

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/176252/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Walton, David B., Featherston, Carol , Coggon, Simon and Kundu, Abhishek 2025. A Bayesian approach to robust early-stage aircraft wing design under uncertainty with reduced order surrogate models. Presented at: AIAA SCITECH 2025 Forum, Orlando, FL, USA, 6-10 January 2025. 10.2514/6.2025-2816

Publishers page: <http://dx.doi.org/10.2514/6.2025-2816>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



A Bayesian approach to robust early-stage aircraft wing design under uncertainty with reduced order surrogate models

Bryn Walton* and Carol Featherston

School of Engineering, Queen's Buildings, The Parade, Cardiff University, Cardiff CF24 3AA, UK

Simon Coggon and Abhishek Kundu†

*School of Engineering, Queen's Buildings, The Parade, Cardiff University, Cardiff CF24 3AA, UK
Airbus Operations Limited, Aerospace Avenue, Filton, Bristol BS34 7PA, UK*

Industrial design of aerostructures is a complex process stemming from challenges associated with quantifying and managing the uncertainty in loading and operating conditions, variable levels of design immaturity and the potential variability in manufacturing and assembly. Setting the margins on design parameters, particularly at the early design stage of aerostructures poses significant challenges. For aircraft wing design, the loads envelopes are generated with a set of industry standard simulations of aircraft maneuver and aerodynamic loads, but these simulations are expensive and time consuming and unsuitable for rapid exploration of the design space. The study presented here develops an industrial design workflow with the aims of providing a holistic Bayesian approach for design optimization and uncertainty management in early stages of design. In order to explore the limits of the design space, it is essential to create a Bayesian data-driven surrogate model that maps the design parameters to vector-valued outputs of load-bearing and aerodynamic performance. The surrogate model is then utilized in an inverse problem setting to optimize a set of important design parameters based on user-defined operational constraints. Data compression methods have been employed to reduce the dimensionality of high-dimensional vector-valued outputs in order to effectively manage the computational overhead and ensure optimal fitting of multi-output Gaussian process regression model. Optimization of the parameters of interest is performed with Bayesian inference with novel objective functions to accommodate the variability introduced due to the presence of stochastic variables. The proposed approach is applied to an aircraft wing, parametrized by jig twists and bending and torsional stiffnesses across its span. The training data is produced as a design of experiment from coupled structural mechanics and computational fluid dynamics models that generate coefficients of lift over drag, as well as shear force, bending moment, and torque responses at a selection of flight envelopes and discrete points along the span. The results demonstrate multiple design scenarios with performance objectives and constraints for various flight envelope. It has been shown that it is possible to produce probabilistic estimates of design parameters with a Bayesian approach, while leveraging the computational efficacy of a pre-trained a stochastic surrogate model. The proposed framework therefore provides industrial design teams with information regarding potential limits of particular design choices, and help with setting design margins for later detailed design stages.

I. Introduction

THE data used in this study is provided from the Airbus robust wing twist optimization academic challenge. This data set is a normalized data set for academic use of high fidelity simulations for the aerodynamic drag and external loads for a representative long range aircraft. This forms the basis of the surrogate, which has been created for the robust optimization of wing twist (jig shape) to balance aerodynamic performance with the risk of exceeding the design loads margins under various operational loads and aircraft manoeuvre scenarios.

In recent years, data-driven surrogate models have been used to provide insight into the correlation between model

*Email: waltondb@cardiff.ac.uk

†Email: kundua2@cardiff.ac.uk

parameters and quantities of interest (QoI) over a full range of potential designs at a reduced cost [1, 2]. There is great interest within the aerospace industry in developing and understanding the application of this technology [3–5]. Bayesian methods can address the challenge of understanding the epistemic uncertainty introduced through surrogate modelling and, by extension, understand the confidence level in model predictions [6–8].

For complex systems operating under diverse, high-stress conditions, detailed information about QoI at important locations is desirable, posing the challenge of mapping design of experiments (DOE) to a vector-valued output. Methods have been developed to formulate multi-output Gaussian process regression (GPR) models [9–11]. However, these approaches require inference of a coregionalization matrix, which becomes prohibitively large for high-dimensional problems. Alternatively, dimensionality reduction techniques such as principal component analysis can be employed to reduce the number of independent GPR models required, though they may introduce additional uncertainty into the model form [12–14].

The development of surrogate models relies on collecting data from high-fidelity numerical simulations, which provide accurate insights into multiphysics design but are computationally expensive. Such models are widely used in structural analysis (e.g., finite element analysis) [15], computational fluid dynamics [16], and thermal analysis [17]. While they expand the design space and allow the exploration of diverse options [18], their high computational cost is a limiting factor in iterative optimization workflows [19]. Optimization algorithms, including genetic algorithms, particle swarm optimization, and Markov Chain Monte Carlo (MCMC) methods, often require numerous evaluations of models, making surrogate models indispensable in reducing computational cost.

This paper explores Bayesian optimization of a subset of wing design parameters, focusing on their coupled aerodynamic and load responses. GP surrogate models, trained on data from FEM simulations, are utilized to optimize wing twist while maintaining computational efficiency. MCMC sampling is employed to optimize the parameters of interest, incorporating a novel likelihood function formulation to enforce bounded constraints. The surrogate model is further leveraged to evaluate the impact of design variations across a range of operating conditions, providing valuable insights into system behavior under scenarios not explicitly included in the optimization process.

II. Problem Statement and Setup

Investigated in this paper is the early-stage design of an aircraft wing, where the design parameters (θ) include the jig twists and wing stiffness. The jig twists, $\theta_{1,2}$, illustrated in Figure 1, are the primary parameters being optimized. These define the wing’s jig shape—its unloaded configuration—designed to allow aerodynamic forces to deform the wing in flight, achieving the desired angle of attack (α). This deformation also depends on the bending and torsional stiffness (EI and GJ respectively) of the wing at specific spanwise locations, represented by $\theta_{3,\dots,10}$. These stiffness values are treated as uncertain and modeled as random variables that follow predefined distributions during optimization.

