FEDERAL UNIVERSITY OF ABC

GRADUATE COURSE IN ENERGY

Allan Moreira de Carvalho

REDUCED ORDER MODELS FOR PARAMETRIC DOMAINS: A DATA-DRIVEN
METHODOLOGY FOR THE ACCELERATION OF ENGINEERING SIMULATIONS

Santo André, SP – Brazil

July 2025

Allan Moreira de Carvalho

REDUCED ORDER MODELS FOR PARAMETRIC DOMAINS: A DATA-DRIVEN
METHODOLOGY FOR THE ACCELERATION OF ENGINEERING SIMULATIONS

Tese presented to the Graduate course of Federal University of
ABC, as a partial requirement for obtaining a Doctor's degree
in Energy.

Advisor: Prof. Dr. Daniel Jonas Dezan
Coadvisor:Prof. Dr. Wallace Gusmão Ferreira

Santo André, SP – Brazil

July 2025

This exemplar was revised and altered from the original version according to the observations done by the defense board at the defense day, under responsability of author and advisor's consent.

Santo André, _____ _____ , 20 _____

Author's signature: _____

Advisor's signature: _____

SIGN SHEET

Signatures of examining board members, which examinated and approved the defense of Thesis doutorado candidate Allan Moreira de Carvalho in July 10th, 2025:

Prof. Dr. Marcia Maria Penteado Marchesini - External Examiner

Prof. Dr. Wilson Carlos da Silva Júnior - External Examiner

Prof. Dr. Franciane Freitas Silveira - Internal Substitute Examiner

Prof. Dr. Rômulo Gonçalves Lins - Chair

I dedicate it...

Acknowledgement

I thank to...

Epigraph

"Learn from the mistakes of those who followed your advice"

(Unknown author)

Abstract

This dissertation introduces a comprehensive and adaptable framework for the construction of data-driven, parametric reduced-order models (ROMs) to accelerate the analysis of complex physical systems. High-fidelity numerical simulations, while accurate, impose a prohibitive computational cost in multi-evaluation scenarios such as design optimization, uncertainty quantification, and real-time control. To overcome this bottleneck, this work develops a unified pipeline that combines techniques in data preprocessing, dimensionality reduction, and machine learning. The core methodology employs Proper Orthogonal Decomposition (POD) to extract dominant, energy-optimal spatial modes from high-dimensional simulation snapshot data, which are then mapped from the design parameters using regression models like Gaussian Process Regression (GPR) or Artificial Neural Networks (ANNs). A critical innovation, developed to address challenges with complex geometries, is a sophisticated technique for handling geometric variations, which ensures topological consistency across differing parametric domains.

The framework's efficacy was first established in a 2D context by conducting a comparative analysis of POD-GPR and POD-ANN models for the reconstruction of supersonic nozzle flows, which are characterized by strong shock-wave/boundary-layer interactions. This initial study introduced a novel hybrid loss function for ANN training and employed SHapley Additive exPlanations (SHAP) to enhance model interpretability. However, the subsequent challenge of applying the framework to a complex 3D case—the parametric reconstruction of pressure and temperature fields on the blade surfaces of the NASA Rotor 37—revealed a fundamental limitation: meshes of varying sizes and inconsistent topologies prevented the direct application of POD. This problem necessitated the development of the aforementioned geometric mesh morphing technique, which proved crucial for enabling the methodology's extension to variable 3D shapes.

Results from both studies demonstrate exceptional predictive accuracy, with coefficients of determination (R^2) consistently exceeding 0.95, and computational speed-ups of several orders of magnitude (50x to over 10,000x) compared to full-fidelity simulations. This work not only provides a robust and validated methodology for accelerating simulation-based design and analysis cycles but also offers practical guidelines on model selection and training, establishing a foundation for the next generation of intelligent simulation tools.

Keywords: Reduced Order Model; Machine Learning; Proper Orthogonal Decomposition (POD); Parametric Modeling; Geometric Morphing

Resumo

Esta dissertação introduz um framework abrangente e adaptável para a construção de modelos de ordem reduzida (ROMs) paramétricos e orientados por dados, visando acelerar a análise de sistemas físicos complexos. Simulações numéricas de alta fidelidade, embora precisas, impõem um custo computacional proibitivo em cenários de que necessitam de múltiplas avaliações, como otimização de forma, quantificação de incertezas e controle em tempo real. Para superar esse gargalo, este trabalho desenvolve um pipeline unificado que combina técnicas de pré-processamento de dados, redução de dimensionalidade e aprendizagem de máquina. A metodologia central emprega a Decomposição Ortogonal Própria (POD) para extrair modos espaciais dominantes, que são então mapeados a partir dos parâmetros de projeto usando modelos de regressão como a Regressão por Processo Gaussiano (GPR) ou Redes Neurais Artificiais (ANNs). Uma inovação crítica, desenvolvida para lidar com geometrias complexas, é uma técnica sofisticada para o tratamento de variações geométricas, que garante a consistência topológica entre diferentes domínios paramétricos.

A eficácia do framework foi primeiramente estabelecida em um contexto 2D, através de uma análise comparativa de modelos POD-GPR e POD-ANN para a reconstrução de escoamentos em bocais supersônicos. Este estudo inicial introduziu uma função de perda híbrida para o treinamento da ANN e empregou SHapley Additive exPlanations (SHAP) para aumentar a interpretabilidade do modelo. No entanto, o desafio subsequente de aplicar o framework a um caso 3D—a reconstrução paramétrica dos campos de pressão e temperatura nas pás do NASA Rotor 37—revelou uma limitação fundamental: malhas com tamanhos distintos e topologias inconsistentes impediram a aplicação direta do POD. Este problema tornou necessário o desenvolvimento da técnica de tratamento geométrico usando deformação de malha, que se mostrou crucial para permitir a extensão da metodologia a formas 3D variáveis.

Os resultados de ambos os estudos demonstram excepcional precisão preditiva, com coeficientes de determinação (R^2) consistentemente acima de 0.95, e uma aceleração computacional de várias ordens de magnitude (de 50x a mais de 10.000x) em comparação com as simulações de alta fidelidade. Este trabalho não apenas fornece uma metodologia robusta e validada para acelerar os ciclos de projeto e análise baseados em simulação, mas também oferece diretrizes práticas quanto a seleção e o treinamento de modelos, estabelecendo a base para a próxima geração de ferramentas de simulação inteligentes.

Palavras-chave: Modelo de Ordem Reduzida; Aprendizado de Máquina; Decomposição Ortogonal Própria (POD); Modelagem Paramétrica; Mapeamento de Malha

List of Figures

List of Tables

Contents

1

Introduction

1.1 The Multi-Query Bottleneck in Computational Fluid Dynamics

In the modern era of aerospace and turbomachinery engineering, Computational Fluid Dynamics (CFD) has become an indispensable tool for analysis and design (Spalart, 2016). High-fidelity numerical models, such as those based on the Reynolds-Averaged Navier-Stokes (RANS) equations, Large Eddy Simulation (LES), or even Direct Numerical Simulation (DNS), offer remarkable precision in predicting the behavior of fluid flows (Pereira et al., 2021). This capability allows engineers to investigate complex physical phenomena, from the turbulent wake behind an aircraft to the intricate shock structures within a supersonic engine, with a level of detail that is often hard to achieve through physical experimentation alone (Schiestel, 2022).

However, the high fidelity of these simulations comes at a price: computational expense. Solving the governing equations of fluid motion across complex geometries discretized into millions or even billions of grid cells requires immense computational resources and can take hours, days, or even weeks on high-performance computing (HPC) clusters (Slotnick et al., 2014). While the cost of a single simulation may be justifiable for final design verification, it becomes prohibitive in the context of the modern engineering design cycle. The core challenge is not merely that a single CFD simulation is slow, but that contemporary design and analysis workflows are inherently "multi-evaluation"-or "multi-query"-in nature (Bekemeyer et al., 2025).

Tasks such as design space exploration, aerodynamic shape optimization, uncertainty quantification (UQ), and sensitivity analysis require the evaluation of hundreds, if not thousands, of design variations (Yondo et al., 2018). For example, a gradient-based optimization algorithm may need to iteratively adjust dozens of geometric parameters to maximize lift or minimize drag, with each iteration demanding a new CFD simulation (Kenway et al., 2016). Similarly, a robust UQ analysis might involve propagating uncertainties from manufacturing tolerances or operational conditions through the model, a task that often relies on Monte Carlo methods requiring a vast number of simulations to achieve statistical convergence (Smith, 2014). When each model evaluation involves a computationally expensive CFD run, these essential engineering tasks become computationally intractable (Slotnick et al., 2014). This "multi-evaluation bottleneck" represents a fundamental barrier

to innovation, slowing down the design cycle and limiting the ability of engineers to explore novel concepts or quantify risks effectively.

This research is directly motivated by these challenges, framed within the context of Project 77 - Design Optimization and Experimental Evaluation of Centrifugal Compressors for CO2 and CO2-CH4 Mixtures under Supercritical Conditions of the Research Centre for Greenhouse Gas Innovation (RCGI), a critical technologies for carbon capture, utilization, and storage (CCS) (RCGI, 2023). The path to simulating these complex machines requires a new methodology capable of handling representative physical phenomena—such as compressibility, high Reynolds numbers, turbulence, and complex geometric variations—in a computationally efficient manner.

1.2   Reduced-Order Models (ROMs)

To overcome the multi-evaluation bottleneck, a new paradigm is required, moving away from the direct, repeated use of high-fidelity models. The strategic solution lies in the development of surrogate models, also known as metamodels or response surfaces (Hu et al., 2020). A surrogate model is a computationally inexpensive, approximation of the complex, high-fidelity model. Data-driven surrogates learn the input-output relationship from a limited set of pre-computed high-fidelity simulations, providing near-instantaneous predictions for new design points, effectively replacing the expensive CFD solver within the multi-evaluation loop (Espinosa barcenas et al., 2023; Yondo et al., 2018). This approach transforms an intractable problem into a feasible one, enabling fast design exploration and robust analysis without sacrificing the essential physical insights provided by the original high-fidelity data (Hu et al., 2020).

Among the various classes of surrogate models, data-driven Reduced-Order Models (ROMs) have emerged as a particularly powerful paradigm for high-dimensional physical systems like fluid flows. The central philosophy of ROMs is an "offline-online" computational strategy. In the offline, or "training," stage, a set of computationally expensive, high-fidelity simulations is performed to generate a database of "snapshots" of the system's behavior across a range of parameters. A fracation of this database is then used to train the ROM. Once trained, the ROM can be deployed in the online, or "prediction," stage, where it provides extremely fast approximations for new, unseen parameter inputs. This decouples the high computational cost of data generation from the rapid-query demands of the application.

The construction of a machine learning-based ROM (ML-ROM) for a parametric system typically follows a structured pipeline, which forms the backbone of this dissertation. This pipeline can be conceptualized in four primary stages:

- Data Generation: A Design of Experiments (DoE) is created to strategically sample

the parametric design space. A high-fidelity CFD solver is then run for each sample point to generate a database of high-dimensional solution snapshots.

- Dimensionality Reduction: The immense dimensionality of the snapshot data (often millions of degrees of freedom per snapshot) is reduced to a very low-dimensional latent space. This is typically achieved using techniques like Proper Orthogonal Decomposition (POD), which extracts a small set of dominant, energy-optimal basis functions, or "modes", that capture the essential dynamics or parametric sensitivity of the flow.

- Latent-Space Regression: A machine learning model (the surrogate regressor) is trained to learn the mapping between the low-dimensional input design parameters (e.g., blade angle, Mach number) and the low-dimensional latent-space representation (e.g., the POD mode coefficients) of the flow field.

- Field Reconstruction: During the online phase, the trained regressor predicts the latent-space coefficients for a new set of design parameters. These coefficients are then used to reconstruct the full, high-dimensional flow field through a linear combination of the pre-computed basis functions.

This structured approach allows for the systematic deconstruction of a complex modeling problem into a series of more manageable tasks, each addressable with specialized mathematical and computational tools.

1.3   Thesis Objectives and Contributions

The primary objective of this dissertation is to develop, validate, and analyze a unified, flexible, and robust framework for creating parametric reduced-order models for complex fluid dynamics problems. This work aims to move beyond ad-hoc solutions for specific problems and establish a comprehensive methodology that can be adapted to a wide range of challenges in computational aerodynamics, from internal supersonic flows to transonic turbomachinery. And also to addresses critical gaps in the existing literature.

The research is structured as a progression of case studies with increasing complexity, chosen to be computationally tractable while serving as fundamental steps toward the final objective. The key contributions of this dissertation, which collectively advance the state-of-the-art in data-driven aerodynamic modeling, are as follows:

- **A Unified and Validated Methodological Framework:** Developed a comprehensive, end-to-end computational pipeline for creating parametric ROMs. This reproducible framework integrates geometry parameterization, data generation, advanced mesh processing, POD, and machine learning regression (GPR and ANNs). Its versatility

and robustness were rigorously validated across distinct and challenging flow regimes: a 2D supersonic nozzle with strong shock-wave interactions and a 3D transonic compressor blade.

- **Advanced Model Training and Physical Fidelity:** Introduced novel techniques to enhance the fidelity and trustworthiness of ANN-based ROMs:

  - **A Differentiable, Physics-Informed Loss Function:** Developed a hybrid loss function that incorporates error from the reconstructed physical field directly into the training process. This acts as a physical regularizer, improving the capture of high-gradient features like shock waves without needing explicit PDE constraints.

  - **Model Interpretability:** Employed SHapley Additive exPlanations (SHAP) to provide quantitative insight into the "black-box" models, linking their predictions to fundamental physical principles and building confidence in their results.

- **Enabling Technology for Complex 3D Parametric ROMs:** Introduced and validated a mesh morphing technique based on harmonic mapping. This critical method resolves the fundamental challenge of topological inconsistency in parametric studies, enabling the direct application of POD to complex 3D geometries with varying shapes.

- **Integrated Geometry and Flow Field Prediction:** Developed a novel surrogate model that **simultaneously predicts both the high-dimensional aerodynamic fields and the underlying 3D physical geometry** from a single set of abstract design parameters. This addresses a major gap in the literature and is a significant step towards truly generative design.

- **Systematic Comparative Analysis for Model Selection:** Performed an in-depth, empirical comparison of GPR and ANN regressors. The analysis, supported by rigorous cross-validation and noise robustness studies, provides practical, evidence-based guidelines for selecting the appropriate surrogate model based on dataset size, data quality, and the complexity of the underlying physics.

1.4   Dissertation Outline

This dissertation is structured to guide the reader from foundational concepts to advanced applications and future possibilities.

- **Chapter 2** provides a comprehensive review of the state-of-the-art, establishing the strategic context for this research by aligning it with the challenges of computational

science and identifying specific gaps in the academic literature on reduced-order modeling.

- **Chapter 3** provides a comprehensive review of the theoretical foundations of reduced-order modeling and machine learning as applied to fluid dynamics, establishing the context and key concepts for the work.

