EXPERIMENTAL DESIGN FOR AERODYNAMIC DATASET MODELLING USING BAYESIAN NEURAL NETWORKS AND BALD SCORING

Ferdinand Dunkes*, Christian Breitsamter*
* Technical University of Munich, Chair of Aerodynamics and Fluid Mechanics, Boltzmannstr. 15, 85748 Garching, Germany

Abstract

Bayesian Experimental Design (BED) provides a practical framework for selecting an experimental design that maximizes the expected usage. Within the scope of the present work, we apply the BED paradigm to high-dimensional aerodynamic stability and control datasets. A Bayesian Neural Network (BNN) serves as a surrogate model of an initial dataset, approximating the posterior distribution via mean-field Variational Inference (VI). The inherent uncertainty quantification capabilities of the BNN enable the decomposition of uncertainty into aleatoric and epistemic components utilizing a dual-head architecture. Leveraging this, the Bayesian Active Learning by Disagreement (BALD) method provides a scoring criterion to estimate the Expected Information Gain (EIG) of hypothetical, unlabeled data points. The candidate pool is generated using a Sobol sequence. In the context of Aerodynamic Dataset Modelling (ADM), a prudent design of new experiments is driven by economic constraints. Data acquisition techniques predominantly focus on Wind-Tunnel Testing (WTT), supplemented by numerical simulations and flight testing. The proposed methodology aims to iteratively identify the subsequent optimal measurements, thereby optimizing the test matrix setup. A greedy plus diversity strategy ensures comprehensive coverage of the input space while maintaining a high EIG. To evaluate the performance and applicability of the implemented methods, data from WTT campaigns of the DLR-F17 and DLR-F19 Unmanned Aerial Vehicles (UAVs), commonly known as SACCON, is utilized. This configuration features a highly swept, low-observable lambda wing planform without a separate empennage, representing a high-agility aircraft configuration. Results demonstrate that the introduced approach reliably identifies potential reductions in data acquisition needs, thus decreasing costs and resource demands of WTT campaigns, numerical computations, or flight testing, while maintaining the predictive accuracy of the BNN surrogate model.

Keywords

Bayesian Neural Network, Uncertainty Quantification, Bayesian Active Learning by Disagreement, Bayesian Experiment Design, Adaptive Sampling, Test Matrix Optimization

NOMENCLATURE

Subscripts

al	Aleatoric uncertainty
cand	Candidate sample
ep	Epistemic uncertainty
l	Rolling moment
LIB	Left inboard
LOB	Left outboard
m	Pitching moment
n	Yawing moment
new	New sample
ref	Reference
RIB	Right inboard
ROB	Right outboard
$root$	Wing root
$*$	Prediction
tip	Wing tip
tot	Total uncertainty
X	Axial force
Y	Side force
Z	Normal force

Symbols

α	Angle of attack	[°]
β	Angle of sideslip	[°]
b	Wingspan	[m]
β	Annealing factor	

doi: 10.25967/650293

C	Coefficient	
c	Chord length	[m]
$\mathcal{D}$	Dataset	
$\mathbb{E}$	Expectation	
η	Control surface deflection	[°]
$f(\cdot)$	Function	
$\mathbb{H}$	Entropy	
$\mathbb{I}$	Mutual Information	[bits]
$\mathcal{L}$	Evidence Lower Bound (ELBO)	
λ	Diversity penalty	
m	Mass	[kg]
Ma	Mach number	
μ	Mean	
$\mathcal{N}$	Normal distribution	
$p(\cdot)$	Probability	
ϕ	Variational parameter	
$q(\cdot)$	Posterior approximation	
$\mathbb{R}$	Real numbers	
σ	Standard deviation	
σ^2, $\mathbb{V}$, Var	Variance	
θ	Network parameters	
φ	Sweep angle	[°]
w	Weight	
$\mathbf{x}$	Input vector	
$\mathcal{X}$	Observation/sample space	
$\mathbf{y}$	Output vector	

Acronyms

ADM	Aerodynamic Dataset Model
AoA	Angle of Attack
AoS	Angle of Sideslip
BALD	Bayesian Active Learning by Disagreement
BED	Bayesian Experimental Design
BNN	Bayesian Neural Network
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V.
EIG	Expected Information Gain
ELBO	Evidence Lower Bound
FCS	Flight Control System
KDE	Kernel Density Estimate
KL	Kullback-Leibler (divergence)
MC	Monte Carlo
MI	Mutual Information
MSE	Mean Square Error
NLL	Negative Log Likelihood
FNN	Feed-forward Neural Network
SACCON	Stability And Control CONfiguration
S and C	Stability and Control
UAV	Unmanned Aerial Vehicle
VI	Variational Inference
WTT	Wind-Tunnel Testing

1. INTRODUCTION

Accurate and sufficient knowledge of the aerodynamic characteristics of next-generation high-agility aircraft systems and Unmanned Aerial Vehicles (UAV) is essential during the development phase throughout a multitude of disciplines. Additionally, a Flight Control System (FCS) is required to offer care-free handling qualities in vastly unstable flight conditions. To leverage its full potential, the FCS must be designed as closely as possible to the aircraft's limits and capabilities. Therefore, a continuous, accurate, and comprehensive Aerodynamic Dataset Model (ADM) is generated.

The ADM is typically based on Wind-Tunnel Testing (WTT) data and shall cover the entire flight envelope, all relevant flight conditions, control surface settings, and respective combinations. While indispensable, traditional WTT remains time-consuming and resource-intensive. Maintaining cost-effectiveness while reducing time to generate is a key challenge in aerodynamic dataset modeling. Due to clearance, safety, and regulatory requirements, an ADM must precisely reflect real-world aerodynamic behavior. However, the underlying aerodynamic phenomena necessitate handling high-dimensional input spaces and coping with intensive nonlinear dependencies. Quantifying data and model uncertainty becomes of paramount importance in this context. In order to meet these challenges, modern data-driven engineering approaches, such as surrogate modeling, are increasingly being investigated for their potential to improve the efficiency of the ADM generation process, as in [1] and [2].

In many engineering applications, active learning is not employed as an abstract sampling criterion, but rather as an iterative workflow integrating surrogate model training, uncertainty quantification, candidate selection, and dataset augmentation. Such closed-loop pipelines have been successfully applied in

computational fluid dynamics [3] or variable-fidelity modelling [4], where the primary objective is the efficient allocation of expensive simulation or experimental resources. A closely related perspective is provided by applied Bayesian experimental design [5]. In this context, schematic representations of the experimental loop are commonly used to highlight how model updates, uncertainty assessment, and experimental planning interact under practical constraints. In aerodynamic dataset generation, the application of such adaptive strategies is complicated by the structured nature of test matrices, high-dimensional input spaces, and operational constraints. Measurements are typically conducted along polar sweeps or Mach-number campaigns, rendering point-wise experimental design strategies impractical. Consequently, active learning approaches must be embedded into workflow-oriented concepts that respect both physical and operational constraints.

