# How to Exploit What We Know About Input and Model: A Trans-probabilistic Approach to the 2022 AIAA UQ Challenge

Peter O. Hristov*
*GATE Institute, Sofia, Bulgaria*
*and*
*University of Liverpool, Liverpool, UK*

Ioanna Ioannou†,and Scott Ferson‡
*University of Liverpool, Liverpool, L69 3GH*

**This paper addresses the aerodynamics uncertainty challenge problem presented in [1]. The principles of an information-based approach to wrangle uncertainty are presented and applied to problems associated with the propagation of stochasticity, ignorance and numerical uncertainty. Prior knowledge that the engineer has about the model is exploited to efficiently quantify the uncertainty in the model's output without the use of surrogate models. Indeed, the approach presented in this paper uses a simple surrogate model only during the quantification of numerical uncertainty, yet achieves model evaluation levels, typical of a surrogate-assisted uncertainty propagation effort. By carefully distinguishing between aleatory and epistemic uncertainty from characterisation to propagation, the adopted methodology provides the engineer with the means to objectively assess the trust they can put in model predictions for the intended application and to help them take evidence-based decisions. The approaches adopted in this paper are trans-probabilistic in that they utilise probability only when needed but no further and instead opt to propagate uncertainties as efficiently and appropriately as possible.**

## Nomenclature

| | | |
|---|---|---|
| $c_l$ | = | coefficient of lift |
| $c_m$ | = | coefficient of pitching moment |
| $d$ | = | input dimensionality |
| $F_b$ | = | $b$-th focal element |
| $\underline{F}(x)$ | = | left bound of p-box |
| $F(x)$ | = | cumulative distribution function |
| $\overline{F}(x)$ | = | right bound of p-box |
| $n_{int}$ | = | number of intervals |
| $n_b$ | = | number of p-box variables |
| $n_F$ | = | number of focal elements |
| $n_s$ | = | number of samples |
| $\underline{p}^e_{c_x}$ | = | left bound for excursion probability of coefficient |
| $\overline{p}^e_{c_x}$ | = | right bound for excursion probability of coefficient |
| $Re$ | = | Reynolds number |
| $\underline{x}$ | = | left bound on $x$ |
| $\overline{x}$ | = | right bound on $x$ |
| $x^I$ | = | interval-valued variable $x$ |
| $x_{tb}$ | = | lower surface flow-transition location |
| $x_{tt}$ | = | upper surface flow-transition location |
| $\alpha$ | = | angle of attack |

*Research Associate, GATE Institute and Institute for Risk and Uncertainty
†Research Associate, Institute for Risk and Uncertainty
‡Professor, Institute for Risk and Uncertainty

$\delta_f$ = flap deflection angle

# I. Introduction

## A. Background and motivation

The relative inexpensiveness and efficiency of computer models, in comparison to physical experiments, have increased their utility in modern engieering. Yet, despite continuous improvements to the fidelity of computer models and the growing range of phenomena they are capable of representing, their role in engineering is still limited to non-critical, usually well-understood stages of design. Aerospace engineering is an excellent demonstrator of this trend. Where it is needed the most, namely in providing ways to explore and assess the safety of bold, non-standard, data-scarce solutions to efficiency and environmental problems for the next generation of aircraft, modelling is only used in a limited way. Very often the final driver in approving a particular design, is whether or not this design can be feasibly tested and certified experimentally. This conservatism is rightfully due to the inability to objectively assess the amount of trust that design and decision-making teams can put in model predictions. At the same time, the extreme intensity and costs associated with safety-critical aspects of aircraft development, most notably certification, have led to an increasing interest in certification by simulation [2]. It has already been recognised that uncertainty quantification (UQ), and not extreme model precision is the key enabler to the use of modelling in the safety-critical stages of design. This is because, regardless of its complexity and precision, the model invariably remains only an approximation to reality, due to computational and epistemic reasons. On the other hand, a careful and honest quantification of uncertainty can provide an imprecise, but rigorous prediction, along with a measure of confidence in this prediction [3].

Despite the rapid academic advances in its field, UQ is seeing a relatively slow adoption in the aerospace industry, much like the models, whose utility it attempts to increase. This is owed to three main reasons. First, UQ requires specialised training to be implemented correctly and efficiently. This is a problem, because UQ expertise is not a core capability in the aerospace industry and thus requires additional investment on part of different companies. Second, UQ methods are inherently resource-demanding, as they unavoidably rely on multiple model evaluations (or experimental replications, if applied to physical testing). Third, the transfer of methodologies from the academic to the industrial setting requires access to sensitive data and information restricted due to intellectual and industrial property rights. For these reasons industrially-led challenge problems are of extreme importance to furthering the adoption of UQ in the aerospace industry. Albeit necessarily idealised, especially due to the third reason stated above, when designed carefully, such challenge problems can contain many of the traits of a larger-scale uncertainty quantification modelling effort, necessary to check the state of the field against problems relevant to the industry (see e.g., [4, 5]).

This paper presents a solution to the *Uncertainty Quantification Challenge Problem for Aerodynamics*, posed in [6]. The challenge focuses on the problem of assessing the performance of a particular airfoil as measured by its lift and moment coefficient and controlled by five physical flow and configuration characteristics. The respondents are asked to propagate the uncertainty about the nominal values of the five variables through the high-order panel method XFOIL [7] to the coefficients, and to quantify the uncertainty due to the discrete nature of the computation. The goal is to assess whether different uncertainties will drive the performance of the airfoil outside of some safe interval. Given the quantified output uncertainty, the participants are asked to provide advise on whether the airfoil can safely be used in subsequent design. Thus, the challenge addresses both the methodological part of UQ and its main goal - to provide support to decision-makers. The five variables, along with their uncertainty characterisation are presented in Table 1. In the remainder of this paper, only those details of the challenge problem that are relevant to the presented solution methodologies will be included. For a complete presentation of the problem itself the reader is referred to [1, 6].

## B. Review of existing methodologies

The main focus of the challenge problem is uncertainty propagation (UP). The topic of UP is central to model-based UQ, as it lays the foundation and determines the validity of all other activities that can be seen as parts UQ, such as model verification, validation and predictive capability estimation [8], sensitivity analysis [9, 10], calibration [11], and reliability analysis [12]. It has been known for a long time that uncertainty arises due to many sources and comes in different types [13, 14]. Two of these types that are most commonly seen in practice are aleatory and epistemic uncertainty. There is a general consensus among engineers and practitioners that the two kinds of uncertainty are fundamentally different and should not be lumped together [15, 16]. However, there is still a disagreement about whether and how these two types of uncertainty should be described mathematically and subsequently propagated through computer models. Despite evidence to the fact that probability theory is not well-suited to dealing with epistemic uncertainty [17], there are researchers who insist that probability theory can adequately capture both kinds of uncertainty [18].

Similar trends are spreading among the aerospace research and engineering community. Even though it may be difficult to distinguish aleatory from epistemic uncertainty, the fact that they seek to express fundamentally different aspects of design has been agreed upon [19], but not reflected in computational work and uncertainty quantification [20–24]. Such a practical mischaracterisaition can often lead to grossly incorrect results and misleading conclusions. The authors of the present paper offer three main reasons why this confounding of uncertainty persists. Firstly, the majority of the available literature on UQ adopts a probability-based approach to propagating epistemic uncertainty. This serves as an evidence to engineers that do not specialise in the field that the probabilistic treatment must be the one to go for. Indeed, theoretical and practical guidance on correctly working with epistemic uncertainty has been around for several decades [3, 25–27], but has somehow not gained sufficient popularity among practitioners. Secondly, the vast majority of computer models in use for engineering purposes are either black-box or crystal-box models, either of which means that their source code can not be (easily) modified. At the same time, epistemic uncertainty propagation methods, such as interval analysis [28, 29] and probability bounds analysis [30, 31] require the expressions of the computer code to be modified to provide the mathematically guaranteed bounds they are so appealing for. Finally, the application of these methods requires thoughtful consideration about what is known and what is only assumed, and about various aspects of the model and inputs. This in turn requires specialised training and dedicated time on part of the engineer who needs to consult the UQ expert in their work. Not all three roadblocks can be addressed fully. However, some recent effort to bring attention to these issues includes the work presented in [32] and [33]. The challenge this paper addresses seems to have been designed to carefully distinguish between aleatory and epistemic uncertainty, both philosophically and methodologically as it requires the model output uncertainty to be quantified under an all-aleatory, all-epistemic and a mixed uncertainty scenarios, as well as to compare an interval to a probabilistic treatment of epistemic uncertainty. This is an important step towards removing the first and addressing, at least partially, the second roadblock above.

