CSMA 2026
17ème Colloque National en Calcul des Structures
18-22 Mai 2026, Presqu'île de Giens (Var)

A probabilistic validation framework for structural models using sparse test data

Victor Estavoyer[1], Nathalie Bartoli[2,3], Christian Gogu[4], Jean-Philippe Navarro[5]

[1] *AIRBUS, ICA, ONERA Toulouse, victor.estavoyer@airbus.com*
[2] *DTIS, ONERA, Université de Toulouse, 31000, Toulouse, France, nathalie.bartoli@onera.fr*
[3] *Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, 31000, Toulouse, France*
[4] *Institut Clément Ader (ICA), Université de Toulouse, ISAE-SUPAERO, UPS, CNRS, INSA, Mines Albi, Toulouse christian.gogu@gmail.com*
[5] *AIRBUS, Toulouse, jean-philippe.navarro@airbus.com*

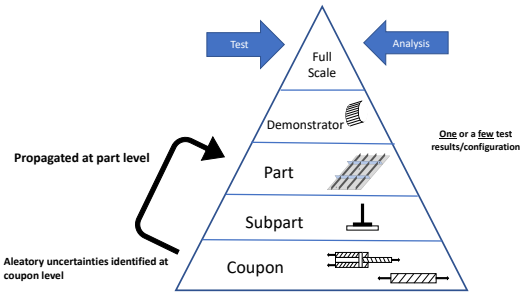
Abstract — This research aims to develop a methodology allowing to extend the validation domain of aeronautical structural models to reduce reliance on costly physical testing while maintaining the current required safety standards . The core challenge is quantifying the risk of applying models to untested configurations. Using composite panel buckling as a case study, the approach employs a multi-fidelity strategy. This leverages low-fidelity analytical models to guide the extrapolation of sparse high-fidelity experimental data. Non-dimensional parameters are used to reduce the high-dimensional input space associated with composites multi-ply laminates. Finally, a probabilistic validation framework propagates uncertainties for a robust comparison between simulation and experiment.
Keywords — Validation, Uncertainty Quantification (UQ), Domain of Validation (DoV), multi-fidelity, buckling, composites

1 Introduction

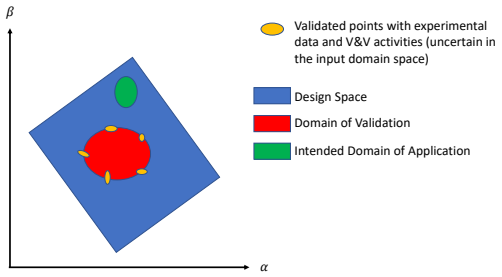
Aeronautical structural analysis can rely on numerical models for demonstrating the compliance of design with safety standards, from a regulatory standpoints these models must be validated by comparing their predictions against physical test results [2]. These limited physical tests combined with Verification & Validation (V&V) activities establish the model's "validation domain" [8]. In an effort to mitigate the temporal and financial constraints associated with the development of novel aircraft, Airbus aims to optimize the number of physical tests by enhancing the application of numerical models that possess an established experimental validation. The aim of this paper is to establish a preliminary methodology for delineating the validation domain of a numerical model based on the numerical resolution of partial differential equations, as well as for assessing the risks linked to the application of the model beyond this specified domain. To achieve this objective, we shall endeavor to establish quantitative metrics [5], such as robustness indicators, to validate the assessment of the outcomes produced by a model when employed outside its validation domain. This metric must encompass all pertinent uncertainties (including numerical discrepancies, measurements uncertainty, etc.), the sensitivity of physical phenomena with respect to the parameters delineating the model's application field, and the distance of previously tested configurations. The complexities of these numerical models are enlarged by the materials they seek to simulate. This challenge is particularly evident with the shift to advanced materials, such as the fibre-reinforced composites that have revolutionized modern aerospace engineering [1].

