AIAA SciTech Forum
8-12 January 2024, Orlando, FL
AIAA SCITECH 2024 Forum

10.2514/6.2024-0172

Certified interval model updating using scenario optimization

Robin R. P. Callens * and David Moens†
KU Leuven Department of Mechanical engineering Division LMSD, St.-Katelijne-Waver 2860 Belgium.

Matthias G. R. Faes‡
TU Dortmund University, Chair for Reliability Engineering, Department of Mechanical Engineering, Dortmund 44227 Germany.

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

Interval model updating is typically performed when gathering data is expensive, time-consuming, or complex and only a limited amount of data is available to perform non-deterministic model updating. In these situations, the fitted intervals will only provide an estimate of the exact interval bounds. This is because the limited data available is unlikely to include any samples that fall precisely on the interval boundaries. In these situations, an analyst could use a metric to assess the accuracy of identified model uncertainties against unseen missing data. Furthermore, when this metric is able to estimate the required amount of data for accurate uncertainty quantification, data-gathering efforts are minimized. This paper defines this metric as the reliability of a data-enclosing set as the probability that future unseen data will fall within the set. Recently, Crespo et al. [1] presented a scenario optimization approach to determine a lower bound for this reliability without having to characterize the underlying distribution of the data generation mechanism. To calculate the reliability, the scenario optimization approach needs the number of hyper-parameters to fit the data enclosing set, the number of samples, and the dimension of the data enclosing set. Once these are obtained, and a confidence level is determined, the approach calculates the lower bound of the reliability. Additionally, analysts can calculate the number of samples required to fit the data enclosing set with predefined lower bound reliability before the measurement campaign. The goals of this paper are to develop the certified interval model updating based on scenario optimization and to apply this to a dynamical modal analysis of a structural finite element model. A four-level building numerical model is used to illustrate the accuracy and the practical application of the developed methodologies.

I. Introduction

Interval model updating is a non-deterministic method that is used to update the uncertainty in models based on indirect measurement data. This method considers the potential range of values for each measurement, rather than assuming that the measured values are precise or conform to a predefined probabilistic distribution. Accounting for the uncertainty in model updating is especially crucial when the measurements are prone to significant uncertainty or noise. Intervals are powerful in cases where gathering data is difficult, time-consuming, or expensive [2, 3]. Therefore, several procedures for interval model updating have been introduced in the literature [4–6]. Fang et al. [7] used interval response surface models to save computational cost and to limit the typical overestimation of interval model updating approaches. Additionally, methodologies employing a Kriging predictor model were proposed [8, 9]. Faes et al. [10] described an efficient multivariate interval model updating for high dimensional models under scarce data availability. Alternatively, when data collection is straightforward, analysts can use probabilistic techniques to describe and update the uncertainty, such as the popular probabilistic model updating technique, Bayesian model updating [11], even though in this case care should be taken with the definition of the prior distributions [12].

Regardless of whether interval or probabilistic uncertainty is considered, having a metric for the accuracy of the identified model uncertainties against unseen missing data would provide the analyst with a tool to assess the "quality" of the identified uncertainty. Furthermore, it can effectively reduce the data-gathering effort to a strictly required minimum. This paper defines such a metric utilizing the concept of the lower bound reliability of a data enclosing set. Here, reliability is defined as the probability that future unseen data will fall in the data-enclosing interval set. Recently,

*PhD Student, Department of Mechanical Engineering Division LMSD, St.-Katelijne-Waver 2860 Belgium.
†Professor, Department of Mechanical engineering Division LMSD, St.-Katelijne-Waver 2860 Belgium, and AIAA senior member.
‡Full Professor, Department of Mechanical Engineering Chair for Reliability Engineering, Dortmund 44227 Germany.

Copyright © 2024 by the American Institute of Aeronautics and Astronautics, Inc. All rights reserved.

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

Crespo et al. [1] presented a scenario optimization approach to determine a lower bound value for this reliability without having to characterize the underlying distribution of the data generation mechanism. To calculate the lower bound reliability, the scenario optimization approach needs the number of hyper-parameters to fit the data enclosing set, the number of samples, and the dimension of the data enclosing set. Together with a confidence level, the lower bound of the reliability is calculated. Additionally, analysts can calculate the number of samples required for fitting the data enclosing set with predefined lower bound reliability prior to the measurement campaign.

The goals of this paper are (1) to develop the certified interval model updating based on scenario optimization, (2) to show the accuracy of this methodology in a numerical application of an interval model updating based on a modal analysis, and (3) to illustrate the practical application of the interval model updating. Hereto the scenario optimization is used to calculate the required number of data measurements for interval fitting. Scenario optimization can also be used during the interval analysis procedure to calculate the required number of data samples for interval analysis. In case the interval model updating is a convex optimization problem that yields the global minimum, the certified interval model updating provides the total lower bound reliability with a known confidence level on the updated input interval set.

The paper is organized as follows. Section II introduces interval analysis and solution techniques based on optimization and vertex analysis. Section III presents the scenario optimization theory and the application to the interval model updating. Section IV illustrates the interval model updating on a four-level building and makes an in-depth accuracy analysis. Section V concludes the paper by summarizing the key results of this paper.

II. Interval analysis

An interval scalar $x^I \in \mathbb{IR}$ represents an uncertain parameter x that has a fixed but unknown value, bounded by the lower bound $\underline{x}$ and upper bound $\overline{x}$. x^I is the interval set defined as $\left[\underline{x}, \overline{x}\right]$, which is a set of real numbers that satisfies $\underline{x} \leq x \leq \overline{x}$, and $\mathbb{IR}$ is the domain of closed, real-valued intervals. The midpoint x_m and width x_w of an interval scalar x^I are defined as

$$x_m = \frac{\overline{x} + \underline{x}}{2}, \tag{1}$$

$$x_w = \frac{\overline{x} - \underline{x}}{2}. \tag{2}$$

An interval vector $\boldsymbol{x}^I \in \mathbb{IR}^{d_X}$ with independent interval scalars x_i^I, where $i = 1, \ldots, d_X$, is defined as $\boldsymbol{x}^I = x_1^I \times x_2^I \times \ldots \times x_{d_X}^I$, with $\times$ denoting the Cartesian product. The interval vector $\boldsymbol{x}^I$ can also be denoted using a set notation as:

$$\boldsymbol{x}^I = \left\{x_1^I, x_2^I, \ldots, x_{d_X}^I\right\}^T = \left\{\boldsymbol{x} \in \mathbb{R}^{d_X} \mid x_i \in x_i^I\right\}. \tag{3}$$

