

Bayesian Models for Uncertainty Estimation in Aerodynamic Databases of Reusable Launch Vehicles

Sven Krummen^{*}, Jan Michael Schraad[†], Marco Sagliano[‡]
DLR, German Aerospace Center, 28359 Bremen, Germany

Tobias Ecker[§], Moritz Ertl[§]
DLR, German Aerospace Center, 37073 Göttingen, Germany

Bodo Reimann[¶]
DLR, German Aerospace Center, 38108 Braunschweig, Germany

Josef Klevanski^{**}, Johannes Riehmer^{**}
DLR, German Aerospace Center, 51147 Köln, Germany

Silas Eichel^{††}
DLR, German Aerospace Center, 70569 Stuttgart, Germany

Lâle Evrim Briese^{‡‡}
DLR, German Aerospace Center, 82234 Wessling, Germany

The definition of Aerodynamic Databases is an important yet very complex and labor-intensive task during the design of new aerospace vehicles. This is particularly true for Reusable Launch Vehicles, as it has been observed during the development of CALLISTO, a demonstrator for a Vertical-Takeoff Vertical-Landing first stage which is jointly developed, manufactured and tested by DLR, JAXA and CNES. In this paper, we present an Inference-based methodology to define various types of Bayesian models exemplarily for a subset of CALLISTO’s Aerodynamic Database to assess their usability and prediction qualities. First, a short introduction to the underlying aerodynamic dataset will be given which has been aggregated from various Computational Fluid Dynamics simulations and Wind Tunnel Test campaigns. Then, the different Bayesian models will be defined and their inference results compared against each other, according to common error metrics. It will be shown that, within the limits and assumptions of this study, several types of Bayesian models provide better accuracy in the

^{*}Systems Engineer, Institute of Space Systems, Robert-Hooke-Straße 7. Corresponding Author: Sven.Krummen@dlr.de.

[†]Student, Institute of Space Systems, Robert-Hooke-Straße 7.

[‡]GNC Engineer, Institute of Space Systems, Robert-Hooke-Straße 7.

[§]Aerothermodynamics Engineer, Institute of Aerodynamics and Flow Technology, Bunsenstraße 10.

[¶]Aerodynamics Engineer, Institute of Aerodynamics and Flow Technology, Lilienthalplatz 7.

^{**}Aerodynamics Engineer, Institute of Aerodynamics and Flow Technology, Linder Höhe.

^{††}Mechanical Engineer, Institute of Structures and Design, Pfaffenwaldring 38-40.

^{‡‡}Research Engineer, Institute of System Dynamics and Control, Münchener Straße 20.

A previous version of this manuscript had been presented at the AIAA SciTech 2024 Forum, 8th January 2024, Orlando, FL.

prediction of uncertain aerodynamic coefficients compared to classical expert-fitted models for the given CALLISTO dataset. Generally, it can be concluded that Bayesian models provide a promising approach for the definition of Aerodynamic Databases and could also find many potential applications in other engineering domains.

Acronyms

AEDB	Aerodynamic Database
ALS	Approach and Landing System
AoA	Angle of Attack
AoR	Angle of Roll
ATDB	Aerothermodynamic Database
AUSMDV	Advection Upstream Splitting Method
CALLISTO	Cooperative Action Leading to Launcher Innovation for Stage Toss-back Operations
CFD	Computational Fluid Dynamics
CNES	Centre National d'Études Spatiales
CSG	Guiana Space Centre
DLM	Doublet Lattice Method
DLR	Deutsches Zentrum für Luft- und Raumfahrt
DNW	German-Dutch Wind Tunnels
ELV	Expendable Launch Vehicle
ESS	Effective Sample Size
FCS/A	Aerodynamic Flight Control System
FCS/R	Reaction Flight Control System
FCS/V	Thrust Vectoring Flight Control System
FEM	Finite Element Method
GLM	Generalized Linear Model
GNC	Guidance, Navigation & Control

GP	Gaussian Process
HDI	Highest Density Interval
HiFi	High-Fidelity
HST	High-Speed Wind Tunnel
JAXA	Japan Aerospace Exploration Agency
LH2	Liquid Hydrogen
LoFi	Low-Fidelity
LOX	Liquid Oxygen
$M_{1/2}$	Matern-1/2 Kernel
$M_{5/2}$	Matern-5/2 Kernel
MAD	Median Absolute Deviation
MCMC	Markov Chain Monte Carlo
MCSE	Monte Carlo Standard Error
MECO	Main Engine Cut-Off
MEIG	Main Engine Ignition
MRE	Maximum Residual Error
MRO	Maintenance, Repair & Overhaul
NUTS	No-U-Turn Sampler
PDF	Probability Density Function
RANS	Reynolds-Averaged Navier-Stokes
ReFEx	Reusability Flight Experiment
RLV	Reusable Launch Vehicle
RMSE	Root Mean Squared Error
RV-X	Reusable Vehicle Experiment
SA	Spalart-Allmaras

SE	Squared-Exponential Kernel
TMK	Trisonic Wind Tunnel
VEB	Vehicle Equipment Bay
VI	Variational Inference
VLM	Vortex Lattice Method
VTVL	Vertical-Takeoff Vertical-Landing
WN	White Noise
WTT	Wind Tunnel Test

I. Introduction

The successful application of Reusable Launch Vehicle (RLV) stages, exemplified by SpaceX's Falcon 9 rocket, has marked a turning point for the global space launcher industry. In response, German Aerospace Center (DLR), French National Centre for Space Studies (CNES), and Japan Aerospace Exploration Agency (JAXA) have joined forces to develop a reusable rocket stage to maintain competitiveness in the European and Japanese space sectors, leading to the inception of the Cooperative Action Leading to Launcher Innovation for Stage Toss-back Operations (CALLISTO) project in 2017. This initiative focuses on advancing the technical aspects and quantifying the economic viability of a reusable Vertical-Takeoff Vertical-Landing (VTVL) first stage demonstrator [1–4].

The CALLISTO vehicle, which is currently developed in this trilateral project, is designed to perform a flight campaign with up to 10 flights, including intermediate Maintenance, Repair & Overhaul (MRO) operations. Figure 1 presents an overview of its main components [5, 6]. The vehicle is propelled by a throttleable and re-ignitable Japanese Reusable Vehicle Experiment (RV-X) engine, utilizing cryogenic Liquid Oxygen (LOX) / Liquid Hydrogen (LH2) and operating in an expander-bleed cycle. Its flight control capabilities are ensured by a Thrust Vectoring Flight Control System (FCS/V), a monergolic Reaction Flight Control System (FCS/R), and an Aerodynamic Flight Control System (FCS/A) with four deployable fins. Landing is enabled by a four-legged Approach and Landing System (ALS) at the aft bay [7].

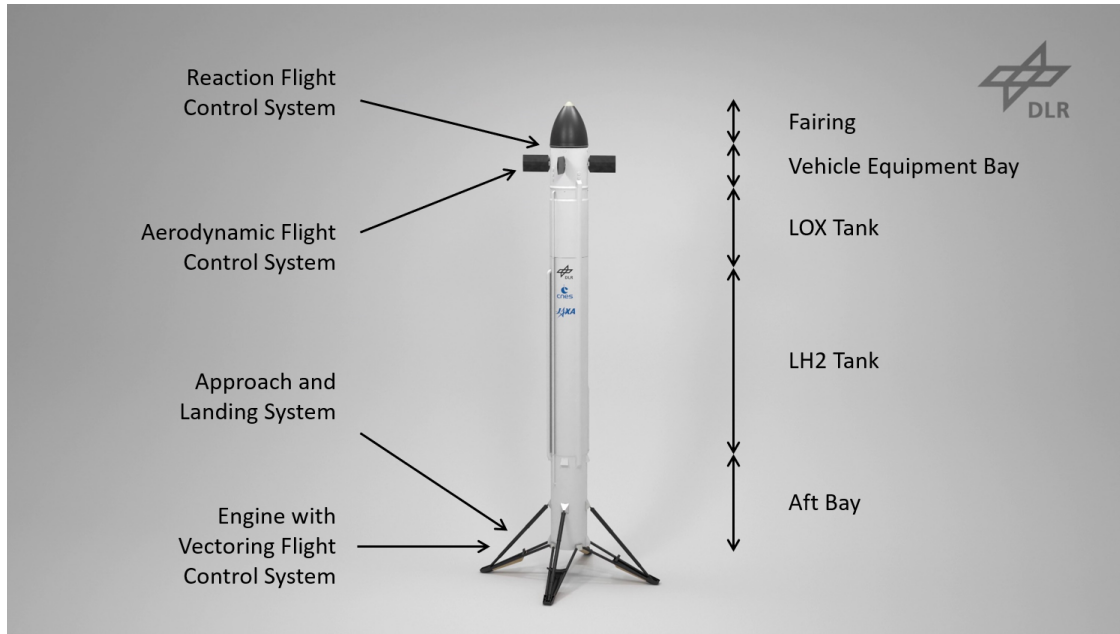


Figure 1 Overview of the major structures and subsystems of the CALLISTO vehicle. Reprinted with permission from [5]. Copyright 2023 DLR.

The planned flight campaign at Guiana Space Centre (CSG) will incrementally progress from low-risk to high-performance flights, gradually increasing apogee and flight velocity for each mission [8]. As illustrated in Figure 2 for a final high-energy mission, several maneuvers are considered for demonstration and final qualification purposes, such as engine reignition, change of downrange landing point, crossing the transonic regime, attitude change, propellant management during maneuvers, and aerodynamically controlled descent [5]. This strategy will test key features of RLVs, supported by extensive vehicle and ground data collection, which is crucial for the development of future operational vehicles.

A critical element in CALLISTO's development is creating an Aerodynamic Database (AEDB) model to predict the vehicle's aerodynamic characteristics during flight. Traditionally, these databases merge Computational Fluid Dynamics (CFD) simulations and Wind Tunnel Test (WTT) measurements via simple empirical-heuristic models. However, for RLVs like CALLISTO, this task is more complex due to the varied vehicle configurations and wide range of flight conditions. Furthermore, the need to perform a pin-point landing at the end of mission constitutes the necessity for a robust Guidance, Navigation & Control (GNC) system, which requires precise quantification of aerodynamic uncertainty. This complexity requires, in a classical approach, manual analyses and expert assessments, and often results in slow, laborious, and non-reproducible processes.

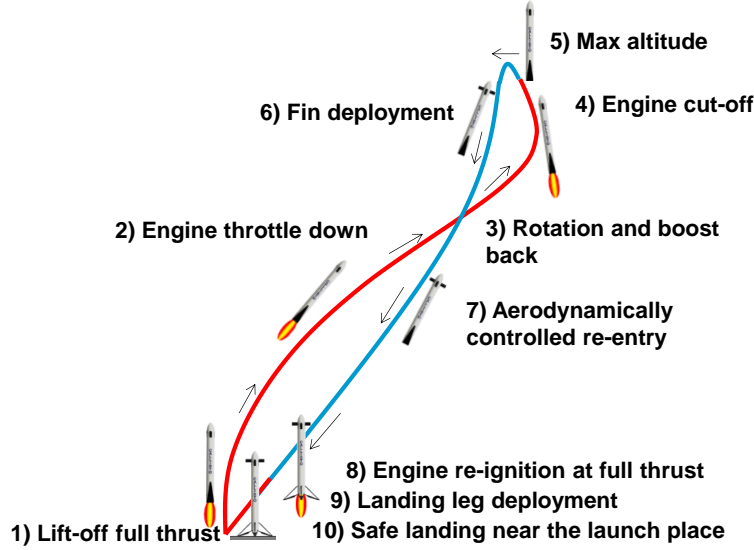


Figure 2 Sketch of the flight profile with main events for the final demo mission of CALLISTO. Reprinted with permission from [5]. Copyright 2023 DLR.

To streamline this process, a Bayesian approach for generating AEDB models has been developed in a previous study and exemplarily tested on CALLISTO's aerodynamic data [9]. This approach leverages Bayesian Inference to predict probability distributions over aerodynamic coefficients, based on existing aerodynamic data and prior knowledge. By doing so, it provides a more traceable and repeatable process with the potential to reduce the workload on human experts.

In this study we apply Bayesian methods to develop various types of AEDB models and test them on a subset of CALLISTO's data. The goal is to evaluate the usability and prediction qualities of these models, as well as to compare them with traditional AEDB models and with each other. The central research question is: *Which Bayesian models are most suitable to quantify uncertainties in the AEDB of typical RLVs, such as CALLISTO?*

Subsequent chapters are structured to provide an exhaustive exploration of this topic. Chapter II delves into the specifics of the aerodynamic data sources and their characteristics pertinent to CALLISTO followed by Chapter III outlining the methodology for constructing and assessing Bayesian AEDB models. Chapter IV applies these methodologies across various classes of Bayesian AEDB models tailored for CALLISTO. Eventually, an in-depth discussion of the inference and prediction outcomes, incorporating a thorough comparison of the diverse models is offered by Chapter V. Finally, Chapter VI synthesizes the key findings of the research and provides perspective on future avenues for exploration in this domain.

II. Aerodynamic Dataset

Throughout CALLISTO's development, extensive aerodynamic studies and experiments were performed across various vehicle configurations to ascertain its aerodynamic properties. These findings are compiled into an AEDB and Aerothermodynamic Database (ATDB), to be used for downstream development tasks like GNC design, mission planning, and structural sizing of components. This AEDB forms the basis of the Bayesian uncertainty models detailed in Chapters IV and V. Subsequent sections outline CALLISTO's aerodynamic design and flight range, the methodologies employed for aerodynamic coefficient determination, and the resultant dataset utilized in this study. For in-depth analyses of CALLISTO's aerodynamics, refer to [10–20].











A. CALLISTO's Aeroshape

The primary structures of CALLISTO, which are determining its aerodynamic shape, consist of the nose fairing, the Vehicle Equipment Bay (VEB) which accommodates the FCS/A and the FCS/R, the integrated LOX and LH2 tanks, the aft bay housing the engine and the FCS/V, and the ALS. When assembled, CALLISTO resembles a slender, almost rotationally symmetric body, akin to traditional rockets or missiles. However, as depicted in Figure 1, it also features several protrusions, such as external fluid lines and cable ducts, as well as external mechanisms (FCS/A and ALS), which are uncommon for conventional Expendable Launch Vehicles (ELVs). Due to CALLISTO's relatively small size compared to conventional launch vehicles, the aerodynamic influence of these protrusions is significantly more pronounced, making them major contributors to the vehicle's overall aerodynamic characteristics.

These unconventional shape features of CALLISTO are typical for RLVs [21, 22], but increase significantly the complexity of the required aerodynamic analyses. On the one hand, detailed simulations and experimental setups are necessary to reproduce the aerodynamic impact of these protuberances. On the other hand, due to the deployable and controllable mechanisms, many aerodynamic configurations occur during flight which need to be individually considered for analysis. The number of system parameters, that have to be included in an AEDB for RLVs, is therefore significantly larger compared to ELVs.

The aerodynamic configuration of CALLISTO is mainly influenced by the state of the deployable and deflectable fins (FCS/A), the deployable legs (ALS), and the engine plume, which varies with thrust settings. During a single flight, this configuration undergoes multiple shifts, requiring also transition analyses between configurations. Table 1 defines the individual aerodynamic configurations in relation to specific flight phases. Beyond these discrete configurations, it is crucial to also account for continuous alterations in the aeroshape, that are resulting from changes in fin deflection and thrust settings.

Table 1 Aerodynamic configurations of CALLISTO

Config	FFO	FFN	UFN	UFO	UUO	UUN	FUFN	FUFO	FUUO	FUO
Picture										
Flight Phase	<u>Ascent:</u> MEIG#1 (Lift-Off) ↓ MECO#1	<u>Ballistic:</u> MECO#1 ↓ Fin Depl.	<u>Aerodyn. Descent:</u> Fin Depl. ↓ MEIG#2	<u>Approach:</u> MEIG#2 ↓ Leg Depl.	<u>Landing:</u> Leg Depl. ↓ MECO#2 (Touch-down)	<u>Ground:</u> After MECO#2	<u>Transition:</u> FFN ↓ UFN	<u>Transition:</u> FFO ↓ UFO	<u>Transition:</u> FUO ↓ UUO	<u>Flights A & B</u>
Fins	Folded	Folded	Unfolded	Unfolded	Unfolded	Unfolded	Transition	Transition	Transition	Folded
Legs	Folded	Folded	Folded	Folded	Unfolded	Unfolded	Folded	Folded	Unfolded	Unfolded
Thrust	On	Off	Off	On	On	Off	Off	On	On	On

Since development began in 2017, multiple iterations of CALLISTO’s aeroshape have been developed and examined, each exhibiting progressively enhanced maturity and fidelity. Figure 3 presents an overview of the key evolutionary stages. Although recent shape developments more accurately represent CALLISTO’s final flight configuration, earlier datasets remain valuable. They offer low-fidelity approximations for flight conditions not (yet) explored with newer models.