This work applies a Bayesian paradigm to WTT data acquired for the UAV platforms DLR-F17 and its successor, DLR-F19, commonly known as the Stability And Control Configuration (SACCON). We propose a Bayesian Neural Network (BNN) approach due to its inherent capabilities for quantifying uncertainty. In order to estimate the posterior distribution, statistical Variational Inference (VI) is employed. Bayesian Active Learning by Disagreement (BALD) provides a foundation to estimate the Expected Information Gain (EIG) from new, unlabeled sample points. During the development process, the WTT measurement settings that yield the most significant gain in predictive accuracy can be iteratively identified. For new measurements, the proposed approach suggests the next steps in data acquisition if desired. This enables data acquisition, which is tailored to the specific needs of the ADM generation process. Ultimately, these methods promise to reduce the cost of experimental campaigns while maintaining the fidelity required for reliable aerodynamic predictions.

Summarizing the targeted contributions of this work, we aim to:

- Develop a BNN-based framework for high-dimensional aerodynamic datasets, capable of uncertainty quantification and quasi-active learning.
- Integrate BALD to estimate the EIG from new, unlabeled sample points, guiding the selection of future measurements.
- Evaluate the performance and plausibility on the DLR-F17 and DLR-F19 aerodynamic datasets, demonstrating the effectiveness of the proposed methodology qualitatively.

2. EXPERIMENTAL SETUP AND DATA

We select a highly swept, low-observable UAV configuration to provide a test case and verify and validate the proposed methodology. Within the NATO STO AVT-201 (Extended Assessment of Reliable Stability and Control Prediction Methods for NATO Air Vehicles) task group, which builds upon the AVT-161 task group, the SACCON platform was developed and intensively investigated. A payload mass of $m = 2000\,\mathrm{kg}$ was defined in addition to other boundary conditions, such as design range, operational and certification requirements, as part of a dedicated design mission [6]. The general work package includes determining the ability of computational methods to accurately predict static and dynamic stability characteristics of air and sea vehicles [6]. During the AVT-201 process, additional WTT models were developed, allowing for research into the influence and effects of control surfaces. Consequently, a stability and control (S&C) database was constructed using experimental and numerical methods.

The SACCON UAV features a lambda planform with a leading-edge sweep angle of $\varphi = 53°$. Fig. 1 illustrates an overview of the configuration. As visualized, three main sections can be distinguished: the fuselage, the wing, and the wingtip section. Ensuring low observability, the leading edges are arranged parallel to the wing trailing edges. Accordingly, the wingtips are designed parallel to the respective trailing edges of the fuselage. The overall flying-wing concept is summarized in [6]. Three different airfoils define the UAV's contour, which are located in the symmetry plane c_{root}, at the transition from fuselage to wing c_{ref}, and at the outer wing zone c_{tip}. An additional twist of $5°$ around the leading edge is applied, reducing aerodynamic loads and shifting the onset of flow separation [7]. Consequently, a sharp leading edge contour transitions to a more rounded shape in the spanwise direction. All six available control surfaces η, located at the wingtips and trailing edges, are depicted in Fig. 1, including their respective nomenclature.

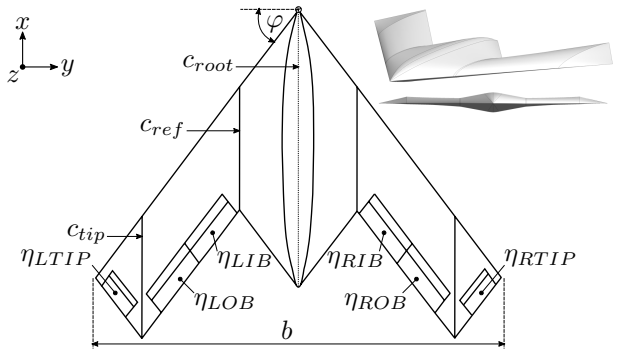


FIG 1. SACCON features a lambda-shaped planform and six control surfaces with the parameters root chord $c_{root} \approx 1.06\,m$, reference chord $c_{ref} \approx 0.48\,m$, wingtip chord c_{tip}, and wingspan $b \approx 1.53\,m$. Graphics adapted from [8] and [9].

Force and moment measurements relevant to this work were acquired for three different wind-tunnel models in four wind tunnels, spanning subsonic and transonic regimes. A review of this research can be found in [8]. Before applying all data from all models, geometrical alterations, and wind tunnels to the methodology presented hereinafter, we conduct a literature review comprising an analysis of aerody-

namic similarity between the alterations. Comparing the acquired measurements of forces and moments for different tunnels, models, and test campaigns, the authors of [8] describe a good general agreement. Two different control surface depths and various hinge line locations were investigated on the DLR-F19 wind-tunnel model. Both parameters showed no considerable effect on the integral coefficients, disregarding subtle changes in pitching moment at an Angle of Attack (AoA) $8° \leq \alpha \leq 10°$ [8] and [10].

In order to assess the suitability and applicability of the SACCON aerodynamic S&C database for the presented use case, a summary of the governing aerodynamic characteristics is executed. Moreover, understanding the UAV's flow phenomena, aerodynamic trends, and dependencies facilitates an improved interpretation of the predictive surrogate model output.

Vortex breakdown alongside instabilities in rolling or yawing requires precise prediction. Therefore, due to the highly nonlinear nature of such phenomena, special attention should be paid to affected regions. Via an integrated approach, described in references [7] and [11], the formation and breakdown of three defining vortex systems are illustrated. At low AoA, the apex vortex and the thickness vortex are positioned adjacently, while a distinct tip vortex forms near the wingtip, separated by a clearly defined zone of flow detachment. With increasing AoA, this separation zone progressively shifts upstream, resulting in all three vortical structures aligned in proximity without clear separation beyond the onset position. At higher AoA, the flow reorganizes into two dominant vortex structures. At approximately $\alpha > 20°$, equilibrium is established before vortex breakdown can be observed from $\alpha \approx 25°$. Between $16° \leq \alpha \leq 20°$, a temporary reversal in pitching moment occurs, attributed to changes in surface suction induced by the evolving vortex system.

