Data fusion for Uncertainty Quantification with Non-Intrusive Polynomial Chaos

Nick Pepper[1], Francesco Montomoli[1], and Sanjiv Sharma[2]

[1] *UQlab, Dept of Aeronautics, Imperial College of London*
[2] *Airbus, United Kingdom*

Abstract

This work presents a framework for updating an estimate of a probability distribution, arising from an uncertainty propagation using Non-intrusive Polynomial Chaos (NIPC), with scarce experimental measurements of a Quantity of Interest (QoI). In recent years much work has been directed towards developing methods of combining models of different accuracies in order to propagate uncertainty, but the problem of improving uncertainty propagations by considering evidence from both computational models and experiments has received less attention.

The framework described here uses the Maximum Entropy Principle (MEP) to find an updated, least biased estimate of a probability distribution by maximising the entropy between the original and updated estimates. A constrained optimisation is performed to find the coefficients of a Polynomial Chaos Expansion (PCE) that minimise the Kullback-Leibler (KL) divergence between estimates, while ensuring that the new estimate conforms to constraints imposed by the available experimental measurements of the QoI. In this work a novel constraint is used, based upon the Dvoretzky-Kiefer-Wolfowitz inequality and the Massart bound (DKWM), as opposed to the more commonly used moment-based constraints. Such a constraint allows scarce experimental data to be used in informing the updated estimate of the probability distribution.

Keywords: `Data fusion, Uncertainty Quantification, Polynomial Chaos Expansions, Maximum Entropy Principle`

1. Introduction

Engineers must frequently produce designs that can yield good performance in a wide variety of operating conditions. For this reason, Uncertainty Quantification (UQ) techniques are used to evaluate the performance of candidate designs. To save both time and expense the performance of these candidate designs are typically tested using complex computational models. Non-intrusive Polynomial Chaos (NIPC) is a popular technique for propagating the effect of aleatoric uncertainties through complex computational models. NIPC has been shown to yield results that are consistent with the more established Monte Carlo (MC) sampling techniques but at a much reduced computational cost, with NIPC requiring many times fewer model evaluations than traditional MC methods (see, e.g. [1]).

In recent years there has been great interest within the field of UQ in using multi-fidelity techniques to carry out uncertainty propagations. The accuracy of uncertainty propagations are fundamentally limited by the accuracy of the computational model that is used, however more accurate models are usually very computationally expensive. For instance, in fluid dynamics a Direct Numerical Simulation (DNS) can give a highly accurate resolution of a flow field but usually takes several days to run, making the repeated model evaluations necessary for UQ impractical. On the other hand, Reynolds-Averaged Navier-Stokes (RANS) simulations are less accurate due to assumptions made in the Reynolds stress closure but also far less computationally expensive and more suitable to repeated evaluations. This has the led to the development of Machine learning [2] and Bayesian [3] approaches which use RANS to leverage the results of DNS simulations. A bi-fidelity approach for combining RANS models of different fidelities was introduced by Doostan et al., in which the low-fidelity model is used to learn a reduced basis, which is used together with an interpolation rule and a limited number of high fidelity evaluations to approximate the high fidelity solution [4]. A multi-fidelity extension to NIPC was introduced in Eldred et al. [5]. Methods such as Multi-level Monte Carlo [6, 7, 8], co-kriging [9] and other Gaussian processes [10, 11, 12] are used widely in the UQ community to leverage the results of a small number of high-fidelity simulations with many low fidelity model evaluations. A more comprehensive review of Multi-fidelity methods in Uncertainty Quantification may be found in the review paper of Peherstorfer et al. [13].

However, while there has been great interest in developing methods that allow models of different fidelities to be combined, there has been relatively little attention paid to combining the results of uncertainty propagations with experimental data. This work introduces a framework by which the results of an uncertainty propagation using a computational model may be adjusted based on the evidence of a limited set of experimental measurements of a Quantity of Interest (QoI). An example of where this problem is relevant is in the design process of composite materials, where thousands of coupon tests are required in order to authorise the use of a new composite material. The material properties of composite materials can vary significantly due to defects introduced in the manufacturing process [14], thus materials are typically characterised by statistical quantities such as the b-basis [15]. A current topic of research is how computational simulations of composite materials can be used to leverage the results of experimental coupon tests.

It should be noted that while relatively under researched in the Engineering sciences, the problem of combining evidence from different sources has been treated in other disciplines. For instance, Dempster-Shafer theory has been developed to combine evidence from multiple sources to reach a subjective judgement [16, 17]. It has been applied in software failure detection [18] and in Machine learning, where the accuracy of a decision is improved by combining the evidence of multiple classifiers [19]. In econometrics and weather forecasting density forecasts from different experts are frequently combined directly [20, 21]. The earliest and most popular attempts at forecast combination has been the 'Linear Opinion Pool', in which the resultant forecast is simply the weighted sum of the individual forecasts [22]. For more complete reviews of the probability distribution combining literature see, e.g. [20, 23, 24, 25]. A drawback of the Linear Opinion Pool, shown by Hora [26], is that if each predictive distribution is calibrated then any non-trivial linear combination will itself be un-calibrated [27, 28]. To this end, methods of non-linear combination have been developed in recent years. Gneiting and Ranjan introduce the Beta calibrated linear pools, in which the beta transform of the traditional linear pool is used to find the resultant distribution [29]. Bassetti et al. placed the Beta calibrated linear pool in a Bayesian framework [30] while Casarin et al. compared the performance of linear, harmonic and logarithmic opinion pools in a Bayesian beta mixture model but found the schemes to be substantially equivalent [31]. Kapetanios et al. introduced Generalised density forecast combinations, in which the weight functions of the Linear Opinion Pool are allowed to vary spatially [32]. One of these schemes could well be an appropriate way to directly combine experimental and computational probability distributions of a QoI, but this would assume that there was firstly a sufficient number of experimental results to estimate a probability distribution with sufficient confidence and secondly that some limitation in the design of the experiment meant that it too was an abstract representation of the system being designed, in the same way that a computational model is an abstract or idealised version of a system. In this work it is assumed that the experimental data is too scarce to meaningfully estimate a probability distribution when taken in isolation, but that the experiment measures the QoI to a higher degree of accuracy than the computational model. Additionally, we assume that there is insufficient related evidence, for instance comparisons of experimental and simulation results of previous test cases, to formulate belief functions that could be combined according to Dempster-Shafer Theory.