After initial trade-offs on size, shape, and fin design, CALLISTO’s first consolidated aeroshape was established with version CAL1B. This shape facilitated the development of the first complete AEDB, which was used for CALLISTO’s preliminary design phase. At this stage, a symmetric layout which neglected many surface protuberances allowed for efficient CFD simulations using just a half-model [10].

The aeroshape evolved to CAL1C during refinements in the preliminary design phase, incorporating details like cable ducts and fluid lines. This more intricate shape defined the baseline for the creation of a comprehensive AEDB for the detailed design phase. As discussed in Section II.C, CAL1C is the most extensively studied shape of CALLISTO and serves as the analytical foundation for this paper.

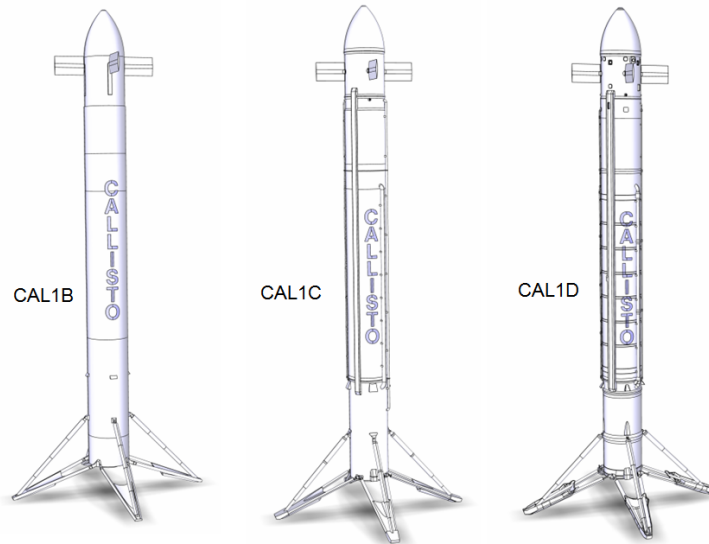


Figure 3 Evolution of CALLISTO’s aerodynamic shape: CAL1B (2018), CAL1C (2019), CAL1D (2022).

The latest, likely final aeroshape, CAL1D, emerged from design evolutions during the detailed design phase. Its highly detailed surface features precluded the creation of a full AEDB, given the demanding computational requirements for CFD simulations and impracticality of repeated WTTs. Consequently, only delta analyses were conducted: limited CFD simulations for critical flight conditions aimed to validate the ability of CAL1C’s AEDB to approximate CAL1D’s behavior with acceptable uncertainty.

B. CALLISTO’s Flight Domain

The aerodynamic characterization of CALLISTO relies not only on its aerodynamic shape but also on its intended flight domain. Designed for an incremental flight campaign of up to 10 flights, CALLISTO will begin with low-altitude, low-risk hop flights, before gradually advancing to more complex high-energy trajectories, as shown in Figure 2. The flight envelope of CALLISTO is extensive and intricate, encompassing a wide range of possible flight conditions.

CALLISTO’s flight domain spans from nearly zero velocity and dynamic pressure during initial hover flights up to low supersonic speeds in the final demonstration flight. It includes both forward and backward flight phases, which are exposing different vehicle surfaces to the incoming flow, across a broad spectrum of altitudes and velocities. During the transition from forward to backward flight, and during initial hover flights, the vehicle will encounter almost arbitrary aerodynamic Angle of Attack (AoA) and Angle of Roll (AoR). Furthermore, the variable interactions between the engine plume and the vehicle during throttling and re-ignition need to be considered, as well as the continuous and independent deflection of the fins.

In contrast to this broad flight domain which implies large variations in the flow topology around the vehicle, high accuracy and confidence in predicting CALLISTO’s aerodynamic coefficients is crucial. This precision is essential for achieving a soft precision landing following the aerodynamically controlled descent phase. Therefore, a comprehensive AEDB is indispensable for the development of CALLISTO, ensuring sufficient coverage for all flight conditions and vehicle configurations.

C. Generation of Aerodynamic Data

To ascertain CALLISTO’s aerodynamic coefficients, numerous CFD and WTT campaigns were conducted based on its aeroshape and flight domain. These campaigns employed a variety of numerical models and experimental setups, creating complementary sets of data to integrate into the AEDB. This diverse approach aimed to minimize systematic errors inherent in individual experimental and numerical methods. Additionally, integrating Low-Fidelity (LoFi) and High-Fidelity (HiFi) methods balanced the need to manage computational, experimental, and human resources while ensuring adequate coverage of CALLISTO’s extensive flight domain with sufficient accuracy.

The subsequent sections summarize the aerodynamic data generation processes for the individual data sources of CALLISTO’s AEDB, with a focus on potential uncertainties in the data. Parts of this discussion are also presented in [9].

1. Computational Fluid Dynamics

Numerical simulations for CALLISTO were conducted using DLR’s TAU code [23, 24]. TAU is a finite volume CFD code that solves Reynolds-Averaged Navier-Stokes (RANS) equations on hybrid grids, combining tetrahedral elements for dealing with complex geometries and prismatic elements for the boundary layer with structured hexahedral elements for the plume refinement. For AEDB generation, simulations utilized second-order accuracy in space, employing an Advection Upstream Splitting Method (AUSMDV) upwinding or central difference scheme with least-squares gradient reconstruction and an explicit Runge-Kutta scheme. To enhance low Mach number computations, a modified variable reconstruction by Thornber [25] and the one-equation Spalart-Allmaras (SA) turbulence model [26] were used. HiFi simulations considered the flow expansion in the nozzle by using TAU’s chemical equilibrium solver with combustion chamber-based inlet conditions. Plume modeling in engine-on scenarios (e.g., UFO, UVO) utilized gas mixtures representing exhaust and air with a frozen chemistry assumption.

Mesh size was categorized into two fidelity classes to balance large parameter space investigation and computational accuracy: TAU HiFi and TAU LoFi. The need for an extensive number of simulations necessitated a smaller LoFi grid for faster simulation times, while accuracy requirements demanded a larger, more detailed HiFi grid. The AEDB generation logic combining these classes is detailed in Chapter IV.

The HiFi grid, with 23 million points, ensures a dimensionless wall distance $y^+ < 1.0$ in the first boundary layer cell. The LoFi grid, with 3 million elements, aims for $y^+ \leq 2.0$ in complex areas, which has been tailored towards acceptable accuracy and computational efficiency based on HiFi simulations. For CAL1D’s delta-aerodynamics, additional shape details increased the number of grid points to 64 million for HiFi UFO simulations [18]. HiFi simulations take about 24 hours on 512 cores to achieve convergence, while LoFi simulations on a local workstation complete in approximately 20 minutes, enabling more parameter variations. Figure 4 visualizes some exemplary TAU HiFi results. Further details on numerical settings and data analysis are available in [15, 17, 27].

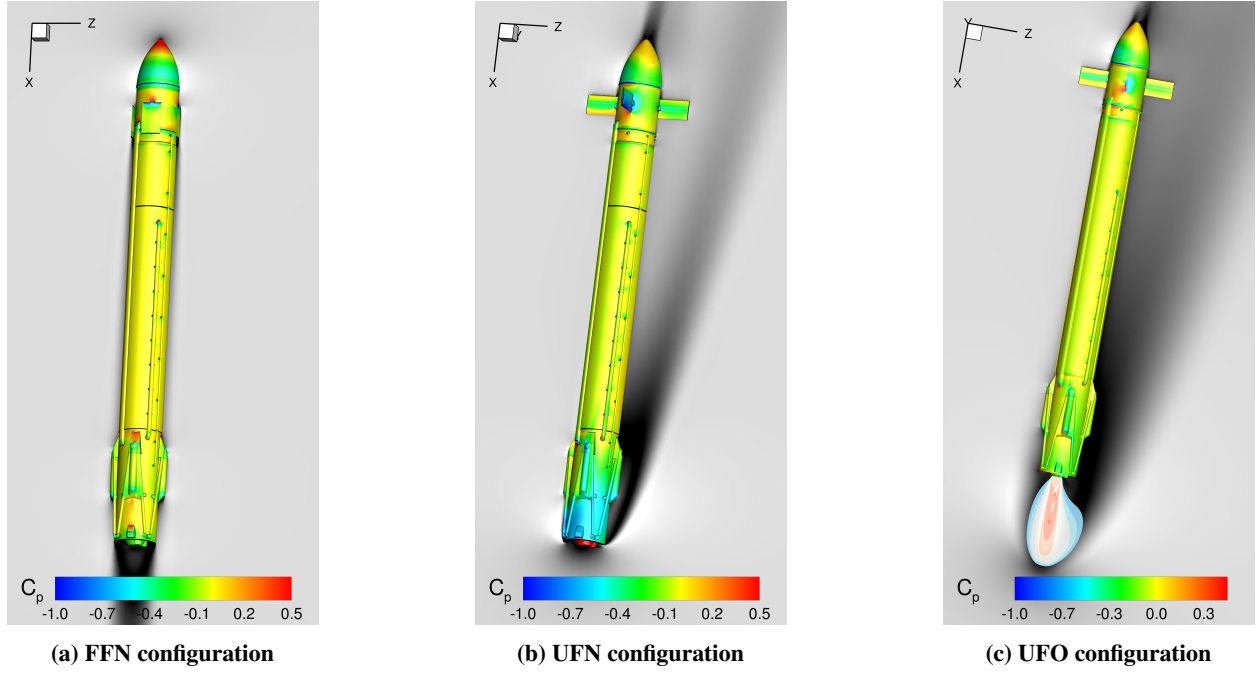


Figure 4 Visualization of CFD results on the TAU HiFi grid for $Ma = 0.7$.

Various error sources and uncertainties are inherent in CFD studies. Recognized solver-related errors include round-off, iterative convergence, and discretization errors [28]. While some errors are mitigated by code design, others require systematic trials and user expertise. Discretization errors stem directly from the grid generation process. For simple grids, these errors are quantifiable through systematic grid convergence studies [29]. However, for complex geometries involving unstructured grids with non-uniform cell sizes and cell growth rates, these studies become more intricate. Our previous grid convergence index studies on similar flows have shown that discretization errors are within the asymptotic range [30].

Additional uncertainties arise from physical modeling choices. The simulations, conducted as steady-state RANS, primarily face turbulence modeling errors, particularly with engine plume involvement. Steady-state assumptions overlook unsteady phenomena and transient effects between trajectory points. Chemical modeling of the engine plume is another uncertainty source. Both, the impacts of turbulence modeling and chemistry on plume characteristics are detailed

in [27]. While chemical effects mainly influences the thermal load estimation due to unburnt hydrogen at the nozzle exit, they are less significant for aerodynamic coefficient estimation. Therefore, the assumption of frozen chemistry and two-species mixtures representing air and exhaust gases is a reasonable simplification for AEDB generation.

2. Wind Tunnel Tests

Multiple wind tunnel campaigns were conducted for CALLISTO's CAL1C aeroshape [11–13, 16], exploring various vehicle configurations and flight phases. These campaigns comprised two test series in the Trisonic Wind Tunnel (TMK) at DLR Cologne and the High-Speed Wind Tunnel (HST) at German-Dutch Wind Tunnels (DNW) Amsterdam, with experimental setups shown in Figure 5.

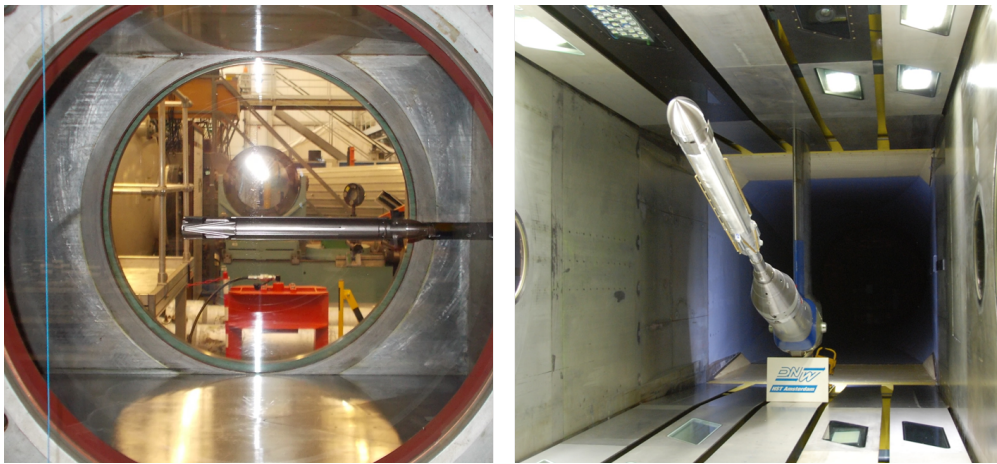


Figure 5 UFN configuration inside the TMK wind tunnel (left), and FFN configuration inside the HST wind tunnel (right).

Flow conditions in both campaigns aimed to closely replicate the flight profile in Mach and Reynolds numbers, ensuring also condition overlap between both wind tunnels. TMK conditions ranged from $Ma = 0.5 \dots 2$ and $Re = 0.38 \dots 1.11 \times 10^6$, while HST covered $Ma = 0.2 \dots 1.3$ and $Re = 0.14 \dots 4.10 \times 10^6$. The focus was on force and moment measurements, supplemented by local pressure measurements and Schlieren imaging. Key differences between both campaigns included the model scale, with HST using a more detailed 1:10 scale model and TMK a 1:35 scale model. Additionally, HST's continuous closed-loop operation allowed to measure 490 combined AoA and AoR polars, whereas TMK's blow-down mode restricted the measurement time to 118 AoA polars. A comprehensive description of these tests is available in [16].

Measurement uncertainties in the WTTs can be determined with high accuracy, as detailed in [16]. However, identifying systemic errors during the WTT measurements presents a larger challenge, with the sting exerting the most significant influence, particularly in backward flight configurations. Other systematic factors include wind tunnel walls

and facility-induced vibrations, though these are less critical. Systematic uncertainties also arise from WTT models themselves, due to minor deviations from the CALIC reference aeroshape in surface details, wall roughness, and wall temperatures. However, based on prior experience, these impacts are comparatively minor.

The parallel approach with two wind tunnels was specifically implemented to quantify these systematic uncertainties, although a detailed quantification is pending. Therefore, the Bayesian models under investigation provide a promising approach to approximate and mitigate these unknown systemic uncertainties in the development of CALLISTO's AEDB.

3. Simplified Models

In addition to the main aerodynamic datasets from CFD simulations and WTT experiments for CALLISTO's AEDB, supplementary data have been sourced from simplified semi-empirical models. These models, developed from extensive experimental data, estimate aerodynamic coefficients based on a basic parametrization of the vehicle's geometry. Due to their low computational demands but high uncertainty, these models are typically used in the conceptual design phase only [31]. For CALLISTO, this data aids in cross-validation of data, especially where better datasets lack coverage.

The dataset was generated using Missile DATCOM [32], analyzing FFN and UFN configurations over the Mach range $Ma = 0.5 \dots 2$ for all AoAs and AoRs. While fin deflections and engine-on configurations are theoretically modelable, these influence have not been included in this dataset (yet). However, configurations with deployed legs are unlikely to be modeled due to limitations in the software's geometry parametrization.

Experience indicates significant uncertainties in these semi-empirical models, especially for complex aeroshapes, like CALLISTO or ELVs in general, and for flight conditions close to $\alpha \approx 90^\circ$. This is largely because tools like Missile DATCOM rely on simple mathematical models fitted to generic aeroshape data under common flight conditions [32]. Compared to CFD or WTT data, these estimates are less reliable for CALLISTO. However, they often adequately represent the qualitative behavior of aerodynamic coefficients.

