

A Unified Framework for Non-Intrusive, Data-Driven Reduced-Order Modeling of Complex Parametric Flows

Allan Moreira de Carvalho

July 20, 2025

Contents

Abstract

This dissertation presents a unified, non-intrusive framework for the development of data-driven reduced-order models (ROMs) capable of rapidly and accurately reconstructing high-dimensional flow fields in complex engineering systems. The prohibitive computational expense of high-fidelity Computational Fluid Dynamics (CFD) simulations motivates this work, particularly for many-query applications such as design optimization, uncertainty quantification, and real-time control. The proposed methodology leverages Proper Orthogonal Decomposition (POD) for dimensionality reduction, coupled with advanced machine learning regressors, to create surrogate models that bridge the gap between computational efficiency and predictive fidelity.

Two significant challenges in ROM development are addressed through distinct, yet complementary, investigations. First, a robust surrogate modeling pipeline is developed to capture the highly nonlinear physics of supersonic nozzle flows, characterized by shock wave-boundary layer interactions. This study introduces a novel hybrid loss function for Artificial Neural Networks (ANNs) that combines errors in both the latent and physical spaces, enhancing reconstruction accuracy. A systematic hyperparameter optimization using Bayesian Optimization with Hyperband (BOHB) is performed, and a comprehensive comparative analysis between ANNs and Gaussian Processes (GPs) reveals a crucial trade-off: ANNs offer superior robustness to data scarcity and noise, while GPs provide higher precision in data-rich scenarios.

Second, the framework is extended to tackle the challenge of significant geometric variability, a common hurdle in engineering design. A novel mesh morphing pipeline, based on harmonic mapping and structured interpolation, is introduced to regularize geometric and field data from parametrically deformed compressor blades (NASA Rotor 37). This enables the effective application of POD and a subsequent GP regression (GPR) model to predict aerodynamic fields and blade geometry with exceptional accuracy ($R^2 > 0.98$, NRMSE $< 3\%$).

Model interpretability is addressed using SHapley Additive exPlanations (SHAP), confirming that the learned models align with physical principles and revealing how ANNs leverage low-energy POD modes to resolve sharp flow features. The developed surrogates achieve computational speed-ups of several orders of magnitude (up to 12,000x) over traditional CFD solvers, demonstrating their profound potential to accelerate the engineering design cycle. This work culminates in a unified, validated, and interpretable framework that significantly advances the state-of-the-art in data-driven surrogate modeling for complex parametric flows.

Chapter 1

Introduction

1.1 The Challenge of Computational Cost in Modern Fluid Dynamics

In modern engineering disciplines such as aerospace, energy systems, and automotive design, high-fidelity Computational Fluid Dynamics (CFD) simulations have become an indispensable tool.[1, 2] By numerically solving the governing equations of fluid motion, such as the Reynolds-Averaged Navier-Stokes (RANS) or Large Eddy Simulation (LES) equations, engineers can gain detailed insights into complex physical phenomena, from turbulent flows over airfoils to shock waves in rocket nozzles. However, the accuracy of these simulations comes at a steep price. High-fidelity models often require millions of degrees of freedom and can take days or even weeks of CPU time on massively parallel computing resources to complete a single simulation.[1]

This "computational bottleneck" renders the direct use of high-fidelity CFD impractical for many-query applications, which are central to the modern engineering design cycle.[3, 4] Tasks such as parametric design exploration, multi-objective optimization, real-time control, and uncertainty quantification require hundreds or thousands of model evaluations, making the reliance on full-order models (FOMs) prohibitively expensive and time-consuming.[5, 6] This computational barrier has motivated a significant research effort toward the development of surrogate models—computationally inexpensive approximations that aim to replicate the input-output behavior of the high-fidelity FOM with sufficient accuracy.[7, 8]

1.2 Reduced-Order Modeling as a Solution

Among the various classes of surrogate models, Reduced-Order Models (ROMs) have emerged as a particularly powerful and structured approach for complex physical systems.[5] The fundamental principle of a ROM is to project the high-dimensional dynamics

of the FOM onto a low-dimensional subspace, or manifold, that captures the most dominant or energetically significant features of the system.[3, 4] This projection drastically reduces the degrees of freedom, enabling near real-time predictions.

The construction and deployment of a ROM typically involve two distinct stages.[3, 4] The first is a computationally intensive "offline" stage, during which a set of high-fidelity FOM simulations is performed to generate a database of "snapshots" of the system's behavior across a range of parameters. This data is then used to construct the low-dimensional basis and train the surrogate model. The second is a rapid "online" stage, where the trained ROM is deployed to make new predictions for unseen parameter inputs at a fraction of the cost of the original FOM.[5] The significant upfront investment in the offline stage is justified by the massive computational savings achieved during the online stage, especially in many-query contexts.

1.3 Research Gaps and Thesis Objectives

Despite significant progress, the development of robust and generalizable non-intrusive, data-driven ROMs for fluid dynamics still faces several critical challenges. This dissertation aims to address four key research gaps:

1. **Robustness in Complex Physics:** Standard ROM techniques can struggle to accurately capture highly nonlinear and multi-scale phenomena, such as shock waves, turbulence, and strong boundary layer interactions, which are ubiquitous in engineering flows.[1] The ability of a ROM to faithfully reconstruct these features is paramount for its practical utility.
2. **Handling Geometric Parametricity:** A major limitation of classical projection-based ROMs, such as those using Proper Orthogonal Decomposition (POD), is the requirement of a consistent mesh topology across all data snapshots. This assumption is fundamentally violated in many design optimization problems where the shape of the computational domain itself is a variable, thus hindering the application of ROMs to a vast class of important engineering problems.[9]
3. **Model Selection and Optimization:** The integration of machine learning (ML) into ROMs has opened up a vast design space of potential models (e.g., Artificial Neural Networks, Gaussian Processes) and hyperparameters. However, there is a lack of systematic guidelines for selecting and tuning these models for specific fluid dynamics problems, often leading to ad-hoc choices and suboptimal performance.[9]
4. **Interpretability:** Many advanced ML models function as "black boxes," making it difficult to understand the physical reasoning behind their predictions. This lack

of transparency can erode trust and prevent the discovery of new physical insights, which is a key goal of scientific inquiry.[10, 11]

The central objective of this dissertation is to develop, validate, and synthesize a unified, non-intrusive ROM framework that systematically addresses these challenges. The goal is to create a methodology that is not only computationally efficient but also robust to both physical and geometrical complexity, and whose predictions are both accurate and interpretable.

