

Antonios Kamariotis
 Giulia Antinori
 Iason Papaioannou
 Fabian Duddeck

Mixed aleatory-epistemic uncertainty quantification and sensitivity analysis

Abstract: This work investigates the modelling of epistemic input parameter uncertainties, and the numerical techniques for uncertainty quantification and uncertainty-based sensitivity analysis in the presence of both aleatory and epistemic uncertainties. Two different approaches are used, the interval valued probability (IVP) method and the Bayesian probabilistic (BP) method. In both cases, a double loop method is used to computationally separate the two different uncertainty types and propagate them within the model. These two approaches are successfully applied on a high-dimensional jet engine secondary air system model from aerospace engineering. The different outputs obtained by the two approaches are interpreted and compared. For the global sensitivity analysis of the epistemic variables, an empirical “pinching” strategy is applied when using the IVP method. With the BP method, variance-based global sensitivity analysis of the epistemic variables is performed. Novel expressions for the Sobol indices of a statistic of a response, conditional on the epistemic variables, are presented and interpreted.

1 Introduction

Uncertainty quantification (UQ) and uncertainty-based sensitivity analysis (SA) is an area of continuously increasing importance in computational engineering. Uncertainties are often classified into two categories, aleatory and epistemic, e.g. [3]. Aleatory uncertainty is due to randomness related to a physical process. This type of uncertainty is considered irreducible, provided that an adequate statistical description is available. Probability theory is used for modeling this type of uncertainties, and there are well established sampling or stochastic expansion methods for the UQ and SA of response quantities involving this type of uncertainties [7] [13]. Epistemic uncertainty is due to incomplete knowledge and/or lack of useful data, and it is considered reducible, e.g. by gathering additional data or by refining the models. The use of probability theory for modeling epistemic uncertainty is put into question in literature and so-called imprecise probability methods [2] have emerged as alternative mathematical tools for modeling this type of uncertainty. This opens a debate between different research groups as to which is the best mathematical tool to model epistemic uncertainty [9]. It is more often than not that both types of uncertainties coexist in the engineering computational model universe. Therefore, appropriate tools to model mixed aleatory-epistemic uncertainties, and computational methods to propagate these uncertainties within the model need to be investigated. When both types of uncertainty are present, it is of great importance to have efficient tools and methods at hand, which can separate, both conceptually and computationally, the influence of the two uncertainty types, and which can propagate them within

the same model. There are few attempts to date for UQ and SA in a mixed uncertainty setting in real high-dimensional engineering models [10]. Such an attempt is presented in this work. The main focus is on the model input parameter uncertainty. The type of epistemic uncertainty which is of interest herein is the case where the quantity is described through intervals, e.g. obtained from expert judgments or technical drawings, and no measurement data is available.

The structure of the paper is as follows: In Section 2 we discuss two different mixed aleatory-epistemic UQ approaches used in this work. Section 3 discusses the SA approaches used in the mixed uncertainty setting, and novel Sobol indices of a statistic of a response are presented, based on a generalization of concepts presented in [4]. Section 4 presents the secondary air system model of a jet engine, followed by the application of the presented UQ and SA methods on this model. Finally, Section 5 summarizes the results and concludes this work.

2 Mixed aleatory-epistemic UQ approaches

Consider a model in the form:

$$Y = g(\mathbf{x}) \quad (1)$$

where $\mathbf{x} \in \mathcal{R}^n$ denotes possible outcomes of a vector \mathbf{X} that collects all uncertain input parameters of the model. In the mixed uncertainty case, the vector \mathbf{X} is decomposed as $\mathbf{X} = [\mathbf{X}_a, \mathbf{X}_e]$, with \mathbf{X}_a containing the aleatory variables, and \mathbf{X}_e containing the epistemic variables. The following two subsections present in detail two mixed aleatory-epistemic UQ approaches, which nest one forward uncertainty propagation within another. In both approaches, the outer loop is associated with the treatment of the epistemic input variables \mathbf{X}_e , while the aleatory input variables \mathbf{X}_a lie in the inner loop (Fig. 1). The difference lies in the way how the epistemic variables are modeled in the outer loop. These two approaches are also available in the open source Dakota software [12].