In addition to the numerical, experimental, and semi-empirical methods mentioned above, the Vortex Lattice Method (VLM) and the Doublet Lattice Method (DLM), based on linearized full potential equations, were employed for the rapid computation of steady and unsteady aerodynamic characteristics of CALLISTO's aerodynamic control surfaces. However, these methods are only valid within the subsonic flight regime and neglect key physical flow properties such as boundary layer effects, flow separation, and locally induced supersonic regions. A comparison of the lift distributions computed with TAU and with VLM, considering also aeroservoelastic effects, shows that VLM provides a more conservative estimation of aerodynamic loads. This discrepancy is particularly pronounced at higher angles of attack due to the method's reliance on linear extrapolation. Consequently, the aerodynamic datasets obtained from VLM and DLM were not incorporated into the AEDB definition, but instead were used for dynamic stability studies [20] and for the rapid assessment of FCS/A-level requirements, particularly regarding the structural deformation of the fins due to bending or torsion loads [19].

D. Dataset Scope and Restrictions

The complete AEDB for CALLISTO encompasses analyses of three aeroshape versions (CAL1B, CAL1C, CAL1D) using the experimental and numerical methods described previously. This study focuses solely on CAL1C data, comprising five distinct data sources:

- CFD-sourced data: TAU HiFi, TAU LoFi
- WTT-sourced data: HST, TMK
- Simplified data: DATCOM

These sources collectively provide estimates for aerodynamic force coefficients (c_x, c_y, c_z) and moment coefficients (c_{Mx}, c_{My}, c_{Mz}) across various flight conditions and vehicle configurations. Key independent variables within the AEDB include:

- Mach number Ma ,
- Altitude H (encoding the air density ρ and Reynolds number Re),
- Angle of Attack α ,
- Angle of Roll ϕ ,
- Fin deflection angles $\delta_{j=1..4}$,
- Fin deployment angles $\eta_{j=1..4}$,
- Leg deployment angles $\lambda_{j=1..4}$, and
- Engine thrust level τ .

If the transition phases between aerodynamic configurations (e.g., during fin deployment or engine start-up) are neglected, the continuous variables $\eta_{j=1..4}$, $\lambda_{j=1..4}$, and τ can be aggregated into a single categorical variable *config*, which represents the vehicle configuration as defined in Table 1. Additional categorical variables *src* and *shape* can be introduced to specify the experimental or numerical method and the aeroshape version associated with each data point.

In summary, the AEDB currently contains over 25000 data points with 25 feature columns, or 17 when using the *config* parameterization. In order to limit the computational demands and overall complexity, this study focuses on a representative subset of 297 samples for Bayesian Inference analysis, formally denoted as the *observed dataset* or *training set* \mathcal{D} , which is constrained to the following conditions:

$$\begin{aligned}
Ma &= 0.5 \\
H &= 1000\text{m} \\
\alpha &\in [0^\circ \dots 360^\circ] \\
\phi &= 0^\circ \\
\delta_{j=1..4} &= 0^\circ \\
config &= \text{UFN}
\end{aligned} \tag{1}$$

This subset aligns with a critical phase of CALLISTO’s aerodynamically controlled descent trajectory, which is essential for precision landing. Robust uncertainty quantification in this domain is therefore crucial for GNC system design. Furthermore, this subset provides ample data from all CFD and WTT sources, supporting cross-validation of inference results, and thereby enhancing the capabilities to assess fitting and predictive accuracies of the individual models.

III. Methodology

For most development tasks, direct use of raw aerodynamic data from the AEDB is impractical, as aerodynamic coefficients are only available for individual trajectory points covered by simulations or experiments, rather than for arbitrary conditions across the flight domain. Additionally, inherent simplifications in the data generation methods introduce small but inevitable systematic discrepancies among the data sources of the AEDB. To facilitate downstream development tasks, an *AEDB function* or *model* is therefore needed to provide a functional relationship between possible flight conditions and aerodynamic coefficients while also quantifying uncertainties, either as confidence intervals or probability distributions.

An early example of a complex AEDB that accounts for different flight domains, vehicle configurations, and quantified uncertainties is the Aerodynamic Design Data Book of the Space Shuttle [33]. Unlike CALLISTO, the Space Shuttle’s AEDB was derived solely from extensive WTT campaigns across multiple facilities and test setups. The uncertainties in these datasets were classified into two levels: a lower uncertainty bound or *tolerance* accounting for scatter in the measurements, and an upper bound or *variation* determined by a team of aerodynamic experts to account for potential systemic errors, particularly for inaccuracies in preflight aerodynamic predictions in uncharted flight regimes [34, 35].

Modern vehicle development requires fewer WTT campaigns than the Space Shuttle program, as CFD advancements have significantly improved aerodynamic characterization across the entire flight envelope. Instead, WTTs are primarily used to validate CFD results or to analyze flight conditions that CFD methods may not fully capture. Such approaches

were implemented for the ReFEx [36] and HEXAFLY-INT [37] experiments. Conversely, the AEDB for the expendable launcher VEGA-C [38] was primarily based on WTT measurements, with CFD data used for validation and uncertainty estimation. These combined approaches can reduce costs and accelerate the development of the vehicle [9].

Similar to the approach taken for the Space Shuttle [34], AEDB models have traditionally been created through expert assessments, where aerodynamicists manually analyze datasets to derive suitable functional relationships based on heuristics and heritage knowledge. These models typically use low-degree inter- and extrapolation of selected AEDB subsets, with test conditions chosen to facilitate model definition. However, this process lacks standardization and independent reproducibility due to its manual approach.

During the development of CALLISTO, it became evident that constructing a classical AEDB model from multiple sources, including CFD, WTT, and simplified aerodynamic models, is particularly challenging for RLVs. The large number of independent variables and their extensive parameter ranges make manual uncertainty quantification difficult, as exhaustive human effort is required to inspect and analyze the multidimensional datasets. Due to the involvement of human experts, these assessments are inherently subjective and hard to automate.

To address these challenges, a Bayesian approach for generating AEDB models has been developed [9], complementary to the definition of a classical AEDB model for CALLISTO [15, 19]. The applied Bayesian methods, though already well-established in various scientific disciplines [39–41], have seen limited application in aerospace engineering so far [42–44]. For instance, Gaussian Processes (GPs) were used to replace Finite Element Method (FEM) models to reduce the computational cost while maintaining acceptable accuracy [42]. In the aerodynamics domain, Bayesian optimization based on GP regression was applied to different CFD problems, such as shape optimization of the wall of a channel flow, amongst others [43]. Furthermore, Bayesian methods were used by [44] to combine WTT and CFD data to predict the desired aerodynamic properties. However, this work focused on the local pressure distribution along an airfoil, while this paper addresses the prediction of aerodynamic coefficients including the quantification of uncertainties for an entire vehicle, covering different flight domains and vehicle configurations. For a detailed discussion about the application of Bayesian Inference on AEDBs and other use cases, refer to [9].

This paper applies Bayesian inference to construct and evaluate multiple types of Bayesian AEDB models for a representative subset of CALLISTO’s AEDB. The goal is to compare their data fitting and predictive capabilities, benchmark them against classical expert-driven approaches, and assess their suitability for future RLVs development. The subsequent sections formalize the AEDB model generation problem, summarize the application of Bayesian Inference for this purpose, and introduce metrics for assessing model quality.

A. Problem Definition

From a mathematical perspective, the generation of an AEDB model can be seen as a classical curve fitting task, aiming to find a regression function for the aerodynamic datasets. As outlined in Chapter II, the outputs of this function are aerodynamic force and moment coefficients, which can be concatenated for convenience to the *coefficient vector* \mathbf{c} :

$$\mathbf{c} := (c_x, c_y, c_z, c_{M_x}, c_{M_y}, c_{M_z}) \quad (2)$$

These coefficients depend on flow conditions, vehicle orientation, aeroshape parametrization, and error characteristics of the data source. Without loss of generality these dependencies can be subsumed to a *generalized input vector* \mathbf{x} . The parametrization of this vector is not uniquely determined, so for CALLISTO's AEDB the following description has been chosen:

$$\mathbf{x} := (Ma, H, \alpha, \phi, \delta_{j=1..4}, config, src) \quad (3)$$

Here, the categorical variable $config \in \mathcal{C}$ defines the aeroshape configuration (as in Table 1), while $src \in \mathcal{S}$ accounts for the different data sources. To enable predictions for the yet unobserved flight setup of CALLISTO, an empty FLIGHT category is formally introduced as additional data source (see Chapter V.B for details).

$$config \in \mathcal{C} := \{\text{FFO}, \text{FFN}, \text{UFN}, \text{UFO}, \text{UUO}, \text{UUN}, \text{FUFN}, \text{FUFO}, \text{FUUO}, \text{FUO}\} \quad (4)$$

$$src \in \mathcal{S} := \{\text{TAU HiFi}, \text{TAU LoFi}, \text{HST}, \text{TMK}, \text{DATCOM}, \text{FLIGHT}\} \quad (5)$$

Together with a *generalized parameter vector* $\boldsymbol{\theta}$, which contains any regression parameters, the searched AEDB function f can then be formalized as:

$$\mathbf{c} = f(\mathbf{x}; \boldsymbol{\theta}) \quad (6)$$

Given the observed aerodynamic dataset $(\tilde{\mathbf{c}}_i, \tilde{\mathbf{x}}_i) \in \mathcal{D}$ with cardinality $|\mathcal{D}|$ (as defined in Chapter II.D) the AEDB generation process, also known as *fitting*, *training* or *inference*, can be formulated as the problem of finding a function f and parameters $\boldsymbol{\theta}$ that best fit the observed data \mathcal{D} :

$$\begin{aligned}
& \text{Select } f, \theta \\
& \text{such that } \tilde{c}_i \approx f(\tilde{x}_i; \theta) \\
& \text{for all data } (\tilde{c}_i, \tilde{x}_i) \in \mathcal{D}
\end{aligned} \tag{7}$$

Please note that neither a solution algorithm nor a quality metric for the regression is imposed by this formulation. Consequently, in practice, this fitting problem is often resolved manually through expert assessments without a formalized mathematical approach. However, within this paper, we utilize Bayesian Inference to address this problem, as detailed in Section III.B.

Once an appropriate AEDB model is established, characterized by f and θ , it can be *evaluated to predict* the aerodynamic coefficients \hat{c} for unobserved flight conditions \hat{x} . This predictive capability is requested from various other technical domains and is a primary motivation for developing AEDBs models:

$$\hat{c} = f(\hat{x}; \theta) \tag{8}$$

Typically, the problem (7) is solved in a way that the prediction function (8) provides a point estimation for the nominal aerodynamic coefficients, which is often sufficient during the conceptual design phase of launch vehicles. Mathematically, this represents a *deterministic model* where f is a deterministic function and θ contains only exact values.

As a launch vehicle design progresses, quantified uncertainties of the estimated aerodynamic coefficients become necessary, particularly for designing robust GNC systems and ensuring reliable mission architectures. This requirement can be met by treating function (6) as a *probabilistic model*, where f is a random function, and θ consists of random variables following specific probability distributions. Hence, the focus of the fitting problem (7) shifts to determining a probability distribution over the coefficient vector c , given the input vector x and random parameter vector θ :

$$c = f(x; \theta) \sim p(c \mid x; \theta) \tag{9}$$

Depending on the specific application, such a probabilistic AEDB model can directly be used in Monte Carlo simulations or for establishing confidence intervals around nominal aerodynamic coefficients.

B. Bayesian Inference

Bayesian Inference is an application of Bayes' Law of conditional probabilities for fitting or updating probabilistic models with observed data. First, a Bayesian AEDB model needs to be defined, whose parameter distribution $p(\theta)$ will subsequently be inferred from the given dataset. The structure and parametrization of such a model is commonly developed via definition of a *prior predictive distribution*, describing the expected distribution of aerodynamic coefficients before any data is known:

$$p(c | x; \epsilon) = \int p(c | x; \theta, \epsilon) p(\theta | \epsilon) d\theta \quad (10)$$

The *likelihood function* $p(c | x; \theta, \epsilon)$ relates here the observed data to model parameters, while $p(\theta | \epsilon)$ is the *prior distribution*, the expected probability of model parameters before inference. For completeness, a vector of *hyperparameters* ϵ is introduced which is defined a-prior and not fitted during inference, contrary to the *parameters* vector θ .

At this stage, expert knowledge can be encoded into the model via the structure of likelihood function and prior distribution, or by selection of hyperparameters. Weakly informative priors, which put little probability mass on unreasonable solutions but still retain high entropy, are generally recommended to ensure that most information is inferred from the given dataset [45, 46].

Given such a Bayesian model definition, the *posterior distribution* of the model parameters can be inferred from the provided aerodynamic dataset $(\tilde{c}_i, \tilde{x}_i)_{i=1..|\mathcal{D}|}$ by application of Bayes' Law:

$$p(\theta | (\tilde{c}_i, \tilde{x}_i)_{i=1..|\mathcal{D}|}, \epsilon) = \frac{p(\tilde{c}_{i=1..|\mathcal{D}|} | \tilde{x}_{i=1..|\mathcal{D}|}, \theta, \epsilon) p(\theta | \epsilon)}{\int p(\tilde{c}_{i=1..|\mathcal{D}|} | \tilde{x}_{i=1..|\mathcal{D}|}, \theta, \epsilon) p(\theta | \epsilon) d\theta} \quad (11)$$

Once the posterior distribution is calculated, it can be used to predict unobserved aerodynamic coefficients \hat{c} for arbitrary flight conditions \hat{x} via the *posterior predictive distribution*:

$$p(\hat{c} | \hat{x}; (\tilde{c}_i, \tilde{x}_i)_{i=1..|\mathcal{D}|}, \epsilon) = \int p(\hat{c} | \hat{x}; \theta) p(\theta | (\tilde{c}_i, \tilde{x}_i)_{i=1..|\mathcal{D}|}, \epsilon) d\theta \quad (12)$$

In comparison to Eq. (9), this fitted posterior predictive model can then be used as an AEDB function for subsequent development tasks:

$$c \sim p(c | x; (\tilde{c}_i, \tilde{x}_i)_{i=1..|\mathcal{D}|}, \epsilon) \quad (13)$$

It should be noted that Eqs. (10) – (12) can usually not be solved analytically in a closed form. However, several efficient approximation techniques and algorithms are known for this type of problem, such as Markov Chain Monte Carlo (MCMC) sampling or Variational Inference (VI). Within the scope of this study, all Bayesian AEDB models

have been implemented with the software library PyMC [47]. For inference and prediction the No-U-Turn Sampler (NUTS) from PyMC and NumPyro [48] have been used, implementing state-of-the-art variants of Hamiltonian MCMC sampling [49].

The main advantage of this Bayesian approach is that model inference and prediction can be fully automated, potentially reducing the need of expert involvement. This way frequent refitting of the AEDB function would become feasible whenever new data is available. Also, complex relationships and correlations of the uncertainty contributors can be inferred directly from the data, depending on very broad a-priori knowledge only.

However, translating expert knowledge into the mathematical form of prior distributions is more challenging than in the classical approach, whose model structures are more often intuitive. Inference results can be complex and harder to interpret, increasing the risk of hidden issues in the model or fitting process. Even though less human workforce is needed to fit a Bayesian AEDB model, it requires significant computational resources for inference and prediction, and is therefore likely unsuitable for real-time use cases in simulations or flights.

C. Error Metrics

To evaluate the fitting and predictive accuracy of AEDB models, this study employs various *error metrics*. These metrics are sometimes referred to as *scoring rules* or *scoring functions* in the literature [50, 51]. They aim to compare the predictions from the fitted model $\hat{\mathbf{c}}$ with multiple observed data points in the AEDB $\tilde{\mathbf{c}}$, known as the *validation set* \mathcal{D}_v , and quantify the deviation as a score. A lower score signifies superior predictive accuracy of the model under examination.

