



# Propagation of epistemic uncertainties using dempster-shafer theory in bepu evaluation

M. Marques, H. Ynuhui, A. Marrel

## ► To cite this version:

M. Marques, H. Ynuhui, A. Marrel. Propagation of epistemic uncertainties using dempster-shafer theory in bepu evaluation. ANS Best Estimate Plus Uncertainty International Conference (BEPU 2018), May 2018, Lucca, Italy. hal-02415473

**HAL Id: hal-02415473**

**<https://hal.science/hal-02415473>**

Submitted on 17 Dec 2019

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

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

## PROPAGATION OF EPISTEMIC UNCERTAINTIES USING DEMPSTER-SHAFER THEORY IN BEPU EVALUATION

**M. Marquès, Y. Hou and A. Marrel**

CEA, DEN, Centre de Cadarache, F-13108 Saint-Paul-Lez-Durance, France

[michel.marques@cea.fr](mailto:michel.marques@cea.fr), [yunhui.hou@insa-cvl.fr](mailto:yunhui.hou@insa-cvl.fr), [amandine.marrel@cea.fr](mailto:amandine.marrel@cea.fr)

### ABSTRACT

BEPU evaluation is generally based on computer simulators such as thermal-hydraulic system codes with expansive computational cost. Furthermore, two types of uncertainties are present in the BEPU evaluation: the aleatory uncertainty which describes the natural variability of random events and the epistemic uncertainty due to lack of knowledge. The former is usually modelled by probability theory where some conditions on data quantity and quality must be satisfied for probability density functions (pdfs) fitting. However, while the epistemic uncertainty is taken into account, with less and imprecise available data (i.e. parameters of physical model and correlations present in the computer simulator), the use of probabilistic methods on uncertainty modelling and propagation cannot be always justified.

The Dempster-Shafer Theory (DST) of evidence provides an adapted framework for representing the parameters with epistemic uncertainty, when it is not possible to build a coherent probabilistic model from the available knowledge. In this theory, instead of pdfs, the parameters are modeled by focal sets with associated degrees of belief. Input uncertainty modeled by DST can be propagated within a BEPU evaluation by mapping input focal sets to the output space. The main problem is how to control the computation cost because the mapped results are obtained by finding the optimal values of the output in each input focal set.

In this paper, we propose a new scheme to propagate epistemic uncertainties modelled by DST through a time-consuming computer simulator. Besides classical Monte Carlo method using Cartesian product input mass construction method, we propose a novel procedure using vacuous dimension extension and mass combination rules (e.g. Dempster's combination rule for independent input variables ) after mapping the input focal sets into the output space. This method significantly reduces the number of input focal sets. The obtained output function has less focal sets, but larger range. Thus, a complete coverage of the output focal set space can be provided. Epistemic uncertainty is estimated globally in a conservative point of view. As a result of a trade-off between calculation cost, estimation accuracy and the quantity of details on epistemic uncertainty presented, our proposal enables propagation of epistemic uncertainties modeled by DST with a very limited computation budget making possible its practical use within a BEPU evaluation.

## 1. INTRODUCTION

The treatment of uncertainties in the analysis of complex system is essential for determining possible ranges on responses of interest such as safety margins or probabilities of exceeding failure criteria. Uncertainties can be categorized according to the character of their sources as either aleatory uncertainties, which represent the intrinsic randomness of a phenomenon, and are irreducible in nature, or epistemic uncertainties, which are reducible uncertainties resulting from lack of knowledge [1, 2, 3]. For aleatory uncertainties, sufficient data can generally enable the definition of input probability distributions and the use of probabilistic methods. By cons, for epistemic uncertainties, data is generally too sparse to enable precise probabilistic descriptions and consequently the uncertainties are often modelled by variation intervals deduced from expert judgment. In this case, the intervals are assimilated to uniform distributions, which enable a probabilistic treatment of epistemic uncertainties.

However in order to skip the limitations of the probabilistic framework, different methods, that we will name extra-probabilistic approaches, have been proposed to model epistemic uncertainties such as fuzzy sets [4], Dempster-Shafer theory of evidence [5, 6], possibility theory [7], probability box (p-box) [8] and random set theory [9]. When both aleatory and epistemic uncertainties are present in the analysis of a complex system, it is often desirable to maintain distinction between both types of uncertainties. A common approach to quantifying the effects of mixed aleatory and epistemic uncertainties is the so-called Second-Order Probability (SOP) analysis [10]. The idea of SOP is to treat separately the aleatory variables and the epistemic variables and to perform a two-stage Monte Carlo simulation, typically by sampling the epistemic variables on the outer loop, then by sampling the aleatory variables on the inner loop. If the result of the inner loop is an empirical distribution of the response of interest, the SOP process yields a family of distributions. Then relevant statistics, such as mean or percentiles, may be computed based on this ensemble and the intervals obtained on the statistics can be interpreted as possible ranges for the statistics given the epistemic uncertainties. In the SOP analysis, the aleatory variables are treated generally by probabilistic methods and the epistemic variables may be considered either by probabilistic methods or by extra-probabilistic approaches, such as interval analysis [10, 11] or Dempster-Shafer theory [11] depending on the assumptions taken for these epistemic variables.