A number of works in the literature that have addressed the problem of updating a probability distribution based on additional sources of data have done so using the Maximum Entropy Principle (MEP). First developed by Jaynes for estimating a canonical ensemble distribution constrained by particle number and total energy [33, 34], the MEP states that the least biased next estimate of a probability distribution is the distribution that maximises the relative entropy between the old and new estimate of the distribution, while at the same time conforming to a constraint imposed by the experimental data. Typically, the maximum entropy approach involves finding a set of Lagrange multipliers to reweight the initial estimate of the distribution and the imposed constraint is usually the experimental mean. This is the approach adopted in Cesari et al. to combine Molecular Dynamics simulations with experimental results [35] and also in Lou and Cukier [36]. The Maximum entropy approach was expanded in another paper by Cesari et al. to allow experimental errors to be included in the formulation [37]. Such an approach has a number of limitations. Firstly, the Lagrange multiplier formulation makes it very difficult to change the modality of the new estimate. Secondly, the use of moment based constraints is inappropriate for a limited set of experimental results as the statistical moments are unlikely to have converged, thus the approach taken in these works would force the new estimate to converge to a set of statistical moments which are unlikely to reflect the 'true' distribution. The limitations of moment based constraints are also discussed in Baggenstoss, where the joint distribution of a set of scalar measurements is used as a constraint, as opposed to the average values of each measurement [38]. In this work a p-box based constraint is used to update the results of an uncertainty propagation using NIPC. In this way scarce experimental data can be used to inform the estimate of the QoI's probability distribution, without placing constraints which are harsher than the weight of the experimental evidence implies.

2. Non-Intrusive Polynomial Chaos and Uncertainty Quantification

Non-Intrusive Polynomial Chaos (NIPC) is a popular technique used widely in Uncertainty Quantification to propagate aleatoric input uncertainties through computational models (see e.g. [39]). A metamodel for the stochastic response of a model, $w(\vec{\xi})$, is constructed as a function of N independent stochastic input variables $\vec{\xi} = \xi_1, \xi_2, ..., \xi_N$. $w(\vec{\xi})$ is expanded into a basis of orthogonal polynomials to give a spectral representation of uncertainty:

$$w(\vec{\xi}) = \sum_{k=1}^{P} \lambda_k \Psi_k(\vec{\xi}), \tag{1}$$

where Ψ_k refers to a multivariate orthogonal polynomial, given as the product of a set of univariate, orthogonal polynomials ψ. A single realisation of the Polynomial Chaos Expansion (PCE) may be generated by sampling the components of the input random vector $\vec{\xi}$ from probability distributions for the N uncertain system inputs. For a PCE of order p, there are $P = \frac{(N+p)!}{N!p!}$ linear combination terms and the multivariate orthogonal polynomial may be calculated as:

$$\Psi_k(\vec{\xi}) = \prod_{i=1}^{N} \psi^{(i)}_{I^{(i)}_k}(\xi_i), \; k \in [1, P], \tag{2}$$

where the rows of the index matrix I contain the multi-index and the superscript index $\psi^{(i)}$ refers to the input random variable. The multi-index is defined such that $\sum_{i=1}^{N} I^{(i)}_k \leq P$. In the case where $N = 2$ and $p = 1$ the index matrix is denoted:

$$I = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{3}$$

with the third row, for instance, yielding the polynomial product $\Psi_3 = \psi_0^{(1)}\psi_1^{(2)}$. Note that as a product of orthogonal polynomials, and due to the statistical independence of the inputs, Ψ_k is itself orthogonal. There are optimal choices for the univariate orthogonal polynomials that correspond to the probability distributions from which the stochastic input variables are drawn; Xiu and Karniadakis proved that members of the Askey scheme of orthonormal polynomials are optimal for certain types of stochastic processes such as the uniform, normal, and exponential processes [40]. Alternatively, the orthogonal polynomials may be derived directly from the input data. Arbitrary Polynomial Chaos (aPC) uses the statistical moments of the input data to calculate the PCE. This is a useful formulation for engineering purposes as it makes it possible to propagate uncertainties free of assumptions in the case that the input data is scarce [41]. The aPC method used here calculates the orthogonal polynomial for each input by first forming the Hankel matrix of moments:

$$M = \begin{bmatrix} \mu_0 & \mu_1 & \dots & \mu_p \\ \mu_1 & \mu_2 & & \mu_{p+1} \\ \vdots & & \ddots & \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p+1} \end{bmatrix} \tag{4}$$

where μ_i represents the i^{th} statistical moment of the input data. Computing the Cholesky decomposition $M = R^T R$ and calculating the inverse of R produces an upper triangular matrix, the elements of which form an orthogonal system of polynomials according to the Mysovskih theorem. Using the analytical formulas of Rutishauser a three term recurrence can be found for these polynomials in terms of the matrix elements R_{ij}:

$$\psi_{j-1}(\xi) = b_{j-1}\psi_{j-2}(\xi) + a_j\psi_{j-1}(\xi) + b_j\psi_j(\xi), \tag{5}$$

where:

$$a_j = \frac{R_{j,j+1}}{R_j, j} - \frac{R_{j-1,j}}{R_{j-1,j-1}},\ b_j = \frac{R_{j+1,j+1}}{R_{j,j}}, \tag{6}$$

and $R_{0,0} = 1$ and $R_{0,1} = 0$ [42]. The roots of the polynomial of order p may then be found by finding the eigenvalues of the symmetric, tri-diagonal Jacobi matrix:

$$J = \begin{bmatrix} a_1 & b_1 & & & & \\ b_1 & a_2 & b_2 & & & \\ & b_2 & a_3 & b_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & b_{p-2} & a_{p-1} & b_{p-1} \\ & & & & b_{p-1} & a_p \end{bmatrix}. \tag{7}$$

More details on this aPC formulation may be found in [43, 44, 45]. Having determined the orthogonal polynomials, the coefficients of the PCE are found non-intrusively using evaluations of the computational model. These coefficients may be found through the solution of the integral

$$\lambda_k = \frac{1}{\langle \Psi_k^2 \rangle} \int_{\xi \in \Omega} w(\vec{\xi}) \Psi_k(\vec{\xi}) \rho(\vec{\xi}) d\vec{\xi}, \tag{8}$$

using either Galerkin projection, collocation or numerical integration [46] . In the above integral $\rho(\vec{\xi}) = \prod_{i=1}^{N} \rho_i(\xi_i)$ represents the tensor product form of the joint probability density function, while the denominator, the squared norm of the multivariate orthogonal polynomial, is defined as the product of the square norms of the univariate orthogonal polynomials:

$$\langle \Psi_k^2 \rangle = \prod_{i=1}^{N} \langle \psi^{(i)}_{I_k^{(i)}}, \psi^{(i)}_{I_k^{(i)}} \rangle. \tag{9}$$

As noted in Eldred (2009) there are simple, closed form expressions for the univariate inner products if the univariate polynomials belong to the Askey scheme [5, 47]. Alternatively, the coefficients may be found through a linear regression; the computational model is evaluated at each of N_{sp} sample points, allowing the coefficients to be found through:

$$\mathbf{\Psi} \hat{\lambda} = Y, \tag{10}$$

where $Y = [w(\vec{\xi_1}), w(\vec{\xi_2}), \ldots, w(\vec{\xi}_{N_{sp}})]^T$ is a vector of function evaluations, $\hat{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_P]^T$, and the matrix $\mathbf{\Psi}$ is defined as:

$$\mathbf{\Psi} = \begin{bmatrix} \Psi_1(\vec{\xi_1}) & \ldots & \Psi_P(\vec{\xi_1}) \\ \vdots & \ddots & \vdots \\ \Psi_1(\vec{\xi}_{N_{sp}}) & \ldots & \Psi_P(\vec{\xi}_{N_{sp}}) \end{bmatrix}. \tag{11}$$

Note that collocation based approaches exploit the orthonormal properties of the multivariate polynomials in their formulation, which is not the case in linear regression. The cardinality of P is such that the number of PCE coefficients grows exponentially with N, the so-called 'curse of dimensionality'. One strategy to mitigate this cost is to use Smolyak's algorithm to generate a sparse grid on which to sample w [43]. Other methods to exploit sparsity in high dimensions include: least angle regression [48], the alternating Least squares scheme [49], compressive sampling [50, 51, 52], and variational inference [53].