This paper applies the following three error metrics:

$$\text{MRE}(\tilde{\mathbf{c}}_i, \hat{\mathbf{c}}_i)_{i=1..|\mathcal{D}_v|} = \max_{i=1..|\mathcal{D}_v|} (|\tilde{\mathbf{c}}_i - \hat{\mathbf{c}}_i|) \quad (14)$$

$$\text{RMSE}(\tilde{\mathbf{c}}_i, \hat{\mathbf{c}}_i)_{i=1..|\mathcal{D}_v|} = \sqrt{\frac{1}{|\mathcal{D}_v|} \sum_{i=1}^{|\mathcal{D}_v|} (\tilde{\mathbf{c}}_i - \hat{\mathbf{c}}_i)^2} \quad (15)$$

$$\text{MAD}(\tilde{\mathbf{c}}_i, \hat{\mathbf{c}}_i)_{i=1..|\mathcal{D}_v|} = \text{median}_{i=1..|\mathcal{D}_v|} (|\tilde{\mathbf{c}}_i - \hat{\mathbf{c}}_i|) \quad (16)$$

The Maximum Residual Error (MRE) calculates the maximum absolute difference between observations and corresponding predictions in the validation set. This approach is analogous to the common engineering practice of designing systems for the most demanding scenarios. However, in probabilistic contexts, this metric disproportionately emphasizes outliers and tail risks, limiting its practical applicability.

Conversely, the Median Absolute Deviation (MAD) computes the median of absolute differences across the validation set, offering a robust metric less influenced by outliers. While such robustness is beneficial for model evaluation, it overlooks the significant development risks posed by outliers.

The Root Mean Squared Error (RMSE), widely used in scientific and engineering disciplines, aggregates deviations based on the Euclidean norm. This metric accounts for all data points, making it sensitive to outliers without giving them undue weight. Hence, the RMSE is currently regarded as the most representative metric for assessing the quality of an AEDB model, unless a more suitable metric emerges from the specific needs of AEDB model users.

Notably, for Bayesian models, the predictions \hat{c} are random variables according to the posterior predictive distribution (12). Thus, the calculated error metrics MRE, RMSE, and MAD also follow a probability distribution. Unlike simple point estimates typical in non-Bayesian models, this probabilistic nature provides further insights into the variability and significance of the estimated error.

One important problem in the evaluation of error metrics is that the validation set \mathcal{D}_v is a subset of the AEDB, which has already been used for model fitting. This overlap may lead to an overestimation of the predictive quality, favoring overfitted models, due to the dual use of validation data. To mitigate this, the validation set can be excluded during model training, enabling independent cross-validation. For details about how cross-validation has been applied to this study, please refer to Section V.A.

IV. Model Definition

The main objective of this study is to evaluate the effectiveness and quality of various Bayesian models for CALLISTO's AEDB through comparative analysis. For this purpose, over 70 model types and variations have been implemented and tested on different AEDB subsets [9]. This paper however limits its focus on models predicting the aerodynamic normal force coefficient c_z in dependence on the Angle of Attack α , and on the data source category $src \in \mathcal{S}$. A representative subset of the AEDB has been chosen accordingly in Section II.D, ensuring comprehensive data coverage while keeping all other independent variables constant. The coefficients vector (2) and the input vector (3) are thus simplified to:

$$\begin{aligned} \mathbf{c} &= (c_z) \\ \mathbf{x} &= (\alpha, src) \end{aligned} \tag{17}$$

This simplification aims to maintain an interpretable model structure, so that differences and similarities between various model types can be emphasized better. More complex models have been explored but omitted here to limit computational demands and visualization challenges, which would occur due to the curse of dimensionality. Therefore,

this restriction helps to perform a broader survey over different Bayesian models and supports better understanding of the general applicability to Bayesian methods. The models analyzed in this study provide a baseline for future combinations and refinements to develop more sophisticated Bayesian AEDB models.

To conduct Bayesian Inference, each model needs to be defined in terms of a likelihood function (18) and a prior distribution of parameters (19):

$$c_z(\alpha, src; \theta, \epsilon) \sim p(c_z \mid \alpha, src; \theta, \epsilon) \quad (18)$$

$$\theta(\alpha, src; \epsilon) \sim p(\theta \mid \alpha, src; \epsilon) \quad (19)$$

Since this formulation imposes little structure on the allowed models, many common model types can be re-used within the Bayesian framework. For model definition, these probability distributions are often composed of parametrized sets from common distribution families, like the Normal(μ, σ), and deterministic relationships between parameters. To enhance interpretability, standardized distribution families are employed in a location-scale parametrization. For instance, Gamma(μ, σ), characterized by a distribution mean μ and standard deviation σ , is utilized over alternative parameterizations.

In order to better incorporate expert knowledge into the model structure, all models are conceptually designed to consist of a nominal functional dependency $f(\alpha)$, which describes the searched relationship of the aerodynamic coefficients, and an error function $e(\alpha, src)$, enabling the specification of source-dependent error characteristics. This division allows for encoding of different trust levels in data subsets, with HST and TAU HiFi data considered most representative, and DATCOM data the least:

$$c_z(\alpha, src; \theta, \epsilon) = f(\alpha; \theta, \epsilon) + e(\alpha, src; \theta, \epsilon) \quad (20)$$

Within the scope of this study, three classes of Bayesian models are analyzed: Spline Models, Fourier Series Models, and Gaussian Process Models. These classes have been chosen for their widespread use, in other domains or in non-Bayesian contexts, and capability to model typical AEDB features. Additionally, results from a non-Bayesian Expert-fitted Models are included for comparative purposes.

The following sections specify the individual model classes as well as selected model instances, whose results are presented in this paper.

A. Spline Models

Spline interpolation, particularly piecewise-linear interpolation, is a widely used method in science and engineering for creating continuous functions from discrete data. Thus, spline models were among the first used for addressing the AEDB modeling problem (7). In fact, also the expert-fitted Classical model resembles a non-Bayesian Spline model, as shown in Section IV.D.

Spline models are a subclass of Generalized Linear Models (GLMs), whose covariates are transformed by local polynomial basis functions ϕ_k with compact support around dedicated knots $k \in [1, \dots, K]$ [52]. By design, splines form smooth functions in C^n , where n denotes the degree of the polynomials used. Their primary advantage lies in their capacity to represent local features effectively, and their straightforward structure, enhancing interpretability and computational efficiency. However, modeling highly complex features can be challenging, as high-degree polynomial solutions may become unstable, while the minimum knot spacing is constrained by the data resolution.

All analyzed spline models are based on a spline function with random coefficients β_k for the nominal function $f(\alpha)$. The error function $e(\alpha, src)$ is modeled as additive white noise with a standard deviation σ_{src} that depends on the data source category. While more complex physical models for the error behavior are possible, this simple approach facilitates comparability with the other model classes.

$$\begin{aligned} f(\alpha) &= \sum_{k=1}^K \beta_k \phi_k(\alpha) \\ e(\alpha, src) &\sim \text{Normal}(0, \sigma_{src}) \end{aligned} \tag{21}$$

Following this design, all investigated Bayesian Spline Model share a common mathematical structure:

$$\begin{aligned} \beta &\sim \text{Normal}(\beta_\mu, \beta_\sigma) \\ \mu(\alpha) &= \sum_{k=1}^K \beta_k \phi_k(\alpha) \\ \sigma &\sim \text{Gamma}(\sigma_\mu, \sigma_\sigma) \\ c_z(\alpha, src) &\sim \text{Normal}(\mu, \sigma_{src}) \end{aligned} \tag{22}$$

As can be seen, the parameter vector θ and hyperparameter vector ϵ can hereby be summarized as:

$$\begin{aligned} \theta &= (\beta, \sigma) \\ \epsilon &= (\beta_\mu, \beta_\sigma, \sigma_\mu, \sigma_\sigma) \end{aligned} \tag{23}$$

Since the number of coefficients in β corresponds to the number of spline knots K , and σ defines a dedicated standard deviation for each source category in S , the total number of Bayesian parameters whose distributions are updated during inference is $\dim(\theta) = K|S|$.

The choice of prior distributions for the parameters is grounded in experience and best practices for Bayesian models [45, 46]. The values for hyperparameters are selected to represent weakly informative priors and are consistent across all models. By specifying unique σ_μ values for each data source category, additional a-priori knowledge about the data sources' reliability is integrated into the models.

In Chapter V, inference results of four representative spline models are shown and discussed. For reference, the following subsections define the actual parametrization of these model instances.

1. *LS(K=12) Model*

Piecewise-linear Spline Model of degree 1, as defined in Eq. (22), with cyclical linear basis functions ϕ_k on $K = 12$ knots. All knots have an equidistant spacing of 30° , starting from 0° .

2. *LS(K=72) Model*

Piecewise-linear Spline Model of degree 1, as defined in Eq. (22), with cyclical linear basis functions ϕ_k on $K = 72$ knots. All knots have an equidistant spacing of 5° , starting from 0° .

This model has strong structural similarity to the expert-fitted Classical model, implementing the same knot locations and polynomial degree. Therefore, it might be interpreted as a Bayesian extension of the non-Bayesian Classical model.

3. *CS(K=12) Model*

Cubic Spline Model of degree 3, as defined in Eq. (22), with cyclical cubic basis functions ϕ_k on $K = 12$ knots. All knots have an equidistant spacing of 30° , starting from 0° .

4. *CS(K=72) Model*

Cubic Spline Model of degree 3, as defined in Eq. (22), with cyclical cubic basis functions ϕ_k on $K = 72$ knots. All knots have an equidistant spacing of 5° , starting from 0° .

B. Fourier Series Models

Another kind of GLM that is analyzed in this study are Fourier Series models. Fourier series models are prevalent in science and engineering for approximating periodic functions. They have a structure similar to Spline Models, but employ sine and cosine terms as basis functions ϕ_n . These functions have global support, making Fourier models apt for modeling global periodic behavior, even with a limited number of terms. However, they may struggle with approximating highly local features, as high-order Fourier series can induce oscillatory behavior on a global scale.

The investigated Fourier models are designed analogously to spline models (21). In these models, the nominal function $f(\alpha)$ is now represented as a Fourier series, while the error function $e(\alpha, src)$ remains source-dependent white noise. The common structure of all Fourier models is defined as follows:

$$\begin{aligned}
\mathbf{a} &\sim \text{Normal}(\mathbf{a}_\mu, \mathbf{a}_\sigma) \\
\mathbf{b} &\sim \text{Normal}(\mathbf{b}_\mu, \mathbf{b}_\sigma) \\
\mu(\alpha) &= \sum_{n=1}^N a_n \cos(n\alpha) + \sum_{n=1}^N b_n \sin(n\alpha) \\
\sigma &\sim \text{Gamma}(\sigma_\mu, \sigma_\sigma) \\
c_z(\alpha, src) &\sim \text{Normal}(\mu, \sigma_{src})
\end{aligned} \tag{24}$$

Given this parametrization, the parameter and hyperparameter vectors for the Fourier models can be summarized as:

$$\begin{aligned}
\boldsymbol{\theta} &= (\mathbf{a}, \mathbf{b}, \sigma) \\
\boldsymbol{\epsilon} &= (\mathbf{a}_\mu, \mathbf{a}_\sigma, \mathbf{b}_\mu, \mathbf{b}_\sigma, \sigma_\mu, \sigma_\sigma)
\end{aligned} \tag{25}$$

Since the number of coefficients in \mathbf{a} and \mathbf{b} corresponds to the order of the Fourier series N , and σ defines a standard deviation for each source category in \mathcal{S} , the total number of Bayesian parameters is $\dim(\boldsymbol{\theta}) = 2N|\mathcal{S}|$. Again, commonly used prior distributions are applied to all Bayesian parameters, and the hyperparameters are selected to provide weakly informative priors.

As can be seen in Eq. (24), the sine-cosine parametrization of Fourier series has been preferred in this study over the amplitude-phase parametrization, due to better its convergence behavior during inference. However, for interpretability, conversion to amplitude A_n and phase ϕ_n form is possible during post-processing:

$$\begin{aligned}
A_n &= \sqrt{a_n^2 + b_n^2} \\
\phi_n &= \text{atan2}(b_n, a_n)
\end{aligned} \tag{26}$$

Inference results from two representative Fourier models are presented and discussed in Chapter V. For reference, their specific parametrizations are defined in the following sections.

1. $F(N=6)$ Model

Fourier Series Model, as defined in Eq. (24), with maximum order $N = 6$. Please note that this model has the same number of parameters as LS(K=12) and CS(K=12).

2. $F(N=36)$ Model

Fourier Series Model, as defined in Eq. (24), with maximum order $N = 36$. Please note that this model has the same number of parameters as LS(K=72) and CS(K=72).

C. Gaussian Process Models

GPs represent another flexible class of models, which have already been explored for modeling AEDB functions [9]. They are commonly employed in nonparametric regression problems.

In distinction to previously used GLMs, GPs are defined as probability distributions over function spaces, such that any finite set of function values follows a multivariate normal distribution [53]. This nonparametric nature means that the model's parameter space is theoretically infinite-dimensional, so that the functions cannot be described in closed form. For practical purposes, the models are however evaluated only on a finite set of data points, so that, by definition, GPs can here be approximated by multivariate normal distributions.

GPs are characterized by their *mean function* $m(\mathbf{x})$ and *covariance function* $k(\mathbf{x}, \mathbf{x}')$, also known as the *kernel*. The kernel is significantly more influential on the GP's functional properties, and thus, the mean function is often set to constant zero. A variety of elementary kernels and generation rules are known in literature, so that a tailored covariance function can easily be constructed to impose specific functional characteristics on the GP [53, 54].

The main advantage of GPs lies in their adaptability and the range of functions they can model through appropriate kernel selection. However, they are computationally intensive, as their parameter space scales with dataset size, considering also correlations between all observed data points.

For the scope of this study, the following general structure for GP models has been developed. Here, custom kernels can be inserted to describe the expected nominal functional behavior (via k_f) and error behavior (via k_e) of the AEDB model. l_f and l_e are length-scale parameters which are typically used by these kernels. In this context, $\delta_{\mathbf{x}, \mathbf{x}'}$ denotes the Kronecker delta function, so that the influence of the error kernel k_e is individually scaled for each data source category by σ_{src} .

$$\begin{aligned}
l_f &\sim \text{InvGamma}(l_{f,\mu}, l_{f,\sigma}) \\
l_e &\sim \text{InvGamma}(l_{e,\mu}, l_{e,\sigma}) \\
\eta &\sim \text{InvGamma}(\eta_\mu, \eta_\sigma) \\
\sigma &\sim \text{InvGamma}(\sigma_\mu, \sigma_\sigma) \\
k_\sigma(\text{src}, \text{src}') &= \sigma_{\text{src}} \delta_{\text{src}, \text{src}'} \sigma_{\text{src}'} \\
k((\alpha, \text{src}), (\alpha', \text{src}')) &= \eta^2 k_f(\alpha, \alpha'; l_f) + k_\sigma(\text{src}, \text{src}'; \sigma) k_e(\alpha, \alpha'; l_e) \\
m(\alpha, \text{src}) &= 0 \\
\mu(\alpha, \text{src}) &\sim \text{GP}(m, k) \\
\nu &\sim \text{Gamma}(\nu_\mu, \nu_\sigma) \\
c_z(\alpha, \text{src}) &\sim \text{Normal}(\mu, \nu)
\end{aligned} \tag{27}$$

For a finite set of data points $(\alpha_i, \text{src}_i)_{i=1..|\mathcal{D}|}$ the Gaussian process $\mu \sim \text{GP}(0, k)$ is approximated by a multivariate normal distribution $\boldsymbol{\mu} \sim \text{MvNormal}(\mathbf{0}, \mathbf{K})$, where $\mathbf{K} = k((\alpha_i, \text{src}_i), (\alpha_j, \text{src}_j))_{i,j=1..|\mathcal{D}|}$ is the covariance matrix obtained by evaluating the kernel function k over the finite dataset \mathcal{D} .

Given this parametrization, nominal function f and error function e can be interpreted as GPs dependent on the respective nominal kernel k_f and error kernel k_e :

$$\begin{aligned}
f(\alpha) &\sim \text{GP}(0, \eta^2 k_f) \\
e(\alpha, \text{src}) &\sim \text{GP}(0, k_\sigma k_e + \nu^2 \delta_{\alpha, \alpha'})
\end{aligned} \tag{28}$$

It should be noted, that this model includes a global white noise term $\nu^2 \delta_{\alpha, \alpha'}$ to ensure numerical stability for the sampling algorithms and to account for any residual measurement or truncation errors in the aerodynamic dataset. However, its impact is relatively minor compared to other uncertainties, as demonstrated in Chapter V.

