Karlsruhe Institute of Technology

Institute of Fluid Mechanics

東京大学
生産技術研究所
Institute of Industrial Science,
The University of Tokyo

Deep Operator Networks for Hemodynamics Modeling of Mitral Valve Flow

Master Thesis of

Jakob Marcel Hoffmann

At the Department of Mechanical Engineering,
Institute of Fluid Mechanics

Advisors: Prof. Dr.-Ing. Bettina Frohnapfel
Prof. D. Eng. Yosuke Hasegawa (UTokyo-IIS)
Dr.-Ing. Alexander Stroh

Duration: July 2025 – January 2026

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, January 15, 2026

.....

(**Jakob Marcel Hoffmann**)

Acknowledgment

I would like to express my gratitude to everyone who supported me throughout the development of this thesis.

Special thanks go to my supervisor, Alex, who I've had the pleasure of continuously working with since even before my bachelor's thesis. He was paramount in enabling and shaping this research in cooperation with Prof. Yosuke Hasegawa at the Institute of Industrial Science of the University of Tokyo.

Hasegawa-sensei welcomed me into his laboratory, and together with Alex, steadily supported and guided me. This interdisciplinary and multicultural environment proved highly conducive through invaluable social and academic engagement.

Specifically Osawa-san, Tomizawa-san and Shimodaira-san frequently assisted me with all kinds of crucial technical and administrative matters, but many more friendships have emerged in addition. I am grateful to the laboratory for their financial support through the UTokyo-IIS internship support program and to all lab members for their kindness and warm hospitality during my stay.

Additionally, this international experience was made possible in part by the generous support of the DAAD PROMOS scholarship, and my parents, who have always fostered my academic work.

Furthermore, I am thankful to Maximilian Dreisbach for sharing his expertise and making insightful observations and suggestions that improved my understanding and application of machine learning.

I extend my gratitude to Prof. Bettina Frohnapfel for her excellent academic leadership, as well as the researchers at both KIT and University Hospital Heidelberg who have contributed to the overarching research project.

Abstract

Accurate characterization of hemodynamic flows is essential for diagnosing and treating cardiovascular pathologies such as mitral regurgitation. While 4D Flow MRI provides volumetric data with limited resolution, Computational Fluid Dynamics (CFD) offers high fidelity but incurs significant computational costs. This thesis addresses the gap between these methods by developing machine learning frameworks to reconstruct transient volumetric (4D) flow fields from sparse experimental observations obtainable in an in-vitro setting. The primary goal is to solve the ill-posed problem of inferring full 4D hemodynamics using limited inputs: time-resolved pressure boundary conditions and planar velocity slices from two-component 2D Particle Image Velocimetry (2D2C PIV). The study evaluates two architectural strategies: instance-specific Physics-Informed Neural Networks (PINNs) and generalized Deep Operator Networks (DeepONets). Synthetic ground truth data was generated using Unsteady Reynolds-averaged Navier-Stokes (URANS) simulations of a hemodynamic simulator with various mitral regurgitation orifice phantoms. Two simplified datasets, representing steady-state and transient 2D channel flows, were used for hierarchical model development. The study compares purely data-driven models, physics-augmented models using RANS loss functions, and a hybrid "test-time-adaptation" strategy where a pre-trained DeepONet is fine-tuned on sparse target observations.

The PINN approach consistently failed to resolve the ill-posed reconstruction problem, struggling with flow topology and temporal evolution across all complexity levels. In contrast, the DeepONet framework successfully leveraged training distributions to act as a robust surrogate. The hybrid training strategy proved critical, significantly reducing error for out-of-distribution steady-state cases. In transient 3D applications, the hybrid DeepONet significantly outperformed PINNs, achieving physically plausible reconstructions and correcting jet trajectories for unseen geometries. However, fine-tuning improvements remained spatially localized to the supervision window, where they are least needed due to the availability of measurement data, and led to disjointed volumetric artifacts. Eventually, the hybrid DeepONet framework has been established as a superior and computationally efficient method for reconstructing 4D hemodynamics from sparse data. While significant challenges regarding volumetric coherence remain to be solved, the approach offers a promising pathway for reduced-cost, high-fidelity flow analysis in cardiovascular geometries.

Kurzfassung

Die präzise Charakterisierung hämodynamischer Strömungen ist für die Diagnose und Behandlung kardiovaskulärer Pathologien wie der Mitralklappeninsuffizienz unerlässlich. Während 4D-Flow-MRI volumetrische Daten mit begrenzter Auflösung liefert, bietet die numerische Strömungsmechanik (CFD) eine hohe Genauigkeit, verursacht jedoch erhebliche Rechenkosten. Diese Arbeit adressiert die Lücke zwischen diesen Methoden durch die Entwicklung von Frameworks für maschinelles Lernen, um transiente volumetrische (4D) Strömungsfelder aus spärlichen experimentellen Beobachtungen zu rekonstruieren, die in einer In-vitro-Umgebung gewonnen werden können. Das primäre Ziel besteht in der Lösung des schlecht gestellten Problems, die vollständige 4D-Hämodynamik unter Verwendung begrenzter Eingangsdaten abzuleiten: zeitaufgelöste Druckrandbedingungen und planare Geschwindigkeitsfelder aus der zwei-Komponenten 2D Particle-Image-Velocimetry (2D2C PIV). Die Arbeit evaluiert zwei Architekturstrategien: instanzspezifische physik-informierte neuronale Netze (PINNs) und generalisierte Deep Operator Networks (DeepONets). Synthetische Referenzdaten wurden mittels instationärer Reynolds-gemittelter Navier-Stokes-Simulationen (URANS) eines hämodynamischen Simulators mit verschiedenen Phantomen für Mitralinsuffizienz-Öffnungen generiert. Zwei vereinfachte Datensätze, die stationäre und transiente 2D-Kanalströmungen repräsentieren, dienten der hierarchischen Modellentwicklung. Die Studie vergleicht rein datengetriebene Modelle, physik-erweiterte Modelle mit RANS-Verlustfunktionen und eine hybride „Test-Time-Adaptation“-Strategie, bei der ein vortrainiertes DeepONet anhand spärlicher Zielbeobachtungen feinjustiert wird.

Der PINN-Ansatz scheiterte konsistent an der Lösung des schlecht gestellten Rekonstruktionsproblems und wies Schwierigkeiten bei der Strömungstopologie sowie der zeitlichen Entwicklung über alle Komplexitätsstufen hinweg auf. Im Gegensatz dazu nutzte das DeepONet-Framework erfolgreich Trainingsverteilungen, um als robustes Surrogatmodell zu fungieren. Die hybride Trainingsstrategie erwies sich als entscheidend und reduzierte den Fehler bei stationären Fällen außerhalb der Trainingsverteilung erheblich. In transienten 3D-Anwendungen übertraf das hybride DeepONet die PINNs deutlich, erreichte physikalisch plausible Rekonstruktionen und korrigierte Strahlverläufe bei unbekannten Geometrien. Die Verbesserungen durch die Feinabstimmung blieben jedoch räumlich auf das Überwachungsfenster begrenzt, in dem sie aufgrund der verfügbaren Messdaten am wenigsten benötigt werden, und führten zu unzusammenhängenden volumetrischen Artefakten. Letztlich hat sich das hybride DeepONet-Framework als eine überlegene und recheneffiziente Methode zur Rekonstruktion der 4D-Hämodynamik aus spärlichen Daten etabliert. Obwohl noch erhebliche Herausforderungen hinsichtlich der volumetrischen Kohärenz zu lösen sind, bietet der Ansatz einen vielversprechenden Weg für eine kosteneffiziente, hochgenaue Strömungsanalyse in kardiovaskulären Geometrien.

Contents

1. Introduction

1.1. Motivation

Mitral regurgitation (MR) is a prevalent cardiovascular condition in which the mitral valve (MV) fails to close adequately during systole, causing retrograde blood flow from the left ventricle (LV) back into the left atrium (LA). MR generally progresses and causes the heart to compensate for the increasing regurgitant volume, which results in poor outcomes when the condition becomes severe. Quantification of MR severity, along with the determination of its cause and mechanism, is the basis for clinical decision-making for MR treatment [12, 24, 64].

A hemodynamic simulator developed at the University Hospital Heidelberg by Karl et al. in cooperation with Karlsruhe Institute of Technology (KIT) provides avenues for clinician training, patient-specific planning, and fundamental research related to the condition [27, 28]. Motivating factors for this effort are the limitations of current diagnostic methods such as Transesophageal Echocardiography (TEE), resulting in an underestimation of MR severity and operator variability. Two-component 2D Particle Image Velocimetry (2D2C PIV) measurements were conducted within the hemodynamic simulator to enhance comprehension of flow dynamics in MR. PIV was used to quantitatively assess the accuracy of clinical diagnostic methods and to capture the pulsatile flow patterns with high spatiotemporal resolution. This enables the discovery of connections between heart pathologies and the dynamics of the flow, such as the size, shape, and direction of the regurgitation jet that forms from the retrograde flow across different leaky valves [27, 28, 29, 43].

Since the PIV measurements are limited to fluid velocity components in optically accessible areas that are recorded one plane at a time, transient computational fluid dynamics (CFD) simulations of the hemodynamic simulator were conducted to gain insights into other flow quantities and parts of the domain that cannot be captured via PIV. The simulations use unsteady Reynolds-averaged Navier-Stokes (URANS) equations for turbulence modeling. This produced a small CFD dataset encompassing various MV geometries and boundary conditions (BCs), thereby facilitating the exploration of machine learning (ML) techniques with CFD ground truth data for validation.

Both Physics-Informed Neural Networks (PINNs) and Deep Operator Networks (DeepONets, DONs) are appealing approaches as they work with sparse datasets. They can be combined into Physics-Informed Deep Operator Networks (PI-DeepONets). By incorporating the underlying physical equations, physics-informed techniques may generally achieve improved accuracy compared to purely data-driven training and even avoid data supervision completely (self-supervised) [1, 18, 37].

1.2. Objective

The goal of this thesis is to explore PINNs and DeepONets to model the pulsatile flow phenomena related to MR, investigating the ability to overcome the following challenges of research in the hemodynamic simulator:

1. The PIV measurements are limited to a confined area or volume visible through windows in the simulator. The simulator's design requirements prevent the arbitrary placement of these optical access points, restricting the possible coverage.
2. Acquiring a high-resolution three-dimensional (3D) velocity distribution proves time-consuming with the 2D2C PIV method employed in the present recordings, as it necessitates the sequential recording of multiple two-dimensional (2D) planes to resolve the third dimension.
3. CFD offers a powerful tool for detailed blood flow simulation but can be computationally intensive, particularly for patient-specific models or extensive parameter studies.

Therefore, using the CFD dataset, artificial neural networks (ANNs) are developed to output the full 3D transient (4D) flow quantities based on observable information of the simulator known before and measured during experiments, such as the geometry of the simulator and MV, the pressure values in the LV and LA and the spatially constrained PIV recordings.

The PINNs are designed as physics-driven surrogates, given the governing equations and aforementioned BCs during training for a single flow configuration. The DeepONets include training data of multiple configurations, making them primarily data-driven surrogates with fast inference. However, the inclusion of those BCs and governing equations for the unseen validation case is also discussed, creating a hybrid approach. This distinction is elaborated in the Methodology Section 2.2.

The thesis proceeds in successive stages:

1. Initially, the previously developed 3D URANS CFD simulations of the hemodynamic simulator are established and validated against available experimental data.
2. The CFD setup is reduced to a 2D channel configuration with a vertical orifice, resulting in two larger datasets representing steady-state (RANS) and transient (URANS) regurgitant jets thanks to the reduced cost of a single configuration.
3. Based on these three datasets, PINNs and (hybrid) DeepONets are implemented in order of increasing CFD simulation complexity and validated against data excluded from the training datasets.
4. The resulting framework is validated against available experimental PIV data from the KIT/Heidelberg hemodynamic simulator to investigate its feasibility for reduced-cost, physics-consistent modeling.

1.3. State of the Art

ML has rapidly become integral across diverse aspects of valvular heart disease research and care, being applied to diagnosis, disease severity assessment, procedural planning, outcome prediction, and fundamental methods [40]. This includes data-driven and physics-driven surrogates for CFD modeling, for which various methods have been proposed [56].

For instance, Pajaziti et al. utilized purely data-driven deep ANNs, trained on 3,000 CFD simulations derived from a statistical shape model of patient aortas, to accurately predict 3D pressure and velocity flow fields based on aorta shape input, achieving considerable speedup over conventional CFD [46]. Deep learning is also being used to improve 4D Flow Magnetic Resonance Imaging by overcoming limitations related to noise, acquisition time trade-offs and low spatiotemporal resolution [5, 7, 14], and for MR severity quantification on a 4-step ordinal scale based on echo-cardiograms [35, 40].

1.3.1. Physics-Informed Neural Networks

PINNs, as first coined by Raissi et al. [50], can solve ill-posed problems and also address the challenges of high data collection costs and limited generalization or robustness of data-driven models. This is achieved by incorporating the residuals of the governing physical laws on collocation points, described by the Navier-Stokes equations (NSE) in the case of CFD, into the ANN's loss function during training. Vanilla PINNs use a standard Multilayer Perceptron (MLP) architecture [1, 4, 50, 56].

Cai et al. show how PINNs can be used to reconstruct transient 3D flow fields when given just a few planes of two-dimensional, two-component velocity observations [4], which are analogous to the PIV measurements in the hemodynamic simulator. Other authors discussed several methodologies for incorporating turbulence models into PINNs, including the utilization of established two-equation models, or simpler approaches such as the mixing length model and treating the turbulence closure parameters directly as outputs of the neural network [10, 16, 19, 47, 62]. Ur Rehman et al. demonstrate an application of PINNs coupled with the RANS equations for pulsatile aortic flow in Marfan-syndrome aneurysms [54]. Their network predicts velocity, pressure, and wall-shear-stress fields while minimizing a composite loss combining partial differential equation (PDE) residuals and data-loss terms based on CFD reference. The CFD supervision is applied at randomly sampled spatiotemporal points across the 4D domain, providing volumetric guidance.

For MR specifically, Ling et al. proposed using both a self-supervised PINNs and a physics-guided supervised network as alternatives to traditional 2D flow mapping algorithms, demonstrating comparable performance in reconstructing intraventricular flow [34]. Maidu et al. also focused on 2D flow, but developed a PINN framework that incorporates the NSE to not only reconstruct velocity but also to derive fluctuating pressure fields and achieve temporal super-resolution, inferring flow at time points without Doppler data [38]. Wong et al. advanced this to 3D, using a Coupled Sequential Frame PINN to reconstruct 3D velocity and pressure fields from 3D color Doppler, showing that their time-marching approach could accurately capture detailed hemodynamics even from sparse and noisy data [60].

1.3.2. Deep Operator Networks

While there are also parametrized PINNs that can infer solutions for different conditions as prescribed via additional network inputs rather than requiring training an ANN for each one [9, 15], operator networks have emerged to address this use-case of fast inference for many configurations. Operator networks learn the mapping between input functions and output functions, for example, between discretized BCs and the flow field inside the domain. Among the proposed frameworks, DeepONets are particularly flexible, as they can handle input functions defined on unstructured data. Their strong inductive bias is realized through a split trunk-branch architecture that separates the input function (branch) from the inference coordinate inputs (trunk). This provides robustness, ensures consistency when calculating partial derivatives, which simplifies integrating physics-informed constraints, and is supported by the universal approximation theorem for operators. DeepONets are notably data-efficient, capable of learning from sparse datasets where other models may require dense grids. PI-DeepONets further enhance this, achieving high accuracy with fewer or even zero paired training samples. DeepONets generalize well to unseen data from the same distribution; their error can grow exponentially for out-of-distribution inputs, which may require further, albeit light, training [18, 30, 37, 56, 57].