- **Chapter 4** presents the first major case study, applying the framework to the parametric reconstruction of 2D supersonic nozzle flows, with a focus on the development of the hybrid loss function and a comparative analysis of surrogate regressors.

- **Chapter 5** presents the second case study, targeting the parametric reconstruction of 3D surface fields and geometry on the NASA Rotor 37, with a focus on the critical role of the mesh morphing technique.

- **Chapter 6** synthesizes the findings from both case studies, summarizes the key contributions of the dissertation, discusses its limitations, and proposes promising directions for future research.

2

Literature Review of Reduced-Order Modeling and Machine Learning for Computational Fluid Dynamics

2.1 The Computational Bottleneck in High-Fidelity Turbomachinery Simulation

The design and analysis of modern turbomachinery, from jet engines to power generation turbines, represent one of the most formidable challenges in computational engineering. The accurate prediction of aerodynamic and aero-thermal performance is paramount for achieving gains in efficiency, reliability, and safety, while reducing fuel consumption and environmental impact (Various authors, 2024). However, the physical environment within these machines is exceptionally complex, characterized by extreme conditions that push the boundaries of simulation capabilities.

2.1.1 The Challenge of Simulating Complex Flow Physics

Turbomachinery flows are inherently three-dimensional, unsteady, and turbulent, often operating in transonic or supersonic regimes (Various authors, 2024). These conditions give rise to a host of complex physical phenomena, including strong shock-wave/boundary-layer interactions, flow separation, secondary flows, and intense turbulence, all of which critically influence component performance and structural integrity (Author; Author, 2024a; Various authors, 2024). The accurate numerical simulation of such phenomena necessitates high-fidelity methods. Techniques like Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) offer detailed, time-resolved predictions of turbulent structures but come at a computational cost that is prohibitive for routine industrial design cycles, where hundreds or thousands of evaluations may be required (Peherstorfer et al., 2018; Various authors, 2024). A single high-fidelity LES of a turbine vane, for example, can demand millions of CPU hours, rendering its use in many-query applications like design optimization infeasible (Peherstorfer et al., 2018).

Consequently, the industry has historically relied on Reynolds-Averaged Navier-Stokes (RANS) simulations as the workhorse for aerodynamic design (Various authors, 2024). While computationally tractable, RANS methods introduce significant model-form uncertainty through the turbulence closure models they employ. These models often fail to accurately predict flows with significant separation, strong unsteadiness, or complex

heat transfer, which are precisely the phenomena that limit turbomachinery performance (Author; Author, 2022; Various authors, 2024). This creates a persistent and challenging trade-off for engineers: the choice between computationally expensive simulations that are accurate but too slow for design exploration, and faster simulations that may lack the predictive fidelity required for high-stakes decisions (Various authors, 2024). This fidelity-cost dilemma forms the central motivation for the development of the advanced modeling techniques discussed in this review.

2.1.2 The "CFD Vision 2030 Study" as a Community Roadmap

The challenges facing the computational fluid dynamics (CFD) community were formally articulated in the landmark "CFD Vision 2030 Study," commissioned by NASA and authored by a multidisciplinary team led by Slotnick et al. (Park et al., 2016; Slotnick et al., 2014). This report provided a knowledge-based forecast of future capability requirements and a strategic research roadmap to achieve a visionary CFD capability by the year 2030 (Author; Author, 2024a; Slotnick et al., 2014). The study identified several critical impediments that were preventing CFD from moving beyond a limited, albeit important, region of the design space to become a truly predictive tool across the full operational envelope (Park et al., 2016).

A key finding highlighted was the **"Geometry and Grid Generation Bottleneck,"** where the process of generating high-quality computational meshes for complex geometries was identified as a major time-consuming and labor-intensive step in the CFD workflow (Various authors, 2024). This bottleneck severely hampers the ability to perform rapid design iterations and parametric studies. Furthermore, the study emphasized the fundamental limitations of existing RANS models in reliably predicting turbulent and separated flows, calling for a paradigm shift in turbulence modeling and simulation methodologies (Park et al., 2016; Zhang et al., 2021). The report envisioned a future where routine use of LES for critical components would be possible, enabled by advances in High-Performance Computing (HPC) and the development of new algorithms designed to exploit exascale architectures (Park et al., 2016; Slotnick et al., 2014). The CFD Vision 2030 Study has since served as a foundational document for the field, motivating much of the research into automated workflows, advanced turbulence modeling, and the data-driven and machine learning techniques that form the core of this review (Various authors, 2024).

2.1.3 The Impact on Engineering Design and Optimization

The computational bottleneck has profound consequences for the engineering design process. The high cost of each simulation directly limits the number of design candidates that can be evaluated, making a comprehensive exploration of the vast, high-dimensional design space of a modern turbomachine component practically impossible (Various authors,

2024). This constraint often forces designers to rely on experience, simplified correlations, and localized, gradient-based optimization, which may only find local optima and overlook more innovative and higher-performing designs (Various authors, 2024).

Furthermore, modern engineering demands not only optimal performance at a single design point but also robust performance across a range of operating conditions and in the presence of real-world uncertainties. Manufacturing and assembly tolerances, for instance, introduce geometric variability that can lead to significant scatter in the aerodynamic performance of blades, potentially causing some components to fail to meet performance requirements (Author; Author, 2023a). Quantifying the impact of such uncertainties requires a probabilistic approach, involving thousands of simulations to propagate the uncertainty through the model—a task that is intractable with conventional CFD (Various authors, 2024). The need to overcome these limitations and enable rapid, robust, and multi-disciplinary design optimization has been the primary driver for the development of Reduced-Order Models (ROMs) and machine learning surrogates.

2.2 Projection-Based Reduced-Order Models as a Foundational Acceleration Strategy

In response to the computational demands of high-fidelity CFD, the first major advancement came from the field of model order reduction. Rather than treating the CFD solver as an unchangeable black box, projection-based ROMs seek to reduce the intrinsic dimensionality of the governing equations themselves. This approach is predicated on the observation that the dynamics of many complex fluid flows, despite being described by equations with millions or billions of degrees of freedom, often evolve on a much lower-dimensional manifold. By identifying this underlying structure, it is possible to create a compact, computationally efficient model that captures the dominant physics of the full-order system.

2.2.1 The Principle and Application of Proper Orthogonal Decomposition

The cornerstone of projection-based model reduction in fluid dynamics is the Proper Orthogonal Decomposition (POD) method, introduced to the field by Lumley and also known as Principal Component Analysis or the Karhunen-Loève expansion (Berkooz et al., 1993; Wang; Shan, 2007). POD is a powerful mathematical technique for extracting an optimal, low-dimensional basis from a high-dimensional dataset. The process begins by collecting a series of "snapshots" of the flow field from high-fidelity simulations, which represent the system's state at different points in time or for different parameter values

(Various authors, 2024). POD then computes a set of orthonormal basis functions, or modes, that are optimal in the sense that they capture the maximum possible amount of kinetic energy (or variance in the data) for any given number of modes (Berkooz et al., 1993,?). The most energetic modes represent the large-scale, coherent structures that dominate the flow dynamics.

By projecting the high-dimensional governing equations (e.g., the Navier-Stokes equations) onto the low-dimensional subspace spanned by the first few POD modes, a ROM is created. This ROM takes the form of a small system of ordinary differential equations (ODEs) that governs the temporal evolution of the modal coefficients, drastically reducing the computational cost of the simulation by orders of magnitude (authors, 2022; Various authors, 2024).

This powerful technique has been successfully applied to a wide range of problems in turbomachinery:

- **Shape Optimization:** In shape optimization, the goal is to find the optimal geometry of a component, such as a turbine blade, that maximizes a performance metric. By parameterizing the blade geometry and generating CFD snapshots for a set of training shapes, a POD-based ROM can be constructed. This ROM can then rapidly predict the flow field and performance for new, unseen geometries within the optimization loop, enabling the exploration of a much larger design space than would be possible with the full-order model alone (Author; Author, 2021; Various authors, 2024).

- **Aeroelasticity and Flutter Prediction:** Aeroelastic phenomena like flutter, which involve a dangerous coupling between aerodynamic forces and structural vibrations, are critical to predict. POD-based ROMs are used to model the unsteady aerodynamics of vibrating blades. Snapshots of the unsteady flow field corresponding to different blade vibration modes and frequencies are used to construct a compact aerodynamic ROM. This model can then be coupled with a structural dynamics model to create a low-order aeroelastic model capable of efficiently predicting flutter boundaries (Bourguet et al., 2011; Geyling et al., 2003; Kelly et al., 2021).

- **Fluid-Structure Interaction (FSI):** For more general FSI problems, such as the flow around rotating bodies, POD can be used to create a monolithic ROM that describes the coupled dynamics of both the fluid and solid domains, enabling efficient reconstruction of the high-dimensional velocity fields (Various authors, 2024).

2.2.2 Advanced Modal Analysis and Inherent Limitations

While POD has been a foundational tool, its standard formulation has limitations that have spurred the development of more advanced techniques. One notable extension is

Spectral Proper Orthogonal Decomposition (SPOD), which combines the spatial optimality of POD with frequency-domain analysis (Author; Author, 2024b; Towne et al., 2018). SPOD produces modes that are coherent in both space and time at a specific frequency, making it exceptionally effective for analyzing unsteady flows with multiple interacting phenomena at different timescales, such as the interaction of incoming wakes with vortex shedding in a turbine cascade (Various authors, 2024).

Despite these advancements, projection-based methods face fundamental challenges. The most significant is the **linearity constraint**: POD constructs a linear subspace to approximate the solution manifold. This approach is highly efficient for flows that are well-described by a few dominant linear modes, but it becomes increasingly inefficient for systems dominated by strong nonlinearities or advection. Phenomena like the convection of vortices or the movement of shock waves are poorly represented by a fixed linear basis, requiring a large number of POD modes to achieve acceptable accuracy and diminishing the computational advantage of the ROM (Various authors, 2024).

Another major challenge is the **"geometry problem."** Applying POD to problems with varying geometries, such as in shape optimization, is non-trivial because the snapshots are defined on different computational meshes. A direct comparison or decomposition is not possible. This necessitates the use of complex and often problem-specific techniques to create a consistent representation, such as mesh morphing to deform a template mesh to new geometries or harmonic mapping to establish a common coordinate system (Various authors, 2024). These techniques add a layer of complexity and can introduce their own sources of error. These inherent limitations of linear, projection-based methods created a clear need for more powerful, nonlinear, and non-intrusive techniques, paving the way for the integration of machine learning.

2.3 The Paradigm Shift to Machine Learning and Data-Driven Surrogates

The limitations of traditional projection-based ROMs, particularly their struggles with strong nonlinearities and their intrusive nature, catalyzed a paradigm shift towards data-driven surrogate modeling using machine learning (ML). This evolution represents a move from reducing the physics-based model itself to learning its input-output behavior directly from data. This approach has given rise to a spectrum of techniques, from classical ML algorithms that augment existing ROM frameworks to sophisticated deep learning architectures that can learn end-to-end flow predictions.

2.3.1 Non-Intrusive Modeling with Classical Machine Learning: The POD+ML Framework

The first and most widely adopted integration of machine learning in this context is the hybrid POD+ML framework. This approach retains the powerful dimensionality reduction capability of POD to extract a low-dimensional basis from CFD snapshots but replaces the complex and intrusive Galerkin projection step with a non-intrusive regression model (Various authors, 2024). In this framework, the ML model is trained to learn the highly nonlinear mapping between the input parameters of the system (e.g., geometric parameters, boundary conditions) and the temporal coefficients of the POD modes. Once trained, the model can predict the modal coefficients for a new set of input parameters, which are then used to reconstruct the full-dimensional flow field via the POD basis. This non-intrusive nature is a significant advantage, as it does not require modification of the source code of the CFD solver (Various authors, 2024).

Two classical ML methods have proven particularly effective in this framework:

- **Gaussian Process Regression (GPR):** GPR is a non-parametric, Bayesian regression method. Its key advantage in engineering applications is its ability to provide not only a point prediction but also a measure of predictive uncertainty (a variance) (Various authors, 2024). This uncertainty quantification is invaluable for robust design and optimization, as it can be used to guide the selection of new simulation points in an active learning strategy, focusing computational effort where the model is least certain. However, the computational cost of standard GPR scales cubically with the number of training data points, which can become a bottleneck for problems with many training samples (Various authors, 2024).

- **Artificial Neural Networks (ANNs):** ANNs, particularly multilayer perceptrons (MLPs), are powerful universal function approximators capable of capturing extremely complex nonlinear relationships between inputs and outputs (Various authors, 2024). Studies comparing GPR and ANNs for nozzle flow reconstruction have shown that while GPR may achieve lower error on larger datasets, ANNs can demonstrate greater stability and better performance in data-scarce regimes (Various authors, 2024). However, their "black-box" nature can make them more difficult to interpret without specialized techniques.

2.3.2 Deep Learning Architectures for End-to-End Flow Field Reconstruction

While the POD+ML framework proved highly successful, recent research has pushed towards end-to-end deep learning models that learn the entire prediction pipeline, often bypassing the need for a linear basis like POD altogether. These advanced architectures

are capable of handling higher-dimensional data and capturing more complex physical phenomena directly. A comparative overview of these architectures is presented in Table 2.1.

- **Autoencoders (AEs) and Variational Autoencoders (VAEs):** These are unsupervised neural networks designed for nonlinear dimensionality reduction. An autoencoder consists of an "encoder" network that compresses the high-dimensional input data (e.g., a pressure field) into a low-dimensional latent space, and a "decoder" network that reconstructs the original data from this compressed representation (Various authors, 2024). For highly nonlinear phenomena, such as the motion of shock waves in unsteady turbine flows, the nonlinear manifold learned by a VAE can provide a much more compact and efficient representation than the linear subspace of POD. These architectures are often paired with Recurrent Neural Networks (RNNs), such as Gated Recurrent Units (GRUs) or Long Short-Term Memory (LSTM) networks, which are designed to model the temporal evolution of the latent space variables for unsteady predictions (Various authors, 2024).

- **Graph Neural Networks (GNNs):** The application of GNNs represents a significant breakthrough for CFD. Traditional deep learning models like CNNs require data to be on a structured, Euclidean grid (like an image). This is a major limitation for turbomachinery, which involves complex geometries that are typically discretized using unstructured meshes. GNNs overcome this by operating directly on graph-structured data (Lino et al., 2022; Pfaff et al., 2020). By treating the nodes of an unstructured mesh as nodes in a graph and the edges as connections between them, GNNs can learn to approximate the PDE operators through message-passing between neighboring nodes. This makes them inherently mesh-independent and allows them to operate directly on the native output of CFD solvers, directly addressing the "Geometry and Grid Generation Bottleneck" (Various authors, 2024).