In this paper, we focus on Dempster-Shafer theory of evidence for epistemic uncertainty modeling. Dempster Shafer theory (DST), also known as the theory of belief functions or evidence theory, has been firstly introduced in [5] and then developed in [6]. Epistemic uncertainties are described by sets to which are associated degrees of belief. The challenge is to propagate the input uncertainties modeled by DST into the output space through either numerical simulators or metamodels (in case of too expensive simulators). A local optimization problem is necessary to find the upper and lower bounds of the output for a given set of inputs. For kriging metamodels, efficient global optimization (EGO) algorithm [12] has been proposed as one of the most effective method to solve the global optimization problem. Metamodel is optimized by adding sequentially new points suggested by the algorithm to its learning sample and then updated. But the same problem of computational cost rises as the new points have to be calculated by the expensive simulators. Consequently, the main issue that we have to deal with using EGO is how to reduce the number of new points demanded to be calculated while propagating DST input uncertainty into the output space. For this, we propose a novel scheme to propagate uncertainties modeled by Dempster-Shafer theory through expensive black-box

simulators. The main idea is first to identify the most influential input epistemic uncertainties using new sensitivity indexes proposed in this paper and then to propagate in a detailed way this important parameters while taking into account the other parameters in a more global way. This paper is organized as follows: Dempster-Shafer theory is introduced in Section 2; propagation of mapping function with DST is presented in section 3 with the two possibilities of dimension extension: Cartesian product and vacuous extension; in Section 4, new sensitivity indexes are proposed to study the impact of epistemic uncertainties on the output; the new scheme to propagate epistemic uncertainties modelled by DST through a time-consuming computer simulator is presented in Section 5; finally a numerical application is given in Section 6 to illustrate the efficiency of the proposed method.

## 2. DEMPSTER-SHAFFER THEORY

The Dempster-Shafer theory of evidence (DST) [5,6] provides a flexible framework to represent and combine uncertainty information. The basic assessment function in DST is called the mass function which assigns a degree of belief (mass) to each subset of the whole definition domain  $\Omega$  (called a frame of discernment). By definition,  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  is a mutually exclusive set of hypotheses for which exactly one hypothesis  $\omega_i$  is true.  $\Omega$  will contain  $2^n - 1$  non-empty subsets (including  $\Omega$  itself) as well as the empty set.

Then, given a variable  $X$  defined on  $\Omega$ , its mass function is a mapping function:

$$m^\Omega: 2^\Omega \rightarrow [0,1]$$

such that

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \subseteq \Omega} m(A) = 1$$

Where  $2^\Omega$  denotes the set containing all subsets of  $\Omega$ ,  $\emptyset$  and  $\Omega$ . All subsets with positive mass are called focal sets. Mass function  $m(A)$  measures the belief that  $X$  exactly belongs to  $A$  but not to any subsets of  $A$ . The belief affected to  $\Omega$  is called the degree of ignorance. A mass function is called vacuous if  $m(\Omega) = 1$ , which indicates total ignorance on the studied variable. The belief function (*bel*) and plausibility function (*pl*) over  $\Omega$  are defined respectively as:

$$\begin{aligned} bel(A) &= \sum_{B \subseteq A} m(B), \quad \forall A \subseteq \Omega, \\ pl(A) &= \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Omega. \end{aligned}$$

Both functions summarize the information from the mass function and interpret it in different ways: the belief function tells the degree of justified support that  $X$  lies in  $A$ ; the plausibility function gives the degree of maximal potential support of the same situation  $X$  in  $A$ . Moreover, the interval  $[bel(A), pl(A)]$  becomes the lower and upper bounds of belief on  $X$  in  $A$ . For  $X \in \mathbb{R}$ , the two functions can be presented as a function of  $x$  by fixing the form of input sets. The most used ones take the same interval as the definition of cumulative distribution function  $F(x)$ ,  $]-\infty, x]$ ,  $x \in \mathbb{R}$  and are defined as

$$\begin{aligned} bel(x) &= bel(]-\infty, x]) \\ pl(x) &= pl(]-\infty, x]) \end{aligned}$$

so that they can also be seen as the bounds of  $F(x)$ :

$$bel(x) \leq F(x) \leq pl(x).$$

Under the framework of DST (in form of mass functions), it is also possible to combine different information sources describing the same variable according to the source reliability and the relation between sources. Among different rules of combination, Dempster's rule is the most popular one: let  $m_1$  and  $m_2$  be two reliable independent distinct mass functions describing the same variable  $X$  on  $\Omega$ ; the new mass function obtained by combining information from  $m_1$  and  $m_2$  using Dempster's rule of combination is given by:

$$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - k} \quad \forall A, B, C \subseteq \Omega$$

where  $k = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  represents the conflict between the two mass functions.

Other rules such as cautious rule and disjunctive rule exist to satisfy different conditions [13].