Having obtained a PCE, Monte Carlo simulation can be used to generate a dataset of outputs by randomly sampling the N input distributions N_{MC} times to create a dataset $W \in [w(\vec{\xi_1}), w(\vec{\xi_2}), \ldots, w(\vec{\xi}_{N_{MC}})]$. A Probability Density Function (PDF) for the QoI, $f^0(x)$, may then be estimated using a Kernel estimation method to fit a kernel function to the dataset:

$$f^0(x) = \frac{1}{N_{MC} \times h} \sum_{i=1}^{N_{MC}} K\left(\frac{x - W_i}{h}\right). \tag{12}$$

Possible choices for the kernel function K are discussed in Scott and Silverman [54, 55]. In this work Gaussian kernel functions are used, which is a common choice of kernel function [56]. It is widely regarded that the estimation of the bandwidth, h, is far more crucial than the choice of kernel function [57] and as such a number of methods for performing bandwidth estimation are available (for more details see the review by Jones et al. [58]). In this work a strategy for bandwidth estimation based on cross validation using the Kullback-Leibler divergence has been used [59]. The process of using NIPC with collocation to propagate uncertainty from N input distributions through a computational model to formulate a PCE for the QoI, which is subsequently used to estimate a PDF for the QoI $f^0(x)$, is illustrated in Figure 1.

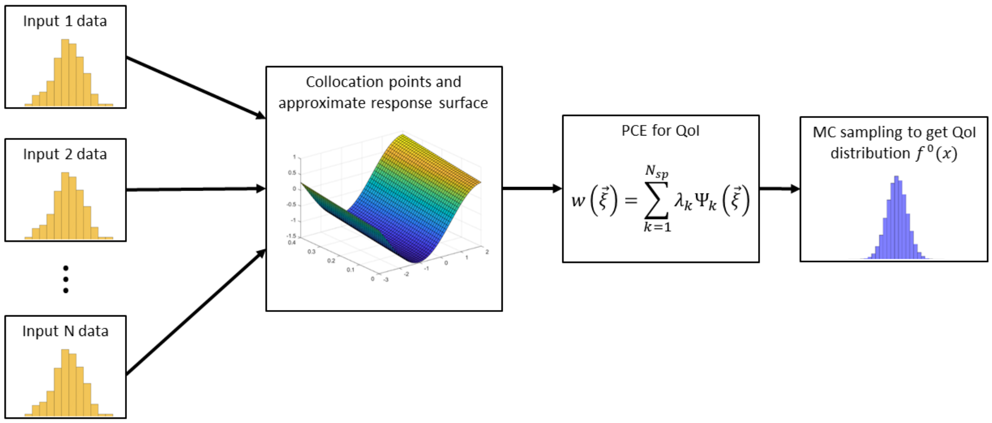

Figure 1: Schematic illustrating a forwards propagation of uncertainty with NIPC. Model evaluations at the collocation points are used to derive the coefficients of the PCE, which approximates the response surface of the model.

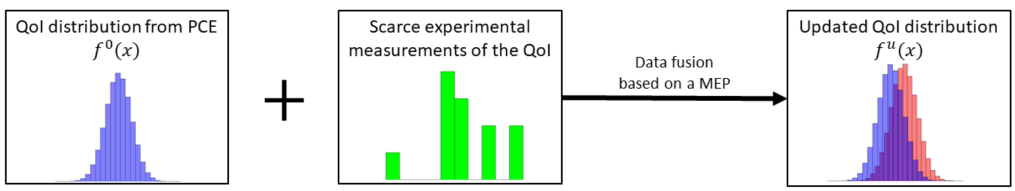

Figure 2: Having determined an initial estimate of the probability distribution for the QoI, scarce experimental measurements of the QoI can be used to update the coefficients of the PCE using a data fusion algorithm based on a MEP. The PCE can be then be resampled to generate an updated probability distribution.

3. Algorithm for data fusion of Polynomial Chaos with experimental results

The algorithm proposed here allows an initial estimate of a probability distribution to be updated through the fusion of an uncertainty propagation with NIPC and the evidence of a scarce experimental dataset. The initial estimate of the QoI probability distribution is arrived at by using an NIPC method to derive a PCE to approximate the response surface of a computational model, $w(\vec{\xi})$:

$$w(\vec{\xi}) = \sum_{k=1}^{P} \lambda_k^0 \Psi_k(\vec{\xi}). \tag{13}$$

A forwards propagation of uncertainty using NIPC is illustrated schematically in Figure 1. In a forwards uncertainty propagation, a probability distribution for the outputs of a system is estimated based on its uncertain inputs. Conversely, an inverse uncertainty propagation in which probability distributions for a systems' inputs are deduced based upon measurements of its uncertain outputs. The algorithm introduced here is general can be applied to a PCE arising from either a forwards or inverse propagation of uncertainty, although only the forwards case is considered here. Having estimated a PCE for the QoI, Monte Carlo evaluations of this PCE are then used to generate an initial estimate of the distribution. The algorithm updates the coefficients of the PCE such that the new estimate of the distribution maximises the cross-entropy between the initial estimate and the new estimate, while at the same time satisfying a constraint on the new estimate imposed by the available experimental data. This is illustrated schematically in Figure 2.

The Kullback-Leibler (KL) divergence is used to quantify the statistical distance between the updated PDF $f^u(x)$ and the initial estimate of the PDF $f^0(x)$. For a continuous set of normalised probability distributions, the Kullback-Leibler divergence is defined as:

$$D_{KL}(f^u||f^0) = \int_{-\infty}^{\infty} f^u(x) \log \frac{f^u(x)}{f^0(x)} dx, \tag{14}$$

where x refers to the space of the QoI [59]. The Kullback-Leibler divergence is defined as the negative of the cross entropy, hence minimising the KL divergence between two probability distributions is the equivalent of maximising the cross-entropy. Minimising the KL divergence between the updated distribution and the initial estimate produces a distribution which introduces the minimum amount of new information with respect to the prior knowledge [36, 37]. The steps of the data fusion algorithm are illustrated in Figure 3. The initial estimate of the PCE coefficients in equation (13), λ_k^0, are then replaced with the updated set of coefficients, λ_k^u, which are the solution of the optimisation problem:

$$\lambda_k^u = \underset{\lambda_k^u}{\mathrm{argmin}} D_{KL}(f^u(x)||f^0(x)) \quad \text{subject to}\, \underline{F}(x) \leq F^u(x) \leq \bar{F}(x). \tag{15}$$

In this paper a p-box placed around the experimental data is used to constrain the Cumulative Density Function (CDF) of the updated distribution $F^u(x)$. It should be noted that the formulation is flexible and can admit a range of user chosen constraints. For instance, moment based constraints could be used such as the experimental mean or the experimental standard deviation. However, given the scarcity of the experimental data the statistical moments of the experimental data are unlikely to have converged to the true distribution, making constraints based on the statistical moments of the experimental data unsuitable. p-boxes themselves are discussed further below. In general, a six sigma criterion is applied in reliability problems (see, e.g. [61]). Montomoli and Massini demonstrated that for aerospace accidents this criterion is not accurate enough and tends to underestimate the likelihood of accidents [62]. Given the scarcity of the experimental data, accurate estimation of the tails of the distribution tails is not possible using existing methods in the literature (see, e.g. [63, 64]). In this paper we focus on a more general case, where constraints are imposed that are likely to only affect the first two or three statistical moments in order to make the experimental data consistent with the probability distribution estimated from the PCE.