Recent works have successfully established DeepONets as a viable tool for hemodynamic modeling in one-dimensional space. For instance, Hong et al. demonstrated the ability to

map an input flow rate waveform directly to an output pressure waveform [20]. Li et al. showed that the framework can learn the complex operator linking abstract physiological signals like electrocardiograms and photoplethysmograms to complete blood pressure and flow dynamics, highlighting its potential for cuffless blood pressure estimation [31].

Applying these principles to full-field hemodynamics, Cruz-González et al. developed a multi-input PI-DeepONet to predict steady-state 3D velocity and pressure fields in an abdominal aortic aneurysm [6]. In their work, the operator learns to map the system's boundary conditions, namely the inlet velocity and outlet pressure, to the complete internal flow field. This demonstrates the framework's capability to act as a surrogate for solving the forward problem governed by the NSE. The authors' comparative analysis highlights a key trade-off: incorporating physical laws consistently improves model accuracy at the cost of increased training time. A PINN, for instance, was more accurate but five times slower to train than a standard deep ANN, with a similar trend observed between the PI-DeepONet and its data-driven counterpart. The authors tested how sparse supervised data, configured as randomly scattered points or cross-sectional or longitudinal 2D slices, impacts model performance. They found that a random distribution of data consistently yielded the highest accuracy for both PINNs and DeepONets. Crucially, their results imply that when supervised data is well-distributed, the additional constraint from the physics loss is less critical than when the data is highly localized and less representative of the overall field. While the paper does not directly compare PINN against DeepONet results, it positions them for distinct use cases: PINNs are applied to solve a single simulation instance, while DeepONets learn a parametric operator capable of solving for new BCs without retraining.

To bridge the gap between simulation and clinical reality, Velikorodny et al. introduced a multifidelity DeepONet that combines many low-fidelity CFD simulations with a smaller set of high-fidelity in-vitro experimental measurements [55]. Their work highlights a pathway for leveraging sparse but accurate physical data to anchor and improve models primarily trained on simulations. To the best of the author's knowledge, there are no publications documenting the application of DeepONets to transvalvular blood flow, making this thesis a novel contribution.

In anticipation of the detailed discussion presented in Section 2.2, the input function to the DeepONet in this work is defined as the PIV measurement, representing a spatially masked 2D subset of the target output field. A limitation of using a standard MLP for the input function network is that the 2D input fields must be flattened into a 1D vector. This process inherently discards spatial information. To address this, Mei et al. proposed a Fully Convolutional DeepONet (FC-DeepONet) for predicting seismic travel-time fields while preserving spatial relationships of the input [41]. Bai et al. employed a convolutional branch network to predict the temporal evolution of unsteady flow by taking a sequence of 2D velocity and pressure fields as input [2]. To effectively capture long-range spatial dependencies across the entire flow field, identified as critical for accurate, stable long-term predictions, they augmented their architecture with a self-attention mechanism.

As for the hybrid training step mentioned in the Objective Section 1.2, several authors investigate the application of a two-step, hybrid approach that combines the strengths of data-driven operators and physics-informed methods. This paradigm is referred to as "pre-train and fine-tune" or "Test-Time Adaptation" (TTA). The core idea is to first pre-train a general operator network, such as a DeepONet, on a training dataset to capture the fundamental dynamics of the system across different configurations. In a second step, before inferring the results for a new, unseen validation case, this pre-trained network is rapidly fine-tuned using information specific to that new case. This adaptation step makes the model more accurate and robust, especially for out-of-distribution inputs [33, 66].

Several recent works have focused on using the governing PDEs as the source of information for this fine-tuning step. The Physics-Informed Neural Operator (PINO) framework proposed by Li et al. formalizes this with an "instance-wise fine-tuning" phase [32]. Similarly, Zhang et al. demonstrate this concept as "zero-shot fine-tuning", where a pre-trained model is adapted to new tasks using physics-informed losses, highlighting that a good initialization from pre-training enables rapid adaptation with minimal data [65]. To make this process computationally efficient, Wu proposes an approach that freezes the majority of the pre-trained DeepONet and fine-tunes only a small fraction of the parameters (FTO-PINN), accelerating the adaptation process while maintaining high fidelity [61].

Alternatively, sparse observations, analogous to the PIV measurements in the hemodynamic simulator, can guide the fine-tuning. Addressing the more complex challenge of conditional shift, where the relationship between inputs and outputs changes between training and testing, Goswami et al. introduce a sophisticated hybrid loss function [17]. Their method combines a standard regression loss on the few available labeled target samples, sparse observations, with a loss term that minimizes the statistical distance between the model's predictions and the unlabeled target data, allowing the model to adapt effectively to a new domain with very limited information. The work by Zhu et al. proposes and evaluates methods to reliably fine-tune a pre-trained DeepONet when either the governing PDEs or sparse new observations are available for a new, unseen input [66]. They analyze the extrapolation behavior of DeepONets and demonstrate that both types of case-specific information can substantially improve prediction accuracy for out-of-distribution inputs.

2. Methodology

2.1. Computational Fluid Dynamics

All CFD simulations are conducted in the commercial software Simcenter STAR-CCM+ v2406 (19.04.009). The fluid properties are prescribed to match the 30 vol% glycerol 70 vol% water mixture used in the simulator by Karl et al. [27, 28]. Matching the assumptions of their work, the fluid is modeled as an incompressible Newtonian fluid. At 20 °C the density and dynamic viscosity amount to constant values of $\rho = 1086.0\,\mathrm{kg/m^3}$ and $\mu = 2.9961 \cdot 10^{-3}\,\mathrm{Pa\,s}$ [59]. The available segregated flow solver of second order is used for the mass and momentum balance. It solves the balance equations sequentially and utilizes a SIMPLE algorithm. The under-relaxation is set to 0.8 and 0.2 for velocity and pressure, respectively. These are conservative values that ensure convergence in most cases [52]. In the transient cases, the implicit unsteady scheme is used with a maximum CFL number of 0.5 that is maintained for second-order methods through adaptive time steps. The calculation moves to the subsequent time step once an asymptotic criterion based on the orifice mass flow is met. This criterion ensures the mass flow has stabilized within a specified relative range before the simulation advances, using a time delta that satisfies the CFL condition.

In all simulations, the (U)RANS approach is followed, which describes the flow by solving for its time-averaged fields rather than time-dependent instantaneous fields with chaotic fluctuations in turbulent cases. This method is practical for technical applications and parameter studies because it is associated with drastically less numerical effort than scale-resolving simulations like Direct Numerical Simulations (DNS) or Large Eddy Simulations (LES) at the cost of accuracy. The time-averaging process introduces unknown terms on the NSE, the Reynolds stresses $\tau_{ij}^{\mathrm{turb}}$, that require additional turbulence models to close the system of equations. For incompressible flow, the governing equations take the following form [8]:

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial \left(\overline{u}_i \overline{u}_j\right)}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j \partial x_j} - \frac{\partial \tau_{ij}^{\mathrm{turb}}}{\partial x_j} \quad \text{and} \quad \frac{\partial \overline{u}_j}{\partial x_j} = 0 \tag{2.1}$$

In the steady RANS equations, the first time-derivative term is simply equal to zero. The overline $\overline{\square}$ denotes the time-averaged value and stems from the Reynolds decomposition. In the unsteady URANS case, $\overline{u}_i$ and $\overline{p}$ remain time dependent. For simplicity, this overline is omitted in the rest of this work. The Menter k-ω-SST-model [42] is chosen as a robust and inexpensive two-equation turbulence model with adaptive wall-modeling capability. The STAR-CCM+ user guide contains details of these implementations [52]. Initially, a fully wall-resolved DNS was pursued; however, maintaining the CFL in the cells near the orifice wall during high jet velocities proved costly. Additionally, the experimental PIV data used for validation consists of phase-averaged images to reduce noise and capture the cyclic-steady state [29, 43], a theoretical similarity to URANS. Readers interested in a more detailed discussion of CFD theory and turbulence modeling are directed to the fundamental works by Ferziger et al. [13] and Pope [48].

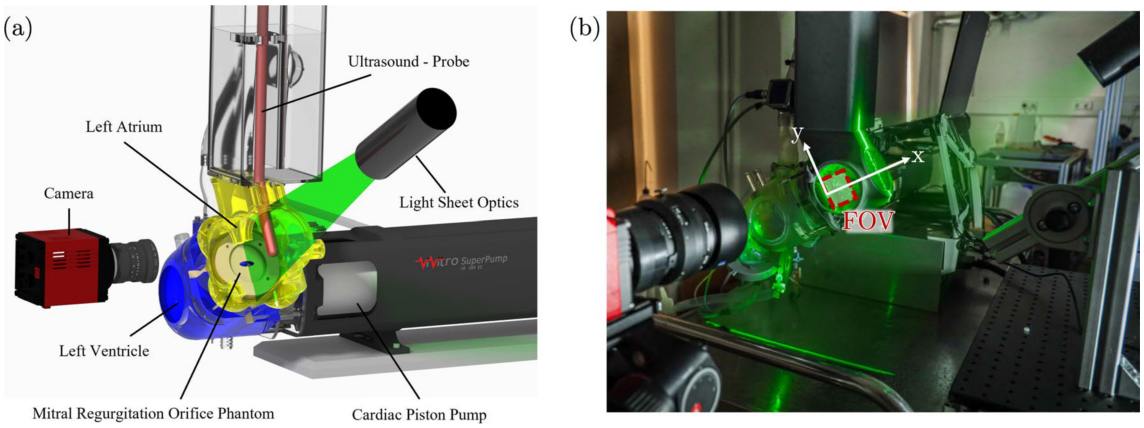

Figure 2.1.: Hemodynamic simulator schematic (a) and photograph of PIV setup (b). [29]

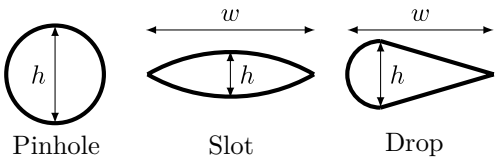

Figure 2.2.: MROP shapes with height and width markers. Adapted from [29].

2.1.1. 3D-Transient Hemodynamic Simulator

The hemodynamic simulator and its CFD simulation, referred to as "3D-Transient", are discussed first since the simplified setups, particularly the BCs and geometry in the channel geometry, are derived from here. Please refer to the development studies by Karl et al. [27, 28] and the flow analyses by Leister et al. [29, 43] for information extending beyond the following discussion. Additionally, the ViVitro Labs Inc. cardiac piston pump's function is described in a collection of user manuals [21, 22, 23]. The reader is referred to the corresponding publications for further details.

2.1.1.1. Setup

The simulator test rig for PIV measurements is depicted in Figure 2.1. Between the LV and LA, a so called Mitral Regurgitation Orifice Phantom (MROP) is installed. MROPs are fixed orifices, manufactured from a 0.5 mm thin PVC film, replicating the insufficient closure of the MV during mid-systole. They are created in various generic shapes (Pinhole, Slot, Drop) and severity sizes (S, M, L) as shown in Figure 2.2. Their purpose is to create reproducible regurgitation jets and to assess the impact of different orifice shapes on them. The main downside is that MROPs simulate only a fixed orifice state, as they cannot open and close like a natural MV. This necessitates removing the aortic valve in the heart simulator setup to allow for left ventricle refill more easily during diastole.

Table 2.1.: MROP dimensions for CFD simulations, based on [29]. For eccentric configuration, the values describe the oval that is created on the flat surface as a result of the angle.

Orifice	PinholeS	PinholeM	PinholeL	SlotS	SlotM	SlotL
$h \times w$ in mm	4.7	8.7	12.2	3.3×11.1	4.5×14.0	7.3×22.9
Orifice	DropS	DropM	DropL	DropXL	SlotL-Bent	EccJet
$h \times w$ in mm	4.3×9.7	6.5×13.7	9.0×19.8	13.0×27.6	7.3×22.9	8.0×10.0

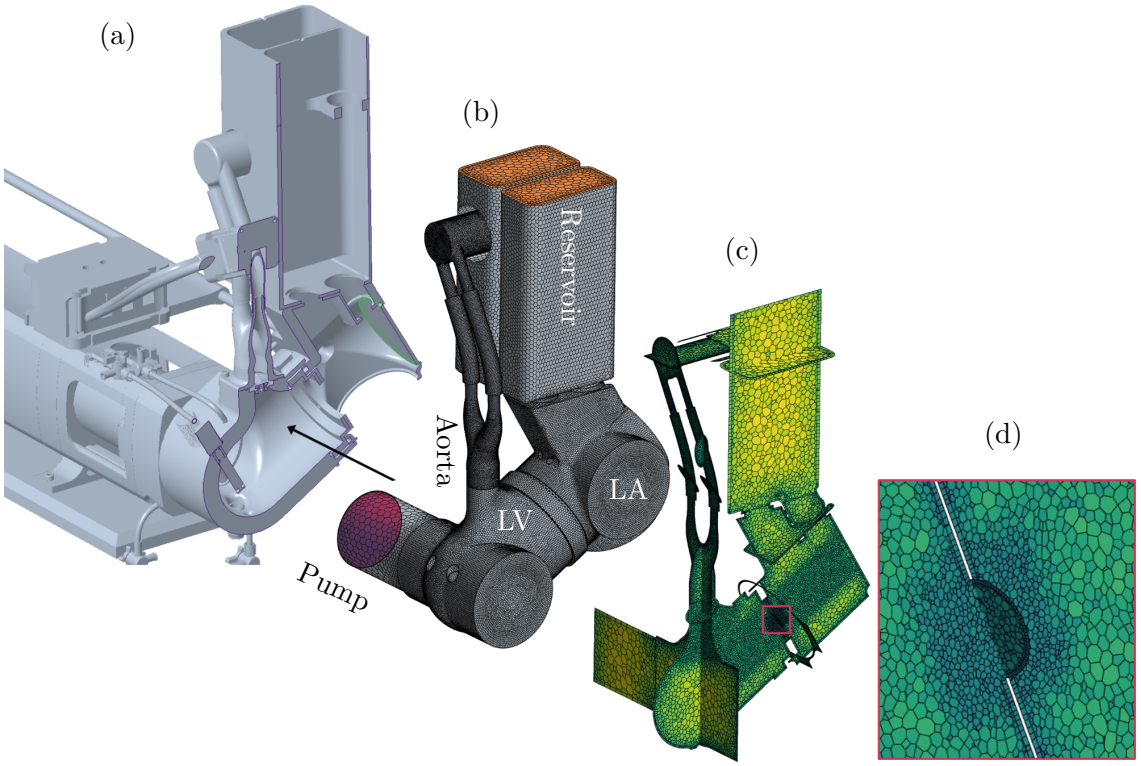

Figure 2.3.: Hemoydynamic simulator geometry. (a) Cut open CAD model, (b) meshed internal fluid domain, (c) mesh slices, (d) magnified PinholeL orifice mesh.