Linear dependencies in control surface interactions or between flight conditions and control surface deflections yield a convenient advantage: Observed effects can be superpositioned and, thus, resemble an expectedly less complex challenge for the BNN model. By identifying these regions of the flight envelope, a consistent and balanced test scenario can be derived in section 5. In reference [10], the effect of inboard flap deflection in low-speed tests ($\mathrm{Ma}\,0.15$) is studied. Due to the decreased lever arm, significant changes only occur in the moment coefficients for large deflections ($\eta_{IB} > 20°$). Forces exhibit no considerable influence. The inboard flap ensures slightly more efficiency than the outboard flap. This holds for high-speed WTT. Superpositioning is valid solely for single deflections per side throughout the entire AoA range, capturing nonlinear phenomena as well. Adjacent flaps exhibit extensive interactions except for drag C_{Drag} and side force C_Y. Neither full span control surface deflections can be superposed. Highly nonlinear, stochastic effects resulting from control surface deflection are evident for increased AoAs in transonic measurements.

Based on the aforementioned aerodynamic phenomena, we can construct a training and test set, as well as a validation metric, as introduced in section 5. As commonly applied in Machine Learning (ML) tasks, a data split of 80% training and 20% testing purposes ensures robust model evaluation [12]. The dataset consists of 261 polars with varying discrete measurements, totaling approximately $n \approx 26.000$ for $i = 1, ..., n$ samples. As elaborated below, the partitioning occurs based on polar sample collections for data coverage reasons. The BNN model attempts to learn a mapping $\boldsymbol{y} = f(\boldsymbol{x})$ of $m = 7$ for $j = 1, ..., m$ features $\boldsymbol{x}_1, ..., \boldsymbol{x}_m$ towards six targets $\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_6$ represented by the integral aerodynamic coefficients in the body-fixed coordinate system:

(1)
$$\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, ..., \boldsymbol{x}^{(n)}\}^T = = \begin{Bmatrix} \alpha^{(1)} & \beta^{(1)} & \mathrm{Ma}^{(1)} & \eta_1^{(1)} & \cdots & \eta_4^{(1)} \\ \alpha^{(2)} & \beta^{(2)} & \mathrm{Ma}^{(2)} & \eta_1^{(2)} & \cdots & \eta_4^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{(n)} & \beta^{(n)} & \mathrm{Ma}^{(n)} & \eta_1^{(n)} & \cdots & \eta_4^{(n)} \end{Bmatrix},$$

and

(2)
$$\{\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(n)}\}^T = = \begin{Bmatrix} C_X^{(1)} & C_Y^{(1)} & C_Z^{(1)} & C_l^{(1)} & C_m^{(1)} & C_n^{(1)} \\ C_X^{(2)} & C_Y^{(2)} & C_Z^{(2)} & C_l^{(2)} & C_m^{(2)} & C_n^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_X^{(n)} & C_Y^{(n)} & C_Z^{(n)} & C_l^{(n)} & C_m^{(n)} & C_n^{(n)} \end{Bmatrix}.$$

$\eta_1, ..., \eta_4$ corresponde to the control surface deflections $\eta_{LOB}, \eta_{LIB}, \eta_{RIB}, \eta_{ROB}$.

Z-score normalization is applied to ensure consistent feature scaling, comparability, and interpretability, while stabilizing convergence [13]. In order to grasp the underlying patterns and relationships within the design domain $\mathcal{D} \subset \mathbb{R}^k$, data coverage and distribution are analysed using the violin plots in Fig. 2. Test and training sets possess similar Kernel Density Estimates (KDE), ensuring a valid performance assessment.

Generally, a heterogeneous distribution for the input features can be observed. Multimodality is only pronounced for the flight condition originating from sweeps in AoA and Angle of Sideslip (AoS) as well as multiple Ma clusters. Asymmetries, as reflected in skewed tails, amplify extrapolative demands at the distribution boundaries, leading to increased uncertainty. Additionally, we can identify a strong focus on $\beta \approx \{-5°, 0°, 5°\}$, leaving regions outside of $\beta \in [-5°, 5°]$ underrepresented. Leveraging the axis-symmetric design of SACCON, only port-side flap settings exhibit measured data for varying clusters. The illustrated data coverage and distribution stem from a conventional WTT campaign, manifesting

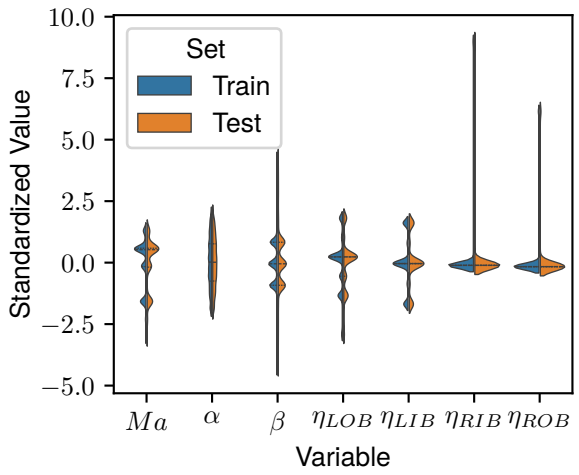


FIG 2. Violin plots of the dataset input features displayed for training and test set illustrating data coverage and distribution via KDE.

the need for BED approaches to further improve surrogate model input data.
Furthermore, we use a diagonal correlation matrix to visualize the intensity and direction of linear relationships between pairs of features and targets. The correlation matrix is displayed in Fig. 3.

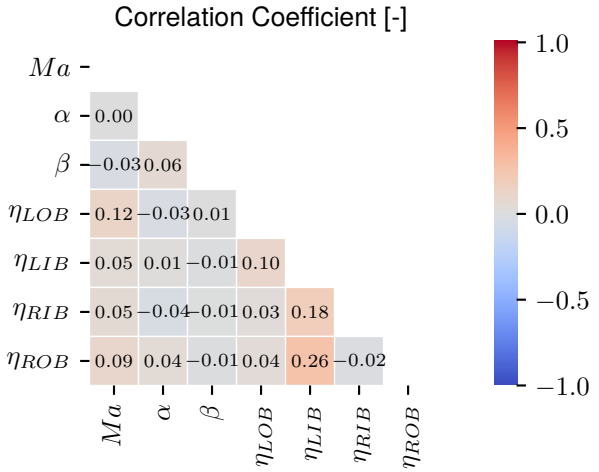


FIG 3. Heatmap of the dataset features displaying a diagonal correlation matrix with correlation coefficients $r \approx 1$ representing linear or vice versa nonlinear relations.

All three features representing the flight condition exhibit no significant linear correlations. Weak linear correlations of η_{LIB} with η_{ROB}, η_{RIB}, as well as η_{LOB} indicate measurements at individual flap settings carrying individual information. Strong positive correlation coefficients attest to symmetric flap deflection, whereas negative coefficients imply differential settings. However, this supports the reliability of the database used for the superposition analysis in section 2.

3. METHODOLOGY

To create a continuous, comprehensive surrogate model of the above-described dataset, we employ a BNN approach with Variational Inference (VI) to approximate the posterior distribution. BALD enables the quantification of the EIG of hypothetical unlabeled additional data points. The following subsections introduce the theoretical background and workflow of the proposed methodology.