1.4 Key Contributions and Thesis Outline

This work makes several key contributions to the field of data-driven reduced-order modeling:

1. A comprehensive comparative surrogate modeling framework is developed for physically complex flows, featuring a novel hybrid loss function for ANNs and a rigorous hyperparameter optimization strategy that provides clear guidelines on the trade-offs between different ML regressors.[9]
2. A novel mesh morphing pipeline is introduced to handle complex geometric variations. This technique regularizes data from inconsistent meshes, enabling the successful application of POD-based ROMs to parametric turbomachinery design problems for the first time in this context.[9]
3. A thorough evaluation of model performance, robustness to noise, and interpretability using SHAP is conducted. This analysis provides not only validation but also deeper physical insights into how the models function, offering practical guidelines for the application of ML-ROMs in engineering.

The remainder of this dissertation is structured as follows. Chapter 2 lays the theoretical foundations of fluid dynamics, reduced-order modeling, and the machine learning techniques used throughout this work. Chapter 3 presents the first major investigation, focusing on the development of a robust surrogate modeling framework for supersonic nozzle flows with complex shock structures. Chapter 4 details the second major investigation, introducing the mesh morphing pipeline to extend the framework to problems with high-dimensional geometric variability, using the NASA Rotor 37 compressor blade as a benchmark. Chapter 5 synthesizes the findings from both investigations into a unified framework and discusses the broader implications. Finally, Chapter 6 concludes the dissertation with a summary of contributions, a discussion of limitations, and an outlook on future research directions.

Chapter 2

Foundations of Reduced-Order Modeling and Machine Learning in Fluid Dynamics

2.1 Governing Equations of Fluid Motion

The motion of a viscous, heat-conducting, compressible fluid is governed by the Navier-Stokes equations, which represent the conservation of mass, momentum, and energy. In conservative form for a two-dimensional domain, these equations can be written as [9]:

$$\frac{\partial U}{\partial t} + \nabla \cdot F^c(U) - \nabla \cdot F^v(U, \nabla U) = S$$

where U is the vector of conserved variables, F^c is the convective flux tensor, F^v is the viscous flux tensor, and S is a source term vector. These are defined as:

$$U = \begin{bmatrix} \rho \\ \rho \mathbf{v} \\ \rho E \end{bmatrix}, \quad F^c = \begin{bmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + pI \\ (\rho E + p)\mathbf{v} \end{bmatrix}, \quad F^v = \begin{bmatrix} 0 \\ \tau \\ \tau \cdot \mathbf{v} + \kappa \nabla T \end{bmatrix}$$

Here, ρ is the density, \mathbf{v} is the velocity vector, E is the specific total energy, p is the pressure, I is the identity matrix, τ is the viscous stress tensor, κ is the thermal conductivity, and T is the temperature. To close this system, an equation of state, such as the ideal gas law, is required, along with models for viscosity and thermal conductivity.[9]

For many engineering applications involving turbulent flows, direct numerical simulation (DNS) of these equations is computationally intractable. Instead, the Reynolds-Averaged Navier-Stokes (RANS) equations are often employed. In RANS, the flow variables are decomposed into mean and fluctuating components, and the resulting equations are solved for the time-averaged flow field. This introduces the Reynolds stress tensor,

which must be modeled using a turbulence model. The studies in this dissertation utilize the k- ω Shear Stress Transport (SST) model, a widely used two-equation model that provides a good balance of accuracy and computational cost for a broad range of aerodynamic flows.[9, 9]

In the limit of high Reynolds number ($Re \rightarrow \infty$), where viscous and thermal effects become confined to thin boundary layers, the Navier-Stokes equations simplify to the Euler equations. This is achieved by setting the viscous stress tensor and heat flux to zero ($\tau = 0, \kappa = 0$). The Euler equations govern inviscid, compressible flow and are capable of capturing key phenomena like shock waves. This simplification provides the basis for the dual-fidelity approach used in Chapter 3, where a quasi-one-dimensional Euler solver serves as a computationally cheap, low-fidelity model.[9]

2.2 Paradigms of Reduced-Order Modeling

2.2.1 Intrusive vs. Non-Intrusive ROMs

ROMs can be broadly categorized into two families: intrusive and non-intrusive.[3, 5] **Intrusive ROMs**, such as those based on Galerkin projection, directly manipulate the governing equations of the FOM. The full-order equations are projected onto the low-dimensional basis, resulting in a much smaller system of differential equations that describes the evolution of the modal coefficients. While this approach can yield highly accurate and physically consistent models, it is "intrusive" because it requires access to and modification of the FOM's source code. This makes intrusive ROMs difficult to implement, often solver-specific, and challenging to apply to complex, industrial-grade CFD codes.[5]

Non-intrusive ROMs (NIROMs), in contrast, treat the FOM as a "black box." They rely entirely on the data generated by the FOM (the snapshots) to learn a mapping from the input parameters to the output solution in the reduced-order space. This is typically accomplished using data-driven regression techniques from machine learning. The key advantage of NIROMs is their flexibility and ease of implementation; they can be built on top of any existing simulation code without modification. This portability makes them highly attractive for engineering applications. The framework developed in this dissertation is exclusively non-intrusive.[3]

2.2.2 Dimensionality Reduction with Proper Orthogonal Decomposition (POD)

Proper Orthogonal Decomposition (POD) is the cornerstone of linear dimensionality reduction in fluid dynamics and is central to the framework presented herein.[3, 12] First

introduced to the fluid dynamics community by Lumley (1967) as a method to objectively extract "coherent structures" from turbulent flows, POD provides a systematic way to find the most efficient linear basis for representing a given dataset.[9, 12]

Mathematically, the goal of POD is to find an orthonormal basis $\{\phi_i\}$ that is optimal in the sense that it maximizes the average projection of the data onto the basis vectors. This is equivalent to minimizing the average reconstruction error for a given number of basis vectors (modes). This optimal basis can be found by solving an eigenvalue problem on the two-point correlation tensor of the flow field. In practice, using the method of snapshots proposed by Sirovich (1987), this is accomplished by performing a Singular Value Decomposition (SVD) on the snapshot matrix $X \in \mathbb{R}^{n \times m}$, where n is the number of degrees of freedom in a single snapshot (e.g., mesh points) and m is the number of snapshots.[9, 9] The SVD of X is given by:

$$X = U \Sigma V^T$$

The columns of the matrix $U \in \mathbb{R}^{n \times n}$ are the left singular vectors, which are the POD modes (ϕ_i). These modes form an orthonormal basis for the flow fields. The matrix $\Sigma \in \mathbb{R}^{n \times m}$ is a diagonal matrix containing the singular values σ_i , and their squares (σ_i^2) are proportional to the amount of "energy" (in an L_2 sense) captured by the corresponding mode. The columns of $V \in \mathbb{R}^{m \times m}$ are the right singular vectors.

