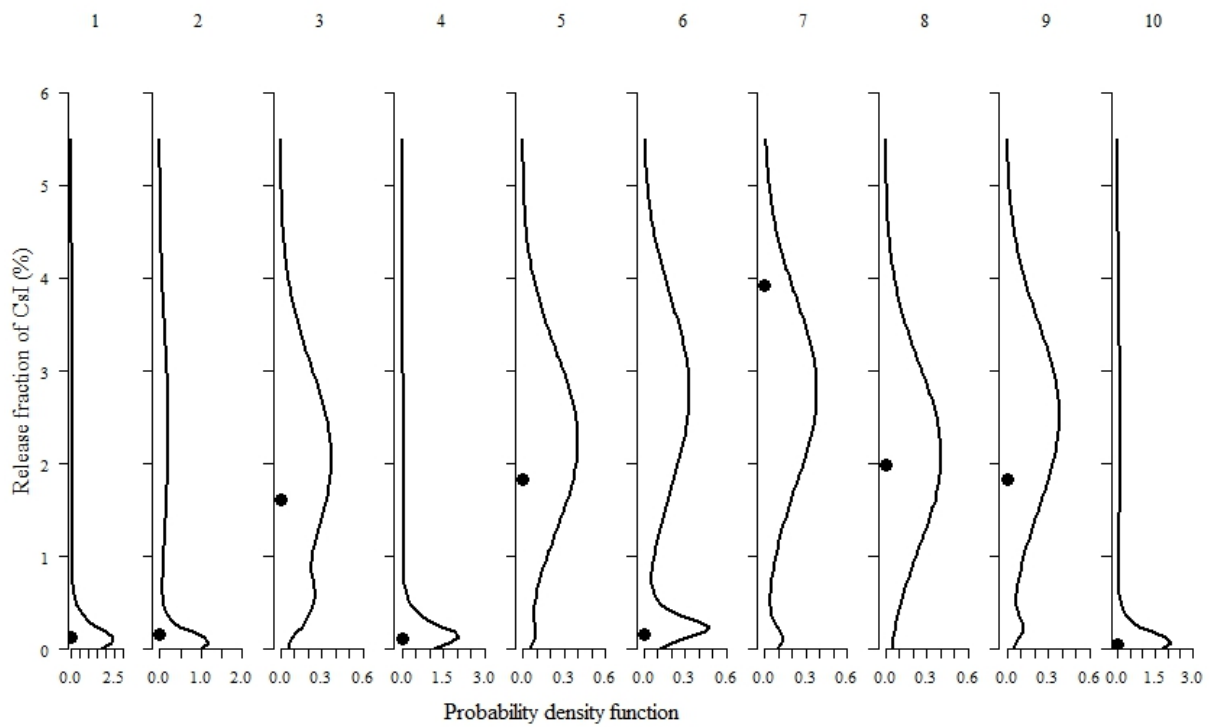


Figure 6. Cross-validation of simulated posterior predictive distributions (solid curves) with actual MELCOR results (single black dots)

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents

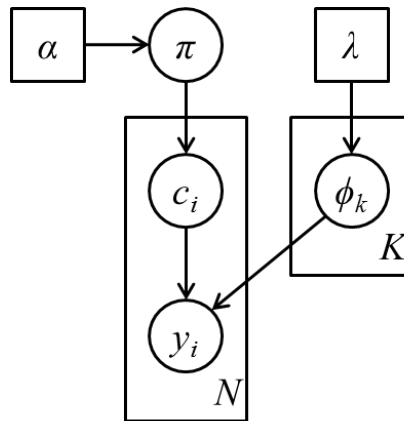


Figure 7. A finite mixture model

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents

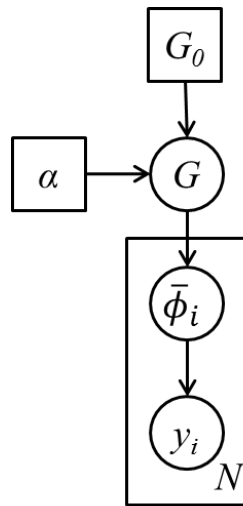


Figure 8. A Gaussian process mixture model

X. Zheng

An integrated approach to source term uncertainty and sensitivity analysis for nuclear reactor severe accidents