2 Verification, Validation and Uncertainty Quantification (VVUQ)

A central theme in the VVUQ literature is the development of frameworks, from comprehensive and modular to highly domain specific. On one hand of the literature, studies such as [10] propose a comprehensive modular framework designed to cover diverse engineering fields, in [11] they offer a complete framework for scientific computing that covers both types of uncertainties: aleatory and epistemic. While these comprehensive frameworks provide essential tools, they do not always address the specific constraints of validation at the higher tiers of the structural testing pyramid (see Figure 1a). This "building block approach," common in aerospace industries, is fundamentally characterized by sparse test data at the component or demonstrator level, where experiments are complex and expensive. The problematic is further compounded because these few, precious high-fidelity data points are themselves defined by uncertain inputs. The uncertainties[1] linked to the material properties and geometric imperfections mean that a single test does not represent a single point, but rather a probabilistic distribution in the input space (see figure1b). Our use case, therefore, demands a highly specific framework focused on data fusion and extrapolation, capable of leveraging low-fidelity data to maximize the value of scarce, uncertain high-level tests.



(a) Test pyramid in aerospace industry



(b) Design space, Domain of Validation, Domain of Application

3 Application to local skin buckling of composite stiffened panels

The use case addressed in this work concerns the local skin buckling of composite stiffened panels, using an Ω-stiffened fuselage panel as the foundational case study. We employ a "superstringer" model, which leverages the Rayleigh-Ritz method [7][6] to discretize the governing partial differential equations for the central bay's buckling behavior (see Figure 2). As indicated on the figure 2, to determine the local buckling load of a panel stiffened with Ω-stringers, we only consider the stringer foot attached to the skin plate and the skin plate [14]. As this local buckling is a typical design-sizing criterion, validating its predictive model is crucial.

However, building a validation domain over the fundamental physical inputs (e.g., material properties, geometry, and especially the stacking sequence of the skin and stringers) creates a high-dimensional problem that is computationally intractable. This combinatorial explosion renders direct validation infeasible. Therefore, a major challenge is to identify physics-informed, low-dimensional parameters that effectively capture the influence of these complex combinations and serve as relevant driving variables for the structure buckling behavior.

[1]The distinction between aleatory and epistemic uncertainties will be addressed in a separate, forthcoming study.

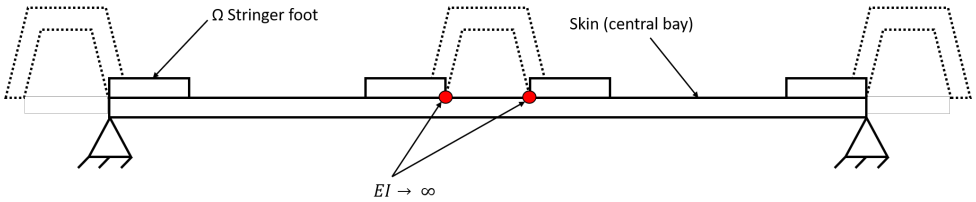


Figure 2: Superstringer idealization for a Ω-composite stiffened panel

The core of the superstringer model is the buckling behavior of the central skin bay (see figure 2). For this first approach the problem will be simplified to a simply supported anisotropic plate, the stringer foot interaction is initially not taken into account. For a symmetric rectangular plate of width b, defined by an $x-y$ coordinate system the plate buckling behavior is governed by [13]:

$$D_{11}\frac{\partial^4 w}{\partial x^4}+2(D_{12}+2D_{66})\frac{\partial^4 w}{\partial x^2\partial y^2}+D_{22}\frac{\partial^4 w}{\partial y^4}+4D_{26}\frac{\partial^4 w}{\partial x\partial y^3}+N_x\frac{\partial^2 w}{\partial x^2}+N_y\frac{\partial^2 w}{\partial y^2}+2N_{xy}\frac{\partial^2 w}{\partial x\partial y}=0 \tag{1}$$

where D-terms are the flexural stiffnesses (ABD-matrix), w is the out-of-plane displacement, and N quantities are the in-plane stress resultants.

To address the high-dimensionality challenge, we follow [13] and [4] by non-dimensionalizing Eq. (1). This process isolates the laminate's complex behavior into four key non-dimensional parameters governing the skin:

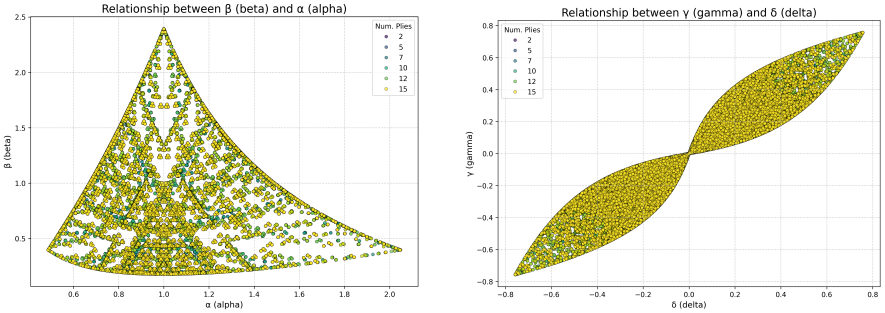
$$\alpha=\frac{\lambda}{b}\sqrt[4]{\frac{D_{22}}{D_{11}}};\quad \beta=\frac{D_{12}+2D_{66}}{\sqrt{D_{11}D_{22}}};\quad \gamma=\frac{D_{16}}{\sqrt[4]{D_{11}^3D_{22}}};\quad \delta=\frac{D_{26}}{\sqrt[4]{D_{22}^3D_{11}}} \tag{2}$$

These parameters represent the panel aspect ratio (α), orthotropy (β), flexural anisotropy (γ,δ). Substituting these into the governing equation, along with non-dimensional loads (n_x,n_y,n_{xy}), yields the non-dimensional form:

$$\frac{1}{\alpha^2}\frac{\partial^4 w}{\partial\xi^4}+2\beta\frac{\partial^4 w}{\partial\xi^2\partial\eta^2}+\alpha^2\frac{\partial^4 w}{\partial\eta^4}+4\frac{\gamma}{\alpha}\frac{\partial^4 w}{\partial\xi^3\partial\eta}+4\delta\alpha\frac{\partial^4 w}{\partial\xi\partial\eta^3}+\pi^2\{n_x\frac{\partial^2 w}{\partial\xi^2}+\alpha^2 n_y\frac{\partial^2 w}{\partial\eta^2}-2\alpha n_{xy}\frac{\partial^2 w}{\partial\xi\partial\eta}\}=0 \tag{3}$$

$$n_x=\frac{N_xb^2}{\pi^2\sqrt{D_{11}D_{22}}};\quad n_y=\frac{N_yb^2}{\pi^2D_{22}};\quad n_{xy}=\frac{N_{xy}b^2}{\pi^2\sqrt[4]{D_{11}D_{22}^3}} \tag{4}$$

Figures 3a and 3b illustrate the design space projection onto these four skin parameters.



(a) Design space on α and β (b) Design space on δ and γ

Figure 3: The design space of anisotropic rectangular composite panels

While parameters ($\alpha,\beta,\gamma,\delta$) capture the skin behavior, they neglect the boundary conditions imposed by the stiffeners, those additional specific parameters will be presented in the final presentation. The central hypothesis is that those latent variables input space is sufficient to characterize the buckling response, and it will serve as the basis for constructing the validation domain.

3.1 Validation framework for typical test pyramid approach

Domain of Validation

The global validation framework is illustrated in Figure 4. The primary objective is to rigorously delineate the Domain of Validation (DoV) for a given numerical model, specifically within a data-sparse context. Each experimental test configuration i is defined by a vector of physical parameters $\mathbf{x}_i$, which includes material properties, geometric parameters, boundary conditions and loading conditions. This vector is inherently uncertain; we denote it as a random vector $\mathbf{X}_i$ following a distribution $p_i(\mathbf{x})$. This distribution p_i encapsulates uncertainties (variability in material modulus E_{11}, panel thickness t or ply angle characterized at the coupon level). These physical uncertainties are then propagated into the space of the model's latent input variables $\boldsymbol{\theta} = [\alpha, \beta, \gamma, \delta]^T$ (see Eq. (2)), resulting in a specific probability distribution $P_i(\boldsymbol{\theta})$ for each test configuration.