3.1. Bayesian Neural Networks

The structure of a BNN closely resembles that of a standard feed-forward Neural Network (FNN). However, a FNN can be prone to overfitting and lacks the ability to quantify uncertainty in its predictions [14]. Using the Bayesian formulation, the deterministic network parameters of the FNN θ are treated as random variables with a prior distribution $p(\theta)$. Given a training dataset $\mathcal{D}$, Bayes' Theorem provides a rule to update the prior beliefs about a distribution according to

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}, \tag{3}$$

where the posterior $p(\theta|\mathcal{D})$ equals the prior $p(\theta)$ times the likelihood $p(\mathcal{D}|\theta)$ over the evidence $p(\mathcal{D})$. The evidence is often intractable, as it requires integrating over all possible parameter values θ [15]. Therefore, we approximate the posterior with variational parameters ϕ by introducing a variational distribution $q_\phi(\theta)$. The posterior approximation $q_\phi(\theta)$ is modeled as a factorized Gaussian, so that each weight and bias have parameters μ_j and ρ_j. The variance is parametrized via the softplus transform:

$$\sigma_j = \log(1 + \exp(\rho_j)). \tag{4}$$

Since both prior and variational posterior are Gaussian, the Kullback-Leibler (KL) divergence can be computed analytically in closed form, which yields the univariate case [16]. The difference of the model probability distribution to the true distribution is measured by the KL divergence:

$$\begin{aligned} &\mathrm{KL}(\mathcal{N}(\mu_q, \sigma_q^2) \| \mathcal{N}(\mu_p, \sigma_p^2)) = \\ &= \log \frac{\sigma_p}{\sigma_q} + \frac{\sigma_q^2 + (\mu_q - \mu_p)^2}{2\sigma_p^2} - \tfrac{1}{2}. \end{aligned} \tag{5}$$

To reduce the KL divergence between the true posterior and the variational distribution, the Evidence Lower Bound (ELBO) is maximized. The ELBO is defined as

$$\mathcal{L}(\phi) = \mathbb{E}_{q_\phi(\theta)}[\log p(\mathcal{D}|\theta)] - \mathrm{KL}(q_\phi(\theta) \| p(\theta)). \tag{6}$$

Bayes by Backprop provides a method to optimize the ELBO using Monte Carlo (MC) estimation of stochastic gradient descent by reparameterization of $\theta = t(\epsilon, \phi), \epsilon \sim \mathcal{N}(0, 1)$ [17]. The predictive distribution of a new input x^* is obtained by marginalizing

over the posterior by approximation via MC sampling:

$$p(y^\star|x^\star,\mathcal{D}) = \int p(y^\star|f_\theta(x^\star))q_\phi(\theta)d\theta$$

$$\Downarrow$$

$$p(y^\star|x^\star,\mathcal{D}) \approx \frac{1}{S}\sum_{s=1}^{S} p(y^\star|f_{\theta^{(s)}}(x^\star)),\ \theta^{(s)} \sim q_\phi(\theta)$$

To improve the computational efficiency, we utilize mini-batch optimization, splitting the training set $\mathcal{D}$ into M equally-sized mini-batches $\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_M$. β-annealing further improves stability and prevents poor local minima in early training stages by gradually introducing the KL term in the loss function [18]. From equation 6 follows:

$$\mathcal{L}(\phi) = \mathbb{E}_{q_\phi(\theta)}[\log p(\mathcal{D}|\theta)] - \beta \cdot \mathrm{KL}(q_\phi(\theta)\|p(\theta)). \tag{7}$$

3.2. Uncertainty Estimation

The total uncertainty, associated with the prediction of a BNN, encompasses two main components expressed by the law of total variance [19]:

$$\begin{aligned}\mathrm{Var}(y|x,\mathcal{D}) &= \\ &= \underbrace{\mathrm{Var}_{q_\phi(\theta)}[\mathbb{E}(y|x,\theta)]}_{\text{epistemic uncertainty}} + \underbrace{\mathbb{E}_{q_\phi(\theta)}[\mathrm{Var}(y|x,\theta)]}_{\text{aleatoric uncertainty}}.\end{aligned} \tag{8}$$

In order to account for both uncertainties in the heteroscedastic setting, we employ a dual-head likelihood formulation. Two outputs per target dimension emerge from this architecture, yielding a predictive mean $\mu_\theta(x)$ and a predictive log variance $\log\sigma_\theta^2(x)$.
Aleatoric uncertainty arises from data-driven uncertainty caused by inherent noise in the dataset. Consequently, high aleatoric uncertainty indicates the model's inability to capture the input-output relationship due to noisy or variable data. A dual-head architecture is conventionally applied to account for aleatoric uncertainty in regression tasks [20]. This structures the output layer with $2 \times O$ neurons, where O represents the number of desired outputs. Thus, the model implicitly learns aleatoric uncertainty, as the aerodynamic dataset does not directly contain information on noise.
In contrast, epistemic uncertainty captures the model's intrinsic limitations, emanating from factors such as limited data, insufficient training, or inadequate model complexity. Hence, a high epistemic uncertainty corresponds to incomplete learning, given the training set and the chosen model architecture. Acquiring more data in regions of increased epistemic uncertainty, alongside adapting the model complexity, mitigates this uncertainty contribution. To utilize this observation, we introduce possibilities to obtain the most informative new samples in the subsequent sections.

3.3. Bayesian Experimental Design

Within the scope of Bayesian Experimental Design (BED), we extend the above-described BNN surrogate model to incorporate the capability of identifying potential new input samples x_{new} that provide the most considerable reduction of epistemic uncertainty. For the use-case scenario motivated in section 1, we provide an instrument to estimate where to acquire additional labeled data. Formally, this next experiment is chosen by maximizing the EIG [21], [22]:

$$\begin{aligned}x_{\text{new}} &= \\ &= \arg\max_{x\in\mathcal{X}} \mathbb{E}_{y\sim p(y|x,\mathcal{D})}\Big[\mathrm{KL}\big(p(\theta \mid \mathcal{D},x,y)\,\|\,p(\theta \mid \mathcal{D})\big)\Big],\end{aligned} \tag{9}$$