The manner in which the PCE coefficients are updated in the algorithm will depend on the optimisation strategy used. The use of an inequality constraint such as a p-box, in addition to the lack of an analytical relationship between the PCE coefficients and $f^u(x)$, precludes the use of Lagrange multipliers, which have been used in other works which use an MEP to update a probability distribution. There are a number of optimisation techniques in the literature with good global search properties such as particle swarm optimisation [60] or genetic algorithms [65] that are appropriate. In this paper an interior point algorithm was used, initialised at a point in PCE space within the feasible region. It was found that the variation of D_{KL} in the P-dimensional PCE parameter space was sufficiently smooth for the algorithm to avoid becoming trapped in a local minima [66, 67].

3.1. p-boxes as a constraint on the optimisation

A p-box is defined by two curves $\bar{F}(x)$ and $\underline{F}(x)$ which act as upper and lower bounds to the updated CDF $F^u(x)$ respectively. The bounds are defined such that the true but unknown CDF lies between $\underline{F}(x)$ and $\bar{F}(x)$, i.e.:

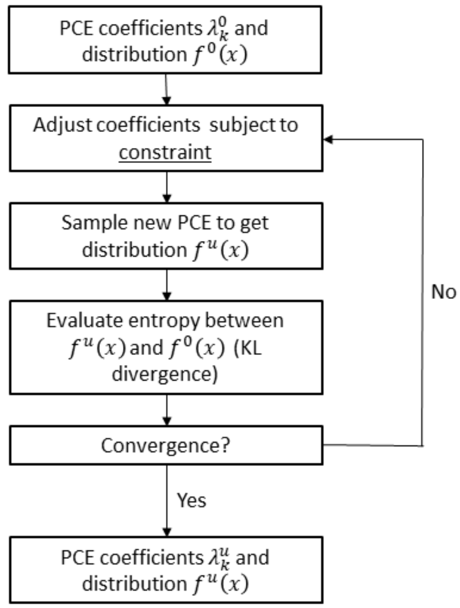

Figure 3: Data fusion algorithm to adjust the coefficients based on an Maximum Entropy Principle, subject to constraints imposed by the available experimental data

$\underline{F}(x) \leq F^{true}(x) \leq \bar{F}(x)\ \forall x$. p-boxes may be non-parametric or parametric. Parametric p-boxes arise from distributions which are known to conform to a particular probability distribution (normal, exponential, Weibull, uniform, etc.) but with parameters which are specified imprecisely as intervals. Non-parametric p-boxes are distribution free and make no assumptions about the form of a distribution and so are a more appropriate choice as a bound when there is insufficient data to specify a family of distributions to which the experimental data for the QoI belongs. In this work a non-parametric technique based upon the Dvoretzky-Kiefer-Wolfowitz inequality and the Massart bound (referred to as the DKWM bound) is used to place a bound around the experimental data for the QoI [68]. Bounds based on the DKWM bound place a confidence interval around the Kolmogorov Smirnov statistic between the unknown, true distribution F^{true} and the distribution for the experimental data F^{exp} [69, 70]:

$$P[\sup_x |F^{exp}(x) - F^{true}(x)| \leq 2e^{-2n\epsilon^2}] = \alpha. \tag{16}$$

With a confidence level α the true CDF does not differ from the experimental CDF by more than the error, ϵ. Rearranging the RHS of equation (16) yields an expression for ϵ given a desired confidence α:

$$\epsilon = \sqrt{\frac{-\log\left(\frac{\alpha}{2}\right)}{2n}}, \tag{17}$$

where n is the number of experimental data points available. Note that the inverse dependence on n means that the bound becomes tighter if more experimental points are available. By translating the experimental CDF vertically by $\pm\epsilon$, a p-box may be created, i.e.:

$$F^{exp}(x) - \epsilon \leq F^{true}(x) \leq F^{exp}(x) + \epsilon. \tag{18}$$

The CDF of the updated distribution, $F^u(x)$ is constrained such that it lies within the p-box. As the p-box is created through a vertical translation of F^{exp} it is defined only over the support of F^{exp} and so does not constrain the tails of $F^u(x)$ directly. As has been discussed, this is considered acceptable as the intention is to place a constraint that affects the first two or three statistical moments of the updated probability distribution. The tails of a distribution are typically determined by higher order statistical moments.

4. Carbon fibre composite test case

The algorithm for the fusion of simulations and scarce experimental measurements of a QoI was tested using a dataset comprised of laboratory measurements of carbon fibre composites and a computational model based on the equations of classical laminate theory. A carbon fibre composite is constructed of layers of unidirectional carbon fibres held together by a resin, referred to as the matrix. A single layer is referred to as a laminae or ply. Note that there are two sets of co-ordinates in a composite, (x, y) co-ordinates refer to global co-ordinates of the laminate as a whole, while the local co-ordinates $(1, 2)$ refer to directions parallel with the fibres and transverse to the fibres in each laminae. A rotation is needed to move from global to local co-ordinates, which will depend on the orientation of the fibres in the laminae. For more details on carbon fibre composites see, e.g. [71, 72, 73]. As has been discussed above, significant variance is observed in the material properties of composite materials when measured in coupon tests due to the complexity of the manufacturing process [14]. Certifying a new composite material can require thousands of coupon tests, which is both expensive and time consuming, providing motivation to replace physical tests with simulations wherever possible. Previous works have attempted to reduce the computational cost of these simulations using non-intrusive polynomial chaos approaches (see e.g. [74, 75, 76]).

The stacking of individual, unidirectional plies to form a composite is illustrated in Figure 4. The figure also illustrates the directions in which the loads N_x and N_{xy} are applied. A rotation of θ degrees must be applied to move from the global co-ordinates (x, y) to the local co-ordinates $(1, 2)$.