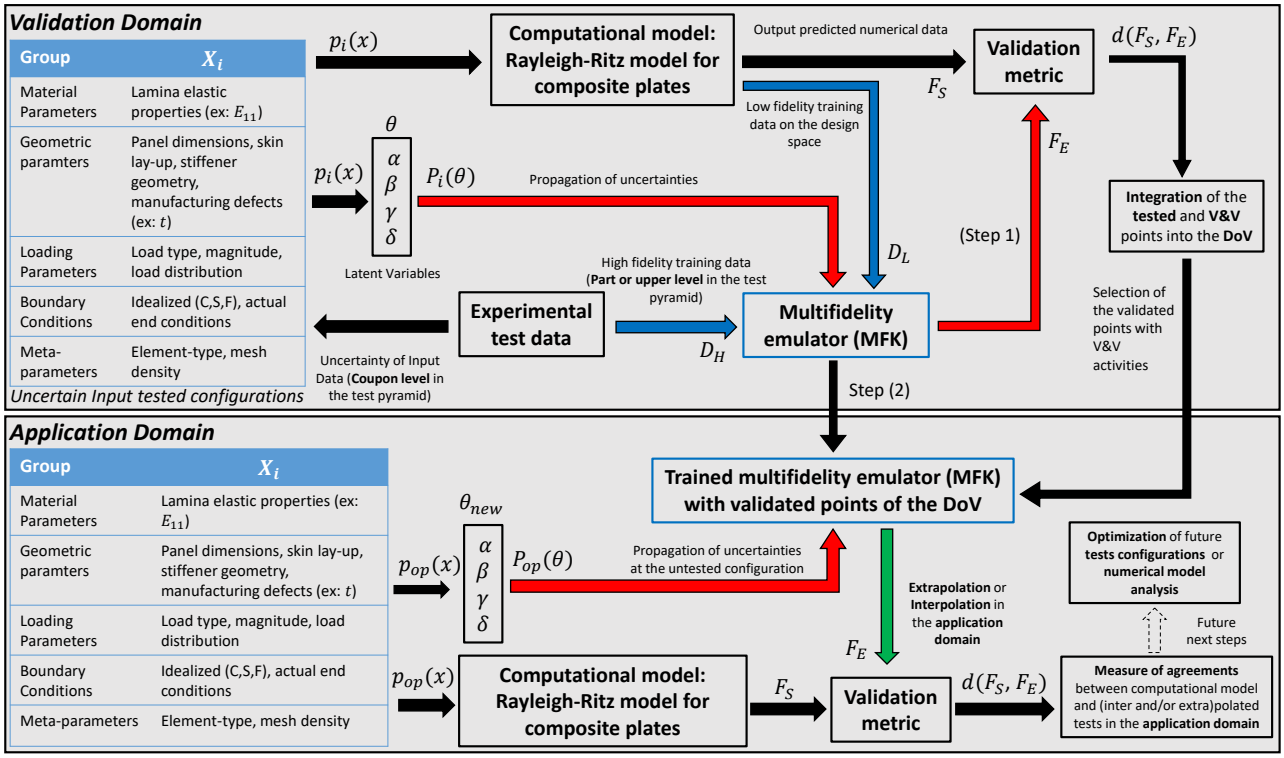


Figure 4: Validation framework from validation domain construction to application domain uncertainty assessment

This defines the High-Fidelity (HF) training set $\mathcal{D}_H = \{(P_i(\boldsymbol{\theta}), y_{H,i})\}_{i=1}^{N_H}$, where $P_i(\boldsymbol{\theta})$ is the uncertain input distribution and $y_{H,i}$ is the single, fixed experimental outcome. To "boost" this sparse HF data, we leverage a larger Low-Fidelity (LF) dataset $\mathcal{D}_L = \{(\boldsymbol{\theta}_{L,j}, y_{L,j})\}_{j=1}^{N_L}$ generated by the numerical model itself. A multi-fidelity emulator (see Section 3.2 for the emulator construction approach) M_{MFK} [3][12], likely instantiated as a family of metamodels to accommodate the input uncertainty in $\mathcal{D}_H$, is trained on both $\mathcal{D}_H$ and $\mathcal{D}_L$. This emulator serves as a probabilistic emulator for the experimental ground truth.

To assess the numerical model validity at a tested configuration i, we propagate the input uncertainty $P_i(\boldsymbol{\theta})$ via Monte Carlo experiments through both the numerical model (LF) and the trained M_{MFK}. This yields two empirical cumulative distribution functions (CDFs): $F_S(\cdot)$ for the simulation and $F_E(\cdot)$ representing the experiment. We quantify the probabilistic error using a validation metric [5], such as the area validation metric [11]:

$$d(F_S, F_E) = \int_{-\infty}^{\infty} |F_S(y) - F_E(y)|\, dy$$

where y is the system response quantity. An expert-defined threshold ϵ_{val} is set, and the numerical model is considered validated at configuration i if $d(F_S, F_E) < \epsilon_{val}$. By applying this process iteratively to all N_H test configurations, we rigorously define the DoV as the subspace of $\boldsymbol{\theta}$ where the numerical model demonstrates sufficient predictive accuracy.