The optimization process considers several Quantities of Interest (QoIs), including the in-flight load responses (shear force, bending moment, and torque) along the wing span, which are dependent on the design parameters. The objective is to find a distribution over $\theta_{1,2}$ that satisfies constraints on the lift-to-drag ratio (L/D) while also adhering to limits on in-flight force responses (F).

The framework, inspired by the work of [20, 21], involves generating data from FEM analysis of the wing at a predetermined number of DOE points that uniformly sample the design space. This data is then used to train a number of separate GP surrogate models, mapping the design parameters θ to the L/D as $f_A : \theta \mapsto L/D$, and to F as $f_L : \theta \mapsto F$, and $f = \{f_L, f_A\}$.

Optimization of θ is performed using Bayesian inference [22] using MCMC sampling [23] with surrogate models employed for the evaluation of likelihood function. Points are accepted based on their posterior probability of satisfying the specified constraints captured in the likelihood function.

$$\underbrace{\pi(\theta|f, \mathcal{D}, \mathcal{M}, \tau)}_{\text{Posterior}} \propto \underbrace{\mathcal{L}(f(\theta)|\mathcal{M}, \tau)}_{\text{Likelihood}} \underbrace{\pi(\theta|\mathcal{D})}_{\text{Prior}} \quad (1)$$

where $\mathcal{D}$ and $\mathcal{M}$ are the data and model respectively, and τ is a set containing the targets associated with L/D (τ_A) and F (τ_L) such that $\tau = \{\tau_A, \tau_L\}$. $\pi(\cdot)$, has been used to define probability distributions of priors and posteriors while $\mathcal{L}(\cdot)$ is the likelihood function.

The Bayesian framework is demonstrated through experiments that enforce two distinct types of constraints: point and bound constraints. These constraints are incorporated into the likelihood function of the inference model. For the point-constrained task, the likelihood is formulated as a Gaussian probability density function (PDF). In contrast, the bound constraint requires a more novel approach, utilising cumulative distribution functions (CDFs) to define the bounds within the optimization process. The techniques used to create the Bayesian framework are outlined in Section III. The experimental results and discussion are presented in Section IV, followed by concluding remarks in Section V.

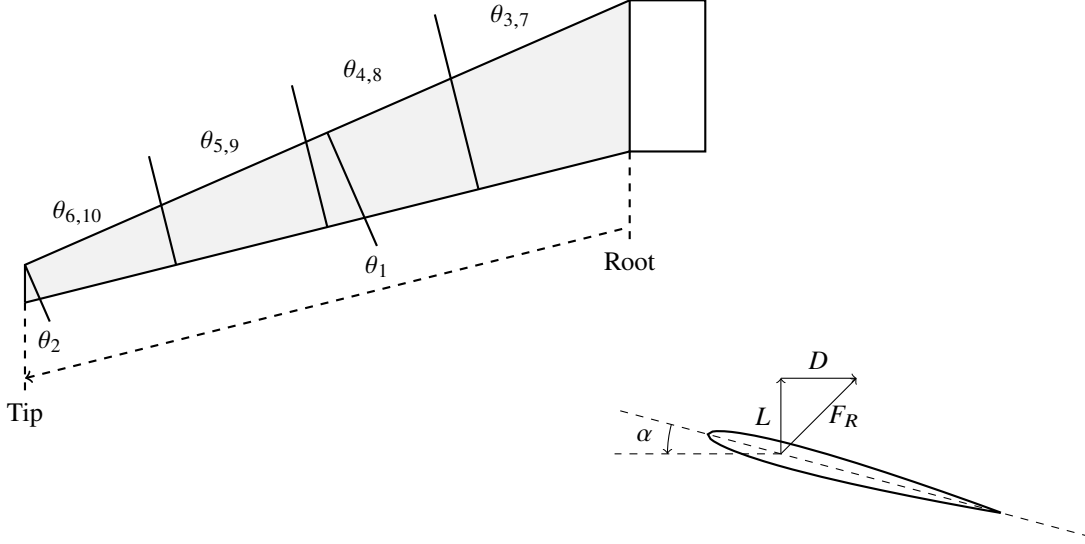


Fig. 1 Approximate location of design parameters on aircraft wing. $\theta_{1,2}$ are located at a point midway between the root and the tip, and at the tip itself. $\theta_{3,...,10}$ are parameter values between the the root, three separating lines, and the tip. In the bottom right hand corner is a standard depiction of an airfoil with applied aerodynamic forces (L, D, F_R) and angle of attack α for reference when reading the text.

III. Methodology

The Bayesian frameworks for the optimization tasks have two main elements, the forward models, and the inverse models. The forward models are data-driven surrogate models in the form of trained Gaussian processes (Section III.A). These GPs provide the model functions for the Bayesian inverse models (Section III.B). The inverse models sample from the posterior distribution over the random variables $\theta_{1,2}$ using the No-U-Turn Sampler (NUTS) which is an adaptive MCMC algorithm based on Hamiltonian Monte Carlo (HMC) (Section III.C).

A. Gaussian Process Regression for Surrogate Modelling

The coupled computational fluid dynamics and computational structural mechanics model is emulated by separate Gaussian processes [24] so they can be queried individually at new design points during MCMC sampling. This method gives an estimate of the output as a normally distributed random variable and can be expressed as follows: The surrogate is a map (f) from the multi-dimensional input space to the output $f : \mathbb{R}^d \mapsto \mathbb{R}$. From Equations (2) and (3) it can be seen that a GP is a normal distribution over functions that is parametrized by a mean function ($M(\theta)$), and covariance function or covariance kernel ($K(\theta, \theta')$).

$$f(\theta|\mathcal{D}) \sim \mathcal{GP}(M(\theta), K(\theta, \theta')) \quad (2)$$

$$\mathcal{GP}(M(\theta_i), K(\theta, \theta'_i)) \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad (3)$$

For vector-valued outputs and multitask learning, independent GPs can be trained on each output to create a set of surrogate models for the entire system. Alternatively, the covariance structure between outputs can be exploited, and the multi-output GP surrogate can be expressed in the Linear Model of Coregionalization (LMC) form [25].