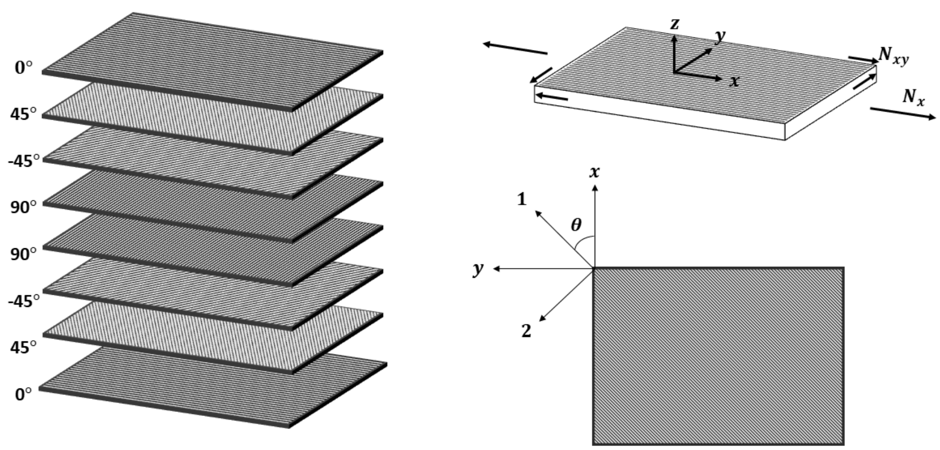

Figure 4: A schematic illustrating how a carbon fibre composite is constructed from layers of unidirectional plies (left). The directions of the global co-ordinates are indicated, together with the directions of the compression load N_x and shear load N_{xy} (top right). The angle θ defines the rotation necessary to move from global co-ordinates (x, y) to the local co-ordinates of the ply $(1, 2)$ (bottom right)

A deterministic model, based on the equations of classical laminate theory, was used as a computational model of the physical coupon tests. The coupons themselves consisted of 48 laminae in a $(([45/90/-45/0]_2)_s)_s$ stacking sequence. The applied stresses N_x and N_{xy} were incremented in ratio until a failure criterion was met at any of the laminae:

1. Initialise N_x and N_{xy}
2. Apply loads N_x and N_{xy} ($N_x = 3.33N_{xy}$)
3. Find the mid surface strains ϵ_x,ϵ_y and γ_{xy}
4. Apply rotations to find the local strains in each ply (ϵ_1,ϵ_2,γ_{12})
5. Evaluate the Yamada-Sun failure criterion $f(f_1, f_{12}, \sigma_1, \tau_{12})$ for each ply
6. If the failure criterion is met at any of the plies then the laminate is considered to have met the first failure stress, otherwise increment the applied load and return to step 2

The relationship between the mid surface strains and the local stresses in a given ply is represented using the reduced stiffness matrix Q, the components of which are determined by the material properties of the laminate:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{22} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \gamma_{12} \end{bmatrix}, \tag{19}$$

where σ_1 is refers to tension/compression in the principal fibre direction, σ_2 is the tension/compression in the transverse direction and τ_{12} the shear stress. The elements of the reduced stiffness matrix are dependent on the material properties of the ply, determined by the following relations:

$$Q_{11} = \frac{E_1}{1-\nu_{12}\nu_{21}},\ Q_{12} = \frac{\nu_{12}E_2}{1-\nu_{12}\nu_{21}} = \frac{\nu_{21}E_1}{1-\nu_{12}\nu_{21}},\ Q_{66} = G_{12}, \tag{20}$$

where E_1 represents the Young's modulus in the fibre direction, E_2 the Young's modulus in the transverse direction, G_{12} the shear modulus, and ν_{12}

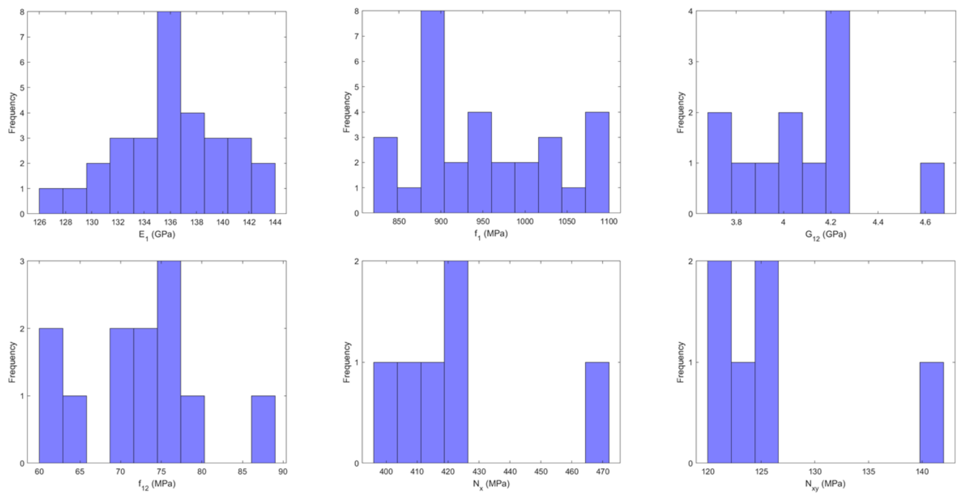

Figure 5: Histograms for the four stochastic material properties in the dataset $(E_1, f_1, G_{12}, f_{12})$ and the compression and shear loadings responsible for first failure of the laminate(N_x and N_{xy}). Note that the data for N_x and N_{xy} is particularly scarce

Poisson's ratio. A linear strain distribution is assumed through the thickness of the laminate:

$$\begin{bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{bmatrix} = \begin{bmatrix} \epsilon_x^0 \\ \epsilon_y^0 \\ \gamma_{xy}^0 \end{bmatrix} + z \begin{bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix}, \tag{21}$$

where ϵ_x^0, ϵ_y^0, and γ_{xy}^0 represent the mid-surface strains and κ_x, κ_y, and κ_{xy} the curvature in the laminate. The applied loads (N_x, N_y, N_{xy}) and moments (M_x,M_y, M_{xy}) are related to the mid-surface strains and curvatures through:

$$\begin{bmatrix} N_x \\ N_y \\ N_{xy} \\ M_x \\ M_y \\ M_{xy} \end{bmatrix} = \begin{bmatrix} [A] & [B] \\ [B] & [D] \end{bmatrix} \begin{bmatrix} \epsilon_x^0 \\ \epsilon_y^0 \\ \gamma_{xy}^0 \\ \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{bmatrix}, \tag{22}$$

where the elements of the ABD matrices depend on a summation across all

n plies in the laminate:

$$A_{ij} = \sum_{k=1}^{n} (\bar{Q}_{ij})_k (z_k - z_{k-1}),\ B_{ij} = \sum_{k=1}^{n} \frac{1}{2} (\bar{Q}_{ij})_k (z_k^2 - z_{k-1}^2),\ D_{ij} = \sum_{k=1}^{n} \frac{1}{3} (\bar{Q}_{ij})_k (z_k^3 - z_{k-1}^3), \tag{23}$$

where $\bar{Q}_{ij}$ represents the ij$^{\text{th}}$ element of the reduced stiffness matrix of the k^{th} ply when it has been rotated to lie in the (x,y) co-ordinate system. It is possible to greatly simplify equation (22) through symmetry considerations, for instance $[B] = [D] = 0$ and $A_{16} = A_{26} = 0$ due to the symmetrical layup of the laminate. In the test case only horizontal and shear loads are applied, meaning there is no curvature due to the applied moments. This allows a further simplification of equation (22):

$$\begin{bmatrix} N_x \\ N_y \\ N_{xy} \end{bmatrix} = \begin{bmatrix} N_x \\ 0 \\ N_{xy} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{12} & A_{22} & 0 \\ 0 & 0 & A_{66} \end{bmatrix} \begin{bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_x y \end{bmatrix} \tag{24}$$

The Yamada Sun failure criterion was used to determine whether failure had occurred in an individual laminae [77]. Equations (19), (20) and (24) were used to relate the applied load to the mid surface strains and in turn the local stresses in each ply. The laminate was considered to have reached first failure when the condition:

$$\left(\frac{\sigma_1}{f_1}\right)^2 + \left(\frac{\tau_{12}}{f_{12}}\right)^2 = 1, \tag{25}$$

was met in any ply. Note that each of these linear transformations is dependent on the material properties of the laminate.