In this paper, the work described in [33] is extended with a new method for the propagation of epistemic uncertainty, some methods to propagate aleatory uncertainty and ways to quantify numerical discretisation error. The remainder of the paper is structured as follows. Section II briefly reviews the computational methods used in addressing the challenge and includes a practical new method for increasing computational efficiency in propagating probability boxes through black-box models. Section III presents an exploratory analysis of the model output and the main results of the paper, along with a discussion of what these results mean for the decision-maker. Section IV presents an approach to deal with numerical uncertainty stemming from the discrete nature of the code. Finally, Section V draws some conclusions and outlines directions for future work.

## II. Methodological overview

### A. Methods for epistemic uncertainty propagation

Epistemic uncertainty is supposed to reflect true ignorance or lack of knowledge about a quantity or subject. Thus, a quantity $x$ that is epistemically uncertain is best described with the use of a mathematical interval, $x^I = [\underline{x}, \overline{x}]$. Intervals should be constructed in such a way as to bound the reasonably expected value of the quantity, $x$. This begs the question, about the choice of the bounds of the interval. If $x^I$ is supposed to reflect lack of knowledge, how are $\underline{x}$ and $\overline{x}$ chosen. It

**Table 1 Input variables for the challenge problem along with their aleatory and epistemic uncertainty characterisations.**

| Variable | Symbol | Unit | Aleatory distribution | Epistemic interval |
|---|---|---|---|---|
| Angle of attack | $\alpha$ | ° | N(0, 0.1) | [−0.3, 0.3] |
| Reynolds number | $Re$ | - | N($5 \times 10^5$, $2.5 \times 10^3$) | [$4.925 \times 10^5$, $5.075 \times 10^5$] |
| Flap deflection | $\delta_f$ | ° | N(0, 0.08) | [−0.24, 0.24] |
| Upper surface flow-transition location | $x_{tt}$ | - | N(0.3, 0.015) | [0.255, 0.345] |
| Lower surface flow-transition location | $x_{tb}$ | - | N(0.7, 0.021) | [0.637, 0.763] |

is true that this choice must also be justified, but it is usually much simpler to make than choosing an infinite-dimensional marginal distribution shape over the real line, or a dependency structure over the input space. Moreover, subject matter experts can much more easily provide two bounding numbers than a complex description of variation. Intervals thus arise for several reasons:

1) theoretical or assumed limits about values of a parameter,
2) insufficient data to prescribe a distribution,
3) measurement uncertainty, non-detects and data censoring, periodic observation, sensor precision, and
4) uncertainty introduced in the modelling process or other bounding studies.

Once constructed, the best way to propagate an interval through a computer model is to use interval arithmetic [28]. This form of uncertainty calculus provides rigorous computation results (that is, if the operands of a particular binary operation are sure to enclose the true value of their respective quantity, then the result of the operation is guaranteed to enclose the true answer of the output quantity) but requires the mathematical expressions of the source code to be rewritten with their interval arithmetic counterparts. In addition to the rigour, interval arithmetic results are also best-possible, in that the output interval of the operation cannot be made any smaller without tightening the operand intervals. These properties provide what some authors refer to as *verified computation* [3].

In the absence of an accessible source code, the engineer must resort to approximations for the intervals, based on sample propagation through the black-box. This fact causes a great deal of confusion with aleatory uncertainty, which is usually propagated via sampling. The difference between working with the two types of uncertainty is that when propagating aleatory uncertainty, the engineer preserves and characterises the computer model output uncertainty by the empirical distributions (either as cumulative distribution function (CDF) or probability density function (PDF)) constructed with these samples. In contrast, non-intrusive interval propagation methods are only concerned with the bounds of the output, since any other uncertainty characterisation will be tantamount to injecting information during the propagation process.

In the most general case, sample-based interval propagation solves two constrained optimisation problems as

$$\begin{aligned} y^I &= [\underline{y}, \overline{y}] \\ &= \left[\min\left(f\left(x^I\right)\right), \max\left(f\left(x^I\right)\right)\right] \end{aligned} \tag{1}$$

Samples are usually not generated beforehand, leaving it to the optimisation routine to choose the points according to some fitness function. Depending on the anticipated complexity of the surface of the model output, simple local optimisation methods, such as the Neldear-Mead simplex method [34], or more complex evolutionary and heuristic algorithms [35] can be used. When the use of optimisation is warranted, the propagation can easily require on the order of tens of thousands of model evaluations, which is why surrogate models are often employed to reduce the cost of the analysis [36]. This choice adds an extra layer of uncertainty and is discussed in more detail in Section III.

If there is sufficient evidence to believe that the code behaves monotonically over the prescribed range of uncertainty, *vertex propagation* can be used. The vertex propagation method [37] is a straightforward way to project intervals through the code, by projecting a number of input combinations given by the Cartesian product of the interval bounds. This results in a total of $n_s = 2^d$ evaluations, where $d$ is the number of interval-valued inputs. In the case of two intervals, $x^I$ and $y^I$, the code, $f(\cdot)$ must be evaluated four times at $f(\underline{x}, \underline{y})$, $f(\underline{x}, \overline{y})$, $f(\overline{x}, \underline{y})$, $f(\overline{x}, \overline{y})$. The main advantage of the method is its simplicity and rigour, given the monotonicity assumptions hold. The exponential computational cost of this method means it is limited to a relatively low-dimensional problems or else also bounded to surrogate modelling.

If the function encoded in the computer model is not monotonic over the input intervals, one can use *subinterval reconstitution* [31] to break up each interval into smaller subintervals over which the monotonicity assumption is more realistic (provided the underlying function is not pathologically rough) and propagate those, either using the vertex method or some other way. In the end, the union of all output intervals is taken to represent the overall interval for the response of the model. Subinterval reconstitution suffers from an even higher cost than vertex propagation, because there are multiple intervals to propagate per dimension.

For functions that are linear this paper introduces a convenient simplification to the vertex method is termed here *extreme-point propagation*. The method works by choosing a base point, here taken as the lowest vertex of the input space and then running the code at this point. Next the code is run, while changing one input to its high value at a time in a factorial fashion. The difference between the runs corresponding to each input point and the run at the base value, provides an estimation to the first derivatives of the function with respect to each input and most importantly the signs of these derivatives. These signs can then be used to inform the engineer which input combinations will provide a rigorous result for the output interval. The total cost of the method is $d + 3$ code evaluations, $d + 1$ of which are needed to compute the derivatives and 2 to propagate the final extreme points through the code. This method is demonstrated in Section III, where it is adopted as the primary uncertainty propagation approach. It is noted that even though the extreme-point propagation method is a combination of established numerical procedures, it has not been applied to the propagation problem in the literature, to the best of the authors' knowledge.

Other methods to propagate epistemic uncertainty through black-box models exist, but these are not detailed here for conciseness. The interested reader is referred to [33].

### B. Methods for aleatory uncertainty propagation

Unlike for epistemic uncertainty, probability theory is ideally suited to characterising stochastic variability. Once characterised, the distributions can be propagated through the black-box model using random or one of a number of systematic sampling methods. Examples of such methods are equiprobable sampling, Latin hypercube sampling [38], or low-discrepancy sequences. In the case the input is characterised by a non-standard distribution with a known CDF, the probability integral transform can be used to generate correctly distributed samples.

Aleatory uncertainty propagation methods are straightforward to implement, but the variance of their estimates decreases as $\sqrt{n_s}$. Moreover, (quasi-) random sampling requires on average $n_e$ samples to discover (sample ones) from a region with probability of occurrence $p^e$ [39], which means $n_e$ can grow rapidly if the engineer wishes to sample close to the tails of the output distribution.

## III. Uncertainty propagation

The main goal of the challenge is to assess whether the lift and moment coefficients of a NACA 2412 airfoil with a simple trailing edge flap, and subject to uncertain flow conditions will exceed the critical values, $c_l = [0.155, 0.265]$ and $c_m = [-0.050, -0.044]$. The challenge is divided into 5 sub-problems as follows:

1) Aleatory uncertainty propagation - consider all inputs to be affected by the aleatory uncertainty shown in Table 1,
2) Epistemic uncertainty propagation - consider all inputs to be affected by the epistemic uncertainty shown in Table 1,
3) Discretisation error quantification - estimate the discretisation error due to using a low number of panels for the solver and its impact on the uncertainty of $c_l$ and $c_m$,
4) Uniform distribution propagation - consider all inputs to be affected by aleatory uncertainty characterised as uniform distributions with bounds equal to the epistemic intervals in Table 1.
5) Mixed uncertainty propagation - consider that a) only the uncertainty of the flap angle, $\delta_f$ is characterised as epistemic; b) that the upper and lower transition point, $x_{tt}$ and $x_{tb}$, respectively are also characterised as having epistemic uncertainty.