Intuitively, we seek x_{new} where the expected posterior divergence after a hypothetical observation y_{new} is maximized. For every candidate x and all possible outcomes y, a posterior update $p(\theta \mid \mathcal{D},x,y)$ needs to be computed, which is computationally intractable [20].
For this reason, the EIG is reformulated as the conditional Mutual Information (MI) I between parameters θ and future observations y,

$$\begin{aligned}I[y,\theta \mid x,\mathcal{D}] &= \\ &= H[y \mid x,\mathcal{D}] - \mathbb{E}_{\theta\sim q(\theta)}\big[H[y \mid x,\theta]\big],\end{aligned} \tag{10}$$

with predictive entropy $H[y \mid x,\mathcal{D}]$ and the expected conditional entropy $\mathbb{E}_{\theta\sim q(\theta)}\big[H[y \mid x,\theta]\big]$. BALD takes advantage of this reformulation by selecting values that maximize the difference of these entropies representing the epistemic uncertainty.
The BNN outputs a mean $\mu_\theta(x)$ and predictive variance $\sigma_\theta^2(x)$ per output dimension d. From multiple posterior samples, a variance decomposition is obtained:

$$\mathrm{Var}_{\text{ep},d}(x) = \mathrm{Var}_\theta\big[\mu_\theta^{(d)}(x)\big], \tag{11}$$

$$\mathrm{Var}_{\text{al},d}(x) = \mathbb{E}_\theta\big[\sigma_\theta^{2,(d)}(x)\big], \tag{12}$$

$$\mathrm{Var}_{\text{tot},d}(x) = \mathrm{Var}_{\text{ep},d}(x) + \mathrm{Var}_{\text{al},d}(x). \tag{13}$$

From there, the entropies can be approximated by assuming a Gaussian distribution [23]:

$$H[y_d \mid x,\mathcal{D}] \approx \tfrac{1}{2}\log\big(2\pi e\,\mathrm{Var}_{\text{tot},d}(x)\big), \tag{14}$$

$$\mathbb{E}_\theta[H[y_d \mid x,\theta]] = \tfrac{1}{2}\log\big(2\pi e\,\mathrm{Var}_{\text{al},d}(x)\big). \tag{15}$$

Hence, the BALD score per output is phrased as

$$\mathrm{BALD}_d(x) \approx \tfrac{1}{2}\log\left(1+\frac{\mathrm{Var}_{\text{ep},d}(x)}{\mathrm{Var}_{\text{al},d}(x)}\right). \tag{16}$$

For D outputs, the BALD scores can be aggregated as

$$\mathrm{BALD}(x) = \sum_{d=1}^{D} w_d\,\mathrm{BALD}_d(x), \tag{17}$$

where $w_d \geq 0$ are user-defined output weights (default: $w_d = 1$) [21].

4. MODEL SETUP

When acquiring $k > 1$ new points from the candidate set simultaneously, simple top-k selection may yield redundant candidates. We employ a greedy acquisition strategy with diversity penalization, as described in references [24] and [25], to address this issue. Firstly, the highest-scoring points are selected. A diversity penalty based on proximity to already selected points is used for subsequent selections. If S is the set of already selected points and $d_{min}(x, S)$ is the minimum Euclidian distance to S, then each candidate's penalized score is

$$\tilde{s}(x) = s(x) - \frac{\lambda}{\epsilon + d_{min}(x, S)}, \tag{18}$$

where $s(x)$ is the original BALD score and λ is a tunable diversity weight controlling the trade-off between BALD score and diversity. ϵ represents a regularization constant, effectively preventing division by zero by setting it to minimal values, e.g., $\epsilon = 10^{-6}$. This approach enables comprehensive coverage of diverse regions of the input space. k alongside the generation of candidates $\mathcal{X}_{\text{cand}}$ represent a design choice and are driven by resource constraints, such as available budget or time for additional measurements. $\mathcal{X}_{\text{cand}}$ directly determines where the design can search for new samples. A problem-aware generation of candidates is therefore crucial. In the present work, we generate the candidate set using a Sobol low-discrepancy sequence as an extendable base, ensuring a space-filling distribution within the limits of the flight envelope. More sophisticated methods can account for aerodynamic, mechanical, or technical constraints. Surrogate constraints can be assigned to suspend infeasible WTT model control settings or to consider flight-scheduled control surfaces. A denser refinement yields a heterogeneous candidate distribution at boundaries or in zones of increased interest. A hybrid data pool can cover these considerations. For instance, the space-filling Sobol set serves as a basis, while boundary-boosting and physics-windows increase local grid refinement. History-aware BALD-proxies leverage information about regions of higher epistemic uncertainty in quasi-active learning scenarios. In summary, the greedy iteration picks x_{new} from the candidate pool according to

$$x_{\text{new}} = \arg \max_{x \in \mathcal{X}_{\text{cand}}} \tilde{s}(x). \tag{19}$$

Table 1 summarizes all relevant hyperparameters of the BNN model.
These parameters were determined via systematic parameter studies and subsequent manual refinement. As for the BED parametrization, we select $k = 1000$ new samples corresponding to 25 polars for a discrete step size of $\alpha = 1°$ within $-10° < \alpha < 30°$ from a candidate set of size $|\mathcal{X}_{\text{cand}}| = 400.000$.
Fig. 4 illustrates the overall workflow of the proposed methodology.

TAB 1. Hyperparameters of the BNN model.

Hyperparameter	Value
Learning Rate	$1 \cdot 10^{-4}$
Learning Rate (var-head)	$2 \cdot 10^{-5}$
Batch Size	256
Hidden Layers	3
Neurons per Layer	256, 256, 128
Activation Function	LeakyReLU
Prior Distribution	Layer-scaled Gaussian
Variational Distribution	$\mathcal{N}(\mu, \sigma^2), \sigma = \log(1 + \exp(\rho))$
Optimizer	Adam
β_{final}	0.28
$\beta_{warmup-frac}$	0.6
MC Samples (predict)	100
MC Samples (train)	25

Starting with an initially available training set $\mathcal{D}$, the BNN is trained as introduced in section 3.1. The obtained best state model $\hat{q}(\theta)$ is subsequently used to approximate the posterior samples. After a forward pass of all candidate points $\mathcal{X}_{\text{cand}}$, the predictive uncertainty per output can be decomposed into its epistemic, aleatoric, and total components as described in section 3.2. Furthermore, the entropies and hence the aggregated BALD score for each candidate are computed. To close the quasi-active learning loop, we select k new samples x_{new} from the candidate set using the *Greedy + Diversity* strategy. These samples are then labeled and added to the training set $\mathcal{D}$. The entire process can be iterated until a desired performance is achieved or available resources are exhausted.
The proposed flowchart is not intended to introduce a novel theoretical active learning criterion, but rather to formalize an engineering-oriented workflow for integrating Bayesian uncertainty quantification and information-based sample acquisition into aerodynamic dataset modelling.