Experimental measurements were available for four material properties of a batch of carbon fibre composite coupons: the in plane longitudinal and shear strengths of the composite (f_1 and f_{12}), the shear modulus (G_{12}), and the Young's modulus in the fibre direction (E_1). The Young's modulus transverse to the fibre direction (E_2) was treated as a constant. The Quantities of Interest (QoI) of the experiment were the first failure stresses of the laminate, N_x and N_{xy}, where N_x refers to the stress in compression and N_{xy} the stress in shear. In addition to experimental measurements of the material properties, the dataset also included a limited number of measurements of the first failure stresses. Histograms of the experimental dataset are displayed in Figure 5.

An aPC formulation was employed to propagate the uncertainties in the material properties to produce probability distributions of the first failure stresses of the laminate, N_x and N_{xy}. As discussed above, an orthogonal set of polynomials was calculated from the statistical moments of the input data. In addition, 4 one-dimensional quadrature rules were derived from the data, from which a sampling grid was formed using a level 1 Smolyak quadrature. The computational model was evaluated at each of the N_{sp} sample points in the Smolyak quadrature and the coefficients of the PCE numerically approximated as:

$$\lambda_k^0 = \frac{\sum_{i=1}^{N_{sp}} w(\vec{\eta}_i)\Psi_k(\vec{\eta}_i)\theta_i}{\sum_{i=1}^{N_{sp}} \Psi_k(\vec{\eta}_i)\theta_i}, \tag{26}$$

where η_i and θ_i represent the sparse grid collocation points and weights [45]. As the compression and shear stresses are increased in ratio, a single PCE can be used to generate both distributions. An initial estimate for the shear load probability distribution, $f^0(N_{xy})$, was generated from 10^6 Monte Carlo realisations of this PCE.

The data fusion algorithm was then used to adjust the coefficients of the PCE based on the evidence of the available experimental measurements of the failure stresses. Given the relative simplicity of the computational model used, we expect the model error to result in a discrepancy between the estimate of the failure stress distribution and the 'true' but unknown distribution, which we assume can only be found with many experimental tests. The aim of the data fusion algorithm is to find improve the initial estimate provided by NIPC, based on the evidence of a small number of experiments.

The results of the uncertainty propagation and subsequent data fusion are displayed in Figure 6. The blue line in the left panel represents the CDF of the initial estimate of the probability distribution from propagating the uncertainties in the material properties with aPC, $F^0(N_{xy})$. The green line indicates the CDF of the 6 available experimental results, $F^{exp}(N_{xy})$. The DKWM bound was used to place a p-box around $F^{exp}(N_{xy})$ with a 95% confidence. An interior point algorithm was used to find a set of updated PCE coefficients which minimised the KL divergence between the initial and updated estimate of the probability distribution of N_{xy}, while at the same time ensuring that the updated probability distribution lies within the bounds set by the p-box. The algorithm was initialised at a point in the PCE parameter space that satisfied the bounds and was found to converge quickly, requiring only 74 function evaluations. The CDF of the updated distribution is represented in red in the left panel of Figure 6. The right panel of Figure 6 illustrates the kernel density estimates of the initial estimate of the probability distribution $f^0(N_{xy})$ and the updated probability distribution $f^u(N_{xy})$. Kernel density estimates were made for $f^0(N_{xy})$ and $f^u(N_{xy})$ based on 10^6 Monte Carlo realisations of the PCE. As can be seen the updated distribution is weighted more heavily to the left, giving a more conservative estimate of the failure strength distribution. The potential of the data fusion algorithm to adjust the modality of the updated distribution is demonstrated, with some peaks in $f^0(N_{xy})$ becoming smoothed out in the updated estimate as a result of the MEP. These peaks likely arise due to the effect of the scarce input data on the uncertainty propagation using aPC.

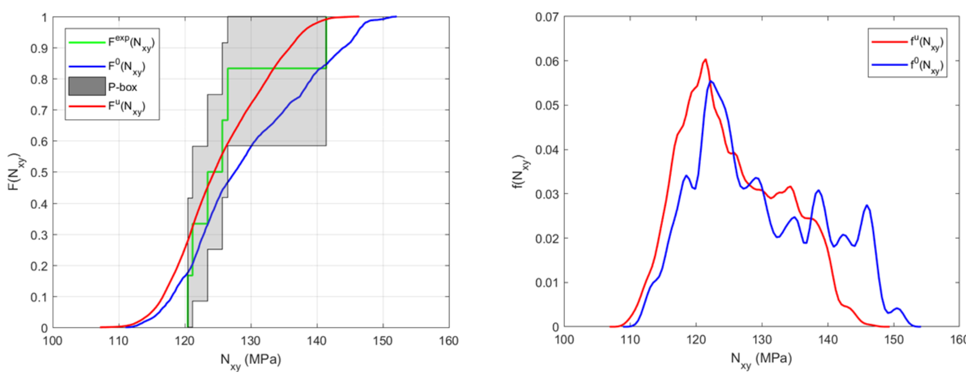

Figure 6: A plot showing the CDFs of the original estimate of the probability distribution arising from the aPC (blue); the available experimental data (green) and the updated estimate of the probability distribution (red). The p-box, generated by the DKWM bound, is shaded in grey (left subplot). Kernel density estimates of the original and updated estimates- note how the maximum entropy estimator has smoothed the peaks in the original estimate arising from the scarce input data (right subplot)