Both types of multi-output GPs are employed in this work. The outputs of interest for the optimization task are represented by independent GP surrogates, while a specific instance of LMC is used to construct the complete multi-output GP across

the broader output space, enabling an assessment of the impact of the optimized parameters across flight envelopes and locations on the wing.

B. Bayesian Inference for Parameter Estimation

Bayesian inference is grounded in Bayes' rule (Equation (4)), which relates conditional probabilities through the joint, marginal, and conditional properties of probability distributions.

$$\pi(\mathbf{y}_A|\mathbf{y}_B) = \frac{\mathcal{L}(\mathbf{y}_B|\mathbf{y}_A)\pi(\mathbf{y}_A)}{p(\mathbf{y}_B)} \quad (4)$$

Here, $\mathbf{y}_A$ and $\mathbf{y}_B$ are disjoint sets partitioned from a set of random variables $y_1, \dots, y_n$. This relationship forms the basis for Bayesian inference by conditioning one variable on another. In Bayesian inference, the goal is to compute the *posterior* distribution ($\pi(\mathbf{y}_A|\mathbf{y}_B)$) of a variable of interest, given noisy data modeled as a function of independent variables. Below is a breakdown of the key terms in Equation (4):

The *prior* ($\pi(\mathbf{y}_A)$) represents prior assumptions about the variable of interest, described by a probability distribution. For instance, a Gamma distribution may be used for strictly positive variables, while uninformed priors, such as flat distributions, are used when no assumptions are available. The *likelihood* ($\pi(\mathbf{y}_B|\mathbf{y}_A)$) quantifies the fit between the model and the data. While not necessarily a probability function, a Gaussian likelihood is common in simple models, parameterized by its mean, and standard deviation which represents the data noise. If the noise level is unknown, a prior can be placed on the standard deviation, enabling inference of its posterior along with the variables of interest. The *marginal likelihood* ($\pi(\mathbf{y}_B)$) is the normalizing constant ensuring the posterior is a valid probability distribution. It is obtained by integrating the product of the prior and likelihood over the variable space. Since it serves as a normalization term, Bayes' rule is often expressed without it, as it is in Equation (1).

Bayesian inference provides a generative framework to model data and uncertainty, enabling robust parameter estimation and prediction through posterior sampling.

C. Sampling the Posterior: No-U-Turn Sampler (NUTS)

The No-U-Turn Sampler (NUTS), introduced by [26], is an adaptive extension of HMC that eliminates the need to specify a fixed trajectory length by dynamically adjusting the number of leapfrog steps. This prevents "U-turns" in the Hamiltonian path, ensuring efficient sampling without manual tuning, especially in high-dimensional spaces with complex posteriors.

HMC employs Hamiltonian dynamics to propose transitions in parameter space by treating the parameters as "positions" with associated "momentum" variables, p , sampled from a Gaussian distribution. The joint probability density of θ and p is expressed by the Hamiltonian function as

$$H(\theta, p) = U(\theta) + K(p), \quad (5)$$

where $U(\theta)$ is the potential energy, corresponding to the negative log-posterior such that regions of high probability correspond to low potential energy. $K(p)$ is the kinetic energy, which is equivalent to the step size taken in parameter space, where the value of K is dependent on the realization of p and a mass matrix associated with it that scales p across all dimensions [27]. The evolution of this system follows Hamilton's equations

$$\frac{d\theta}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial \theta}, \quad (6)$$

where dt is the evolution of steps in the Markov chain rather than time. Numerical integration, such as the leapfrog method, approximates these continuous equations in discrete steps. However, traditional HMC requires a pre-specified number of steps, which is difficult to tune.

NUTS resolves this by incrementally building a binary tree of trajectory states, dynamically doubling its length until a U-turn is detected. A U-turn occurs if $(\theta_j - \theta_0)p_j < 0$, where θ_0 is the starting state and θ_j is a proposed state. This approach balances computational efficiency with effective exploration, avoiding redundant computations. The Metropolis-Hastings criterion (Equation (7)) ensures acceptance of proposed states.

$$\alpha = \min(1, \exp(H(\theta_{\text{current}}, p_{\text{current}}) - H(\theta_{\text{proposed}}, p_{\text{proposed}}))) \quad (7)$$

By automatically adapting trajectory lengths and step sizes, NUTS simplifies tuning and enhances robustness, making it well-suited for exploring complex, high-dimensional posterior distributions.

D. PDF and CDF Likelihood Functions for Applying Constraints

Two types of optimization targets are presented here, point constrained, and bound constrained, therefore, the likelihood functions take on different forms. The point constraint requires a target $\tau_{(\cdot)}$ for which the likelihood is in the form of a Gaussian distribution PDF

$$\mathcal{L}(f(\theta) = \tau_{(\cdot)}) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-((\tau_{(\cdot)} - f(\theta)) - \mu)^2}{2\sigma^2}. \quad (8)$$

Taking the logarithm of Equation (8) and assuming independence gives the final point-constrained likelihood $\mathcal{L}_p$, which is to be maximized. In this formulation, the subscripts L and A refer to the load and aerodynamic responses, respectively, while the superscripts B , S , and T correspond to the bending moment, shear force, and torque, respectively,

$$\ln \mathcal{L}_p = \ln \mathcal{L}_L + \ln \mathcal{L}_A, \quad (9)$$

$$\text{where } \mathcal{L}_L = \mathcal{L}_L^B \mathcal{L}_L^S \mathcal{L}_L^T. \quad (10)$$

For the bound constrained optimization the normal CDF (Π) in Equation (11) applies a lower bound

$$\mathcal{L}_A(f(\theta) \geq \tau_A) = \Pi = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{f(\theta)} \exp \frac{-(t - \mu)^2}{2\sigma^2} dt, \quad (11)$$

and the survival function is used to apply an upper bound in Equation (12)

$$\mathcal{L}_L(f(\theta) \leq \tau_L) = 1 - \Pi \quad (12)$$

where $t = (\tau_{(\cdot)} - f(\theta))$.