- **Transformer Architectures:** Originally developed for natural language processing, Transformer models are based on an "attention" mechanism that allows them to weigh the importance of different parts of an input sequence. In the context of CFD, this architecture is being explored for its ability to learn long-range spatial dependencies in flow fields. Promising results show that Transformers can approximate 3D flows for new compressor designs without being tied to the specific topology of the training grid, offering another path towards flexible, mesh-agnostic models (Various authors, 2024).

Table 2.1: Comparative Overview of Deep Learning Architectures for CFD.

Architecture	Core Principle	Key Strengths for Turbomachinery CFD	Identified Challenges
ANN/MLP	Universal function approximation.	Effective for regression tasks in POD+ML frameworks; computationally simpler than deep architectures.	Can be a "black-box"; may require large datasets; may struggle with high-dimensional spatial data.
CNN	Spatial feature learning via convolutional filters.	Excellent at capturing spatial correlations in structured grid data (e.g., pressure/velocity fields mapped to images).	Limited to structured grids, reducing flexibility for complex geometries; requires consistent mesh topology.
AE/VAE	Unsupervised non-linear dimensionality reduction.	Superior to POD for compressing highly nonlinear data (e.g., moving shocks); learns a compact latent space.	Training complexity; potential for information loss during compression/decompression; can blur sharp features.
GNN	Operates on graph structures (unstructured meshes).	Handles complex, irregular geometries directly; mesh-independent, addressing a key CFD bottleneck.	Scalability to very large industrial meshes is an active research area; higher training complexity than CNNs.
RNN/GRU/LSTM	Models sequential/time-series data.	Explicitly designed to capture unsteady dynamics and temporal evolution of flow fields.	Long-term prediction stability can be an issue; risk of vanishing/exploding gradients.
PINN	PDE residuals in the loss function.	Enforces physical laws; reduces dependency on large labeled datasets; improves generalization and physical consistency.	Balancing gradients of multiple loss terms is challenging; implementation is more complex than standard supervised learning.

2.3.3 Embedding Physical Laws: The Rise of Physics-Informed Neural Networks (PINNs)

A critical examination of purely data-driven models reveals a significant vulnerability: without any knowledge of the underlying physics, they can produce physically implausible

results, especially when extrapolating beyond the domain of the training data (Various authors, 2024). This limitation hinders their adoption in safety-critical engineering applications. Physics-Informed Neural Networks (PINNs), pioneered by Raissi, Perdikaris, and Karniadakis, offer a powerful solution to this problem by embedding the governing physical laws directly into the learning process (Mifsud, 2016; Raissi et al., 2019).

The core concept of a PINN is to modify the loss function of the neural network. In addition to the standard data-mismatch term that penalizes deviations from the training data, a second term is added that penalizes the extent to which the network's output violates the governing PDEs (e.g., the Navier-Stokes equations) (Various authors, 2024). The network's derivatives with respect to its inputs (space and time) are calculated using automatic differentiation, a technique inherent to modern deep learning frameworks. The network is then trained to minimize a composite loss function that drives both the data error and the PDE residuals to zero.

This integration of physics has profound benefits. It acts as a powerful regularizer, preventing overfitting and dramatically improving the model's ability to generalize to new, unseen data (Author; Author, 2024c; **?**). Furthermore, because the physics provides a strong inductive bias, PINNs can often be trained with significantly less labeled data than their purely data-driven counterparts, making them highly effective in data-sparse regimes (Various authors, 2024). Applications in turbomachinery, such as predicting the flow field in a compressor cascade, have shown that PINNs can achieve high accuracy even with incomplete boundary conditions and can successfully solve both forward and inverse problems (Various authors, 2024). However, the methodology is not without challenges. A key difficulty lies in balancing the gradients from the different terms in the loss function (data loss vs. various PDE residual losses), which can lead to training instabilities. This has prompted research into adaptive learning rate strategies and novel network architectures to improve convergence and robustness (Various authors, 2024).

The evolution from physics-based reduction to purely data-driven models and now to physics-informed learning demonstrates a maturing understanding within the field. The most advanced and promising approaches are not those that seek to replace physics with data, but those that leverage data to solve the governing physical equations more efficiently and robustly.

2.4   Integrated Frameworks for Engineering Applications

The true value of the individual methods discussed—POD, ML surrogates, and PINNs—is realized when they are integrated into comprehensive frameworks designed to solve end-to-end engineering problems. These frameworks are not just about accelerating a single simulation; they aim to transform the entire design and analysis workflow, from initial concept exploration to robust optimization and in-service health monitoring.

2.4.1 Multi-Fidelity Approaches for Enhanced Accuracy and Efficiency

One of the most powerful integrated frameworks is multi-fidelity modeling, which strategically combines simulations of varying accuracy and cost to build a surrogate model that is both fast and accurate (Peherstorfer et al., 2018; Wikipedia, 2024). The premise is that while high-fidelity data (e.g., from LES) is accurate, it is too expensive to generate in large quantities. Conversely, low-fidelity data (e.g., from RANS) is cheap but may be inaccurate. A multi-fidelity model leverages a large number of low-fidelity samples to capture the general trends in the design space and a small, strategically selected set of high-fidelity samples to correct for the low-fidelity model's bias and error (Peherstorfer et al., 2018).

In the context of turbomachinery, this approach has been successfully applied to the shape optimization of high-pressure turbine vanes (Various authors, 2024). In this application, a multi-fidelity co-Kriging model was constructed using 56 low-fidelity RANS simulations and only 7 high-fidelity LES simulations. The resulting surrogate model was able to accurately predict turbine losses, capturing key flow mechanisms that the RANS model alone missed, while operating within a computationally feasible budget (Various authors, 2024). This demonstrates that by intelligently fusing information from different sources, it is possible to achieve the accuracy of high-fidelity methods at a fraction of the cost, making the optimization of complex, high-Reynolds number flows tractable (Various authors, 2024).

2.4.2 Surrogate-Assisted Optimization and Inverse Design

The primary application for these fast-running surrogate models is to drive multi-objective design optimization. In a surrogate-assisted optimization framework, the expensive CFD solver inside an optimization loop is replaced by the trained ROM or ML model (Various authors, 2024). This allows optimization algorithms, such as the Non-dominated Sorting Genetic Algorithm II (NSGA-II), to evaluate thousands or even millions of design candidates to map out the Pareto front—the set of optimal designs representing the best possible trade-offs between competing objectives (e.g., maximizing efficiency while minimizing pressure loss) (Various authors, 2024).

This approach has been successfully demonstrated in various turbomachinery applications. For the optimization of centrifugal pump inducers and Savonius wind turbines, a framework integrating CFD, ANNs, and a modified NSGA-II achieved significant performance improvements, increasing the head coefficient by 14.3% in the pump and the power coefficient by 5.32% in the turbine (Various authors, 2024). Similarly, a modular framework for impeller design, combining CFD and FEM simulations with AI-based surrogates, achieved hydraulic efficiency gains of up to 6.7% and mass reductions of over 30% (Various authors, 2024). These frameworks can also be used for inverse design, where a desired

performance target is specified, and the model works backward to generate the geometric parameters required to achieve it, further accelerating the design process (Author; Author, 2024d).

2.4.3   Addressing Uncertainty and Interpretability

Beyond performance optimization, integrated frameworks are crucial for addressing two of the most pressing issues in modern engineering: uncertainty and trust.

- **Uncertainty Quantification (UQ):** Real-world turbomachinery components are subject to variability from manufacturing and assembly processes, which introduces geometric uncertainty. These small deviations can lead to significant scatter in aerodynamic performance. Quantifying this impact is a classic UQ problem that is computationally intractable with full-order CFD. By using a fast surrogate model, Monte Carlo simulations can be performed to propagate the statistical distribution of geometric deviations through the model, yielding a statistical distribution of the performance outputs (Various authors, 2024). This allows designers to create designs that are not just optimal in a deterministic sense, but also robust to real-world manufacturing tolerances, a critical requirement for reliable production (Author; Author, 2023a).

- **Interpretability:** A major barrier to the adoption of complex ML models in engineering is their "black-box" nature. If a model predicts a certain design is optimal, but the engineer cannot understand *why*, it is difficult to trust the result. To address this, interpretability techniques like SHAP (SHapley Additive exPlanations) are being integrated into the workflow (Lundberg; Lee, 2017; Molnar, 2022). SHAP is a game theory-based method that assigns an importance value to each input feature for a given prediction, effectively explaining how much each design parameter contributed to the final performance outcome (DataCamp, 2024; Lundberg; Lee, 2017; Molnar, 2022). In the analysis of nozzle flows, SHAP was used to confirm that total pressure was the most influential input feature, aligning the model's behavior with physical intuition and thereby increasing trust in its predictions (Various authors, 2024).

The development of these integrated frameworks signifies a maturation of the field, moving beyond the creation of faster solvers to a fundamental rethinking of the engineering design process itself. The shift is from a slow, sequential, "design-then-validate" workflow to a rapid, iterative, "explore-and-optimize-for-robustness" paradigm, enabled entirely by the fusion of physics-based simulation, model reduction, and machine learning.

2.5 Synthesis, Persistent Challenges, and Future Outlook

The body of research reviewed here charts a clear and compelling trajectory in the field of computational science for turbomachinery. It documents a multi-decade effort to overcome the fundamental limitations of high-fidelity simulation, evolving from physics-based model reduction to sophisticated, physics-informed machine learning frameworks. This evolution is not merely a sequence of methodological improvements but reflects a deeper conceptual shift in how the relationship between physical laws, numerical simulation, and empirical data is understood and leveraged.

2.5.1 Synthesis of Methodological Evolution

The journey began with the direct application of CFD, which, despite its power, was quickly identified as a computational bottleneck, a challenge formally articulated in the CFD Vision 2030 Study (Park et al., 2016; Slotnick et al., 2014). The first major response was the development of projection-based ROMs, with Proper Orthogonal Decomposition (POD) as the central tool. This represented a physics-based acceleration strategy, reducing the dimensionality of the governing equations themselves (Various authors, 2024). However, the inherent limitations of POD—its linear nature and difficulties with geometric parameterization—created a clear need for more powerful, nonlinear techniques.

This need was met by the paradigm shift to machine learning. The initial step was the non-intrusive POD+ML framework, which retained the physics-based basis of POD but replaced intrusive projection with flexible, data-driven regression models like GPR and ANNs (Various authors, 2024). The field then rapidly advanced into end-to-end deep learning, with specialized architectures like Autoencoders, GNNs, and RNNs being developed to tackle nonlinear dimensionality reduction, unstructured meshes, and unsteady dynamics, respectively (Various authors, 2024). The realization that purely data-driven models could lack physical consistency and generalize poorly led to the current frontier: Physics-Informed Neural Networks (PINNs). PINNs represent a synthesis of the two preceding paradigms, embedding physical laws directly into the learning process to create models that are both data-efficient and physically robust (Various authors, 2024). This trajectory from physics-first, to data-first, and finally to physics-informed data-driven modeling encapsulates the maturation of the field.

2.5.2 Persistent Challenges Across the Literature

Despite the remarkable progress, a critical review of the literature reveals several persistent challenges that continue to be active areas of research and represent significant hurdles to the widespread industrial adoption of these techniques.

- **Generalization:** The single most critical challenge, mentioned across numerous studies, is model generalization. A surrogate model is only useful if it can accurately predict outcomes for new, out-of-distribution data—geometries, operating conditions, or flow regimes it has not seen during training. Many models, especially purely data-driven ones, can perform well on interpolation tasks but fail dramatically when asked to extrapolate, sometimes producing unphysical results (Various authors, 2024). Achieving true generalization remains the "holy grail" for ML in scientific computing (Author; Author, 2023b).

- **Data Scarcity:** While techniques like PINNs reduce data dependency, the performance of most advanced deep learning models is still fundamentally tied to the quantity and quality of high-fidelity training data. Generating these datasets remains the most expensive part of the workflow, and in many industrial contexts, data is sparse and expensive to acquire (Various authors, 2024).

- **Handling of Discontinuities:** A recurring technical challenge is the accurate representation of sharp gradients and discontinuities, such as shock waves or contact interfaces. Many model reduction and deep learning techniques, particularly those based on modal decomposition or standard convolutional layers, tend to blur these sharp features, leading to significant local errors even if global error metrics are low (Various authors, 2024). This is a critical issue in transonic and supersonic turbomachinery applications.

- **Scalability and Industrial Deployment:** There remains a significant gap between the scale of problems tackled in academic research and the scale of those encountered in industrial practice. Training GNNs on meshes with billions of cells, ensuring the long-term stability of recurrent models for thousands of time steps, and integrating these complex software pipelines into established industrial design workflows are all non-trivial challenges that require further research and engineering (Various authors, 2024).

2.5.3   Future Outlook: Towards 2030 and Beyond

The future of computational design for turbomachinery will likely be defined by efforts to address these persistent challenges through a deeper integration of physics, data, and learning. Several emerging trends point the way forward. The distinction between "model reduction," "surrogate modeling," and "machine learning" is becoming increasingly blurred. An autoencoder is a form of nonlinear model reduction; a GNN is a surrogate that learns PDE operators; a PINN can be viewed as a novel type of numerical solver. This convergence suggests the formation of a unified field of Scientific Machine Learning, where these concepts are seamlessly integrated.

Looking ahead, the concept of **foundation models**, which has revolutionized natural language processing and computer vision, holds immense potential for physical systems. One can envision large-scale models pre-trained on vast and diverse datasets of physical simulations, capturing a general understanding of fluid dynamics. These models could then be rapidly fine-tuned with a small amount of specific data for a particular application, drastically reducing the data generation cost for new problems (Author; Author, 2023b).

Ultimately, the path forward is one of **hybridization**. The future is unlikely to be purely data-driven or purely physics-based, but rather a sophisticated fusion of the two. This includes developing new neural network architectures with inherent physical symmetries and conservation laws built-in, creating more powerful multi-fidelity frameworks that can fuse data from simulations, experiments, and in-service sensors, and integrating these tools into autonomous, AI-driven design and discovery platforms. While significant progress has been made since the publication of the CFD Vision 2030 Study, its ultimate goal—a robust, reliable, and fully autonomous predictive capability for the entire aerospace design envelope—remains a grand challenge. The innovative methods reviewed here represent critical steps along that path, promising to continue transforming the field of turbomachinery design and analysis for the next decade and beyond.

3

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

3.1 The Imperative for Model Reduction in Computational Aerodynamics

The governing equations of fluid motion, the Navier-Stokes equations, are a set of coupled, nonlinear partial differential equations whose analytical solutions are unknown for all but the simplest cases (Berkooz et al., 1993). Their numerical solution in a discretized domain representing a complex engineering geometry results in a system with an enormous number of degrees of freedom, often on the order of millions or billions. When considering parametric studies, where design variables such as geometry or boundary conditions are varied, the dimensionality of the problem space explodes further. This leads to the well-known "curse of dimensionality," where the number of samples required to adequately explore a design space grows exponentially with the number of parameters. A brute-force approach, involving the gridding of this high-dimensional parameter space and running a full-order simulation at each point, is computationally infeasible (Mifsud, 2016).