Domain of Application

The second step is to quantify the numerical model reliability within its intended Domain of Application (DoA), as depicted in Figure 1b. The DoA may be located relative to the DoV such that predictions require interpolation, extrapolation, or a combination of both. The relevance of using Multi-Fidelity Kriging (MFK) [3] is particularly significant here. The MFK predictive mean $\mu_{MFK}(\boldsymbol{\theta})$ is co-kriged with the abundant LF data. Consequently, in data-sparse regions or extrapolation, the prediction does not revert to an arbitrary trend (e.g., constant mean) but remains correlated with the physical behavior predicted by the numerical (LF) model.

For any new, untested configuration $\boldsymbol{\theta}_{new}$ within the DoA, we define its associated operational input uncertainties, $P_{op}(\boldsymbol{\theta})$ (coupon level in the test pyramid Figure 1a). By propagating this distribution P_{op} through both the numerical model (LF) and the MFK emulator, we again compute their respective CDFs, F_S and F_E. The validation metric $d(F_S, F_E)$ is then re-calculated, providing a quantitative assessment of the model's reliability for this specific application, fully accounting for all aleatory and epistemic uncertainties. This metric serves as a decision-making input to assess if the model's reliability is sufficient or if additional test experiments are necessary to expand the DoV [9].

3.2 Construction of the multi-fidelity emulator

The emulator construction process is divided into data formulation and training of K models.

- **Phase 1: Model and data formulation**
  - **Define fidelity levels**: High-Fidelity (HF) "test" $Y_H = f_H(\boldsymbol{x})$ and Low-Fidelity (LF) "simulation" $Y_L = f_L(\boldsymbol{x})$.
  - **Define training data sources**:

    (1) **HF data**: N_H pairs $\mathcal{D}_H = \{(X_{H,i}, y_{H,i})\}_{i=1}^{N_H}$, with uncertain inputs $X_{H,i} \sim p_i(\boldsymbol{x})$ and fixed outputs $y_{H,i}$.[2]

    (2) **LF Data**: Generated dynamically for each model, combining N_A additional LHS points with nested $f_L(\boldsymbol{x})$ evaluations at the sampled HF locations.

- **Phase 2: Emulator ensemble construction**
  - **Inputs:**

    (1) The HF dataset $\mathcal{D}_H$ (specifically the outputs $\boldsymbol{y}_H$) and input distributions $p_i(\boldsymbol{x})$.

    (2) The LF model function $f_L(\boldsymbol{x})$ (used as a "black box" to generate LF outputs).

    (3) Data generation parameters (LHS, N_A).

    (4) The number of models to create: K
  - **Outputs:**

    (1) The complete ensemble of K trained MFK emulators: $\mathcal{E} = \{M^{(k)}\}_{k=1}^{K}$.

    (2) Each model $M^{(k)} \in \mathcal{E}$ is a function capable of providing a prediction (mean $\mu_k(\boldsymbol{x})$) and an associated statistical variance ($\sigma_k^2(\boldsymbol{x})$) for any new input $\boldsymbol{x}$.

[2] However, this method also works for uncertain output variables: $Y_{H,i}$.

– **Emulator generation**: Build an ensemble of K MFK models, $\mathcal{E} = \{M^{(k)}\}_{k=1}^{K}$, to capture the epistemic uncertainty from the $p_i(\boldsymbol{x})$ distributions.

– **Emulator training loop**: For each model $k \in [1, K]$:

1. **Generate dynamic datasets**:
   (1) Sample HF inputs $\boldsymbol{x}_H^{(k)} \sim p_i(\boldsymbol{x})$ and additional LF inputs $\boldsymbol{x}_A^{(k)} \sim$ LHS.
   (2) Construct HF Set $\mathcal{D}_H^{(k)} = \{(\boldsymbol{x}_H^{(k)}, \boldsymbol{y}_H)\}$ (using fixed $\boldsymbol{y}_H$).
   (3) Construct LF Set $\mathcal{D}_L^{(k)} = \{(\boldsymbol{x}_L^{(k)}, f_L(\boldsymbol{x}_L^{(k)}))\}$, where $\boldsymbol{x}_L^{(k)} = \boldsymbol{x}_H^{(k)} \cup \boldsymbol{x}_A^{(k)}$.
2. **Train model**: Train $M^{(k)}$ using the generated $\mathcal{D}_L^{(k)}$ and $\mathcal{D}_H^{(k)}$.