5. Conclusion

In recent years much attention has been paid to developing methods of combining models of different accuracies in order to propagate uncertainty, but relatively little attention has been paid to improving uncertainty propagations by considering evidence from both computational models and experiments. The growing accuracy of computational models offers industries the possibility of reducing the amount of physical testing needed in the design phase of a product by supplementing physical tests with simulations, making this research problem increasingly relevant. Data fusion allows for evidence from multiple sources to be considered in order to estimate a probability distribution. The framework introduced here is intended for cases where a probability distribution must be estimated based on the evidence of simulations, which are relatively plentiful but are considered less reliable than experimental measurements as a result of the epistemic uncertainties present in the models (model error). The experimental data, while more reliable, is expensive or time consuming to obtain and as a consequence there are not enough points to fit a probability distribution to the data without making significant assumptions. The principle of Maximum Entropy is a key aspect of the framework. This assumption, together with the data driven NIPC formulation used to propagate uncertainties, means that the method is non-parametric as no assumptions are made as to the family to which the underlying distribution belongs. This makes the framework well suited to handling scarce data. As can be seen from Figure 6, the Maximum Entropy Principle tends to have a smoothing effect in practice and can eliminate spurious peaks which arise in the original estimate of the probability distribution due to the scarcity of the input data. The use of a p-box as a constraint, as opposed to moment based constraints, is another feature of the framework which allows it to handle scarce data efficiently. Scarce experimental data can be used to inform the updated estimate of the probability distribution, without placing a set of onerous moment based constraints on the estimate which are unlikely to reflect the moments of the true distribution. The use of the DKWM to construct the p-box allows the experimental data to impose a non-parametric constraint, however, it should be noted that this does not preclude the use of a parametric method to construct the p-box should that be considered more appropriate. An extension to the method could be to include estimates of the bounds in the p-box. In this work it was assumed that the experimental data was too scarce for maximum likelihood methods of tail estimation such as Generalized Pareto Tails to be appropriate, however the addition of constraints on the tail of the updated distribution could be an interesting extension of the framework. In this work we did not place too much emphasis on the optimisation aspect of the framework, this is another area that could be explored further in the future.

Funding data

This work was jointly funded by Airbus and the EPSRC through iCASE voucher number 17000099

Acknowledgement

The authors would like to thank Philippe Peters for his helpful comments and for providing the experimental data used in this work

References

[1] S. Hosder, R. W. Walters, and R. Perez. 2006 A Non-Intrusive Polynomial Chaos Method For Uncertainty Propagation in CFD Simulations. In *44th AIAA Aerospace Sciences Meeting and Exhibit*, 1–19

[2] J. L. Wu, H. Xiao, and E. G. Paterson. 2018 Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Phys. Rev. Fluids* **3**.

[3] H. Xiao, J. X. Wang, R. Sun, and C. Roy. 2016 Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier-Stokes simulations: A data-driven, physics-informed bayesian approach. *J. Comput. Phys.* **324**, 115–136.

[4] A. Doostan, G. Geraci, and G. Iaccarino. 2016 A Bi-fidelity for Uncertainty Quantification of Heat Transfer in a Rectangular Ribbed Channel. In *Proceedings of ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition*, 1–9.

[5] M. S. Eldred, .L.W.T Ng, M. F. Barone, and D. S.P. 2017 Multifidelity Uncertainty Quantification Using Spectral Stochastic Discrepancy Models. In *Handbook of Uncertainty Quantification*, R.G., D.H., and O. H. Eds. Springer.

[6] M. Giles. 2015 Multilevel Monte Carlo methods. *Acta Numer.* **24**, 259–328.

[7] J. Charrier, R. Scheichl, and A. L. Teckentrup. 2013 Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and Its Application to Multilevel Monte Carlo Methods. *SIAM Journal on Numerical Analysis* **51**, 322-352.

[8] M. Pisaroni, F. Nobile, and P. Leyland. 2017 A Continuation Multi Level Monte Carlo (C-MLMC) method for uncertainty quantification in compressible inviscid aerodynamics. *Comput. Methods Appl. Mech. Eng.* **326**, 20–50.

[9] M. C. Kennedy and A. O'Hagan. 2001 Bayesian calibration of computer models. *J. Roy. Stat. Soc. Ser. B* **63**, 425–464.

[10] D. Crevillén-García, R. D. Wilkinson, A. A. Shah, and H. Power. 2017 Gaussian process modelling for uncertainty quantification in convectively-enhanced dissolution processes in porous media. *Adv. Water Resour.* **99**, 1–14.

[11] P. T. Roy et al. 2018 Comparison of polynomial chaos and Gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. *Stoch. Env. Res Risk Assess* **32**, 1723–1741.

[12] I. Bilionis and N. Zabaras. 2015 Bayesian Uncertainty Propagation Using Gaussian Processes. In *Handbook of Uncertainty Quantification*, Springer.

[13] B. Peherstorfer, K. Willcox, and M. Gunzburger. 2018 Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. *SIAM Rev.* **60**, 550–591.

[14] K. Potter. 2009 Understanding the origin of defects and variability in composites manufacture. In *Proceedings of the 17th international conference on composite materials.*

[15] O. Vallmajó et al. 2018 Virtual calculation of the B-value allowables of notched composite laminates. *Compos. Struct.* **212**, 11–21.

[16] A.P. Dempster. 1967 Upper and lower probabilities induced by a multi-valued mapping. *The Annals of Mathematical Statistics* **38**, 325–339.

[17] G. Shafer. 1976 A Mathematical Theory of Evidence, Princeton University Press.

[18] A. Paksoy and M. Göktürk. 2011 Information fusion with dempster-shafer evidence theory for software defect prediction, *Procedia Computer Science*, **3** 600-605.

[19] B. Quost, M.H. Masson and T. Denœux. 2011 Classifier fusion in the Dempster–Shafer framework using optimized t-norm based combination rules, *International Journal of Approximate Reasoning*, **52**, 353-374

[20] S. Hall and J. Mitchell. 2007 Combining density forecasts. *Int. J. Forecast.* **23** 1–13.

[21] A. Garratt, K. Lee, H. Pesaran, and Y. Shin. 2003 Forecast Uncertainties in Macroeconomic Modeling: An Application to the U.K. Economy. *J. Am. Stat. Assoc.* **98**, 829–838.

[22] M. Stone. 1961 The Opinion Pool. *Ann. Math. Stat.* **32**, 1339–1342.

[23] C. Genest and J. Zidek. 1986 Combining probability distributions: A critique and an annotated bibliography. *Stat. Sci.* **1**, 114–135.

[24] R. Clemen and R. Winkler. 1999 Combining Probability Distributions From Experts in Risk Analysis. *Risk Anal.* **19**, 187–203.

[25] T. Wallsten, D. Budescu, I. Erev, and A. Diederich. 1997 Evaluating and Combining Subjective Probability Estimates. *J. Behav. Decis. Mak.* **10**, 243–268.

[26] S. Hora. 2004 Probability judgements for continuous quantities: Linear combinations and calibration. *Manage. Sci.* **50**, 597–604.

[27] R. Ranjan and T. Gneiting. 2010 Combining probability forecasts. *J. R. Stat. Soc.* **72**, 71–91.

[28] A. Allard, A. Comunian, and P. Renard. 2012 Probability aggregation methods in geoscience. *Math. Geosci.* **44**, 545–581.

[29] T. Gneiting and R. Ranjan. 2013 Combining predictive distributions. *Electron. J. Stat.* **7**, 1747–1782.

[30] F. Bassetti, R. Casarin, and F. Ravazzolo. 2018 Bayesian Nonparametric Calibration and Combination of Predictive Distributions. *J. Am. Stat. Assoc.* **113**, 675–685.

[31] R. Casarin, G. Mantoan, and F. Ravazzolo. 2016 Bayesian Calibration of Generalized Pools of Predictive Distributions. *Econometrics* **4**.

[32] G. Kapetanios, J. Mitchell, S. Price, and N. Fawcett. 2015 Generalised density forecast combinations. *J. Econom.* **188**, 150–165.