By arranging the POD modes in descending order of their corresponding singular values, we obtain a basis that is ordered by energetic significance. A low-dimensional representation is then achieved by truncating this basis, keeping only the first k modes ($k \ll n$) that capture a desired amount of the total energy, typically quantified by a cumulative explained variance criterion [9]:

$$CEV_k = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{j=1}^m \sigma_j^2}$$

Any flow field in the original dataset can then be approximated as a linear combination of these k modes. The projection of a snapshot onto this truncated basis yields a low-dimensional vector of modal coefficients, which becomes the target for the machine learning regressor.

2.3 Machine Learning for Surrogate Modeling

2.3.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs), and specifically the feedforward Multilayer Perceptron (MLP), are powerful function approximators inspired by the structure of biological

nervous systems.[13] Due to their ability to model highly complex and nonlinear relationships, they are a natural choice for surrogate modeling in NIROMs.[9] An MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer contains a number of nodes (neurons), and each neuron in a layer is connected to all neurons in the subsequent layer. The computation in a hidden layer l involves an affine transformation of the output from the previous layer, followed by a nonlinear activation function h^l :

$$\mathbf{a}^l = h^l(W^l \mathbf{a}^{l-1} + \mathbf{b}^l)$$

where W^l and \mathbf{b}^l are the weight matrix and bias vector for layer l , respectively. These weights and biases are the trainable parameters of the network. The choice of activation function (e.g., sigmoid, ReLU, swish) introduces the nonlinearity necessary to approximate complex functions.[9]

The network is trained by minimizing a loss function that measures the discrepancy between the network's predictions and the true target values from the training data. This optimization is typically performed using gradient-based methods, where the gradients of the loss with respect to the parameters are calculated efficiently via the backpropagation algorithm.[14] In recent years, the concept of Physics-Informed Neural Networks (PINNs) has gained traction. PINNs embed physical knowledge, such as the governing partial differential equations (PDEs), directly into the loss function as a regularization term. This encourages the network to produce solutions that are not only data-consistent but also physically plausible.[15, 16] This concept provides important context for the novel hybrid loss function developed in Chapter 3, which can be seen as a form of structural, rather than differential, physical constraint.

2.3.2 Gaussian Processes (GPs)

Gaussian Process (GP) regression offers a different, probabilistic approach to surrogate modeling.[9, 17] A GP is a non-parametric, Bayesian method that defines a distribution over functions. It is fully specified by a mean function $m(\mathbf{x})$ and a covariance function, or kernel, $k(\mathbf{x}, \mathbf{x}')$. [17] The kernel is the crucial component of a GP, as it encodes our prior assumptions about the function being modeled, such as its smoothness and length-scales. It defines the similarity between data points: if two points \mathbf{x} and \mathbf{x}' are "close" according to the kernel, their function values $f(\mathbf{x})$ and $f(\mathbf{x}')$ are assumed to be highly correlated.[18]

A widely used kernel in engineering applications is the squared exponential (SE) kernel, also known as the Radial Basis Function (RBF) kernel [9]:

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2l^2} \|\mathbf{x} - \mathbf{x}'\|^2\right)$$

where the hyperparameters σ_f^2 (signal variance) and l (characteristic length-scale) are typically learned from the data by maximizing the log marginal likelihood.[18] The seminal text by Rasmussen and Williams (2006) provides a comprehensive treatment of GPs for machine learning and is a foundational reference for this work.[9, 17, 19]

A key advantage of the GP framework is its ability to provide principled uncertainty quantification. For any new prediction, a GP provides not only a mean value (the prediction) but also a predictive variance, which quantifies the model’s confidence in that prediction.[9]

2.4 Model Evaluation and Interpretation

2.4.1 Performance Metrics

To quantitatively assess the performance of the surrogate models, two standard regression metrics are used throughout this work. The **Coefficient of Determination (R^2)** measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s). An R^2 value of 1 indicates a perfect fit. The **Normalized Root Mean Square Error (NRMSE)** provides a measure of the average prediction error, normalized by the range of the true data to allow for comparison across different physical fields.[9, 9]

2.4.2 Robustness and Generalization

A good surrogate model must not only perform well on the data it was trained on but also generalize to new, unseen data. To assess generalization and mitigate the risk of overfitting, **k-fold cross-validation** is employed. This technique involves partitioning the dataset into k subsets, iteratively training the model on $k-1$ subsets, and validating on the remaining one. This provides a more robust estimate of the model’s performance and stability.[9] Furthermore, to simulate the uncertainties inherent in real-world applications, **noise robustness tests** are conducted by injecting controlled levels of Gaussian noise into the model inputs and observing the degradation in prediction accuracy.[9]

2.4.3 Interpretability with SHapley Additive exPlanations (SHAP)

Addressing the “black-box” nature of complex ML models is crucial for building trust and gaining scientific insight.[10, 11] SHapley Additive exPlanations (SHAP) is a powerful model-agnostic technique for explaining individual predictions.[9] Rooted in cooperative game theory, SHAP assigns an “importance” value (the Shapley value) to each input feature, representing its marginal contribution to pushing the model’s output away from the baseline average prediction.[20] SHAP provides both **local explanations** (for a single

data point) and **global explanations** (by aggregating the Shapley values across the entire dataset), making it an invaluable tool for understanding feature importance and validating that the model’s learned relationships are physically meaningful.[20]

2.5 Handling Geometric Variability in CFD

A significant hurdle in applying POD-based ROMs to design optimization is geometric variability. Since POD requires a one-to-one correspondence between points across all snapshots, any change in the domain geometry breaks this requirement, rendering the standard approach unusable.[9] In CFD-based optimization, this is often addressed with mesh deformation or morphing techniques, which adapt an existing mesh to a new geometry instead of performing a full remeshing, thereby saving significant time.[21] Common methods include Radial Basis Functions (RBF) and Inverse Distance Weighting (IDW).[22]

For complex, non-trivial deformations, more advanced parameterization techniques are needed. **Harmonic mapping** is a method from differential geometry that creates a low-distortion mapping of a 3D surface onto a 2D parametric domain (e.g., a unit square) by solving Laplace’s equation.[23, 24] This technique is particularly attractive because it can ”flatten” complex surfaces while preserving local angles and minimizing area distortion, providing a powerful basis for the novel data regularization pipeline introduced in Chapter 4.