### 3. PROPAGATION OF MAPPING FUNCTION WITH DST

#### 3.1 Propagation using Dimension Extension with Cartesian Product

Consider a continuous function  $f(x): R^d \rightarrow R$  taking  $d$  independent input variables  $x = (x_1, \dots, x_d)$ . Each input variable  $x_i$  is modeled by a mass function  $m^{x_i}: 2^{\Omega^{x_i}} \rightarrow [0,1]$  where  $\Omega^{x_i} \subseteq R$  denotes the definition field of  $x_i$  so that the mass function of the input  $m^x: 2^{\Omega^x} \rightarrow [0,1]$  is given by:

$$m^x(s^x) = \prod_{i=1}^d m^{x_i}(s^{x_i}) \quad \forall s^{x_i} \subseteq \Omega^{x_i}, \quad i = 1, \dots, d$$

where  $s^x$  is a Cartesian product of subsets on all input dimensions, i.e.  $s^x = s^{x_1} \times \dots \times s^{x_d}$  and the definition field of  $x$ ,  $\Omega^x = \Omega^{x_1} \times \dots \times \Omega^{x_d}$ . We notice that the focal sets of input mass function are the Cartesian products of focal sets of mass functions on each input dimension.

Then the mass function of the output  $y = f(x)$  can be defined as follows

$$m^y(g(s^x)) = m^x(s^x)$$

where function  $g: 2^{\Omega^x} \rightarrow 2^{\Omega^y}$  maps a set of input values to its correspondent output space which includes all possible output values given by the input set, i.e.

$$\{y = f(x) | \forall x \in s^x\} \subseteq g(s^x).$$

With the continuity of  $f$ , we can easily define the set mapping function  $g$  as follows:

$$g(s^x) = [\min_{x \in s^x} f(x), \max_{x \in s^x} f(x)] \quad \forall s^x \subseteq \Omega^x.$$

#### 3.2 Propagation using Dimension Extension with Vacuous Extension

An alternative method is to propagate the mass function of each variable to the output space with vacuous mass functions on the other dimensions and then merge the partial output mass functions with Dempster's rule of combination (shown in Algorithm 0)

The number of multidimensional input focal sets to be propagated to the output space is the sum of number of focal sets of all unidimensional input variables. Compared to the number obtained by the former method which is equal to the product of the focal set number of all variables, here the number of optimization searches demanded during the propagation is largely reduced. *For example, if there are 4 input variables defined each by 3 focal sets, the number of multidimensional input focal sets to be propagated will be  $3^4 = 81$  with Cartesian product and only  $4 \times 3 = 12$  with vacuous extension.* Vacuous extension avoids the fast calculation cost increase as more input variables are taken into account, which is very important for expensive optimization operations.

**Algorithm 0: initial proposal with vacuous extension**

**Step 1:** construct partial multidimensional input mass function for each input variable  $x_i$  ( $i = 1, \dots, d$ )

$$m_i^x(s_i^x) = \begin{cases} m^{x_i}(s_i^{x_i}) & \text{if } s_i^x = \Omega^{x_1} \times \dots \times s_i^{x_i} \times \dots \times \Omega^{x_d} \\ 0 & \text{otherwise} \end{cases}$$

**Step 2:** construct the partial output mass functions  $m_i^y$  by propagating separately  $m_i^x$  to the output space ( $i = 1, \dots, d$ )

$$m_i^y(g(s_i^x)) = m_i^y\left(\left[\min_{x \in s_i^x} f(x), \max_{x \in s_i^x} f(x)\right]\right) = m_i^x(s_i^x) \quad (1)$$

**Step 3:** merge all partial output mass functions  $m_i^y$  ( $i = 1, \dots, d$ ) into one final output function  $m^y$  using Dempster's rule of combination

$$m^y = m_1^y \otimes \dots \otimes m_d^y.$$

#### 4. SENSITIVITY ANALYSIS WITH EPISTEMIC UNCERTAINTY

In this section, we propose firstly two measures of epistemic uncertainty and then new sensitivity indexes useful within DST using the propagation results introduced in the previous section.

##### 4.1 Existing Sensitivity Indexes

Sensitivity analysis studies the impact of the variation of input variables on the output. In probability theory, variance analysis, where variance is used as a measure of dispersion, is widely applied. The first-order sensitivity index of an input variable  $X_i$ , introduced by I.M. Sobol [14], is defined as follows:

$$S_i = \frac{Var(E(Y|X_i))}{Var(Y)}$$

It measures the effect on output variance of varying  $X_i$  alone, but averaged over variations in other input parameters; therefore it represents the contribution to the output variance of the main effect of  $X_i$ . So its value is large when the contribution of  $X_i$  is important.

Similar to variance method, for imprecise probability, a p-box method is proposed in [15] which quantifies the aleatory and epistemic uncertainties by surface between the upper and lower cumulative distribution functions of the output. The p-box sensitivity index is given by:

$$S_i^p = \frac{A_i}{A} = \frac{\int_{\Omega'} (\bar{F}_i(y) - \underline{F}_i(y)) dy}{\int_{\Omega'} (\bar{F}(y) - \underline{F}(y)) dy}$$

where  $\bar{F}$  and  $\underline{F}$  are the upper and lower cumulative distribution functions estimated over the whole variation domains of all input variables and  $\bar{F}_i$  and  $\underline{F}_i$  are the ones obtained over the whole variation domains of variables  $X_i$  with given default values for the other input variables. It indicates the global importance in terms of uncertainty of  $X_i$  on the output.