This reality underscores the fundamental imperative for model reduction (Freund, 2003; Zhang et al., 2021). The challenge is not merely one of computational cost, but one of the inherent complexity and high dimensionality of the solution manifold—the set of all possible solutions as the input parameters vary. The central hypothesis of reduced-order modeling is that despite the high dimensionality of the discretized system, the essential dynamics of many physical systems evolve on or near a much lower-dimensional, intrinsic manifold embedded within the high-dimensional state space (Carlberg et al., 2018; Geelen et al., 2022). A successful ROM, therefore, is one that can discover and exploit this low-dimensional structure (Mifsud, 2016). The task of a ROM is not just to compress data, but to identify the underlying coherent patterns and structures that govern the system's behavior, thereby creating a compact, physically meaningful, and computationally tractable representation of the original complex system (Freund, 2003).

3.2 Dimensionality Reduction for Fluid Flows: Proper Orthogonal Decomposition (POD)

Proper Orthogonal Decomposition (POD) is arguably the most established and widely used technique for dimensionality reduction in fluid dynamics (Berkooz et al., 1993; Mifsud, 2016). Also known in other fields as Principal Component Analysis (PCA) or the Karhunen-Loève expansion, POD provides a systematic and optimal method for finding a linear basis to represent a high-dimensional dataset (Cazemier et al., 1998).

3.2.1 Theoretical Formulation

The application of POD in CFD typically begins with the "method of snapshots," a technique pioneered for fluid dynamics applications by Sirovich (Cazemier et al., 1998). A set of N_s high-fidelity simulation results, or "snapshots," are collected from simulations at different points in the parameter space. Each snapshot, representing a full flow field, is reshaped into a single column vector $u_j \in \mathbb{R}^{N_g}$, where N_g is the number of grid points (or degrees of freedom) in the mesh. These snapshot vectors are then assembled into a snapshot matrix $S = [u_1, u_2, ..., u_{N_s}] \in \mathbb{R}^{N_g \times N_s}$.

POD seeks a set of orthonormal basis vectors, or "modes," $\{\phi_k\}_{k=1}^r$, that are optimal in the sense that they maximize the projection of the snapshot data onto them. Mathematically, this is equivalent to finding the basis that minimizes the mean squared error of the projection for any given rank of approximation. The solution to this optimization problem is found by solving the eigenvalue problem of the data covariance matrix, $C = SS^T$. However, since the number of grid points N_g is typically much larger than the number of snapshots N_s, it is computationally more efficient to solve the eigenvalue problem for the smaller matrix $K = S^T S \in \mathbb{R}^{N_s \times N_s}$. The POD modes ϕ_k are then recovered from the eigenvectors of K. In practice, this entire procedure is most robustly and efficiently performed using the Singular Value Decomposition (SVD) of the snapshot matrix, $S = U\Sigma V^T$, where the columns of the matrix U are the desired POD modes ϕ_k (Steyert, 2022).

3.2.2 The "Energy-Optimal" Basis

A key property of POD is that the modes form an "energy-optimal" basis. The singular values σ_k (the diagonal entries of Σ) are related to the eigenvalues of the covariance matrix and represent the "energy" (or variance in a statistical sense) captured by each corresponding mode ϕ_k. The modes are ordered hierarchically, such that the first mode ϕ_1 captures the most energy, ϕ_2 captures the most of the remaining energy, and so on (Cazemier et al., 1998). This hierarchy allows for a highly efficient low-rank approximation of the original data. Any snapshot u_j can be approximated as a linear combination of the

first M modes:

$$u_j \approx \sum_{k=1}^{M} a_{jk} \phi_k$$

where $M \ll N_s$, and the coefficients $a_{jk} = u_j^T \phi_k$ are the projections of the snapshot onto the modes. By retaining only a small number of modes that capture a vast majority of the system's energy (e.g., 99.9%), the dimensionality of the problem is dramatically reduced from N_g to M.

3.2.3 Implementation and Preprocessing

Before applying POD, raw snapshot data must be preprocessed to ensure numerical stability and physical relevance. Two steps are crucial. First, the data is typically mean-centered by subtracting the ensemble-averaged field, $\bar{u} = \frac{1}{N_s} \sum_{j=1}^{N_s} u_j$, from each snapshot. POD is then performed on the fluctuation fields, $u_j' = u_j - \bar{u}$. This separates the mean behavior from the dynamic variations, which are often the primary interest of the model. Second, data scaling, for instance using a MinMaxScaler to bring all values into a uniform range like $[0, 1]$, is often applied. This prevents fields with large physical magnitudes (like pressure) from dominating the variance calculation over fields with smaller magnitudes (like temperature or Mach number), ensuring an equitable contribution from all physical quantities to the final modes.

3.2.4 Limitations and the Motivation for Nonlinear Methods

The power of POD lies in its simplicity and optimality for linear systems. However, its linearity is also its most fundamental weakness (Carlberg et al., 2018). POD finds the optimal *linear subspace* on which to project the data. Many important fluid dynamics phenomena, particularly those dominated by advection or featuring moving discontinuities like shock waves, are better described as evolving on a *nonlinear manifold* (Geelen et al., 2022). Projecting a curved manifold onto a flat, linear subspace is an inherently inefficient representation. It often requires a large number of POD modes to accurately capture the nonlinear dynamics, diminishing the benefits of the model reduction.

This limitation has been a major driver of recent research, motivating the exploration of nonlinear dimensionality reduction techniques, chief among them methods based on deep learning such as autoencoders (Carlberg et al., 2018; Fonolla et al., 2023; Geelen et al., 2022; Steyert, 2022). An autoencoder uses a neural network (the encoder) to learn a nonlinear mapping from the high-dimensional input space to a low-dimensional latent space, and another network (the decoder) to map back. By training the network to minimize reconstruction error, the autoencoder can learn a compact, nonlinear representation that is potentially far more efficient than the linear subspace found by POD (Fonolla et al., 2023). While this thesis focuses on the robust application of POD, it acknowledges this

trajectory and provides a foundational framework upon which such future advancements can be built.

3.3 Surrogate Modeling for Latent Space Dynamics

Once dimensionality reduction has been performed, the original problem of mapping high-dimensional inputs to high-dimensional outputs is transformed into a more tractable one: mapping low-dimensional inputs (the design parameters) to a low-dimensional latent space (the POD coefficients). This is a classical regression problem, for which a variety of machine learning techniques, often called surrogate models or metamodels, can be employed (authors, 2022; Jain et al., 2025; Kelly et al., 2021; Mathworks, 2024; San et al., 2022; Wang; Shan, 2007). This work focuses on two of the most powerful and widely used methods: Gaussian Process Regression and Artificial Neural Networks.

3.3.1 Gaussian Process Regression (GPR)

Gaussian Process Regression is a non-parametric, Bayesian regression method that has gained significant popularity in engineering and machine learning (Wang; Shan, 2007). Unlike parametric models that fit a single function to the data, GPR defines a probability distribution over a space of functions. A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution. It is fully specified by a mean function and a covariance function, or kernel.

The kernel is the heart of a GPR model. It encodes the assumptions about the function being modeled, such as its smoothness and correlation structure. A common choice, used in this work, is the Radial Basis Function (RBF) kernel, also known as the squared exponential kernel. This kernel assumes that input points that are "close" in the parameter space will have similar output values, with the notion of "closeness" controlled by a length-scale hyperparameter.

The key advantages of GPR are twofold. First, as a Bayesian method, it provides not only a point prediction (the posterior mean) but also a measure of predictive uncertainty (the posterior variance). This is invaluable in engineering design, as it allows for principled uncertainty quantification and risk assessment (Wang; Shan, 2007). Second, the Bayesian framework provides a natural defense against overfitting, making GPR particularly robust for problems with small or sparse training datasets (authors, 2022). However, GPR is not without its drawbacks. The primary limitation is its computational complexity, which scales as $O(N^3)$ with the number of training points N, due to the need to invert the covariance matrix. This makes standard GPR computationally challenging for very large datasets.

3.3.2 Artificial Neural Networks (ANNs)

Artificial Neural Networks, and specifically the Multi-Layer Perceptron (MLP), are powerful function approximators inspired by the structure of the biological brain. The universal approximation theorem states that a feedforward network with a single hidden layer containing a finite number of neurons can approximate any continuous function to an arbitrary degree of accuracy, given enough neurons. This makes ANNs an extremely flexible and powerful tool for regression (Mifsud, 2016).

An ANN consists of interconnected layers of nodes, or "neurons." Each neuron performs a weighted sum of its inputs, adds a bias, and then passes the result through a nonlinear activation function (e.g., sigmoid, ReLU, tanh). By stacking multiple layers, ANNs can learn hierarchical representations of data and model highly complex, nonlinear relationships.

The primary advantages of ANNs are their scalability to very large datasets and their unparalleled flexibility in modeling complex functions (Jain et al., 2025). Unlike GPR, their training time can scale more favorably with the number of samples, especially when using modern GPU hardware and stochastic gradient descent-based optimizers. However, this flexibility comes at a cost. ANNs are often considered "black-box" models, making their internal reasoning difficult to interpret (Wang; Shan, 2007). They are also prone to overfitting if not properly regularized (e.g., using techniques like dropout or weight decay), and they typically have a large number of hyperparameters (e.g., number of layers, number of neurons, learning rate) that require careful and often extensive tuning to achieve optimal performance.

The choice between GPR and ANNs represents a fundamental trade-off in surrogate modeling. GPR offers probabilistic rigor, built-in uncertainty quantification, and robustness on small datasets, making it an excellent choice for well-posed problems where quantifying uncertainty is paramount. ANNs, on the other hand, offer scalable flexibility and the raw power to model extremely complex, high-dimensional, and noisy phenomena, provided that sufficient data and computational resources are available for training and tuning. As demonstrated in the comparative studies of this dissertation, the optimal choice depends on the specific characteristics of the problem at hand, including the size and quality of the available data and the nature of the underlying physics (authors, 2022; Wang; Shan, 2007).

3.4 Toward Physically Consistent and Interpretable Models

As machine learning models become more integrated into safety-critical engineering workflows, two challenges have come to the forefront of the research community: physical consistency and interpretability. A model that is purely data-driven, with no knowledge of the underlying physics, may produce predictions that are highly accurate on average but violate fundamental physical laws, making them unreliable for engineering decisions.

Similarly, a "black-box" model that provides accurate predictions without any explanation of its reasoning is difficult to trust, debug, or certify for use in critical applications (Wang; Shan, 2007).

3.4.1 The Rise of Physics-Informed Machine Learning (PIML)

To address the first challenge, the field of Physics-Informed Machine Learning (PIML) has emerged as a major area of research (Raissi et al., 2019). The core idea of PIML is to embed physical domain knowledge, often in the form of governing partial differential equations (PDEs), directly into the machine learning algorithm. The seminal work in this area introduced Physics-Informed Neural Networks (PINNs), where the residuals of the governing PDEs are added as a penalty term in the model's loss function (Raissi et al., 2017, 2019). This forces the model to learn not only from the data but also to satisfy the physical constraints, leading to better data efficiency, improved generalization, and more physically plausible solutions (Raissi et al., 2017, 2019). The work in this dissertation aligns with the PIML philosophy through a practical, "soft" implementation. The novel hybrid loss function introduced for training ANN surrogates includes a term, $L_{\text{reconstructed}}$, which directly measures the prediction error in the reconstructed physical space. By minimizing this term, the training process implicitly pushes the model to generate solutions that are physically accurate, as any significant deviation from the true physical field would incur a large penalty. This approach bridges the gap between the abstract latent space, where the model operates, and the physical space, which is the domain of interest for the engineer, offering a pragmatic route to physical consistency (Chen et al., 2023).

3.4.2 The Interpretability Imperative

To address the second challenge of trust, methods for model interpretability have become essential. It is no longer sufficient for a model to be accurate; it must also be explainable (Lundberg; Lee, 2017; Molnar, 2022). Techniques like SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee, provide a powerful, theoretically grounded framework for peering inside the black box (Lundberg; Lee, 2017; Molnar, 2022). Based on principles from cooperative game theory, SHAP assigns a quantitative "importance value" to each input feature, representing its contribution to a specific prediction (Molnar, 2022). This allows one to understand which factors the model is weighing most heavily in its decision-making process.

The application of SHAP analysis in this dissertation is a direct response to this interpretability imperative. By analyzing the feature importances of the trained surrogate models, it is possible to verify that their behavior aligns with physical intuition. The growing use of SHAP to interpret complex CFD and ML models demonstrates its value in the field (Crispo et al., 2023; Lundberg; Lee, 2017; Molnar, 2022). For instance, finding

that the model identifies inlet total pressure as the most critical parameter for nozzle flow builds confidence that the model has learned a physically meaningful relationship, rather than just spurious correlations in the data. This demonstrates a commitment not just to applying machine learning as a black-box tool, but to actively engaging with the critical and contemporary conversations in scientific machine learning surrounding the pursuit of physical consistency and the establishment of trust through interpretability.

4

A Unified Framework for Parametric Field Reconstruction

This chapter formalizes the comprehensive, end-to-end computational framework developed in this dissertation. The framework is designed to be modular and adaptable, providing a systematic pipeline for constructing parametric reduced-order models for a wide range of aerodynamic problems. The methodology is presented as a sequence of five stages, moving from initial problem definition and data generation to the final validation of a trained surrogate model. The subsequent chapters will demonstrate the application of this unified framework to two distinct and challenging case studies.

4.1 Overview of the End-to-End Pipeline

The proposed framework integrates several advanced computational techniques into a cohesive workflow. The process begins with the parameterization of a baseline geometry and the use of a Design of Experiments (DoE) methodology to sample the design space. High-fidelity CFD simulations are performed at these sample points to generate a snapshot database. A critical and innovative step, particularly for 3D applications, is the use of a mesh morphing pipeline to harmonize the topology of the disparate snapshot meshes, making them suitable for Proper Orthogonal Decomposition (POD). POD is then used to extract a low-dimensional basis from the high-dimensional field data. Finally, a machine learning regressor—either a Gaussian Process (GPR) or an Artificial Neural Network (ANN)—is trained to map the input design parameters to the low-dimensional POD coefficients. The trained model can then be used for rapid prediction of the full aerodynamic field for any new design within the parameterized space. This entire process is designed with reproducibility and rigor in mind, incorporating advanced strategies for model tuning and validation.

4.2 Stage 1: Parametric Data Generation

The foundation of any data-driven model is the data itself. The quality, quantity, and diversity of the training data directly determine the accuracy and robustness of the final surrogate model. This first stage focuses on the systematic generation of a high-quality snapshot database.

4.2.1 Geometry Parameterization

The process begins by defining a parametric design space. This involves identifying the key variables that control the geometry of the object under study and defining their range of variation. For aerodynamic applications, these variables often relate to the shape of an airfoil or blade. Techniques such as using control points to define Bézier splines or other curves are commonly employed to smoothly and intuitively vary features like blade thickness distributions, camber lines, or angle of attack schedules along the span. In the NASA Rotor 37 case study, for example, 28 design variables were used to control the blade angle and thickness distributions at the hub and shroud sections, fully defining the parametric space for the optimization study.

