



Efficient Prediction of Laminar-Turbulent Transition in Hypersonic Flows Using Surrogate Modeling

Alexander Theiss¹, Ana Teresa Leal Faustino¹, Paul Hoffmann¹, Stefan Hein¹, Alexander Wagner¹

Abstract

The research presented in this paper represents a foundational first step toward the strategic objective of creating a powerful, multi-fidelity, and multi-modal surrogate tool that provides a holistic and computationally efficient yet physically sound solution for hypersonic transition prediction. This objective is driven by the fact that accurate prediction of laminar-turbulent transition is critical for hypersonic vehicle design. Yet established physics-based methods like Linear Stability Theory (LST) are computationally intensive and require significant expert intervention, hindering their use in automated design cycles. The development of surrogate models to overcome these limitations is particularly challenging for hypersonic boundary layers; for blunt-nosed vehicles, strong non-similar flow effects caused by the entropy layer render traditional, simplified profile parameterizations inadequate, complicating the surrogate modeling process. To address these challenges, this paper presents a methodology for creating high-fidelity surrogate models by training them exclusively on a comprehensive database from laminar simulations of a 7° blunt cone, focusing on the dominant second Mack mode instability. Two surrogate modeling frameworks are employed and compared: a Radial Basis Function (RBF) interpolation model and an eXtreme Gradient Boosting (XGBoost) machine learning framework. The performance of both models is validated against flight test data from the MF-1 experiment, demonstrating excellent agreement with the N-factor envelopes derived from direct LST. Furthermore, a quantitative assessment of computational performance reveals a key advantage of the surrogate approach, with both models providing predictions significantly faster than direct LST and the XGBoost model being the computationally most performant.

Keywords: Hypersonic flow, Boundary-layer transition, Surrogate modeling, Second Mack mode, LST

Nomenclature

Latin

$c.c.$ – Complex conjugate
 d – Maximum depth of a decision tree (XGBoost)
 d_{ip} – Wall-normal distance of the impedance inflection point
 f_k – Function representing the k -th tree in an XGBoost ensemble
 g_i – First-order gradient statistic of the loss function (XGBoost)
 H – Height above ground
 h_i – Second-order gradient statistic of the loss function (XGBoost)
 h_t – Total enthalpy
 H_{12} – Boundary-layer shape factor
 K – Total number of trees in an XGBoost ensemble

ble

L – Reference length
 M – Mach number
 m – Azimuthal mode number
 N – N-factor, logarithmic amplification factor
 N_{end} – N-factor envelope
 n – Total number of sample points in a dataset
 S – Entropy
 Obj – Objective function (XGBoost)
 $\hat{q}$ – Vector of amplitude variables
 $\tilde{q}$ – Vector of perturbation variables
 r – Euclidean distance between two points in the input space
 Re – Reynolds number
 s – Arc length

¹German Aerospace Center (DLR), Institute of Aerodynamics and Flow Technology, Department of High Speed Configurations, Göttingen, Germany. Corresponding author: alexander.theiss@dlr.de

T – Temperature	(XGBoost)
T_l – Number of leaves in an XGBoost tree	ρ – Density
t – Time	σ – Spatial amplification rate ($-\alpha_i$)
U – Streamwise velocity components	ϕ – Radial Basis Function (e.g., thin-plate spline)
w_i – Interpolation weight for the i -th sample (RBF)	Ω – Regularization function (XGBoost)
w_j – Score (weight) of the j -th leaf in a decision tree (XGBoost)	ω – Angular frequency of a disturbance
$\mathbf{x}$ – Input feature vector	(ξ, θ, ζ) – Streamwise, azimuthal, and Wall-normal coordinate
y – True (ground truth) output value	Subscripts
$\hat{y}$ – Predicted output value from a surrogate model	0 – Neutral stability point or initial value
(X, Y, Z) – Cartesian coordinates	$crit$ – Critical value for transition
Greek	e – Boundary-layer edge condition
α – Complex streamwise wavenumber ($\alpha_r + i\alpha_i$)	ip – Value at the impedance inflection point
γ – Regularization parameter for the number of leaves (XGBoost)	r – Real part of a complex number
δ_1 – Displacement thickness	i – Imaginary part of a complex number, Indices for samples or data points
δ_{99} – 99% velocity-thickness of the boundary layer	j – Indices for samples or data points
λ – Regularization parameter for leaf weights	k – Index for a specific tree (iteration) in an XGBoost ensemble
	w – Wall condition
	∞ – Freestream condition

1. Introduction

The transition of the boundary layer from a laminar to a turbulent state is a critical phenomenon in hypersonic flight, profoundly impacting surface heat transfer, skin friction drag, and overall vehicle performance. An accurate prediction of the transition location is therefore essential for the design of effective thermal protection systems and for optimizing aerodynamic efficiency [1]. Among the various instability mechanisms present in hypersonic flows, the second Mack mode is often the dominant driver of transition for slender vehicles at zero or small angles of attack [2, 3].

The e^N method, which couples Linear Stability Theory (LST) with semi-empirical criteria, is a well-established physics-based approach for transition prediction [4, 5]. However, its direct application within industrial design cycles is often hindered by significant challenges. LST computations are computationally expensive, require highly resolved boundary-layer profiles, and demand considerable user expertise to ensure robust and accurate results, making automation difficult [6]. To overcome these limitations, surrogate-based e^N methods have emerged as a powerful and efficient alternative. These models, trained on pre-computed databases of instability characteristics, can predict disturbance amplification rates with a fraction of the computational cost of direct LST [4].

At the German Aerospace Center (DLR), a robust framework using surrogate models has been successfully developed and applied for predicting Tollmien-Schlichting (TS) and stationary crossflow (CF) instabilities on transonic swept-wing aircraft [7, 8, 9]. This paper represents a foundational step in extending DLR's surrogate modeling expertise to the hypersonic flight regime.

Acknowledging the complexity of hypersonic transition—which can involve multiple competing mechanisms such as first and second modes, crossflow, and Görtler instabilities—this initial work deliberately focuses on the second Mack mode. Specifically, we address two-dimensional planar waves, as they are often the most amplified and dominant mechanism for the canonical geometries under consideration [1, 2].

Recent advancements in surrogate modeling for hypersonic flows provide a strong foundation for this work. Radial Basis Function (RBF) models, for instance, have been successfully applied by Nie et al. [4] for creating an e^N method for compressible boundary layers. Their work systematically showed that while simple self-similar parameters were insufficient for blunt cones, using a combination of four boundary-layer edge parameters and a shape factor (H_{12}) could achieve accuracy comparable to using

16 discrete profile points. More advanced machine learning models have also shown immense promise. Zafar et al. [6] and Paredes et al. [5] used Convolutional Neural Networks (CNNs), which can automatically extract physically relevant features from raw boundary-layer profiles. Their approach demonstrated not only excellent agreement with LST but also superior robustness when using under-resolved CFD input data, a crucial advantage for practical applications.