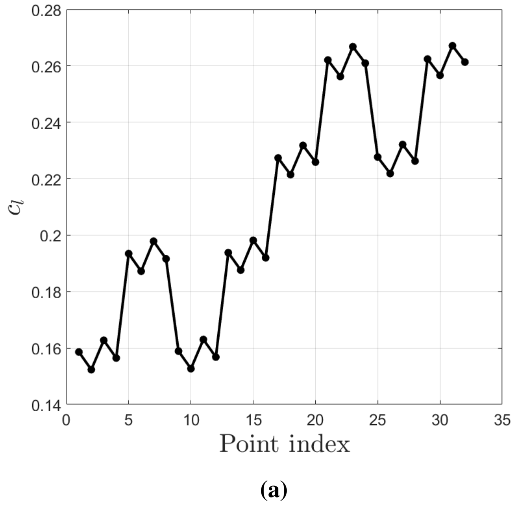
Anticipating the results in the remainder of this section, a conscious choice to avoid the use of surrogate models is made as this will complicate the uncertainty one has to deal with. Instead, a careful analysis of the model and prior physical knowledge is used to reduce the cost of the analyses. It must be noted, however, that the authors are not in any way subjectively against the use of surrogate models. They merely advocate that such models are not used automatically without giving due consideration to other options. In fact, surrogate models are often the only option for uncertainty propagation.

In order to maximise the insight into the model's behaviour, while keeping the computational budget as low as possible, the epistemic analysis is conducted first. This is because an uncertainty propagation effort with an all-epistemic uncertainty is similar in nature to an *engineering control sensitivity analysis*[*] which is used to study the model before further analyses are carried out.

### A. Model analysis and epistemic uncertainty propagation

It is well-known that the lift and moment coefficients for airfoils at small angles of attack, $\alpha$, are a linear function of $\alpha$ [40, 41]. Moreover, given the relatively small magnitude of uncertainty around the nominal values, an initial assumption of a linear, or at least monotonic relationship between inputs and outputs is not too extreme. To test this assumption, the epistemic uncertainty was propagated using the vertex method, which only assumes there are no inflection points over the input intervals. The authors thought such an assumption entirely reasonable. The lift and moment coefficients for the $n_s = 2^5 = 32$ evaluations of the Cartesian product are shown in Fig. 1. To produce these

[*] See the upcoming DAWS Report on Sensitivity Analysis for Computer Models, currently hosted at https://sites.google.com/view/dawsreports/sa.



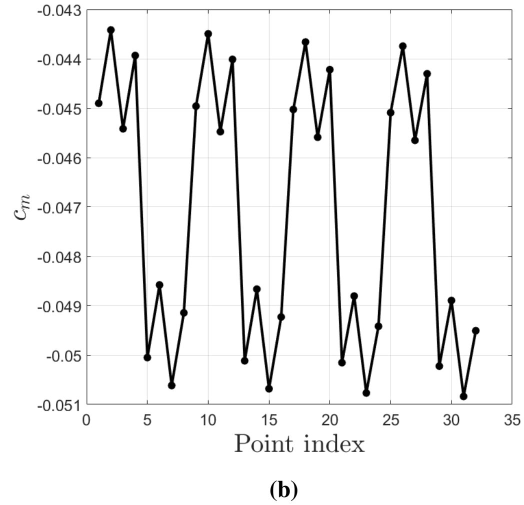




**Fig. 1 Variation of (a) lift and (b) pitching moment over the** $32$ **combinations of the Cartesian product for vertex propagation.**

results, the inputs were arranged as they are given in Table 1 and were varied such that $x_{tb}$ was altered at every input combination, $x_{tt}$ at every other combination, and so on, with $\alpha$ altered only once. Thus the high-frequency variation in both Fig. 1(a) and Fig. 1(b) correspond to changes in $x_{tb}$. Two main things can be observed from these figures. First, for the price of 32 model evaluations one can estimate the effects of all inputs and their interactions on the variations of the model output. The main effects of the inputs on $c_l$ and $c_m$ are $[0.0687, 0.0003, 0.0348, 0.0041, -0.0063]$ and $[0.0015, -0.0005, -0.0051, -0.0001, -0.0001]$, respectively ordered as in Table 1. It can be immediately seen that all inputs but $x_{tb}$ have positive influence on the lift and negative on the moment coefficients. It is also apparent that $\alpha$ is by far the strongest driver for $c_l$, while this role is given to $\delta_f$ for $c_m$. The roles of the two inputs switch for the lift and moment when it comes to influence runner-ups. Both outcomes align with known physical principles. Strong second-order interactions between $\alpha$ and $\delta_f$ are also present. Second, the overall influence of each input is linear to the resolution of the two plots. The numerical results, in conjunction to the physical evidence strongly suggest that the input-output function for both coefficients is monotonic.

Since this challenge is meant to serve as a demonstrator and the code can be run many times, an optimisation solution was run to test if the assumption would fail in an obvious way. In addition, an extreme-point propagation was also performed to see if the initial assumption, based only on physical insights would differ from the other two solutions. The results from all three methods are shown on Fig. 2 and presented numerically in Table 2. Despite the fact that the intervals look identical on the figures, Table 2 shows there is a minor difference between the results of the Nelder-Mead optimisation and the other two methods. This outcome is numerically insignificant, but provides an important reminder that random sampling will almost always produce an *inner approximation* to the output interval. Furthermore, the results from the vertex and extreme point propagation methods are identical. Even though computationally this is no surprise as the latter uses a subset of the evaluations of the former, this result goes to show that the function indeed exhibits strong linear behaviour, which can lead to an appreciable reduction in the number of code evaluations. Therefore performing the epistemic uncertainty propagation first, as described in this section, will allow the rest of the analyses to use this linearity information to increase their efficiency.

Both coefficients exceed their safe limits and given that all approaches are in good agreement about this, under the current state of knowledge, the airfoil should not be approved for further flight testing. It may generally be recommended that additional tests are requested, to determine the excursion probabilities with finer detail. Such an analysis is presented in the next section.

### B. Aleatory uncertainty propagation

Three factors should be taken into consideration from the outset, when propagating aleatory uncertainty. First, since the model is quasi-linear and Gaussian distributions are closed under linear transformations, the tails of the

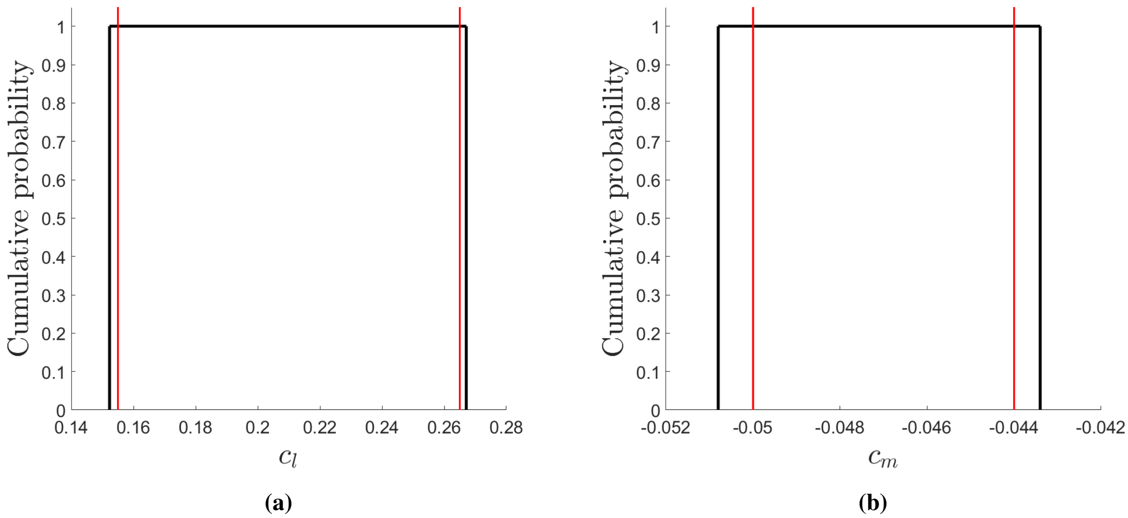
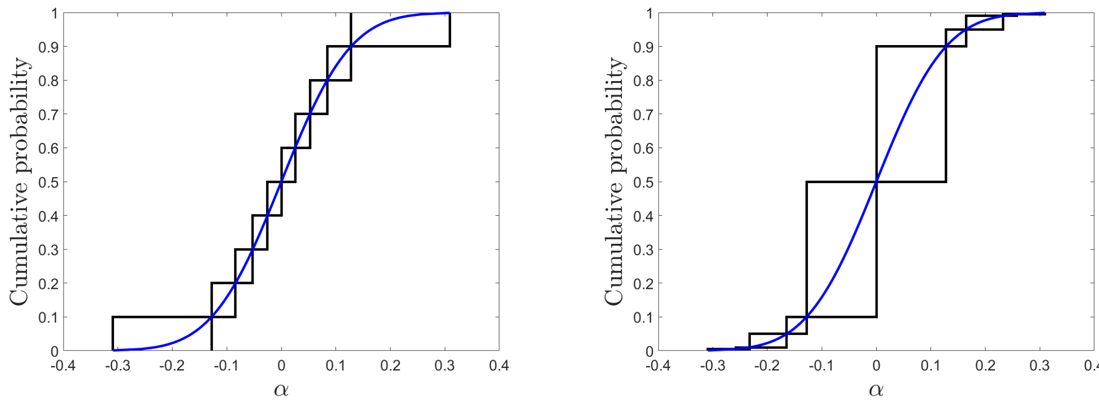