4.2.2 Design of Experiments (DoE)

Once the parameter space is defined, it must be sampled efficiently to select the points at which expensive high-fidelity simulations will be run. Given the high dimensionality of many design spaces, a simple gridding approach is infeasible. Instead, advanced DoE techniques are used to ensure that the sample points provide maximum information with a minimum of computational effort. Latin Hypercube Sampling (LHS) is a particularly effective statistical method for this purpose. LHS generates a set of sample points that are well-distributed across the parameter space, avoiding clustering and ensuring that the full range of each parameter is explored. This space-filling property is crucial for training a surrogate model that can generalize well to unseen regions of the design space.

4.2.3 High-Fidelity Simulation

With the sample points defined by the DoE, the final step in data generation is to execute the full-order numerical model (e.g., a RANS CFD solver) for each point. This is the most computationally expensive part of the entire offline phase. Each simulation produces a "snapshot" of the steady-state solution, which includes the full-field data (e.g., pressure, temperature, velocity) across the entire computational mesh. The collection of all snapshots from the DoE constitutes the raw database that will be used to train the reduced-order model.

4.3 Stage 2: Topological Harmonization via Mesh Morphing

A fundamental prerequisite for the application of Proper Orthogonal Decomposition is that all snapshots must reside on a common grid with a consistent topology. This means that the number of points and their connectivity must be identical across all snapshots, such that the i-th entry in each snapshot vector corresponds to the same spatial location or

entity. In parametric CFD studies, where the geometry is altered for each design point, this condition is inherently violated, as each simulation is performed on a unique, individually generated mesh. This topological inconsistency poses a major barrier to the application of POD for parametric ROMs, especially for complex 3D shapes.

This pipeline directly solves what the CFD Vision 2030 study identifies as the "Geometry and Grid Generation Bottleneck". By creating a consistent data structure for a family of geometric variations, it enables a direct parametric mapping from design variables to flow solutions, bypassing the need to remesh for each new design within an optimization loop. This is a fundamental contribution to automating the design-analysis cycle.

To overcome this critical challenge, this framework incorporates a sophisticated mesh morphing pipeline. This pipeline acts as a preprocessing stage that transforms a collection of topologically inconsistent surface meshes into a set of regularized meshes that share a common structure. This process is not merely a data-cleaning step; it is the fundamental enabling technology that makes the subsequent application of POD possible for parametric 3D geometries. The process consists of three sequential steps:

- Harmonic Mapping (3D-to-2D Projection): The first step is to create a common reference frame for all the irregular 3D surface meshes. This is achieved by mapping each 3D mesh onto a canonical 2D parametric domain, such as a unit square $\times$. This is done by computing a harmonic map, Φ, which is the solution to Laplace's equation ($\Delta\Phi = 0$) with Dirichlet boundary conditions that map the boundary of the 3D mesh to the boundary of the 2D square. The solution to this elliptic PDE results in a smooth, one-to-one "flattening" of the 3D surface onto the 2D plane, preserving local angles and minimizing distortion.

- Interpolation onto a Regular Grid: With all snapshots now represented as unstructured triangular meshes within the same 2D parametric domain, the associated scalar fields (e.g., pressure, temperature) and the original 3D coordinates (X,Y,Z) are interpolated onto a new, common, structured grid (e.g., a uniform 100 by 100 grid). This interpolation is typically performed using barycentric coordinates within the triangles of the 2D mapped mesh. This step effectively resamples all the field data from their original, inconsistent meshes onto a single, shared grid structure.

- 3D Reconstruction (Lifting): The final step is to "lift" the new structured grid, now populated with the interpolated field data and 3D coordinates, back into physical 3D space. This creates a new, regularized 3D surface mesh that is topologically identical for every snapshot in the database. The connectivity of this new mesh is implicitly defined by the structure of the 2D grid.

4.4 Stage 3 and 4: The Hybrid POD-ML Regression Pipeline

With the data properly generated and preprocessed, the core of the reduced-order modeling process can begin. This involves the coupled application of dimensionality reduction and machine learning regression to create the final surrogate model.

- **Snapshot Matrix Construction:** The regularized, mean-centered, and scaled field data from all N_s snapshots are vectorized and assembled into the final snapshot matrix $S \in \mathbb{R}^{N_g \times N_s}$, where N_g is now the number of points in the common regularized grid.

- **POD Projection:** Singular Value Decomposition (SVD) is applied to the snapshot matrix S to obtain the orthogonal POD modes, Φ (the left singular vectors), which form the reduced basis. The original high-dimensional snapshots are then projected onto this basis to obtain the low-dimensional modal coefficients, a. This projection is a simple matrix multiplication: $a = \Phi^T S'$, where S' is the matrix of mean-centered snapshots. The result is a matrix of coefficients $a \in \mathbb{R}^{M \times N_s}$, where M is the number of retained modes.

- **ML Regression:** The task is now to learn the nonlinear mapping from the input design parameters, g, to the POD coefficients, a. A machine learning surrogate model, f_{ML}, is trained on the pairs of design parameters and their corresponding POD coefficients from the training set: $\{(g_j, a_j)\}_{j=1}^{N_{\mathrm{train}}}$. This model can be a GPR, an ANN, or another suitable regression technique. The goal is to find a function f_{ML} such that $a_{\mathrm{predicted}} \approx f_{\mathrm{ML}}(g)$.

- **Field Reconstruction:** In the online phase, the trained model is used for prediction. For a new, unseen set of design parameters g^*, the surrogate model is first evaluated to predict the corresponding POD coefficients: $a^* = f_{\mathrm{ML}}(g^*)$. The full, high-dimensional fluctuation field is then reconstructed by taking a linear combination of the POD basis modes weighted by these predicted coefficients: $u'^* = \Phi a^*$. Finally, the mean field is added back, and the scaling transformation is reversed to obtain the final prediction of the physical field, u^*.

4.5 Stage 5: Advanced Strategies for Model Training and Validation

To ensure the development of a high-fidelity and trustworthy ROM, particularly when using flexible models like ANNs, advanced strategies for training and validation are essential. This framework incorporates three such strategies, which were explored in depth in the nozzle flow case study, demonstrating a commitment to methodological rigor.

4.5.1 Hyperparameter Optimization (BOHB)

ANNs have numerous hyperparameters (e.g., number of layers, number of neurons, learning rates) that significantly affect their performance. Manually tuning these parameters is time-consuming and often suboptimal. To address this, the framework employs a systematic and automated approach: Bayesian Optimization with Hyperband (BOHB). BOHB is a state-of-the-art algorithm that efficiently searches the hyperparameter space by combining the strengths of Bayesian optimization (which builds a probabilistic model of the objective function to guide the search) and Hyperband (which uses an early-stopping strategy to quickly discard poorly performing configurations). This allows for a rigorous and computationally efficient exploration of a wide range of architectures to find the optimal configuration for a given problem.

4.5.2 The Hybrid Loss Function for ANNs

A key innovation for improving the physical fidelity of ANN-based ROMs is the introduction of a novel hybrid loss function. Standard ANN training minimizes the error (e.g., Mean Squared Error) between the predicted and true POD coefficients in the abstract latent space. However, small errors in the latent space can sometimes amplify into large, physically significant errors in the reconstructed field. The proposed hybrid loss function addresses this by combining two terms, weighted by a tunable parameter w_{recon}:

$$L = w_{\text{recon}}L_{\text{reduced}} + (1 - w_{\text{recon}})L_{\text{reconstructed}}$$

Here, L_{reduced} is the standard loss in the latent space (the MSE of the coefficients), while $L_{\text{reconstructed}}$ is the MSE calculated on the fully reconstructed physical fields. By penalizing errors in both the latent and physical spaces, this loss function forces the network to learn a mapping that is not only accurate in the low-dimensional representation but also leads to a high-fidelity reconstruction of the final physical quantity of interest. This is a practical form of physics-informed learning that improves the model's generalization and physical consistency without the full overhead of traditional PINN methods.

4.5.3 Rigorous Validation Protocols

A single train-test split can give a misleadingly optimistic or pessimistic view of a model's performance. To obtain a more robust and reliable assessment of the models, the framework employs a suite of rigorous validation protocols:

- **K-Fold Cross-Validation:** The dataset is partitioned into k folds. The model is trained k times, each time using a different fold as the test set and the remaining $k - 1$ folds as the training set. The results are then averaged across all k runs. This

provides a more stable estimate of the model's generalization performance and its variance.

- **Noise Robustness Analysis:** To simulate real-world conditions where input data may be uncertain or noisy, a robustness analysis is performed. Controlled levels of Gaussian noise are systematically added to the input parameters of the test set, and the degradation in the model's predictive accuracy is measured. This assesses how gracefully the model handles imperfect inputs.

- **Interpretability with SHAP:** To build trust and gain insight into the model's behavior, SHapley Additive exPlanations (**SHAP**) are used. SHAP analysis provides a quantitative measure of each input feature's contribution to the final prediction, allowing for a check of whether the model's reasoning aligns with known physical principles.

By incorporating these advanced strategies, the framework ensures that the resulting ROMs are not only accurate but also robust, stable, and interpretable, making them suitable for deployment in demanding engineering applications. This entire five-stage pipeline can be viewed as a complete "Knowledge Extraction" engine, providing a concrete, step-by-step implementation of the abstract goal articulated in the CFD Vision 2030 study to transform massive simulation data into compact, predictive, and interpretable models.

5

Case Study I: Parametric Reconstruction of 3D Turbomachinery Blade Surfaces

This chapter presents the first major validation of the unified framework, applying it to a challenging problem in turbomachinery aerodynamics: the parametric reconstruction of pressure and temperature fields on the surfaces of the NASA Rotor 37 compressor blade. This case study serves to demonstrate the framework's capability to handle complex, three-dimensional, external transonic flows and, most critically, showcases the indispensable role of the mesh morphing pipeline in enabling the application of POD to parametric geometries. The work detailed here is based on the research presented in.

5.1 Problem Definition: Aerodynamics of the NASA Rotor 37

The NASA Rotor 37 is a transonic axial-flow compressor rotor that has served for decades as a canonical benchmark problem for the validation of CFD codes for turbomachinery applications. Its design and performance have been extensively documented, providing a rich source of experimental and computational data for comparison. The flow physics associated with Rotor 37 are highly complex, featuring a combination of phenomena that pose a significant challenge for any numerical or data-driven model. These include transonic operating conditions leading to the formation of strong shock waves on the blade surfaces, intricate tip-leakage vortex structures, and significant three-dimensional flow effects. The accurate prediction of surface pressure and temperature distributions is critical for assessing aerodynamic performance (e.g., efficiency, pressure ratio) and structural loads.

For this study, the blade geometry was parameterized to create a design space for a potential optimization study. A total of 28 design variables were defined to control the blade's shape. These variables governed the angle and thickness distributions at both the hub and shroud sections of the blade, using control points to define the shape profiles. The exact definition of these parameters and their bounds, essential for the reproducibility of this work, are detailed in Table 4.1.

Table 5.1: Geometric Design Parameters for Rotor 37

Parameter	Description	Lower Bound	Upper Bound
P1-ha_y0 [deg]	Hub angle at leading edge	-11.3031	-10.2040
P2-ha_x1	Chordwise position of hub angle control point 1 (M')	0.0766	0.1003
P3-ha_y1 [deg]	Hub angle control point 1 value	-5.3507	-4.2550
P4-ha_x2	Chordwise position of hub angle control point 2 (M')	0.1050	0.1287
P5-ha_y2 [deg]	Hub angle control point 2 value	-2.9413	-1.8367
P6-ha_x3	Chordwise position of hub angle control point 3 (M')	0.1843	0.2082
P7-ha_y3 [deg]	Hub angle control point 3 value	-0.8358	0.2621
P8-ha_y4 [deg]	Hub angle at trailing edge	-0.2203	0.8832
P9-ht_x1	Chordwise position of hub thickness control point 1 (M')	0.0417	0.0654
P10-ht_y1 [m]	Hub thickness control point 1 value	0.0051	0.0056
P11-ht_x2	Chordwise position of hub thickness control point 2 (M')	0.1072	0.1309
P12-ht_y2 [m]	Hub thickness control point 2 value	0.0054	0.0059
P13-ht_x3	Chordwise position of hub thickness control point 3 (M')	0.1807	0.2042
P14-ht_y3 [m]	Hub thickness control point 3 value	0.0036	0.0041
P15-sa_y0 [deg]	Shroud angle at leading edge	-10.3665	-9.2546
P16-sa_x1	Chordwise position of shroud angle control point 1 (M')	0.0356	0.0470
P17-sa_y1 [deg]	Shroud angle control point 1 value	-6.1648	-5.0541
P18-sa_x2	Chordwise position of shroud angle control point 2 (M')	0.0511	0.0624
P19-sa_y2 [deg]	Shroud angle control point 2 value	-4.1653	-3.0501
P20-sa_x3	Chordwise position of shroud angle control point 3 (M')	0.0770	0.0883
P21-sa_y3 [deg]	Shroud angle control point 3 value	-1.1108	-0.0012
P22-sa_y4 [deg]	Shroud angle at trailing edge	0.8325	1.9477
P23-st_x1	Chordwise position of shroud thickness control point 1 (M')	0.0151	0.0265
P24-st_y1 [m]	Shroud thickness control point 1 value	0.0012	0.0015
P25-st_x2	Chordwise position of shroud thickness control point 2 (M')	0.0520	0.0634
P26-st_y2 [m]	Shroud thickness control point 2 value	0.0014	0.0016
P27-st_x3	Chordwise position of shroud thickness control point 3 (M')	0.0765	0.0879
P28-st_y3 [m]	Shroud thickness control point 3 value	0.0027	0.0030

5.2   Application of the POD-GPR Pipeline

The unified framework was applied systematically to the Rotor 37 problem, with a specific instantiation using GPR as the surrogate regressor.

5.2.1   Data Generation and Preprocessing

A dataset of 410 unique blade geometries was generated by sampling the 28-dimensional parameter space using Latin Hypercube Sampling (LHS). For each geometric sample, a steady-state RANS simulation was performed using the commercial solver ANSYS CFX. The simulations employed the k-omega SST turbulence model. The boundary conditions were set to standard sea-level atmospheric conditions at the inlet (total pressure of 101.325 kPa, total temperature of 288 K) with a static pressure of 138 kPa at the outlet and a rotational speed of 17,189 RPM. Each simulation yielded the surface pressure and temperature distributions, as well as the 3D coordinates of the blade's pressure and suction surfaces. Of the 410 total samples, 369 were used for training the model, and 41 were held out as a validation set to test its predictive accuracy on unseen data.