Chapter 3

A Robust Surrogate Modeling Framework for Parametric Internal Flows

3.1 Problem Statement: Supersonic Nozzle Flow

The first investigation in this dissertation focuses on developing a robust surrogate modeling framework for a physically complex problem: the steady-state, two-dimensional, compressible, viscous flow of hot air through a convergent-divergent nozzle.[9] This problem was chosen as it presents a formidable challenge for surrogate modeling due to its rich and highly nonlinear physics. The flow is characterized by high Reynolds numbers (on the order of 10^7) and features complex shock wave-boundary layer interactions (SWBLI) in the divergent section of the nozzle. The formation, position, and strength of these shock structures are extremely sensitive to the operating conditions, making this an excellent benchmark for assessing a ROM's ability to capture nonlinear, discontinuous phenomena.[9]

The problem is parameterized by three key variables that govern the flow behavior: the inlet total pressure (p_0), the inlet total temperature (T_0), and the nozzle divergence angle (θ_{div}). The ranges for these parameters were chosen to span a variety of flow regimes, including over-expanded and under-expanded conditions, ensuring the formation of complex shock patterns.[9]

3.2 Dual-Fidelity Data Generation

A dual-fidelity approach was adopted for data generation to explore different input representations for the surrogate models.

- **High-Fidelity Data:** The ground-truth data was generated using the open-source

SU2 CFD solver to perform two-dimensional RANS simulations with the k- ω SST turbulence model. To ensure the numerical rigor of the high-fidelity dataset, a comprehensive verification and validation process was undertaken. A Grid Convergence Index (GCI) study was performed on three mesh levels, demonstrating monotonic and asymptotic convergence with a GCI value near unity, confirming grid independence. Furthermore, the solver was validated against experimental pressure distribution data for the baseline nozzle geometry, showing good agreement across multiple operating points.[9]

- **Low-Fidelity Data:** A corresponding low-fidelity dataset was generated using a custom-built, in-house finite volume solver for the quasi-one-dimensional Euler equations. This model captures the essential compressibility effects, including shock formation, but neglects viscosity and thermal effects, making it computationally orders of magnitude faster than the RANS solver. The Euler solver was also validated against established benchmark results from the literature.[9]

Using these two solvers, a series of datasets was constructed using an optimized Latin Hypercube Sampling (LHS) method to ensure uniform coverage of the parametric design space. The datasets varied in size (small, medium, large) and were designed to test two distinct input snapshot formats for the surrogate models: (1) a simple vector of the scalar input parameters (T), and (2) a vector containing the field solution from the low-fidelity quasi-1D Euler solver.[9]

Table 3.1: Nomenclature for the numerical experiments, categorized by boundary conditions, dataset size, and input format.[9]

B.C.	Case Study Name	Samples (Train, Val, Test)	Input Snapshot	Output Snapshot
Prescribed T_w	PTSS	Small (128,17,16)	T	p, T, M
	PTSF		T	
	PTMS	Medium (323,41,40)	T	p, T, M
	PTMF		T	
	PTLS	Large (644,81,81)	T	p, T, M
	PTLF		T	
Adiabatic Wall	ADSS	Small (130,17,16)	T	p, T, M
	ADSF		T	
	ADMS	Medium (332,42,41)	T	p, T, M
	ADMF		T	
	ADLS	Large (652,82,82)	T	p, T, M
	ADLF		T	

3.3 The ML-ROM Pipeline

3.3.1 Preprocessing and Dimensionality Reduction

Prior to model training, the snapshot data undergoes a two-step scaling process to improve numerical stability. First, a custom slice-wise mean centering is applied, where the mean is subtracted from each physical field (pressure, temperature, Mach number) independently. This prevents distortion from variables with vastly different physical scales. Second, a MaxAbsScaler normalizes each feature to the range $[-1, 1]$. [9]

Following scaling, POD is applied to the high-fidelity RANS snapshots to perform dimensionality reduction. The number of retained POD modes is determined by a strict cumulative explained variance threshold of 99.99% to ensure minimal information loss. This process reduces the dimensionality of each snapshot from 7878 (for the 2D fields) to a much smaller number of modal coefficients (e.g., from 16 to 154, depending on the dataset size). [9]

3.3.2 Innovation: A Hybrid Loss Function for ANNs

A key innovation of this study is the development of a novel hybrid loss function for training the ANN-based surrogates. Standard data-driven ROMs train the regressor to minimize the error between the predicted and true modal coefficients in the low-dimensional latent space. However, this does not guarantee accuracy in the reconstructed high-dimensional physical space. To address this, a loss function was formulated as a weighted sum of two terms: the Mean Squared Error (MSE) in the latent space ($\mathcal{L}_{reduced}$) and the MSE in the reconstructed physical space ($\mathcal{L}_{reconstructed}$) [9]:

$$\begin{aligned}\mathcal{L}_{reduced} &= \frac{1}{m} \sum_{i=1}^m \|\Phi(\bar{X}_i) - \bar{y}_i\|^2 \\ \mathcal{L}_{reconstructed} &= \frac{1}{m} \sum_{i=1}^m \|(\Phi(\bar{X}_i) - \bar{y}_i)V_y^T\|^2 \\ \mathcal{L} &= w_{recon}\mathcal{L}_{reduced} + (1 - w_{recon})\mathcal{L}_{reconstructed}\end{aligned}$$

Here, Φ is the ANN, \bar{X} and \bar{y} are the latent-space inputs and outputs, V_y is the POD basis for the output data, and w_{recon} is a tunable hyperparameter.

This formulation can be interpreted as a form of structural physical constraint. While it does not enforce the governing PDEs directly like a PINN, it penalizes deviations from the physically plausible solution space defined by the POD basis. The POD modes are derived from snapshots that are solutions to the RANS equations, and thus they inherently contain the physics of the problem. By including the reconstruction error, the loss function ensures that the model's output, when projected back to the full-order

space, remains consistent with the dominant physical structures captured by POD. This "POD-Informed" or "Structure-Informed" approach provides a practical bridge between purely data-driven models and more complex, fully physics-informed models, enhancing generalization and reconstruction fidelity.[9]

3.3.3 Rigorous Hyperparameter Optimization

To avoid ad-hoc model design, a systematic and automated hyperparameter optimization was performed using the Bayesian Optimization with Hyperband (BOHB) algorithm. BOHB efficiently explores a large search space by combining the early-stopping strategy of Hyperband with the model-based search of Bayesian Optimization.[9] A comprehensive search was conducted over the number of layers, neurons per layer, activation functions, learning rate, weight decay, and the reconstruction weight w_{recon} .