The thin polymer is flexible and temporarily deforms into a spherical dome due to the pressure difference between the LV and LA. This dynamic is not modeled in CFD due to the stability impact of Fluid Structure Interaction (FSI) on the solver. Instead, the foil is considered a regular, rigid wall. To assess the impact of this, an additional simulation with the maximum spherical deformation during systole as the rigid wall is completed for the largest slot geometry. Additionally, a fourth, extra-large size for the drop geometry was computed. Finally, a 12 mm thick 3D-printed MROP incorporating an eccentric circular through-hole with a diameter of 8 mm is analyzed. The hole's axis is tilted 37.5° relative to the surface normal of the MROP, which is the jet direction in all previous cases. This produces a tubular passage that exits the opposite side, which faces the LV interior off-center. The thickness allows the flow to evolve, exiting into the LA from the same location as the thin MROPs. The orifice dimensions of all 12 3D-Transient CFD configurations are listed in Table 2.1. Of those, DropXL and SlotL-Bent are selected for the ML validation dataset, enabling the assessment of out-of-distribution or "extrapolation" cases. Note that, contrary to what Figure 2.1(a) displays and Figure 2.2 suggests, the shapes are actually aligned vertically in the PIV and therefore CFD data.

The 3D model for the internal fluid domain is derived from the CAD model of the hemodynamic simulator, which is shown in Figure 2.3(a) and (b). It is simplified in a few noncritical locations and meshed in STAR-CCM+ using a polyhedral mesher with a single prism-layer on wall surfaces. Localized mesh refinements are employed to optimize the trade-off between spatial resolution and computational expense, prioritizing higher cell densities in regions of interest and areas characterized by high velocity magnitudes or steep flow gradients. This is displayed in Figure 2.3(b) to (d). The reservoir above the LA and the cylinder connecting the LV to the pump are meshed more coarsely since they are not regions of primary interest and low flow velocities are expected. Meanwhile, the aorta connecting the LV to the reservoir has a finer mesh to accommodate the small diameter of

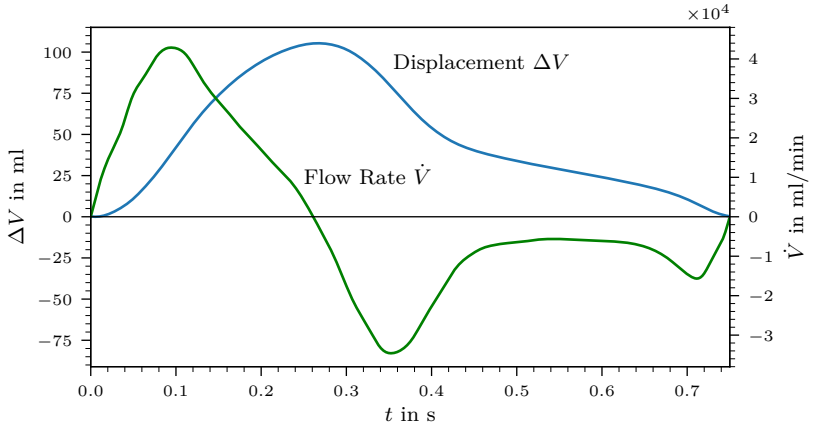

Figure 2.4.: Cardiac pump displacement (blue) and flow rate (green) for PinholeL.

the tubes, and the cylindrical zones in the LV and LA where jets form due to the orifice are locally refined to resolve this flow feature accurately. Additional refinement is applied around the opening to support accurate flow development near the sharp geometry. A mesh independence study has been carried out and is described in the Appendix A.1. Based on the study, each mesh for the different MROPs contains roughly 400,000 cells and about 2,500,000 faces. To characterize the near-wall mesh resolution relative to the boundary layer structure, the non-dimensional wall distance y^+ is defined as:

$$y^+ = \frac{u^\star\, y}{\nu} \tag{2.2}$$

where y is the normal distance to the wall. Consistent with the adaptive wall modeling approach employed in this work, $u^\star$ is a hybrid velocity scale. In the viscous sublayer, this scale recovers the standard definition of the friction velocity based on the wall shear stress ($u^\star \approx u_\tau = \sqrt{\tau_w/\rho}$); elsewhere, it blends with a turbulent scale based on kinetic energy across the buffer and logarithmic regions. The specific proprietary formulation of this blending is omitted here, but full implementation details are available in the STAR-CCM+ user guide [52]. The SlotL mesh, for example, includes 443,353 cells, where at each timestep, the spatial maximum y^+, meaning the highest y^+ value among all wall-adjacent cells at that instant, varies between 4 and 27 on the foil wall, depending on the phase of the periodic flow. On all other walls, the spatial maximum y^+ ranges between 11 and 61.

During the in-vitro experiment, physiological pressures are created by the cardiac piston pump. The pump's cylinder movement is controlled by a waveform, one cardiac cycle of which is shown in Figure 2.4. All cases feature a heart rate of 80 beats per minute (BPM) which corresponds to a period of 0.75 s. For each geometry, the displacement amplitude is adjusted until a peak ventricle pressure of 120 mmHg is reached during systole in the experiment. Larger orifices produce lower resistance and therefore require a higher mass flow to reach the same peak pressure. In CFD, the pump is modeled as a mass flow BC on the fixed circular surface highlighted in red in Figure 2.3(b). For this, the waveform is scaled to match the displacement amplitude as required in the experiment, for instance, 106.6 ml in the PinholeL case. The scaled waveform is then derived over time to obtain the volume flow rate, also shown in Figure 2.4. This is converted to mass flow with ρ and prescribed at the pump BC surface via a table interpolated by STAR-CCM+ to match the simulation's phase within the cycle. The reservoir is simplified as a pressure outlet, which are the orange surfaces in Figure 2.3(b). All other surfaces are no-slip walls. The simulations are initiated with zero values for all quantities. Therefore, at least two full cardiac cycles need to be calculated to create a starting point for reasonable results and

looping playback in the second one. This behavior is also shown for 2D-Transient in 2.10. Using all cores on an AMD EPYC 7662 64-Core Processor, one cycle requires approximately 60 hours, depending on the MROP case and corresponding mesh. CFD data is exported every 2.5 ms, creating 300 discrete snapshots over a single cycle.

2.1.1.2. Validation

The quality of the 3D-Transient CFD simulations is assessed using several quantities measured during the experiments. Firstly, the phase-resolved PIV recordings allow the comparison of velocity fields. For the PinholeL MROP, a snapshot of velocity magnitudes for PIV and CFD is shown in Figure 2.5(a) and (b). The other MROPs where such PIV data is available are PinholeL, SlotL, DropXL, and EccJet. Similar snapshots of these are presented in A.2, including a comparison between the SlotL and SlotL-Bent results. Among these, the velocity magnitude of the regurgitation jet tends to be overestimated by CFD, while the shape generally agrees well. Moreover, the jet velocity usually rises earlier and drops later than observed with PIV. PinholeL represents the best result, where the velocity magnitude of PIV is closely matched by the URANS simulation. The accuracy of other time-steps is approximated by the comparison of maximum velocity magnitudes plotted over the cardiac cycle in (d) of these figures. After around 0.25 s, these lines diverge due to PIV not being able to capture the opposite jet that forms during diastole inside the LV.

For the comparisons, the phases of PIV and CFD are synced up manually, so some tolerance for phase misalignment should be considered. This also applies to Figure 2.7, in which all Slot MROP results are compared based on mass flow and pressure measurements. The magnitudes of these are similar between all three sizes, since the same physiological pressure was established as previously discussed. The aorta of the hemodynamic simulator branches off into two rubber tubes through which the mass flow is measured ultrasonically. The general shape and amplitudes of these measurements are matched by CFD results shown in 2.7(a). The aorta mass flow during regurgitation in systole, seen in the negative peak, is consistently underestimated. Pressure values are recorded in one wall location each for the LV and LA. The ventricle pressure is shown in Figure 2.7(b). Similarly to the flow rate, the systole pressure peak is smaller, and the diastole peak is larger for the simulation. Here, the experimental data also shows more local extrema, but the overall trend is resolved. Conversely, the simulation features more local extrema for the atrium pressure, as shown in figure 2.7(c). It seems possible that the flexible pinhole foil mediates these fluctuations in the experiment, while the stiff wall is unable to do so in the current simulation setup.

The higher velocities, lower aorta mass flow, and ventricle pressure suggest that the MROPs create less resistance in the CFD cases. This also manifests in the overall regurgitation volume (RVol) being consistently overestimated compared to PIV and clinical techniques, as shown in Figure 2.6. However, the simulations follow the trend of increasing severity based on MROP size closely. In conclusion, while there are quantifiable differences, the URANS model is reasonably accurate and captures the key transient and topological features of the experimental flow. Therefore, it serves as a valid proxy for the real-world physics for the purpose of developing the ML frameworks in the subsequent chapters.

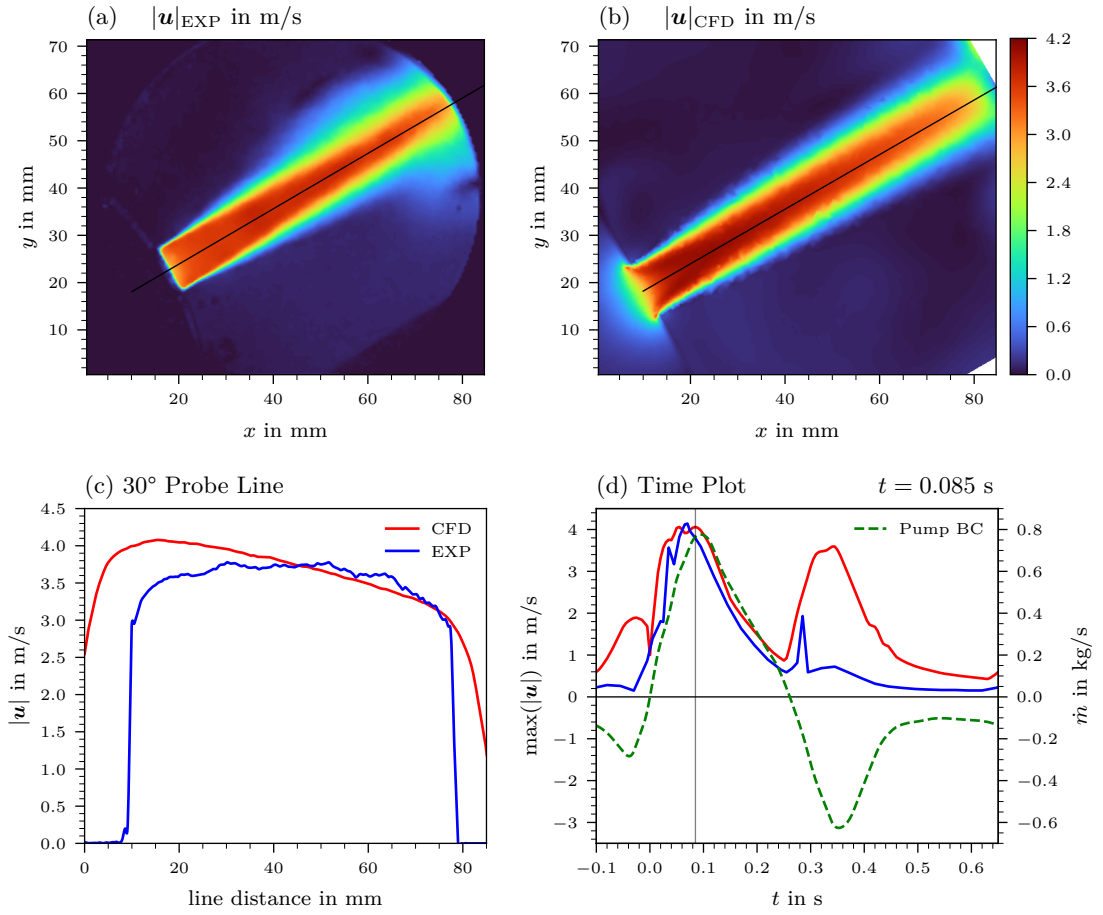

Figure 2.5.: Comparison of instantaneous experimental PIV and CFD results for PinholeL at $t = 0.085\,\mathrm{s}$. A line in direction of the regurgitation jet is positioned in (a) and (b) to probe the velocity magnitudes, starting in the center of the orifice in the foil's undeformed state. The values probed along them are shown in (c). The pump BC and the resulting maximum velocity magnitudes are shown in (d). A vertical line marks the time-step of (a) to (c).

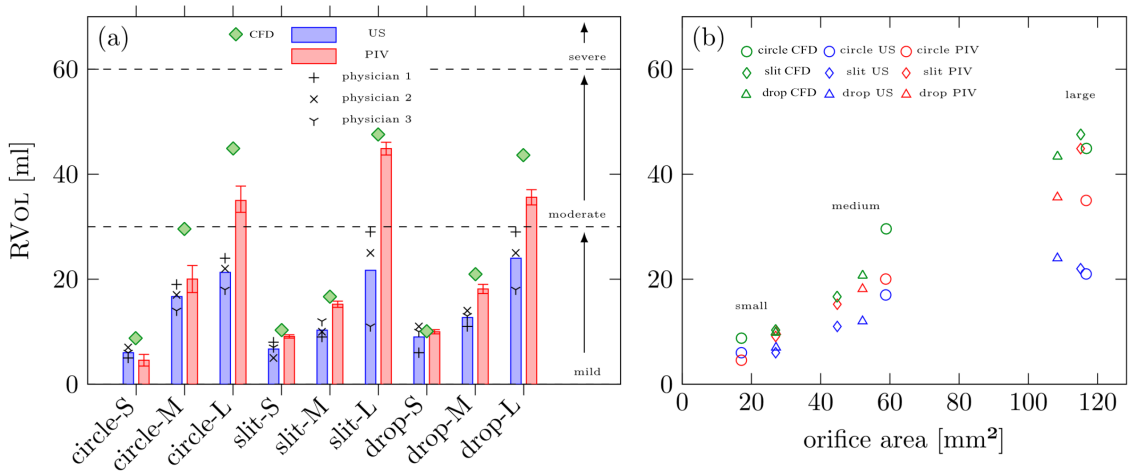

Figure 2.6.: Regurgitation volume (RVol) of each MROP measured by ultrasound (US) based on three physicians and PIV with uncertainty confidence intervals. Adapted from Leister et al. [29] to include CFD results in green.