## 4.2 Proposed Measures of Epistemic Uncertainty

We propose to measure the epistemic uncertainty based on the surface constituted by the mass function or by the surface bounded by the belief and plausibility functions on  $] - \infty, x]$ ,  $\forall x \in \Omega \subseteq \mathbb{R}$ , i.e.

$$A^{mass} = \int_{a \in \Omega} \int_{b \in \Omega} m([a, b])(b - a) db da$$

$$A^{belpl} = \int_{x \in \Omega} [pl(]-\infty, x]) - bel(]-\infty, x])] dx.$$

The measure  $A^{mass}$  depends on the range and mass value of the focal sets of the mass function  $m$ . The value of  $A^{mass}$  will increase if the range of any focal set becomes larger for an unchanged mass value. The measure  $A^{belpl}$  has the same property as it just summarizes the information provided by the mass function.

$A^{mass}$  and  $A^{belpl}$  measure the epistemic uncertainty in a mass function: smaller (or larger)  $A^{mass}$  and  $A^{belpl}$  values signify more (or less) precise information and less (or more) epistemic uncertainty provided by the mass function; the maximum values ( $A_{\Omega}^{mass} = A_{\Omega}^{belpl} = \max_{x \in \Omega} x - \min_{x \in \Omega} x$ ) come from a vacuous mass function, i.e.  $m(\Omega) = 1$ , which corresponds the notion of total ignorance (maximal degree of epistemic uncertainty); probability cumulative distributions, which can also be seen as mass functions whose focal sets are only singletons ( $\{x\}$ ), produce the minimums of both measures ( $A_F^{mass} = A_F^{belpl} = 0$ ).

## 4.3 New Sensitivity Indexes

The initial idea of our proposed sensitivity indexes is to compare the difference on degree of epistemic uncertainty over the obtained output mass function with and without information provided by the mass function describing a certain input variable. The degree of epistemic uncertainty is measured by  $A^{mass}$  or by  $A^{belpl}$ , defined previously.

In order to avoid fixing a default values for the input variable, as in the existing indexes (§4.1), we propose to characterize the effect of variable  $X_i$  by the partial output function  $m_i^y$  obtained in Step 2 of algorithm 0. Indeed, each focal set of  $m_i^y$  is obtained by searching the optimums of the output over one corresponding focal set for  $X_i$  and the whole field for the other input variables.

According to (1), the range of the output focal sets shows the impact of variation of  $X_i$  on the output in term of degree of epistemic uncertainty.

Therefore we define the ratios  $\frac{A_{y_i}^{mass}}{A_{\Omega^y}^{mass}}$  and  $\frac{A_{y_i}^{belpl}}{A_{\Omega^y}^{belpl}}$  where  $A_{y_i}^{mass}$  or  $A_{y_i}^{belpl}$  denote respectively the surfaces created by  $m_i^y$  and  $bel_i^y/pl_i^y$ , i.e.

$$A_{y_i}^{mass} = \int_{\Omega^y} m_i^y(s) |s| ds = \sum_{s^{x_i} \in \Omega^{x_i}} m(s^{x_i}) \left[ \max_{x \in (\times_{j \neq i} \Omega^{x_j}) \times s^{x_i}} M(x) - \min_{x \in (\times_{j \neq i} \Omega^{x_j}) \times s^{x_i}} M(x) \right]$$

$$A_{y_i}^{belpl} = \int_{x \in \mathbb{R}} pl_i^y([-\infty, x]) - bel_i^y([-\infty, x]) dx$$

and  $A_{\Omega}^{mass} = A_{\Omega}^{belpl}$  is the surface constituted by the mass function of the response obtained by vacuous extension.. They represent the ratios of output epistemic uncertainty given only the mass function of  $X_i$  and with total ignorance on all input variables. They measure the diminution of epistemic uncertainty on the output while the available information on  $X_i$  is added in DST. As the measures become larger, the focal sets of  $m_i^y$  cover larger range and show less impact on limiting the output value variation. On the contrary, a smaller measure indicates that the focal sets of  $m^{x_i}$  plays an important role on reducing the epistemic uncertainty and variation of the output.

Finally the sensitivity indexes that we propose are defined as follows:

$$S_{mass}^{y_i} = 1 - \frac{A_{y_i}^{mass}}{A_{\Omega^y}^{mass}},$$

$$S_{belpl}^{y_i} = 1 - \frac{A_{y_i}^{belpl}}{A_{\Omega^y}^{belpl}}$$

$S_{mass}^{y_i}$  and  $S_{belpl}^{y_i}$  can be seen as sensitivity measures of epistemic input uncertainty on output within the framework of DST.

## 5. MIXED SCHEME PROPOSAL

The ranking of the  $S_{mass}^{y_i}$  and/or the  $S_{belpl}^{y_i}$  over all input variables ( $i = 1, \dots, d$ ), makes possible the selection of the most influential epistemic variables, that is the ones which reduce significantly the range output focal sets and provide more precise information on the output. We know that dimension extension with Cartesian product gives more input focal sets and details but is more costly; and that the vacuous method needs less calculation resources but provides very conservative result and larger epistemic uncertainty estimate. In order to improve the algorithm (Algorithm 0) initially proposed while keeping the advantage of vacuous extension and adding more precise information, an improved algorithm (Algorithm 1) is proposed using both methods: for input variables with evident impact on output mass function, dimension extension using Cartesian product is used; for other less influential variables, vacuous extension is applied. The selection of influential input variables is determined according to the ranking the  $S_{mass}^{y_i}$  and/or the  $S_{belpl}^{y_i}$ .