Building on these insights, this paper details the development of surrogate models based on two distinct data-driven approaches. First, we employ the established Radial Basis Function (RBF) method, following the successful implementation by Nie et al. [4]. Second, recognizing that stability databases for hypersonic flows can become exceptionally large, we explore XGBoost (Extreme Gradient Boosting) as a powerful alternative. The selection of XGBoost is motivated by its increasing and successful application in the aerodynamics community. Its effectiveness has been demonstrated broadly for predicting aerodynamic performance metrics, such as lift and drag coefficients, for supersonic aircraft [10]. More specifically within the hypersonic domain, XGBoost-based models have been developed to predict surface heat flux, successfully incorporating complex geometric features to handle nonlinear thermal loads [11]. Most pertinently to the present study, the algorithm has been applied to fine-grained boundary-layer transition prediction from wall pressure data, where it was shown to outperform other ensemble methods in capturing the nonlinear characteristics of transition onset [12].

The aim of this paper is therefore to develop and assess a robust methodology for creating surrogate models capable of predicting second-mode instability in complex, non-similar hypersonic flows. This task is particularly challenging for blunt-nosed vehicles, where the flow similarity assumed by simpler methods breaks down. A strong entropy layer generated at the nose significantly alters the boundary layer downstream, rendering stability predictions based on simplified, self-similar profiles unreliable. As demonstrated by Paredes et al. [5], such approaches can lead to errors in the predicted transition location of up to 40% when compared to full Navier-Stokes solutions. To this end, this work focuses exclusively on a database generated from high-fidelity CFD simulations of a 7° blunt cone. The performance of the resulting RBF and XGBoost models is then evaluated against direct LST calculations and validated against real-world flight conditions from the MF-1 experiment [13, 14].

The remainder of this paper is structured as follows: Section 2 details the theoretical foundations, covering Linear Stability Theory and the architectures of the Radial Basis Function and XGBoost surrogate models. Section 3 describes the CFD-based database generation, including the cone model and simulation setup. The development of the surrogate models is then detailed in Section 4, which covers the input parameter selection and the specific training configurations for both the RBF and XGBoost approaches. In Section 5, the predictive capabilities of these models are demonstrated through a verification against LST data and a validation against the MF-1 flight test cases. Finally, Section 6 provides a conclusion and an outlook on future work.

2. Theory

This section outlines the theoretical and numerical methods employed in this study. It details the fundamentals of Linear Stability Theory (LST), the method used to generate the ground-truth instability data, and the feature extraction process based on the boundary-layer impedance inflection point. Subsequently, the architectures of the two surrogate modeling techniques used to approximate the LST results, namely Radial Basis Function (RBF) interpolation and XGBoost, are presented.

2.1. Local linear stability theory (LST) and e^N method

The instability characteristics of the generated basic flows are computed using Linear Stability Theory (LST) to provide the ground-truth data for training the surrogate models. LST analyzes the behavior of infinitesimal disturbances superimposed on a steady laminar base flow. The disturbances are assumed to be wave-like normal modes of the form:

$$\tilde{q}(\xi, \theta, \zeta, t) = \hat{q}(\zeta) \exp[i(\alpha\xi + m\theta - \omega t)] + c.c., \quad (1)$$

where $\hat{q}(\zeta)$ is the disturbance amplitude function in the wall-normal direction ζ , and (α, m) are the wavenumber components in the streamwise (ξ) and azimuthal (θ) directions, respectively. In the spatial

stability framework used here, the frequency ω is real, while the streamwise wavenumber $\alpha = \alpha_r + i\alpha_i$ is complex. The spatial amplification rate is given by the negative of its imaginary part, $\sigma = -\alpha_i$.

The numerical solution of the stability equations is performed using a DLR in-house instability code, as detailed in Theiss [15]. The spatial discretization in the wall-normal direction is achieved using a high-order finite-difference scheme of order eight (FD-q8) [16]. Based on extensive experience, 201 grid points are used in the wall-normal direction, which provides more than sufficient resolution for the instability modes under investigation. Isothermal, no-slip boundary conditions are imposed at the wall, while Dirichlet boundary conditions are applied in the far-field to ensure that all disturbance amplitudes decay to zero.

The resulting amplification rates are integrated along the streamwise direction to obtain the N-factor, which serves as the metric for transition prediction:

$$N = \int_{\xi_0}^{\xi} \sigma(\xi'; \omega, m) d\xi', \quad (2)$$

where ξ_0 is the neutral point where disturbances begin to amplify ($\sigma > 0$). To establish the link to transition prediction via the e^N method, the N-factor envelope can be constructed by taking the maximum N-factor over all relevant frequencies and wavenumbers at each streamwise location:

$$N_{\text{env}}(\xi) = \max_{\omega, m} [N(\xi; \omega, m)]. \quad (3)$$

Transition is then predicted to occur at the location where this envelope reaches a prescribed threshold value, N_{crit} , which can be determined experimentally.

2.2. Feature Extraction and the Boundary-Layer Edge Challenge

A critical aspect of building an effective surrogate model is the robust extraction of physically relevant features from the basic flow. In non-similar hypersonic flows, particularly over blunt bodies, the presence of a strong entropy layer complicates the unambiguous detection of the boundary-layer edge, rendering traditional metrics like the 99% velocity thickness (δ_{99}) unreliable. This issue is a known challenge for both direct stability analysis and surrogate modeling.

Previous researchers have addressed this challenge using various criteria based on the total enthalpy profile. For instance, Paredes et al. [5] defined the edge based on the total enthalpy ratio reaching 99.5% of the freestream value ($h_t/h_{t,\infty} = 0.995$), while Nie et al. [14] used a definition based on the average of the maximum and freestream total enthalpy. While effective, such definitions can be sensitive to the profile shape and numerical resolution, potentially introducing uncertainty into the extracted features.

To circumvent the potential ambiguities and sensitivities associated with edge detection, this work adopts a different approach for feature extraction. Instead of relying on a traditional edge definition, we extract reference values at the boundary-layer impedance inflection point. This location is chosen based on the thermoacoustic interpretation of the second Mack mode instability proposed by Kuehl [17]. According to this theory, the second mode behaves as a resonant standing wave trapped within an acoustic impedance well, which is bounded by the wall and a secondary impedance peak near the sonic line where the flow velocity equals the disturbance propagation speed. The inflection point of this impedance profile is a unique, physically significant feature that is intrinsically linked to the instability mechanism itself.