The final joint likelihood of the bound constraint $\mathcal{L}_b$ (Equation (13)) is then calculated by the product of weights $\mathbf{w} = [w_1, \dots, w_4]$ and $\mathbf{L} = [\mathcal{L}_L^B, \mathcal{L}_L^S, \mathcal{L}_L^T, \mathcal{L}_A]$, and the log likelihood is evaluated with a normal distribution centered around two. $\mathbf{w}$ ensures balance between output dimension during optimization. In this specific case, the weight vector has been chosen such that $w_1 = w_2 = w_3 = 1/3$ and $w_4 = 1$. This ensures that the sum of all three load CDF likelihoods within $\mathcal{L}_L$ has a maximum value of one, and $\mathcal{L}_A$ also achieves a maximum value of one. Hence, the maximum value of $\mathbf{w}^T \mathbf{L} = 2$, which maximizes Equation (13).

$$\ln(\mathcal{L}_b(f(\theta)|\tau_{L,A}) = \ln \left(\frac{1}{\sigma\sqrt{2\pi}} \right) + \ln \left(\frac{-(\mathbf{w}^T \mathbf{L} - 2)^2}{2\sigma^2} \right) \quad (13)$$

E. Conditioning of Likelihood Functions

In order to evaluate the likelihoods as a function of just the important parameters $\theta_{1,2}$ it is necessary to marginalize out the effect the stiffness parameters have on $f(\theta_i)$ at each step in the chain with a precalculated $n \times d_s$ Latin Hypercube sample S^{LHS} , and an $n \times d_t$ matrix T^{θ_i} that has $T_n^{\theta_i} = \theta_i, \forall n \in \{1, 2, \dots, n\}$. Concatenation then gives $\hat{\theta} = [T^{\theta_i}, S^{LHS}]$ which is the $n \times d$ argument of $f(\theta_i)$ in the chain that produces an $n \times 1$ dimensional output $\mathbf{y}_i$. The final marginalization step requires taking the mean of $\mathbf{y}_i$ thus giving the approximated $f(\theta_{1,2})$

$$f(\theta_{1,2}) = \mathbb{E} [f(\hat{\theta} | \theta_{1,2})] \quad (14)$$

To limit bias in the likelihood function induced by the difference in scale of the variance between outputs, a normalization step before evaluation but after $f(\theta_{1,2})_i$ is obtained in the Markov chain is used giving $\hat{f}(\theta_{1,2})$. This ensures that all

output variances are ≈ 1 while retaining the relationship between points. The normalization in Equation (15) is of a standard score form

$$\hat{f}(\theta_{1,2}) = (f(\theta_{1,2}) - \mu_{\mathcal{D}}) \Sigma_{\mathcal{D}}^{-\frac{1}{2}}, \quad (15)$$

where $\mu_{\mathcal{D}}$ is the vector containing the mean of each output dimension, and $\Sigma_{\mathcal{D}}$ is the covariance between outputs. Both $\mu_{\mathcal{D}}$ and $\Sigma_{\mathcal{D}}$ are pre-calculated from the data before sampling begins, along with the normalised target values, where $f(\cdot)$ in Equation (15) is replaced with $[\tau_L, \tau_A]$.

IV. Results and Discussion

The results of the experiments are presented here under the following conditions. A flight envelope representing the greatest load responses is selected for optimization. The aim is to infer the posterior distribution over the twist

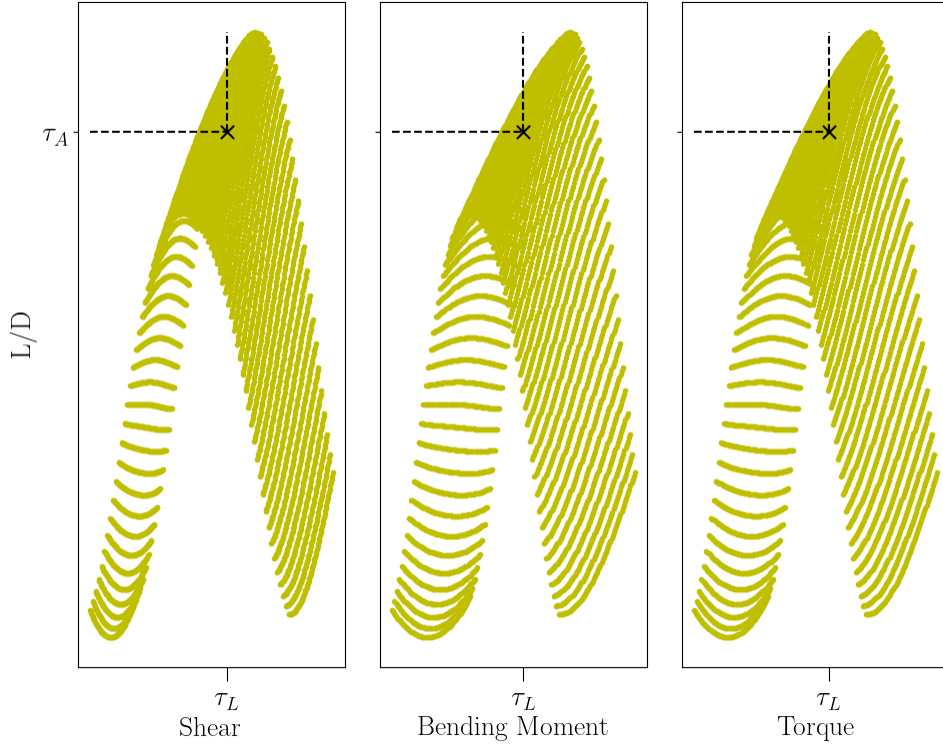


Fig. 2 Distribution of the surrogate model predictions on Latin Hypercube samples of the design space (in yellow). Point-based design targets ('x'). Boundary-based design targets (-).

parameters and recover the predictive posterior distribution over the coefficients of lift over drag, shear force, bending moment, and torque at the root of the wing, given the surrogate model, constraints, and marginalization of the stochastic stiffness parameters. To apply appropriate constraints it is first necessary to define the subspace of reachable regions to ensure the prescribed constraints are valid. In Figure 2 the reachable region is shown in yellow. These data points are the result of taking separate Latin Hypercube samples of the twist and stiffness parameters and applying the method in Section III.E (Equation (14)), with the exception that the use of MCMC to draw samples from the prior of θ_1 and θ_2 is replaced by iterating through the elements of the twist Latin Hypercube. Also shown in Figure 2 are the point and bound constraints. The point constraint 'x' is representative of the response of the wing when cruising at a common altitude and regular gust condition. Similarly, the bound constraint defines the area above the L/D for these conditions and below the load responses.