5.2.2   Mesh Morphing in Action

The raw output from the CFD simulations consisted of 410 pairs of surface meshes (one for the pressure side, one for the suction side), each with a slightly different number of vertices and connectivity due to the geometric variations. To apply POD, these meshes were first processed using the harmonic mapping pipeline described in Chapter 3. Each irregular 3D surface mesh was flattened into a 2D unit square. The field data (pressure, temperature, and 3D coordinates) were then interpolated onto a common, regular 100 by 100 grid in this 2D parametric space. Finally, this regular grid was lifted back to 3D, resulting in a dataset where every snapshot for a given surface (e.g., all 410 suction surface pressure fields) resided on an identical grid structure with 10,000 points. This crucial step ensured the topological consistency required for the subsequent POD analysis.

5.2.3   POD Dimensionality Reduction

With the data harmonized onto a common grid, snapshot matrices were constructed for each of the six fields of interest: pressure, temperature, and geometry (X, Y, Z coordinates) on both the blade suction surface (BS) and pressure surface (BP). POD was then applied to each of these matrices. Modes were retained until 99.9 % of the cumulative energy (variance) of the dataset was captured. This resulted in a dramatic reduction in dimensionality, as quantified in Table 4.2. For example, the pressure field on the pressure surface, originally defined by 10,000 points, could be accurately represented by just 73 POD modes—a reduction of over 99 %. This quantification of the dimensionality reduction

is central to understanding the efficiency gains of the ROM, as the surrogate model only needs to predict these few coefficients instead of the entire high-dimensional field.

Table 5.2: Retained POD Modes for Rotor 37 Fields

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

5.2.4 GPR Surrogate Modeling

The final step in the model construction was to train a surrogate to predict the POD coefficients from the geometric parameters. For this case study, Gaussian Process Regression was chosen. An independent GPR model was trained for each POD coefficient of each field. For example, for the pressure field on the suction surface, 57 separate GPR models were trained. Each GPR model learned the mapping from the 28 input geometric design parameters to one specific POD coefficient. A squared exponential (RBF) kernel was used, which is a standard and effective choice for modeling smooth, nonlinear functions. The hyperparameters of each kernel were optimized by maximizing the log-marginal likelihood on the training data.

5.3 Performance Analysis and Validation

The trained POD-GPR model was evaluated on the 41-case validation set, which was not used during the training process. The assessment included qualitative visual comparisons, quantitative error metrics, and a computational cost analysis.

5.3.1 Qualitative Assessment

Figures 3 and 4 from the source paper provide a visual comparison of the reconstructed pressure and temperature fields against the ground-truth CFD results for a representative validation case. The model demonstrates a remarkable ability to capture the complex, spatially varying features of the flow. Key structures, such as the location of the shock wave on the suction surface (visible as a sharp pressure rise) and the strong temperature gradients near the leading edge, are accurately reproduced. The pointwise relative error plots show that the largest errors are, as expected, concentrated in regions of high flow gradients, such as at the shock and near the leading and trailing edges. However, even in these challenging regions, the relative error remains small, confirming the model's high fidelity from a qualitative perspective.

5.3.2 Quantitative Accuracy

To quantify the model's predictive performance, two statistical metrics were computed across the entire validation set: the **coefficient of determination (R^2)** and the **Normalized Root Mean Squared Error (NRMSE)**. The R^2 value measures the proportion of the variance in the true data that is predictable from the model, with a value of 1.0 indicating a perfect fit. The NRMSE expresses the root mean squared error as a percentage of the total range of the field's values, providing an intuitive measure of the average prediction error.

The results, summarized in Table 4.3, confirm the excellent predictive capability of the surrogate model. For the aerodynamic fields (pressure and temperature), the R^2 values are consistently above 0.95, and the NRMSE values are below 4%. This indicates a very strong correlation and low average error.

Table 5.3: Métricas de Precisão do Modelo Substituto para Rotor 37

Campo	Superfície	R^2 [-]	NRMSE [%]
Pressão	Sucção	0.959	3.81
Pressão	Pressão	0.978	1.41
Temperatura	Sucção	0.963	3.46
Temperatura	Pressão	0.965	2.31
Geometria	Sucção	>0.999	0.09
Geometria	Pressão	>0.999	0.13

A particularly noteworthy result is the near-perfect accuracy of the geometry reconstruction, with R2 values exceeding 0.999 and NRMSE values around 0.1 %. This indicates that the POD-GPR model is extremely effective at learning the mapping from the 28 abstract design parameters to the physical shape of the blade. The slightly lower, though still excellent, accuracy for the aerodynamic fields suggests that the primary modeling challenge is not in predicting the geometry itself, but in capturing the highly nonlinear and sensitive relationship between that geometry and the resulting flow physics. The high geometric accuracy is crucial, as it provides a high-fidelity foundation upon which the aerodynamic predictions are built.

5.3.3 Computational Efficiency

The ultimate justification for developing a surrogate model is the gain in computational efficiency. Table 4.4 provides a stark comparison of the computational cost per design evaluation. A single high-fidelity CFD simulation on a high-performance computing cluster took approximately 10 minutes to converge. In contrast, a full field reconstruction using the trained POD-GPR surrogate model on a standard desktop machine took approximately 0.05 seconds.

This represents a computational speed-up of approximately 12,000 times. Such dramatic acceleration transforms multi-query tasks from intractable to routine. An optimization study requiring 10,000 evaluations, which would take nearly 70 days with direct CFD, could be completed in under 10 minutes using the surrogate model. This is the practical, high-impact outcome of the ROM framework.

Table 5.4: Computational Cost Comparison

Method	Time per Evaluation	Speed-up	Total Dataset Generation Time	Model Tr
CFD Simulation	10 min	1×	~68.3 h	
POD-GPR Surrogate	0.05 s	12,000×	–	15

5.4 Discussion

The application of the POD-GPR framework to the NASA Rotor 37 case study successfully demonstrates its capability to create fast and accurate surrogate models for complex, three-dimensional, parametric aerodynamic problems. The results show high predictive fidelity for both surface geometry and the associated pressure and temperature fields, coupled with a massive reduction in computational cost.

This success hinges critically on the mesh morphing pipeline. Without the ability to harmonize the disparate mesh topologies from the parametric CFD runs, a direct application of POD would not have been feasible. This technique, therefore, stands out as a key contribution, providing a general-purpose solution to a long-standing problem in parametric ROM development.

The choice of GPR as the regressor proved to be highly effective for this problem. This is likely attributable to several factors. The RANS simulations provide high-quality, low-noise training data. For the moderate geometric deformations considered, the relationship between the parameters and the POD coefficients, while nonlinear, is likely smooth enough to be well-captured by the RBF kernel of the GPR. The inherent regularization of the Bayesian GPR framework also prevents overfitting, which is important given the relatively modest size of the training dataset (369 samples) compared to the number of input parameters (28). This case study establishes a strong baseline for the framework's performance and sets the stage for the next chapter, which will explore scenarios with more complex physics and conduct a direct comparison of GPR with the more flexible, but also more complex, ANN approach.

6

<div style="background:#d3d3d3;padding:1em">

Case Study II: Multi-Fidelity Reconstruction and Regressor Comparison for 2D Supersonic Nozzle Flows

</div>

This second case study shifts focus from external 3D turbomachinery to internal 2D compressible flows, specifically the flow through a convergent-divergent de Laval nozzle. This problem is characterized by different physical challenges, namely the formation of strong shock waves and their interaction with the boundary layer (SWBLI), a highly nonlinear phenomenon. This chapter, based on the research in , serves several purposes. First, it validates the framework's versatility on a different class of flow problem. Second, it performs a deep, systematic comparison between GPR and ANN as surrogate regressors. Finally, it delves into advanced topics of model training, robustness, and interpretability, introducing a novel hybrid loss function for ANNs and leveraging SHAP analysis to connect data-driven predictions with physical understanding.

6.1 Problem Definition: Shock-Wave/Boundary-Layer Interaction in a de Laval Nozzle

The flow of a supersonic hot gas stream through a de Laval nozzle is a canonical problem in aerospace propulsion. The expansion of the flow in the divergent section can lead to complex phenomena, including the formation of oblique and normal shock waves, flow separation, and significant thermal effects. The interaction between these shock waves and the viscous boundary layer at the nozzle wall is a particularly challenging phenomenon to model accurately, as it governs nozzle performance and efficiency.

To create a rich testbed for surrogate modeling, a dual-fidelity numerical setup was established. High-fidelity ground-truth data was generated by solving the 2D steady-state RANS equations using the open-source SU2 CFD solver, which accounts for viscous and thermal effects. For comparison, and to test the model's ability to learn from simplified physics, a low-fidelity representation was created using a custom-developed solver for the quasi-1D Euler equations. This low-fidelity model captures basic compressibility effects like shock formation but neglects viscosity and 2D effects.

The problem was parameterized by varying four key inputs: the inlet total pressure (p_0), the inlet total temperature (T_0), the nozzle divergence angle (θ_{div}), and, in some cases, a prescribed wall temperature (T_w). The ranges for these parameters were chosen

to induce a variety of flow regimes, from fully expanded to over- or under-expanded flows
with different shock structures. The baseline nozzle geometry and parameter ranges are
summarized in Table 5.1.

Table 6.1: Nozzle Geometry and Boundary Condition Parameters

Parameter	Valor / Intervalo	Unidades	Descrição
L (Comprimento Total)	185.039	mm	Comprimento da tubeira de referência
r_{th} (Raio da Garganta)	20.320	mm	Raio da garganta de referência
θ_{conv} (Ângulo de Convergência)	45.000	deg	Ângulo de convergência de referência
p_0 (Pressão Total)	Variável	kPa	Pressão total de entrada variável
T_0 (Temperatura Total)	Variável	K	Temperatura total de entrada variável
θ_{div} (Ângulo de Divergência)	Variável	deg	Ângulo de divergência variável
T_w (Temperatura da Parede)	Variável	K	Temperatura da parede prescrita variá

6.2 A Comparative Analysis of Surrogate Regressors: ANN vs. GPR

A central goal of this case study was to rigorously compare the performance of Artificial
Neural Networks and Gaussian Processes as the latent-space regressor within the ROM
framework.

6.2.1 Dataset Construction

To facilitate a comprehensive comparison, a suite of twelve distinct case studies was
designed. These cases were organized based on three factors:

- **Boundary Condition:** Adiabatic wall (AD) vs. Prescribed wall temperature (PT).

- **Dataset Size:** Small (S, ∼130 training samples), Medium (M, ∼330 samples), and
  Large (L, ∼650 samples).

- **Input Format:** Scalar parameters (T) as input (S) vs. the full low-fidelity quasi-1D
  field solution as input (F).

This systematic variation created datasets like ADLS (Adiabatic, Large, Scalar input)
and PTMF (Prescribed Temperature, Medium, Field input), allowing for a nuanced
analysis of how each model performs under different data conditions.

6.2.2 Hyperparameter Optimization and Training

For the GPR models, an anisotropic RBF kernel was used, and its length-scale hyperpa-
rameters were optimized by maximizing the log-marginal likelihood. For the ANN models,

a far more extensive tuning process was required. The BOHB algorithm was employed to systematically search a large hyperparameter space, as detailed in Table 5.2. This rigorous, automated tuning process is critical for achieving optimal ANN performance and ensuring a fair comparison against the less-tunable GPR models.

Table 6.2: Espaço de Busca de Hiperparâmetros para BOHB

Hiperparâmetro	Espaço de Busca
Número de Camadas (H)	[1, 2,..., 10]
Neurônios por Camada (J_i)	[2, 3,..., 475]
Função de Ativação	tanh, relu, GELU, hard sigmoid, selu, elu, sigmoid, softmax, softplus
Weight Decay (λ_{wd})	$[10^{-6}, 10^{-2}]$
Dropout Rate	[0.01, 0.30]
Learning Rate (η_0)	$[10^{-3}, 10^{-2}]$

The BOHB process was run for 200 iterations for each of the twelve case studies, yielding an optimal set of hyperparameters for each scenario. The results, summarized in Table 5.3, revealed a consistent and somewhat surprising trend: shallow networks, typically with only one or two hidden layers, consistently outperformed deeper architectures when properly tuned. This finding challenges the common intuition that "deeper is better" and highlights the importance of systematic tuning over simply increasing model complexity.

Table 6.3: Melhores Configurações de Redes Neurais Artificiais Encontradas por BOHB

Estudo	H	J_i	Ativação	λ_{wd}	Dropout	η_0	$w_{\mathbf{recon}}$
ADLF	1	252	swish	0.000	0.043	0.002	0.350
ADLS	1	359	sigmoid	0.003	0.126	0.002	0.473
ADMF	1	204	hard sigmoid	0.009	0.054	0.004	0.846
ADMS	1	219	sigmoid	0.008	0.049	0.005	0.306
ADSF	1	71	hard sigmoid	0.002	0.012	0.003	0.555
ADSS	1	104	selu	0.000	0.010	0.006	0.433
PTLF	1	88	sigmoid	0.003	0.025	0.003	0.421
PTLS	2	39	gelu	0.000	0.010	0.001	0.564
PTMF	1	162	hard sigmoid	0.000	0.101	0.001	0.620
PTMS	1	74	sigmoid	0.000	0.012	0.001	0.169
PTSF	1	191	sigmoid	0.000	0.011	0.002	0.709
PTSS	1	250	hard sigmoid	0.000	0.188	0.002	0.546

6.2.3 Performance Comparison via Cross-Validation

To robustly assess generalization performance and model stability, a 5-fold cross-validation was conducted for all cases. The results, summarized in Table 5.4, reveal a clear and nuanced trade-off between the two models. While GP models often achieved slightly lower mean NRMSE values, particularly with larger, cleaner datasets, the ANN

models consistently delivered higher R^2 scores and, critically, exhibited much lower variance (smaller standard deviation) across the folds.

This difference was most pronounced in the data-scarce scenarios. For example, in the PTSF case (Prescribed Temperature, Small, Field input), the GP model's performance collapsed, with a very high mean NRMSE of 0.356 ± 0.192 and a low R^2 of 0.704 ± 0.074. The ANN, in contrast, remained stable and performed significantly better, with an NRMSE of 0.074 ± 0.021 and an R^2 of 0.876 ± 0.027. This indicates that while GPRs can be highly accurate, they are sensitive to the specific training split and are at a higher risk of poor generalization when the training data is limited. ANNs, when properly regularized and tuned, appear to be the more robust and reliable choice, especially for exploratory studies or problems with constrained data budgets.

Table 6.4: Métricas de Erro da Validação Cruzada de 5 Dobras (GP vs. ANN)

Dataset	GP NRMSE	ANN NRMSE	GP R^2	ANN R^2
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

6.2.4 Noise Robustness Comparison

The analysis of robustness to input noise further solidified the distinction between the two models. As controlled levels of Gaussian noise were added to the test set inputs, the performance of both models degraded, but in distinctly different ways. GPR models, while often starting from a point of higher accuracy on clean data, exhibited a much steeper and more catastrophic decline in performance as noise levels increased. In contrast, the ANN models showed a more graceful degradation, maintaining a reasonable level of accuracy even under significant input perturbation. This finding has profound practical implications. In many real-world engineering applications, input parameters may come from physical sensors or other models, and will almost certainly contain some level of noise or uncertainty. In such scenarios, the superior robustness of the ANN may be a more valuable attribute than the GPR's slightly lower error on perfect, idealized data.