The results of this extensive tuning process yielded a significant finding: across all datasets and input types, the optimal ANN architectures were consistently shallow, typically consisting of only one or two hidden layers with a moderate number of neurons. Activation functions like sigmoid, hard sigmoid, and swish were found to be most effective. This contradicts the "deeper is better" mantra often associated with deep learning and underscores that for this class of regression problems, simpler, well-tuned architectures can outperform more complex ones, while also being more computationally efficient to train.[9]

Table 3.2: Best ANN hyperparameter configurations found through BOHB optimization for each case study.[9]

Study	NRMSE	R^2	H	J_i	Activation	λ_{wd}	Dropout	η_0	w_{recon}
ADLF	0.026	0.959	1	252	swish	0.000	0.043	0.002	0.350
ADLS	0.026	0.972	1	359	sigmoid	0.003	0.126	0.002	0.473
ADMF	0.047	0.966	1	204	hard_sigmoid	0.009	0.054	0.004	0.846
ADMS	0.035	0.963	1	219	sigmoid	0.008	0.049	0.005	0.306
ADSF	0.038	0.956	1	71	hard_sigmoid	0.002	0.012	0.003	0.555
ADSS	0.034	0.947	1	104	selu	0.000	0.010	0.006	0.433
PTLF	0.036	0.971	1	88	sigmoid	0.003	0.025	0.003	0.421
PTLS	0.032	0.969	2	39	gelu	0.000	0.010	0.001	0.564
PTMF	0.060	0.960	1	162	hard_sigmoid	0.000	0.101	0.001	0.620
PTMS	0.048	0.959	1	74	sigmoid	0.000	0.012	0.001	0.169
PTSF	0.057	0.932	1	191	sigmoid	0.000	0.011	0.002	0.709
PTSS	0.055	0.944	1	250	hard_sigmoid	0.000	0.188	0.002	0.546

3.4 Results and Analysis

3.4.1 Generalization Performance

A 5-fold cross-validation was conducted to rigorously evaluate the generalization performance and stability of the tuned ANN and GP models. The results revealed a clear and consistent trade-off between the two model types. GP models generally achieved a lower mean NRMSE, indicating higher average pointwise accuracy, particularly on the larger, cleaner datasets. However, ANN models consistently achieved higher R^2 scores, suggesting they were better at capturing the overall variance and trends in the data. Critically, the ANNs demonstrated significantly greater stability, with much lower variance in performance across the validation folds, especially in data-scarce regimes. In contrast, the GP models were highly sensitive to the amount of training data, showing large performance variations and degradation on the smaller datasets.[9]

Table 3.3: Error metrics (mean \pm std. dev.) from 5-fold cross-validation, comparing GP and ANN models on unseen test data.[9]

Dataset	NRMSE		R^2	
	GP	ANN	GP	ANN
ADLF	0.022 ± 0.002	0.026 ± 0.007	0.967 ± 0.007	0.972 ± 0.011
ADMF	0.027 ± 0.003	0.035 ± 0.004	0.962 ± 0.004	0.969 ± 0.004
ADSF	0.065 ± 0.027	0.074 ± 0.015	0.931 ± 0.018	0.938 ± 0.006
ADLS	0.022 ± 0.001	0.029 ± 0.002	0.965 ± 0.007	0.962 ± 0.012
ADMS	0.025 ± 0.004	0.037 ± 0.008	0.960 ± 0.004	0.961 ± 0.004
ADSS	0.038 ± 0.009	0.039 ± 0.007	0.943 ± 0.012	0.934 ± 0.024
PTLF	0.027 ± 0.002	0.030 ± 0.003	0.954 ± 0.027	0.969 ± 0.010
PTMF	0.051 ± 0.039	0.042 ± 0.008	0.939 ± 0.015	0.955 ± 0.015
PTSF	0.356 ± 0.192	0.074 ± 0.021	0.704 ± 0.074	0.876 ± 0.027
PTLS	0.023 ± 0.001	0.030 ± 0.003	0.963 ± 0.010	0.967 ± 0.010
PTMS	0.025 ± 0.005	0.045 ± 0.008	0.950 ± 0.014	0.944 ± 0.017
PTSS	0.046 ± 0.009	0.068 ± 0.022	0.884 ± 0.012	0.886 ± 0.022

3.4.2 Robustness to Input Noise

To assess model performance under more realistic conditions, a noise robustness study was performed by adding controlled levels of Gaussian noise to the test set inputs. The results reinforced the findings from the cross-validation. ANN models exhibited a much more graceful degradation in performance as noise levels increased. GP models, while superior in noise-free scenarios, were more sensitive to input perturbations, with their accuracy dropping more sharply. This suggests that ANNs are the more robust choice for practical engineering applications where input data from sensors or low-fidelity models

may be inherently noisy or uncertain.[9]

3.4.3 Interpretability through SHAP

SHAP analysis was employed to open the "black box" of the trained models and validate their physical consistency. For models using scalar inputs, the SHAP analysis confirmed physical intuition: the inlet total pressure (p_0) was overwhelmingly the most influential feature for both ANN and GP models, as it is the primary driver of the compressible flow. Conversely, the wall temperature (T_w) had a negligible impact, which helps explain the poorer performance of models trained on datasets that included this less-influential parameter.[9]

A more profound finding emerged from the analysis of models using field-based inputs (the POD coefficients of the 1D Euler solution). The SHAP analysis for the GP model showed that the feature importance strictly followed the energetic hierarchy of the POD modes—the first few, highest-energy modes were the most important. The ANN, however, displayed a different behavior. While it also relied heavily on the high-energy modes, it assigned significant importance to several mid- and low-energy POD modes as well. In fluid dynamics, it is known that high-energy POD modes capture large-scale, global flow structures, while low-energy modes are associated with small-scale, localized features. In this nozzle problem, the most critical localized features are the shock waves. The ANN's ability to leverage these low-energy modes explains its superior performance in reconstructing the sharp, discontinuous shock structures that the GP model, relying only on the smoothest, highest-energy modes, tended to smear out. This demonstrates a symbiotic relationship: POD provides a structured, hierarchical basis, and the ANN learns to selectively use modes from across the energy spectrum to reconstruct both the smooth and discontinuous parts of the flow field.[9]

3.4.4 Flow Field Reconstruction and Computational Speed-up

Qualitative analysis of the reconstructed 2D flow fields confirmed the quantitative and interpretability results. For a challenging test case with a prominent Mach disc, both models captured the overall flow structure. However, the ANN prediction provided a much sharper and more accurate resolution of the shock wave and post-shock flow features. Pointwise error maps revealed that the GP's prediction errors were larger and more diffuse, whereas the ANN's errors were smaller and highly localized to the immediate vicinity of the shock itself.[9]