A. Point Constraint Optimization

This case study attempts to recover the posterior distribution over the twists at a point that exists in the training data. For reference it is a point where all the design parameters are equal to zero. The joint distribution in Figure 3 shows that the expectations of the Gaussian marginals are located at $\theta_1 \approx 0.139$ and $\theta_2 \approx -0.147$ which is an approximate 14% and -15% divergence from the target given that the prescribed limits of the design parameters are between -1 and 1. In Figure 4 the divergence in twist from the target is reflected in the expected value of the posterior predictive distribution in red that is not located directly over the target area. Contributors to this error may include the surrogate models, the residual effects of the stochastic stiffness parameters, and the number of sampling steps. However, further research is required to determine the relative contribution of each factor.

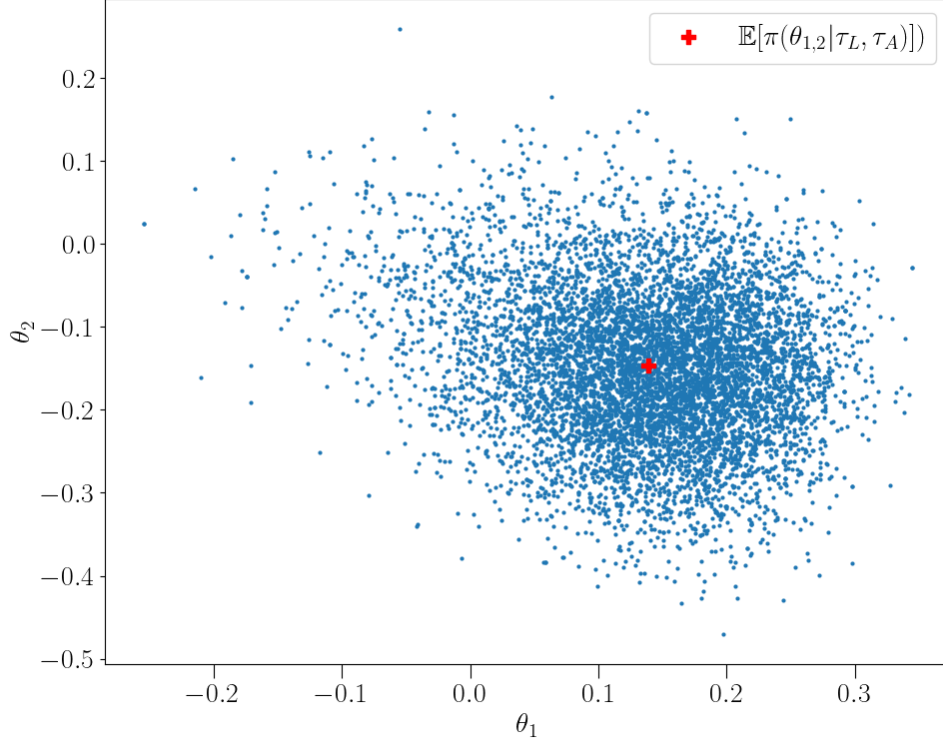


Fig. 3 Joint posterior distribution over θ_1 and θ_2 with marginals for point-based design targets on loads and aerodynamic performance.

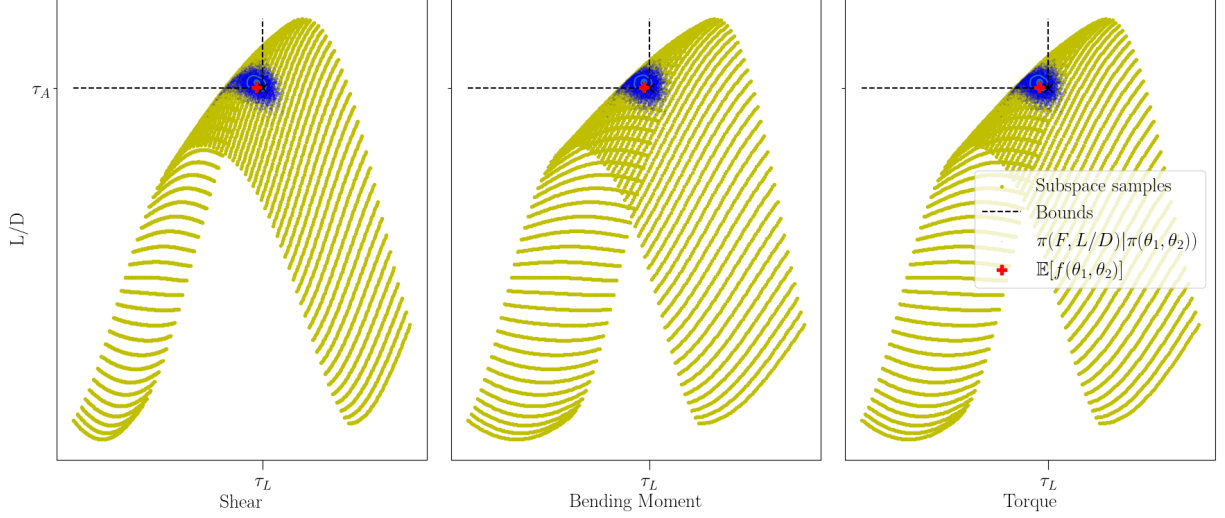


Fig. 4 Posterior predictive distributions on the aerodynamic L/D performance and wing load at the root ("Bending Moment, Torque, Shear Force") for point-based design targets on loads and aerodynamic performance (in blue). Distribution of the surrogate model predictions on Latin Hypercube samples of the design space (in yellow). Point-based design targets ('x'). Boundary-based design targets (-).