2.1 Interval valued probability (IVP) approach

In the IVP method [5] [12], the model input variables which are identified as epistemic, \mathbf{X}_e , for which there is only interval information and no data is available, are in the outer loop, and are simply treated as single intervals. Therefore, there is no distribution information for the epistemic variables; any sample within the provided interval bounds is a possible realization. The aleatory variables are in the inner loop, and

probability distributions are assigned for them. There are different combinations of methods that can be used in either loop in order to propagate the uncertainty, however here only the nested sampling approach will be mentioned, which involves using sampling techniques in both loops, as shown in Fig.1. Following this approach, first samples from the outer loop epistemic variables are taken through generating uniform random samples within the specified intervals. Subsequently, these epistemic variable values are kept as fixed values in the inner loop, where the aleatory input variables sampling takes place, based on their defined probability distributions. At this point, all the model input variables have obtained a value, therefore the real model evaluation can take place. Multiple model evaluations take place in the inner loop, as many as the number of inner loop samples, conditional on a specific realization of the outer loop epistemic variables vector, which is kept constant in the inner loop. Therefore, each time in the inner loop a statistical quantity of interest (QoI) (e.g. mean value, variance) of the model response can be computed, conditional on fixed values of the epistemic variables vector. In the end of this nested sampling approach, one obtains multiple values of the statistical QoI (as many as the number of the outer loop samples). The final goal of the IVP method is to compute output intervals bounding the statistical QoI of the response, which quantify the effect of epistemic uncertainty within the model. This is simply done by keeping the minimum and maximum computed values of the statistical QoI. Additionally, at each epistemic sample one obtains a sample estimate of the cumulative distribution function (CDF) or one complementary CDF (CCDF). Collecting all CCDF estimates in the same plot gives the so-called horsetail plot [8], which can be used to observe the effect of epistemic uncertainty and to infer intervals, as is presented in Section 4.

2.2 Bayesian probability (BP) approach

The BP method (also referred to as second order probability method in literature [5] [12]) uses the exact same idea with the IVP method, in the sense that both methods separate the variables in outer epistemic and inner aleatory loops. The difference lies in the way how the outer loop epistemic variables are treated. In the BP method, they are no longer treated simply as intervals, and subjective (Bayesian) probability assignments are made for them by the modeler based on the available information. We note here that assigning a fixed interval on the epistemic variables as is done in the IVP method based solely on an expert judgment is also a subjective modeling choice. The BP approach enables expressing a subjective degree of belief on the expert interval through the assignment of a probability distribution. Exactly the same nested sampling procedure is followed like the one described in Section 2.1. However, in the BP case the final result will be the full distribution and epistemic statistics (mean, variance, CDF) of the statistical QoI of the response, and not in the form of intervals bounding the statistical QoI of the response as in the IVP method. Therefore, it is clear that the mathematical structure of the output is different between the two different methods. Within this work, the authors would

like to demonstrate that, also when using the BP method, it is possible to obtain intervals bounding the statistical QoI of the response. Having obtained multiple samples of the statistical QoI of the response in the outer loop, one can use them to compute Bayesian credible intervals (BCI) which bound the result with a certain probability. More specifically the highest posterior density Bayesian credible interval is computed herein.

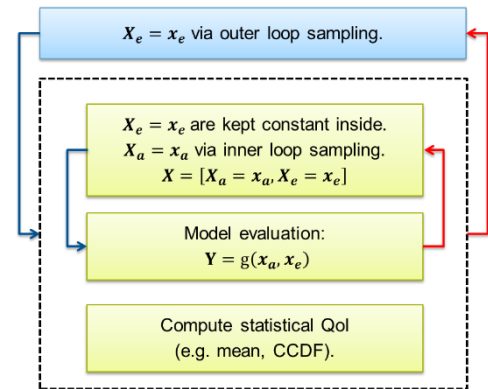


Fig 1: Nested sampling approach used in the IVP and BP methods.

3 Mixed aleatory-epistemic SA approaches

Global SA investigates how much the variability of the output depends on each of the input parameters, taken individually or in combinations with each other. The goal of an SA in the mixed aleatory-epistemic uncertainty setting is to obtain an importance ranking only of the epistemic variables, having integrated out the effect of aleatory uncertainty. This is helpful if one considers that epistemic uncertainty can be potentially reduced, e.g. by performing measurements to obtain more data. The epistemic variable importance ranking could then drive the decision making with regards to collecting additional information to reduce uncertainty.