5. RESULTS AND DISCUSSION

In order to evaluate the performance, effectiveness, and feasibility of the approach described in Fig. 4 regarding experimental design, we construct a validation case based on the SACCON platform. Considering the data coverage and distribution as well as combinations among the features, the left outboard flap η_{LOB} is defined as a dependent quantity for analysis as elaborated in section 2. η_{LOB} possesses the highest variability in measured settings out of all control surfaces. Due to its extensive impact on rolling moment, capturing the underlying patterns and relationships is of interest. Therefore, we focus the result analysis hereinafter on the planes defined in Tab 2

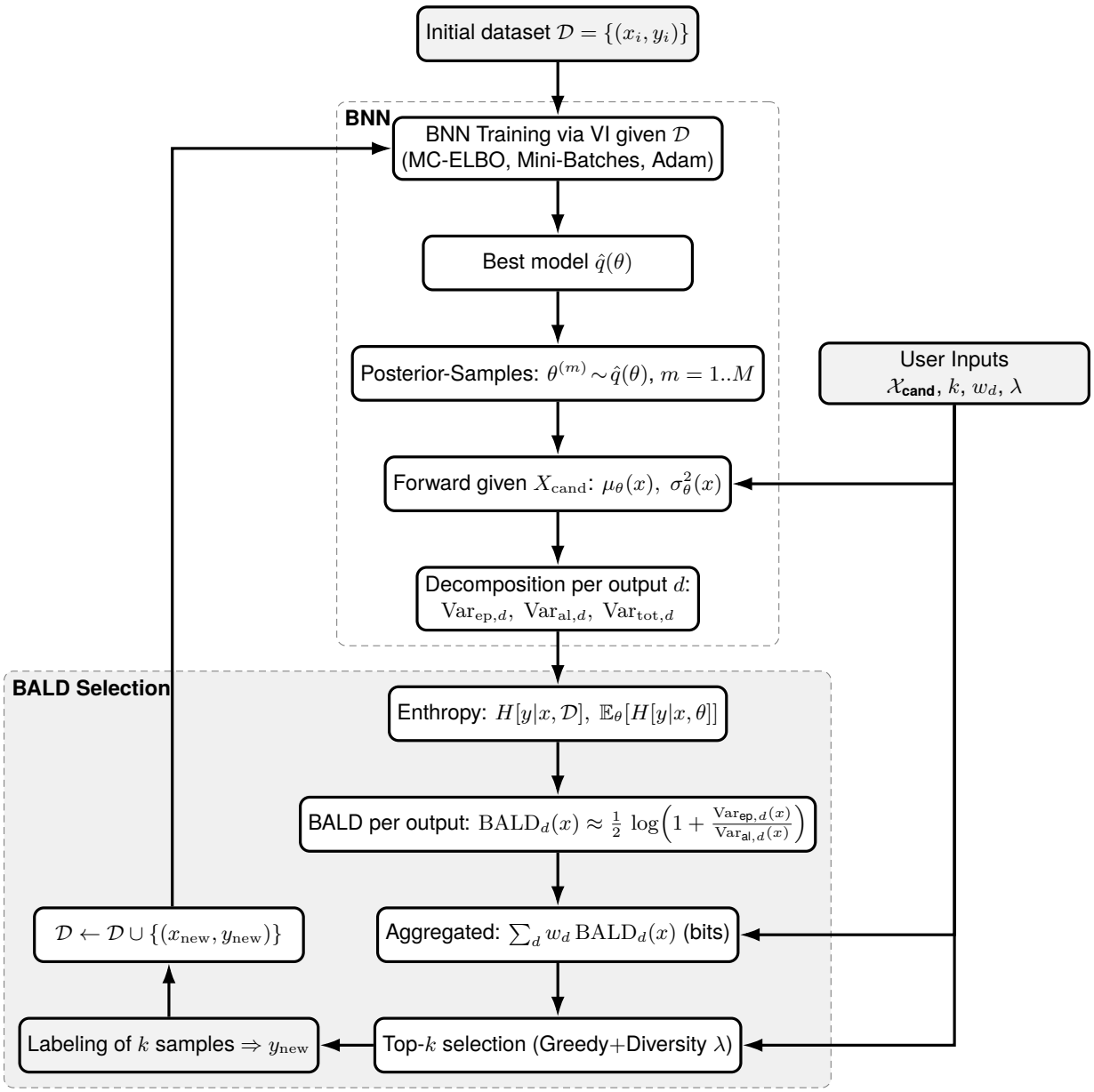


FIG 4. Overall workflow of the proposed BALD extended BNN methodology for experimental design.

within the multi-dimensional input space. A dash indicates the dependent features.

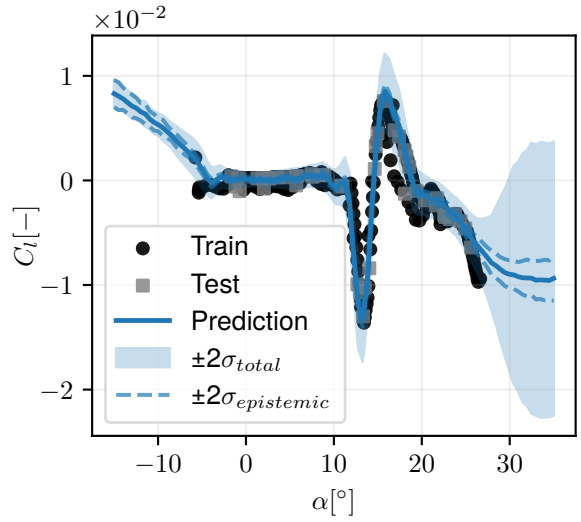
TAB 2. Validation case definition. Each row represents a 1 or 2-dimensional plane for result analysis. Dashes mark dependend quantities.

α	β	Ma	η_{LOB}	η_{LIB}	η_{RIB}	η_{ROB}
$0°$	$0°$	0.8	-	$0°$	$0°$	$0°$
-	$0°$	0.8	-	$0°$	$0°$	$0°$
$0°$	-	0.8	-	$0°$	$0°$	$0°$

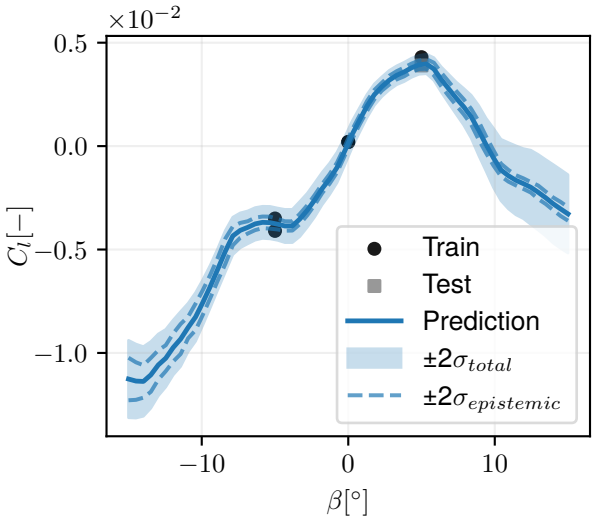
At first instance, we assess the BNN's predictive performance on the defined validation case. Fig. 5a and Fig. 5b illustrate the predicted mean and total as well as epistemic uncertainty of the rolling moment coefficient C_l alongside the training and test samples depending on AoA and AoS, respectively.