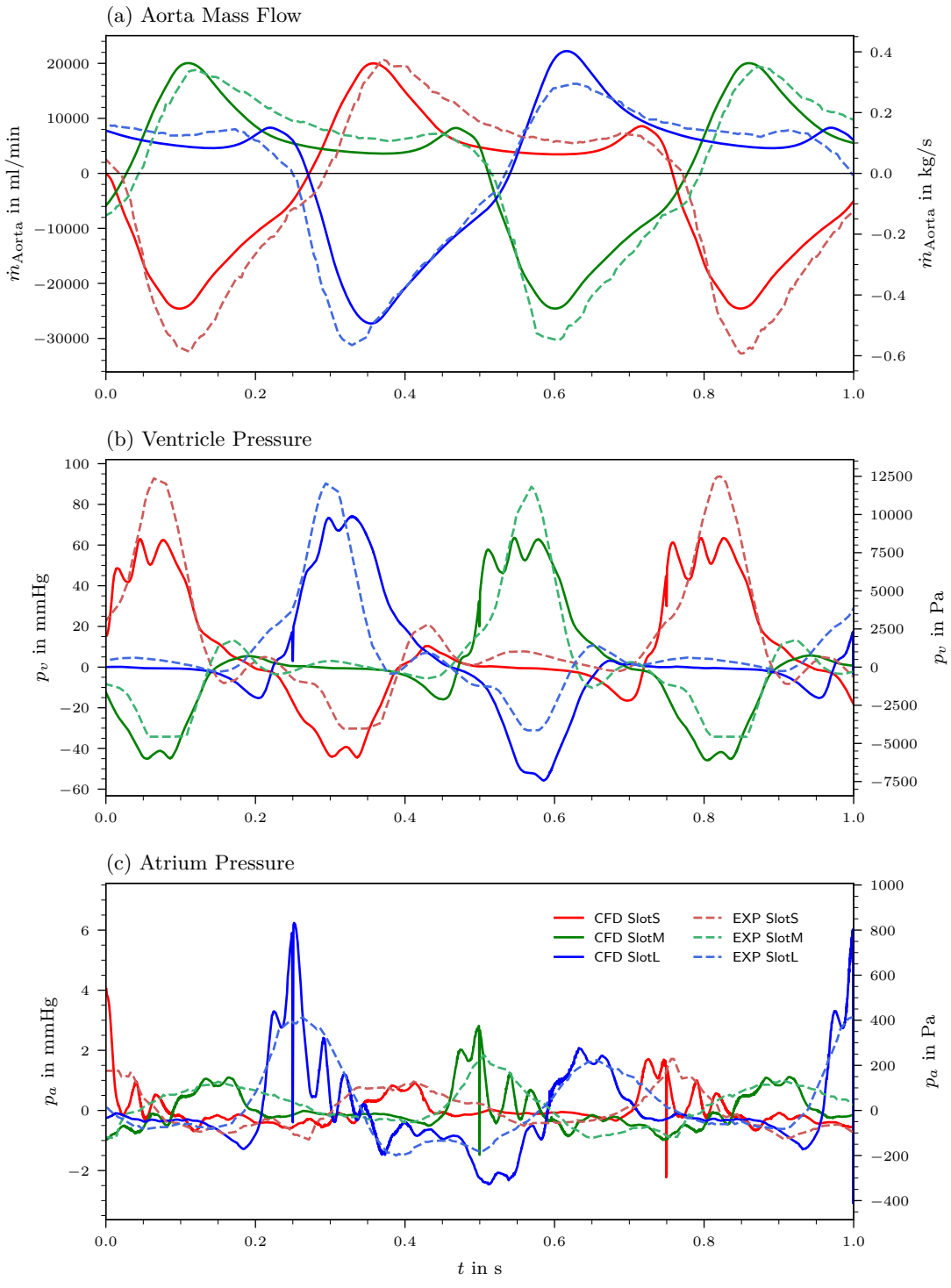

Figure 2.7.: Combined mass flow measured through the two aorta tubes and pressures observed in the LV and LA in the experiment and by CFD for the three sizes of slot MROPs. Two unit scalings are shown for each. Sizes are offset by 0.25 s to make the presentation clearer.

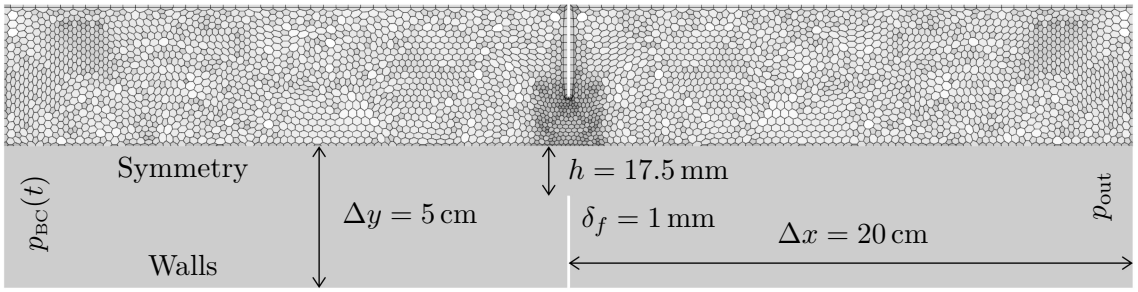

Figure 2.8.: 2D-Transient mesh for exemplary half-orifice-height h displayed in the upper half, BCs and dimensions in the lower half.

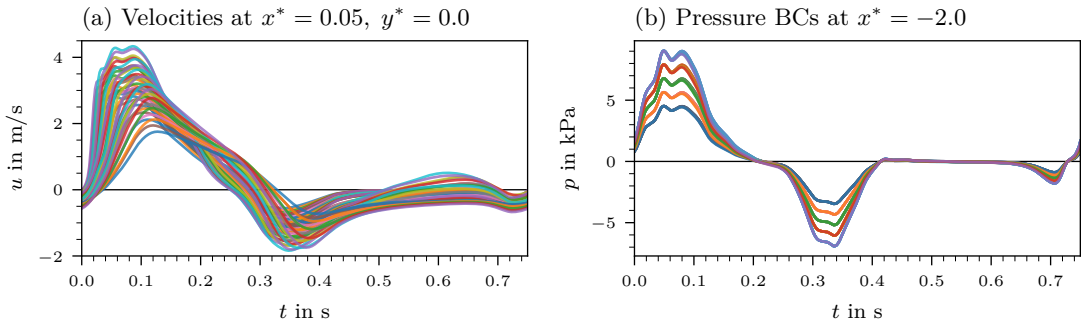

Figure 2.9.: Velocity component in the x-direction sampled near the orifice opening and displayed in (a) for all 45 2D-Transient cases. Pressure BCs at the inlet are also shown in (b). Coordinates non-dimensionalized by $l_c = 0.1\,\text{m}$.

2.1.2. 2D-Transient Channel

The purpose of this setup, referred to as "2D-Transient", is to provide a dataset to train and validate models on pulsatile flow that mimics the cardiac cycle while reducing the complexity and associated computation cost. The domain consists of a channel with a central orifice of varying size. By using a symmetry BC along the x-axis, only half of the geometry needs to be modeled. The variable h describes the half-orifice-height and covers the range $h \in [2.0\,\text{mm},\ 17.5\,\text{mm}]$ in 9 steps, thereby slightly extending the minimum and maximum 3D-Transient MROP dimensions. The geometry is meshed in STAR-CCM+ using a polygonal mesher with a single prism-layer on wall surfaces. Again, the area around the opening is refined to support accurate flow formation near the sharp geometry, with the refinement area size automatically adjusted according to h. For the largest h, this results in 3,960 cells and 11,433 faces, which is displayed in Figure 2.8. All defining dimensions and BCs are also shown. The foil thickness δ_f is doubled to 1 mm to improve mesh quality at the selected coarseness.

Initial tests with a prescribed mass flow waveform like in 3D-Transient showed stability issues due to the short CFD domain. The setup was revised to use a pressure drop curve instead. The ventricle and atrium pressure curves, analogous to 2.7(b,c), are extracted for the PinholeL case since it is the most accurate case according to the validation of 3D-Transient. The pressure difference is calculated by subtraction and then scaled such that the total amplitude of the pressure waveform corresponds to a selected $\Delta p \in [8\,\text{kPa},\ 16\,\text{kPa}]$ value, creating 5 different $p_{BC}(t)$. These are displayed in Figure 2.9(b). The unscaled Δp for PinholeL is 15.598 kPa. The outlet pressure p_{out} is fixed at zero.

Together with the varying orifice sizes, this creates a distribution of 45 unique URANS

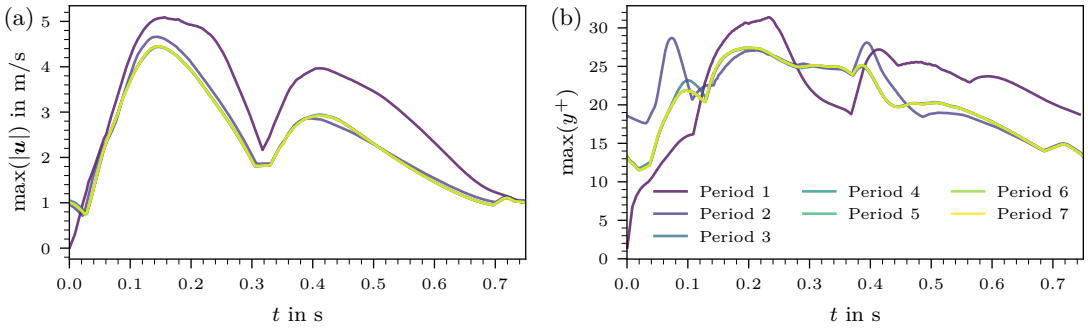

Figure 2.10.: Maximum velocity magnitude (a) and wall y^+ values (b) over 7 cardiac cycles for the 2D-Transient $h = 17.5\,\text{mm}$, $\Delta p = 14\,\text{kPa}$ case.

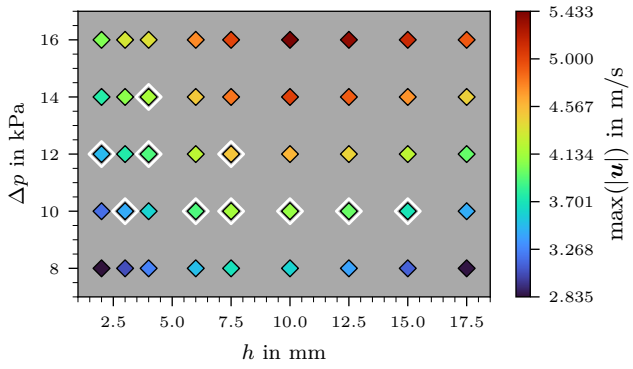

Figure 2.11.: Distribution of 2D-Transient cases with varying pressure BC ranges and orifice heights. The highest velocity magnitude of all time-steps decides the case's marker color. Training set highlighted by white border.

simulations of pulsatile flow. Matching 3D-Transient, this setup assumes 80 BPM, saves 300 snapshots per case and requires multiple cardiac cycles to evolve from the zero-value initial condition. The seventh period is used to achieve strong similarity between the start and end. How the jet velocity and wall y^+ for the mesh evolve is shown in Figure 2.10.

To approximate the number of available 3D-Transient cases, only 10 are included in the ML training dataset. They are selected manually to bias towards a specific pressure amplitude since that aspect is largely matched in the hemodynamic simulator and to allow for validation inside and far outside the training distribution. This selection and the resulting maximum velocity magnitudes for all cases are displayed in Figure 2.11. Figure 2.9(a) also illustrates how the temporal characteristics vary between cases, such as changes in flow direction or differences in the timing of peak values, as well as variations in jet velocities.

2.1.3. 2D-Steady-State Channel

This setup is a further reduction in complexity and referred to as "2D-Steady". It establishes a foundational test case for the ML models in a steady-state environment, isolating the spatial challenge from temporal complexities. Largely the same domain and BCs are used as in 2D-Transient. The inlet pressure BC is now a stationary value and so Δp refers to the static pressure drop between the inlet and zero-pressure outlet. Furthermore, the

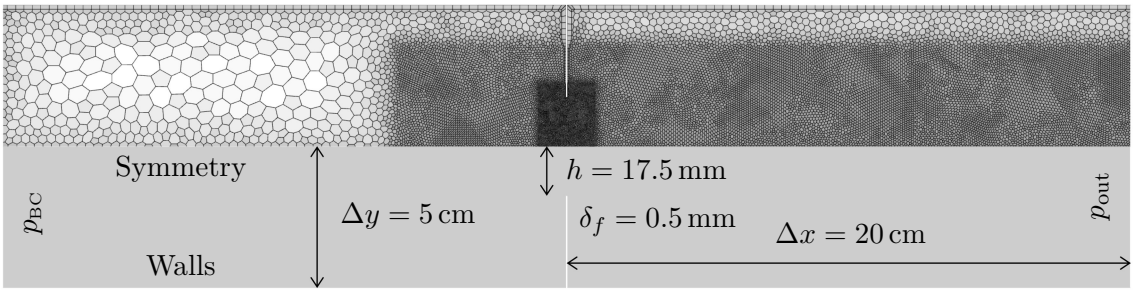

Figure 2.12.: 2D-Steady mesh for exemplary half-orifice-height h displayed in upper half, BCs and dimensions in lower half.

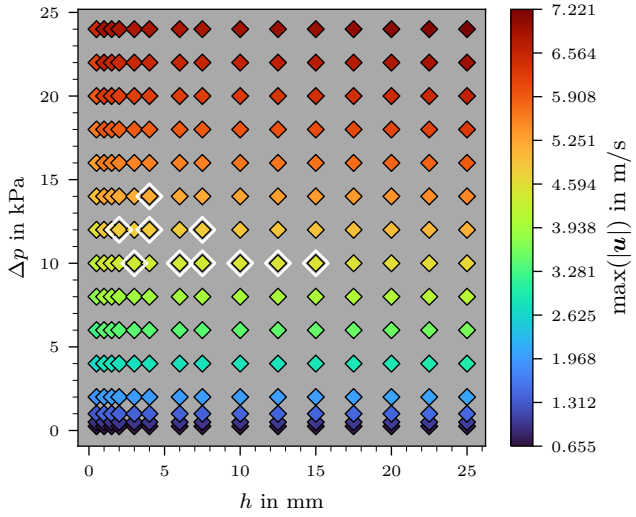

Figure 2.13.: Distribution of 2D-Steady cases with varying pressure BCs and orifice heights. The maximum velocity magnitude decides the case's marker color. White borders highlight the training set. The lowest, harder-to-discern parameters are $\Delta p = [250\,\text{Pa},\ 500\,\text{Pa}]$ and $h = [0.5\,\text{mm},\ 1.0\,\text{mm},\ 1.5\,\text{mm}]$.

foil thickness matches 3D-Transient again at $\delta_f = 0.5\,\text{mm}$, since the mesh can easily be refined for higher resolution at little overhead. Because the locations of large velocity gradients persist, the 2D-Steady meshes feature more heterogeneous refinement compared to 2D-Transient. This slightly adjusted domain is shown in Figure 2.12; the specific example mesh has 17,560 cells and 51,844 faces.

Thanks to the low computation cost, the distribution of cases is expanded to cover $h \in [0.5\,\text{mm},\ 25\,\text{mm}]$ and $\Delta p \in [250\,\text{Pa},\ 24\,\text{kPa}]$ in 15 steps each, creating 225 unique RANS results. Figure 2.13 gives an overview of the resulting regurgitation jet velocity magnitudes. The same 10 parameter combinations are chosen as the primary training dataset as before. Note that the Δp definition differs slightly compared to 2D-Transient, where it means the distance between the maximum and minimum value of the waveform. Here, it is more analogous to just the maximum value, so at the same Δp-value, the maximum jet velocity is expected to be higher.

2.2. Physics-Informed Machine Learning

The core of this thesis is the development and comparison of ML models to reconstruct fluid flow fields from sparse data. Two primary approaches are investigated: Physics-Informed Neural Networks and (Physics-Informed) Deep Operator Networks, both implemented using the NVIDIA PhysicsNeMo Symbolic framework (v1.5.0, formerly NVIDIA Modulus Symbolic) [45]. Readers are encouraged to consult the PhysicsNeMo Sym sub-module documentation [44]. It introduces the theoretical foundations, provides practical examples, and cites academic literature for the various architectures and physics-informed training methods implemented within the framework. Note that this resource covers integrated tools rather than a comprehensive survey of all state-of-the-art methods. While subsequent references to individual methods cite only their original sources, they also appear in the repository documentation. For additional background on ML theory and principal methods, readers may consult the works by Bishop [3], James et al. [26], and Rabczuk et al. [49].