By using this inflection point as the reference for extracting local flow quantities, we anchor our feature set to a location with direct physical relevance to the second mode. This provides a more robust and physically grounded foundation for the surrogate models, reducing the sensitivity to the far-field portion of the boundary-layer profile.

2.3. Radial Basis Function (RBF) interpolation

The RBF-based surrogate model follows the methodology successfully applied for compressible boundary layers by Nie et al. [14] and for transonic flows by Theiss and Hein [7]. RBF is a powerful technique

for multidimensional data interpolation. For a given set of input parameters $\mathbf{x}_c$, the surrogate model predicts an output $\hat{y}$ (e.g., σ) as a weighted sum of basis functions:

$$\hat{y}(\mathbf{x}_c) = \sum_{i=1}^n w_i \phi(\|\mathbf{x}_c - \mathbf{x}_i\|_2), \quad (4)$$

where n is the number of sample points in the training database, $\mathbf{x}_i$ are the stored input feature vectors, w_i are the interpolation weights, and ϕ is the radial basis function. The interpolation weights are precomputed for the entire dataset by solving the linear system:

$$y_j = \sum_{i=1}^n w_i \phi(\|\mathbf{x}_i - \mathbf{x}_j\|_2). \quad (5)$$

To ensure numerical stability and prevent the basis function matrix from becoming ill-conditioned, the sample points are normalized by their maximum absolute value prior to this precomputation step. Consistent with prior work, the thin-plate spline function, $\phi(r) = r^2 \log r$, is used, where r is the Euclidean distance. To manage the large database size and reduce computational effort, a sample partitioning method is employed, which divides the full dataset into several smaller, localized subsamples. An independent RBF model is then constructed for each partition [9].

2.4. eXtreme Gradient Boosting (XGBoost)

XGBoost (eXtreme Gradient Boosting) is a highly efficient and scalable machine learning algorithm based on an ensemble of decision trees, specifically Classification and Regression Trees (CART) [18]. It operates on the principle of gradient boosting, where models are constructed in a stage-wise, additive manner. For a dataset with n samples, the prediction $\hat{y}_i$ for sample i after k iterations is the sum of the predictions from all k trees:

$$\hat{y}_i^{(k)} = \sum_{j=1}^k f_j(x_i) = \hat{y}_i^{(k-1)} + f_k(x_i) \quad (6)$$

where f_k is the new decision tree added at the k -th iteration. The core of the algorithm is to train this new tree f_k to correct the residual errors of the previous ensemble of $k - 1$ trees.

To determine the optimal structure of the new tree, XGBoost minimizes a regularized objective function. This function is approximated using a second-order Taylor expansion of the loss function, which allows for rapid optimization:

$$\text{Obj}^{(k)} \approx \sum_{i=1}^n \left[g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \Omega(f_k), \quad (7)$$

where g_i and h_i are the first and second-order gradient statistics of the loss function. The term $\Omega(f_k)$ is a regularization function that penalizes the complexity of the tree:

$$\Omega(f_k) = \gamma T_l + \frac{1}{2} \lambda \sum_{j=1}^{T_l} w_j^2, \quad (8)$$

where T_l is the number of leaves in the tree, w_j is the score of the j -th leaf, and γ and λ are regularization parameters. This regularization is crucial for preventing overfitting and improving the model's ability to generalize to new data.

In the context of the present work, the XGBoost model is trained to learn the mapping from the vector of local boundary-layer parameters to the instability growth rate σ . The decision-tree-based nature of the algorithm is particularly well-suited for this task, as it can effectively capture the complex, non-linear interactions between the input features without requiring a predefined functional form.

3. Database Generation

The foundation of any robust surrogate model is a comprehensive and high-fidelity training database. For blunt hypersonic vehicles, the presence of a strong entropy layer creates a highly non-similar boundary layer where simplified, self-similar models are known to be inadequate [5]. Therefore, to accurately capture the relevant flow physics for second-mode instability, this study focuses exclusively on a database generated from high-fidelity CFD simulations of a canonical blunt cone.

3.1. CFD Simulation Setup

All laminar basic flow computations were performed using the DLR flow solver TAU. A second-order upwind scheme was employed on a shock-adapted, structured grid to ensure accurate capturing of the bow shock and the ensuing flow field, as illustrated in Fig. 1.

The computational grid consists of 621 points in the streamwise direction and 335 points in the wall-normal direction. To ensure adequate resolution of the viscous region, a minimum of 120 points are clustered within the boundary layer near the wall. The spatial convergence of the solution was rigorously verified through a grid refinement study. The final grid was selected not merely based on the convergence of the mean flow, but on the convergence of the subsequent second-mode instability results from LST, ensuring that the mean flow is sufficiently resolved for the stability analysis itself.

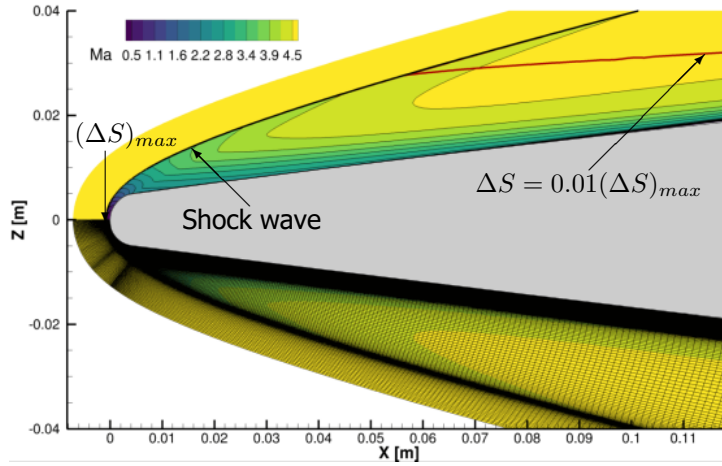


Fig 1. Computational grid and Mach number contours from a representative CFD simulation of the 7° blunt cone ($M_\infty = 4.93$, $T_w = 600$ K, $Re_\infty/L = 1.0 \times 10^7 \text{ m}^{-1}$). The red line indicates the entropy layer edge, defined by the criterion from Nie et al. [14].