In interval analysis, an interval vector $\boldsymbol{x}^I$ defines the hyper-rectangular input-space of a function $\mathcal{M} : \mathbb{R}^{d_X} \mapsto \mathbb{R}^{d_Y}, \boldsymbol{x} \to \boldsymbol{y}$. Herein is $\mathcal{M}$ the numerical model that consists of d_Y deterministic functions $m_i : \mathbb{R}^{d_X} \mapsto \mathbb{R}, \boldsymbol{x} \to y_i$, where $i = 1, \ldots, d_Y$. The output of $\mathcal{M}$ when given the input $\boldsymbol{x}^I$ is represented as a solution set $\boldsymbol{y}^S \in \mathbb{R}^{d_Y}$ bounding the model responses $\boldsymbol{y}$ of interest. This set is explicitly given as follows:

$$\boldsymbol{y}^S = \left\{\boldsymbol{y} \mid \boldsymbol{y} = \mathcal{M}(\boldsymbol{x}), \boldsymbol{x} \in \boldsymbol{x}^I\right\}. \tag{4}$$

Since finding the exact set $\boldsymbol{y}^S$ in the general case constitutes an NP-hard problem, $\boldsymbol{y}^S$ is usually bounded by an interval vector $\boldsymbol{y}^I \in \mathbb{IR}^{d_Y}$: $\boldsymbol{y}^I = \left\{y_1^I \quad y_2^I \quad \ldots \quad y_{d_Y}^I\right\}^T$. The individual components of $\boldsymbol{y}^I$ are $y_i^I = \left[\underline{y}_i, \overline{y}_i\right]$ and are determined with an (anti-)optimisation procedure, i.e.,

$$\underline{y}_i = \min_{\boldsymbol{x} \in \boldsymbol{x}^I} \left(m_i(\boldsymbol{x})\right), \tag{5}$$

$$\overline{y}_i = \max_{\boldsymbol{x} \in \boldsymbol{x}^I} \left(m_i(\boldsymbol{x})\right). \tag{6}$$

This optimization procedure finds the components $\underline{y}_i, \overline{y}_i$ of each interval y_i^I of the interval vector $\boldsymbol{y}^I$ independently, resulting in an approximation of the solution set $\boldsymbol{y}^S$ as a conservative hyper-rectangle. The optimization approach requires a total of $2d_Y$ optimization problems to be solved, each requiring potentially numerous model evaluations of $\mathcal{M}$. For a model $\mathcal{M}$ that is monotonic in $\boldsymbol{x}^I$ the interval of each component y_i^I of $\boldsymbol{y}^I$ can be found by vertex analysis. In this approach, each component is found independently by minimizing/maximizing the model responses of the set of vertex points $\mathcal{V}$ of $\boldsymbol{x}^I$. This set contains all possible input parameter combinations located at vertex points of the interval input vector $\boldsymbol{x}^I$, yielding 2^{d_X} combinations. For large $d_{\mathcal{X}}$ such an approach becomes computationally demanding, even when one evaluation of $\mathcal{M}$ is computationally fast to solve. In the situation of a large $d_{\mathcal{X}}$ and a linear model, the authors recently introduced a computationally efficient alternative with Multilevel Monte Carlo for interval analysis [13]. Also, [14] introduced a triple-engine parallel Bayesian global optimization for efficient interval analysis for expensive to evaluate numerical models and $d_{\mathcal{X}} < 20$.

III. Certified Interval model updating

A. Reliability of data enclosing sets

In this paper, the reliability of a data-enclosing set is defined as the probability that future unseen data points will fall in the data-enclosing set. The recently presented scenario theory [1] enables bounding this reliability without having to characterize the underlying distribution of the data generation mechanism. As such, it provides a distribution-free measure for the quality of the fitted interval bounds of the governing uncertain quantity.

The cost function of a data-driven constrained optimization is defined as $\mathcal{J} : \Theta \to \mathbb{R}$ with design variable $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, and n_θ the number of design variables. Also, the set of design points satisfying the design requirements for scenario (data samples) $\delta \in \Delta$ are denoted as h_δ. For each individual scenario $\delta^{(i)} \in \mathbf{D}$ this is $h_{\delta^{(i)}}$. Then, consider the constrained, data-driven scenario program

$$\theta^*(\mathbf{D}) = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \mathcal{J}(\theta) : \theta \in \bigcap_{i=1}^{n} h_{\delta^{(i)}} \right\}, \tag{7}$$

where the data $\mathbf{D} = \delta^{(1)}, \dots, \delta^{(n)}$ is sampled from a stationary data generating mechanism. The unknown probability measure governing the underlying data generation mechanism is defined as P. The reliability of θ^* is $1 - V(\theta^*)$, with $V(\theta^*)$ being the violation defined as

$$V(\theta^*) = P\left[\delta \in \Delta | \theta^* \notin h_\delta\right]. \tag{8}$$

The lower the violation, the higher the reliability of θ^*. As the data set $\mathbf{D}$ is chosen randomly out of infinitely many possible data sets of size n, the θ^* and thus also $V(\theta^*)$ is random. This randomness can be quantified by using

$$P^n\left[V(\theta^*) \leq \epsilon\right] \geq 1 - \beta. \tag{9}$$

This equation defines that the probability $P^n = P \times \cdots \times P$ of the violation of θ^* being less or equal to $\epsilon \in [0, 1]$ is greater than $1 - \beta$. Herein represents $\beta \in [0, 1]$ the confidence and ϵ called the reliability. Note that θ^* is a random element that depends on n randomly chosen samples from P. Therefore, the violation probability $V(\theta^*)$ can be greater than ϵ for some random observations but not for others, and β refers to the probability P^n of observing one of those bad sets of n samples. As a result, the confidence β is thus key to obtaining results that are guaranteed independently of P.

As a result, scenario theory allows evaluating (9) without making any assumption about P. When the optimization program (7) is convex, the hyperparameters are obtained with a convex optimization from the data set, $V(\theta^*)$ is dominated by a beta distribution [15, 16], and ϵ can be calculated from

$$\binom{k + n_\theta - 1}{k} \sum_{i=0}^{k+n_\theta-1} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \leq \beta, \tag{10}$$

where $k < n - n_\theta$ is the number of data points/outliers removed from the data set $\mathbf{D}$ before θ^* was calculated. Equation (10) allows the analyst to compute a lower bound reliability estimate before θ^* is computed. As a result, the lower bound reliability can be calculated before data is drawn from the data generation mechanism [1].