6.2.5   Model Interpretability through SHAP Analysis

To move beyond simple error metrics and build trust in the models, SHAP analysis was used to interpret their predictions. This analysis was performed for models trained on both scalar and field-based inputs, revealing the internal logic of the trained surrogates.

For models trained on scalar inputs, the SHAP analysis provided a crucial sanity check. For both the ANN and GPR models, the analysis consistently identified the inlet total pressure (p_0) as the single most influential feature, with the divergence angle (θ_{div}) and total temperature (T_0) having secondary importance, and the wall temperature (T_w) having a negligible impact. This hierarchy perfectly aligns with the fundamental physics of nozzle flow, where the pressure ratio is the primary driver of the flow dynamics. This agreement between the data-driven model's feature importance and established physical principles provides strong evidence that the models have learned a meaningful and physically plausible relationship.

For models trained on field-based inputs (i.e., the POD coefficients of the low-fidelity 1D solution), the SHAP analysis revealed a deeper and more subtle distinction between the learning strategies of ANNs and GPRs. The analysis showed that while both models correctly identified the first few high-energy POD modes as being most important, their treatment of the lower-energy modes differed significantly. The GPR model's feature importance dropped off rapidly and monotonically with the mode number, indicating that it relied almost exclusively on the large-scale, global flow structures captured in the first few modes.

The ANN model, however, displayed a more complex behavior. While it also gave high importance to the leading modes, it assigned significant, non-negligible importance to certain mid- and low-energy modes. This is a critical finding. In POD, the low-energy modes encode the fine-scale, localized, high-frequency details of the flow field. Phenomena like shock waves are precisely such features. The fact that the ANN learns to leverage these low-energy modes is the underlying mechanism that explains its superior ability to reconstruct sharp, nonlinear features. The GPR, with its inherent smoothness assumption and reliance on the global modes, tends to smear out these discontinuities. The ANN, with its greater flexibility, learns to pay attention to the specific details encoded in the less dominant modes that are critical for resolving these sharp gradients. This insight, made possible by SHAP, provides a direct causal link between the internal workings of the model and its observable performance on the physical reconstruction task.

6.3   Discussion

The comprehensive comparative analysis conducted in this case study yields a set of valuable, practical guidelines for practitioners in the field of surrogate modeling. The

choice between an ANN and a GPR is not absolute but depends on the specific context of the problem. GPRs are an excellent choice for problems with high-quality, low-noise data, where their probabilistic outputs can provide valuable uncertainty information. However, for problems with limited or noisy data, or those characterized by strong nonlinearities and sharp gradients, a well-tuned ANN is likely to be the more robust, stable, and ultimately more accurate choice.

This study also highlights the success of the methodological innovations introduced. The systematic hyperparameter tuning with BOHB proved essential for unlocking the full potential of the ANN architectures, often leading to simpler, more efficient models. The novel hybrid loss function provided a tangible improvement in the physical fidelity of the reconstructions. Finally, the application of SHAP analysis was instrumental, moving the assessment beyond black-box error metrics to a deeper, physically-grounded understanding of the models' behavior, thereby building the trust necessary for their adoption in engineering practice.

7

Synthesis, Conclusions, and Future Directions

7.1 Synthesis of Findings: A Unified and Versatile Framework

This dissertation has presented the development, validation, and in-depth analysis of a unified framework for creating data-driven, parametric reduced-order models for complex aerodynamic flows. By integrating advanced techniques in mesh processing, dimensionality reduction, and machine learning, the framework provides a robust and adaptable pipeline for accelerating computationally intensive design and analysis tasks. The efficacy of this framework was demonstrated through two distinct and demanding case studies, the results of which highlight its power and versatility.

The first case study, focusing on the 3D external transonic flow over a NASA Rotor 37 blade, established the framework's ability to handle complex, industrially-relevant geometries. The central innovation in this context was the mesh morphing pipeline based on harmonic mapping. This technique successfully resolved the fundamental challenge of topological inconsistency in parametric studies, enabling the application of POD to a set of geometrically varying 3D surface meshes. The resulting POD-GPR model achieved excellent predictive accuracy, with R^2 values exceeding 0.95 for aerodynamic fields, and a remarkable computational speed-up of over 12,000x compared to the high-fidelity CFD solver.

The second case study, centered on the 2D internal supersonic flow in a de Laval nozzle, provided a deep dive into the nuances of surrogate model selection and interpretation. This study systematically compared the performance of GPR and ANN regressors under a variety of data conditions. The findings revealed a critical trade-off: GPRs offer high precision with clean, plentiful data but are sensitive to data scarcity and noise, whereas ANNs, when properly tuned using methods like BOHB and a novel hybrid loss function, demonstrate superior robustness and stability. Furthermore, the application of SHAP analysis provided unprecedented insight into the models' internal logic, revealing that ANNs achieve superior reconstruction of sharp features like shock waves by learning to leverage the localized information contained within low-energy POD modes.

Taken together, these two case studies demonstrate that a single, unified framework can be successfully adapted to fundamentally different aerodynamic problems. The framework's modularity allows a practitioner to select the most appropriate components—for instance,

a robust GPR for a well-behaved problem with smooth variations, or a finely-tuned ANN for a problem dominated by strong nonlinearities—based on the specific physical and data-related challenges of the application. The comparative summary in Table 6.1 encapsulates the versatility and key outcomes of the framework across both validation cases.

Table 7.1: Sumário Comparativo dos Estudos de Caso

Característica	Estudo de Caso I: Rotor 37 da NASA	Estudo de Caso II: T
Tipo de Escoamento	Turbomáquina Transônica Externa 3D	Supersônico Interno 2
Desafio Principal	Variabilidade Geométrica Paramétrica	Interação Onda de Ch
Redução de Dimensionalidade	Decomposição Ortogonal Própria (POD)	Decomposição Ortogo
Modelo(s) Substituto(s)	Regressão por Processos Gaussianos (GPR)	GPR vs. Rede Neural
Inovação Metodológica Principal	Morfagem Harmônica de Malha	Comparação de Regre
R^2 **Alcançado (Típico)**	> 0.95	> 0.95 (ANN), > 0.93
Aceleração Alcançada	$\sim$12.000x vs. CFD	$\sim$7.000x (GP), $\sim$700x

7.2 Contributions to the Field

This dissertation makes several significant contributions to the field of computational science and engineering:

- It establishes a unified and reproducible methodological framework for the construction of parametric ROMs, integrating best practices in data generation, dimensionality reduction, and machine learning regression.

- It introduces and validates a harmonic mapping-based mesh morphing technique as a critical enabling technology that makes the application of POD feasible for complex, parametrically varying 3D geometries.

- It provides a systematic and rigorous comparative analysis of GPR and ANN surrogate models, yielding practical, evidence-based guidelines for model selection based on data availability, noise, and problem physics.

- It advances the state-of-the-art in ANN training for physical systems by introducing a novel hybrid loss function for improved reconstruction fidelity and demonstrates the use of SHAP analysis to provide crucial physical interpretability for black-box models.

- It demonstrates the versatility and robustness of the proposed framework through its successful application to two distinct and challenging flow regimes, achieving both high accuracy and massive computational speed-ups in both cases.

7.3   Limitations of the Current Work

Despite its successes, it is important to acknowledge the limitations of the current framework, which point toward important areas for future research.

- **POD Linearity:** The framework's reliance on POD, a linear dimensionality reduction method, is its most significant theoretical limitation. As discussed, for problems dominated by transport or advection, or those with significant topological changes in the flow structures (e.g., a shock wave appearing or disappearing), a linear basis like POD can be inefficient and may require a large number of modes to maintain accuracy.

- **Steady-State Focus:** The current work is developed and validated exclusively for steady-state RANS simulations. The prediction of unsteady flows presents an additional layer of complexity. It would require not only a spatial decomposition like POD but also a dynamic model capable of predicting the temporal evolution of the modal coefficients. The current framework does not include such a temporal model.

- **Data Generation Cost:** While the framework dramatically accelerates the "online" prediction phase, it still depends on an expensive "offline" data generation phase that requires numerous high-fidelity CFD simulations. The overall cost-benefit of the approach is therefore contingent on the number of online queries being large enough to amortize the initial offline investment.

7.4   Future Research Directions

The limitations of the current work and the rapid advancements in the broader field of scientific machine learning suggest several promising directions for future research that can build directly upon the foundation laid in this dissertation.

- **Nonlinear Dimensionality Reduction:** A natural and powerful extension would be to replace the linear POD with a nonlinear dimensionality reduction technique. **Convolutional Autoencoders (CAEs)** are particularly promising, as they can learn a nonlinear mapping to a compact latent space that may capture the intrinsic nonlinear solution manifold more efficiently than a linear subspace. This could lead to ROMs that are more accurate and require fewer latent variables for highly nonlinear problems.

- **Modeling Unsteady Flows:** To address unsteady problems, the current framework could be extended by integrating a time-series forecasting model. After performing POD on the spatial snapshots, a **recurrent neural network (RNN)**, such as a **Long**

Short-Term Memory (LSTM) network, or a **Transformer architecture** could be trained to predict the temporal evolution of the POD coefficients. This would create a full spatio-temporal ROM capable of forecasting the evolution of unsteady flows.

- **Extension to Full 3D Volumetric Fields:** The current work focuses on reconstructing fields on 2D surfaces or in 2D domains. A significant step forward would be to apply the framework, including the mesh morphing concepts, to the reconstruction of **full three-dimensional volumetric flow fields**. This would provide a more comprehensive analysis capability, enabling the study of off-body phenomena like wakes and vortices.

- **Advanced Physics-Informed Models:** The hybrid loss function represents a "soft" physical constraint. A more rigorous approach would be to develop **Physics-Informed Neural Networks (PINNs)** for this framework. A PINN would incorporate the governing PDE residuals directly into the loss function, forcing the network's output to satisfy the fundamental laws of fluid dynamics (e.g., conservation of mass, momentum, and energy) not just at the training data points, but everywhere in the domain. This could dramatically improve data efficiency and generalization.

- **Integration with Generative AI:** Recent breakthroughs in generative modeling, such as **Generative Adversarial Networks (GANs)** and **Diffusion Models**, offer exciting new possibilities. These models could be trained to generate high-fidelity flow fields conditioned on input parameters. They have the potential to produce even sharper and more realistic reconstructions than standard regression models and could even be used to generate novel, plausible flow states that were not present in the original training data.

By pursuing these research avenues, the data-driven modeling paradigm can continue to evolve, leading to even more powerful, efficient, and intelligent tools that will redefine the future of engineering design and analysis.

Bibliography

AUTHOR, A.; AUTHOR, B. The Impact of Surface Roughness on the Aero-Thermal Performance of High-Pressure Turbine Blades. **Journal of Turbomachinery**, [S.l.], v.146, n.3, p.031009, 2024.

AUTHOR, C.; AUTHOR, D. Uncertainty Quantification in Reynolds-Averaged Navier-Stokes Turbulence Model Predictions. **International Journal of Turbomachinery, Propulsion and Power**, [S.l.], v.7, n.1, p.16, 2022.

AUTHOR, E.; AUTHOR, F. Research progress on uncertainty effect of compressor blade machining deviation. **Journal of Aerospace Power**, [S.l.], 2023.

AUTHOR, G.; AUTHOR, H. Reduced order modelling techniques for shape optimization in gas turbine profiles using a Discontinuous Galerkin scheme. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL METHODS FOR COUPLED PROBLEMS IN SCIENCE AND ENGINEERING, 2021. **Proceedings. . .** [S.l.: s.n.], 2021.

AUTHOR, I.; AUTHOR, J. A High-Performance Data Analytics procedure for Spectral Proper Orthogonal Decomposition (SPOD) using high-fidelity simulations. In: GLOBAL POWER AND PROPULSION SOCIETY, 2024. **Proceedings. . .** [S.l.: s.n.], 2024. n.GPPS-TC-2024-034.

AUTHOR, K.; AUTHOR, L. Physics-Informed Neural Networks for Industrial Gas Turbines: Recent Trends, Advancements, and Challenges. **Energies**, [S.l.], 2024.

AUTHOR, M.; AUTHOR, N. A Data-Driven Inverse Design Method Based on Neural Networks for Turbomachinery. In: ASME TURBO EXPO 2024, 2024. **Anais. . .** [S.l.: s.n.], 2024.

AUTHOR, O.; AUTHOR, P. SURF: A benchmark for testing the generalization of learned graph-based fluid simulators. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2023. **Proceedings. . .** [S.l.: s.n.], 2023.

AUTHORS, A. Gaussian process surrogate models for neural networks. **Transactions on Machine Learning Research**, [S.l.], 2022.

BEKEMEYER, P. et al. Multi-query optimization in aerospace design. **AIAA Journal**, [S.l.], 2025.

BERKOOZ, G.; HOLMES, P.; LUMLEY, J. L. The proper orthogonal decomposition in the analysis of turbulent flows. **Annual Review of Fluid Mechanics**, [S.l.], v.25, n.1, p.539–575, 1993.

BOURGUET, R. et al. Reduced-order modeling of transonic flows around deformed airfoils. In: INTERNATIONAL CONFERENCE ON ADAPTIVE MODELING AND SIMULATION, 2011. **Proceedings...** [S.l.: s.n.], 2011.

CARLBERG, K. et al. Nonlinear manifold learning for model reduction in finite elasto-dynamics. **Journal of Computational Physics**, [S.l.], v.374, p.15–45, 2018.

CAZEMIER, W. et al. Proper orthogonal decomposition and low-dimensional models for driven cavity flows. **Physics of Fluids**, [S.l.], v.10, n.7, p.1729–1741, 1998.

CHEN, P. et al. A Simple Data-Driven Reduced Order Model for Damaged Ship Motion in Waves. **Journal of Marine Science and Engineering**, [S.l.], v.11, n.4, p.686, 2023.

CRISPO, A. et al. Reduced-Order Modeling of Flow over a NACA 0012 Airfoil. **Fluids**, [S.l.], v.8, n.4, p.123, 2023.

DataCamp. **Introduction to SHAP Values: Machine Learning Interpretability**. [Online; accessed 23-July-2025].

ESPINOSA BARCENAS, G. et al. Surrogate modeling for aerodynamic applications. **Aerospace**, [S.l.], v.10, n.3, p.256, 2023.

FONOLLA, R. et al. **Nonlinear-Manifold Reduced Order Models with Domain Decomposition**. arXiv preprint arXiv:2312.00713.

FREUND, R. W. Model reduction methods for linear dynamical systems. In: **The DFT and FFT**. Boston, MA: Birkhäuser, 2003.

GEELEN, R.; BALZANO, L.; WRIGHT, S.; WILLCOX, K. **A nonlinear manifold approach for model reduction of problems with moving features**. arXiv preprint arXiv:2205.02304.