Finally, the practical value of the framework was quantified by the computational speed-up. The trained surrogate models could perform an inference (a full 2D field reconstruction) in under 0.25 seconds. Compared to the high-fidelity SU2 solver, which took nearly 3 minutes per simulation, this represents a speed-up of up to 744x for the

ANN and 7374x for the GP. These dramatic gains in efficiency unlock the possibility of performing many-query analyses that would be intractable with traditional CFD.[9]

Table 3.4: Summary of computational time and speed-up factors.[9]

Task	Average Time
Dataset Generation Time (SU2)	21h 41m 33s
Hyperparameter Tuning Time (Fields)	7h 19m 55s
Model Training Time (Fields)	0h 12m 13s
Inference Time per Sample - SU2	169.60 s
Inference Time per Sample - ANN	0.228 s
Inference Time per Sample - GP	0.023 s
Speed-up over SU2 - ANN	744x
Speed-up over SU2 - GP	7374x

3.5 Chapter Conclusion

This chapter detailed the development and rigorous validation of a non-intrusive ML-ROM framework for a physically complex internal flow problem. Through systematic comparison, it was established that ANNs and GPs present a trade-off between robustness and precision. The introduction of a novel hybrid, structure-informed loss function was shown to improve ANN reconstruction fidelity. Finally, interpretability analysis not only confirmed the physical consistency of the models but also provided a deeper understanding of how the ANN and POD components work together to resolve complex, multi-scale flow features. The demonstrated accuracy and massive computational speed-up validate the framework’s potential for accelerating engineering design and analysis.

Chapter 4

Extending the Framework to Geometrically Complex Turbomachinery Flows

4.1 Problem Statement: Parametric Compressor Blade

Having established a robust framework for flows with fixed geometry but complex physics, the second investigation addresses a different, equally critical challenge: problems with significant geometric variability. This is a ubiquitous scenario in engineering design optimization, where the shape of a component is systematically altered to improve performance. The chosen benchmark for this study is the NASA Rotor 37, a transonic axial compressor blade that is a standard and challenging test case for CFD validation due to its complex 3D flow features, including shocks and tip-leakage vortices.[9, 25]

The blade geometry was parameterized using 28 design variables controlling the thickness and angle distributions at the hub and shroud. The core problem is that each of the 410 unique geometries generated via LHS for the training dataset has its own distinct surface mesh. As established in Chapter 2, standard POD requires a consistent mesh topology with a one-to-one point correspondence across all snapshots. The varying geometries fundamentally violate this condition, making the direct application of the POD-based ROM framework impossible. This chapter introduces a novel data regularization pipeline to overcome this limitation.[9]

4.2 Innovation: A Mesh Morphing Pipeline for Data Regularization

The central methodological innovation of this chapter is a three-stage mesh morphing pipeline designed to transform the collection of disparate, irregular surface meshes into a

set of regularized data fields that share a common topological structure. This pipeline is the key enabling technology that allows the POD-based ROM to be applied to problems with high-dimensional geometric parametricity.[9]

4.2.1 Stage 1: Harmonic Mapping (3D to 2D)

The first stage involves mapping each irregular 3D triangular surface mesh of the blade onto a simple, 2D parametric domain—in this case, a unit square ($\Omega = \times$). This is achieved using a discrete **harmonic map**. A harmonic map is a solution to Laplace’s equation ($\Delta\Phi = 0$) and has the property of being the smoothest possible mapping that satisfies the given boundary conditions, thereby minimizing distortion.[23, 24] The boundary vertices of the 3D mesh are first mapped to the boundary of the 2D square, preserving their relative arc-length distances. Then, the interior vertex positions in the 2D domain are found by solving the discrete Laplace equation. This process effectively ”unwraps” or ”flattens” the complex 3D blade surface onto a simple 2D plane.[9]

4.2.2 Stage 2: Interpolation onto a Regular Grid

Once the blade surface and its associated scalar fields (pressure and temperature) are represented in the 2D parametric domain, the data is interpolated onto a uniform, structured grid (e.g., 100×100 points). This is done by first creating a Delaunay triangulation of the mapped (irregular) 2D vertex locations. Then, for each point on the new regular grid, its barycentric coordinates within the enclosing Delaunay triangle are calculated. These same barycentric weights are then used to linearly interpolate the scalar field values (pressure, temperature) and, crucially, the original 3D vertex coordinates (X, Y, Z) from the values at the triangle’s vertices. This process results in a complete set of fields defined on a common, regular grid structure.[9]

4.2.3 Stage 3: 3D Reconstruction (Lifting)

The final stage is to ”lift” the regularized data back into 3D space. The interpolated 3D coordinate fields (X^*, Y^*, Z^*) on the regular grid now define the vertices of a new, regularized 3D surface mesh. Because every geometric instance from the dataset has been processed through this pipeline, all resulting snapshots—for geometry, pressure, and temperature—now reside on an identical grid structure, with a perfect one-to-one correspondence between grid points. This satisfies the prerequisite for applying POD.[9]



Figure 4.1: Schematic of the mesh morphing pipeline: (a) Original irregular 3D surface mesh. (b) Harmonic mapping to a 2D parametric domain. (c) Interpolation of fields onto a regular structured grid in the parametric domain. (d) Lifting of the regular grid and interpolated fields back to a regularized 3D surface mesh. This process ensures a consistent topology across all geometric variations.[9]

4.3 Application of the POD-GPR Framework

With the data successfully regularized, the ROM framework can be applied. A dataset was generated by running high-fidelity RANS simulations using ANSYS CFX for 410 variations of the Rotor 37 blade geometry, defined by the 28 input design parameters.[9]

POD was then applied independently to the snapshot matrices for the pressure field, the temperature field, and the blade geometry itself (i.e., the X, Y, Z vertex coordinates). The number of modes required to capture 99.9% of the energy was determined for each field.[9]

Table 4.1: Number of POD modes retained to capture 99.9% cumulative energy for fields on the blade suction (BS) and pressure (BP) surfaces.[9]

Field	Suction Surface (BS)	Pressure Surface (BP)
Pressure	57	73
Temperature	86	102
Geometry (Vertices)	21	20

A separate Gaussian Process Regression (GPR) model was then trained for each POD mode of each field. Each GPR model learns the mapping from the 28-dimensional vector of geometric input parameters to a single scalar modal coefficient. For this study, GPR was chosen over an ANN. The powerful mesh morphing pipeline acts as a significant regularization and filtering step, transforming the potentially chaotic geometric variations into a smooth, well-behaved, structured representation. The resulting latent space of POD coefficients is therefore much "cleaner" and more regular than the latent space in the nozzle flow problem, which was contaminated by the physics of the shock. In this clean, regularized space, the strong smoothness assumptions of the GPR model (enforced by its RBF kernel) make it an ideal and highly effective choice. This highlights a key principle: complexity can be handled either in the pre-processing stage (as done here) or in the model stage (as with the ANN in Chapter 3), and the optimal choice of regressor depends on where this complexity is managed.