**Fig. 2 Epistemic interval for (a) lift and (b) pitching moment, propagated with the extreme point and vertex methods, as well as with Nelder-Mead optimisation. Results overlap to the resolution of the plots.**

**Table 2 Epistemic propagation results for the three methods, along with their computational cost.**

| | Extreme point | Vertex | Nelder-Mead |
|---|---|---|---|
| Lift coefficient | $[0.1523, 0.2671]$ | $[0.1523, 0.2671]$ | $[0.1523, 0.2668]$ |
| Moment coefficient | $[-0.0508, -0.0434]$ | $[-0.0508, -0.0434]$ | $[-0.0508, -0.0434]$ |
| Cost | 8 | 32 | 298 |




**(a)** **(b)**

**Fig. 3 Finite samples turn probability distributions into probability boxes. The discretisation can represent the whole distribution equally (a), or emphasise a particular part of it, such as the tails (b).**

distributions will need to be propagated to assess whether and with what probability the outputs exceed their respective safe ranges. Secondly then, this means that any random sampling will require many samples to reliably estimate these excursion levels. Thirdly, this will usually justify the use of a surrogate model for the aleatory propagation part of the challenge. However, since no surrogate is used in this work, an approach which minimises the number of samples without compromising rigour must be adopted.

To keep the number of model evaluations low, an arbitrary choice to use the equivalent to 10 samples to propagate the distributions was made. Such a low sample size will inevitably lead to a large sampling uncertainty. Therefore each distribution was turned to a probability box [30] with 10 discretisation levels. Probability boxes, or p-boxes for short, consist of a pair of CDFs, $[\underline{F}(x), \overline{F}(x)]$, which bound a set of distribution functions, such that for any CDF, $F(x)$, compatible with the p-box the relationship $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$ holds for all $x \in \mathbb{R}$. The p-box resulting from the discretisation of $\alpha$ into 10 equiprobable levels is depicted on Fig. 3(a). Notice that at the level where each sample would have been taken, there is no uncertainty about the value of the distribution. Everywhere else, however, the width of the staircase-looking intervals, called focal elements, bounds the uncertainty due to the lack of samples. It must be noted that the challenge asks that distributions be propagated to the 0.001 and 0.999 quantile, which will have the effect of slightly wider output ranges for the aleatory than for the epistemic uncertainties, whose input intervals correspond to the 0.0015 and 0.9985 quantiles.

The equiprobable discretisation creates focal elements towards the median of the distribution where fine-grained resolution is not needed when focusing on tail propagation. Therefore, a more appropriate, tail-emphasising discretisation is used, which is shown on Fig. 3(b). The focal elements with edges at the $[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99, 0.995, 0.999]$ quantiles represents the sampling uncertainty due to only using 10 samples, albeit spread out preferentially.

To propagate the newly discretised p-boxes, one can use the rules of probability bounds analysis (PBA) [26, 30]. By these rules when performing intrusive propagation, the Cartesian product of focal elements of the pair of p-boxes participating in a binary operation is formed and each element is propagated using interval arithmetic. In the present case, this simply means that any of the methods presented in Section III.A can be used. By construction the Cartesian product entails the propagation of $n_F^2$ intervals, where $n_F$ is the number of focal elements in the operands. The p-box resulting from the operation will thus have $n_F^{out} = n_F^2$ focal elements. If this strategy is repeated, without accounting for the growing number of focal elements, one will end up with $n_F^{out} = 1 \times 10^8$ intervals to propagate after only 3 binary operations, given that the input p-boxes have only $n_F = 10$ focal elements. To keep this growth under control, focal element *condensation* can be used which melts every $n_F^{out}/n_F$ focal elements of the output p-box into one, producing a p-box with $n_F$ focal elements. In this way, one always has to do $n_F^2$ interval projections. This condensation strategy would not work for the challenge problems, because the code of the model is inaccessible and one can at most see the very final result and not the outcomes of individual binary operations. Nominally this means that the total number of intervals to propagate is

$$\begin{aligned} n_{int} &= |C| \\ &= \bigtimes_{b=1}^{n_b} F_b \\ &= \prod_{b=1}^{n_b} n_F \end{aligned} \tag{2}$$

where $n_b$ is the number of p-box inputs to propagate, $\bigtimes$ denotes the Cartesian product of focal elements, $F_b$ and $|\cdot|$ denotes the cardinality of a set. The total cost of propagation will then be

$$n_{tot} = \sum_{c \in C} n_s \tag{3}$$

where $n_s$ is the number of samples to propagate each of the $n_{int}$ intervals. Thus for the challenge problem with $n_b = 5$, $n_F = 10$, and $n_s = 8$ one has to evaluate the code $n_{tot} = 800,000$ times which, if theoretically possible for the current example, would be infeasible for only moderately more expensive models.

To address this issue a new method for range-preserving propagation is adopted which substantially reduces the number of evaluations required. The development of the method is as follows. Consider the case where each input distribution is turned into a p-box, one input a time. All other inputs are replaced with intervals, whose widths are equal to the support of the respective distributions. Because interval-valued inputs have a single focal element, the number of intervals to be propagated through the code will be $n_{int} = n_F$. This propagation must be repeated for each of the $n_b$ inputs described by p-boxes. The resulting $n_b$ output p-boxes can be intersected with each other to give a single output p-box which has the same range as the p-box that would have resulted from a single-pass propagation, but at a cost of $n_{tot} = n_b n_F n_s$. This methods is termed *first-order condensation*, because it considers the variability of the inputs one at a time, i.e. to first order and then it condenses the final result by intersection. First-order condensation gains its efficiency at the cost of loosing information about the internal detail of the output p-box. That is, the final p-box could be wider than necessary.

The method, however, is general and does not require assumptions about the function or the inputs. To see this, consider the following general expression

$$y = e^f - ab + \frac{c}{d} \tag{4}$$

where $a \sim \mathrm{N}([10, 12], [0.5, 1])$, $b \sim \mathrm{Tri}(1, [2, 2.5], 3)$, $c \sim \mathrm{Weibull}([3, 4], [1, 2])$, $d \sim \mathrm{N}([1, 2], 0.2)$, $e \sim \mathrm{Exp}([1, 3])$, $f \sim \mathrm{U}([0.5, 0.9], [0.8, 1.2])$. Fig. 4 compares the intrusively propagated p-box for $y$ (black) to those

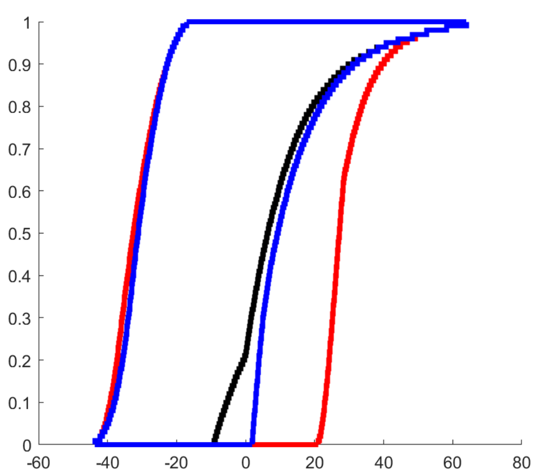
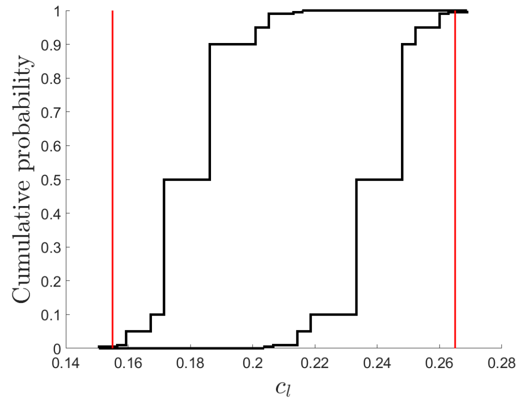