B. Bound Constraint Optimization

The following results pertain to the multiphysics optimization of $\theta_{1,2}$, with a bound constraint on L/D at the known current design's optimum. The bounds on F are chosen to correspond to the same design point as L/D , but for a flight envelope case where stresses are greatest, relative to the remaining unconsidered cases. In Figure 5, the bounds are indicated by a dashed line. The objective of the optimization is to sample from the posterior distribution over θ that maximizes L/D and minimises F within these bounds. The predictive posterior distribution of F and L/D , given $\pi(\theta_{1,2})$, is plotted in blue in Figure 5. The majority of the predictive posterior resides within the bounds. However, some of its variance is observed to be below the L/D bound and above the bound on F . This may be attributed to the negative correlation between shear force and bending moment with the torque, causing them to compete in the optimization process and expressing likelihoods that are nearing optimal values.

An important observation is that the expected values remain entirely within the bounds for all $(F, L/D)$ pairs, indicating successful optimization of $\theta_{1,2}$ given the problem statement. This suggests that an optimized design point with respect to the twist parameters has been found, achieving an improvement in L/D during cruising (potentially enhancing efficiency) without compromising the load response at the wing root.

Considering that $\mathbb{E}[\pi(\theta_{1,2})]$ provides a posterior predictive mean that improves the structural response at the root, it is desirable to investigate whether it has a similar effect across the entire set of flight envelopes. Figure 6 illustrates the $\mathbb{E}[f(\theta_1, \theta_2) | \pi(\theta_{1,2})]$ with $\pm 3 \times \text{STD}$ error bars, plotted alongside the observed data at the bound design point at the root for all flight envelopes. The data points are highlighted with colored ellipses, indicating whether $\mathbb{E}[f(\theta_1, \theta_2) | \pi(\theta_{1,2})]$ is less than the observed data (in green) or greater than the observed data (in red). The vertical diameter of each ellipse is scaled relative to the difference between $\mathbb{E}[f(\theta_1, \theta_2) | \pi(\theta_{1,2})]$ and the observed data.

It is evident from Figure 6 that the optimized twist parameters have had a positive effect on the torque, as all cases with high torque exhibit a reduction. There are, however, five cases in the mid-to-low torque range that show relatively small increases compared to the significant decreases observed in the other cases. The level of correlation between the shear force and bending moment is apparent in Figure 6, with most cases showing the same direction of change. Exceptions include cases 34 and 45, which show minimal change from their original values. Importantly, the six cases with the highest stress (cases 35, 36, 37, 38, 41, and 42) all exhibit a reduction with the optimized twists. However, cases with mid-to-low stresses present more mixed results, with both the shear force and bending moment showing increases in 25 cases and decreases in 15 cases.

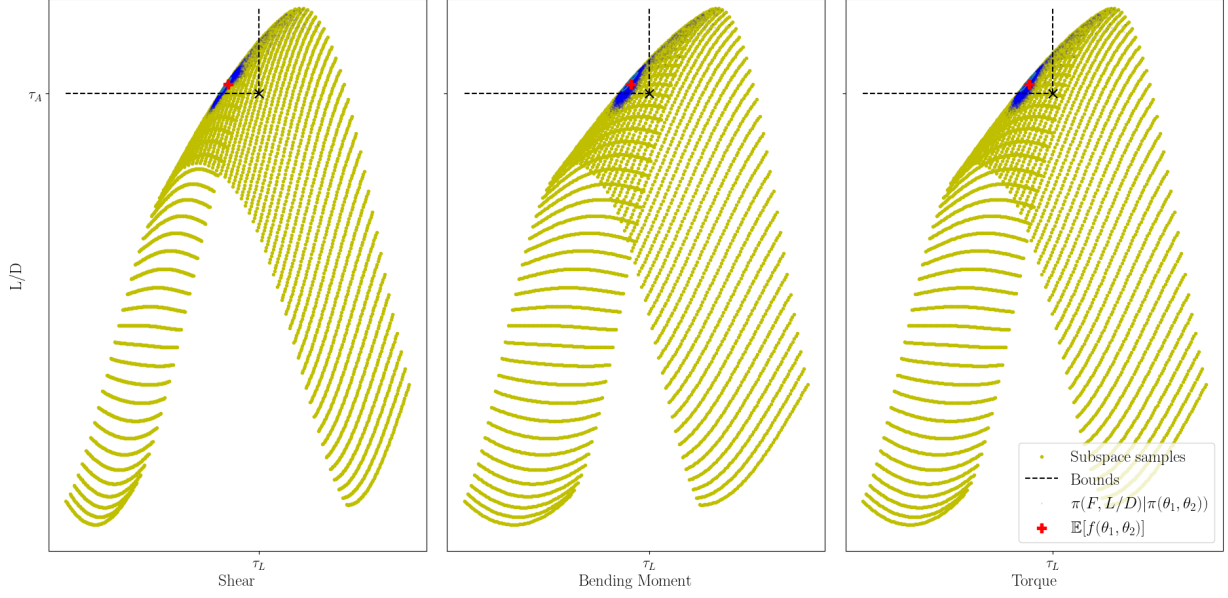


Fig. 5 Results of optimization where the posterior predictive distribution $\pi(F, L/D) | \pi(\theta_1, \theta_2)$ is plotted over samples of the reachable region subspace.