A common set of architectural and training components is used across different models. These are now defined before going into more detail on the two approaches and finally the exact ANNs, which are then evaluated in the following Chapters 3 and 4.

2.2.1. Shared Model Components

The foundational network for all models is a Multilayer Perceptron (MLP, feed-forward and fully-connected) with adaptive [25] SiLU (Swish) [11] activation functions. The AdamW optimizer [36] adjusts the models' internal parameters, i.e. weights and biases, to minimize the composite loss function, thereby making the output more accurate. The initial learning rate of 0.001 is multiplied by 0.95 every 4000 steps, which represents exponential decay. To balance the influence of the competing loss terms, Neural Tangent Kernel (NTK) based dynamic weighting [58] is employed. The weights are updated every 100 training steps. This method based on kernel eigenvalues has the theoretical basis for surpassing empirical gradient-balancing approaches, which it achieved in early tests of this work. Maximum training iterations are set for each model and training phase based on performance in initial trial runs. This manual selection is guided by monitoring of multiple validation metrics, focusing on optimal results and general convergence. This manual termination strategy is chosen to prevent performance degradation from potential overfitting during excessive training. An automated stopping criterion is currently not implemented due to the need to balance complex performance indicators, including both quantitative scalar metrics and qualitative visual assessments of many setups.

Among the research on turbulence modeling for PINNs mentioned in the State of the Art Section 1.3.1, Pioch et al. compare several approaches, one of which is a simple turbulent viscosity formulation [47]. In this setup, the turbulent viscosity ν_t is treated as an additional network output rather than as the solution of a transport equation. This means no extra PDE residuals must be included in the composite loss term, and it keeps the model simple to implement. For instance, introducing the two additional complex PDEs for k and ω and properly handling wall-distance is not trivial and might affect the ANN optimization process. Additionally, the ν_t formulation allows for the inclusion of data supervision from (U)RANS simulations for this ν_t output. When ν_t is not available or supplied as training data, the model still enables closure of the Reynolds stresses; the weights of the ANN are adjusted in the training process such that the predicted output quantities together satisfy the RANS equations. The results of Pioch et al. indicate that the ν_t model performs competitively and could be a viable alternative to the traditional formulations. For these reasons, this equation-free model is selected to address the RANS modeling of the ground truth data.

All physical quantities α are non-dimensionalized using a characteristic value α_c before being supplied to the ANNs for training and inference, meaning all input and output layers use non-dimensionalized values. This step aligns the magnitudes of different quantities and brings their largest values close to unity, which may improve convergence behavior. For PINNs, non-dimensionalization has been shown to result in greater consistency of predictions, offers computational benefits, and may improve extrapolation ability [39]. Denoted by $\square^*$, the non-dimensionalized quantity is then defined as $\alpha^* = \alpha/\alpha_c$. The characteristic values of the present application are chosen as

$$l_c = 0.1\,\mathrm{m}, \quad t_c = 0.75\,\mathrm{s}, \quad u_c = \max(u_i), \quad p_c = \max(p), \quad \nu_{t,c} = \max(\nu_t) \tag{2.3}$$

For example, the highest velocity component value of all CFD snapshots of 3D-Transient defines u_c for that 3D-Transient ANN. For density and fluid viscosity, the dimensionless quantity can be set equal to one, since these quantities are assumed to be constant and the characteristic quantity is therefore equal to the dimensioned quantity:

$$\rho_c = \rho = 1086.0\,\mathrm{kg/m^3}, \quad \mu_c = \mu = 2.9961 \cdot 10^{-3}\,\mathrm{Pa\,s} \quad \text{where} \quad \nu = \mu/\rho \tag{2.4}$$

Based on these, the following non-dimensional numbers are calculated as

$$\mathrm{Re} = \frac{\rho_c\ u_c\ l_c}{\mu_c}, \quad \mathrm{St} = \frac{l_c}{u_c\ t_c}, \quad \mathrm{Eu} = \frac{p_c}{\rho_c\ u_c^2}, \tag{2.5}$$

which then leads to the following constant-density, variable-viscosity form of the NSE

$$\mathrm{St}\,\frac{\partial u_i^*}{\partial t^*} + u_j^*\frac{\partial u_i^*}{\partial x_j^*} = -\mathrm{Eu}\,\frac{\partial p^*}{\partial x_i^*} + \frac{1}{\mathrm{Re}}\frac{\partial}{\partial x_j^*}\left[\mu_{\mathrm{eff}}\left(\frac{\partial u_i^*}{\partial x_j^*} + \frac{\partial u_j^*}{\partial x_i^*}\right)\right] \quad \text{and} \quad \frac{\partial u_j^*}{\partial x_j^*} = 0 \tag{2.6}$$

with the x, y and z-momentum equations for $i \in [1, 2, 3]$ and the continuity equation. These are the PDEs which may individually be included in the composite loss terms. The formulation retains a variable effective viscosity, μ_{eff}, to account for the effects of turbulence within the (U)RANS framework. This is achieved using the Boussinesq hypothesis, which introduces the turbulent viscosity to model the Reynolds stresses by analogy to the viscous stresses [48]. The non-dimensional effective viscosity is therefore the sum of the non-dimensional fluid and turbulent viscosities:

$$\mu_{\mathrm{eff}} = \mu^* + \mu_{\mathrm{t}}^* = 1 + \frac{\nu_{t,c}}{\nu_c} \cdot \nu_t^*(x_i, t) \tag{2.7}$$

where ν_t^* is the non-dimensional spatial and temporal distribution of the turbulent viscosity field predicted by the ANN. If this turbulence model is disabled or there are locally no turbulence effects, μ_{eff} simply becomes 1.

By default, PhysicsNeMo Sym uses L2 loss that computes the sum of squared errors (SSE) between the model's predictions $\hat{y}_k$ and the ground truth target data y_k on k discrete points with optional pointwise weights λ_k:

$$\mathcal{L}_{\mathrm{L2}} = \sum_k \lambda_k \left|\hat{y}_k - y_k\right|^2 \tag{2.8}$$

Some trained ANNs discussed in this work use spatial pointwise weighting of losses based on wall-distance. By smoothly lowering the weights to eventually 0 on the wall, this method prevents steep solution gradients from dominating the PDE loss function at sharp corners. This approach may accelerate convergence and improve overall accuracy [44], but it didn't seem to significantly affect results here.

The outputs of the PINNs and DeepONets are differentiable with respect to their input coordinates via automatic differentiation, allowing the inclusion of the PDE equations [18].

By moving all parts of each of the NSE (2.6) to one side, the output target for the loss function becomes zero. The goal is to minimize this resulting PDE residual at collocation points, which generally are specific, often randomly sampled, coordinates in the domain where the PDE satisfaction is promoted, not enforced. Sufficient density of collocation points is required to avoid overfitting. Note that using SSE leads to the absolute loss magnitude depending on the number of points k, which may affect the loss balancing of different terms. These can be categorized in the following way:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} w_i \mathcal{L}_i(\theta) \subseteq \{\mathcal{L}_{\text{data}},\ \mathcal{L}_{\text{PDE}},\ \mathcal{L}_{\text{BC}}\} \tag{2.9}$$

$$\subseteq \{\mathcal{L}^{\text{train}},\ \mathcal{L}^{\text{val}}\} \tag{2.10}$$

with θ being the internal ANN parameters and w_i the effective weights assigned by the dynamic NTK calculation. While a base weight can be assigned before the NTK process, only very large or small values significantly sway the final effective weights. The subset symbol $\subseteq$ indicates that the elements making up the left-hand side loss term are part of the right-hand side set. Each of the data, PDE and BC sets can be further qualified as either concerning the training cases or validation cases. Thereby, $\mathcal{L}^{\text{val}}_{\text{data}}$ represents the sparse observable information of the simulator known before and measured during experiments for the otherwise unseen application (validation) case, while $\mathcal{L}^{\text{train}}_{\text{data}}$ contains the full supervision of all cases and quantities found in the CFD training dataset. Depending on the domain dimensionality, ANN approach and training phase, loss terms from the following selection are enabled individually for training and validation cases:

$$\mathcal{L}_{\text{data}} \subseteq \{\mathcal{L}_u,\ \mathcal{L}_v,\ \mathcal{L}_{\text{w}},\ \mathcal{L}_p,\ \mathcal{L}_{\nu_t}\} \tag{2.11}$$

$$\mathcal{L}_{\text{PDE}} \subseteq \{\mathcal{L}_{\text{mom,x}},\ \mathcal{L}_{\text{mom,y}},\ \mathcal{L}_{\text{mom,z}},\ \mathcal{L}_{\text{conti}}\} \tag{2.12}$$

$$\mathcal{L}_{\text{BC}} \subseteq \{\mathcal{L}_{\text{noslip}},\ \mathcal{L}_{\text{sym}}\} \tag{2.13}$$

The BC group contains no-slip, meaning zero-velocity on geometry walls, and symmetry terms for 2D-Steady and 2D-Transient:

$$0 = v = \frac{\partial u}{\partial y} = \frac{\partial p}{\partial y} = \frac{\partial \nu_t}{\partial y} \tag{2.14}$$

The no-slip loss is not used with the training cases because the same zero-velocity information is already part of the full flow field supervision.

2.2.2. PINN and DeepONet Architectures

The composition of the validation data loss term, $\mathcal{L}^{\text{val}}_{\text{data}}$, is dictated by the experimental setup of the hemodynamic simulator described in Section 2.1.1. It specifically comprises the sparse quantities obtainable during in-vitro experiments: the phase-resolved PIV velocity fields (u, v) restricted to the optical window, the scalar pressures (p) recorded by sensors in the LV and LA, and potentially the mass flow rate measured ultrasonically at the aorta. Crucially, the distinction between training and validation scopes extends to the physics loss terms as well. $\mathcal{L}^{\text{val}}_{\text{PDE}}$ denotes the residuals of the governing equations calculated for the specific, otherwise unseen application case. This term can support the reconstruction of the full flow field from the sparse $\mathcal{L}^{\text{val}}_{\text{data}}$ by promoting physical consistency across the extensive 4D domain where experimental data is absent. Conversely, $\mathcal{L}^{\text{train}}_{\text{PDE}}$ refer to the physics losses enabled simultaneously for all training cases. These are interwoven with the full supervision data points $\mathcal{L}^{\text{train}}_{\text{data}}$ across the spatiotemporal domain, aiding the network in learning the general solution operator during the initial training phase.

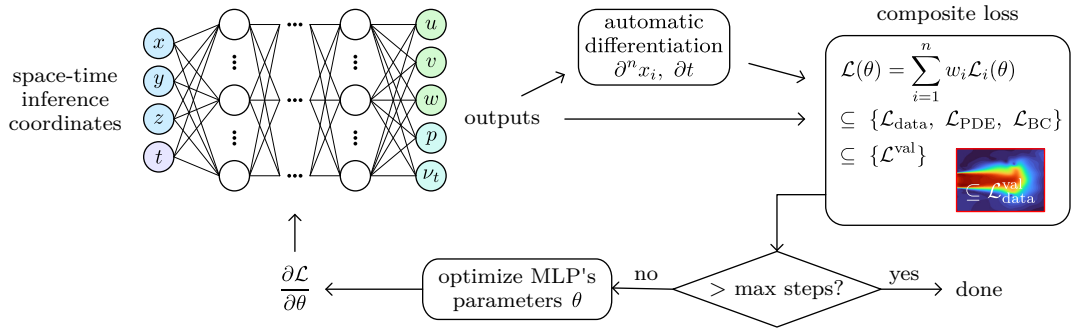

Figure 2.14.: Schematic of PINNs used in this work. Losses concerning a single validation case, including the constrained u, v-velocity field, supplied to solve the ill-posed problem. All inputs and outputs are non-dimensional; notation $\square^*$ is omitted for legibility.

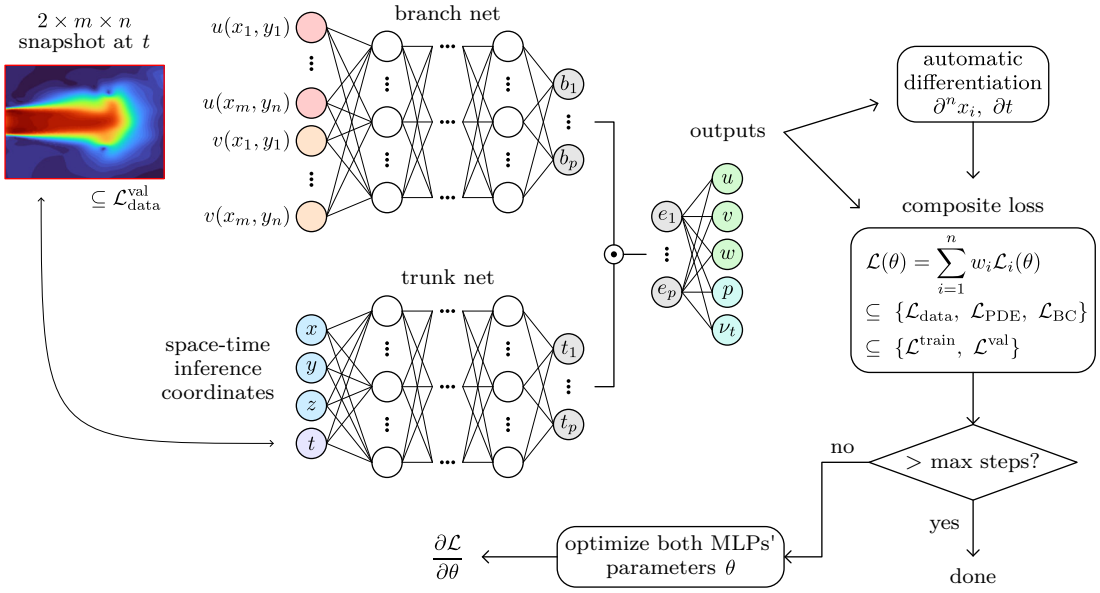

Figure 2.15.: Schematic of DeepONets used in this work. The constrained u, v-velocity field is supplied to the branch net as flattened input for all training cases during training and validation cases during inference of the unseen case. The same constrained field may be included in hybrid training for partial supervision of the observable validation case data. The branch net snapshot's time step is the trunk net input t. Branch and trunk outputs b_i and t_i are multiplied element-wise to form e_i. This output layer size of the branch and trunk is matched to the hidden layers' neurons per layer. All inputs and outputs are non-dimensional; notation $\square^*$ is omitted for legibility.

The manner in which these data and physics terms are utilized constitutes the fundamental distinction between the two machine learning approaches investigated in this work:
In the PINN approach, the model functions as a physics-driven surrogate dedicated to a single flow configuration. Here, the problem is severely ill-posed due to the sparsity of the experimental observations. The PINN architecture employs the governing equations in the form of $\mathcal{L}_{\text{PDE}}^{\text{val}}$ to generate the full 3D transient flow fields based on the limited information provided by $\mathcal{L}_{\text{data}}^{\text{val}}$ and the known boundary conditions $\mathcal{L}_{\text{BC}}^{\text{val}}$. As illustrated in Figure 2.14, the network maps the space-time inference coordinates (x, y, z, t) directly to the flow quantities of interest. A verification of this physics-based implementation using simple 2D channel flow without orifices is provided in the Appendix B.2.