**Fig. 4 Comparison of black-box condensation strategies for the expression in Eq. (4). Original p-box in black, first-order condensation in red, second order condensation in blue. The range of the original p-box is always preserved.**



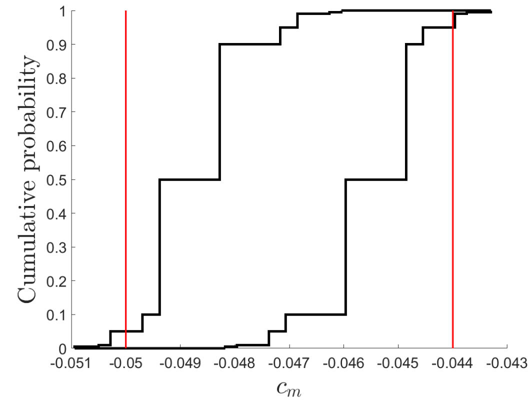




**Fig. 5 Final output p-boxes for (a) the lift coefficient and (b) the moment coefficient. Both outputs violate their safe ranges, but in contrast to the epistemic propagation, the probability of these violations can now be inspected.**

computed via first-order (red) and second-order condensation (blue). It can be seen that all p-boxes have the same range, with higher-order condensations tightening the internal width of the uncertain number.

Propagating the $n_b = 5$ input p-boxes with extreme-point propagation would have resulted in $n_{tot} = 100$ samples for computing the p-box of the lift coefficient and as many for the moment coefficient. In the case of the challenge, however, it is observed that the extreme points for the lift and moment coefficients are exact reverses of one another, e.g., the input combination which produces the left bound of the lift coefficient also produces the right bound of the moment coefficient. Additionally, because the p-boxes result from discretisation of distributions all interior vertices of the focal elements are repeated and do not need to be evaluated twice. Thus the total number of evaluations for the computation of the final output p-boxes is $n_{tot} = 92$ points. These p-boxes are shown in Fig. 5. It can be seen there that the excursion probability for the lift coefficient is bounded at $\underline{p}^e_{c_l} < 0.01$ and $\overline{p}^e_{c_l} > 0.995$. A similar excursion trend, although with slightly larger probabilities is observed for the moment coefficient with $\underline{p}^e_{c_m} < 0.1$ and $\overline{p}^e_{c_m} > 0.99$.

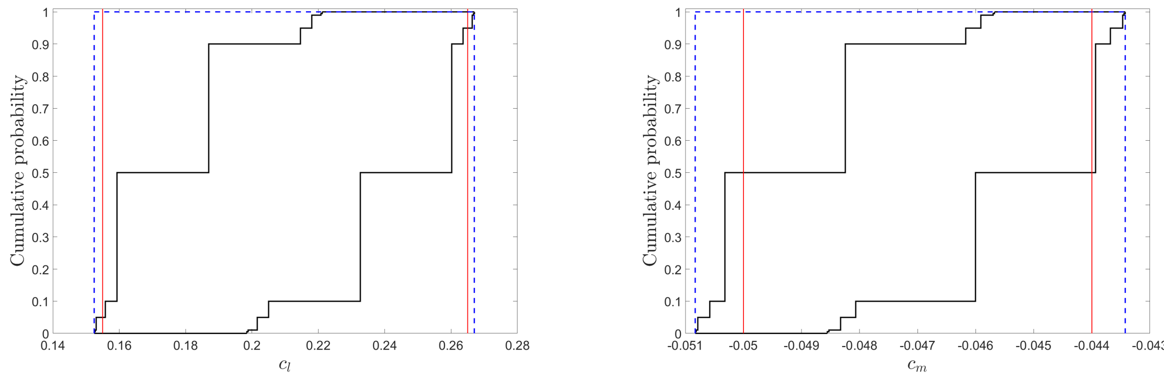
Given the results from the aleatory uncertainty propagation, and assuming the choice of input distributions is justified, the excursion levels for both $c_l$ and $c_m$ are within a tolerable 5% limit. The airfoil can thus be approved for further testing, perhaps under the condition that a notice to for the possible deviations is issued to test pilots.

### C. Propagation of alternative input uncertainty characterisations

As outlined in the introduction to this section, the challenge asks the participants to repeat the aleatory propagation with inputs characterised by uniform distributions with support over the epistemic intervals. The p-boxes for $c_l$ and $c_m$ are computed using the strategies presented in Section III.B. They are graphically presented in Fig. 6. A small increase in the excursion probability for $c_l$ is observed under the assumption for a uniform input to $\underline{p}^e_{c_l} < 0.1$ and $\overline{p}^e_{c_l} > 0.95$. A much more substantial increase is observed for the excursion probability of $c_m$ to $\underline{p}^e_{c_m} < 0.9$ and $\overline{p}^e_{c_m} > 0.5$, which effectively means that, given the current discretisation, a violation of the safe values is expected. However, the important outcome of this analysis is the difference between the results from the epistemic propagation, carried out in Section III.A and the uniform propagation presented in this section. The results are entirely different. In fact, the only similarity is the fact that the range of the interval is the same as the support of the p-box. This is important, because, as discussed in Section I, very often epistemic uncertainty is represented using uniform distributions. The results form this analysis are another evidence that such a treatment leads to severely skewed results.

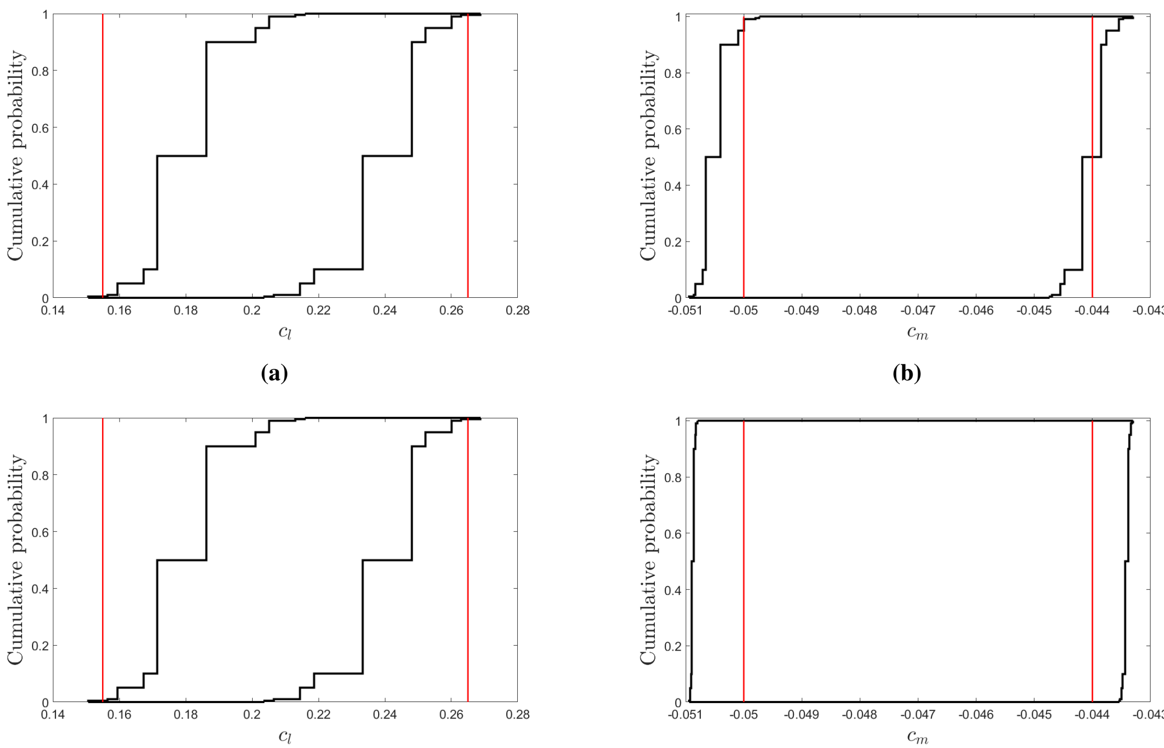
Under the new information, and assuming the choice of uniform input distributions is justified, the excursion levels for $c_m$ are unacceptably high, making the airfoil unfit for use in test flights.