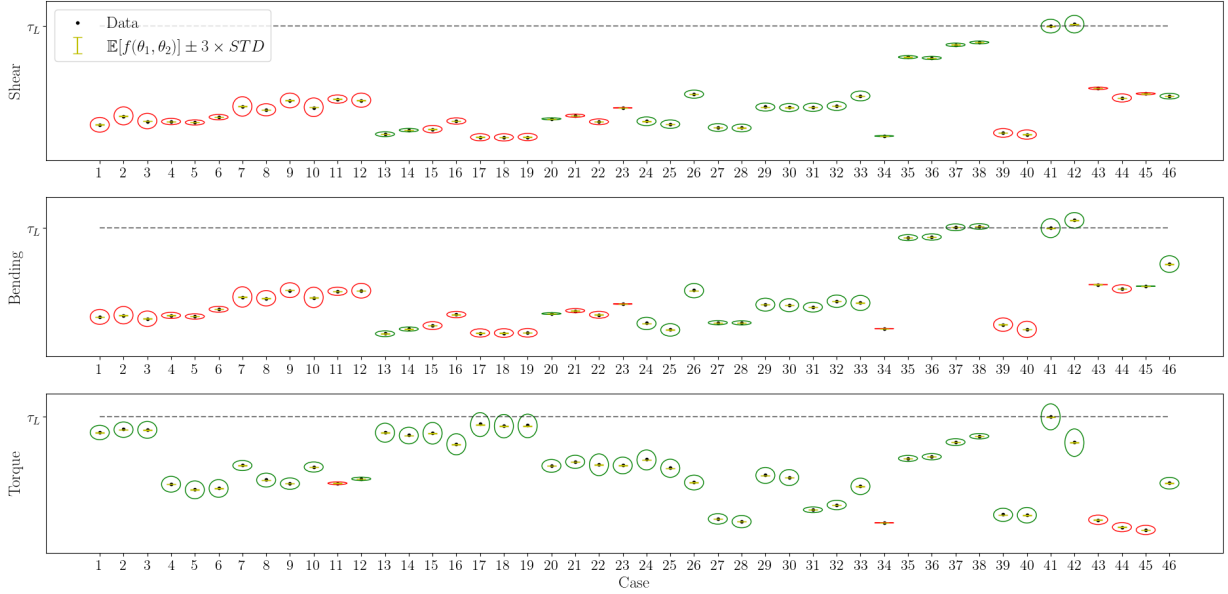


Fig. 6 Observed load responses at the wing root across all flight envelope cases are plotted alongside the expected values of the responses, as predicted by the surrogate model with the optimized twists.

V. Conclusion

This paper presents an approach for using reduced surrogate modelling techniques which can be integrated with industrial analysis tools and the overall design process of aerostructures. The significant contribution of this study lies in the application of the proposed robust design framework to real-world case studies produced by Airbus. The results from the two case studies exemplify the potential use cases and nature of the information produced by the proposed framework, and highlights further areas of study, such as, investigating the individual contributions of the surrogate model, marginalization, and response negative correlations, as well as, further manipulation of the likelihood function to address biases. Furthermore, results regarding the generation of posterior predictive distributions across the entire set of flight envelopes and wing spans via the reduced order multi-output Gaussian process surrogate model is not presented. Further work would produce and analyze graphs that have practical use for structural design engineers for critical flight envelope e.g plotting the posterior bending moment against span to assess the likelihood of exceeding collapse moments. This would require further investigation into the expansion of the likelihood functions to include risk terms for design failure.

Acknowledgement

Airbus' support is acknowledged for providing the aircraft load and aerodynamic datasets. AK acknowledges the funding provided by EPSRC (project number EP/V055577/1) to support this research.

References

- [1] Sudret, B., Marelli, S., and Wiart, J., "Surrogate models for uncertainty quantification: An overview," *2017 11th European Conference on Antennas and Propagation, EUCAP 2017*, 2017, pp. 793–797. <https://doi.org/10.23919/EUCAP.2017.7928679>.
- [2] Alizadeh, R., Allen, J. K., and Mistree, F., "Managing computational complexity using surrogate models: a critical review," *Research in Engineering Design* 2020 31:3, Vol. 31, 2020, pp. 275–298. <https://doi.org/10.1007/S00163-020-00336-7>, URL <https://link.springer.com/article/10.1007/s00163-020-00336-7>.
- [3] Iorga, L., Malmedy, V., Stodieck, O., Loxham, J., and Coggon, S., "Preliminary Sizing Optimisation of Aircraft Structures - Industrial Challenges and Practices," *6th Aircraft Structural Design Conference, At: Bristol, UK*, 2018.
- [4] Brunton, S. L., Kutz, J. N., Manohar, K., Aravkin, A. Y., Morgansen, K., Klemisch, J., Goebel, N., Buttrick, J., Poskin, J., Blom-Schieber, A. W., Hogan, T., and McDonald, D., "Data-driven aerospace engineering: Reframing the industry with machine learning," *AIAA Journal*, Vol. 59, 2021. <https://doi.org/10.2514/1.J060131>.
- [5] Clainche, S. L., Ferrer, E., Gibson, S., Cross, E., Parente, A., and Vinuesa, R., "Improving aircraft performance using machine learning: A review," *Aerospace Science and Technology*, Vol. 138, 2023, p. 108354. <https://doi.org/10.1016/j.ast.2023.108354>, URL <https://doi.org/10.1016/j.ast.2023.108354>.
- [6] Draper, D., "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 45–70. <https://doi.org/10.1111/j.2517-6161.1995.tb02015.x>.
- [7] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T., "Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors)," *Statistical Science*, Vol. 14, No. 4, 1999, pp. 382–417. <https://doi.org/10.1214/ss/1009212519>, URL <http://projecteuclid.org/journals/statistical-science/volume-14/issue-4/Bayesian-model-averaging--a-tutorial-with-comments-by-M/10.1214/ss/1009212519.full>, publisher: Institute of Mathematical Statistics.
- [8] Yin, J., Medellín-Azuara, J., Escrivá-Bou, A., and Liu, Z., "Bayesian machine learning ensemble approach to quantify model uncertainty in predicting groundwater storage change," *Science of the Total Environment*, Vol. 769, 2021. <https://doi.org/10.1016/j.scitotenv.2020.144715>.
- [9] Álvarez, M. A., Rosasco, L., and Lawrence, N. D., "Kernels for Vector-Valued Functions: A Review," *Foundations and Trends® in Machine Learning*, Vol. 4, No. 3, 2012, pp. 195–266. <https://doi.org/10.1561/22000000036>.
- [10] Konomi, B. A., and Lin, G., "LOW-COST MULTI-DIMENSIONAL GAUSSIAN PROCESS WITH APPLICATION TO UNCERTAINTY QUANTIFICATION," *International Journal for Uncertainty Quantification*, Vol. 5, 2015, pp. 375–392. <https://doi.org/10.1615/INT.J.UNCERTAINTYQUANTIFICATION.2015007941>, URL <https://www.dl.begellhouse.com/journals/52034eb04b657aea,6976a15960773171,4a5168bd7f15bc31.html>.