B. Randomized interval analysis

As discussed in section II, interval analysis (see the discussion after (5) and (6)) is computationally demanding when $d_{\mathcal{Y}}$ is high or/and when $d_{\mathcal{X}}$ is high. The (anti-)optimization requires $2 * d_{\mathcal{Y}}$ optimization procedures from which the individual optimization procedures possibly require numerous model evaluations. The vertex analysis required in total 2^{d_X} model evaluations to solve the interval analysis. Yet, it is only applicable for models $\mathcal{M}$ that are monotonic in

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

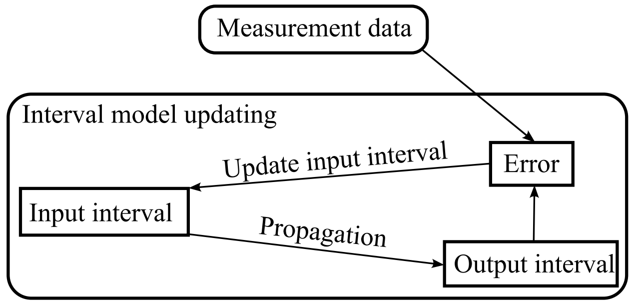

Fig. 1 Interval model updating flowchart

$\boldsymbol{x}^{\boldsymbol{I}}$. The vertex analysis has the advantage over the (anti-) optimization procedure to always yield the exact result and to be less computationally demanding. However, the vertex analysis requires numerous model evaluations when the number of input dimensions $d_{\mathcal{X}}$ is moderately high. For example, for an interval analysis solved with vertex analysis and a model with $d_{\mathcal{X}} = 16$ a total of 65.536 model evaluations are required.

Another new approach is the randomized interval analysis. Here the input interval vector $\boldsymbol{x}^{\boldsymbol{I}}$ is sampled randomly from a known distribution. These samples are then propagated through the numerical model $\mathcal{M}$. On the output side, an approximation to the exact output intervals $\boldsymbol{y}^{\boldsymbol{I}}$ is independently found by the minimal and maximal model response. However, the obtained output intervals are inherently an inner approximation of the exact output intervals.

The crucial question in this new approach is "*How many samples from the input interval vector* $\boldsymbol{x}^{\boldsymbol{I}}$ *are required to guarantee that the resulting intervals are close to the exact output intervals* $\boldsymbol{y}^{\boldsymbol{I}}$*?*". The scenario optimization discussed in paragraph III.A answers this question, with (10). For example, for the case with $d_{\mathcal{X}} = 16$ and a lower bound reliability level of $\epsilon = 0.95$ and confidence $(1 - \beta) = 0.99$ a total of 945 samples are required. The vertex method requires for the same example 65.536 model evaluations. This example illustrates the increase in computational efficiency of randomized sampling over vertex analysis. However, the computationally efficient randomized sampling technique is not free as it does not yield the exact output intervals $\boldsymbol{y}^{\boldsymbol{I}}$. Still, thanks to the scenario optimization approach a lower bound reliability level ϵ is guaranteed with a known confidence level β.

It is also important to note that the results obtained from (10) are independent of the sampling distribution used for taking samples of the input interval vector $\boldsymbol{x}^{\boldsymbol{I}}$. As a result, it is also applicable in the situation where the distribution is unknown which is the case for interval analysis. More powerfully, it allows the user to select any sort of input distribution depending on the case at hand, and use this to estimate the output interval vector $\boldsymbol{y}^{\boldsymbol{I}}$ with its lower bound reliability level ϵ and confidence β.

C. Certified interval model updating

Interval model updating is the process of updating the input interval vector $\boldsymbol{x}^{\boldsymbol{I}}$ of the numerical model $\mathcal{M}$ based on the output $\boldsymbol{y}$ of the numerical model and measurement data $\boldsymbol{y_m}$. Figure 1 visualizes the interval model updating process as a flowchart.

The measurement data is generated first. In this stage, the analyst is typically confronted with the question: "*How many samples do I have to take?*". To answer this question for measurement data used to fit an interval, equation 10 of the scenario optimization methodology is used. Fitting an interval around a measurement data-set $\mathbf{Y} \in \mathbb{R}^{d_{\mathcal{Y}}}$ containing n samples, in paragraph III.A referred to as scenarios $\boldsymbol{\delta}^{(i)}$ here $\boldsymbol{y}^{(i)}$ with $i = 1, \cdots, n$ in $d_{\mathcal{Y}}$ dimensions is defined as

$$\boldsymbol{y}^{\boldsymbol{I}} = \begin{bmatrix} \underline{y_1} = min(y_1) & \overline{y_1} = max(y_1) \\ \vdots & \vdots \\ \underline{y_{d_{\mathcal{Y}}}} = min(y_{d_{\mathcal{Y}}}) & \overline{y_{d_{\mathcal{Y}}}} = max(y_{d_{\mathcal{Y}}}) \end{bmatrix}. \tag{11}$$

For each dimension $d_{\mathcal{Y}}$ of $\boldsymbol{y}^{(i)}$, two interval bounds are calculated. As a result, the total number of design variables is $n_\theta = 2 * d_{\mathcal{Y}}$.

The lower bound reliability estimation calculated with equation 10 requires a random data sampling strategy and a convex optimization procedure to find all $\boldsymbol{x}^{\boldsymbol{I},*}$ of the interval fitting. With superscript ,* denoting the input interval vector obtained after model updating. Fitting intervals satisfies this requirement as finding the interval set of a randomly

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

sampled data set yields the exact interval bounds for that data set. This, as fitting an interval set around randomly sampled data is simply finding the minimum and maximum value of the data.