3.2. Parametric Space and Data Extraction

The parametric space for the CFD simulations is identical to that used in the work of Nie et al. [14], covering 280 unique basic flow conditions over a 7° cone of 1 m length with a 5 mm nose radius at an angle of attack of 0°. The simulations span a comprehensive range of flight conditions relevant to hypersonic transition, where Mach numbers (M_∞) vary from 4.3 to 5.5 (in increments of 0.2), unit Reynolds numbers (Re_∞/L) range from $1.4 \times 10^7 \text{ m}^{-1}$ to $3.5 \times 10^7 \text{ m}^{-1}$ (in increments of $0.3 \times 10^7 \text{ m}^{-1}$), and wall temperatures (T_w) cover 500 K to 600 K (in increments of 25 K). This set of CFD-generated basic flows forms the core of our training database for capturing non-similar, blunt-body effects. Following the convergence of each CFD simulation, the DLR in-house tool BLEXIT was utilized to extract the boundary-layer profiles and create the input files for the subsequent instability analysis.

3.3. Validation Case

To validate the performance of the trained surrogate models, this study uses flight test data from the MF-1 blunt cone experiment performed by the China Aerodynamic Research and Development Center in 2015 [19]. The original experimental data and flight conditions are detailed by Tu et al. [13]. This experiment serves as an ideal validation case, as the vehicle geometry (7° half-angle cone with a 5 mm nose radius) is identical to the one used for generating the CFD database. The specific freestream conditions

for the six flight instants of the ascent phase analyzed in this paper are detailed in Table 1.

Table 1. Freestream conditions for the MF-1 flight experiment ascent phase. Data sourced from the original flight test study by Tu et al. [13].

t [s]	H [km]	M_∞	$Re_\infty/L[\text{m}^{-1}]$	T_w [K]	T_∞ [K]
17	11.20	4.54	3.333×10^7	518	216.65
18	12.24	4.87	3.036×10^7	531	216.65
19	13.35	5.23	2.739×10^7	575	216.65
20	14.50	5.25	2.295×10^7	595	216.65
21	15.63	5.08	1.860×10^7	595	216.65
22	16.71	4.93	1.524×10^7	600	216.65

4. Surrogate Modeling

This section details the development and evaluation of the surrogate models for predicting second-mode instability. The discussion covers the selection of input parameters, and the specific configurations of the Radial Basis Function (RBF) and XGBoost models.

4.1. Input Parameter Selection

A crucial step in developing any surrogate model is the selection of input features that effectively capture the underlying physics. As established in Section 2.2, the challenges associated with defining a consistent boundary-layer edge in non-similar hypersonic flows motivate an approach anchored to more robust physical features. Based on the thermoacoustic interpretation of the second mode [17], we use the boundary-layer impedance inflection point as a unique and physically relevant location for feature extraction.

The surrogate models utilize a set of six non-dimensional and dimensional parameters as inputs: the Mach number at the impedance inflection point (Ma_{ip}); a local Reynolds number based on the streamwise velocity and wall-normal distance of the inflection point (Re_{ip}); the non-dimensional circular frequency (ω_{ip}); the ratio of the impedance inflection point's wall-normal distance to the local displacement thickness (d_{ip}/δ_1); the dimensional density at the inflection point (ρ_{ip}); and the ratio of wall temperature to the temperature at the impedance inflection point (T_w/T_{ip}). While a thorough feature evaluation is a subject for future work, this initial study found that this physically-grounded set of parameters provides a strong basis for accurate predictions, as will be demonstrated in the results.

4.2. Surrogate Model Configurations and Training

To compare the performance of different data-driven approaches, two distinct surrogate models were developed using the CFD-based database. Both models are constructed to predict the planar ($m = 0$) second-mode spatial amplification rate scaled with the wall-normal height of the impedance inflection point, σ/d_{ip} . The instability data for the training sets were initially computed on a fine grid (with 300 samples) for the non-dimensional circular frequency, ω_{ip} , ranging from 1.8 to 3.1, which comprehensively covers the amplified frequency spectrum for all cases.

4.2.1. RBF Model Configuration

For the RBF model, specific data preparation steps were necessary to manage the computational load associated with its construction. The creation of an RBF model involves inverting a dense $n \times n$ kernel matrix, leading to a time complexity of $\mathcal{O}(n^3)$ and a memory requirement of $\mathcal{O}(n^2)$ (for both constructing and more important evaluating the model), which becomes prohibitive for very large datasets.

To make this approach computationally feasible, the dataset was first downsampled in two ways: the frequency range was discretized into 65 segments, and data from only every third streamwise station was used. Subsequently, the model employs the sample partitioning method previously utilized by Theiss and Hein [7]. The downsampled dataset is divided into 25 subsets based on the values of

ω_{ip} and Re_{ip} . The boundaries for these partitions are strategically chosen to ensure that each subset contains a roughly equal and computationally manageable number of points, approximately 20000 per partition. An independent RBF model is then constructed for each of these 25 partitions by precomputing its unique set of interpolation coefficients.

4.2.2. XGBoost Model Configuration

In contrast, leveraging the superior scalability of the XGBoost algorithm (with training complexity of approximately $\mathcal{O}(K \cdot d \cdot n \log n)$), where K is the number of trees and d the maximum depth of the trees. The XGBoost model was trained using data from all available streamwise stations. This allows for a more data-rich training environment. To optimize its predictive performance, a hyperparameter tuning process was conducted using the 'GridSearchCV' utility from scikit-learn [20] with threefold cross-validation, optimizing for the R^2 score, a statistical measure of the goodness of fit, where a value of 1.0 represents a perfect prediction of the data's variance.

A crucial aspect of the training was the application of sample weighting to focus the model's learning on the physically most important regions of instability. A sigmoid function was used to assign significantly higher weights to samples with positive growth rates ($\sigma > 0$), effectively forcing the model to prioritize accuracy in the amplified region over the damped region. This ensures that the model is most accurate where it matters most for transition prediction.

While an exhaustive optimization of all possible hyperparameters was beyond the scope of this initial study, the grid search identified a robust set of parameters for the present work, including `n_estimators` of 1500, a `max_depth` of 15, a `learning_rate` of 0.15, and both `subsample` and `colsample_bytree` ratios of 0.8.

It is important to note that the performance of the XGBoost model is dependent on the chosen hyperparameters; however, this set was found to provide excellent performance for the present work. The following section will compare the predictive capabilities of these two distinct model configurations.

5. Results and Discussion

This section presents a comprehensive evaluation of the surrogate models developed from the high-fidelity CFD database. The objective is threefold: first, to verify that the trained RBF and XGBoost models can accurately reproduce the ground-truth LST data; second, to validate their predictive performance against realistic boundary layers from the MF-1 hypersonic flight experiment; and third, to quantify the computational efficiency gained by the surrogate approach.

5.1. Model Verification against LST Data

Before application to flight data, the models are first verified against a held-out portion of the CFD database. This step confirms their ability to accurately predict instability growth rates for a known case not included in their training.