The second approach utilizes DeepONets to learn a parametric operator capable of generating solutions for new configurations without full retraining. This framework is primarily data-driven, relying on a distribution of full simulation results $\mathcal{L}_{\text{data}}^{\text{train}}$ to learn the mapping between an input function and the output field. In this training phase, $\mathcal{L}_{\text{PDE}}^{\text{train}}$ can be included to regularize the solution across the distribution of cases. As shown in Figure 2.15, the architecture splits into a trunk net, which encodes the coordinates, and a branch net, which encodes the input function, defined here as a spatially masked subset of the CFD velocity field analogous to a PIV measurement. While the MLPs for trunk and branch can be tuned completely separately, the hidden layer count and size were kept equal in this work to reduce the hyperparameter space for possible optimizations. As mentioned in Section 1.3.2, utilizing a standard MLP for the branch net necessitates flattening the 2D input fields into a 1D vector, discarding spatial relationships. A CNN represents the more suitable approach for such structured inputs. However, this work proceeds with the MLP architecture as a pragmatic decision to facilitate implementation within the PhysicsNeMo framework. In practice, the flattening takes place by sampling the velocity component values on $m \times n$ discrete points, first for u and then for v. The selection of these points for the discretization of the input function is arbitrary but must remain consistent between cases, since no information of the input function beyond the flattened velocity magnitudes themselves is supplied to the network. The DeepONets learn to map these simple features, representing the flow inside the subset of the domain, with the full flow fields of the entire domain.
Alternative definitions for the branch network input were considered, specifically the utilization of a geometric parametrization of the MROPs or a multi-branch DeepONet architecture capable of fusing heterogeneous data sources such as pressure, PIV, and shape. A single-branch configuration based on the PIV recordings was selected for this work, as multi-branch architectures are not directly implemented in the PhysicsNeMo version used. The experimental velocity field serves as likely the richest source of information, capturing the specific transient state of the flow physics at each instant, anchoring the prediction in observed data. A distinct characteristic of this snapshot-based approach is that the model can only infer full fields for time steps where an input recording exists. Conversely, a geometric input would require complex parameterization to account for the 3D and flexible nature of the orifices.

While the base model infers results near-instantly, a hybrid "test-time adaptation" step can be employed. In this step, the pre-trained operator is fine-tuned using the sparse observations $\mathcal{L}_{\text{data}}^{\text{val}}$ and, optionally, the case-specific physics $\mathcal{L}_{\text{PDE}}^{\text{val}}$ with $\mathcal{L}_{\text{BC}}^{\text{val}}$, effectively bridging the two approaches. It is important to note that in this specific context, the velocity part of the sparse observable data $\mathcal{L}_{\text{data}}^{\text{val}}$ corresponds to the same spatiotemporal subset used to define the input function. Consequently, this dataset serves a dual purpose during adaptation: the flattened velocity magnitudes continue to drive the branch net as the condition, while the specific spatial coordinates associated with these sampling points are now explicitly supplied to the trunk net. This allows the model to evaluate the error

between its predictions and the observed "PIV" values at those exact locations, enabling the supervision.

For flexibility and consistency, development was conducted on networks that always output p and ν_t in addition to the velocity components, even though they are only strictly necessary when using the PDE losses. So, for specific configurations where this is not the case, some improvements may be achievable by reducing the output vector size. Additionally, with full supervision of training cases ($\mathcal{L}_{\text{data}}^{\text{train}}$), any other quantities of interest from CFD could be learned as an additional output.

The subsequent sections from 2D-Steady to 3D-Transient introduce the technical details specific to the various models, for example, what loss terms are enabled for how many training steps. The reasons for certain evolutionary steps and settings are explored in the Validation Chapter 3. The shown models are a subset of tested configurations, chosen to present key findings and ultimately the best results achieved within this work. The reported calculation times resulted from training on a system with an AMD EPYC 7313 16-Core Processor, an NVIDIA RTX 6000 Ada GPU with 48GB VRAM, and 504GB RAM. Due to costly and frequent validation steps during the training used for the development, the training times are not necessarily representative of what a fully optimized framework might achieve.

2.2.2.1. 2D-Steady

The characteristic quantities of 2D-Steady DeepONets, meaning the maximum values observed in the entire CFD dataset, round to:

$$u_c = 7.221\,\text{m/s}, \quad p_c = 23998.7\,\text{Pa}, \quad \nu_{t,c} = 3.794 \cdot 10^{-3}\,\text{m}^2/\text{s} \tag{2.15}$$

Similarly, the maximum values for the single PINN case are:

$$u_c = 5.268\,\text{m/s}, \quad p_c = 13830.4\,\text{Pa}, \quad \nu_{t,c} = 2.772 \cdot 10^{-3}\,\text{m}^2/\text{s} \tag{2.16}$$

name	layers	basis	steps	time	losses
PINN RANS	5×512	-	$300,000$	10.12 h	train: - val: data, noslip, sym, conti, mom
Dist20	4×128	-	$200,000$	5.424 h	train: data val: -
Fourier	4×128	-	$200,000$	12.47 h	train: data val: -
DD	4×128	-	$200,000$	2.814 h	train: data val: -
Sym	4×128	DD	$+100,000$	+2.051 h	train: data, sym val: -
Conti	4×128	DD	$+100,000$	+16.23 h	train: data, sym, conti val: -
RANS	4×128	DD	$+100,000$	+18.552 h	train: data, sym, conti, mom val: -
Hybrid Exact	4×128	Exact	$+2,000$	+0.144 h	train: data val: data, noslip, sym
Hybrid Conti ×3	4×128	Conti	$+2,000$	+0.48 h	train: data, sym, conti val: data, noslip, sym, conti

Table 2.2.: Training configurations for 2D-Steady ANNs. All except "PINN" are DeepONets, for which the hidden layer specification applies to the trunk and branch each. The basis describes which prior model the training continues on, hence the "+" for number of training steps and elapsed time.

With these, the evaluation plots can be re-dimensionalized to real units. Naturally, the architectures shown in Figures 2.14 and 2.15 are adapted so that the inputs and outputs match the steady-state 2D domain by omitting z, t, w.

CFD data is interpolated onto a structured 548×102 grid spanning the domain $[-2, 2] \times [0, 0.5]$. The discretization employs anisotropic spacing ($\Delta x^* \approx 7.31 \times 10^{-3}$, $\Delta y^* \approx 4.95 \times 10^{-3}$). To enhance fidelity near the orifice, three supplementary vertical grid lines are integrated at the centerline ($x^* = 0$) and the immediately adjacent intervals ($x^* = \pm\Delta x^*$), resulting in a total of 56202 points. Even though DeepONets and PINNs can ingest unstructured or raw point cloud data, interpolating onto a structured grid is used here to prevent the spatial bias from the highly heterogeneous CFD meshes. To construct the flattened input vector for the branch network, the interpolated velocity fields are sampled within a restricted subdomain. This region is defined by a spatial mask spanning $x^* \in [0.025, 0.525]$ and $y^* \in [0, 0.25]$, effectively cropping the grid to an area similar to the 3D-Transient experimental PIV field of view. This consistent masking procedure isolates $N = 3,519$ discrete points, which serve as the fixed sampling locations for the input velocity components. Accounting for the two velocity components, twice that number is the total branch net input dimension. The scalar evaluation metric NRMSE, as defined in Appendix B.1, is calculated using all grid points across the 225 cases.

Table 2.2 lists the trained models for which results are discussed. To reiterate, a train-data-loss refers to the supervision of full flow fields from training cases, while val-data-loss refers to the constrained "PIV" velocity field and pressure measurements of the application case. For 2D-Steady and 2D-Transient, the two pressure nodes are positioned in the top left and top right corners, respectively. The Dist20 DeepONet uses a different distribution of 20 instead of 10 training cases. The impact of this change is assessed by comparison against the DD (data-driven) DeepONet. Sets of PDE terms are added in a second training step (Sym, Conti, RANS) to augment the baseline DD model, creating PI-DeepONets. The Fourier DeepONet applies a Fourier feature mapping to the trunk net that transforms low-dimensional input coordinates into a higher-dimensional space using high-frequency sinusoidal functions. This process overcomes the inherent spectral bias of standard MLPs, enabling them to effectively learn high-frequency details and complex functions [53]. The results presented in this work utilize the axis configuration with integer frequencies from 0 to 34, selected for achieving the best performance in preliminary trials, which employs a deterministic sampling strategy along the coordinate dimensions to function as a generalized positional encoding. The loss terms included in the "Hybrid Conti" model result in the best fine-tuning among the tested combinations. For this reason, 3 different cases are test-time-adapted:

name	Extra	Extra2	Inter
h in mm	20	25	6
Δp in kPa	14	24	12

Table 2.3.: Case selection for "Hybrid Conti" 2D-Steady DeepONets.

The "Extra" models evaluate cases that fall outside the training distribution (extrapolation), as shown in Figure 2.13. The "Inter" model tests an interpolation case that is closely surrounded by similar training dataset cases. The same "Extra" case is also test-time-adapted in a DeepONet where the continuity equation is satisfied trivially ("Exact"). Here, the network outputs u, v are replaced by the stream function ψ and instead calculated in a second step via automatic differentiation. With the following velocity component

definitions, the 2D incompressible continuity equation is satisfied automatically:

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x} \quad \Rightarrow \quad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial^2 \psi}{\partial y \partial x} = 0 \tag{2.17}$$

This formulation eliminates the need for a continuity loss term and the associated weighting to balance it against other loss components. However, the approach still requires an additional differentiation pass, which may affect performance and accuracy. The basis model, trained for 200,000 steps using just $\mathcal{L}_{\text{data}}^{\text{train}}$, is not listed in Table 2.2 as its results aren't discussed separately.

2.2.2.2. 2D-Transient

The characteristic quantities of the 2D-Transient DeepONets are:

$$u_c = 5.433\,\text{m/s}, \quad p_c = 9102.3\,\text{Pa}, \quad \nu_{t,c} = 2.488 \cdot 10^{-3}\,\text{m}^2/\text{s} \tag{2.18}$$

The PINN mistakenly uses the same values, resulting in maximum non-dimensional values below rather than exactly 1. This is just cosmetic inconsistency and does not affect the accuracy noteworthily. The network input t is included again compared to the previous flow regime simplification.

For the 2D-Transient dataset, a 228×42 structured grid is employed with anisotropic spacing ($\Delta x^* \approx 1.76 \times 10^{-2}$, $\Delta y^* \approx 1.22 \times 10^{-2}$), resulting in $9,702$ total nodes after including the same centerline refinement. The branch network input mask is also adapted to the same coordinates. Within this restricted subdomain, a diamond-pattern sub-sampling strategy retains every second point. This additional reduction step yields a compact input vector of $N = 305$ fixed sampling locations. The NRMSE metric when concerning the performance over the entire cardiac cycle is calculated using a fifth of the available snapshots, 60 per case. The reduced grid density and temporal subsampling accommodate the coarser source mesh and manage the computational load of the transient dataset (300×45 snapshots).

As shown in Table 2.4, one model is trained using the instance-specific PINN approach as defined in this work but without PDE losses. This gives an additional diagnostic step

name	layers	basis	steps	time	losses
PINN DD	6×512	-	$29,000$	1.271 h	train: - val: data, noslip
PINN RANS	6×512	-	$200,000$	10.96 h	train: - val: data, noslip, sym, conti, mom
DD Smaller	6×256	-	$200,000$	4.927 h	train: data val: -
DD	8×512	-	$200,000$	4.331 h	train: data val: -
Conti	8×512	DD	$+100,000$	+6.96 h	train: data, conti val: -
Hybrid RANS	8×512	Conti	+200	+0.024 h	train: data, conti val: data, noslip, conti, mom
Hybrid DD $\times 4$	8×512	Conti	+200	+0.024 h	train: data, conti val: data

Table 2.4.: Training configurations for 2D-Transient ANNs. All except "PINN" are DeepONets, for which the hidden layer specification applies to the trunk and branch each. The basis describes which prior model the training continues on, hence the "+" for number of training steps and elapsed time.

name	Extra	Extra2	Extra3	Inter
h in mm	17.5	17.5	10	6
Δp in kPa	14	8	16	12

Table 2.5.: Case selection for "Hybrid DD" 2D-Transient DeepONets.

but is not what the literature would refer to as a physics-informed neural network. For the data-driven DeepONets, two different MLP sizes are compared. Again, the continuity equation is given as supplementary supervision of training cases, and then hybrid fine-tuning steps are investigated. The data-driven fine-tuning step in "Hybrid DD" is applied to the extrapolation and interpolation cases listed in Table 2.5.

The first extrapolation case is the same as the PINN case. The second and third extrapolation cases have the worst results of the "Conti" baseline model, and so the fine-tuned results indicate how well the hybrid step can overcome very difficult out-of-distribution situations.

2.2.2.3. 3D-Transient

The characteristic quantities of 3D-Transient are for the DeepONets:

$$u_c = 4.988\,\mathrm{m/s}, \quad p_c = 13439.052\,\mathrm{Pa}, \quad \nu_{t,c} = 3.917 \cdot 10^{-4}\,\mathrm{m^2/s} \tag{2.19}$$

and for the single PINN case, which is DropXL:

$$u_c = 4.354\,\mathrm{m/s}, \quad p_c = 10382.2\,\mathrm{Pa}, \quad \nu_{t,c} = 3.396 \cdot 10^{-4}\,\mathrm{m^2/s} \tag{2.20}$$

The $\mathcal{L}_{\mathrm{data}}^{\mathrm{val}}$ for the PINN model not only includes the constrained velocity fields and pressure measurements but also an integral BC for the aorta mass flow. This is not implemented for use with DeepONets. The pressure nodes are positioned at $\boldsymbol{x}_{\mathrm{LV}}^* = (-0.84, -0.91,\ 0.0)$ and $\boldsymbol{x}_{\mathrm{LA}}^* = (0.085,\ 0.05, -0.4)$, mimicking the sensor locations in the physical experiment. The 3D-Transient ANNs are not trained over the entire domain as depicted in Figure 2.3. Their training is limited to the LV and LA in the following way: with the coordinate system located in the center of the orifice and the x-axis oriented along the 30° regurgitation jet direction, points with $y^* > 0.6$ or $z^* < -0.5$ are omitted.

For 3D-Transient, the dataset retains the original unstructured CFD mesh points rather than interpolating onto a structured grid. This approach captures the complex geometry, whereas a grid would require very high density. To address this, an approach where the no-slip wall points are retained from the CFD mesh and the interior is filled with a grid also seems feasible but isn't used here. For the branch net input, a 2D section is extracted at $z^* = 0$ and interpolated onto a 119×102 equidistant grid spanning $x^* \in [0.005, 0.69]$ and $y^* \in [-0.22, 0.22]$. This rectangular region is geometrically clipped in the top right and bottom right corners using linear boundaries inclined at $-30°$ and $60°$. This polygonal masking isolates the common field of view shared across all experimental cases, as different camera setups and the circular window shape resulted in varying coverages of coordinates. The input vector is thereby constructed from $N = 10738$ fixed sampling locations.