The parameter vector $\boldsymbol{\theta}$ and hyperparameter vector $\boldsymbol{\epsilon}$ are summarized in Eq. (29). Since the components of $\boldsymbol{\mu}$ correspond to evaluations of the GP at each observed data point in \mathcal{D} , and σ defines an independent factor for each source category in \mathcal{S} , the total number of Bayesian parameters is given by $\dim(\boldsymbol{\theta}) = |\mathcal{D}||\mathcal{S}| + 4$. It should be noted that the number of parameters directly depends on the size of the training dataset \mathcal{D} , which is a typical property of nonparametric methods.

$$\begin{aligned}
\theta &= (l_f, l_e, \eta, \nu, \sigma, \mu) \\
\epsilon &= (l_{f,\mu}, l_{f,\sigma}, l_{e,\mu}, l_{e,\sigma}, \eta_\mu, \eta_\sigma, \nu_\mu, \nu_\sigma, \sigma_\mu, \sigma_\sigma)
\end{aligned} \tag{29}$$

Similar to the other Bayesian models, all prior distributions and hyperparameters have been chosen to resemble common weakly informative priors.

In preparation for this study various kernel types and combinations for k_f and k_e were tested. In Chapter V, we present results from the following representative set of GP model instances. These models have been selected because they feature common characteristics of GPs and provide a good fit to the AEDB.

1. GP(SE, WN) Model

Gaussian Process Model, as defined in Eq. (27), with periodic Squared-Exponential (SE) kernel as nominal covariance function $k_f = k_{SE}$ and standard White Noise (WN) kernel as error covariance function $k_e = \delta_{\alpha, \alpha'}$. The periodic SE kernel is constructed from the standard SE kernel via input warping [53].

Due to the selection of a WN error kernel, this model possesses the same source-dependent error characteristics as the Spline Models and Fourier Series Models.

2. GP(SE, SE) Model

Gaussian Process Model, as defined in Eq. (27), with periodic SE kernel as nominal $k_f = k_{SE}$ and as error $k_e = k_{SE}$ covariance function. The periodic SE kernel is constructed from the standard SE kernel via input warping [53].

If the influence of the global WN term $\nu^2 \delta_{\alpha, \alpha'}$ is neglected, this model generates very smooth functions in dependence of α , since the SE kernel generates functions that are infinitely mean-square differentiable [53].

3. GP(SE, $M_{5/2}$) Model

Gaussian Process Model, as defined in Eq. (27), with periodic SE kernel as nominal covariance function $k_f = k_{SE}$ and periodic Matern-5/2 ($M_{5/2}$) kernel as error covariance function $k_e = k_{M_{5/2}}$. Both periodic kernels are constructed from the standard SE and $M_{5/2}$ kernels via input warping [53].

If the influence of the global WN term $\nu^2 \delta_{\alpha, \alpha'}$ is neglected, this model generates smooth functions in dependence of α , since the $M_{5/2}$ kernel generates functions that are 2-times mean-square differentiable [53].

It should be noted that the usability of this model has already been analyzed in a predecessor study [9]. Due to improvements in the parametrization and implementation of this model, the shown numerical results will however slightly differ.

4. $GP(M_{5/2}, M_{1/2})$ Model

Gaussian Process Model, as defined in Eq. (27), with periodic $M_{5/2}$ kernel as nominal covariance function $k_f = k_{M_{5/2}}$ and periodic Matern-1/2 ($M_{1/2}$) kernel as error covariance function $k_e = k_{M_{1/2}}$. Both periodic kernels are constructed from the standard $M_{5/2}$ and $M_{1/2}$ kernels via input warping [53].

If the influence of the global WN term $\nu^2 \delta_{\alpha, \alpha'}$ is neglected, this model generates mean-square continuous functions in dependence of α due to the $M_{1/2}$ kernel [53]. This kernel is also known as Ornstein-Uhlenbeck kernel to describe the Brownian motion of particles [53].

D. Expert-fitted Models

For comparative analysis, this study also reviews a non-Bayesian AEDB model for CALLISTO. This model estimates nominal aerodynamic coefficients and expected uncertainty levels. It has been designed and fitted by a team of aerodynamic experts, the same team behind the CFD simulations and experimental WTT campaigns presented in Chapter II. This model currently sets the baseline for CALLISTO's development and is therefore serving as the primary reference for any alternative approaches. More details on the development of this model are available in [15, 19].

The full AEDB model reflects all functional dependencies between the full input vector (3) and coefficients vector (2). Given the restriction of the problem space (17), the model structure can however be simplified to:

$$\begin{aligned}\mu(\alpha) &= \sum_{k=1}^K \beta_k \phi_k(\alpha) \\ b &\sim \text{Normal}(0, b_z) \\ \sigma(\alpha) &= \max(e_{rel}|\mu|, e_{abs}) \\ c_z(\alpha, src) &= \mu + \sigma b\end{aligned}\tag{30}$$

Here, a piecewise-linear spline $\mu(\alpha)$ with deterministic coefficients β_n is fitted to the TAU LoFi subset of the AEDB, serving as the nominal estimate. The remaining data points of the AEDB are afterwards used to define appropriate absolute and relative error levels, e_{abs} and e_{rel} .

As an expert-fitted non-Bayesian model, there are no probabilistic parameters θ which would be updated during inference. Its deterministic (hyper-)parameters are summarized as:

$$\epsilon = (\beta, e_{rel}, e_{abs}, b_z)\tag{31}$$

The model’s strength lies in its straightforward mathematical structure, previously successful in developing other aerospace vehicles [36–38]. Its simplicity and absence of complex posterior distributions also ensure high computational efficiency. However, it requires significant effort of human resource for development and fitting. Model updates or (hyper-)parameter adjustments are almost impractical with new data, while the (hyper-)parameter selection is inherently subjective due to expert’s assessment. The resource intensity to develop this model was essentially the main motivation to investigate the applicability of Bayesian Inference for AEDB models.

For completeness, the results of this model are referenced as Classical model, as defined below.

1. Classical Model

Non-Bayesian model, as defined in Eq. (30), fitted by human experts to CALLISTO’s AEDB. Nominal estimates are given by piecewise-linear interpolation of the TAU LoFi dataset. Confidence level for the absolute and relative error is set to $b_z = 1/3$, ensuring that the interval $\mu \pm \sigma$ represents the 3-sigma range of the resulting distribution. This model defines currently the baseline for the development of CALLISTO.

V. Inference Results

This study employed a workflow comprising three stages. Firstly, the software implementations of the Bayesian models (Chapter IV) were fitted to the AEDB dataset (Chapter II.D) according to Eq. (11). Subsequently, the posterior predictive distribution was calculated using Eq.(12) for predicting aerodynamic coefficients under new flight conditions. Finally, model validation was conducted using error metrics (Chapter III.C), comparing the predictions with observed AEDB data. Key results from each workflow step are presented in this chapter.

All sampling, either for inference or prediction, was performed on local workstations with 8–16 logical cores and 16–64 GB RAM. A common runtime environment with Ubuntu 22.04 and Python 3.9 was defined and shared between all machines. With this setup, inference runtime ranged from under a minute for Fourier and Spline models to around 30 minutes for GP models. Prediction runtime varied significantly with the number of predicted data points, making a general statement about its computation time impractical.

Due to the usage of sampling algorithms for inference and prediction, all probability distributions are approximated by sets of samples. These results are stored as NetCDF files for further analyses, primarily using the ArviZ Python package [55].

A. Model Fitting

To investigate the usability of the Bayesian models, all models were fitted using three different training setups. In the nominal setup the models are fitted to the entire AEDB dataset, as if they would be used for subsequent development tasks. Additionally, to assess the generalization capabilities of the models via cross-validation, two hold-out training scenarios are established: one excluding specific AEDB data sources, and another excluding subsampled partitions (folds) of the dataset.

1. Nominal Training with Full Dataset

In the Nominal setup, Bayesian models are fitted to the entire AEDB dataset as defined in Section II.D. This maximizes the use of available information contained in the data during inference. Since the resulting posterior predictive models represent the best estimate of aerodynamic coefficients, given the AEDB dataset and a-priori knowledge, they would typically be utilized in downstream development tasks for CALLISTO. In this paper, this setup is referred to as Nominal.

Bayesian Inference for all models has been conducted using MCMC sampling, drawing 1000 samples across four independent parallel chains, after an initial burn-in period. These independent chains allow to perform convergence checks via comparison between the chains. Generally, one can consider convergence to be achieved when no diverging transitions are identified by the sampling algorithm, the \hat{R} -metric is below $\hat{R} \lesssim 1.01$ and the Effective Sample Size (ESS) is above $ESS \gtrsim 400$ [52].

Table 2 Convergence metrics of Bayesian models after posterior inference on Nominal training set.

Model	Divergences	Max \hat{R}	Min ESS_{Bulk}	Min ESS_{Tail}	Max $MCSE_{Mean}$	Max $MCSE_{Std}$
LS(K=12)	0	1.00	3807.52	2510.42	0.0048	0.0044
LS(K=72)	0	1.00	428.45	1020.53	0.0089	0.0063
CS(K=12)	0	1.00	1853.61	2125.17	0.0046	0.0033
CS(K=72)	0	1.02	355.62	847.23	0.0125	0.0089
F(N=6)	0	1.01	1418.43	1871.22	0.0050	0.0035
F(N=36)	0	1.01	1010.15	1650.61	0.0115	0.0082
GP(SE, $M_{5/2}$)	0	1.00	984.65	1528.24	0.0251	0.0178
GP(SE, SE)	0	1.00	2891.80	2189.93	0.0251	0.0184
GP($M_{5/2}$, $M_{1/2}$)	0	1.00	1326.42	1017.99	0.0102	0.0074
GP(SE, WN)	0	1.00	3561.41	2280.59	0.0139	0.0098

Table 2 displays common convergence metrics for the model parameters after Nominal training. For more detailed diagnostics the ESS is separately evaluated for the tail and bulk of the probability distributions. The presented results are indicating satisfactory convergence for all models, except minor deviations for the CS(K=72), which \hat{R} and ESS_{Bulk} is slightly beyond the common threshold. However, these deviations are not deemed to significantly impact the inference validity for this model, as the difference is minimal and could likely be mitigated by increasing the sample size.

Table 2 also reports maximum Monte Carlo Standard Error (MCSE) for mean and standard deviation of predicted coefficients c_z . This error can be interpreted as maximum predictive precision that can be reached due to approximation of the posterior distributions by only a finite amount of samples [52]. It can be observed that the precision of all models is below the first decimal, and for many, even below the second decimal, which is significantly lower than the expected uncertainties in the AEDB. Hence, the approximation of posterior distributions is considered adequate for this study, with only negligible errors from MCMC sampling affecting the results.

The effects of posterior inference are exemplarily illustrated in Figure 6 for the F(N=6) model. Here, the presented plots depict the predictive Probability Density Function (PDF) through color mapping, alongside some exemplary samples in the backdrop, and the mean and Highest Density Interval (HDI) in the foreground. The choice of an 89% HDI is primarily for visualizing dispersion and does not carry any specific statistical significance. Inspired by [56], this value was selected to provide a balanced interpretation while avoiding overemphasis on specific confidence levels. As can be seen, the prior predictive distribution contains practically no information about the AEDB, except the dataset's magnitude and some general functional properties like periodicity and smoothness. This reaffirms the followed modeling principle of employing only weakly informative priors. In contrast, the posterior predictive distribution closely aligns with the aerodynamic data, with significantly reduced, almost invisible uncertainty levels. All information about this distribution was automatically inferred from the given dataset and the prior, without further human interaction. This automatic inference process, which could be demonstrated for all Bayesian models in this study, confirms the efficacy of the Bayesian approach in AEDB modeling.

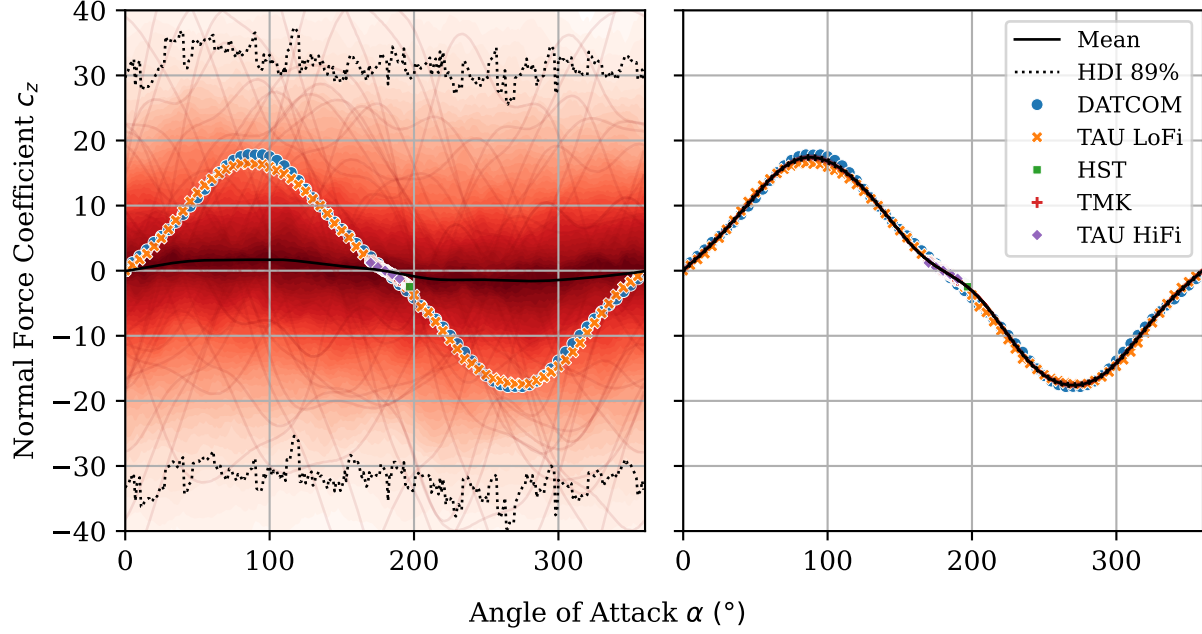


Figure 6 Comparison of prior predictive (left) and posterior predictive (right) distribution of F(N=6) model with the AEDB dataset.

In order to give further insight into the inference process, Figure 7 compares the prior and posterior parameter distributions of the F(N=6) model. Here, amplitude-phase parameterization of the Fourier model has been used according to Eq. (26). It can be observed that two dominant oscillations are inferred from the aerodynamic dataset, corresponding to the Fourier terms $\sin(\alpha)$ and negative $\sin(3\alpha)$, which roughly reflects the global trend of the AEDB dataset. Considering the different scale of the plots, this comparison shows that parameter distributions have become more precise after inference, indicating increased information in the trained model.

2. Training with Hold-Out Set

One major limitation of the Nominal setup is to reliably assess the prediction qualities of the model. The error metrics used to evaluate the prediction qualities (Chapter III.C) compute deviations between a validation set of the AEDB and posterior predictive samples. Since in the Nominal setup the validation sets are also included in the training set, which simply contains all AEDB data, these metrics may overestimate predictive quality for new flight conditions due to in-sample validation.

To better evaluate the generalization qualities of the models, two hold-out setups were defined, excluding either the HST or the TAU HiFi dataset from training. This enables cross-validation of the model independent from the training set, and reflects also the process of predicting yet unobserved FLIGHT data mimicked by HST or TAU HiFi data. The HST and TAU HiFi datasets are selected as suitable hold-out validation sets, because it is assumed that they are most representative for the actual flight configuration of CALLISTO, and also because sufficient alternative data

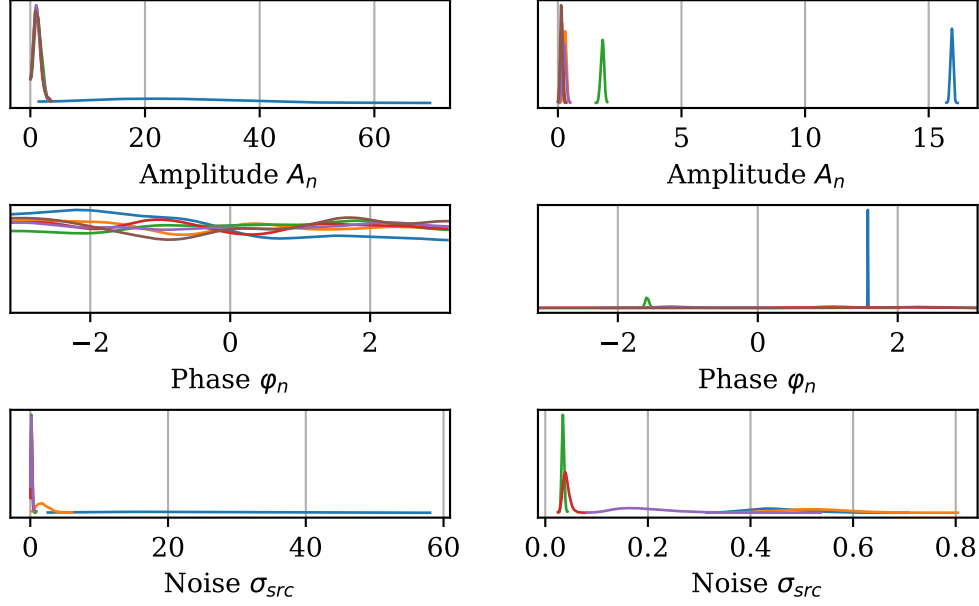


Figure 7 Prior (left) and posterior (right) distributions of F(N=6) model's parameters.

sources are available to conduct meaningful inference. However, since a considerable portion of the AEDB is excluded from training in these setups, it is expected that the prediction qualities are significantly degraded compared to the Nominal-fitted model.