After sampling the data samples, the interval model updating starts. In this paper, the interval model updating minimizes the distance between the fitted output interval vector $\boldsymbol{y}^{\boldsymbol{I}}$ of the numerical model and the fitted interval vector that bounds the measurement data $\boldsymbol{y_m}^{\boldsymbol{I}}$. The total error function consists of two parts

$$e_m = \left\| \frac{\boldsymbol{y}_m - \boldsymbol{y}_{m_m}}{\boldsymbol{y}_{m_m}} \right\|_2, \tag{12}$$

$$e_w = \left\| \frac{\boldsymbol{y}_w - \boldsymbol{y}_{m_w}}{\boldsymbol{y}_{m_w}} \right\|_2, \tag{13}$$

where $||\cdot||_2$ is the $\mathcal{L}_2$ norm the subscripts $_m$ and $_w$ denote respectively the midpoint and width of an interval y_i^I. Equation (12) calculates the error on the midpoint, and Equation (13) calculates the error on the width. Hereto, both equations exploit the $\mathcal{L}_2$ norm of the normalized error between the interval vector around measurement data $\boldsymbol{y}_m^I$ and the interval vector $\boldsymbol{y}^{\boldsymbol{I}}$ obtained using a numerical model $\mathcal{M}$. Both error terms are normalized with the interval vector $\boldsymbol{y}_m^I$ of the measurement data, this to make every individual component $y_{m,i}$ of $\boldsymbol{y}_m$ and y_i of $\boldsymbol{y}_m$ equally contributing to the error terms. Both error terms are combined as

$$error = 0.5\frac{e_m}{e_{m_0}} + 0.5\frac{e_w}{e_{w_0}}, \tag{14}$$

with e_{m_0} and e_{w_0} the error terms that are calculated after an initial guess before the optimization is started. The optimization itself is defined as:

$$\boldsymbol{x}^{\boldsymbol{I},*} = \underset{x^I \in \mathbb{IR}^{d_X}}{\operatorname{argmin}}(error). \tag{15}$$

When the minimization is convex, a gradient-based optimization procedure yields the global minimum up to a numerical error. One of the requirements for convex optimization is using a numerical model $\mathcal{M}$ with a monotonic response to the input intervals $\boldsymbol{x}^{\boldsymbol{I}}$. Note that only this requirement is not enough to obtain a convex optimization, the convexity is case-dependent. For the case of a convex interval model updating, the total lower bound reliability ϵ_{tot} estimation of the interval model updating is

$$\epsilon_{tot} = \epsilon_{IA} * \epsilon_{data} \tag{16}$$

with ϵ_{IA} the lower bound reliability estimation of the interval analysis used during the model updating and ϵ_{data} the lower bound reliability estimation of the data. This equation holds when the model updating is independent of the data and the confidence for both ϵ_{IA} and ϵ_{data} is equal. Next, 3 typically seen cases of model updating are discussed.

Case 1: a convex optimization combined with the vertex analysis. In this situation, a 100% reliability of the interval analysis is obtained. This is as the vertex analysis yields the exact interval vector up to a numerical error and the convex optimization that will obtain the exact minimum up to a numerical error. The total lower bound reliability on $\boldsymbol{x}^{\boldsymbol{I},*}$ for this case is

$$\epsilon_{tot} = \epsilon_{data}. \tag{17}$$

Case 2: a convex optimization combined with the randomized sampling approach. The total lower bound reliability ϵ_{tot} on $\boldsymbol{x}^{\boldsymbol{I},*}$ is in this case defined as:

$$\epsilon_{tot} = \epsilon_{IA} * \epsilon_{data}, \tag{18}$$

where ϵ_{IA} is the lower bound reliability of the randomized sampling approach used during every iteration of the optimization procedure. Similar to the previous case, the optimization converges to the global minimum. However, in this case, the global minimum is not the exact solution as the randomized sampling approach is used. As a result, the lower bound reliability of the interval analysis step ϵ_{IA} is equal to the lower bound reliability of the randomized sampling approach. Note that for this case to have a convex optimization, the randomness must be the same for every iteration of the optimization procedure. This is typically ensured by fixing the random seed to a single value for the complete optimization procedure. When the random seed is not fixed, a noisy non-linear optimization has to be solved which is not convex.

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

IV. Illustration of the Certified Interval Model Updating

This section illustrates the certified interval model updating on a numerical dynamical model of a four-level building. Figure 2 shows the four-level building model. In the model, the mass of the four levels is equal to $m = 4000$ kg $= m_1, m_2, m_3, m_4$. The stiffness values k_1, k_2, k_3, k_4 are different for each connection between the floors and defined as interval uncertainty. The interval vector $\boldsymbol{k}^{\boldsymbol{I}}$ is defined as:

$$\boldsymbol{k}^{\boldsymbol{I}} = \begin{bmatrix} k_1^I \\ k_2^I \\ k_3^I \\ k_3^I \end{bmatrix} = \begin{bmatrix} 4960 & 5040 \\ 4180 & 4620 \\ 4470 & 4730 \\ 5730 & 5870 \end{bmatrix} \text{N/m}$$

This numerical model $\mathcal{M}$ calculates based on the input masses and stiffness, the first four eigenfrequencies $\phi_1, \phi_2, \phi_3, \phi_4$.

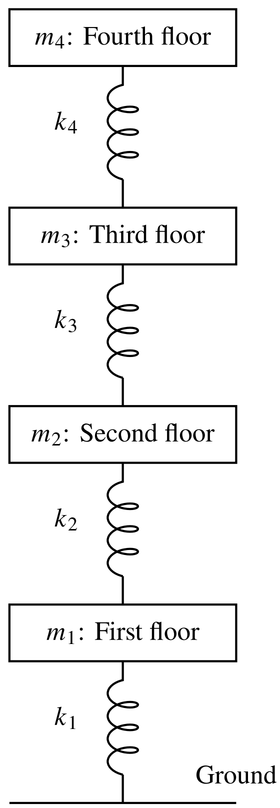

Fig. 2 Dynamic model of a four-level building with masses $\{m_1, m_2, m_3, m_4\}$ **and stiffness values** $\{k_1, k_2, k_3, k_4\}$**.**

A. Virtual measurement data

The first step in this illustration is the creation of a virtual measurement data set. From hereon, $\phi_{i,m}$ denotes an output variable with virtual measurement data. This virtual measurement data set consists of virtual measurements of the eigenfrequencies $\phi_{1,m}, \phi_{2,m}, \phi_{3,m}, \phi_{4,m}$. The virtual measurement data is generated with the numerical model $\mathcal{M}$ (see figure 2). Hereto, uniformly distributed samples are generated from the stiffness interval vector $\boldsymbol{k}^{\boldsymbol{I}}$. These samples are then propagated through the numerical model $\mathcal{M}$ to obtain the virtual measurement data samples $\boldsymbol{\phi}_m^{(i)}$ containing n samples in d_ϕ dimensions. Around the virtual measurement data set, an interval vector $\boldsymbol{\phi}_m^{\boldsymbol{I}}$ is fitted with

$$\boldsymbol{\phi}_m^{\boldsymbol{I}} = \begin{bmatrix} \underline{\phi_{1,m}} = min(\phi_{1,m}) & \overline{\phi_{1,m}} = max(\phi_{1,m}) \\ \vdots & \vdots \\ \underline{\phi_{d_\phi,m}} = min(\phi_{d_\phi,m}) & \overline{\phi_{d_\phi,m}} = max(\phi_{d_\phi,m}) \end{bmatrix}. \tag{19}$$