4.4 Results and Analysis

4.4.1 Field and Geometry Reconstruction Accuracy

The POD-GPR surrogate model was evaluated on a validation set of 41 unseen geometric configurations. The qualitative results, comparing the reconstructed pressure and temperature fields to the CFD ground truth, showed excellent agreement. Pointwise error maps revealed that the largest errors were, as expected, concentrated in regions of high

gradients, such as near the leading edge and where shock waves impinge on the blade surface, but these errors remained small relative to the overall field variation.[9]

The quantitative metrics confirmed the model’s exceptional fidelity. Across all predicted fields, the coefficient of determination (R^2) was consistently high and the NRMSE was low. Most impressively, the model was able to reconstruct the blade geometry itself with near-perfect accuracy ($R^2 > 0.999$), demonstrating that the pipeline robustly captures the mapping from the abstract design parameters to the physical shape of the blade.[9]

Table 4.2: Mean surrogate model accuracy metrics for test cases. NRMSE is presented as a percentage of the field’s value range.[9]

Field	Surface	R^2 [-]	NRMSE [%]
Pressure	Suction	0.959	3.81
Pressure	Pressure	0.978	1.41
Temperature	Suction	0.963	3.46
Temperature	Pressure	0.965	2.31
Geometry	Suction	>0.999	0.09
Geometry	Pressure	>0.999	0.13

4.4.2 Computational Efficiency

The primary motivation for this work was to accelerate the design process, and the results demonstrate a profound success in this regard. A full CFD simulation for a single Rotor 37 geometry required approximately 10 minutes on a high-performance computing cluster. In contrast, the trained POD-GPR surrogate could predict the full pressure, temperature, and geometry fields for a new design in approximately 0.05 seconds on a standard desktop machine. This constitutes a computational **speed-up of 12,000x**. This dramatic reduction in evaluation time transforms the design process, enabling engineers to perform rapid, iterative design space exploration and optimization studies that would be completely intractable using traditional CFD alone.[9]

4.5 Chapter Conclusion

This chapter successfully demonstrated the extension of the non-intrusive ROM framework to one of the most challenging problems in engineering design: high-dimensional geometric parametricity. The novel mesh morphing pipeline, based on harmonic mapping, was shown to be a key enabling technology, effectively bridging the gap between geometrically inconsistent datasets and the requirements of POD-based model reduction. The resulting POD-GPR model achieved exceptional accuracy in predicting both

aerodynamic fields and the underlying geometry, all while delivering a computational speed-up of four orders of magnitude. This validates the methodology as a powerful tool for accelerating the turbomachinery design cycle.

Chapter 5

Synthesis and Discussion

5.1 A Unified View of the ROM Framework

The investigations presented in Chapters 3 and 4, while addressing different primary challenges, can be integrated into a single, unified framework for non-intrusive, data-driven reduced-order modeling. This framework provides a comprehensive workflow for developing surrogate models for complex parametric flows, from data generation to online deployment. The core of the framework consists of dimensionality reduction via POD followed by machine learning-based regression. The key innovations—the mesh morphing pipeline and the hybrid loss function—can be viewed as specialized modules that are activated to handle specific types of complexity.

If the problem involves geometric variability (as in Chapter 4), the mesh morphing pipeline is a mandatory pre-processing step to regularize the data before POD can be applied. If the problem involves fixed geometry but highly nonlinear physics with sharp, localized features like shocks (as in Chapter 3), the hybrid, structure-informed loss function can be employed to enhance the reconstruction fidelity of an ANN-based regressor. This unified view demonstrates the modularity and adaptability of the proposed methodology, allowing it to be tailored to the specific challenges posed by a given engineering problem.

5.2 Comparative Analysis of Surrogate Models and Input Representations

Synthesizing the results from both studies provides a more nuanced understanding of the trade-offs involved in model selection. The choice between an ANN and a GP is not arbitrary but should be informed by the nature of the latent space being modeled.

- In the nozzle flow problem (Chapter 3), the latent space of POD coefficients was

”wild”—it directly encoded the highly nonlinear physics of the shock wave, leading to a complex mapping problem. In this scenario, the flexible, universal approximation capability of the ANN proved superior, especially its ability to leverage low-energy modes to capture discontinuities.

- In the rotor blade problem (Chapter 4), the mesh morphing pipeline acted as a powerful regularizer, effectively ”taming” the latent space. The mapping from the 28 geometric parameters to the POD coefficients was much smoother and well-behaved. Here, the GP, with its strong prior assumption of smoothness (via the RBF kernel), was the more effective and efficient choice.

This leads to a guiding principle: if significant complexity and nonlinearity are expected to be present in the latent space itself, a more flexible model like an ANN is preferable. If the complexity can be handled through data pre-processing and regularization, a more constrained model like a GP can be highly effective.

A similar synthesis can be made regarding the choice of input representation. The nozzle study showed that using a low-fidelity 1D Euler solution as an input did not consistently outperform using simple scalar boundary conditions. The rotor study achieved excellent results using only scalar geometric parameters as inputs. This suggests a principle of diminishing, or even negative, returns for some multi-fidelity approaches. If the physics of the low-fidelity model (e.g., inviscid Euler) are fundamentally inconsistent with the high-fidelity model (e.g., viscous RANS), the machine learning model may have to expend some of its capacity ”unlearning” the incorrect physics before it can learn the correct mapping from the low- to high-fidelity space. Therefore, a direct mapping from the fundamental, independent design and boundary parameters to the high-fidelity latent space is often a more direct, robust, and effective strategy for parametric ROMs.

5.3 The Role of Interpretability in Validating Physical Consistency

The use of SHAP in Chapter 3 was critical not only for building confidence in the model but also for generating deeper understanding. The analysis confirmed that the surrogate models were not merely ”pattern matching” in a way that was divorced from the underlying physics. The fact that total pressure was identified as the most dominant feature aligns perfectly with fundamental principles of compressible fluid dynamics, providing strong evidence that the model learned physically meaningful relationships.[9]

Furthermore, the insight that the ANN actively utilized low-energy POD modes to resolve shock structures moves interpretability beyond simple validation. It provides a mechanistic explanation for *why* the ANN outperformed the GP in that specific task.