The NRMSE metric when concerning the performance over the entire cardiac cycle is calculated using a third of the available snapshots, 100 per case, with 5% of points per snapshot included. This still results in roughly 70,000 points per snapshot being used, depending on the case's CFD mesh density. Additionally, PDE residuals are calculated on the same points according to Appendix B.1.2, which enables the quantification of results

name	layers	basis	steps	time	losses
PINN RANS	6×512	-	$100,000$	15.535 h	train: - val: data, noslip, conti, mom
DD Smaller	8×512	-	$100,000$	18.746 h	train: data val: -
Exact	9×640	-	$188,000$	34.344 h	train: data val: -
DD	9×640	-	$50,000$	12.03 h	train: data val: -
Conti	9×640	DD	$+50,000$	+8.736 h	train: data, conti val: -
Hybrid Conti	9×640	DD	$+1,000$	+0.36 h	train: data val: data, conti
Hybrid DD $\times 6$	9×640	DD	+400	+0.24 h - +1.44 h	train: data val: data

Table 2.6.: Training configurations for 3D-Transient ANNs. All except "PINN" are DeepONets, for which the hidden layer specification applies to the trunk and branch each. The basis describes which prior model the training continues on, hence the "+" for number of training steps and elapsed time.

name	DropXL	SlotL-Bent	PinholeL	SlotL	DropXL	EccJet
source	CFD		Exp			

Table 2.7.: Case selection for "Hybrid DD" 3D-Transient DeepONets.

in the absence of ground truth data across the entire domain, as is the case with the experimental PIV measurements.

Table 2.6 gives an overview of trained models, which allow for the comparison between the PINN and DeepONet approaches, between different network sizes, physics inclusions, and hybrid training steps. The "Hybrid DD" models are fine-tuned for two CFD validation cases and for the four experimental setups where PIV recordings are presently available. These are listed in Table 2.7.

The PINN and "Hybrid Conti" model also concern the CFD DropXL data. The "Exact" continuity approach is implemented for 3D using a vector potential formulation [63]. The ANN predicts the three components of the vector potential ψ_i rather than the velocity components u_i directly. The velocity field is subsequently derived via the curl operation, expressed in index notation using the Levi-Civita symbol ϵ_{ijk}:

$$u_i = \epsilon_{ijk} \frac{\partial \psi_k}{\partial x_j} \quad \left(\text{e.g.} \quad u_1 = \frac{\partial \psi_3}{\partial x_2} - \frac{\partial \psi_2}{\partial x_3} \right) \quad \Rightarrow \quad \frac{\partial u_i}{\partial x_i} = \epsilon_{ijk} \frac{\partial^2 \psi_k}{\partial x_i \partial x_j} = 0 \tag{2.21}$$

While the 2D stream function formulation reduces the network outputs from two to one, the 3D vector potential formulation maps three potential components to three velocity components. This setup maintains the identical number of output variables while automatically satisfying the continuity equation.

3. Validation on Synthetic Data

This chapter presents the performance evaluation of the ML models detailed in the methodology. The different PINN and DeepONet configurations are compared to identify the most effective architecture and training strategy for the flow reconstruction task. All evaluations in this chapter are performed against the synthetic CFD data, which serves as the ground truth.

3.1. 2D-Steady

The initial validation uses the 2D-Steady dataset to assess the models' fundamental capability for spatial reconstruction from sparse data in a stationary environment.

3.1.1. PINN

The main results for the 2D-Steady PINN are shown in Figure 3.1. The red rectangle in (a) highlights the area where the u, v CFD data is supervised via $\mathcal{L}_{\text{data}}^{\text{val}}$. Consequently, the error (c) in this area is visibly lower than the rest of the domain. The colormap in (c) is scaled non-linearly to make low errors visible too, while the absolute error in (d) uses a linear colormap, albeit partially occluded by the streamline comparison. The largest deviation appears upstream of the orifice in what is equivalent to the LV ($x^* < 0$). The velocity magnitude gradually decreases towards the upper wall, while the CFD solution remains largely homogeneous until very close to the no-slip condition. Downstream in the LA, the regurgitation jet velocity decreases more quickly, and the resulting vortex center significantly misaligns with the target solution. The PINN fails to accurately develop the pressure field based on the supplied top left and top right corner values. The orifice results in a pressure drop over a short distance, but the PINN prediction once again develops the gradient over a longer distance. This is shown in the Appendix Figure C.10. Furthermore, the ν_t field is predicted as zero across the entire domain, which does not agree with the CFD results. Similarly, Pioch et al. demonstrate that for all turbulence models examined in their study, including the ν_t-model, training in the absence of labeled data leads to predictions that differ significantly from DNS results [47]. While the PINN produces plausible velocity fields with accurate magnitude, it fails to match the flow topology. The success of the PINN approach relies on the PDE losses to connect the sparse data to the BCs. The abrupt geometry change introduced by the orifice is likely challenging to fit around, and the supplied conditions might be ambiguous. Small training setup changes lead to substantially different predictions. In light of the promising results presented below, the subsequent development concentrated on the DeepONet approach, although additional enhancements of the PINN remain feasible.

3.1.2. DeepONet

The predictive accuracy of the previously introduced DeepONet configurations is summarized in Figure 3.2, which plots the prediction error for all flow variables. The model labels utilized in the evaluation plots correspond to the configurations detailed in the methodology

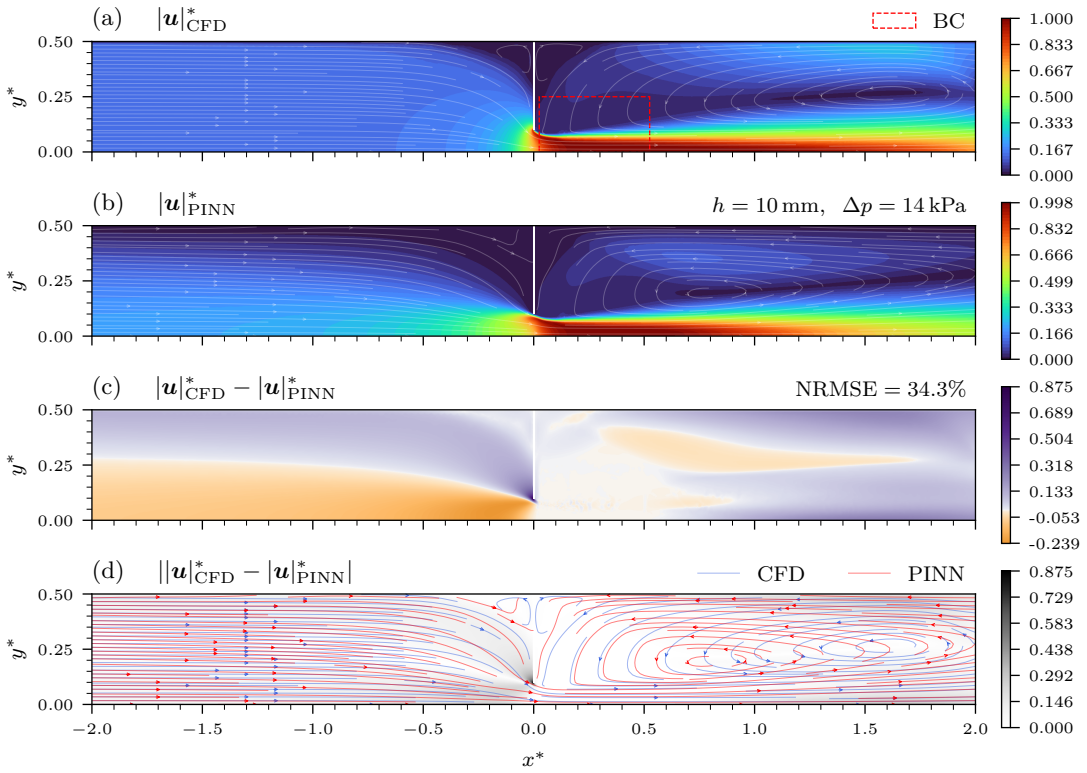

Figure 3.1.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 10\,\text{mm}$, $\Delta p = 14\,\text{kPa}$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

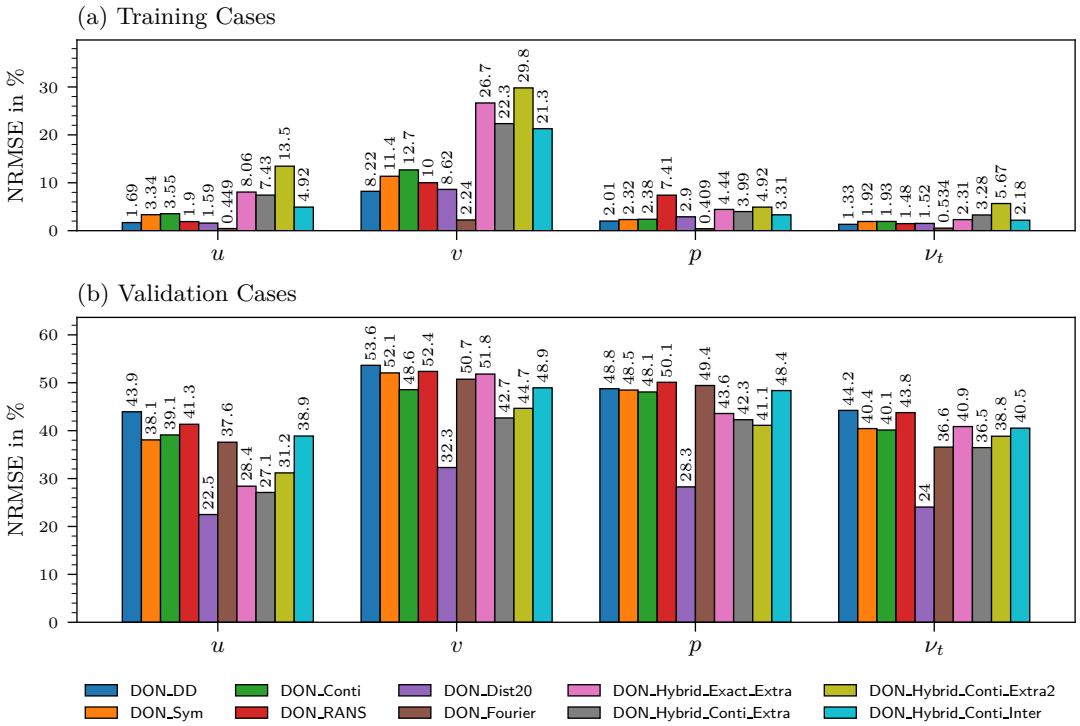

Figure 3.2.: Comparison of NRMSE for each 2D-Steady ANN output across various trained DeepONets. The charts display error percentages for (a) training cases and (b) validation cases.

tables and adhere to the following pattern: "Type_TableName_OptionalHybridTableName". The figure separates the results into training and validation sets to allow for a comparison of the data-fitting versus generalization performance across the different configurations. The network has to learn the relationship between the branch net input and the previously learned values of other cases, which might result in overfitting. The goal is to train a general operator, and so for the pre-trained basis models, a low validation case error (b) is the desired outcome. The hybrid models are also included here to quantify how the overall model is affected by the inclusion of one case's sparse data.

Augmenting the training of the baseline "DD" model with the symmetry and continuity PDE losses results in lower aggregate errors across all fields. The errors on the training examples rise simultaneously, indicating a transition from data-fitting towards improved generalization. A further 100,000 steps on "DD" were included to match the total training budget of the other variants and confirm that the gains arise from the added losses rather than from extended optimization. The inclusion of the momentum equations results in slightly worse accuracy of the "RANS" model compared to "Conti". The "Fourier" model rivals the latter in terms of validation performance, while simultaneously providing the closest fit to the training data.

Detailed error maps for the flow variables are presented in Figure 3.3. Each subplot (a–d) corresponds to a specific output field, displaying the NRMSE for every simulated case within the h vs. Δp parameter space. The non-linear color scale provides a quantitative measure of error, and the white-bordered markers identify the subset of cases used for model training. All remaining samples contribute to the validation error. This figure specifically shows the errors for just the "Conti" model, but the main topological features appear across models with the same training cases: the error distribution centers visibly around the training points and increases gradually towards the extreme ends of the parameter space. The v field consistently displays slightly higher error rates, even for training cases. This possibly results from the nature of the flow field and error definition. A target field with low and highly localized magnitudes produces a small standard deviation. The NRMSE metric, defined in Appendix B.1, divides the absolute error by this relatively smaller scale.

The model maintains acceptable performance near the training data. The interpolation cases located between training points show particularly low errors. One such best-case example is given in Figure 3.4, where the DeepONet manages to nearly instantaneously infer the velocity field very accurately. A recurring limitation of the non-hybrid DeepONets becomes visible immediately around and downstream of the orifice. The velocity gradient along the jet border is sharpest in this region. Consequently, the prediction error reaches its maximum here. This error manifests as small but distinct streaks of valleys and hills in (c). This artifact likely stems from the statistical averaging of the different orifice heights present in the training set. This specific area represents the only region where significant divergence of a streamline is noticeable. Elsewhere, the prediction aligns nearly perfectly with the reference. The out-of-distribution performance remains strong in the direction of increasing pressure drops Δp. The model generalizes less effectively in the direction of larger gap heights h.

As indicated in Figure 3.2, the "Dist20" DeepONet demonstrates improved generalization, which stems from the wider distribution of 20 training cases. Figure 3.5 highlights these specific training points. A larger portion of the validation set now falls within the interpolation regime. The model achieves consistently low error rates in these areas. The spatial distribution of error resembles the trends observed in the previous model: The prediction quality decreases more quickly in the direction of larger h, while performing very well for larger Δp. Additionally, the model continues to exhibit high error rates at the lower extremes of the parameter space. The "Dist20" model demonstrates the significant

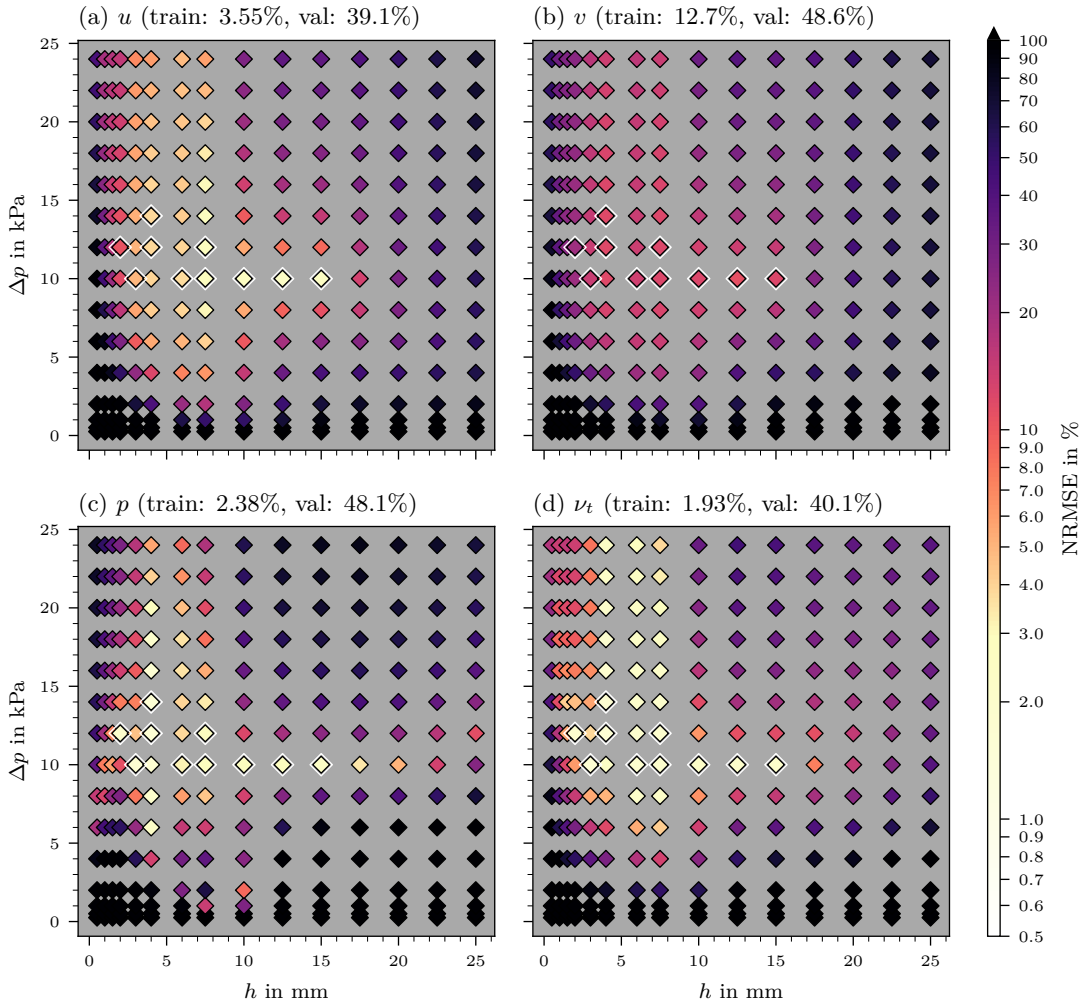

Figure 3.3.: Distribution of NRMSE for 2D-Steady "Conti" for all fields across the BC parameter space defined by orifice height and pressure difference. The color of each marker indicates the error magnitude, with the 10 training cases distinguished by a white border. Values higher than 100% are clipped.