The output interval vector $\boldsymbol{\phi}_m^I$ defines the smallest independent hyper-rectangle around the output samples $\boldsymbol{\phi}_m^{(i)}$. The question here is "*How many virtual data samples are enough to have an accurate interval fitting?*". A rearranged equation (10) provides the minimum number of samples n required to obtain at least the lower bound reliability ϵ_{data} with a known confidence β. The confidence here is the probability that the reliability of the fitted interval vector $\boldsymbol{\phi}_m^I$ is lower than the predefined ϵ_{data}. Yet it is unclear what the influence of the chosen reliably is on the accuracy of the interval model updating. An in-depth discussion is provided in paragraph IV.C.

For now, a few assumptions are made: the lower bound reliability $\epsilon_{data} = 0.95$, the confidence of the reliability $\beta = (1 - 0.999)$, and the number of outliers that will be removed $k = 0$. The number of design variables for fitting the interval set around the virtual measurement data set $\boldsymbol{\phi}_m^I$ is $n_\theta = 2 * 4 = 8$ (see paragraph III.C).

Then based on these assumptions, the rearranged version of equation (10) provides the minimum number of samples $n = 386$. From these samples the obtained output vector $\boldsymbol{\phi}_m^I$ is

$$\boldsymbol{\phi}_m^I = \begin{bmatrix} \underline{\phi_{1,m}} = 0.0593 & \overline{\phi_{1,m}} = 0.0602 \\ \underline{\phi_{2,m}} = 0.1775 & \overline{\phi_{2,m}} = 0.1790 \\ \underline{\phi_{3,m}} = 0.2699 & \overline{\phi_{3,m}} = 0.2737 \\ \underline{\phi_{4,m}} = 0.3251 & \overline{\phi_{4,m}} = 0.3294 \end{bmatrix} \text{Hz.}$$

B. Interval Model Updating

The interval model updating uses an optimization program to minimize the error between the simulated generated data set and the propagated data set based on the to-be-updated input intervals. The minimized error function is equation (15). Each iteration of the model updating uses vertex analysis to propagate the input interval vector $\boldsymbol{k}^I$ to the output interval vector $\boldsymbol{\phi}^I$ through the numerical model $\mathcal{M}$. This is possible as the eigenvalue analysis of the numerical model is monotonic and $d_k = 4$ is low. An advantage of the vertex analysis is that the exact interval bounds are obtained. The convexity of the interval model updating is ensured by the combination of (1) the response of the numerical model to varying stiffness values is monotonic (2) the eigenmodes are tracked with a Modal Assurance Criterium (3) when a switch in eigenmodes is detected, the corresponding eigenvalues are switched as well. As a result, the interval model updating can use a gradient descent algorithm.

The algorithm is chosen to be of the type SQP, yet this algorithm can change without loss of generality of the methodology. For this optimization, the stopping criterion is a maximum 5000 model $\mathcal{M}$ evaluations and a step tolerance of $1e^{-6}$. The convex optimization yields a resulting interval input vector $\boldsymbol{x}^{I,*}$. This convex optimization with the vertex analysis illustrates case 1 discussed in section III. As a result, the total lower bound reliability ϵ_{tot} of the resulting interval input vector $\boldsymbol{x}^{I,*}$ is defined by equation (17) and is equal to the lower bound reliability ϵ_{data} of the virtual measurement interval vector $\boldsymbol{\phi}_m^I$ that is 0.95 with $(1 - 0.999)$ confidence.

The optimization algorithm of the interval model updating stops with a remaining error lower than $1e^{-6}$. Figure 3 shows the interval bounds on the output side, the obtained interval output vector $\boldsymbol{\phi}^{I,*}$ after propagating the interval model updating result $\boldsymbol{k}^{I,*}$ and the interval around the virtual measurement data $\boldsymbol{\phi}_m^I$. From this figure and the low remaining error, it is clear that in the output space the optimization found an optimum.

The obtained interval set on the input side $\boldsymbol{k}^{I,*}$ is summarized in table 1 together with the initially defined interval set $\boldsymbol{k}^I$ to generate the numerical data. This shows that the obtained input intervals $\boldsymbol{k}^{I,*}$ have a small error in comparison to $\boldsymbol{k}^I$. This error originates from the uniform sampling used to generate the virtual measurement data interval vector $\boldsymbol{\phi}_m^I$. The measurement data generated using samples from $\boldsymbol{k}^I$ have a high probability of not including any bounds of $\boldsymbol{k}^I$. The added value of the scenario optimization in this situation is that it gives a certificate that the reliability of the obtained interval bounds $\boldsymbol{x}^{I,*}$ is at least 0.95 with $(1 - 0.999)$ confidence against missing unseen data.

Table 1 Comparison of the obtained stiffnesses and the stiffnesses used for data generation all units are N/m

Stiffness	defined input interval	model updating result
k_1^I	[4960 5040]	[4960 5056]
k_2^I	[4180 4620]	[4194 4587]
k_3^I	[4470 4730]	[4477 4700]
k_4^I	[5730 5870]	[5750 5859]