The cost of this calculation is limited by the number of input variables chosen in **Step 3** and the propagation method in **Step 4**. Since the number of focal sets of the multidimensional mass function obtained in **Step 3** by Cartesian dimension extension is equal to the product of numbers of focal sets of all the selected variables, the number of focal sets to be propagated is controlled by limiting the number of selected variables. It is also possible to use Monte-Carlo method instead of complete propagation over all focal sets in **Step 4** in order to reduce computation cost.

**Algorithm 1:** Improved algorithm (mixed scheme) (see Figure 1))

**Step 1:** calculate the partial mass functions of the output  $m_i^y, (i = 1, \dots, d)$  with vacuous dimension extension, i.e.

$$m_i^y(g(s_i^x)) = \begin{cases} m^{x_i}(s^{x_i}) & \text{if } s_i^x = \Omega^{x_1} \times \dots \times s^{x_i} \times \dots \times \Omega^{x_d} \\ 0 & \text{otherwise} \end{cases}$$

**Step 2:** calculate sensitivity indexes of epistemic uncertainty  $S_{mass}^{y_i}$  and  $S_{belpl}^{y_i}$  and select the input variables (e.g.  $X_j$  and  $X_k$ ) with greatest  $S_{mass}^{y_i}$  and/or  $S_{belpl}^{y_i}$  values

**Step 3:** merge mass functions of the selected input variables by Cartesian dimension extension, for example

**Step 4:** propagate the merged mass function into the output space while the search area of the non-selected parameters is the whole definition field

**Step 5:** combine the mass function obtained in Step 4 with the partial output mass functions obtained in step1 of the non-selected parameters

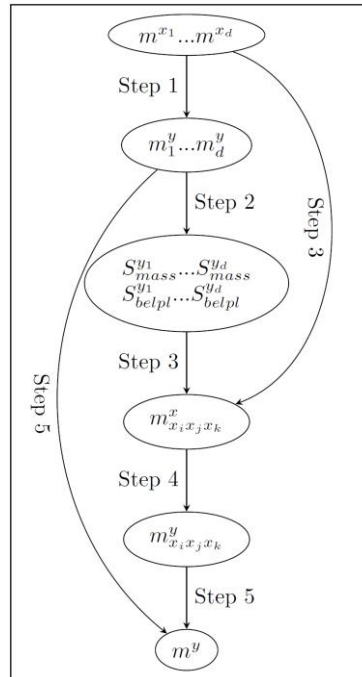


Figure 1 - Illustration of Algorithm 1

## 6. NUMERICAL APPLICATION

The benchmark we used to illustrate and test our proposed methods is the Borehole function [16] with eight independent input variables. This Borehole function models water flow through a borehole. It is widely used for testing a wide variety of methods in computer experiments thanks to its simplicity and quick evaluation. The output is the water flow rate (in m<sup>3</sup>/yr) calculated as follows:

$$f(x) = \frac{2\pi T_u (H_u - H_l)}{\ln(r/r_w) \left(1 + \frac{2LT_u}{\ln\left(\frac{r}{r_w}\right) r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

Instead of modelling the input variables by probability density functions as is usually done in the classical BEPU evaluation, the input variables are here modelled by mass functions (shown in Table 1). For the purpose of this numerical application, the values associated to the mass functions have been arbitrary chosen but in a real application they may have been obtained by expert elicitation. For example, the mass function of the variable  $H_u$  (shown in Figure 2) may correspond to the judgement of one expert giving the input range of the variable and three quantiles (25%, 50% and 75%) or to the combination of several expert judgements. In all cases, this type of representation by mass function enables to handle, in case of epistemic uncertainty, less precise information than probability density functions.

Parameter	Description	Mass function	Parameter	Description	Mass function
$r_w$	Radius of borehole (m)	$m^{r_w}([0.05, 0.075]) = 0.25$ $m^{r_w}([0.075, 0.10]) = 0.25$ $m^{r_w}([0.10, 0.12]) = 0.2$ $m^{r_w}([0.12, 0.15]) = 0.3$	$T_l$	Transmissivity of lower aquifer (m <sup>2</sup> /yr)	$m^{T_l}([63.10, 80]) = 0.25$ $m^{T_l}([80, 86]) = 0.25$ $m^{T_l}([86, 116]) = 0.2$ $m^{T_l}([100, 116]) = 0.3$
$r$	Radius of influence (m)	$m^r([100, 2000]) = 0.3$ $m^r([2000, 2230]) = 0.2$ $m^r([2230, 5000]) = 0.45$ $m^r([2500, 50000]) = 0.05$	$H_l$	Potentiometric head of lower aquifer (m)	$m^{H_l}([700, 760]) = 0.3$ $m^{H_l}([730, 750]) = 0.2$ $m^{H_l}([750, 780]) = 0.2$ $m^{H_l}([760, 820]) = 0.3$
$T_u$	Transmissivity of upper aquifer (m <sup>2</sup> /yr)	$m^{T_u}([63070, 72000]) = 0.25$ $m^{T_u}([72000, 89335]) = 0.25$ $m^{T_u}([89335, 94000]) = 0.25$ $m^{T_u}([94000, 115600]) = 0.25$	$L$	Length of borehole (m)	$m^L([1120, 1200]) = 0.1$ $m^L([1200, 1352]) = 0.4$ $m^L([1352, 1405]) = 0.3$ $m^L([1405, 1680]) = 0.2$
$H_u$	Potentiometric head of upper aquifer (m)	$m^{H_u}([990, 1010]) = 0.25$ $m^{H_u}([1010, 1050]) = 0.25$ $m^{H_u}([1050, 1080]) = 0.25$ $m^{H_u}([1080, 1110]) = 0.25$	$K_w$	Hydraulic conductivity of borehole (m/yr)	$m^{K_w}([9855, 10050]) = 0.25$ $m^{K_w}([10050, 10950]) = 0.25$ $m^{K_w}([10950, 11050]) = 0.25$ $m^{K_w}([11050, 12045]) = 0.25$