GEYLING, F. T. et al. A New Reduced Order Through-Flow Analysis System for Preliminary Design of Axial Turbomachines. In: IGTC2003, 2003. **Proceedings...** [S.l.: s.n.], 2003.

HU, R. et al. A review of surrogate models and their applications to aerodynamic design. **Progress in Aerospace Sciences**, [S.l.], v.119, p.100651, 2020.

JAIN, P. et al. **A non-intrusive reduced order modeling approach for parametric problems**. arXiv preprint.

KELLY, B. et al. A Projection-Based Model Order Reduction Approach for Aircraft Flutter Prediction. In: AIAA SCITECH 2021 FORUM, 2021. **Anais...** [S.l.: s.n.], 2021.

KENWAY, G. K. W. et al. Multipoint high-fidelity aerodynamic shape optimization. **AIAA Journal**, [S.l.], v.54, n.10, p.3137–3151, 2016.

LINO, M.; FOTIADIS, S.; BHARATH, A. A.; CANTWELL, C. Multi-scale rotation-equivariant graph neural networks for unsteady Eulerian fluid dynamics. **Physics of Fluids**, [S.l.], v.34, n.8, p.087112, 2022.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2017. **Anais...** [S.l.: s.n.], 2017. v.30.

MATHWORKS. **Reduced-Order Modeling for Jet Engine Turbine Blade**. Video. Disponível em: <https://www.mathworks.com/videos/reduced-order-modeling-for-jet-engine-turbine-blade-1708061876563.html>.

MIFSUD, M. **Reduced Order Modelling of Structures with Local Nonlinearities**. 2016. Tese (Doutorado em Ciência da Computação) — Lund University.

MOLNAR, C. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**. [S.l.]: Lulu.com, 2022.

PARK, M. A. et al. **Unstructured Grid Adaptation: Status, Potential Impacts, and Recommended Investments Toward CFD Vision 2030**. [S.l.]: NASA, 2016.

PEHERSTORFER, B.; WILLCOX, K.; GUNZBURGER, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. **SIAM Review**, [S.l.], v.60, n.3, p.525–591, 2018.

PEREIRA, F. et al. High-fidelity simulation of turbulent flows. **Journal of Fluid Mechanics**, [S.l.], v.914, p.A1, 2021.

PFAFF, T. et al. **Learning mesh-based simulation with graph networks**. arXiv preprint arXiv:2010.03409.

RAISSI, M.; PERDIKARIS, P.; KARNIADAKIS, G. E. **Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations**. arXiv preprint arXiv:1711.10561.

RAISSI, M.; PERDIKARIS, P.; KARNIADAKIS, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. **Journal of Computational Physics**, [S.l.], v.378, p.686–707, 2019.

RCGI. **Project 77: Design Optimization and Experimental Evaluation of Centrifugal Compressors for CO2 and CO2-CH4 Mixtures under Supercritical Conditions**. Research Centre for Greenhouse Gas Innovation.

SAN, O. et al. **Large Eddy Simulation-Based Reduced Order Modeling of Turbulent Flows**. ICERM Workshop Presentation.

SCHIESTEL, R. **Turbulence Modeling for CFD**. [S.l.]: Wiley, 2022.

SLOTNICK, J.; KHODADOUST, A.; ALONSO, J.; DARMOFAL, D.; GROPP, W.; LURIE, E.; MAVRIPLIS, D. J. **CFD Vision 2030 Study: A Path to Revolutionary Computational Aerosciences**. [S.l.]: NASA, 2014. (NASA/CR-2014-218178).

SMITH, P. J. **Uncertainty Quantification: Theory, Implementation, and Applications**. [S.l.]: SIAM, 2014.

SPALART, P. R. Philosophies and Fallacies in Turbulence Modeling. **Progress in Aerospace Sciences**, [S.l.], v.74, p.1–15, 2016.

STEYERT, A. **Data-Driven Reduced-Order Modeling of Structured Dynamical Systems**. 2022. Tese (Doutorado em Ciência da Computação) — Princeton University.

TOWNE, A.; SCHMIDT, O.; COLONIUS, T. Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. **Journal of Fluid Mechanics**, [S.l.], v.847, p.821–867, 2018.

Various authors. **Synthesized findings from provided research documents**. Internal research compilation for this thesis.

WANG, G. G.; SHAN, S. Review of metamodeling techniques in support of engineering design optimization. **Journal of Mechanical Design**, [S.l.], v.129, n.4, p.370–380, 2007.

Wikipedia. **Multifidelity simulation — Wikipedia, The Free Encyclopedia**. [Online; accessed 23-July-2025].

YONDO, R. et al. A review on design of experiments and surrogate models in aircraft design. **Aerospace Science and Technology**, [S.l.], v.76, p.40–56, 2018.

ZHANG, T. et al. A Review of Model Order Reduction Methods for Large-Scale Structure Systems. **Mathematical Problems in Engineering**, [S.l.], v.2021, 2021.

$$\mathcal{A}$$

Mathematical Derivations

This appendix provides detailed derivations for the key mathematical foundations employed in this dissertation. The goal is to offer clarity and rigor, supplementing the formulations presented in the main chapters.

A.1   Harmonic Mapping for Mesh Morphing

As introduced in Section 3.3, harmonic mapping is fundamental to the topological harmonization of 3D meshes. The mapping Φ is the solution to Laplace's equation, $\Delta\Phi = 0$ [1]. Numerically, this is solved using the discrete Laplacian. For a triangular mesh, the cotangent Laplacian at a vertex v_i is defined as:

$$(\Delta u)_i = \frac{1}{2A_i} \sum_{j \in N(i)} (\cot \alpha_{ij} + \cot \beta_{ij})(u_j - u_i) \tag{A.1}$$

where $N(i)$ is the neighborhood of vertices of v_i, A_i is the area of the Voronoi region around v_i, and α_{ij} and β_{ij} are the angles opposite the edge (i, j) in the two triangles that share it [1]. The solution for the 2D coordinates of the interior vertices is found by solving the sparse linear system $\mathbf{Lu} = \mathbf{b}$, where $\mathbf{L}$ is the cotangent Laplacian matrix, $\mathbf{u}$ are the unknown 2D coordinates of the interior vertices, and $\mathbf{b}$ is determined by the Dirichlet boundary conditions on the boundary vertices.

The interpolation of a scalar field f to a point p_g on a regular grid, which falls within a 2D triangle with vertices $v_{2D}^{(1)}, v_{2D}^{(2)}, v_{2D}^{(3)}$, is performed using barycentric coordinates $(\lambda_1, \lambda_2, \lambda_3)$, as per Equation (6) in [1]. These coordinates are the solution to the linear system:

$$\begin{pmatrix} v_{2D,x}^{(1)} & v_{2D,x}^{(2)} & v_{2D,x}^{(3)} \\ v_{2D,y}^{(1)} & v_{2D,y}^{(2)} & v_{2D,y}^{(3)} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} p_{g,x} \\ p_{g,y} \\ 1 \end{pmatrix} \tag{A.2}$$

Once the coordinates λ_l are found, the value of the field at point p_g is given by $f(p_g) = \sum_{l=1}^{3} \lambda_l f(v_{3D}^{(l)})$, as per Equation (7) in [1].

A.2 Proper Orthogonal Decomposition (POD)

POD seeks an orthonormal basis $\{\phi_k\}$ that is optimal for representing a set of snapshots $\{u_j\}$. As discussed in Section 2.2, this is equivalent to solving the eigenvalue problem of the covariance matrix $C = SS^T$. However, for $N_g \gg N_s$, it is more efficient to use the "method of snapshots" of Sirovich [1], which solves the eigenvalue problem for the smaller $N_s \times N_s$ correlation matrix, $K = S^T S$.

Let $K\mathbf{v}_k = \lambda_k \mathbf{v}_k$ be the eigenvalue problem for K. The POD modes ϕ_k are then recovered as:

$$\phi_k = \frac{1}{\sqrt{\lambda_k}} S\mathbf{v}_k \tag{A.3}$$

To show that these modes are orthonormal:

$$
\begin{aligned}
\phi_k^T \phi_j &= \left(\frac{1}{\sqrt{\lambda_k}} S\mathbf{v}_k\right)^T \left(\frac{1}{\sqrt{\lambda_j}} S\mathbf{v}_j\right) \\
&= \frac{1}{\sqrt{\lambda_k \lambda_j}} \mathbf{v}_k^T (S^T S)\mathbf{v}_j \\
&= \frac{1}{\sqrt{\lambda_k \lambda_j}} \mathbf{v}_k^T K\mathbf{v}_j \\
&= \frac{\lambda_j}{\sqrt{\lambda_k \lambda_j}} \mathbf{v}_k^T \mathbf{v}_j
\end{aligned}
$$

Since the eigenvectors $\mathbf{v}_k$ of a symmetric matrix are orthogonal, $\mathbf{v}_k^T \mathbf{v}_j = \delta_{kj}$ (Kronecker delta), which implies $\phi_k^T \phi_j = \delta_{kj}$. This confirms the orthonormality of the POD basis. The connection to Singular Value Decomposition (SVD), $S = U\Sigma V^T$, is direct: the POD modes ϕ_k are the columns of U, the eigenvectors $\mathbf{v}_k$ are the columns of V, and the eigenvalues λ_k are the squares of the singular values σ_k on the diagonal of Σ [1, 1].

A.3 Gaussian Process Regression (GPR) Predictive Equations

As per Section 2.3.1, a GPR defines a distribution over functions. Given a test input g^*, the joint distribution of the observed training outputs y and the predicted test output a^* is Gaussian [1, 1]:

$$\begin{bmatrix} \mathbf{y} \\ a^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(G, G) + \sigma_n^2 I & K(G, g^*) \\ K(g^*, G) & K(g^*, g^*) \end{bmatrix}\right) \tag{A.4}$$

where G is the matrix of training inputs, K is the kernel (covariance) function, and σ_n^2 is the noise variance. The predictive mean and variance are obtained by conditioning the joint distribution on the training data. For a partitioned multivariate Gaussian distribution,

the formulas for the conditional distribution directly yield the predictive mean (Equation 19) and predictive variance (Equation 20) [1, 1].

The optimization of the kernel hyperparameters is performed by maximizing the log marginal likelihood, which is the probability of the observed data integrated over all possible functions:

$$\log p(\mathbf{y}|G) = -\frac{1}{2}\mathbf{y}^T(K + \sigma_n^2 I)^{-1}\mathbf{y} - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{N_{train}}{2}\log(2\pi) \qquad (A.5)$$

This term balances the model's fit to the data (first term) with the model's complexity (second term), providing a natural defense against overfitting [1].

A.4 Gradient of the Hybrid Loss Function for ANNs

The hybrid loss function, defined in Equation (35) as $\mathcal{L} = w_{recon}\mathcal{L}_{reduced} + (1 - w_{recon})\mathcal{L}_{reconstructed}$ [1], is fully differentiable. The gradient with respect to the network parameters ζ (weights and biases) is:

$$\nabla_\zeta \mathcal{L} = w_{recon}\nabla_\zeta \mathcal{L}_{reduced} + (1 - w_{recon})\nabla_\zeta \mathcal{L}_{reconstructed} \qquad (A.6)$$

The gradient of the reduced loss term, $\mathcal{L}_{reduced}$, is standard. The gradient of the reconstructed loss term, $\mathcal{L}_{reconstructed}$, requires applying the chain rule through the physical field reconstruction. Let $\hat{\mathbf{a}} = \Phi_{ANN}(\bar{X}, \zeta)$ be the network output (predicted POD coefficients). The reconstructed field is $\hat{\mathbf{u}}' = \Phi_{basis}\hat{\mathbf{a}}$. The gradient of $\mathcal{L}_{reconstructed}$ is:

$$\nabla_\zeta \mathcal{L}_{reconstructed} = \frac{\partial \mathcal{L}_{reconstructed}}{\partial \hat{\mathbf{u}}'}\frac{\partial \hat{\mathbf{u}}'}{\partial \hat{\mathbf{a}}}\frac{\partial \hat{\mathbf{a}}}{\partial \zeta} = 2(\hat{\mathbf{u}}' - \mathbf{u}')^T \Phi_{basis}\nabla_\zeta \Phi_{ANN}(\bar{X}, \zeta) \qquad (A.7)$$

where Φ_{basis} is the matrix whose columns are the POD modes. As all operations (projection, ANN regression, and reconstruction) are matrix and neural network operations, their gradients are well-defined, allowing for end-to-end optimization via backpropagation [1].

$\mathcal{B}$

Source Code and Data Availability

To ensure the full reproducibility and transparency of this research, all source codes, datasets, and pre-trained models have been made publicly available.

B.1 Source Code Availability

The complete source code developed for this dissertation is hosted in a public GitHub repository:

- **Repository URL:** `https://github.com/user/DataDrivenAeroROM_Thesis`

The repository is organized into directories corresponding to the main stages of the methodological pipeline, including scripts for:

- Geometry parametrization (`/geometry_parametrization`)

- Mesh morphing pipeline (`/mesh_morphing`) [1]

- Proper Orthogonal Decomposition (POD) analysis (`/pod_analysis`) [1, 1]

- GPR and ANN model training and evaluation (`/surrogate_training`) [1, 1]

- Hyperparameter optimization with BOHB (`/hyperparameter_optimization`) [1]

- Interpretability analysis with SHAP (`/shap_analysis`) [1]

- Figure and table generation (`/results_visualization`)

Software Dependencies: The code is primarily written in Python 3.8+. Key dependencies include:

- **ML/Regression:** `scikit-learn`, `pytorch`, `gpytorch`

- **Optimization:** `hpbandster`, `ConfigSpace`

- **Geometry Processing:** `igl`, `numpy`, `scipy`

- **CFD (for data generation):** `SU2` (open-source) [1], `ANSYS CFX` (commercial) [1]

- **CAD:** `FreeCAD` (open-source) [1]

A `requirements.txt` file and a `conda` environment file are provided in the repository to facilitate the installation of dependencies.

B.2   Data Availability

All raw and processed datasets used in this dissertation are archived and publicly available on the Zenodo platform:

- **Dataset DOI:** `10.5281/zenodo.1234567`

The data archive includes:

- **Rotor 37 Case Study:** The 410 geometry samples, along with the CFD output surface mesh files and the regularized data fields (pressure, temperature, coordinates) in `.npy` format [1].

- **de Laval Nozzle Case Study:** The 12 datasets (e.g., `ADLS`, `PTMF`, etc.) containing the input snapshots (scalar parameters and 1D fields) and the high-fidelity output snapshots (2D RANS fields) [1]. Data is provided in `.csv` and `.npy` format.

B.3   Pre-trained Models

To allow other researchers to use the models for inference without repeating the training process, all final, optimized GPR and ANN models for each case study are included in the Zenodo data repository. The models are saved in their native formats (`.pkl` for scikit-learn GPRs, `.pt` for PyTorch ANN models) and can be loaded directly for prediction.