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

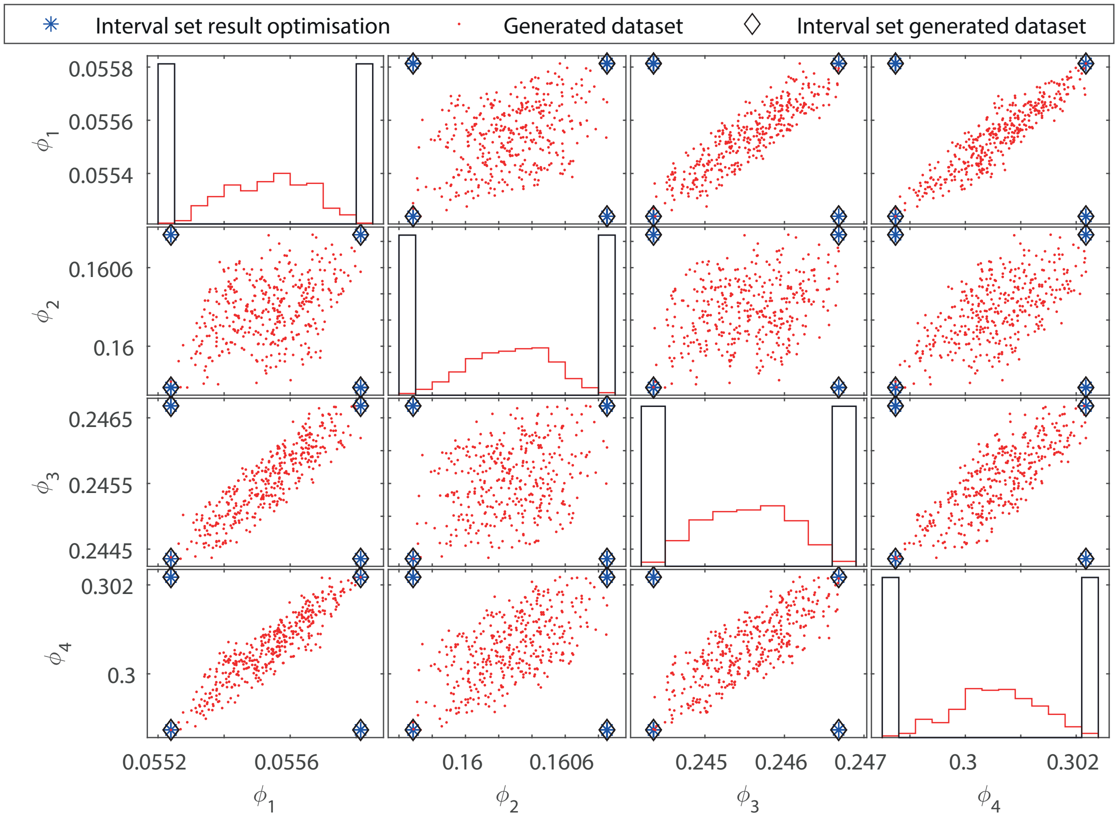

Fig. 3 The numerically generated data set in red "·" with the fitted interval set in black "◊" and the set of the propagated solution in blue big dots "∗".

C. Effect of the amount of virtual measurement data on the accuracy of interval model updating

The main question for this interval model updating is "*How many virtual measurement data samples are required to have an accurate interval model updating?*". This is studied by performing interval model updating, as performed in the previous paragraphs, with varying the lower bound reliability ϵ_{data} from 0.1 to 0.99 for the virtual measurement data. The change in ϵ_{data} is directly changing the required minimum amount of virtual measurement data samples. The accuracy of the obtained input interval set $\boldsymbol{x}^{\boldsymbol{I},*}$ is defined as:

$$d_m = \left\| \frac{\boldsymbol{k}_m - \boldsymbol{k}_m^*}{\boldsymbol{k}_m} \right\|_2, \tag{20}$$

$$d_w = \left\| \frac{\boldsymbol{k}_w - \boldsymbol{k}_w^*}{\boldsymbol{k}_w} \right\|_2, \tag{21}$$

Here, the distances d_m, d_w are defined as the $\mathcal{L}_2$ norm of the normalized difference between the midpoints and widths respectively of the obtained interval vector $\boldsymbol{k}^{\boldsymbol{I},*}$ and the exact interval vector $\boldsymbol{k}^{\boldsymbol{I}}$. The total distance is defined as

$$d_{tot} = 0.5 d_m + 0.5 d_w. \tag{22}$$

This distance is calculated for a range $n = [13, 1956]$ of number of virtual measurement data samples $\boldsymbol{\phi}_m^{(i)}$ and with 50 different random seed numbers. The 50 different random seed numbers are selected to see the mean effect instead of the effect from one single random seed. Figure 4 visualizes the mean "–·"-line and the envelope "·"-line of this distance d_{tot} on the left vertical axis with an increasing number n of virtual measurement data samples $\boldsymbol{\phi}_m^{(i)}$ on the horizontal axis. On the right vertical axis is the lower bound reliability ϵ_{tot} visualized also with the same increasing number of virtual measurement data samples. This figure shows that for a larger number n, the distance is lower and the lower bound reliability becomes higher. For larger lower bound reliability $\epsilon_{tot} > 0.95$ the additional required virtual measurement data samples n increase fast. At the same time, the extra accuracy gains are decreasing. For a real measurement campaign, there will be a trade-off between accuracy (high lower bound reliability) and computational cost.

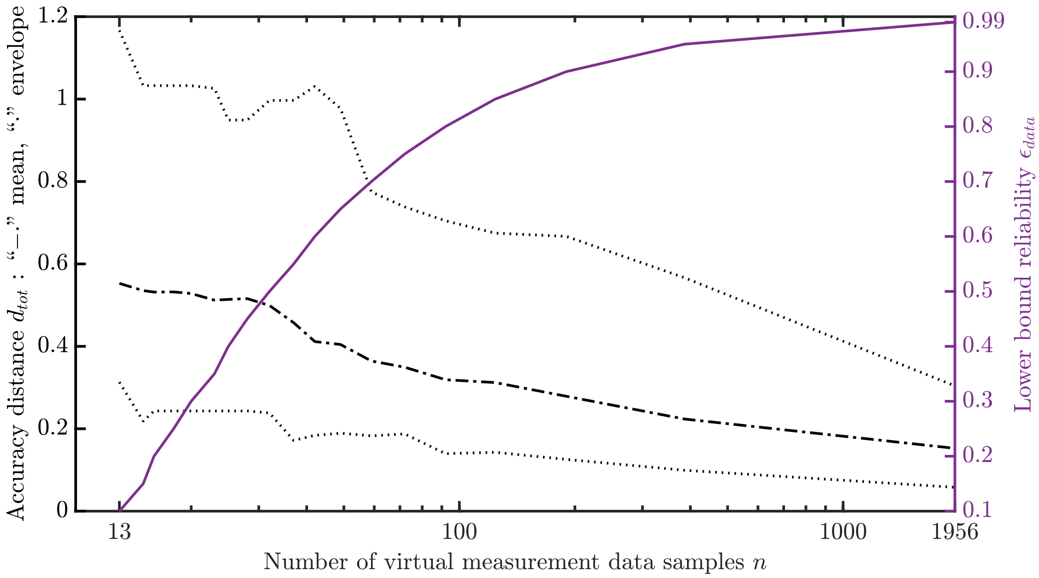

Fig. 4 On the left vertical axis the distance d_{tot} between the obtained input interval vector $\boldsymbol{k}^{I,*}$ and the exact interval vector $\boldsymbol{k}^{I}$ for increasing number of virtual measurement data samples n. With mean "–·"-line and envelope "·"-line. On the right vertical axis the lower bound reliability level ϵ_{tot} for an increasing number of virtual measurement data samples.