Table 1 - Borehole function - Description of the input variables and their mass functions

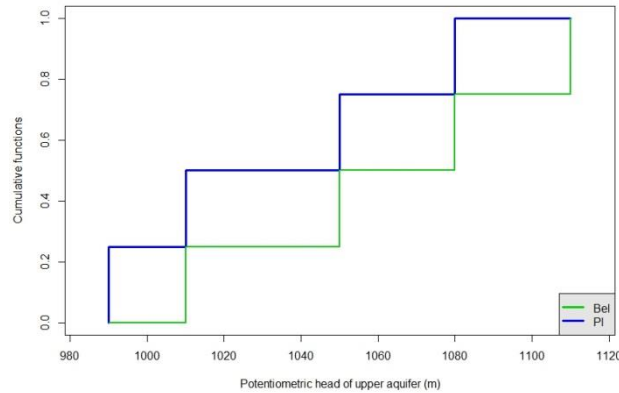


Figure 2 – Mass function of variable  $H_u$  (Bel = belief function and Pl = plausibility function)

In this example, optimization search for the corresponding output focal sets are conducted by a genetic method provided by R package *rgenoud* [17].

After executing **Step 2**, two or three input variables are selected according to their sensitivity measures (Table 2). When three variables are selected, the choices based on one or the other of the two indexes are the same ( $r_w$ ,  $H_u$ ,  $L$ ). When two variables are selected, the choices are slightly different: ( $r_w$ ,  $L$ ) according to  $S_{mass}^{yi}$  and ( $r_w$ ,  $H_u$ ) according to  $S_{belpl}^{yi}$ .

Input variable	$A_{yi}^{mass}$	$A_{yi}^{belpl}$	$S_{mass}^{yi}$	$S_{belpl}^{yi}$
$r_w$	160.5616	212.9625	<b>0.468</b>	<b>0.2942</b>
$r$	301.0632	301.2724	0.002	0.002
$T_u$	301.7553	301.7556	<< 0.001	<< 0.001
$H_u$	263.9355	282.8121	<b>0.125</b>	<b>0.063</b>
$T_l$	301.3379	301.5718	0.001	< 0.001
$H_l$	274.2439	284.0597	0.091	0.059
$L$	263.6856	283.5106	<b>0.126</b>	<b>0.060</b>
$K_w$	277.4571	288.1576	0.081	0.045

Table 2 - Borehole function – Epistemic sensitivity measures

On **Step 4**, the input mass functions  $m_{r_w H_u}^x$  (or  $m_{r_w L}^x$ ) and  $m_{r_w H_u L}^x$  containing the information on the selected input variables extended using Cartesian product method with ignorance on the other variables are propagated to the output space. Then we obtain the corresponding output mass functions  $m_{r_w H_u}^y$  (or  $m_{r_w L}^y$ ) and  $m_{r_w H_u L}^y$ . Compared with the output mass functions  $m_{r_w}^y \otimes m_{H_u}^y$  and  $m_{r_w}^y \otimes m_{H_u}^y \otimes m_L^y$ , intermediate results obtained on Step 3 in Algorithm 0 and containing exactly the same information, these new intermediate results show smaller value on  $A_y^{mass}$  and  $A_y^{belpl}$  and indicate less epistemic uncertainty and a more precise description of the output (see Table 3).

Method	Selected input variables	$A_y^{mass}$	$A_y^{belpl}$
Algorithm 0	$(r_w, H_u)$	149.3	192.3
	$(r_w, L)$	147.4	195.1
	$(r_w, H_u, L)$	144.7	185.5
Algorithm 1	$(r_w, H_u)$	132.5	147.4
	$(r_w, L)$	133.5	147.6
	$(r_w, H_u, L)$	106.6	110.8

Table 3 - Sensitivity measures of intermediate results

Since the next operation (i.e. merging these intermediate results with the other partial output mass functions using Dempster's rule of combination) is the same for both intermediate results it is normal to observe similar results in the final results (shown in Figure 3 and Table 4): the final output mass function obtained by Algorithm 1 is less conservative than the one obtained by Algorithm 0 using vacuous extension. The rapid decrease on epistemic uncertainty measures as more input variables enter into **Step 2** corresponds well to our initial objectives.

Table 4 gives also the number of output focal sets and so an idea of the calculation costs in each case. The mixed scheme proposed in Algorithm 1 can be seen as an intermediate solution between Algorithm 0 with complete vacuous extension and the method with total Cartesian extension before propagation. Starting with Algorithm 0, as more variables are selected, this improved method tends to the latter one, presenting less epistemic uncertainty and more details but becoming more costly.