These setups, referred to as $\text{HoldOut}_{\text{HST}}$ and $\text{HoldOut}_{\text{TAU HiFi}}$ (or HoldOut when contextually clear), used the same MCMC sampling configuration as the Nominal training setup. Evaluation of convergence metrics showed that most models achieved good levels of convergence comparable to the Nominal setup, with no direct evidence of divergent samples. Nevertheless, the models $\text{LS}(K=72)$, $\text{CS}(K=72)$, and $\text{CS}(K=12)$ exhibited notable deviations from the common convergence thresholds in the $\text{HoldOut}_{\text{HST}}$ setup, as well as $\text{F}(N=36)$ and $\text{GP}(\text{SE}, \text{WN})$ within the $\text{HoldOut}_{\text{TAU HiFi}}$ setup. These deviations warrant caution when interpreting the corresponding inference results and request further investigation to identify whether they are caused by the sampling algorithms or by the definition of the model and training setups. Although these cases showed deviations from the recommended convergence criteria, their general behavior was deemed adequately reasonable, thus they were not excluded from subsequent analyses.

3. Training over Subsampled Folds

In addition to the hold-out scenarios, stratified K-Fold Cross-Validation has also been applied for out-of-sample evaluation [57]. Here, the AEDB dataset is divided into five equally sized partitions, called *folds*, with four folds used for training and one for validation. To avoid class imbalance problems, the folds are stratified with respect to the *src*

variable as defined in Eq. 5, ensuring balanced AEDB data source representation in each fold. With this setup, referred to as KFold (or KFold_i for specific folds), more comprehensive dataset utilization for inference is possible, compared to the HoldOut setup.

For this training setup the same MCMC sampling configuration as in the Nominal setup is used, and sufficient convergence has been achieved by most folds of most models, with no clear indication of divergent samples. However, the recommended convergence criteria could not be met by one fold each in the GP(SE, M_{5/2}) and GP(SE, SE) models, and by four folds in the F(N=36) model, so that these cases should be interpreted cautiously. Furthermore, posterior inference could not be completed yet for the LS(K=72) and CS(K=72) models due to computational issues. It is assumed that these difficulties stem from the small knot distance in these models in combination with the KFold training setup, which results in certain spline basis functions being devoid of observed data for inference. This can be considered a mitigable limitation of the current software implementation, since the prior distribution should simply be maintained in the absence of observed data.

B. Model Prediction

In a typical development workflow for CALLISTO, the fitted AEDB models would be used in subsequent development tasks to predict aerodynamic coefficients for arbitrary flight conditions that could occur during the flight campaign. However, for the investigated Bayesian models an explicit dependency on the data source *src* has been modelled to consider different error characteristics during posterior inference. This reliance on explicit data sources necessitated the creation of a new FLIGHT category for the prediction of in-flight condition. The priors on this FLIGHT category are configured to minimize the influence of the error function $e(\alpha, src)$ as outlined in Eq. (20) in comparison to the nominal function $f(\alpha)$. This choice is based on the hypothesis, that the nominal function inferred from the AEDB represents the vehicle's in-flight behavior.

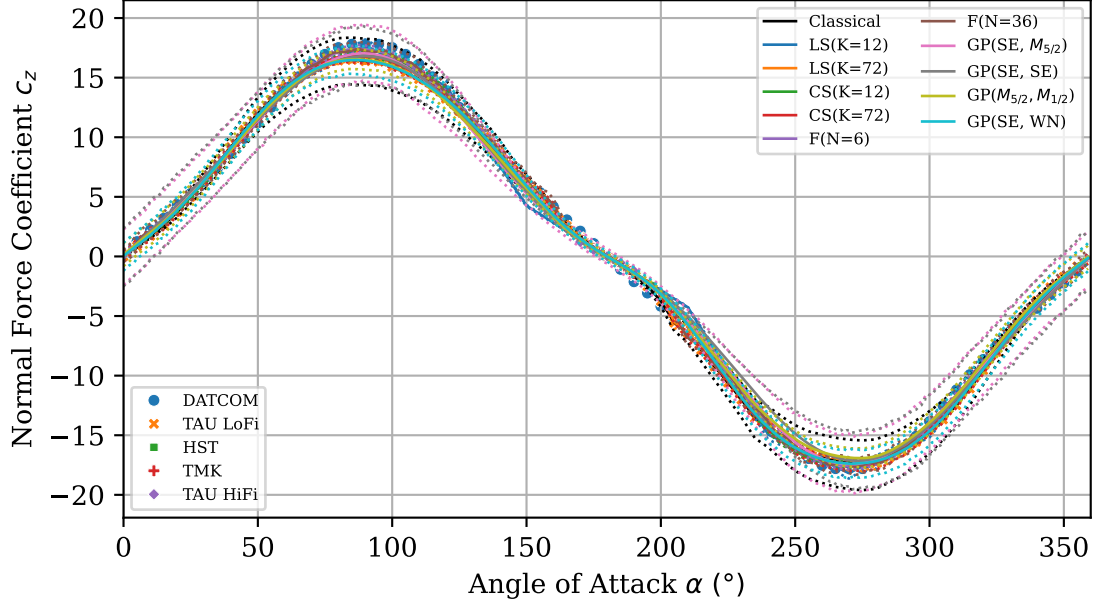


Figure 8 Comparison of posterior predictive distributions with AEDB dataset, after Nominal training and prediction for $src = \text{FLIGHT}$, indicating mean and 89% HDI.

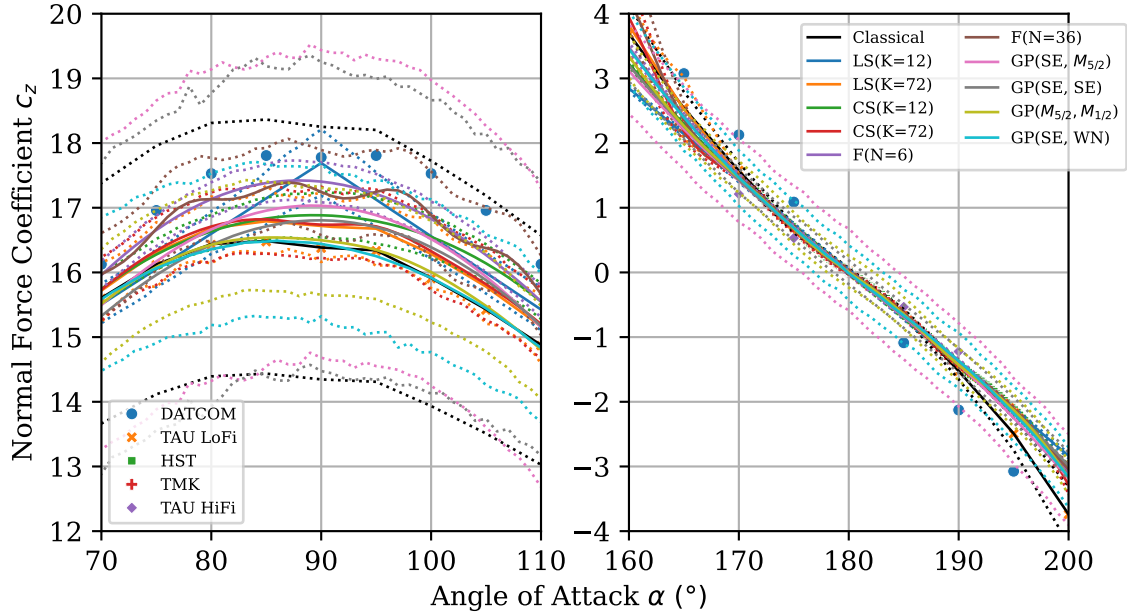


Figure 9 Zoomed comparison of posterior predictive distributions with AEDB dataset, after Nominal training and prediction for $src = \text{FLIGHT}$, indicating mean and 89% HDI.

Figures 8 and 9 provide an overview of the prediction results for all AoA with $src = \text{FLIGHT}$. For all models it can be observed that the mean of the posterior predictive distribution, representing the nominal estimate for aerodynamic coefficients, aligns well with the global behavior of the AEDB. The uncertainties in the coefficients, depicted by the 89%

HDI, narrow around $\alpha \approx 170^\circ \dots 190^\circ$ due to the higher data density in this range. However, the spread of uncertainties varies significantly across models. Notably, in Figure 9, several AEDB data points fall outside the HDIs of certain models.

Upon closer examination, Figure 8 and the left plot in Figure 9 reveal that the LS(K=12) model exhibits several corners in its trend, due to its piecewise-linear nature and relatively wide knot spacing of 30° . This non-physical behavior suggests that this model underfits the AEDB dataset, which has an AoA resolution of 5° or finer. Conversely, the F(N=36) model shows small oscillations around $\alpha \approx 90^\circ$ and $\alpha \approx 270^\circ$, which are particularly visible in the left plot of Figure 9. This behavior may indicate slight overfitting of this model to the dense data region around $\alpha \approx 170^\circ \dots 190^\circ$. For the remaining models no similar behavior could be observed.

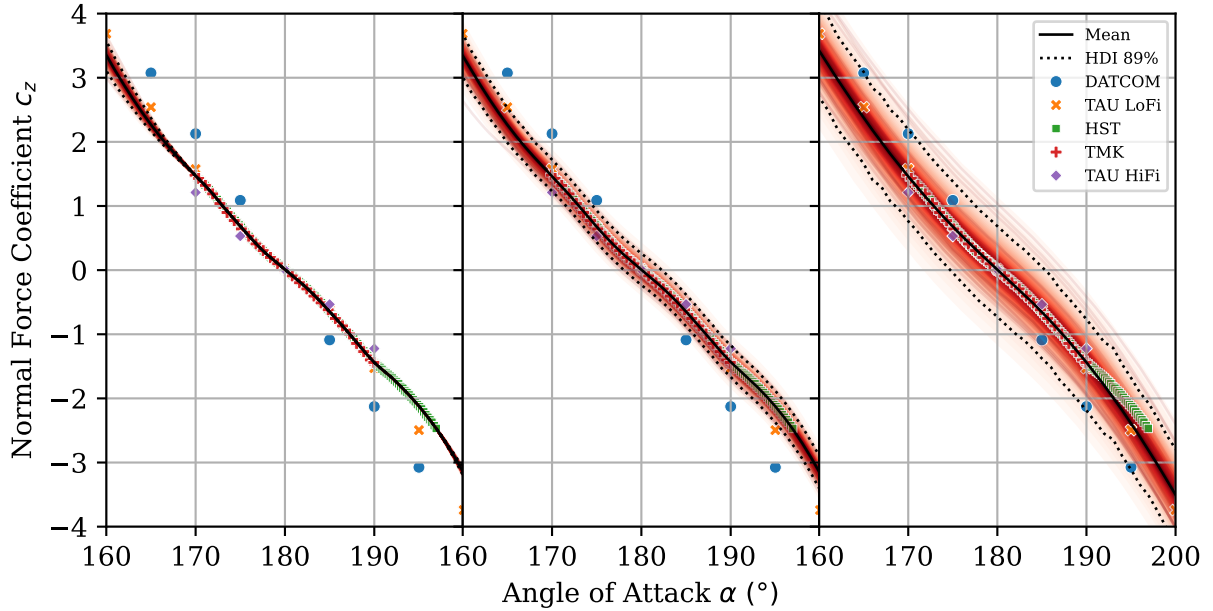


Figure 10 Comparison of posterior predictive distribution for different training and prediction setups of $GP(M_{5/2}, M_{1/2})$ model. Left: Nominal(HST); Center: Nominal(FLIGHT); Right: HoldOut_{HST}(FLIGHT).

Generally, the prediction results of the Bayesian models are heavily influenced by training and prediction setups, as illustrated for the $GP(M_{5/2}, M_{1/2})$ model in Figure 10. The left subplot shows the Nominal trained model predicting aerodynamic coefficient for $src = \text{HST}$, denoted as Nominal(HST). It can be seen that this setup fits the HST data very tightly, since the error behavior of the HST data is explicitly considered for prediction, but lacking generalization to the other datasets. Due to this lack of generalization, such a prediction setup which incorporates also the error behavior of the data source may be used to evaluate the quality of fit, but should not be utilized to assess the prediction quality of a model for flight conditions. The center subplot displays results for the same Nominal training setup but predicting for $src = \text{FLIGHT}$, identified as Nominal(FLIGHT). This setup demonstrates significantly better generalization capability, such that the predicted HDI covers almost all dataset, except DATCOM. Also the right subplot visualizes prediction

results for $src = \text{FLIGHT}$ but for the $\text{HoldOut}_{\text{HST}}$ training setup, denoted as $\text{HoldOut}_{\text{HST}}(\text{FLIGHT})$. It shows broader uncertainty levels and a slight bias between the mean and HST data around $\alpha \approx 190^\circ \dots 200^\circ$, due to the significantly reduced training dataset. Therefore, it is assumed that the HoldOut training setups generally underestimate the predictive quality of a model, due to this non-negligible lack of information at training.

Given the limited amount of AEDB data and the training setups defined in Section V.A, the Nominal(FLIGHT) setup is presumed to most realistically assess model prediction quality in the scope of this study, though the risk of overfitting due to in-sample validation necessitates verification via HoldOut or KFold data.

On the contrary, the non-Bayesian Classical model requires no distinction between training or prediction setups, since training has been conducted by a team of aerodynamic experts and no explicit data source dependency is modelled, which can be interpreted as an implicit $src = \text{FLIGHT}$ setup. However, due to this fixed setup also no out-of-sample validation can formally be performed with the Classical model.

C. Model Validation and Comparison

To validate the usability of the AEDB models, and more specifically to compare their quality to predict aerodynamic coefficients for arbitrary flight conditions, the error metrics MRE, RMSE, and MAD defined in Section III.C are used to quantify the deviations between the observed AEDB data and the predictions. For CALLISTO no clear error metric is specified by the AEDB model users which shall be minimized to reduce the development risk, but it is currently assumed that users' needs are most appropriately reflected by the RMSE because of its balanced treatment of outliers. Considering the previous discussions about the models' training and prediction setups, the Nominal(FLIGHT) setup is therefore deemed most representative for assessing prediction quality, if sufficient generalization capability can be verified by HoldOut or KFold results.

An overview of the error metrics calculated for the HST and TAU HiFi validation sets is displayed in Figures 11 and 12. These figures visualize the mean (markers), interquartile range (thick lines) and 89% HDI (thin lines) of the different metrics across multiple *Training(Prediction)* setups. The HDI of the Classical model is specifically highlighted in green to facilitate visual comparison. The interpretation of these plots is as follows: When comparing the error metric of two models for the same setup, non-overlapping error bars indicate a statistically significant difference in predictive performance concerning the selected metric and setup. In contrast, overlapping error bars do not allow a definitive conclusion but may suggest a discernible trend between the models.

Figures 11 and 12 reveal that the RMSE and MRE of all Spline Models and Fourier Series Models are notably lower than those of the Classical model under the Nominal(FLIGHT) setup. For the MAD this separation is not so clear, however, it is assumed that this metric is not so relevant for the potential model users. Consequently, it can be supposed

that the prediction quality on the HST and TAU HiFi validation sets is better for these Bayesian models than for the Classical model. Among these, the CS(K=72) model generally exhibits the lowest RMSE, but a statistically significant distinction between the Spline and Fourier models is difficult to make.