– **Model properties**: Each resulting $M^{(k)} \in \mathcal{E}$ is a Kriging model predicting $Y|\boldsymbol{x}, M^{(k)} \sim \mathcal{N}(\mu_k(\boldsymbol{x}), \sigma_k^2(\boldsymbol{x}))$, where μ_k is the mean and σ_k^2 is the statistical variance.

3.3 Emulator application for UQ in the Domain of Application

Once the emulator ensemble $\mathcal{E}$ is constructed, it serves as a probabilistic emulator for the true HF response. It is now used to perform uncertainty quantification (UQ) within the intended Domain of Application (DoA).

- **Phase 3: Prediction and uncertainty propagation**

  – **Inputs:**

  (1) The emulator $\mathcal{E}$ (the output of Phase 2).

  (2) The definition of the Domain of Application (DoA): the probability distribution of operational inputs $p_{op}(\boldsymbol{x})$.

  (3) The number of Monte Carlo samples to generate (M).

  – **Outputs:**

  A set of M output samples: $\{y^{(j)}\}_{j=1}^{M}$.[3]

  – **Define the Domain of Application (DoA)**: Define the probability distribution $X \sim p_{op}(\boldsymbol{x})$ for the inputs in the DoA.

  – **Assessing interpolation vs. extrapolation mode**: The emulator's behavior depends on the location of a new input $\boldsymbol{x}_{new} \sim p_{op}(\boldsymbol{x})$ relative to the training data $\mathcal{D}_H$ and $\mathcal{D}_L$.

  (1) **Interpolation**: Occurs when $\boldsymbol{x}_{new}$ falls *within* the convex hull of the training data. Here, predictions $\mu_k(\boldsymbol{x}_{new})$ are reliable and the statistical variance $\sigma_k^2(\boldsymbol{x}_{new})$ is low.

  (2) **Extrapolation**: Occurs when $\boldsymbol{x}_{new}$ falls *outside* this domain. The statistical variance $\sigma_k^2(\boldsymbol{x}_{new})$ increases significantly, correctly quantifying the model's low confidence.

  (3) **MFK extrapolation benefit**: In extrapolation, the MFK mean $\mu_k(\boldsymbol{x}_{new})$ remains correlated with the LF model $f_L(\boldsymbol{x}_{new})$. This provides a physically-informed trend, preventing the prediction from reverting to a non-physical constant.

  – **UQ via Monte Carlo (triple sampling)**: To compute the final output distribution $p(y)$, generate M samples (e.g., $M \gg 10{,}000$). For each sample $j = 1...M$:

  1. **Sample 1 (input uncertainty)**: Draw an input from the operational distribution:

$$\boldsymbol{x}^{(j)} \sim p_{op}(\boldsymbol{x})$$

[3] This is a set of raw "points," resulting from the triple sampling procedure (input, model, statistical). It is not yet an analyzed distribution.

2. **Sample 2 (model uncertainty)**: Draw a model uniformly from the ensemble:

$$M^{(j)} \sim \mathcal{U}(\mathcal{E})$$

3. **Sample 3 (statistical uncertainty)**: Draw the final output from the chosen model's predictive distribution (which is in interpolation or extrapolation mode):

$$y^{(j)} \sim \mathcal{N}\left(\mu_j(\boldsymbol{x}^{(j)}), \sigma_j^2(\boldsymbol{x}^{(j)})\right)$$

- **Phase 4: Output distribution analysis**
  - **Inputs:**

    The set of M output samples $\{y^{(j)}\}_{j=1}^{M}$ (the output of Phase 3).
  - **Outputs:**

    (1) The non-parametric approximation of the final output distribution $p(y)$ (e.g., as a histogram or kernel density estimate).

    (2) Statistical metrics calculated from $p(y)$ (e.g., mean, variance, quantiles, probabilities of exceeding a threshold).

    (3) A quantification of the overall uncertainty, integrating all three sources (operational, epistemic model, and statistical kriging).
  - **Construct distribution**: The collection of samples, $\{y^{(j)}\}_{j=1}^{M}$, forms a non-parametric approximation of the marginal output distribution $p(y)$.
  - **Analyze results**: This final distribution $p(y)$ integrates all three sources of uncertainty:

    (1) The aleatory and epistemic operational uncertainty $p_{op}(\boldsymbol{x})$.