It reveals a sophisticated interplay between the linear decomposition (POD) and the nonlinear regressor (ANN), where the model learns to exploit the full spectrum of the provided basis to reconstruct a complex, multi-scale field. This level of understanding is crucial for the intelligent design and application of future ML-ROM frameworks.

Chapter 6

Conclusions and Future Outlook

6.1 Summary of Contributions

This dissertation has presented a unified, non-intrusive, data-driven framework for the reduced-order modeling of complex parametric flows, making significant contributions to address key challenges in the field. The main achievements are summarized as follows:

1. The development of a comprehensive ML-ROM framework for nonlinear nozzle flows, which was rigorously validated through systematic hyperparameter tuning, k-fold cross-validation, and noise robustness testing. This investigation provided clear, data-driven guidelines on the trade-off between the robustness of ANNs and the precision of GPs.
2. The introduction of a novel hybrid, structure-informed loss function for ANNs that improves reconstruction fidelity by penalizing errors in both the latent and physical spaces. This was shown to be particularly effective for flows with sharp, localized features.
3. The creation of a novel mesh morphing pipeline based on harmonic mapping that successfully extends the ROM methodology to problems with high-dimensional geometric variability. This key enabling technology was validated on the challenging NASA Rotor 37 benchmark, demonstrating its power for accelerating turbomachinery design.
4. The delivery of practical insights into surrogate model selection and the use of SHAP for model interpretability. This work demonstrated that the models learn physically consistent relationships and provided a deeper understanding of how ML models interact with POD bases. The final validated framework achieves computational speed-ups of several orders of magnitude (700x to 12,000x) over high-fidelity CFD.

6.2 Limitations of the Current Framework

Despite the successes of the proposed framework, it is important to acknowledge its limitations, which point the way toward future research:

- **Data Dependency:** As a non-intrusive method, the framework’s performance is fundamentally tied to the quality and quantity of the high-fidelity training data. The generation of this data remains the most computationally expensive part of the workflow.[9]
- **Offline Cost:** While the online inference is extremely fast, the offline stage—including data generation, POD basis construction, and especially the BOHB hyperparameter optimization—still represents a significant upfront computational investment.[9]
- **Surface vs. Volume Fields:** The work presented here focused on the reconstruction of 2D volumetric fields and 3D surface fields. The extension to full 3D volumetric fields is a non-trivial step that would significantly increase data storage and processing requirements.[9]
- **Steady-State Assumption:** The framework was demonstrated exclusively on steady-state problems. Extending it to unsteady flows would require a different modeling approach for the temporal evolution of the POD coefficients, as the current regressors do not account for time-history effects.[3]

6.3 Recommendations for Future Research

Based on the findings and limitations of this work, several promising avenues for future research are recommended:

- **Advanced Dimensionality Reduction:** Future work should explore nonlinear dimensionality reduction techniques, such as convolutional autoencoders (CAEs). CAEs have the potential to learn more compact and expressive latent representations than the linear subspace of POD, which could be particularly advantageous for highly complex, turbulent flows.[3, 9]
- **Advanced Regression Models:** The use of more sophisticated neural network architectures that can intrinsically handle spatial data, such as Convolutional Neural Networks (CNNs) or Graph Neural Networks (GNNs), should be investigated. These models might be able to learn spatial features directly from the data, potentially reducing or even eliminating the need for a separate POD step.[7, 9]
- **Extension to Unsteady and 3D Volumetric Flows:** A natural and important next step is to extend the framework to full 3D, unsteady problems. This would

likely involve combining the spatial reconstruction techniques developed here with a recurrent neural network, such as a Long Short-Term Memory (LSTM) network, to model the temporal dynamics of the latent variables.[3]

- **Application to New Problem Domains:** The validated framework is now ready to be applied to new and challenging industrial problems. A particularly relevant application, motivated by the second study, is the design and optimization of centrifugal compressors for Carbon Capture, Utilization, and Storage (CCS) technologies, a critical area for sustainable energy systems.[9]

Bibliography

- [1] Benini, E. (2004). Three-dimensional multi-objective design optimization of a transonic compressor rotor. *Journal of Propulsion and Power*, 20(3), 559-565. [9]
- [2] Berkooz, G., Holmes, P., & Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25(1), 539-575. [9, 12]
- [3] Botsch, M., Bommers, D., & Kobbelt, L. (2005). Efficient linear system solvers for mesh processing. In *Mathematics of Surfaces XI* (pp. 62-83). Springer, Berlin, Heidelberg. [9]
- [4] Floater, M. S. (1997). Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design*, 14(3), 231-250. [9]
- [5] Forrester, A. I. J., Sobester, A., & Keane, A. J. (2008). *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons. [9]
- [6] Lumley, J. L. (1967). The structure of inhomogeneous turbulent flows. In *Atmospheric Turbulence and Radio Wave Propagation*. Nauka, Moscow. pp. 166-178. [9, 12]
- [7] Metz, B., Davidson, O., de Coninck, H., Loos, M., & Meyer, L. (Eds.). (2005). *IPCC Special Report on Carbon Dioxide Capture and Storage*. Cambridge University Press. [9]
- [8] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., & Tucker, P. K. (2005). Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1), 1-28. [9]
- [9] Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press. [9, 17]
- [10] Reid, L., & Moore, R. D. (1978). *Design and overall performance of four highly loaded, high-speed inlet stages for an advanced high-pressure-ratio core compressor* (NASA Technical Paper 1337). NASA Lewis Research Center. [9]

- [11] Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press. [9]
- [12] Sirovich, L. (1987). Turbulence and the dynamics of coherent structures. Part 1: Coherent structures. *Quarterly of Applied Mathematics*, 45(3), 561-571. [9]
- [13] Suder, K. L., & Celestina, M. L. (1994). *Experimental and computational investigation of the tip clearance flow in a transonic axial compressor rotor* (NASA TM-106711, AIAA-94-2091). NASA Lewis Research Center. [9]
- [14] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*. [9, 20]
- [15] Falkner, S., Klein, A., & Hutter, F. (2018). BOHB: Robust and efficient hyperparameter optimization at scale. In *International Conference on Machine Learning* (pp. 1437-1446). PMLR. [9]
- [16] Dutton, R. W. (1992). *Validation of a computational fluid dynamics code for analysis of subsonic flows through a side-inlet ramjet combustor* (NASA TM-105877). [9]
- [17] Celere, S. V., Nogueira, A. C., & Yanagihara, J. I. (2012). Numerical and experimental investigation of a supersonic air-water separator. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 34, 555-564. [9]

Appendix A

LaTeX Source Code

The entire content of this document constitutes the LaTeX source code requested.