3.1 SA using the IVP approach

To the knowledge of the authors, global SA involving non-probabilistic uncertain models is based mostly on empirical methods. A so-called pinching strategy is used [6]. This strategy assesses how much the output uncertainty would reduce, if knowledge on the value of a variable would become available. This is done by comparing the uncertainty in the output, before and after pinching an uncertain input variable, i.e. replacing it with a fixed deterministic value. The pinching approach is helpful to assess qualitatively the sensitivity of epistemic variables; however, the obtained sensitivities lack a clear interpretation. This method is able to assess individual contributions, and interactions are not taken into account. If one is interested in interactions, then the pinching strategy should be applied to more than one input variables at a time. Within this work, as presented in Section 4, an empirical estimate of the sensitivity of each individual epistemic variable is given by comparing the horsetail plots, before and after removing the uncertainty from one epistemic variable at a time (see Figures 3 and 4).

3.2 SA using the BP approach

When using the BP approach, it is possible to compute sensitivities through decomposing the variance of the statistical QoI, leading to the so-called Sobol sensitivity indices. Classical variance-based SA looks at the contribution of each input random variable, or combinations thereof, to the variance of the model output. The first order Sobol index of Y with respect to the i -th component X_i is given by:

$$S_i = \frac{V_i}{\text{Var}[g(\mathbf{X})]} = \frac{\text{Var}_{X_i}\{E_{X_{-i}}[g(\mathbf{X})|X_i]\}}{\text{Var}[g(\mathbf{X})]} \quad (2)$$

where $E_{X_{-i}}[g(\mathbf{X})|X_i]$ is the expected value of the output with respect to all input variables except X_i which is fixed. Eq. 2 considers only the influence of the contribution of variable X_i , neglecting interactions with other variables. The effect of interactions is described by higher order Sobol indices. The total effect index considers the total influence of the contribution of variable X_i , accounting also for interactions of X_i with other variables. These expressions are based on the Sobol decomposition of the variance of the output given as follows:

$$\text{Var}[g(\mathbf{X})] = \sum_{i=1}^n V_i + \sum_{i=1}^n \sum_{j \neq i}^n V_{ij} + V_{1,2,..,n} \quad (3)$$

where V_{ij} is the combined effect of X_i and X_j on the variance. Monte Carlo sampling or polynomial chaos expansions can be used to compute these Sobol indices. In the mixed aleatory-epistemic uncertainty case, the input variable vector is $\mathbf{X} = [\mathbf{X}_a, \mathbf{X}_e]$. Assume that the statistical QoI to be computed in the inner loop is the mean value of the scalar model response, i.e. the expectation $E_{X_a}[Y|\mathbf{X}_e]$. In this case, the Sobol decomposition is performed on the variance of the mean value of the output, conditional on the epistemic vector realization. The resulting expression for the first order Sobol index of the mean value is:

$$S_{\mu,i} = \frac{\text{Var}_{X_{e,i}}\{E_{X_{e,-i}}[E_{X_a}[Y|\mathbf{X}_e]|X_{e,i}]\}}{\text{Var}_{X_e}[E_{X_a}[Y|\mathbf{X}_e]]} \quad (4)$$

Such an expression can be obtained for other statistical QoIs, such as the variance of the output, $\text{Var}_{X_a}[Y|\mathbf{X}_e]$:

$$S_{var,i} = \frac{\text{Var}_{X_{e,-i}}\{E_{X_{e,i}}[\text{Var}_{X_a}[Y|\mathbf{X}_e]|X_{e,i}]\}}{\text{Var}_{X_e}[\text{Var}_{X_a}[Y|\mathbf{X}_e]]} \quad (5)$$

As noted in [4], in the mixed uncertainty case, the variance-based decomposition of the total variance contributed by \mathbf{X}_e rather than \mathbf{X} is obtained, as can be seen in the denominator of Eqs. (4) and (5). The variance due to the aleatory variable set \mathbf{X}_a is involved in the inner loop computation of the statistical QoI of the response, conditional on the epistemic vector realization, therefore it acts as a weight of the contribution of each realization of \mathbf{X}_e . Expressions for the total effect Sobol indices of a statistical QoI can be obtained accordingly.