Figure 2 provides a direct visual comparison of the predicted spatial amplification rate, σ , against the ground-truth LST data. A striking difference in the qualitative nature of the predictions is immediately apparent. The RBF model (a) produces smooth, continuous contour lines, which is inherent to its design as an interpolation method that constructs a single, continuous function across the parameter space. In contrast, the jagged contours of the XGBoost model (b) are a direct result of its tree-based architecture, which partitions the input space into numerous discrete regions, leading to a piecewise constant prediction.

Despite these visual differences, the strong quantitative agreement of both models with the underlying LST instability results validates that the chosen set of input features effectively parameterizes the complex physics of the problem.

An accurate reproduction of the entire instability landscape is the foundation for reliable transition prediction. The N-factor, which is the ultimate metric for the e^N method, is the cumulative result of integrating these local growth rates along streamwise paths (i.e., for constant frequencies). Figure 3 compares the N-factor curves for a range of disturbance frequencies for the same test case. The RBF

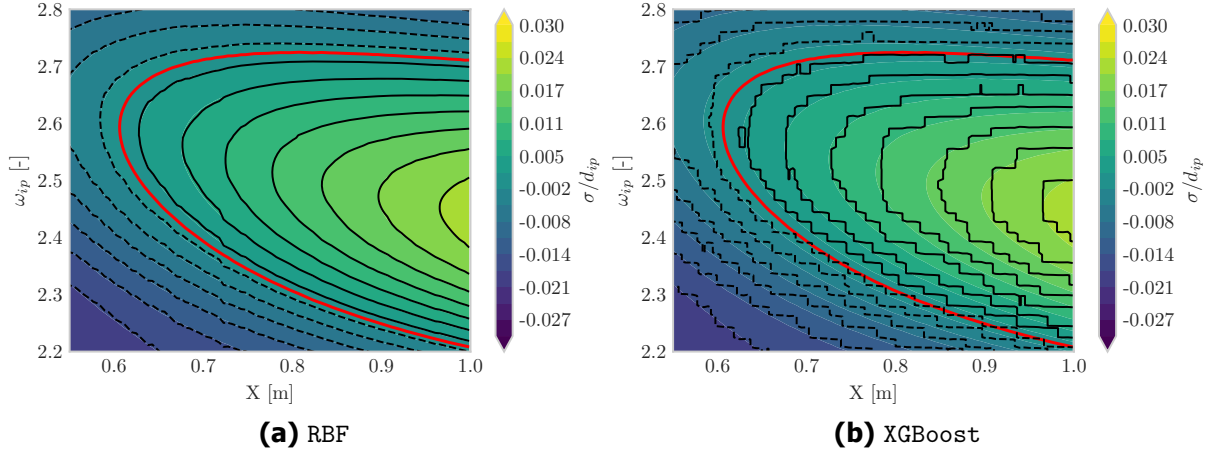


Fig 2. Comparison of predicted spatial nondimensional amplification rate (σ) contours from the surrogate models against the LST ground truth (filled contour) for $M_\infty = 4.55$, $T_w = 555$ K and $Re_\infty = 3.2 \times 10^7$ /m. The red line indicates the neutral curve.

model's predictions are exceptionally precise, with the predicted curves being nearly indistinguishable from the LST ground truth. The XGBoost model also shows excellent agreement, albeit with a minor but consistent deviation visible for the most amplified frequencies. Crucially, these minor local deviations do not accumulate into significant errors. This successful verification provides strong confidence in the ability of both approaches to accurately represent the underlying physics.

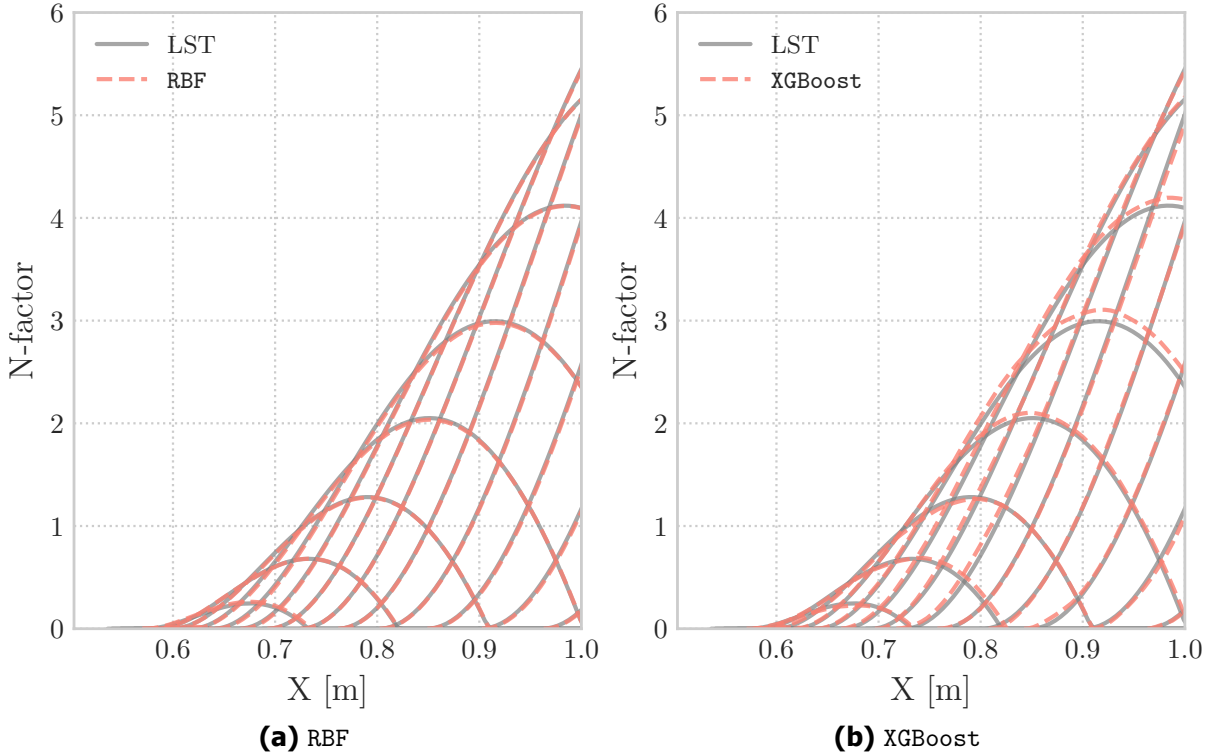


Fig 3. Comparison of predicted N-factors from the surrogate models against the LST ground truth at selected frequencies for $M_\infty = 4.55$, $T_w = 555$ K and $Re_\infty/L = 3.2 \times 10^7$ /m.