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

D. Propagation with the randomized interval analysis

In this paragraph, the added value of the randomized interval analysis is illustrated. Assume that instead of the vertex method the randomized interval analysis would be applied to the four-level numerical model. Then the main question is "*How many samples during interval analysis are required to obtain an accurate updated input interval?*"

This is studied by performing interval model updating with randomized interval analysis by varying the lower bound reliability of the randomized interval analysis ϵ_{IA} from 0.1 to 0.99. The change in ϵ_{IA} is directly changing the required minimum amount of samples used in the interval analysis. The obtained result of the interval model updating $\boldsymbol{k}^{I,*}$ is finally compared with the exact interval set $\boldsymbol{k}^{I}$. The virtual measurement interval vector $\boldsymbol{\phi}_m^I$ is here generated with the vertex analysis, in order to reduce the accuracy error from creating the virtual measurement data interval vector $\boldsymbol{\phi}_m^I$ to zero. As a result, the accuracy of the randomized interval analysis technique is studied based on the accuracy of the full interval model updating.

The accuracy is similarly defined as in paragraph IV.C. Here the accuracy distance d_{tot} is calculated for a range $n = [13, 1956]$ of data samples and with 50 different random seed numbers used for the randomized interval analysis. The 50 different random seed numbers are selected to see the mean effect instead of the effect from one single random seed. Figure 5 visualizes the mean "–·"-line and the envelope "·"-line of the accuracy distance d_{tot} on the left vertical axis with an increasing number of data samples n on the horizontal axis. On the right vertical axis is the lower bound reliability ϵ_{IA} visualized also with the same increasing number of data samples. This figure shows that for a larger number of data samples n, the accuracy distance d_{tot} is lower and the lower bound reliability ϵ_{AI} becomes higher. For lower bound reliability $\epsilon_{AI} > 0.95$ the additional required data samples n increase fast. At the same time, the extra accuracy gains are decreasing.

Similar to the measurement data, in a real situation, there will be a trade-off between accuracy (high lower bound reliability) and computational cost. Note that this paragraph is only intended to showcase the randomized interval. For this low-dimensional $d_{\mathcal{X}} = 4$ case the vertex analysis is more accurate and more computationally efficient.

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172

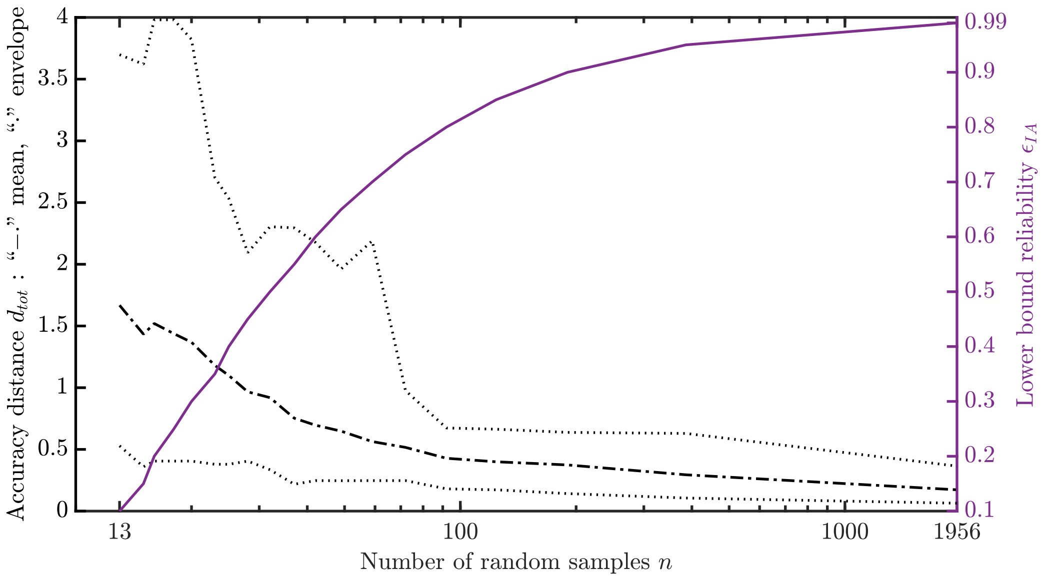

Fig. 5 On the left vertical axis the accuracy distance d_{tot} between the obtained input interval vector $k^{I,*}$ and the exact interval vector k^{I} for increasing number of data samples. With mean "–·"-line and envelope "·"-line. On the right vertical axis the lower bound reliability ϵ_{AI} for an increasing number of data samples n.

V. Conclusions

This paper presents a novel approach that performs an interval model updating where a certificate against unseen missing data is provided. Hereto the approach in this paper utilizes scenario theory to rigorously bound the probability of unseen data falling outside an identified data enclosing set. The lower bound reliability approach provides the analyst with additional insight into the obtained results compared to other existing techniques. For instance, it returns information on whether enough data is generated to get a certain lower bound reliability of the results, and what the lower bound reliability is when only very limited data is available. The scenario theory also allows for a more computationally efficient interval analysis for moderate to large numbers of input intervals compared to interval techniques in the literature. In this paper, the certified interval model updating is illustrated on a modal analysis of a 4-level building. This illustration shows the additional insides of the scenario optimization theory. The minimum amount of measurement data samples is calculated with the scenario optimization theory. For 4 input intervals and a predefined lower bound reliability of 95% with 99, 9% confidence, the minimum amount of samples measurement data samples is 386. The 95% reliability of the fitting interval set is also the total lower bound reliability of the interval model updating. This is due to the fact that the optimization is convex. Also, the randomized interval analysis approach is introduced to reduce the computation cost of interval analysis for medium to large uncertainty dimensions. With this approach, random samples are taken from an interval with a distribution that can be chosen by the analyst. The scenario theory provides here, also, a lower bound reliability depending on the amount of samples. The accuracy from the certified interval model updating increases when a higher lower bound reliability is selected and thus more samples are taken. Yet more computational time or more measurement data is required. In practice, it is concluded that there will always be a trade-off between (1) the cost linked to computational resources and measurement data and (2) the accuracy of the obtained result. Here, the reliability certificate provides the analyst with more information to make this trade-off.