4 Numerical investigations

4.1 The jet engine secondary air system

The two mixed aleatory-epistemic UQ approaches presented herein are applied to the analysis of the secondary air system (SAS) of a three-stage low pressure turbine (LPT) of a jet engine. The jet engine's secondary air system (SAS) is the ensemble of flows, which do not directly contribute to the engine thrust and is used for different purposes like internal unit cooling, prevention of hot gas ingestion into the turbine rotor cavities, sealing of the bearing chamber and control of the axial bearing load [11]. A complete description of the SAS functionality is given in [11]. The SAS is modeled with a 1-D flow network as a succession of chambers linked by flow passages (seals, orifices, pipes, etc.) [1]. Each flow passage has its own pressure loss characteristics. Figure 2 shows an example of an SAS 1-D flow network. The model takes the following input: boundary conditions in terms of pressure and temperature (in yellow in Figure 2) which are dependent on several engine performance parameters, and the geometrical definition of each pressure loss device. After the solution of the network model equations the output is the pressure and temperature level at each chamber (green boxes in Figure 2) and the mass flow rate in each flow line (blue lines in Figure 2).

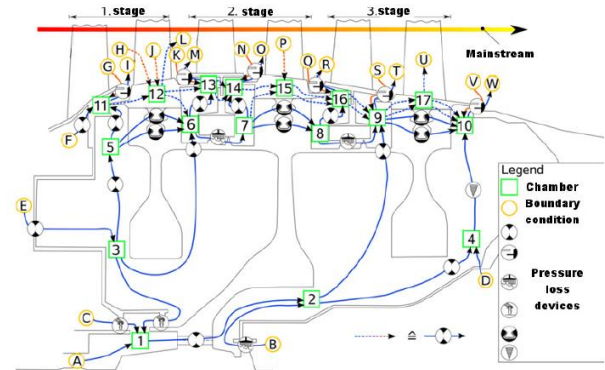


Fig 2: SAS network of an LPT arrangement. Figure taken from [1]

4.2 Uncertainty modeling in the SAS model

There are two types of inputs in the SAS model. The boundary conditions are dependent deterministically on the engine performance parameters, which are pressures and temperatures at critical engine locations. The uncertainty in the performance parameters comes mainly from the uncertainty in the ambient conditions, as well as from engine-to-engine variations. This kind of uncertainty is considered irreducible; these are therefore modeled as aleatory uncertain variables. The information available regarding these variables comes from industrial experts, based on legacy engines, in the form of a mean value and a standard deviation. These variables are modeled with the normal distribution. Uncertainty in the geometrical variables of the different flow elements comes from the uncertainty related to manufacturing tolerances. No measurement data of the actual geometries of any of the flow

elements is available. However, in most cases one could perform measurements on the manufactured parts in order to obtain data for the geometric variables. Thus the uncertainty in the geometric variables is modelled as epistemic, since there is a chance of controlling and reducing this uncertainty. Without any available measurement data, the information available coming from the technical drawings is in the form of a single interval per variable, within which the true value could lie. For the IVP approach these intervals are taken as fixed, whereas for the BP approach the epistemic variables are modeled as uniform random variables on the given intervals. In total, within the used SAS model of Figure 2, there are 63 uncertain input variables, of which 51 are modeled as epistemic uncertain, and 12 as aleatory. All uncertain input variables are considered independent within the current investigation. There is a large number of responses that the SAS produces as an output, however in this study, only the three mass flows B-2 (mass flow 1), E-3 (mass flow 2) and F-11 (mass flow 5), which constitute the ingoing mass flows, (see Figure 2) are of interest. The sum of these three mass flows is the response of interest.

4.3 Mixed aleatory-epistemic UQ and SA of the SAS

4.3.1 IVP approach

As mentioned in Section 2.1, the final goal of the IVP method when performing mixed uncertainty UQ is to compute output intervals bounding the statistical QoI of the response, which quantify the effect of epistemic uncertainty within the model. Therefore, following the nested sampling approach, an interval for the mean model response of interest can be obtained (Table 1).

Table 1: Interval on the mean value of the sum of mass flows 1,2,5. W25 is the reference mass flow in the gas path.