5.2. Validation against the MF-1 Flight Experiment

Having verified the models' accuracy, we now assess their practical utility by validating them against flight test data from the MF-1 experiment [13]. This step tests the models' performance on the complex, non-similar boundary layers encountered in a real-world flight scenario. The analysis begins with the local spatial growth rates, which form the foundation for the integrated N-factor envelopes used in transition prediction.

Figure 4 presents a comparison of the predicted spatial growth rates at a fixed axial location ($X = 0.736882\text{ m}$) for the six flight instants. A key observation is that both the RBF and XGBoost models accurately capture the primary characteristics of the second-mode instability, successfully predicting both the amplified frequency range and the magnitude of the maximum amplification rate across all conditions shown. This high local accuracy is crucial, as it provides the basis for the integrated N-factor envelopes presented in Fig. 5.

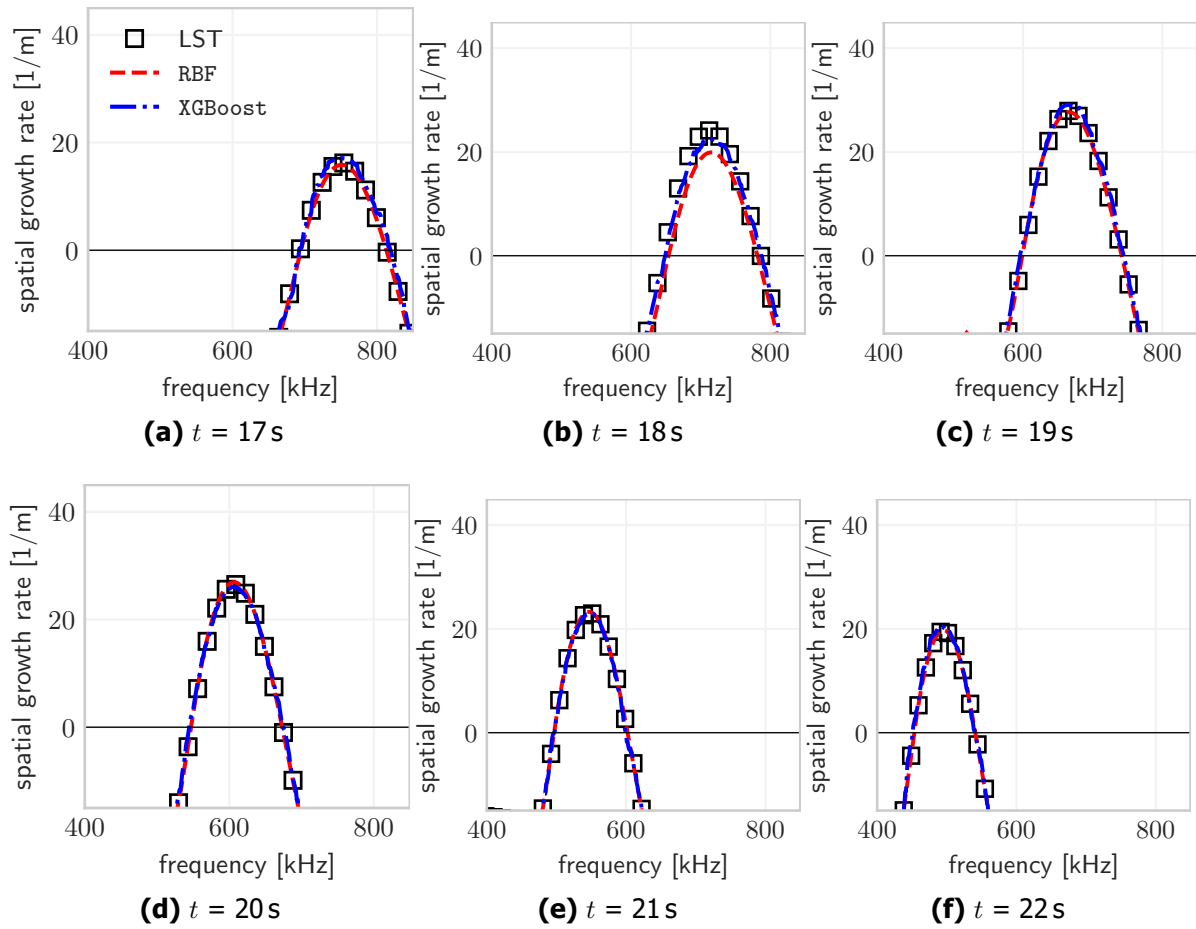


Fig 4. Comparison of spatial growth rates predicted by the RBF and XGBoost surrogate models against the LST benchmark at a fixed axial location ($X = 0.736882\text{ m}$) for the six flight instants of the MF-1 experiment.

The integrated envelopes confirm the strong predictive performance of both surrogate models, with the predicted curves closely tracking the LST benchmark across nearly all flight conditions. This validates their applicability and accuracy for the specific 7° blunt cone configuration. However, a minor but consistent deviation is visible for the RBF model at the $t = 18\text{ s}$ flight condition, both in the local growth rate (Fig. 4b) and the resulting N-factor envelope (Fig. 5b). Since this flight condition lies well within the interior of the training database's parameter space, this small discrepancy is likely an artifact of the

sample partitioning method. As the flow parameters may lie near a boundary between two of the localized RBF interpolants, the prediction can be slightly less constrained. In contrast, the XGBoost model, which was trained globally on the full spatial data without partitioning, remains highly accurate across all conditions, suggesting its more holistic training approach provides slightly more robust generalization capabilities.