The final projection sub-problem of the challenge calls for quantifying the uncertainty in the output coefficients under two different scenarios of mixed uncertainty. Because of the first-order condensation scheme used during the propagation of the original distributions, the required analysis had already been conducted as a byproduct of the work presented in Section III.B. What is left as a final step is after computing the $n_b = 5$ outputs for each coefficient is




**(a)** **(b)**

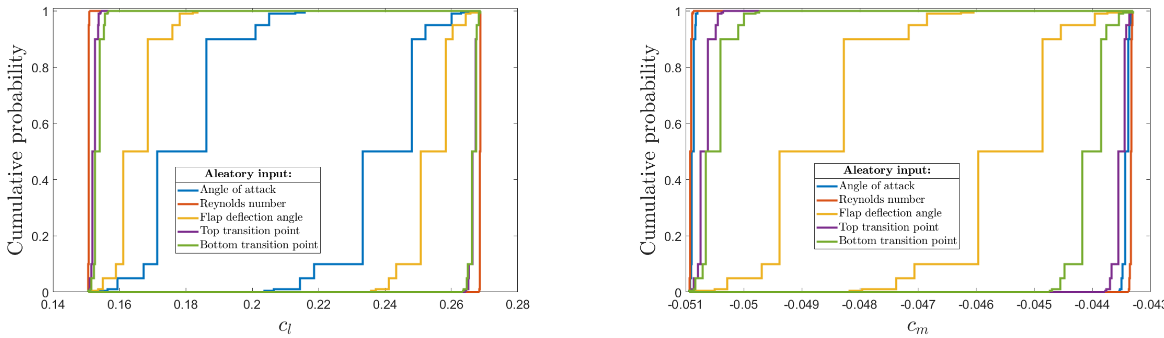
**Fig. 6 Output p-boxes for (a) the lift coefficient and (b) the moment coefficient under the assumption for uniformly distributed input. The epistemic intervals computed in Section III.A are shown as dashed blue lines for comparison.**




**(c)** **(d)**

**Fig. 7 Output p-boxes for the lift coefficient (left column) and the moment coefficient (right column) for the epistemic flap deflection (top row) and epistemic transition points (bottom row).**

to intersect the correct p-boxes. In the case of the epistemic flap deflection $\delta_f$ all p-boxes, except the one produced by propagating $\delta_f$ should be intersected. In the case of epistemic $\delta_f$, $x_{tt}$, and $x_{tb}$, only the p-boxes resulting from the propagation of $\alpha$ and $Re$ should be intersected. The resulting p-boxes are shown in Fig. 7(a) and Fig. 7(b) for epistemic flap deflection and in Fig. 7(c) and Fig. 7(d) for the epistemic transition points. This analysis presents a sort of a sensitivity study for the inputs. The results for $c_l$ confirm the findings from Section III.A that $\alpha$ is by far the




**(a)** **(b)**

**Fig. 8 Output p-boxes for (a) the lift coefficient and (b) the moment coefficient as the p-boxes are gradually replaced with intervals.**

**Table 3 Summary of the excursion probability levels, $\underline{p}^e_{c_x}$ and $\overline{p}^e_{c_x}$ for all propagation sub-problems. The final $c_m$ exhibits certain violation.**

| Excursion probability | All normal | All uniform | Epistemic $\delta_f$ | Epistemic $\delta_f, x_{tt}, x_{tb}$ |
|---|---|---|---|---|
| $\underline{p}^e_{c_l}$ | $< 0.01$ | $< 0.1$ | $< 0.01$ | $< 0.01$ |
| $\overline{p}^e_{c_l}$ | $> 0.995$ | $> 0.95$ | $> 0.99$ | $> 0.99$ |
| $\underline{p}^e_{c_m}$ | $< 0.1$ | $< 0.9$ | $< 0.995$ | $< 1$ |
| $\overline{p}^e_{c_m}$ | $> 0.99$ | $> 0.5$ | $> 0.9$ | $> 0$ |

most important input as there is no change in the p-box for $c_l$ from all p-box inputs to the flap deflection and transition points being fully epistemic. The same thing cannot be said for $c_m$, where $\delta_f$ has a great impact on the width of the p-box and so do $x_{tt}$ and $x_{tb}$, as expected. Because intersection is not an expensive operation and does not require any additional model evaluations, Fig. 8 additionally presents the results for all inputs being gradually changed from p-boxes to intervals.

For all of these results and assuming the choice of distributions for the aleatory inputs is justified, the epistemic uncertainty around the flap angle generates unacceptably high excursion levels for $c_m$. Ascribing epistemic uncertainty to the two transition points does not alter the results, as the flap angle is the strongest uncertainty driver for $c_m$ (as expected). The excursion probabilities for all aleatory sub-problems are presented in Table 3

## IV. Quantifying numerical uncertainty

The challenge asks respondents to assess the inflation of uncertainty due to the use of a low number of panels in the XFOIL code. All results thus far were generated using 256 panels as instructed by the challenge authors, whereas the numerical uncertainty should be evaluated using 100-panel solutions for the bulk of the analyses. Representative convergence traces are shown in Fig. 9. It can be seen that solutions with the required 100 panels are very far from convergence and in fact are separated by any such hypothetical regime by the local solution peak observed at 256 panels. Furthermore, it can be seen in Fig. 9 that simulations seem to never converge within the range of available discretisations, as evidenced by the non-negligible slope at the 2048-panel solution. The convergence traces for all 32 vertex points exhibit very similar behaviour. All of this hinders the use of established verification techniques, such as grid convergence indices [42] and robust verificaiton [43]. Several different regression and interpolation methods, based on splines were also attempted with rather poor results.

Instead a heuristic, obserevation-based approach is proposed in this paper, as follows. From Fig. 9 it seems that despite the fact that simulations do not converge, their variations are confined to within an interval (red dashed lines in Fig. 9) after a certain number of panels. How to determine this number will be discussed shortly. It must be noted here, that this interval differs from the intervals used so far in this paper in that it comes with no guarantees about it

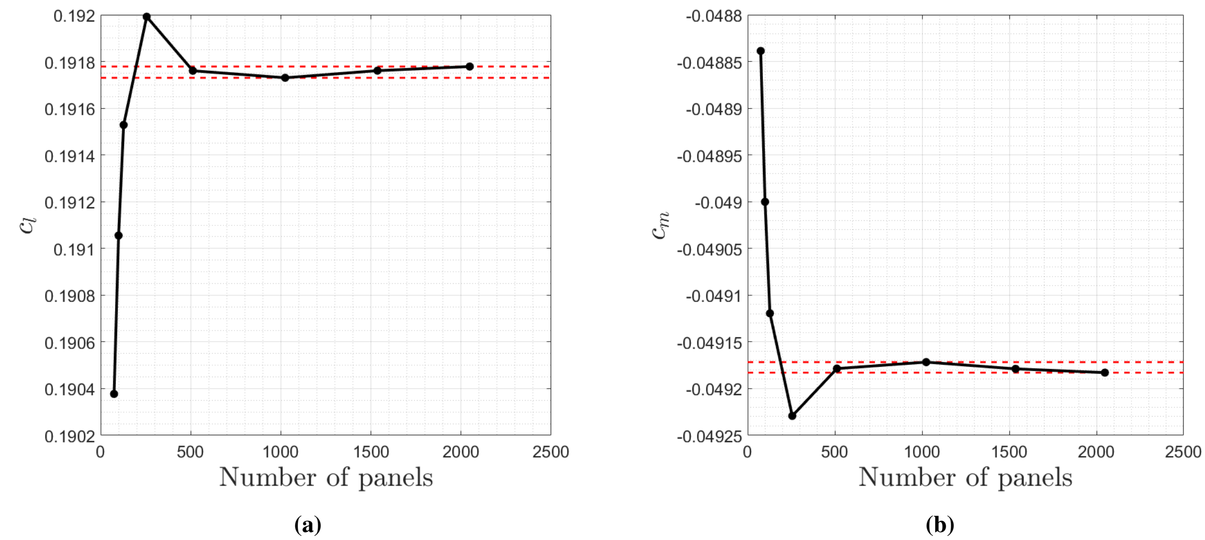



**Fig. 9 Convergence traces for a single input combination. Solutions with the 256 panels prduce extreme values for both $c_l$ and $c_m$.**

containing the true answer of the converged simulations. Knowing the quasi-linear nature of the computer code, two linear regression models, one for the lower bound of the confining interval and one for its upper bound can be trained on a subset of the vertex solutions, in this case 20 points. These surrogate models can be used to map the confining interval throughout the input space and will be used in propagating the normal p-boxes. This will mean that the simulation will have to be run to the maximum resolution only at a handful of input combinations. Once the intervals are computed everywhere the distance between them and 100 panels solutions, computed without any surrogate can be calculated and used as a proxy for the numerical uncertainty.