Outer loop LHS	Inner loop LHS	Interval (%W25)	Time (hrs.)
50 samples	50 samples	[1.1943, 1.6121]	1.60
100 samples	100 samples	[1.1935, 1.6122]	4.88
200 samples	200 samples	[1.1927, 1.6172]	9.69

The horsetail plot is plotted in Figure 3, which is an ensemble of CCDFs, and can be used to infer intervals. Each CCDF corresponds to aleatory uncertainty, conditional on one realization of the epistemic variables vector. By observing how widespread the CCDF ensemble is, one can assess the influence of the epistemic input parameter uncertainty on the model's response metric of interest. Looking at the results of Table 1 and Figure 3, one can observe that the resulting interval is significant. The epistemic uncertainty comes from the manufacturing tolerances of the geometric variables. If the geometric variables would take their nominal values, we would expect a resulting CCDF close to the middle of the

interval shown in Figure 3. Furthermore, the influence of the aleatory uncertainty, coming from the performance parameters, can be observed in the form of each resulting CCDF.

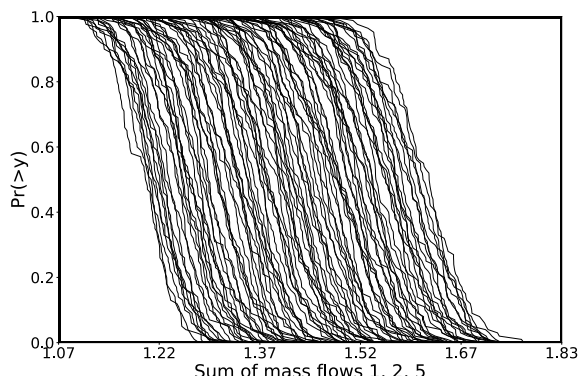


Fig 3: Horsetail plot for the sum of the three mass flows. Multiple aleatory CCDFs of this sum resulting from epistemic uncertainty in the geometric variables. 100 outer loop epistemic vector realizations, and 100 samples in the inner loop to compute the mean value of the response, conditional on each outer loop realization.

Figure 4 illustrates the empirical pinching strategy for sensitivity analysis with the IVP approach. Based on engineering expertise, mass flow 1 is clearly dependent on the value of the effective area of the orifice connecting boundary condition E to chamber 3 in figure 2. Also mass flow 1 obtains much larger values in the sum compared to mass flows 2 and 5. Therefore, we would expect a large reduction in the epistemic uncertainty of the mass flow sum output, if this geometric variable would be fixed to its nominal deterministic value. Hence, we choose to fix the effective area of this orifice to its midpoint value and perform the uncertainty analysis again. The resulting horsetail plot is seen in Figure 4. Comparing it to Figure 3, it can be observed that the resulting output epistemic uncertainty has been reduced significantly.

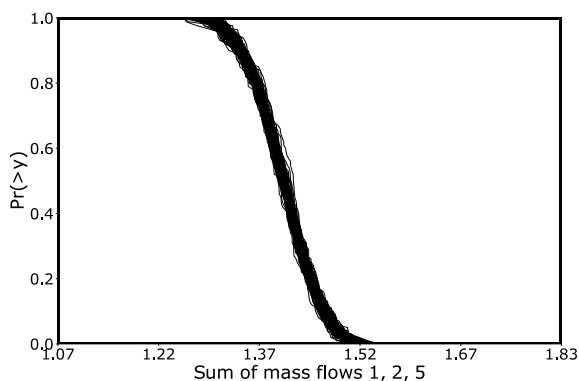


Fig 4: Pinching approach for sensitivity analysis: Horsetail plot when fixing the value of effective area of orifice E-3 to its midpoint value.

4.3.2 BP approach

As discussed in Section 2.2, the final outcome of the BP method for mixed aleatory-epistemic UQ is the full distribution and epistemic statistics of the statistical QoI of the response. For our example, Figure 5 shows the distribution of the mean value of the sum of the responses.

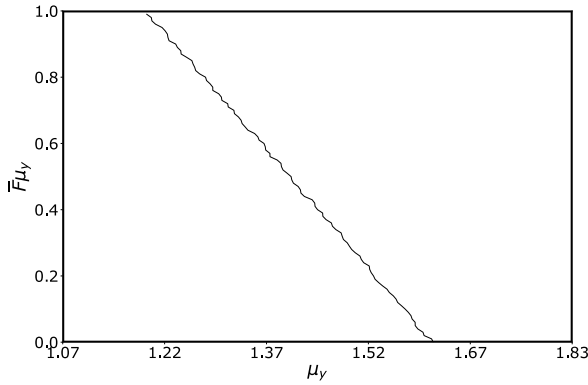


Fig 5: Estimated CCDF of response (sum of three mass flows) mean over aleatory uncertainty. ($\mu_{epist} = 1.4066, \sigma_{epist} = 0.1235$).