As Cartesian product method using complete propagation (or Monte-Carlo simulations) through expensive simulators is generally impossible, this improved method enables a trade-off, within a BEPU evaluation, between the calculation costs and how conservative is the final output mass function.

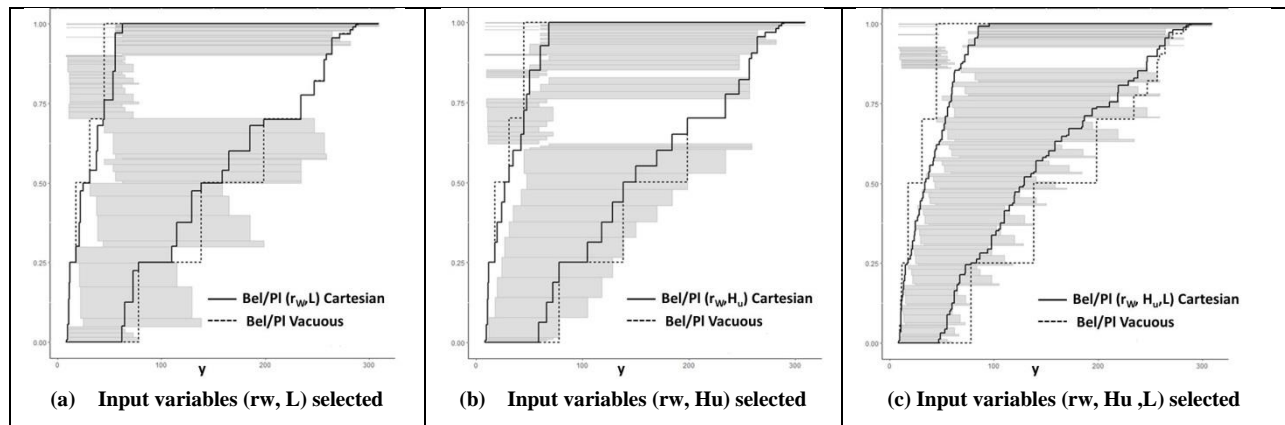


Figure 3 - Borehole function - 1) mass functions (rectangles) and Bel/Pl functions (solid lines) of the output obtained by Algorithm 1 with 2 or 3 input variables selected 2) Bel/Pl functions of the output obtained by Algorithm 0 with vacuous dimension extension (dashed lines)

Method	Selected input variables	$A_y^{mass}$	$A_y^{belpl}$	Number of output focal sets
Algorithm 0	-	143.2	182.6	32
Algorithm 1	(rw, Hu)	126.4	137.7	32+16
Algorithm 1	(rw, L)	127.6	137.5	32+16
Algorithm 1	(rw, Hu ,L)	105.5	108.6	32+64

Table 4 - Borehole function - sensible measures of the finally obtained output mass functions using different proposed methods (Algorithm 0 and Algorithm 1)

## 7. NUCLEAR POWER PLANT APPLICATION

Our method has been applied within the framework of a BEPU analysis of a “Loss Of Coolant Accident” (LOCA) in a pressurized water reactor. The thermal-hydraulic analysis is performed by the CATHARE2 code [18]. In our application, 27 independent CATHARE2 input parameters related to the modeling of physical phenomena (e.g. friction or condensation coefficients) are considered as epistemic uncertainties. The response of interest calculated by CATHARE2 is the maximal peak cladding temperature (PCT), obtained during the LOCA transient. The objective of our analysis is to quantify the epistemic uncertainty of the PCT and to identify among the 27 input parameters the most important contributors to this uncertainty.

The mass functions of the input parameters have been built from the information, obtained by expert judgement, on their ranges of variation and on three quantiles (25%, 50% and 75%). Thus each input parameter is defined by 4 focal sets.

Due to the large cpu-time cost associated to each evaluation by the CATHARE2 code, an improved optimization for the search of the output (i.e. PCT) focal sets has been used. The optimisation scheme used is based on the efficient global optimization (EGO) algorithm [12] where the computer code is surrogated by a kriging metamodel. In order to build this metamodel, CATHARE2 simulations are performed following a design of experiments based on a space-filling LHS. In our approach, the EGO algorithm has been adapted to locally (i.e. at the level of a focal set) improve the accuracy of the metamodel for optimization purpose.

As for the Borehole function (see section 6), we have compared, in Figure 4, the results obtained by vacuous extension and Dempster’s rule (algorithm 0) and by the mixed scheme (algorithm 1). In the mixed scheme, the two or three most important input parameters have been identified thanks to the calculation of the epistemic sensitivity indexes (see section 4.3). For algorithm 0, the number of focal sets to be propagated is equal to 108 (i.e.  $27 \times 4$ ); for algorithm 1, it is 124 (i.e.  $27 \times 4 + 4^2$ ) when two parameters are selected or 172 (i.e.  $27 \times 4 + 4^3$ ) when three parameters are selected. In the last case, the corresponding number of CATHARE2 simulations necessary to evaluate the output mass function, is 698.

This application shows the possibility of applying the DST in the case of a large number of input epistemic variables (27 in our case). Algorithm 0 gives a complete but conservative global representation of the simulator response (i.e. the PCT). In the case of using the mixed scheme, the domain delimited by the belief and plausibility functions is substantially reduced, allowing a less conservative representation of the PCT epistemic uncertainty. The interest of this mixed

scheme for propagating epistemic uncertainties with DST is also to adapt the number of selected input variables (in Step 2) to the calculation budget.