    (2) The epistemic model uncertainty (from the emulator $\mathcal{E}$).

    (3) The statistical uncertainty (Kriging variance σ_k^2), which is naturally larger for samples in extrapolation.

4 Results and conclusions

This paper established a probabilistic validation framework for structural models using sparse, test data. We detailed the construction of a multi-fidelity emulator that handles input uncertainty in high-fidelity (HF) test data while leveraging low-fidelity (LF) simulations, using physics-based non-dimensional parameters to manage dimensionality and comprehensibility.

Ongoing work involves the concrete application of this framework. The objective is to validate the "superstringer" Rayleigh-Ritz model [7] using experimental data. This will allow us to formally define the model's validation domain (DoV) and to provide a quantitative metric to support the decision to use it in a defined application domain. Preliminary results will be exposed in the final presentation.

References

[1] *Annex II - AMC 20-29 - Composite Aircraft Structure*. 2010.

[2] *Certification Specifications for Large Aeroplanes (CS-25)*. 2007.

[3] Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis.

[4] J. Enrique Herencia, Paul M. Weaver, and Michael I. Friswell. *Closed-Form Solutions for Buckling of Long Anisotropic Plates with Various Boundary Conditions under Axial Compression*. In: *Journal of Engineering Mechanics* 136.9 (Sept. 2010), pp. 1105–1114. DOI: 10.1061/(ASCE)EM.1943-7889.0000158.

[5] Yu Liu et al. *Toward a Better Understanding of Model Validation Metrics*. In: *Journal of Mechanical Design* 133.7 (July 1, 2011), p. 071005. DOI: 10.1115/1.4004223.

[6] Christian Mittelstedt. *Explicit local buckling analysis of stiffened composite plates accounting for periodic boundary conditions and stiffener–plate interaction*. In: *Composite Structures* 91.3 (Dec. 2009), pp. 249–265. DOI: 10.1016/j.compstruct.2009.04.021.

[7] Pablo Moreno-García, José V. Araújo dos Santos, and Hernani Lopes. *A Review and Study on Ritz Method Admissible Functions with Emphasis on Buckling and Free Vibration of Isotropic and Anisotropic Beams and Plates*. In: *Archives of Computational Methods in Engineering* 25.3 (July 1, 2018), pp. 785–815. DOI: 10.1007/s11831-017-9214-7.

[8] William L. Oberkampf and Matthew F. Barone. *Measures of agreement between computation and experiment: Validation metrics*. In: *Journal of Computational Physics* 217.1 (Sept. 2006), pp. 5–36. DOI: 10.1016/j.jcp.2006.03.037.

[9] William L. Oberkampf and Matthew F. Barone. *Measures of agreement between computation and experiment: Validation metrics*. In: *Journal of Computational Physics* 217.1 (Sept. 2006), pp. 5–36. DOI: 10.1016/j.jcp.2006.03.037.

[10] Stefan Riedmaier et al. *Unified Framework and Survey for Model Verification, Validation and Uncertainty Quantification*. In: *Archives of Computational Methods in Engineering* 28.4 (June 2021), pp. 2655–2688. DOI: 10.1007/s11831-020-09473-7.

[11] Christopher J. Roy and William L. Oberkampf. *A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing*. In: *Computer Methods in Applied Mechanics and Engineering* 200.25 (June 2011), pp. 2131–2144. DOI: 10.1016/j.cma.2011.03.016.

[12] Paul Saves et al. *SMT 2.0: A Surrogate Modeling Toolbox with a focus on hierarchical and mixed variables Gaussian processes*. In: *Advances in Engineering Software* 188 (Feb. 2024), p. 103571. DOI: 10.1016/j.advengsoft.2023.103571.

[13] Paul M. Weaver and Michael P. Nemeth. *Bounds on Flexural Properties and Buckling Response for Symmetrically Laminated Composite Plates*. In: *Journal of Engineering Mechanics* 133.11 (Nov. 2007), pp. 1178–1191. DOI: 10.1061/(ASCE)0733-9399(2007)133:11(1178).

[14] C El Yaakoubi-Mesbah and C Mittelstedt. *Computational model for local buckling of compressively loaded omega stringer-stiffened panels*.