Further comparison with the RMSE for the HoldOut(FLIGHT) setup indicates significant generalization errors for the LS(K=12) model. As highlighted in Figure 8 and 9, the large knot distance of the LS(K=12) model likely leads to underfitting, accounting for the observed deviation. All other Spline and Fourier models demonstrate notable but minor generalization errors in HoldOut validation. However, the absence of HoldOut data for the Classical model precludes direct comparison. Furthermore, the available KFold(FLIGHT) data resembles the Nominal(FLIGHT) setups very closely, so that good generalization capabilities of the models can be assumed. Based on these observations, it can be concluded that Fourier and Spline Models, with appropriate parameter selection, significantly outperform the expert-fitted Classical model in the definition range of the HST and TAU HiFi validation sets of $\alpha \approx 160^\circ \dots 200^\circ$.

For GP models, such clear distinctions are not evident. GP(SE, SE) and GP($M_{5/2}$, $M_{1/2}$) demonstrate lower RMSE for the Nominal(FLIGHT) setup compared to the Classical model. Considering the HoldOut setups as well, it can be observed that especially the GP(SE, SE) model shows only a small generalization error. However, their superiority over the Classical model on the HST and TAU HiFi validation sets is not statistically significant within this study. Compared to Fourier and Spline models, the GP models show higher RMSE on the HST set but are mostly indifferent on the TAU HiFi set, suggesting that Fourier and Spline models may still have superior prediction quality. The other GP models

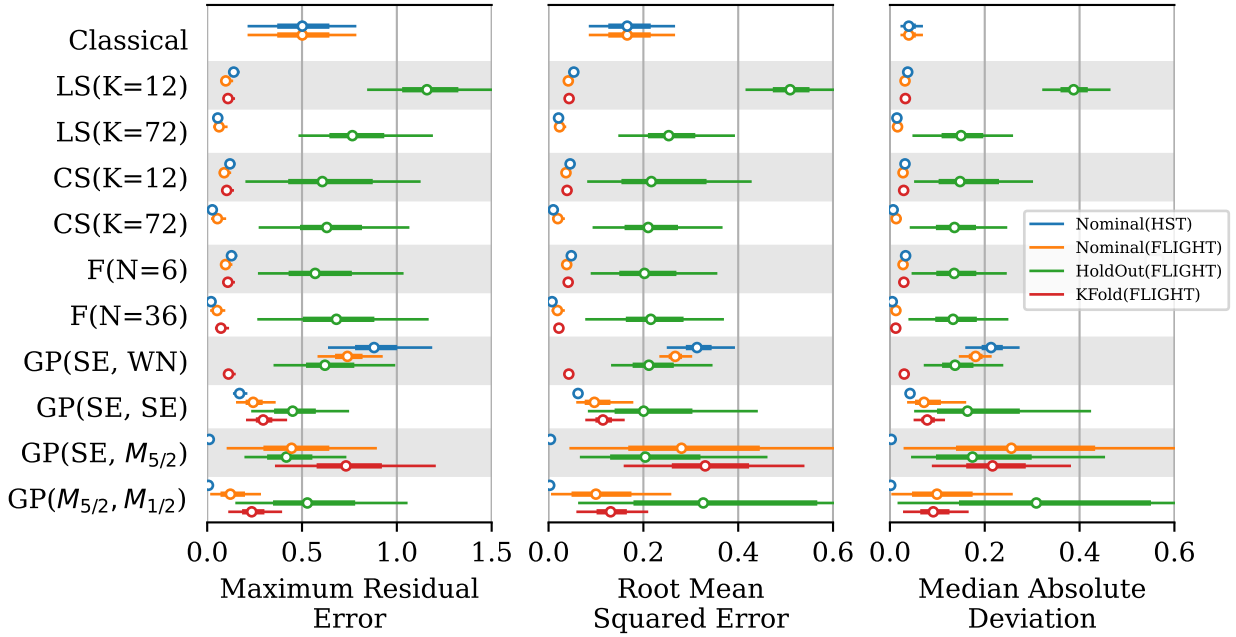


Figure 11 Error metrics calculated on the HST validation dataset for different training and prediction setups.

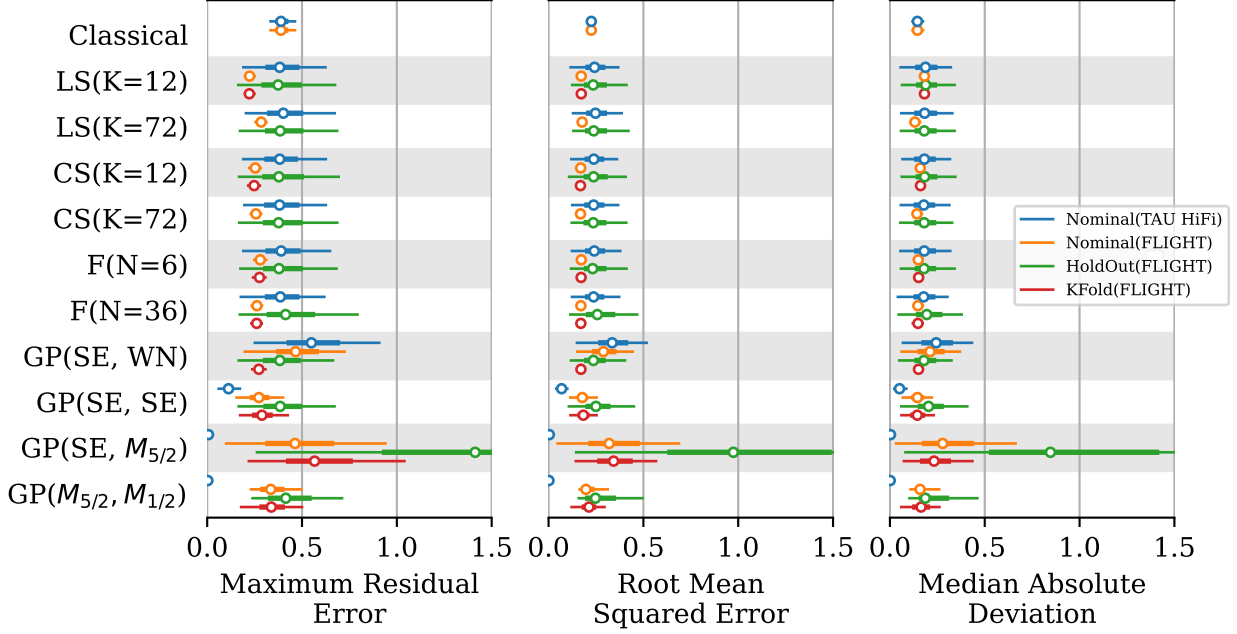


Figure 12 Error metrics calculated on the TAU HiFi validation dataset for different training and prediction setups.

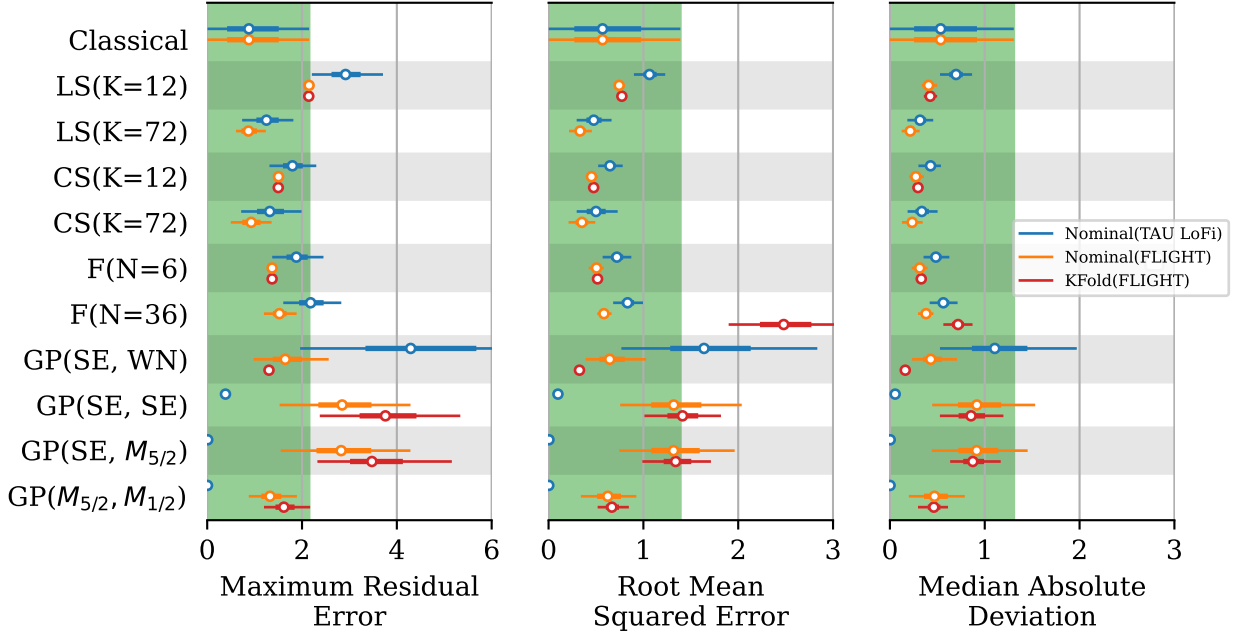


Figure 13 Error metrics calculated on the TAU LoFi validation dataset for different training and prediction setups.

however do not demonstrate clear advantages. Notably, $GP(SE, M_{5/2})$ and $GP(M_{5/2}, M_{1/2})$ show negligible errors for both Nominal(HST) and Nominal(TAU HiFi) setups, which aligns with the conclusions of Section V.B that predictions are overfitting the AEDB if the error behavior of the data source is considered.

To assess the overall predictive quality of Bayesian models, Figure 13 illustrates the error metrics for the TAU LoFi validation set. Due to the absence of HoldOut data for this validation set, the reported global prediction quality might be overestimated. The LS(K=72) and CS(K=72) models demonstrate the lowest RMSE for the Nominal(FLIGHT) setup, marginally outperforming the Classical model, albeit not with statistical significance. Given the direct interpolation of the TAU LoFi dataset by the Classical model, it is assumed that the shown prediction quality for this model is overestimated and not generalizable to flight data. Other Fourier and Spline models, along with the GP(SE, WN) and GP($M_{5/2}$, $M_{1/2}$) models, exhibit prediction qualities on par with the Classical model throughout the AoA range of the TAU LoFi validation set. The consistency between KFold(FLIGHT) and Nominal(FLIGHT) data among most models underscores their generalization capabilities over global AoA scales, except for the F(N=36) model. This model's tendency to overfit was previously noted in Figure 9 and is confirmed by its elevated RMSE in the KFold scenario.

VI. Conclusions and Outlook

This research has demonstrated the efficacy of Bayesian models in enhancing the prediction accuracy and generalizability of AEDB models for RLVs, particularly for the CALLISTO project. The application of Bayesian modeling has shown clear advantages over traditional empirical-heuristic methods in terms of precision and reliability. Especially, the natively supported estimation of uncertainties is a distinct feature of Bayesian AEDB models, which is difficult to achieve with classical approaches. Our findings highlight that Bayesian models, specifically Fourier Series Models and Spline Models with adequately chosen number of parameters, have significantly outperformed the expert-fitted Classical Model in predicting the aerodynamic normal force coefficient c_z of CALLISTO under varying flight conditions.

The comparative analysis of different Bayesian models, utilizing error metrics like RMSE, MRE, and MAD, confirms that these models provide a more accurate representation of the modeled aerodynamic behavior, as evidenced by their lower prediction errors compared to traditional models. Particularly, Spline models (e.g., CS(K=72)) have exhibited promising results in terms of lower RMSE and generalization capabilities. Furthermore, some GP models, notably GP(SE, SE) and GP($M_{5/2}$, $M_{1/2}$), have shown potential, albeit with less statistical significance compared to the Fourier and Spline models.

The study's methodology, involving model definition, fitting, and error analysis, has successfully addressed the complex aerodynamic modeling challenges for RLV systems like CALLISTO. Demonstrating automated parameter inference across various types of Bayesian models showcased the significant potential in alleviating the workload of human experts. Although only a subset of CALLISTO's AEDB has been used to investigate the applicability of Bayesian Inference, it is assumed that the drawn conclusions are likely transferable to the full dataset, including other aerodynamic coefficients, and to AEDBs in general. However, further investigations would be necessary to validate this generalization.

In future studies we intend to investigate further model setups, particularly with more versatile error behavior, and plan to solidify the error analysis via cross-validation. Furthermore, also other dependent and independent variables of CALLISTO's AEDB shall be included in the models, so that a full Bayesian characterization of the vehicle aerodynamics can be achieved. Future research should also focus on exploring the applicability to other aspects of RLV design and operations, leveraging uncertainty information to enhance the precision and reliability of space missions.

In summary, this research demonstrates the applicability and benefits of Bayesian models in aerodynamic modeling for RLVs. By improving predictive accuracy and generalization, these models contribute to the CALLISTO project and offer a promising approach for AEDB development in aerospace engineering. Since Bayesian Inference is a widely applicable methodology, its use in aerodynamic modeling may also inspire further advancements in other engineering domains.

Acknowledgments

The authors would like to thank CNES and JAXA for the exemplary cooperation in the CALLISTO project. Without our joint effort this project would have never been possible. Besides that, we would like to thank Pavan Tummala, Alexander Trey, Etienne Dumont and Jascha Wilken for their support of this study. Furthermore, we would like to acknowledge the assistance of GPT-4 [58] in enhancing the grammar and wording used in this paper.

References

- [1] Dumont, E., Sagliano, M., Eichel, S., Markgraf, M., Glaser, T., Braun, B., Windelberg, J., Petkov, I., Schröder, S., Witte, L., Schneider, A., Panthen, B., Häseker, J. S., Opp, L., Krieger, A. T. E., Stief, M., Woicke, S., Martens, H., and Krummen, S., "CALLISTO: Reusable rocket stage demonstrator: getting ready for implementation," *34th International Symposium on Space Technology and Science*, Fukuoka, Japan, June 2023. <https://elib.dlr.de/197208/>.
- [2] Dumont, E., Ishimoto, S., Illig, M., Sagliano, M., Solari, M., Ecker, T., Martens, H., Krummen, S., Desmariaux, J., Saito, Y., Ertl, M., Klevanski, J., Reimann, B., Woicke, S., Schwarz, R., Seelbinder, D., Markgraf, M., Riehmer, J., Braun, B., and Aicher, M., "CALLISTO: Towards Reusability of a Rocket Stage: Current Status," *Journal of Evolving Space Activities*, Vol. 1, No. 75, October 2023. <https://doi.org/10.57350/jesa.75>.
- [3] Dumont, E., Illig, M., Ishimoto, S., Chavagnac, C., Saito, Y., Krummen, S., Eichel, S., Martens, H., Giagkozoglou, S., Häseker, J. S., Ecker, T., Klevanski, J., Krziwianie, F., Rotärmel, W., Schröder, S., Schneider, A., Grimm, C., Woicke, S., Sagliano, M., Schlotterer, M., Markgraf, M., Braun, B., Aicher, M., Briese, L. E., Petkov, I., Riehmer, J., and Reimann, B., "CALLISTO: A Prototype Paving the Way for Reusable Launch Vehicles in Europe and Japan," *73rd International Astronautical Congress*, Paris, France, September 2022, IAC-22-D2.6.1. <https://elib.dlr.de/188663/>.