As can be seen, the mixed uncertainty output is different between the IVP and the BP method. However, as shown in Table 2, the final samples in the BP outer loop can be used to compute epistemic Bayesian credible intervals for the statistical QoI of the response.

Table 2: Comparing IVP intervals and BP Bayesian credible intervals

Response statistical QoI	Interval	Result (%W25)
Mean mass flow sum	95% BCI	[1.1913, 1.5964]
Mean mass flow sum	99.9% BCI	[1.1913, 1.6194]
Mean mass flow sum	IVP interval	[1.1935, 1.6122]

It can be seen that, a 95% BCI already gives an interval which is very similar to the IVP resulting interval, only a bit underestimated, while taking the 99.9% BCI gives a resulting interval resembling very closely the corresponding IVP interval. This is due to the fact that epistemic variables in the BP approach are modeled as uniform random variables on the given intervals. Therefore, the higher the confidence level in the BCIs, the closer the intervals will be to the ones obtained with the IVP approach.

In Section 3.2, Sobol indices of the statistical QoIs were introduced. The first order Sobol indices of the mean values of the responses 1 and 2 are shown in Figure 6. This first order sensitivity index indicates how much each epistemic variable, when taken into account independently, affects the mean value of the model response of interest. Accordingly, the first order Sobol indices of another statistical QoI (e.g. variance) could be obtained, as shown in Figure 7. In contrast to the main first order Sobol indices, the total order Sobol indices give the contribution of each epistemic variable when taking into account also the interactions with the other epistemic variables. The aleatory variables are integrated out when computing the statistic of interest in the inner loop, and thus no interactions between aleatory and epistemic variables are present. For our SAS application, no difference between main and total order indices of the mean value was observed.

This indicates that there are no interactions between the epistemic variables, i.e. the geometric variables of the different flow elements. This result is expected, since from an engineering point of view we do not expect interactions between the geometric variables of the different flow elements.

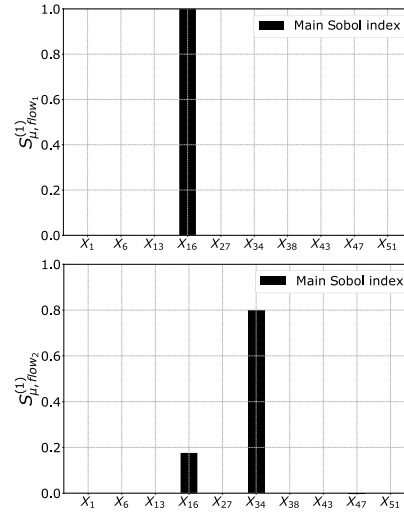


Fig 6: First order Sobol indices of the mean of mass flows 1 and 2. On the x-axis is a selection of the epistemic uncertain variables.

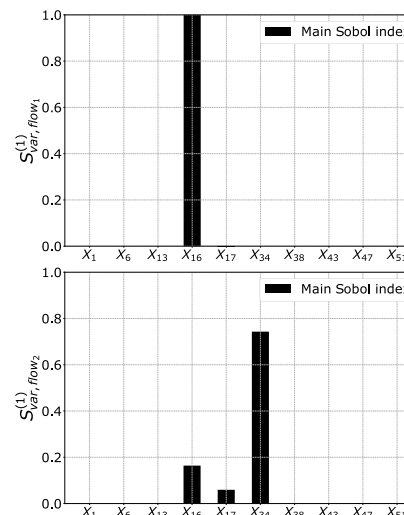


Fig 7: First order Sobol indices of the mean of mass flows 1 and 2. On the x-axis is a selection of the epistemic uncertain variables.

By comparing Figs. 6 and 7, it can be observed that the Sobol indices of the variance are very similar to the Sobol indices of the mean, and we obtain the same important epistemic variables and the same importance ranking. However, this result depends on the behavior of the applied model. As an example, in [14] it is shown that a completely different ranking is obtained for the analytical Ishigami function. It then depends on the purpose of the analysis which Sobol index is of interest.