Both Fig. 5a and Fig. 5b can be characterized by a sufficient general fit of the training and test data by the posterior mean prediction in the depicted plane. All samples lie within the confidence interval. A Mean Squared Error (MSE) of $\text{MSE} = 0.0298$, normalized and Negative Log Likelihood (NLL) of $\overline{\text{NLL}} = -12.289$ substantiate this observation for the global model. The data underlying the prediction consist of two polars in proximity in the Ma domain with diverging characteristics at higher AoA. This poses challenges for the model locally, especially in the high AoA range, which are addressed by incorporating information from other dimensions of the design space.

The resulting decomposition of the used heteroscedastic model appears physically consistent. In less informed regions, the epistemic uncertainty component increases, while aleatoric values remain approximately unchanged. In particular, at $\alpha < -10°$,



(a) BNN prediction of C_l depending on AoA α.



(b) BNN prediction of C_l depending on AoS β.

FIG 5. BNN prediction for the rolling moment coefficient C_l including total σ^2_{tot} and epistemic uncertainty σ^2_{ep} bands alongside training and test samples for $\eta_{LOB} = 0°$. Ma $= 0.8$, the remaining parameters of the design space are set to zero.

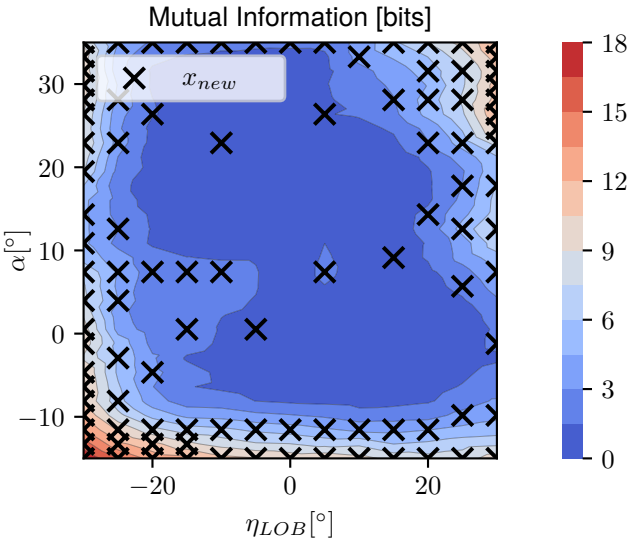


FIG 6. BALD scores depending on AoA α and left outboard flap deflection η_{LOB}. Ma $= 0.8$, the remaining parameters of the design space are set to zero.

and $\alpha > 25°$ as well as $\beta < -10°$, and $\beta > 10°$, aggravating σ^2_{ep} can be observed due to the lack of training information. At exactly these positions, the BALD extension is expected to identify and select candidates yielding a high EIG.

After ensuring a sufficient prediction accuracy, we interpret the BALD scores of the candidate set. The pivotal question for the presented use-case scenario is, for which exact input combinations does the BNN model benefit most from additional labeling? Fig. 6 depicts the BALD scores depending on AoA and deflection of the left outboard flap η_{LOB}, exemplarily. Herefrom, the most informative new experiments for the defined validation case can be deduced.

As expected from Fig. 5, high MI is present for combinations of η_{LOB} and α in boundary areas, high nonlinear aerodynamic regimes, or data-wise underpopulated regions. Additionally, the influence of the candidate pool can be clearly detected, including the restriction to a step size for flap deflections of $\Delta\eta = 5°$. As a consequence of Fig. 6, we preserve measurement points including their quantified information gain for the resulting ADM refinement. For instance, for a new experiment at $\alpha = 0°$, $\eta_{LOB} = \{-30°, -15°, -5°, 30°\}$ yields the most significant improvement, effectively reducing required data acquisition costs and time requirements by eliminating the necessity of homogenously measuring across a large amount of control surface settings.

To gain an understanding of the allocation of the new samples x_{new} selected from $\mathcal{X}_{\text{cand}}$, we visualize the rolling moment coefficient as a function of the left outboard flap in Fig. 7. This confirms the effectiveness of the proposed methodology in identifying the most informative new samples for improved predictive accuracy. As can be seen in Fig. 7, all selected new samples (dashed lines) are physically comprehensible. Only at $\eta_{LOB} = -5°$ a contradiction arises with the existing training sample. This is due to the fact that the specified number k of new samples must be reached, and, in this case, is forced by the penalty factor to be located at a low MI. This can be improved in future developments.

Provided that selected new samples featuring high BALD scores can be labeled, retraining an identically parametrized BNN offers a broader insight into the potential knowledge gain and its effect on predictive accuracy. Comparing the initial model with its successors, which contain the newly labeled data points, in a calibration plot provides a quantitative and reliable evaluation basis. Since this is outside the scope of the presented study, we will limit ourselves to conducting a qualitative analysis of the developed method to

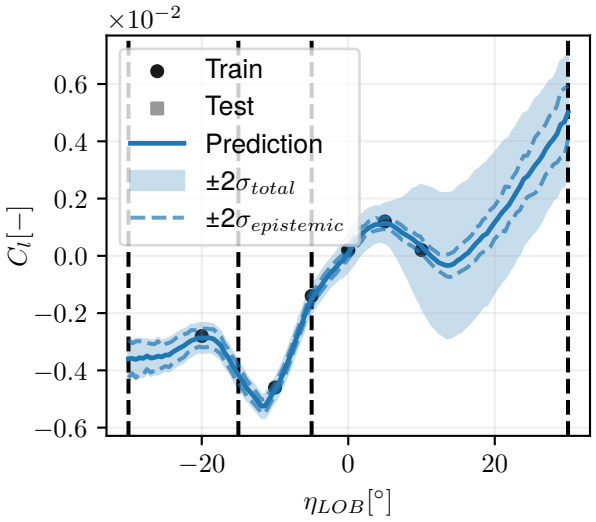


FIG 7. BNN prediction of C_l depending on left outboard flap deflection η_{LOB} including total and epistemic uncertainty bands alongside training, test and newly (dashed lines) acquired samples. $\mathrm{Ma} = 0.8$, the remaining parameters of the design space are set to zero.

address the introduced engineering problem. Analog to Fig. 6, new samples from the candidate pool are obtained primarily at boundary regions and areas lacking training samples. Approaching higher Mach numbers presumably implies steeper and more volatile gradients in the rolling moment coefficients due to increased nonlinearities in transonic aerodynamic conditions. Overall, we confirm the validity of the selected new samples, marked at the bottom of Fig. 7. Future work will focus on a quantified assessment of the achieved improvement.

6. CONCLUSION

In this work, we have developed and demonstrated the application of a Bayesian Active Learning by Disagreement-extended Bayesian Neural Network for modelling high-dimensional aerodynamic stability and control datasets.