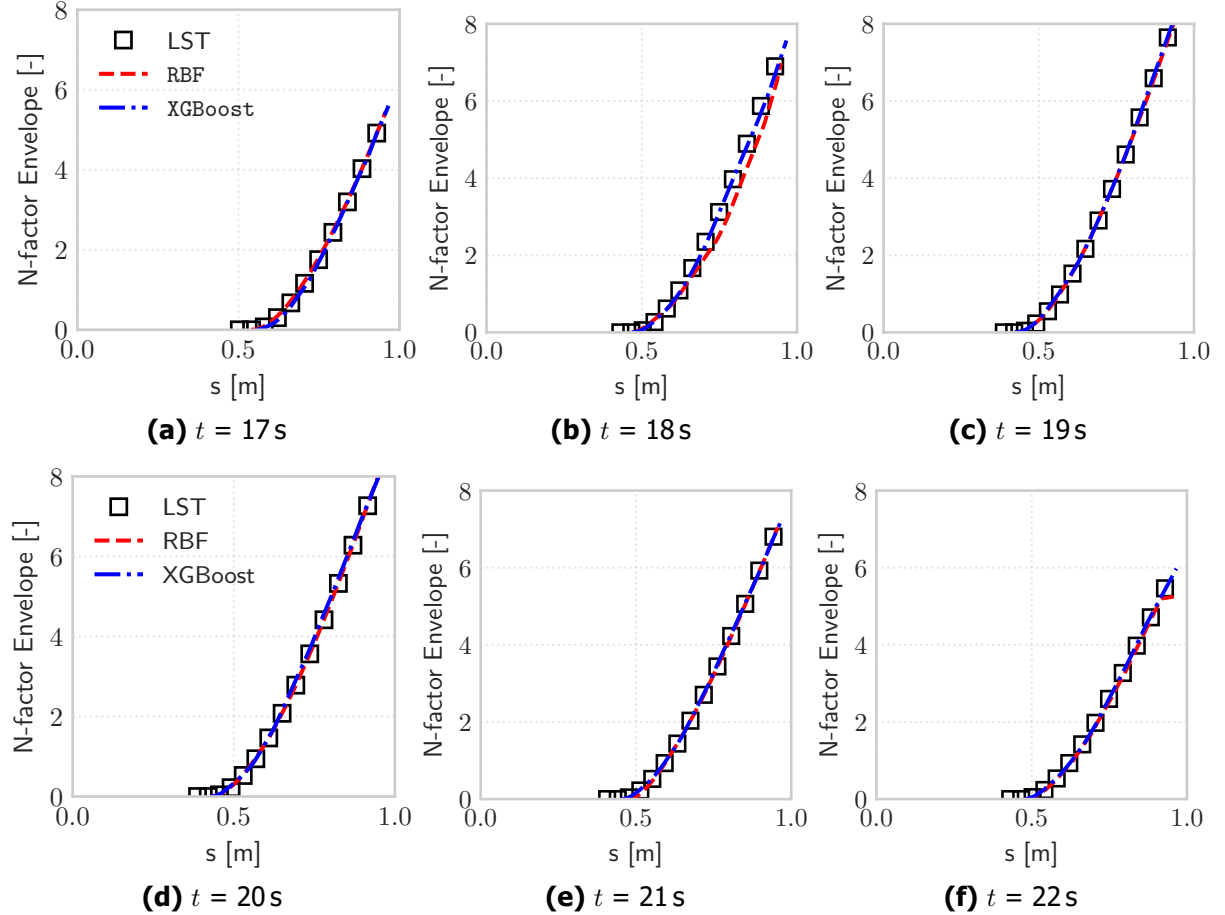


Fig 5. Comparison of N-factor envelopes predicted by the RBF and XGBoost surrogate models against the LST benchmark for six different flight instants of the MF-1 experiment.

5.3. Computational Performance Analysis

The primary motivation for developing surrogate models is to accelerate the transition prediction workflow and make it more robust. To quantify this efficiency gain, the wall-clock time required to generate a full N-factor envelope for a single flight condition was measured. This task involves computing 50 individual N-factor curves across the relevant frequency spectrum.

To ensure a fair and direct comparison of the algorithms' intrinsic speeds, all computations—for LST, RBF, and XGBoost —were performed on a single CPU core. It should be noted that all methods are capable of exploiting multi-core architectures for further acceleration (e.g., by computing N-factor curves in parallel), but the single-core benchmark provides a clear baseline for performance.

The results of this timing analysis are presented in Table 2. As shown, both surrogate models offer a dramatic reduction in computational time compared to the direct LST approach. However, a significant performance difference between the two surrogate models is also evident. The XGBoost model is exceptionally fast, providing a prediction in under a second, whereas the RBF model requires an order of magnitude longer evaluation time.

This difference is a direct consequence of the RBF model's computational complexity, which scales with the number of data points per partition. For the hypersonic flow problem, it was found that a large number of points per partition (approximately 20000) was necessary to maintain high predictive accuracy. While initial trials with a greater number of partitions were conducted to reduce the evaluation time, this consistently resulted in a degradation of the model's predictive performance. Therefore, the current RBF configuration represents a necessary trade-off between its computational cost and predictive accuracy.

This remarkable speed-up, especially for the XGBoost model, makes the surrogate-based approach ideal for applications requiring rapid design iterations, uncertainty quantification, or integration into multi-disciplinary optimization frameworks.

Table 2. Comparison of computational costs for generating a full N-factor envelope (50 frequencies) for the $t = 22$ s flight case on a single CPU core (Intel(R) Xeon(R) W-2155).

Method	Prediction Time (Single Core)
LST Analysis	~240 seconds
RBF Model	~9.5 seconds
XGBoost Model	~0.4 seconds

6. Conclusion and Outlook

6.1. Conclusion

This paper has successfully demonstrated a robust methodology for developing high-fidelity surrogate models for the prediction of second-mode instability on a blunt hypersonic vehicle. By training both Radial Basis Function (RBF) and XGBoost models exclusively on a database of CFD simulations for a 7° blunt cone, the complex, non-similar flow physics of the boundary layer were effectively captured.

The key findings of this work are:

1. Both the RBF and XGBoost surrogate models accurately reproduced the ground-truth N-factors from LST. This success validates that the chosen input features, which are based on the physically-relevant impedance inflection point, effectively capture the underlying instability physics.
2. The models showed strong predictive performance when validated against flight test data from the MF-1 experiment, confirming their applicability and accuracy for the specific 7° blunt cone configuration under real-world flight conditions.
3. A significant computational advantage was quantified, with both surrogate models offering predictions several orders of magnitude faster than traditional LST. The XGBoost model, in particular, demonstrated the highest computational efficiency. The RBF model, while also substantially faster than LST, was found to be computationally limited by the large number of data points required per partition to maintain accuracy, highlighting the superior scalability of tree-based methods for this class of problem. It is crucial to note, however, that this speed-up pertains to the inference stage; the generation of the initial training database remains reliant on a comprehensive set of LST computations.

This work establishes and validates a robust methodology at DLR for creating accurate, configuration-specific surrogate models for hypersonic transition, demonstrating a viable approach for complex flows where simplified, self-similar methods are inadequate.

6.2. Outlook

The research presented in this paper represents a foundational first step toward a larger, strategic objective. The ultimate goal is to establish a comprehensive and versatile database that accurately represents the boundary layers on typical hypersonic flight vehicles and in relevant ground test facilities. Such a database must capture both the non-similar flow physics near the leading edge of blunt bodies

and the potentially more self-similar flow characteristics found further downstream or within wind tunnel nozzles.