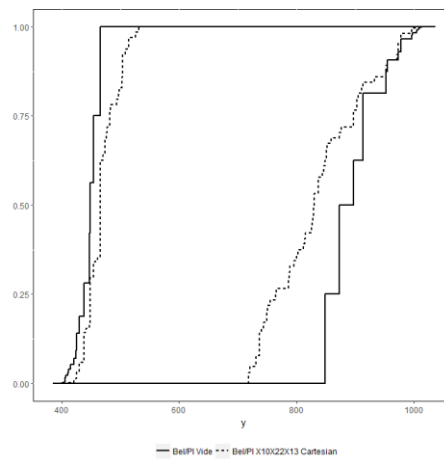


Figure 4 – LOCA application – Belief and plausibility functions of the response (PCT) obtained by the mixed scheme with three selected parameters (dashed lines) and by the algorithm 0 (solid lines)

## 8. CONCLUSION

In this paper, we have first proposed new sensitivity indexes focusing on epistemic uncertainty useful within the framework of DST. Then we have proposed a new scheme to propagate epistemic uncertainties modelled by DST through a time-consuming computer simulator. Starting from a selection of the most influential input variables based on these new sensitivity indexes, the method consists in using dimension extension by Cartesian product for input variables with evident impact on output mass function and to apply vacuous extension for other less influential variables. Finally, mass combination rules (e.g. Dempster's combination rule for independent input variables) are used after mapping the input focal sets into the output space.

Besides classical Monte Carlo method using Cartesian product input mass construction method, this method significantly reduces the number of input focal sets. The obtained output function has less focal sets, but larger range. Thus, a complete coverage of the output focal set space can be provided. Epistemic uncertainty is estimated globally in a conservative point of view. As a result of a trade-off between calculation cost, estimation accuracy and the quantity of details on epistemic uncertainty presented, our proposal enables propagation of epistemic uncertainties modeled by DST with a limited computation budget making possible its practical use within a BEPU evaluation.

## 9. REFERENCES

- [1] A. Der Kiuregian, "Aleatory or epistemic? Does it matter ?" *Special Workshop on Risk Acceptance and Risk Communication*, Stanford University, March 26-27, 2007.
- [2] J.C Helton, "Quantification of margins and uncertainties: conceptual and computational basis". *Reliability Engineering and System Safety*. 96, 976-1013, 2011.

- [3] L.P. Swiler et al, "Epistemic uncertainty in calculation of margins". *50th AI-AA/ASME, Structures, Structural Dynamics, and Materials Conference*, Palm Springs, USA, 4-7 May 2009.
- [4] B. Möller, "Fuzzy randomness ; a contribution to imprecise probability", *ZAMM-Journal of Applied Mathematics and Mechanics* 84 (10-11) (2004) 754-764.
- [5] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *The annals of mathematical statistics* (1967) 325-339.
- [6] G. Shafer, "A mathematical theory of evidence", *Vol. I, Princeton University press*, 1976.
- [7] D. Dubois, H. Prade, "Possibility theory: an approach to computerized processing of uncertainty", *Springer Science & Business Media*, 2012.
- [8] S. Ferson, L. R. Ginzburg, "Different methods are needed to propagate ignorance and variability", *Reliability Engineering & System Safety* 54 (2-3) (1996) 133{144.
- [9] G. Matheron, "Random sets and integral geometry", *Wiley series in probability and mathematical statistics*, Wiley, 1975.
- [10] M.S. Eldred et al, "Mixed aleatory-epistemic uncertainty quantification with stochastic expansions and optimization-based interval estimation", *Reliability Engineering and System Safety* 96, 1092–1113, 2011.
- [11] M. Marquès, "Propagation of aleatory and epistemic uncertainties in quantification of systems failure probabilities or safety margins", *Safety and Reliability: Methodology and Applications* – Nowakowski et al. (Eds), 2015.
- [12] D. R. Jones, M. Schonlau, W. J. Welch, "Efficient global optimization of expensive black-box functions", *Journal of Global optimization* 13 (4) (1998) 455-492.
- [13] T. Doneux, "Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence", *Artificial Intelligence* 172 (2008) 234–264.
- [14] I. M. Sobol. "Sensitivity estimates for nonlinear mathematical models." *Mathematical Modelling and Computational Experiments* 1.4 (1993): 407-414.
- [15] J. Guo, X. Du, "Sensitivity analysis with mixture of epistemic and aleatory uncertainties", *AIAA Journal* 45:9 (2007), 2337-2349.
- [16] M. D. Morris, T. J. Mitchell, D. Ylvisaker, "Bayesian design and analysis of computer experiments: use of derivatives in surface prediction", *Technometrics* 35 (3) (1993) 243-255.
- [17] W. R. Mebane, J. S. Sekhon, "Genetic optimization using derivatives: The rgenoud Package for R", *Journal of Statistical Software*, (2011).
- [18] P. Mazgaj, J.L. Vacher, S. Carneval, "Comparison of CATHARE results with the experimental results of cold leg intermediate break LOCA obtained during ROSA-2/LSTF test 7", *EPJ Nuclear Sciences and Technologies*, 2 (1), 2016.