To choose the point at which the confining interval begins, the normalised relative gradient of the convergence trace can be calculated as

$$D_{rel}(c_x) = \left| \left( \frac{\Delta n_p}{\Delta c_x} \right)_{2048} \frac{\Delta c_x}{\Delta n_p} \right| \tag{5}$$

where $n_p$ is the number of panels and 2048 is the maximum $n_p$. All points below $D_{rel}(c_x) = 10$ belong to the confining interval. Note that this criterion is entirely arbitrary and is only based on observation that solutions with $D_rel$ of the same order of magnitude seem to be the ones that drive the confining interval.

The p-boxes for $c_l$ and $c_m$ with the respective estimation of numerical uncertainty are shown in Fig. 10 and Fig. 11. In each figure, the black lines show the 100 panel solutions and the green lines show the distance between this and the 95% prediction confidence interval for the confining convergence intervals. In neither case can the airfoil be vetted for testing, as excursions occur with probability 1. If on the other hand the 95% prediction confidence intervals for the confining convergence intervals are taken as a measure for numerical uncertainty then the conclusions do not greatly differ from those based on the original results in Section III.B. It is the opinion of the authors that the second, more favourable approach to the quantification of numerical uncertainty represents it more accurately. However, such an approach does not seem to fit with the challenge, as it does not use any 100-panel solutions.

The recommendation to the decision-maker is thus to clarify what is it that the convergence results are telling them and to inspect the model itself before drawing any conclusions from the study.

# V. Conclusions

This paper presented a trans-probabilistic approach to the 2022 AIAA Uncertainty Quantification Challenge Problem for Aerodynamics. The approaches discussed in this paper emphasised utilising the available knowledge about the problem and minimising assumptions whenever possible. It was shown that adopting this approach, uncertainty quantification can be performed efficiently and rigorously. This challenge presents an idealised problem, which however

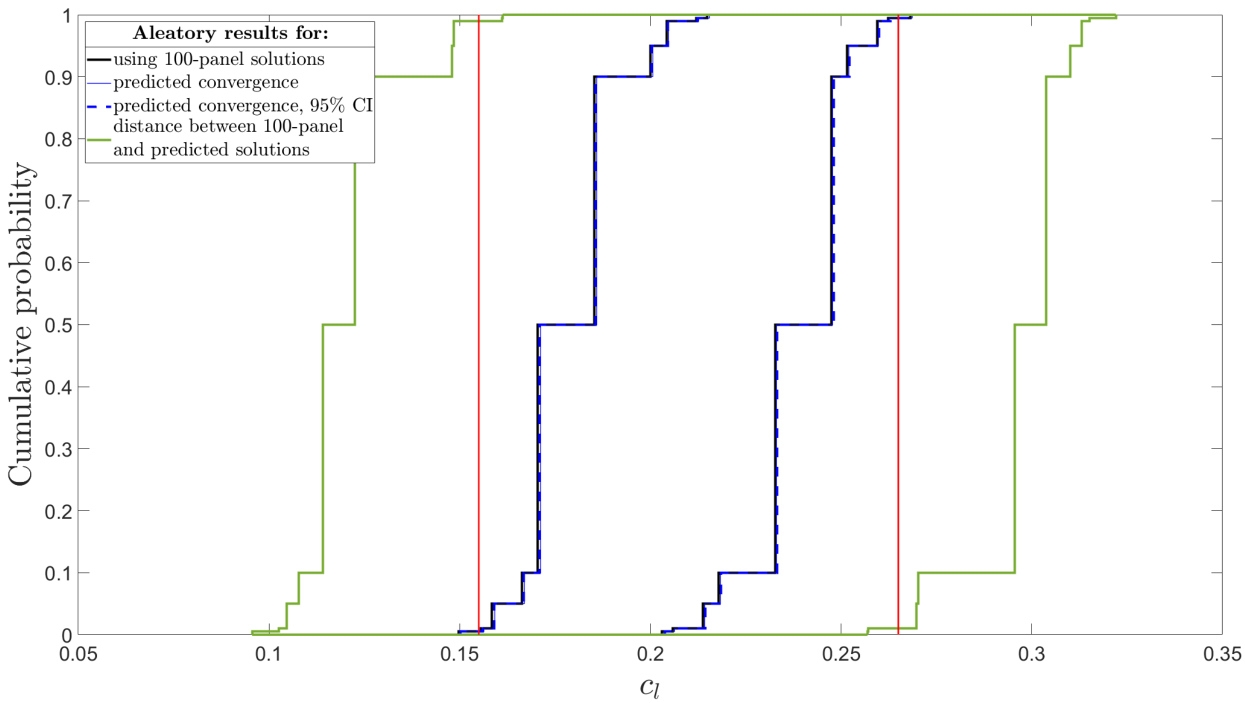



**Fig. 10 Numerical uncertainty estimation for the $c_l$ p-box propagation sub-problem.**

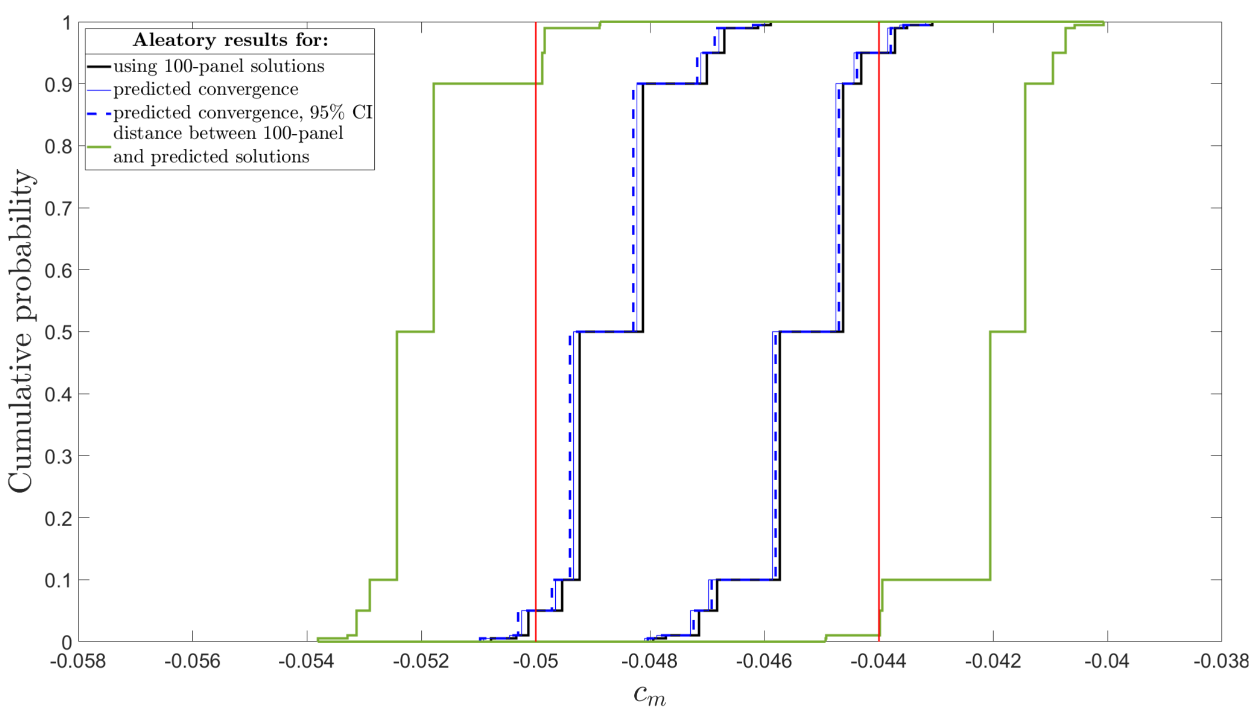



**Fig. 11 Numerical uncertainty estimation for the $c_m$ p-box propagation sub-problem.**

served to demonstrate the value of prior expert information and careful consideration, which seem to have been diminished in recent advances in UQ.

Some directions for future work include further optimisation in the cost-effectiveness of methods for black-box propagation, as well as the development of more robust methods for the quantification of numerical uncertainty in situations which are far from the mathematical limits of established theory.