While more rigorous batch acquisition strategies based on joint mutual information have been proposed in the active learning literature, their computational cost and implementation complexity often limit their applicability in high-dimensional engineering problems. The present work therefore, adopts a pragmatic batch selection strategy combining information-based scoring with diversity promotion, which has proven effective in practice.

The proposed methodology provides efficient capabilities to identify the most informative new samples by indirectly maximizing the Expected Information Gain, thereby improving predictive accuracy. We found that greedy acquisition and diversity penalization are indispensable to ensure comprehensive coverage of the input space. Moreover, we have successfully demonstrated the design-aware generation of candidate points, which problem-specific constraints can further refine. Leveraging the BNN's inherent uncertainty quantification capabilities, a resource and cost-efficient design of new experiments is enabled, providing significant potential to reduce wind-tunnel testing efforts, flight testing campaigns, or numerical simulations. Future work will focus on incorporating physics-based constraints into the candidate generation process and validating the proposed methodology on further experimental data. Additionally, we aim to investigate the impact of candidate scoring based on the Expected Predictive Information Gain on the overall model performance compared to the present BALD strategy.

ACKNOWLEDGMENTS

The funding of all research activities by Airbus Defence and Space GmbH and the scientific guidance by the Technical University of Munich are gratefully acknowledged. The authors want to thank Kai Richter and Andreas Schuette from Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) for providing valuable training data from the SACCON configuration.

Contact address:

ferdinand.dunkes@tum.de

References

[1] Hugo Valayer, Nathalie Bartoli, Mauricio Castaño-Aguirre, Rémi Lafage, Thierry Lefebvre, Andrés F. López-Lopera, and Sylvain Mouton. A python toolbox for data-driven aerodynamic modeling using sparse gaussian processes. *Aerospace*, 11(4), 2024. ISSN: 2226-4310. DOI: 10.3390/aerospace11040260.

[2] Guillermo Suarez, Emre Özkaya, Nicolas R. Gauger, Hans-Jörg Steiner, Michael Schäfer, and David Naumann. Nonlinear surrogate model design for aerodynamic dataset generation based on artificial neural networks. *Aerospace*, 11(8), 2024. ISSN: 2226-4310. DOI: 10.3390/aerospace11080607.

[3] Andrea Garbo, Jigar Parekh, Tilo Rischmann, and Philipp Bekemeyer. Multi-fidelity adaptive sampling for surrogate-based optimization and uncertainty quantification. *Aerospace*, 11(6), 2024. ISSN: 2226-4310. DOI: 10.3390/aerospace11060448.

[4] Mario Stradtner, Carsten M. Liersch, and Philipp Bekemeyer. An aerodynamic variable-fidelity modelling framework for a low-observable ucav. *Aerospace Science and Technology*, 107:106232, 2020. ISSN: 1270-9638. DOI: https://doi.org/10.1016/j.ast.2020.106232.

[5] Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern bayesian

experimental design, 2023. https://arxiv.org/abs/2302.14545.

[6] Russell M. Cummings, Carsten M. Liersch, Andreas Schütte, and Kerstin C. Huber. Aerodynamics and conceptual design studies on an unmanned combat aerial vehicle configuration. *Journal of Aircraft*, 55(2):454–474, 2018. ISSN: 0021-8669. DOI: 10.2514/1.C033808.

[7] Russell M. Cummings and Andreas Schütte. Integrated computational/experimental approach to unmanned combat air vehicle stability and control estimation. *Journal of Aircraft*, 49(6):1542–1557, 2012. ISSN: 0021-8669. DOI: 10.2514/1.C031430.

[8] Dan D. Vicroy, Kerstin C. Huber, Andreas Schütte, Martin Rein, Jonathan P. Irving, Glyn Rigby, Thomas Löser, Andreas-René Hübner, and Trevor J. Birch. Experimental investigations of a generic swept unmanned combat air vehicle with controls. *Journal of Aircraft*, 55(2):475–501, 2018. ISSN: 0021-8669. DOI: 10.2514/1.C033782.

[9] Andreas Schuette, Dietrich Hummel, and Stephan M. Hitzel. Numerical and experimental analyses of the vortical flow around the saccon configuration. In *28th AIAA Applied Aerodynamics Conference*, Reston, Virigina, 06282010. American Institute of Aeronautics and Astronautics. ISBN: 978-1-62410-141-0. DOI: 10.2514/6.2010-4690.

[10] Kerstin C. Huber, Dan D. Vicroy, Andreas Schuette, and Andreas Huebner. Ucav model design and static experimental investigations to estimate control device effectiveness and s&c capabilities. DOI: 10.2514/6.2014-2002.

[11] Andreas Schütte, Kerstin C. Huber, Neal T. Frink, and Okko J. Boelens. Stability and control investigations of generic 53 degree swept wing with control surfaces. *Journal of Aircraft*, 55(2):502–533, 2018. ISSN: 0021-8669. DOI: 10.2514/1.C033700.

[12] V. Roshan Joseph. Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4):531–538, 2022. ISSN: 1932-1864. DOI: 10.1002/sam.11583.

[13] Esra'a Alshdaifat, Doa'a Alshdaifat, Ayoub Alsarhan, Fairouz Hussein, and Subhieh Moh'd Faraj S. El-Salhi. The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance. *Data*, 6(2):11, 2021. DOI: 10.3390/data6020011.

[14] Hao Wang, Xingjian Shi, and Dit-Yan Yeung. Natural-parameter networks: A class of probabilistic neural networks. https://arxiv.org/pdf/1611.00448.

[15] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. ISBN: 9780521518147. DOI: 10.1017/CBO9780511804779.

[16] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. https://arxiv.org/pdf/1312.6114.

[17] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. https://arxiv.org/pdf/1505.05424.

[18] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. *SIGNLL Conference on Computational Natural Language Learning (CONLL)*.

[19] George Casella and Roger L. Berger. *Statistical inference*. Duxbury advanced series. CENGAGE Learning, Andover and Melbourne and Mexico City and Stamford, CT and Toronto and Hong Kong and New Delhi and Seoul and Singapore and Tokyo, second edition edition, 2002. ISBN: 8131503941.

[20] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? https://arxiv.org/pdf/1703.04977.

[21] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. https://arxiv.org/pdf/1112.5745.

[22] David J. C. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992. ISSN: 0899-7667. DOI: 10.1162/neco.1992.4.4.590.

[23] Christopher M. Bishop. *Pattern recognition and machine learning*. Information Science and Statistics. Springer Science+Business Media LLC, New York, NY, 2006.

[24] Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985. ISSN: 03043975. DOI: 10.1016/0304-3975(85)90224-5.

[25] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. https://arxiv.org/pdf/1906.08158.