5 Conclusions

Within this work, two different methods for mixed aleatory-

epistemic UQ have been investigated and applied successfully on a SAS model from aerospace engineering. The IVP approach computes intervals bounding the statistical QoI of the response. The width of these intervals quantifies the effect of epistemic uncertainty. The BP approach on the other hand gives as an output the full distribution of the statistical QoI of the response, and epistemic statistics on this statistical QoI. It has been shown that within the BP method, BCIs can be computed, which quantify epistemic uncertainty in the form of an interval, therefore comparable with the IVP resulting intervals.

Uncertainty-based SA in the mixed aleatory-epistemic uncertainty setting, using the IVP and BP approaches, has also been investigated. An empirical pinching strategy for sensitivity analysis of the epistemic variables when using the IVP method has been applied. Using the BP approach, novel expressions for the Sobol indices of a statistical QoI of the response have been presented and interpreted. These indices compute how much each epistemic variable, when taken into account independently, or when interacting with other epistemic variables, influences the variance of the statistical QoI of the response. Such a sensitivity analysis results in an importance ranking of the epistemic variables, when the effect of aleatory (irreducible) uncertainty has been integrated out.

Bibliography

- [1] ANTINORI, G.: *Uncertainty analysis and robust optimization for low pressure turbine rotors*. Doctoral thesis, Technical University of Munich, Germany (2016).
- [2] BEER, M.; FERSON, S.; KREINOVICH, V.: *Imprecise probabilities in engineering analyses*. Mechanical Systems and Signal Processing 37 (2013), pp. 4-29.
- [3] DER KIUREGHIAN, O.; DITLEVSEN, O.: *Aleatory or epistemic. Does it matter?* Structural Safety 31 (2007), pp. 105-112.
- [4] EHRE, M.; PAPAIOANNOU, I.; STRAUB, D.: *Efficient estimation of variance-based reliability sensitivities in the presence of multi-uncertainty*. Proc. 19th working conference of the IFIP working group 7.5 on Reliability and Optimization of Structural Systems, Zürich, Switzerland (2018).
- [5] ELDRED, M. S.; SWILER, L. P.; TANG, G.: *Mixed aleatory-epistemic uncertainty quantification with stochastic expansions and optimization-based interval estimation*. Reliability Engineering & System Safety 96 (2011), pp. 1092-1113.
- [6] FERSON, S.; TROY TUCKER, W.: *Sensitivity analysis using probability bounding*. Reliability Engineering & System Safety 91 (2006), pp. 1435-1442.
- [7] GHANEM, R. G.; SPANOS, P. D.: *Stochastic finite elements: a spectral approach, 2nd*. Courier Dover Publications, Mineola (2003).

- [8] HELTON, J. C.; JOHNSON, J. D.; OBERKAMPF, W. L.: *An exploration of alternative approaches to the representation of uncertainty in the model predictions*. Reliability Engineering & System Safety 85 (2004), pp. 39-71.
- [9] HELTON, J. C.; OBERKAMPF, W. L.: *Alternative representations of epistemic uncertainty*. Reliability Engineering & System Safety 85 (2004), pp. 1-10.
- [10] OBERGUGGENBERGER, M.; KING, J.; SCHMELZER, B.: *Classical and imprecise probabilities methods for sensitivity analysis in engineering: A case study*. International Journal of Approximate Reasoning 50 (2009), pp. 680-693.
- [11] ROLLS ROYCE: *The jet engine*. (1986).
- [12] SANDIA NATIONAL LABORATORIES.: *Dakota software version 6.4 user's manual* (2016).
- [13] SUDRET, B.: *Global sensitivity analysis using polynomial chaos expansions*. Reliability Engineering & System Safety 93 (2008), pp. 964-979.
- [14] KAMARIOTIS, A.: *Mixed aleatory-epistemic uncertainty quantification and sensitivity analysis*. Master thesis, Technical University of Munich, Germany (2019).

Authors

M.Sc. Antonios Kamariotis
Technische Universität München
Engineering Risk Analysis Group
Arcistrasse 21
80333 München, Germany
antonis.kamariotis@tum.de

Dr.-Ing. Giulia Antinori
MTU Aero Engines AG
Dachauer Strasse 665
80995 München, Germany
giulia.antinori@mtu.de

Dr.-Ing. Iason Papaioannou
Technische Universität München
Engineering Risk Analysis Group
Arcistrasse 21
80333 München, Germany
iason.papaioannou@tum.de

Prof. Dr.-Ing. Fabian Duddeck
Technische Universität München
Computational Mechanics Group
Arcistrasse 21
80333 München, Germany
duddeck@tum.de