- [11] Lin, Q., Hu, J., Zhou, Q., Cheng, Y., Hu, Z., Couckuyt, I., and Dhaene, T., “Multi-output Gaussian process prediction for computationally expensive problems with multiple levels of fidelity,” *Knowledge-Based Systems*, Vol. 227, 2021, p. 107151. <https://doi.org/10.1016/j.knosys.2021.107151>, URL <https://doi.org/10.1016/j.knosys.2021.107151>.
- [12] Liu, Y., Wang, D., Sun, X., Liu, Y., Dinh, N., and Hu, R., “Uncertainty quantification for Multiphase-CFD simulations of bubbly flows: a machine learning-based Bayesian approach supported by high-resolution experiments,” *Reliability Engineering & System Safety*, 2021, p. 107636. <https://doi.org/10.1016/j.ress.2021.107636>, URL [https://www.sciencedirect.com.abc.cardiff.ac.uk/science/article/pii/S0951832021001770#bib0058](https://www.sciencedirect.com/abc.cardiff.ac.uk/science/article/pii/S0951832021001770#bib0058).
- [13] Manfredi, P., “Probabilistic Uncertainty Quantification of Microwave Circuits Using Gaussian Processes,” *IEEE Transactions on Microwave Theory and Techniques*, Vol. 71, 2023, pp. 2360–2372. <https://doi.org/10.1109/TMTT.2022.3228953>.
- [14] Roma, G., Maio, F. D., Bersano, A., Pedroni, N., Bertani, C., Mascari, F., and Zio, E., “A Bayesian framework of inverse uncertainty quantification with principal component analysis and Kriging for the reliability analysis of passive safety systems,” *Nuclear Engineering and Design*, Vol. 379, 2021, p. 111230. <https://doi.org/10.1016/J.NUCENGDES.2021.111230>.
- [15] Grbovic, A., and Rasuo, B., “FEM based fatigue crack growth predictions for spar of light aircraft under variable amplitude loading,” *Engineering Failure Analysis*, Vol. 26, 2012, pp. 50 – 64. <https://doi.org/10.1016/j.engfailanal.2012.07.003>.
- [16] Spalart, P. R., and Venkatakrishnan, V., “On the role and challenges of CFD in the aerospace industry,” *Aeronautical Journal*, Vol. 120, 2016, pp. 209 – 232. <https://doi.org/10.1017/aer.2015.10>.
- [17] Careri, F., Khan, R. H., Todd, C., and Attallah, M. M., “Additive manufacturing of heat exchangers in aerospace applications: a review,” *Applied Thermal Engineering*, Vol. 235, 2023, p. 121387. <https://doi.org/10.1016/j.applthermaleng.2023.121387>.
- [18] Kroll, N., Abu-Zurayk, M., Dimitrov, D., Franz, T., Führer, T., Gerhold, T., Görtz, S., Heinrich, R., Ilic, C., Jepsen, J., Jägersküpper, J., Kruse, M., Krumbein, A., Langer, S., Liu, D., Liepelt, R., Reimer, L., Ritter, M., Schwöppe, A., Scherer, J., Spiering, F., Thormann, R., Togiti, V., Vollmer, D., and Wendisch, J. H., “DLR project Digital-X: towards virtual aircraft design and flight testing based on high-fidelity methods,” *CEAS Aeronautical Journal*, Vol. 7, 2016, pp. 3 – 27. <https://doi.org/10.1007/s13272-015-0179-7>.
- [19] Kafkas, A., Kilimtzidis, S., Kotzakolios, A., Kostopoulos, V., and Lampeas, G., “Multi-fidelity optimization of a composite airliner wing subject to structural and aeroelastic constraints,” *Aerospace*, Vol. 8, 2021, p. 398. <https://doi.org/10.3390/aerospace8120398>.
- [20] Kundu, A., Matthies, H., and Friswell, M., “Probabilistic optimization of engineering system with prescribed target design in a reduced parameter space,” *Computer Methods in Applied Mechanics and Engineering*, Vol. 337, 2018, pp. 281 – 304. <https://doi.org/https://doi.org/10.1016/j.cma.2018.03.041>, URL <http://www.sciencedirect.com/science/article/pii/S0045782518301658>.
- [21] Walton, D. B., Featherston, C. A., Kennedy, D., and Kundu, A., “Inverse design under uncertainty with surrogate models,” *Journal of Physics: Conference Series*, Vol. 2647, 2024, p. 212008. <https://doi.org/10.1088/1742-6596/2647/21/212008>, URL <https://iopscience.iop.org/article/10.1088/1742-6596/2647/21/212008https://iopscience.iop.org/article/10.1088/1742-6596/2647/21/212008/meta>.
- [22] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., “Bayesian Data Analysis,” *Bayesian Data Analysis*, 1995. <https://doi.org/10.1201/9780429258411>, URL <https://www.taylorfrancis.com/books/mono/10.1201/9780429258411>.
- [23] Brooks, S., Gelman, A., Jones, G. L., and Meng, X. L., “Handbook of Markov Chain Monte Carlo,” *Handbook of Markov Chain Monte Carlo*, 2011, pp. 1–592. <https://doi.org/10.1201/b10905>.
- [24] Rasmussen, C. E., *Gaussian processes for machine learning*, Adaptive computation and machine learning, MIT Press, Cambridge, Mass., 2006.
- [25] Alvarez, M. A., Rosasco, L., and Lawrence, N. D., “Kernels for Vector-Valued Functions: a Review,” *arXiv:1106.6251 [cs, math, stat]*, 2012. URL <http://arxiv.org/abs/1106.6251>, arXiv: 1106.6251.
- [26] Hoffman, M. D., and Gelman, A., “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo,” *Journal of Machine Learning Research*, Vol. 15, 2011, pp. 1593–1623. URL <https://arxiv.org/abs/1111.4246v1>.
- [27] Hanada, M., and Matsuura, S., *Other Useful Algorithms*, Springer Nature Singapore, Singapore, 2022, Chap. 4, pp. 75–112. https://doi.org/10.1007/978-981-19-2715-7_5, URL https://doi.org/10.1007/978-981-19-2715-7_5.