[33] E. Jaynes. 1957 Information Theory and Statistical Mechanics. *Phys. Rev.* **106**, 620–630.

[34] E. Jaynes. 1982 On the Rationale of Maximum-Entropy Methods. *Proc. IEEE* **70**, 939–952.

[35] A. Cesari, S. Reißer, and G. Bussi. 2018 Using the Maximum Entropy Principle to Combine Simulations and Solution Experiments. *Computation* **6**, 1–25.

[36] H. Lou and R. Cukier. 2018 Reweighting ensemble probabilities with experimental histogram data constraints using a maximum entropy principle. *J. Chem. Phys.* **149**.

[37] A. Cesari, A. Gil-Ley, and G. Bussi. 2016 Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J. Chem. Theory Comput.* **12**, 6192–6200.

[38] P. M. Baggenstoss. 2018 Beyond Moments: Extending the Maximum Entropy Principle to Feature Distribution Constraints. *Entropy* **20**.

[39] M. Eldred and J. Burkardt. 2009 Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification. *AIAA*.

[40] D., Xiu and G.E. Karniadakis. 2002 The Wiener Askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, **24**, 619-644.

[41] S. Oladyshkin and W. Nowak. 2012 Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering & System Safety*, **106**, 179-190.

[42] H. Rutishauser. 1963 On a modification of the QD-algorithm with Graeffe-type convergence. *Proc. IFIP Congr.*, **62**, 93–96.

[43] R. Ahlfeld, B. Belkouchi and F. Montomoli. 2016 SAMBA: Sparse approximation of moment-based arbitrary polynomial chaos. *J. Comput. Phys.*, **320**, 1–16.

[44] N. Pepper, F. Montomoli and S. Sharma. 2019 Multiscale Uncertainty Quantification with Arbitrary Polynomial Chaos. *Comput. Methods Appl. Mech. Eng.* **357**.

[45] N. Pepper, L. Gerardo-giorda and F. Montomoli. 2019 Meta-modeling on detailed geography for accurate prediction of invasive alien species dispersal. *Scientific Reports* **9**, 1–10.

[46] A. Kaintura, T. Dhaene and D. Spina. 2018 Review of Polynomial Chaos-Based Methods for Uncertainty Quantification in Modern Integrated Circuits. *Electronics* **7**, 1-21.

[47] Abramowitz, M. and Stegun, I. A.. 1965 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover.

[48] G. Blatman and B. Sudret. 2011 Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comp. Phys.* **230**, 2345-2367.

[49] A. Doostan and G. Iaccarino. 2009 A least-squares approximation of partial differential equations with high-dimensional random inputs. *J. Comp. Phys.* **228**, 4332–4345.

[50] J. Peng, J. Hampton and A. Doostan. 2014 A weighted l1-minimization approach for sparse polynomial chaos expansions. *J. Comp. Phys.* **267**, 92-111.

[51] X. Yang and G.E. Karniadakis. 2013 Reweighted l1 minimization method for stochastic elliptic differential equations. *Journal of Computational Physics* **248**, 87-108.

[52] K. Sargsyan, C. Safta, H.N.Najm, B.J. Debusschere, D. Ricciuto and P. Thornton. 2014 Dimensionality reduction for complex models via Bayesian compressive sensing. *International Journal for Uncertainty Quantification* **4**.

[53] P. Tsilifis, I. Papaioannou, D. Straub and F. Nobile. 2020 Sparse Polynomial Chaos expansions using variational relevance vector machines. *J. Comp. Phys.* **416** 109498.

[54] D. Scott. 1992 Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley, New York.

[55] B. Silverman. 1986 Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

[56] J. Hwang, S. Lay, and A. Lippman. 1994 Nonparametric Multivariate Density Estimation: A Comparative Study. *IEEE Trans. Signal Process.* **42**, 2795–2810.

[57] A. Bowman. 1984 An alternative method of cross-validation for the smoothing of density estimates. *Biometrics* **71**, 353–360.

[58] M. Jones, J. Marron, and J. Sheather. 1996 A Brief Survey of Bandwidth Selection for Density Estimation. *J. Am. Stat. Assoc.* **91**, 401–407.

[59] S. Kullback and R. A. Leibler. 1951 On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86.

[60] J. Kennedy and R. Eberhart. 1995 Particle Swarm Optimization. In *Proceedings of IEEE International Conference on Neural Networks*, 1942–1948.

[61] F. Montomoli et al. 2010 Interaction of wheelspace coolant and main flow in a new aeroderivative low pressure turbine. *Journal of turbomachinery* **132**, 1-7.

[62] F. Montomoli and M. Massini. 2013 Gas turbines and uncertainty quantification: Impact of PDF tails on UQ predictions, the Black Swan. In *Proceedings of ASME Turbo Expo 2013: Power for Land, Sea, and Air.*

[63] J. R. M. Hosking and J. R. Wallis. 1987 Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* **23**, 339-349.

[64] Smith, Estimating the tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.

[65] D. E. Goldberg. 1989 Genetic Algorithms. In *Search, Optimization and Machine Learning*, Addison-Wesley Longman Publishing Co.

[66] R. H. Byrd, M.E. Hribar and J. Nocedal. 1999 An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM Journal on Optimization* **9**, 877–900.

[67] R. A. Waltz, J. L. Morales, J. Nocedal and D. Orban . 2006 An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming* **107**, 391–408.

[68] A. Gaymann, M. Pietropaoli, L. G. Crespo, S. P. Kenny, and F. Montomoli. 2018 Random Variable Estimation and Model Calibration in the Presence of Epistemic and Aleatory Uncertainties. *WCX 18 SAE World Congr. Exp.*, 1–14.

[69] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. 1956 Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator. *Ann. Math. Stat.* **27**, 642–669.

[70] P. Massart. 1990 The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *Ann. Probab.* **18**, 1269–1283.

[71] D. Hull. 1981 An Introduction to Composite Materials. Cambridge University Press.

[72] R. M. Jones. 1999 Mechanics of Composite Materials. Taylor & Francis.

[73] I. M. Daniel and O. Ishai. 2005 Engineering Mechanics of Composite Materials. Oxford University Press.

[74] Z.G. Ghauch, V. Aitharaju, W.R. Rodgers, P. Pasupuleti, A. Dereims and R.G. Ghanem. 2019 Integrated stochastic analysis of fiber composites manufacturing using adapted polynomial chaos expansions. *Composites Part A: Applied Science and Manufacturing* **118**, 179-193.

[75] P. Sasikumar, A. Venketeswaran, R. Suresh and S. Gupta. 2015 A data driven polynomial chaos based approach for stochastic analysis of CFRP laminated composite plates. *Composite Structures* **125**, 212-227.

[76] F. Zhang, B. Cosson, S. Comas-Cardona and C. Binetruy. 2011 Efficient stochastic simulation approach for RTM process with random fibrous permeability. *Composites Science and Technology* **71**, 1478-1485.

[77] S. E. Yamada and C. T. Sun. 1987 Analysis of Laminate Strength and its Distribution. *J. Compos. Mater* **12**, 275–284.