To achieve this long-term vision, future work will focus on expanding the methodology demonstrated here:

- **Expanding the CFD Database:** The current CFD database will be broadened to include various nose radii and cone half-angles. This will enhance the surrogate models' ability to generalize across a wider range of blunt-body geometries.
- **Developing a Self-Similar Database:** The ongoing development of a database based on compressible Falkner-Skan-Cooke (FSC) local similarity solutions [21] will be completed to represent boundary layers in regions of local similarity.
- **Creating a Hybrid Surrogate Model:** With both database components in place, future research will focus on pioneering a hybrid surrogate model. The challenge will be to build a single, unified model (likely using a scalable framework like XGBoost or a deep neural network) that can seamlessly leverage high-fidelity, non-similar CFD data alongside data from more general, locally self-similar flow solutions. Such a multi-fidelity model would be capable of correctly applying the appropriate physics based on the local flow conditions.
- **Extension to Three-Dimensional Flows:** The present study was limited to two-dimensional axisymmetric flows. A critical future step is to extend this methodology to three-dimensional (3D) boundary layers to account for effects such as angle of attack. This will require expanding the input features of the surrogate models to include crossflow properties and extending the database to cover relevant 3D flow phenomena.
- **Extending the Framework to Other Instability Mechanisms:** The present study was deliberately focused on planar second Mack modes. Building on the 3D framework, the surrogate modeling methodology will be extended to other relevant transition mechanisms in hypersonic flows, such as first-mode, crossflow, and Görtler instabilities.

By pursuing these parallel efforts, we aim to create a powerful, multi-fidelity, and multi-modal surrogate tool that provides a holistic and computationally efficient yet physically sound solution for hypersonic transition prediction across entire vehicle configurations.

References

- [1] Lee, S., Duan, M., Tu, G., Wan, B., Chen, J., Yuan, X.: Investigation on transition correlation of conical boundary layer in hypersonic wind tunnel experiments. *Acta Astronaut.* 201, 329–339 (2022)
- [2] Borg, M.P., Adamczak, D.W., Tufts, M.W.: HIFLIER Hypersonic Flight Experiment: Ascent Results. In: *AIAA SCITECH 2025 Forum*. AIAA Paper 2025-0734 (2025)
- [3] Barraza, B., Gross, A., Haas, A.P., Hader, C., Fasel, H.F.: Machine-Learning-Based Amplification Factor Transport Equation for Hypersonic Boundary-Layers. In: *AIAA AVIATION 2022 Forum*. AIAA Paper 2022-3681 (2022)
- [4] Nie, H., Song, W., Han, Z., Chen, J., Tu, G.: A surrogate-based eN method for compressible boundary-layer transition prediction. *J. Aircr.* 59(1), 89–102 (2022)
- [5] Paredes, P., Venkatachari, B., Choudhari, M.M., Li, F., Chang, C.L., Zafar, M.I., Xiao, H.: Toward a practical method for hypersonic transition prediction based on stability correlations. *AIAA J.* 58(10), 4475–4484 (2020)
- [6] Zafar, M.I., Xiao, H., Choudhari, M.M., Li, F., Chang, C.L., Paredes, P., Venkatachari, B.: Convolutional neural network for transition modeling based on linear stability theory. *Phys. Rev. Fluids* 5, 113903 (2020)

- [7] Theiss, A., Hein, S.: A Surrogate-Based eN Transition Prediction Method for Three-Dimensional Compressible Boundary Layers. In: Dillmann, A., et al. (eds.) *New Results in Numerical and Experimental Fluid Mechanics XIV, STAB/DGLR Symposium 2022*, pp. 475–485. Springer, Cham (2024)
- [8] Hoffmann, P., Theiss, A., Hein, S.: Correlating the Internal Encoding of Boundary-Layer Profiles: Insights in Neural Networks Used for Boundary-Layer Stability Prediction. In: *AIAA SCITECH 2024 Forum*. AIAA Paper 2024-2684 (2024)
- [9] Hoffmann, P., Theiss, A., Hein, S.: Neural Networks as a Surrogate Model for Linear Stability Analysis of Three-Dimensional Compressible Boundary Layers. In: *AIAA SCITECH 2024 Forum*. AIAA Paper 2024-2684 (2024)
- [10] Liu, Q., Zhang, J., Zheng, Y., Liu, Y.: Supersonic aircraft aerodynamic performance prediction based on machine learning approach. *AIP Adv.* 15, 085111 (2025). <https://doi.org/10.1063/5.0282257>
- [11] Wang, C., Song, Y., Yan, T., Xiao, S., Zhu, L.: Data-driven model for predicting surface heat flux of hypersonic vehicles based on eXtreme Gradient Boosting. *Phys. Fluids* 37, 085173 (2025). <https://doi.org/10.1063/5.0282563>
- [12] Chang, W., Hu, H., Xi, Y., Kloker, M., Teng, H., Ren, J.: Boundary-layer transition in the age of data: from a comprehensive dataset to fine-grained prediction. *Phys. Fluids* (2025, submitted)
- [13] Tu, G., Wan, B., Chen, J., Yuan, X., Chen, J.: Investigation on correlation between wind tunnel and flight for boundary layer stability and transition of MF-1 blunt cone. *Sci. Sin.-Phys. Mech. Astron.* 49, 124701 (2019). <https://doi.org/10.1360/SSPMA-2019-0162>
- [14] Nie, H., Song, W., Han, Z., Tu, G., Chen, J.: Application of surrogate models to stability analysis and transition prediction in hypersonic flows. *Adv. Aerodyn.* 4, 33 (2022). <https://doi.org/10.1186/s42774-022-00120-2>
- [15] Theiss, A.: *Transition Mechanisms on Blunt Re-Entry Capsules With and Without Roughness*. Ph.D. thesis, Technische Universität Braunschweig, Germany (2021)
- [16] Hermanns, M., Hernández, J.A.: Stable high-order finite-difference methods based on non-uniform grid point distributions. *Int. J. Numer. Methods Fluids*. 56(3), 233–255 (2007)
- [17] Kuehl, J.J.: Thermoacoustic interpretation of second-mode instability. *AIAA J.* 56(9), 3585–3592 (2018). <https://doi.org/10.2514/1.J057015>
- [18] Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. ACM, New York (2016)
- [19] Xiao, H., Ou, C., Ji, H., He, Z., Liu, N., Yuan, X.: Low-Cost and Aerodynamics-Aim Hypersonic Flight Experiment MF-1. In: *MATEC Web of Conferences*, vol. 316, p. 04006. EDP Sciences (2020). <https://doi.org/10.1051/mateconf/202031604006>
- [20] 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., Duchesnay, E.: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011)
- [21] Liu, Z.: Compressible Falkner-Skan-Cooke boundary layer on a flat plate. *Phys. Fluids* 33(12), 126109 (2021)
- [22] Qiao, L., Xu, J., Bai, J., Zhang, Y.: Fully local transition closure model for hypersonic boundary layers considering crossflow effects. *AIAA J.* 59(5), 1692–1705 (2021)