## Acknowledgments

Peter O. Hristov is funded by the Bulgarian National Scientific Fund program “Petar Beron i NIE” under the AUDiT project, no. KP06-DB/3. Part of this research is also funded under the Development of Advanced Wing Solutions (DAWS) project, supported by the Aerospace Technology Institute (ATI) Programme, a joint government and industry investment to maintain and grow the UK’s competitive position in civil aerospace design and manufacture.

## References

[1] Cary, A. W., Schaefer, J. A., Duque, E. P. N., Khurana, M. S., and DeCarlo, E. C., “Overview of Fluid Dynamics Uncertainty Quantification Challenge Problem and Results,” *AIAA SciTech Forum*, 2024.

[2] Padfield, G., van’t Hoff S, and L, L., “Preliminary guidelines for the rotorcraft certification by simulation process: update no. 1, March 2023,” Tech. rep., Politecnico di Milano, 2023.

[3] Berleant, D., and Goodman-Strauss, C., “Bounding the Results of Arithmetic Operations on Random Variables of Unknown Dependency Using Intervals,” *Reliable Computing*, Vol. 4, 1998, pp. 147–165.

[4] Ferson, S., Oberkampf, W. L., and Ginzburg, L., “Model validation and predictive capability for the thermal challenge problem,” *Computer Methods in Applied Mechanics and Engineering*, Vol. 197, 2008, pp. 2408–2430.

[5] Gray, A., Wimbush, A., de Angelis, M., Hristov, P., Calleja, D., Miralles-Dolz, E., and Rocchetta, R., “From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information,” *Mechanical Systems and Signal Processing*, Vol. 165, 2022, p. 108210.

[6] Cary, A. W., Schaefer, J., Duque, E. P., Khurana, M., and DeCarlo, E. C., “Overview of Challenges in Performing Uncertainty Quantification for Fluids Engineering Problems,” *AIAA SciTech Forum*, 2022, p. 2357.

[7] Drela, M., “XFOIL: An Analysis and Design System for Low Reynolds Number Airfoils,” *Low Reynolds Number Aerodynamics*, edited by T. J. Mueller, Springer Berlin Heidelberg, Berlin, Heidelberg, 1989, pp. 1–12.

[8] Oberkampf, W. L., and Roy, C. J., *Verification and Validation in Scientific Computing*, Cambridge University Press, 2010.

[9] Ferson, S., and Tucker, T. W., “Sensitivity in risk analyses with uncertain numbers, SAND2006-2801,” Tech. rep., Sandia National Laboratories (SNL), Albuquerque, (NM), United States, 2006.

[10] Borgonovo, E., *Sensitivity Analysis An Introduction for the Management Scientist*, Springer, 2017.

[11] Kennedy, M. C., and O’Hagan, A., “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425–464.

[12] Bedford, T., and Cooke, R., *Probabilistic Risk Analysis: Foundations and Methods*, Cambridge University Press, 2001.

[13] Morgan, M., Henrion, M., and Small, M., *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*, University of Cambridge, 1990.

[14] Rowe, W. D., “Understanding Uncertainty,” *Risk Analysis*, Vol. 14, No. 5, 1994, pp. 743–750.

[15] O’Hagan, T., “Dicing with the unknown,” *Significance*, Vol. 1, No. 3, 2004, pp. 132–133.

[16] Der Kiureghian, A., and Ditlevsen, O., “Aleatory or epistemic? Does it matter?” *Structural Safety*, Vol. 31, No. 2, 2009, pp. 105–112.

[17] Ferson, S., and Ginzburg, L. R., “Different methods are needed to propagate ignorance and variability,” *Reliability Engineering and System Safety*, Vol. 54, No. 2, 1996, pp. 133–144.

[18] O'Hagan, A., and Oakley, J. E., "Probability is perfect, but we can't elicit it perfectly," *Reliability Engineering & System Safety*, Vol. 85, No. 1, 2004, pp. 239–248.

[19] Bianchi, D. D., Sêcco, N., and Silvestre, F., "A framework for enhanced decision-making in aircraft conceptual design optimisation under uncertainty," *The Aeronautical Journal*, Vol. 125, No. 1287, 2021, p. 777–806.

[20] Bashkirov, I., Chernyshev, S., and Veresnikov, G., "Parametric synthesis optimization models for high speed transport aerodynamic design to comply with flight safety and low environmental impact requirements," *Acta Astronautica*, 2022.

[21] Duca, R., Sarojini, D., Bloemer, S., Chakraborty, I., Briceno, S. I., and Mavris, D. N., "Effects of Epistemic Uncertainty on Empennage Loads During Dynamic Maneuvers," *2018 AIAA Aerospace Sciences Meeting*, 2018.

[22] Phillips, B. D., and West, T. K., "Aeroelastic Uncertainty Quantification of a Low-Boom Aircraft Configuration," *2018 AIAA Aerospace Sciences Meeting*, 2018.

[23] Molina-Cristobal, A., Nunez, M., Guenov, M., Laudan, T., and Druot, T., "Black-box model epistemic uncertainty at early design stage. An aircraft power-plant integration case study," *29th Congress of the International Council of the Aeronautical Sciences, ICAS*, 2014.

[24] Zaman, K., and Mahadevan, S., "Robustness-Based Design Optimization of Multidisciplinary System Under Epistemic Uncertainty," *AIAA Journal*, Vol. 51, No. 5, 2013, pp. 1021–1031.

[25] Moore, R. E., *Methods and applications of interval analysis*, SIAM, 1979.

[26] Williamson, R., "Probabilistic Arithmetic," Ph.D. thesis, University of Queensland, 8 1989.

[27] Williamson, R. C., and Downs, T., "Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds," *International Journal of Approximate Reasoning*, Vol. 4, No. 2, 1990, pp. 89–158.

[28] Moore, R. E., Kearfott, R. B., and Cloud, M. J., *Introduction to Interval Analysis*, Society for Industrial and Applied Mathematics, 2009.

[29] Jaulin, L., Kieffer, M., Didrit, O., and Walter, E., *Applied Interval Analysis: With Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer London, 2012.

[30] Ferson, S., Kreinovich, V., Ginzburg, L., and Sentz, K., "Constructing Probability Boxes and Dempster-Shafer Structures, SAND2002-4015," Tech. rep., Sandia National Laboratories (SNL), Albuquerque, (NM), United States, 2003.

[31] Ferson, S., and Hajagos, J. G., "Arithmetic with uncertain numbers: rigorous and (often) best possible answers," *Reliability Engineering & System Safety*, Vol. 85, No. 1, 2004, pp. 135–152.

[32] *Impact of Epistemic Uncertainty on Performance Parameters of Compressor Blades*, Turbo Expo: Power for Land, Sea, and Air, Vol. Volume 10D: Turbomachinery — Multidisciplinary Design Approaches, Optimization, and Uncertainty Quantification; Turbomachinery General Interest; Unsteady Flows in Turbomachinery, 2022.

[33] Ioannou, I., Hristov, P. O., Yong, H. K., Marsh, R., Silva, E., Sobester, A., and Ferson, S., "Towards a Framework for Non-intrusive Uncertainty Propagation in the Preliminary Design of Aircraft Systems," *AIAA SciTech Forum*, 2023, p. 2373.

[34] Nelder, J. A., and Mead, R., "A Simplex Method for Function Minimization," *Comput. J.*, Vol. 7, 1965, pp. 308–313.

[35] Słowik, A., and Kwasnicka, H., "Evolutionary algorithms and their applications to engineering problems," *Neural Computing and Applications*, 2020, pp. 1–17.

[36] Cicirello, A., and Langley, R. S., "Vibro-Acoustic Response of Engineering Structures With Mixed Type of Probabilistic and Nonprobabilistic Uncertainty Models," *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, Vol. 1, 2015.

[37] Dong, W., and Shah, H. C., "Vertex method for computing functions of fuzzy variables," *Fuzzy Sets and Systems*, Vol. 24, 1987, pp. 65–78.

[38] McKay, M. D., Beckman, R. J., and Conover, W. J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics*, Vol. 21, 1979, pp. 239–245.

[39] Zuev, K. M., "Subset Simulation Method for Rare Event Estimation: An Introduction," , 2015.

[40] Anderson, R., *Determination of the Characteristics of Tapered Wings*, NACA R-572, NACA, 1936.

[41] Abbott, I., and von Doenhoff, A., *Theory of Wing Sections: Including a Summary of Airfoil Data*, Dover Books on Aeronautical Engineering, Dover Publications, 1959.

[42] Roache, P., *Fundamentals of verification and validation*, Hermosa publ., 2009.

[43] Rider, W., Witkowski, W., Kamm, J. R., and Wildey, T., "Robust verification analysis," *Journal of Computational Physics*, Vol. 307, 2016, pp. 146–163.