- [4] Krummen, S., Desmariaux, J., Saito, Y., Boldt, M., Briese, L. E., Cesco, N., Chavagnac, C., Cliquet-Moreno, E., Dumont, E., Ecker, T., Eichel, S., Ertl, M., Giagkozoglou, S., Glaser, T., Grimm, C., Illig, M., Ishimoto, S., Klevanski, J., Lidon, N., Mierheim, O., Niccolai, J.-F., Reershemius, S., Reimann, B., Riehmer, J., Sagliano, M., Scheufler, H., Schneider, A., Schröder, S., Schwarz, R., Seelbinder, D., Stief, M., Windelberg, J., and Woicke, S., "Towards a Reusable First Stage Demonstrator: CALLISTO - Technical Progresses & Challenges," *72nd International Astronautical Congress*, Dubai, UAE, October 2021, IAC-21-D2.6.1. <https://elib.dlr.de/147143/>.
- [5] Rickmers, P., Dumont, E., Krummen, S., Redondo Gutierrez, J. L., Bussler, L., Kottmeier, S., Wübbels, G., Martens, H., Woicke, S., Sagliano, M., Häseker, J. S., Witte, L., Sippel, M., Bauer, W., and Peetz, H.-J., "The CALLISTO and ReFEx Flight Experiments at DLR - Challenges and Opportunities of a Wholistic Approach," *Ascension Conference*, Dresden, Germany, September 2023. <https://elib.dlr.de/197696/>.
- [6] "CALLISTO," YouTube, event - dlr - de, October 2020. <https://www.youtube.com/watch?v=wAJ6cwDEme8>, retrieved 15 March 2024.
- [7] Giagkozoglou, S., Eichel, S., Rotärmel, W., Krziwianie, F., Petkov, I., Dumont, E., Schneider, A., Schröder, S., Windelberg, J., Ecker, T., and Ertl, M., "Development of Reusable Structures and Mechanisms for CALLISTO," *Journal of Evolving Space Activities*, Vol. 1, No. 36, October 2023. <https://doi.org/10.57350/jesa.36>.
- [8] Guéron, S., Ishimoto, S., Dumont, E., Tatiossian, P., Chavagnac, C., Desmariaux, J., Monchaux, D., Frenoy, O., Moreno, E. C., Deremaux, C., Lidon, N., Cesco, N., Witte, L., Sagliano, M., Seelbinder, D., Klevanski, J., Ecker, T., Reimann, B., Riehmer, J., Ertl, M., and Krummen, S., "CALLISTO DEMONSTRATOR: Focus on system aspects," *71th International Astronautical Congress*, Dubai, UAE, October 2020, IAC-20-D2.6.1. <https://elib.dlr.de/138808/>.
- [9] Krummen, S., Tummala, P., Wilken, J., Dumont, E., Ertl, M., Ecker, T., Riehmer, J., and Klevanski, J., "Applying Bayesian Inference to Estimate Uncertainties in the Aerodynamic Database of CALLISTO," *2023 IEEE Aerospace Conference*, Big Sky, Montana, USA, March 2023. <https://doi.org/10.1109/AERO55745.2023.10115932>.
- [10] Klevanski, J., Ecker, T., Riehmer, J., Reimann, B., Dumont, E., and Chavagnac, C., "Aerodynamic Studies in Preparation for CALLISTO - Reusable VTVL Launcher First Stage Demonstrator," *69th International Astronautical Congress*, Bremen, Germany, October 2018, IAC-18-D2.6.3. <https://elib.dlr.de/122062/>.
- [11] Riehmer, J., Marwege, A., Klevanski, J., Gülhan, A., and Dumont, E., "Subsonic and Supersonic Ground Experiments for the CALLISTO VTVL Launcher Demonstrator," *International Conference on Flight Vehicles, Aerothermodynamics and Re-entry Missions & Engineering*, Monopoli, Italien, September 2019. <https://elib.dlr.de/140709/>.
- [12] Marwege, A., Riehmer, J., Klevanski, J., Gülhan, A., Ecker, T., Reimann, B., and Dumont, E., "First Wind Tunnel Data of CALLISTO - Reusable VTVL Launcher First Stage Demonstrator," *8th European Conference for Aeronautics and Space Sciences*, Madrid, Spain, July 2019. <https://doi.org/10.13009/EUCASS2019-350>.

- [13] Marwege, A., Riehmer, J., Klevanski, J., Gülhan, A., and Dumont, E., “Wind Tunnel investigations in CALLISTO - Reusable VTVL Launcher First Stage Demonstrator,” *70th International Astronautical Congress*, Washington D.C., USA, October 2019, IAC-19-D2-6x50636. <https://elib.dlr.de/132573/>.
- [14] Schneider, A., Desmariaux, J., Klevanski, J., Schröder, S., and Witte, L., “Deployment dynamics analysis of CALLISTO’s approach and landing system,” *CEAS Space Journal*, Vol. 15, December 2021, pp. 343–356. <https://doi.org/10.1007/s12567-021-00411-2>.
- [15] Klevanski, J., Reimann, B., Krummen, S., Ertl, M., Ecker, T., Riehmer, J., and Dumont, E., “Progress in Aerodynamic Studies for CALLISTO - Reusable VTVL Launcher First Stage Demonstrator,” *9th European Conference for Aeronautics and Space Sciences*, Lille, France, June 2022. <https://doi.org/10.13009/EUCASS2022-4689>.
- [16] Riehmer, J., Kapteijn, K., Klevanski, J., Reimann, B., Krummen, S., Gülhan, A., and Dumont, E., “Wind Tunnel Experiments of the CALLISTO VTVL Launcher in the TMK and HST Wind Tunnels,” *9th European Conference for Aeronautics and Space Sciences*, Lille, France, June 2022. <https://doi.org/10.13009/EUCASS2022-4634>.
- [17] Ertl, M., Ecker, T., Klevanski, J., Dumont, E., and Krummen, S., “Aerothermal analysis of plume interaction with deployed landing legs of the CALLISTO vehicle,” *9th European Conference for Aeronautics and Space Science*, Lille, France, June 2022. <https://doi.org/10.13009/EUCASS2022-4686>.
- [18] Ertl, M., and Ecker, T., “Aerodynamic and aerothermal comparison between the CAL1C and CAL1D geometries for the CALLISTO vehicle,” *10th European Conference for Aeronautics and Space Sciences*, Lausanne, Switzerland, July 2023. <https://doi.org/10.13009/EUCASS2023-431>.
- [19] Klevanski, J., Reimann, B., Krummen, S., Ertl, M., Ecker, T., Riehmer, J., Dumont, E., Briese, L. E., and Kier, T., “Further Progress in Aerodynamic Studies for CALLISTO - Reusable VTVL Launcher First Stage Demonstrator,” *10th European Conference for Aeronautics and Space Sciences*, Lausanne, Switzerland, July 2023. <https://doi.org/10.13009/EUCASS2023-510>.
- [20] Briese, L. E., Kier, T., Petkov, I., Windelberg, J., Heinrich, L., and Krummen, S., “Advanced Modeling and Dynamic Stability Analysis of the Aerodynamic Control Surfaces of CALLISTO,” *AIAA SciTech Forum*, National Harbor, USA, January 2023. <https://doi.org/10.2514/6.2023-2405>.
- [21] Ecker, T., Karl, S., Dumont, E., Stappert, S., and Krause, D., “Numerical Study on the Thermal Loads During a Supersonic Rocket Retropropulsion Maneuver,” *Journal of Spacecraft and Rockets*, Vol. 57, No. 1, October 2019, pp. 131–146. <https://doi.org/10.2514/1.A34486>.
- [22] Laureti, M., and Karl, S., “Aerothermal databases and load predictions for Retro Propulsion-Assisted Launch Vehicles (RETALT),” *CEAS Space Journal*, Vol. 14, January 2022, pp. 501–515. <https://doi.org/10.1007/s12567-021-00413-0>.

- [23] Hannemann, K., Martinez Schramm, J., Wagner, A., Karl, S., and Hannemann, V., “A Closely Coupled Experimental and Numerical Approach for Hypersonic and High Enthalpy Flow Investigations Utilising the HEG Shock Tunnel and the DLR TAU Code,” *NATO Research and Technology Organisation, AVT-186 RTO AVT/VKI Lecture Series*, 2010, pp. 8–1 – 8–66. <https://elib.dlr.de/69521/>.
- [24] Schwamborn, D., Gerhold, T., and Heinrich, R., “The DLR TAU-code: recent applications in research and industry,” *European Conference on Computational Fluid Dynamics*, Egmond aan Zee, The Netherlands, September 2006. <https://elib.dlr.de/22421/>.
- [25] Thornber, B., Mosedale, A., Drikakis, D., Youngs, D., and Williams, R., “An improved reconstruction method for compressible flows with low Mach number features,” *Journal of Computational Physics*, Vol. 227, No. 10, 2008, pp. 4873–4894. <https://doi.org/10.1016/j.jcp.2008.01.036>.
- [26] Spalart, P., and Allmaras, S., “A one-equation turbulence model for aerodynamic flows,” *30th Aerospace Sciences Meeting and Exhibit*, Reno, USA, January 1992. <https://doi.org/10.2514/6.1992-439>.
- [27] Ecker, T., Ertl, M., Klevanski, J., Krummen, S., and Dumont, E., “Aerothermal characterization of the CALLISTO vehicle during descent,” *9th European Conference for Aeronautics and Space Sciences*, Lille, France, June 2022. <https://doi.org/10.13009/EUCASS2022-4680>.
- [28] Versteeg, H. K., and Malalasekera, W., *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*, 2nd ed., Pearson Education Limited.
- [29] Celik, I., Ghia, U., Roache, P., Freitas, C., Coloman, H., and Raad, P., “Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Application,” *Journal of Fluids Engineering*, Vol. 130, No. 7, July 2008, pp. 078001–1 – 078001–4. <https://doi.org/10.1115/1.2960953>.
- [30] Ecker, T., Karl, S., Dumont, E., Stappert, S., and Krause, D., “Numerical Study on the Thermal Loads During a Supersonic Rocket Retropropulsion Maneuver,” *Journal of Spacecraft and Rockets*, Vol. 57, No. 1, October 2019, pp. 131–146. <https://doi.org/10.2514/1.A34486>.
- [31] Krummen, S., and Sippel, M., “Effects of the rotational vehicle dynamics on the ascent flight trajectory of the SpaceLiner concept,” *CEAS Space Journal*, Vol. 11, No. 2, November 2018, pp. 161–172. <https://doi.org/10.1007/s12567-018-0223-7>.
- [32] Rosema, C., Doyle, J., Auman, L., Underwood, M., and Blake, W. B., “Missile DATCOM User’s Manual - 2011 Revision,” Tech. rep., U.S. Army Aviation and Missile Research, Development and Engineering Center, March 2011. <https://apps.dtic.mil/sti/citations/ADA548461>.
- [33] Space Systems Group, “Aerodynamic Design Data Book Orbiter Vehicle STS-1,” Tech. rep., Rockwell International, 1980.
- [34] Young, J. C., and Underwood, J. M., “Development of Aerodynamic Uncertainties for the Space Shuttle Orbiter,” *Journal of Spacecraft*, Vol. 20, No. 6, 1983, pp. 513–517. <https://doi.org/10.2514/3.8580>.

- [35] Weil, J., and G. P. B., “Correlation Of Predicted And Flight Derived Stability And Control Derivatives With Particular Application To Tailless Delta Wing Configurations,” Tech. Rep. NASA TM-81361, NASA, 1981.
- [36] Wartemann, V., Konosidou, N., Flock, A. K., and Merrem, C. H.-J., “Contribution of Numerical and Experimental Flow Simulations to the Aerodynamic Data Base of the DLR Reusable Flight Experiment ReFEx,” *New Results in Numerical and Experimental Fluid Mechanics*, Vol. 151, July 2021, pp. 141–150. https://doi.org/10.1007/978-3-030-79561-0_14.
- [37] Schettino, A., Pezzella, G., Marini, M., Di Benedetto, S., Villace, V. F., Steelant, J., Choudhury, R., Gubanov, A., and Voevodenko, N., “Aerodynamic database of the HEXAFLY-INT hypersonic glider,” *CEAS Space Journal*, Vol. 12, No. 2, 2022, pp. 295–311. <https://doi.org/10.1007/s12567-020-00299-4>.
- [38] Roncioni, P., Vitagliano, P. L., De Gregorio, F., Paglia, F., and Milana, C., “Aerodatabase of Vega-C Launcher Development and Integration,” *8th European Conference for Aeronautics and Space Sciences*, Madrid, Spain, July 2019. <https://doi.org/10.13009/EUCASS2019-130>.
- [39] Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., Norman, A. J., Monrad, J. T., Besiroglu, T., Ge, H., Hartwick, M. A., Teh, Y. W., Chindelevitch, L., Gal, Y., and Kulveit, J., “Inferring the effectiveness of government interventions against COVID-19,” *Science*, Vol. 371, No. 6531, 2021, p. eabd9338. <https://doi.org/10.1126/science.abd9338>.
- [40] Graham, N. A. J., Jennings, S., MacNeil, M. A., Mouillot, D., and Wilson, S. K., “Predicting climate-driven regime shifts versus rebound potential in coral reefs,” *Nature*, Vol. 518, No. 7537, 2015, pp. 94–97. <https://doi.org/10.1038/nature14140>.
- [41] Smith, R. J. E., Ashton, G., Vajpeyi, A., and Talbot, C., “Massively parallel Bayesian inference for transient gravitational-wave astronomy,” *Monthly Notices of the Royal Astronomical Society*, Vol. 498, No. 3, August 2020, pp. 4492–4502. <https://doi.org/10.1093/mnras/staa2483>.
- [42] Hoffer, J. G., Geiger, B. C., and Kern, R., “Gaussian process surrogates for modeling uncertainties in a use case of forging superalloys,” *Applied Sciences*, Vol. 12, No. 3, 2022, p. 1089. <https://doi.org/10.3390/app12031089>.
- [43] Morita, Y., Rezaeiravesh, S., Tabatabaei, N., Vinuesa, R., Fukagata, K., and Schlatter, P., “Applying Bayesian optimization with Gaussian process regression to computational fluid dynamics problems,” *Journal of Computational Physics*, Vol. 449, 2022, p. 110788. <https://doi.org/10.1016/j.jcp.2021.110788>.
- [44] Renganathan, S. A., Harada, K., and Mavris, D. N., “Aerodynamic data fusion toward the digital twin paradigm,” *AIAA Journal*, Vol. 58, No. 9, September 2020, pp. 3902–3918. <https://doi.org/10.2514/1.J059203>.
- [45] Gelman, A., Vehtari, A., Riddell, A., Bales, B., Wilson, P., Lawrence, M., and Betancourt, M., “Prior Choice Recommendations,” GitHub, April 2020. <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>, retrieved 12 September 2023.
- [46] Stan Development Team, *Stan User’s Guide Version 2.29*, Feb 2022. https://mc-stan.org/docs/2_29/stan-users-guide-2_29.pdf.

- [47] Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., and Zinkov, R., “PyMC: A Modern and Comprehensive Probabilistic Programming Framework in Python,” *PeerJ Computer Science*, Vol. 9, September 2023, p. e1516. <https://doi.org/10.7717/peerj-cs.1516>.
- [48] Phan, D., Pradhan, N., and Jankowiak, M., “Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro,” *arXiv Preprint*, December 2019, 1912.11554. <https://doi.org/10.48550/arXiv.1912.11554>.
- [49] 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, No. 47, January 2014, pp. 1593–1623. <http://jmlr.org/papers/v15/hoffman14a.html>.
- [50] Gneiting, T., and Raftery, A. E., “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American statistical Association*, Vol. 102, No. 477, January 2007, pp. 359–378. <https://doi.org/10.1198/016214506000001437>.
- [51] Gneiting, T., “Making and evaluating point forecasts,” *Journal of the American Statistical Association*, Vol. 106, No. 494, March 2011, pp. 746–762. <https://doi.org/10.1198/jasa.2011.r10138>.
- [52] Martin, O. A., Kumar, R., and Lao, J., *Bayesian Modeling and Computation in Python*, 1st ed., Chapman and Hall / CRC, Boca Raton, USA. <https://doi.org/10.1201/9781003019169>.
- [53] Williams, C. K. I., and Rasmussen, C. E., *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, USA. <https://doi.org/10.7551/mitpress/3206.001.0001>.
- [54] Duvenaud, D., “Automatic model construction with Gaussian processes,” Ph.D. thesis, University of Cambridge, November 2014. <https://doi.org/10.17863/CAM.14087>.
- [55] Kumar, R., Carroll, C., Hartikainen, A., and Martin, O., “ArviZ a unified library for exploratory analysis of Bayesian models in Python,” *Journal of Open Source Software*, Vol. 4, No. 33, 2019, p. 1143. <https://doi.org/10.21105/joss.01143>.
- [56] McElreath, R., *Statistical rethinking: A Bayesian course with examples in R and Stan*, 2nd ed., Chapman and Hall / CRC, Boca Raton, USA. <https://doi.org/10.1201/9780429029608>.
- [57] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, Vol. 12, No. 85, November 2011, pp. 2825–2830. <https://www.jmlr.org/papers/v12/pedregosa11a.html>.
- [58] OpenAI, “GPT-4 Technical Report,” *arXiv Preprint*, March 2023. <https://doi.org/10.48550/arXiv.2303.08774>.