Acknowledgments

The authors would like to acknowledge the financial support of the Research Foundation - Flanders in the framework of the PhD fellowship strategic basic research 1SD2423N of Robin R.P. Callens.

References

[1] Crespo, L. G., Colbert, B. K., Kenny, S. P., and Giesy, D. P., "On the quantification of aleatory and epistemic uncertainty using Sliced-Normal distributions," *Systems & Control Letters*, Vol. 134, 2019, p. 104560. https://doi.org/10.1016/j.sysconle.2019.104560, URL https://linkinghub.elsevier.com/retrieve/pii/S0167691119301707.

[2] Ferson, S., and Ginzburg, L. R., "Different methods are needed to propagate ignorance and variability," *Reliability Engineering & System Safety*, Vol. 54, No. 2, 1996, pp. 133–144. https://doi.org/10.1016/S0951-8320(96)00071-3, URL https://www.sciencedirect.com/science/article/pii/S0951832096000713.

[3] Imholz, M., Faes, M., Vandepitte, D., and Moens, D., "Robust uncertainty quantification in structural dynamics under scarse experimental modal data: A Bayesian-interval approach," *Journal of Sound and Vibration*, Vol. 467, 2020, p. 114983. https://doi.org/10.1016/j.jsv.2019.114983, URL https://www.sciencedirect.com/science/article/pii/S0022460X19305450.

[4] Faes, M., and Moens, D., "Recent trends in the modeling and quantification of non-probabilistic uncertainty," *Archives of Computational Methods in Engineering*, 2019, pp. 1–39. Publisher: Springer.

[5] Fedele, F., Muhanna, R. L., Xiao, N., and Mullen, R. L., "Interval-Based Approach for Uncertainty Propagation in Inverse Problems," *Journal of Engineering Mechanics*, Vol. 141, No. 1, 2015, p. 06014013. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000815, URL https://ascelibrary.org/doi/10.1061/%28ASCE%29EM.1943-7889.0000815, publisher: American Society of Civil Engineers.

[6] Deng, Z., Guo, Z., and Zhang, X., "Interval model updating using perturbation method and Radial Basis Function neural networks," *Mechanical Systems and Signal Processing*, Vol. 84, 2017, pp. 699–716. https://doi.org/10.1016/j.ymssp.2016.09.001, URL https://www.sciencedirect.com/science/article/pii/S0888327016303417.

[7] Fang, S.-E., Zhang, Q.-H., and Ren, W.-X., "An interval model updating strategy using interval response surface models," *Mechanical Systems and Signal Processing*, Vol. 60-61, 2015, pp. 909–927. https://doi.org/10.1016/j.ymssp.2015.01.016, URL https://www.sciencedirect.com/science/article/pii/S0888327015000229.

[8] Govers, Y., Haddad Khodaparast, H., Link, M., and Mottershead, J. E., "A comparison of two stochastic model updating methods using the DLR AIRMOD test structure," *Mechanical Systems and Signal Processing*, Vol. 52-53, 2015, pp. 105–114. https://doi.org/10.1016/j.ymssp.2014.06.003, URL https://www.sciencedirect.com/science/article/pii/S088832701400226X.

[9] Khodaparast, H. H., Mottershead, J. E., and Badcock, K. J., "Interval model updating with irreducible uncertainty using the Kriging predictor," *Mechanical Systems and Signal Processing*, Vol. 25, No. 4, 2011, pp. 1204–1226. https://doi.org/10.1016/j.ymssp.2010.10.009, URL https://www.sciencedirect.com/science/article/pii/S0888327010003286.

[10] Faes, M., Cerneels, J., Vandepitte, D., and Moens, D., "Identification and quantification of multivariate interval uncertainty in finite element models," *Computer Methods in Applied Mechanics and Engineering*, Vol. 315, 2017, pp. 896–920. https://doi.org/10.1016/j.cma.2016.11.023, URL https://www.sciencedirect.com/science/article/pii/S0045782516307447.

[11] Lye, A., Cicirello, A., and Patelli, E., "Sampling methods for solving Bayesian model updating problems: A tutorial," *Mechanical Systems and Signal Processing*, Vol. 159, 2021, p. 107760. https://doi.org/10.1016/j.ymssp.2021.107760, URL https://www.sciencedirect.com/science/article/pii/S0888327021001552.

[12] Faes, M., Broggi, M., Patelli, E., Govers, Y., Mottershead, J., Beer, M., and Moens, D., "A multivariate interval approach for inverse uncertainty quantification with limited experimental data," *Mechanical Systems and Signal Processing*, Vol. 118, 2019, pp. 534–548. https://doi.org/10.1016/j.ymssp.2018.08.050, URL https://www.sciencedirect.com/science/article/pii/S0888327018305946.

[13] Callens, R. R. P., Faess, M. G. R., and Moens, D., "MULTILEVEL QUASI-MONTE CARLO FOR INTERVAL ANALYSIS," *International Journal for Uncertainty Quantification*, Vol. 12, No. 4, 2022, pp. 1–19.

[14] Dang, C., Wei, P., Faes, M. G. R., Valdebenito, M. A., and Beer, M., "Interval uncertainty propagation by a parallel Bayesian global optimization method," *Applied Mathematical Modelling*, Vol. 108, 2022, pp. 220–235. https://doi.org/10.1016/j.apm.2022.03.031, URL https://www.sciencedirect.com/science/article/pii/S0307904X22001196.

[15] Campi, M. C., and Garatti, S., "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM Journal on Optimization*, Vol. 19, No. 3, 2008, pp. 1211–1230. Publisher: SIAM.

[16] Campi, M. C., and Garatti, S., "A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality," *Journal of optimization theory and applications*, Vol. 148, No. 2, 2011, pp. 257–280. Publisher: Springer.

Downloaded by Matthias Faes on January 4, 2024 | http://arc.aiaa.org | DOI: 10.2514/6.2024-0172