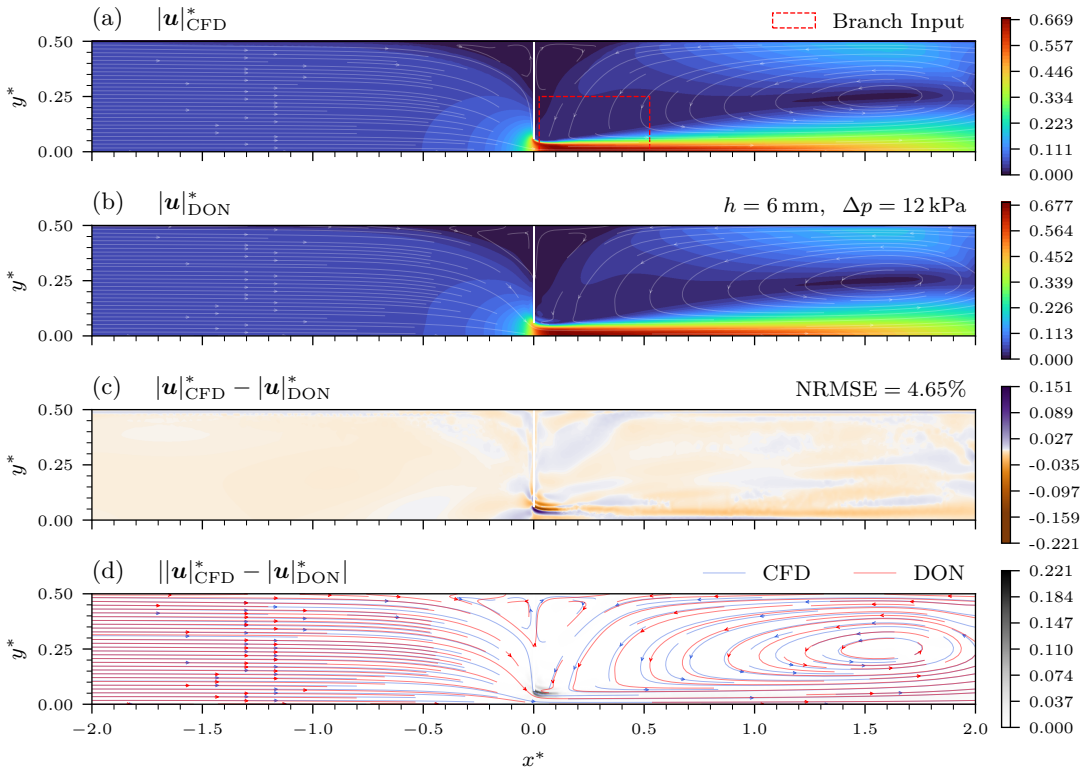

Figure 3.4.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 6\,\text{mm}$, $\Delta p = 12\,\text{kPa}$. (b) "Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

impact of training data availability on the DeepONet approach; increasing the amount of ground truth data serves as a major influence on accuracy. The remaining models in this work utilize a standard set of 10 training cases since this aligns the analysis with the target 3D-Transient setup. The analysis now turns to an extrapolation case. This case lies at the top right corner of the parameter distribution and presents a significant challenge. The "Dist20" model manages an NRMSE for the velocity magnitude of 39.3% here. This result appears in the Appendix Figure C.11. In comparison, the "Conti" model reaches an error of 84.0%. Figure 3.6 displays this outcome.

Here, the predicted velocity magnitude remains too low across the entire domain. The characteristic streaking artifacts persist, and a prominent underprediction of velocity magnitude overlays these artifacts within the jet region. This specific case features a large orifice opening previously absent from the training set. Consequently, the model limits the flow prediction to the maximum gap height encountered during training. Deviations also appear in the flow topology; the center of the recirculation vortex in the LA now shows a distinct misalignment compared to CFD. Visually, the prediction by "DD" is similar but results in an NRMSE of 95.7%, showing that the inclusion of $\mathcal{L}^{\text{train}}_{\text{PDE}}$ may improve baseline performance but not overcome the fundamental challenge for out-of-distribution inference.

This is where the test-time adaptation presents its advantage, such as with the "Hybrid Conti" model. This approach incorporates the sparse observations $\mathcal{L}^{\text{val}}_{\text{data}}$ into the composite loss function during the fine-tuning phase. The results, visualized in Figure 3.7, demonstrate a substantial improvement of predictive fidelity. The NRMSE decreases significantly from 84.0% to 15.2%, which corresponds to an over 5-fold improvement. Qualitatively, the velocity magnitude is largely recovered across the LV, the regurgitation jet, and the downstream LA region. The flow topology improves correspondingly, with the center of

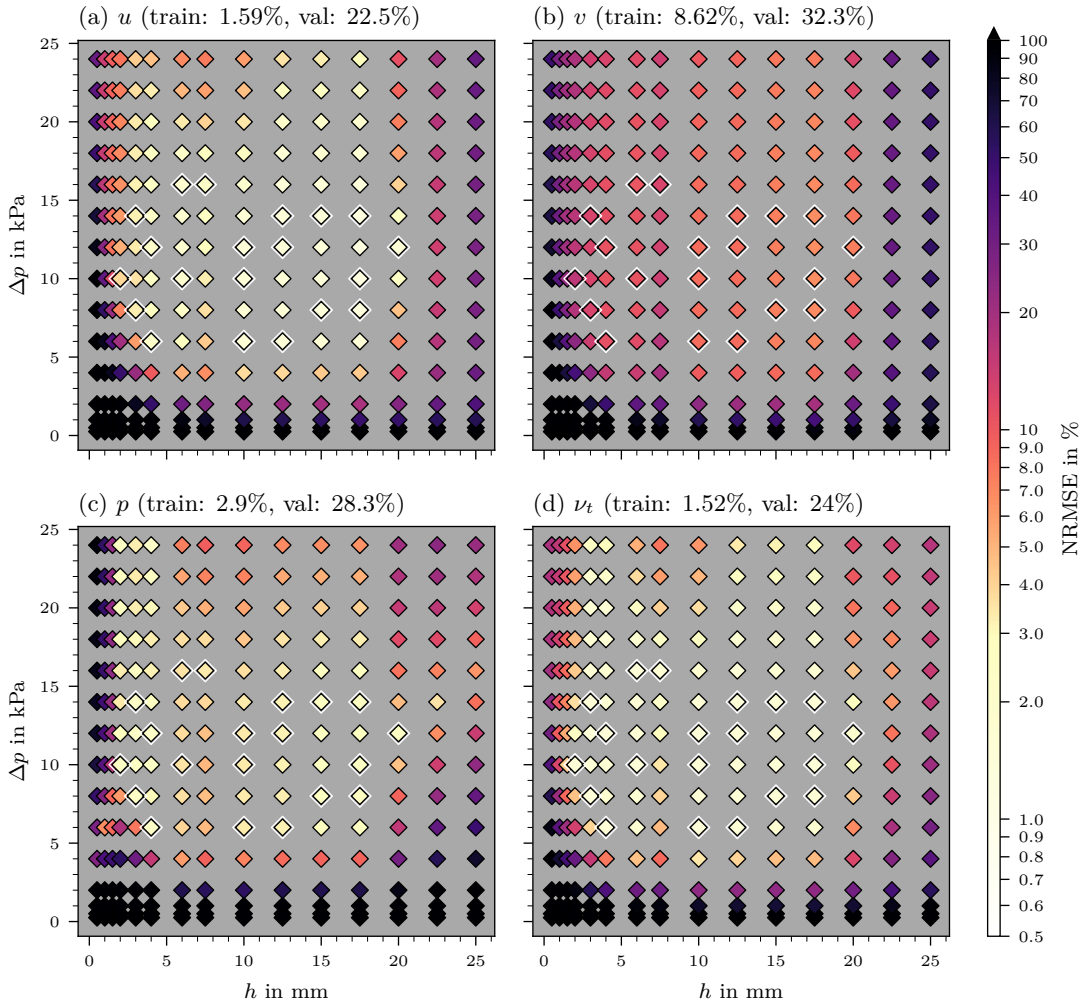

Figure 3.5.: Distribution of NRMSE for 2D-Steady "Dist20" for all fields across the BC parameter space defined by orifice height and pressure difference. The color of each marker indicates the error magnitude, with the 20 training cases distinguished by a white border. Values higher than 100% are clipped.

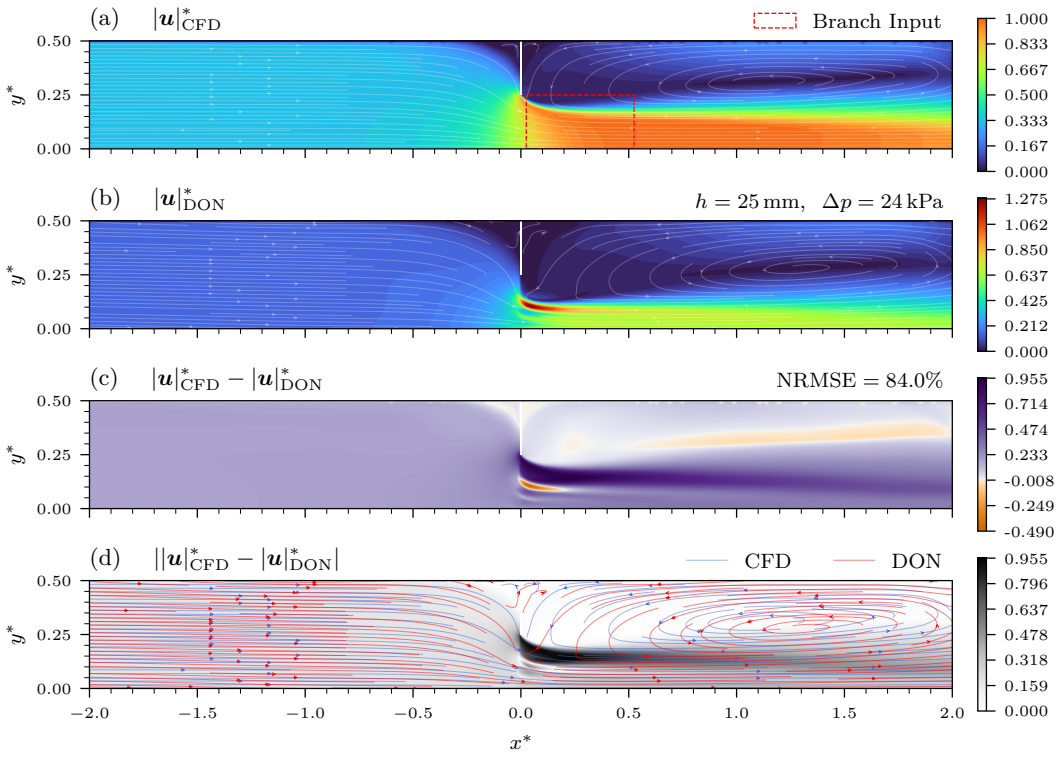

Figure 3.6.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 25\,\mathrm{mm}$, $\Delta p = 24\,\mathrm{kPa}$. (b) "Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

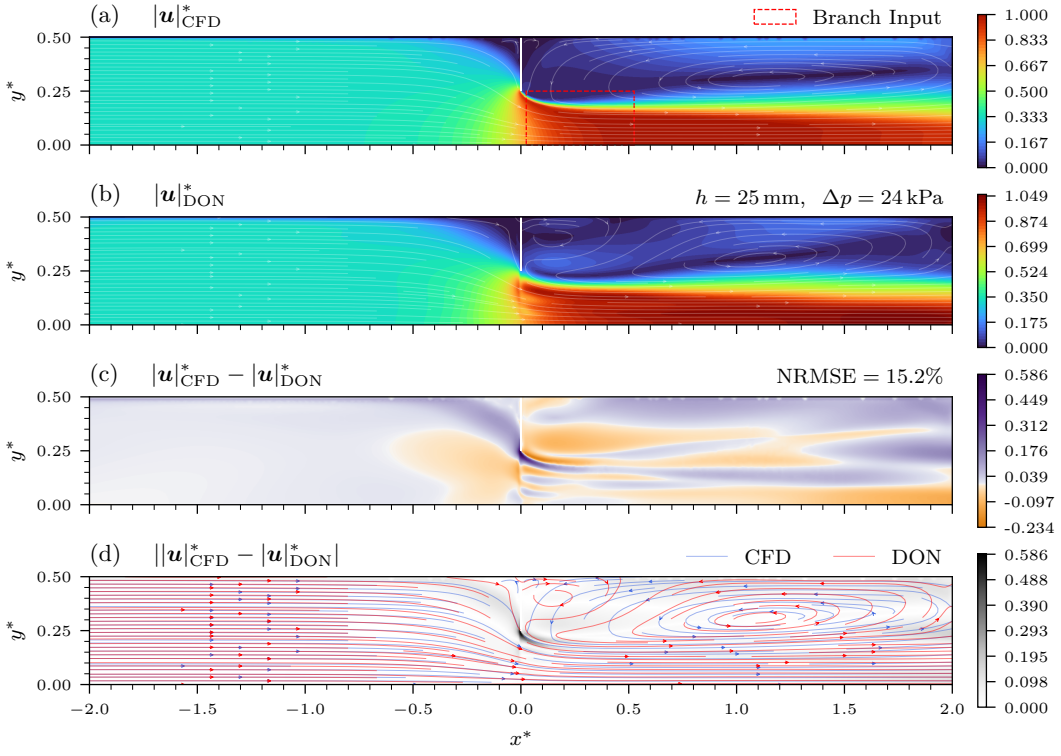

Figure 3.7.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 25\,\mathrm{mm}$, $\Delta p = 24\,\mathrm{kPa}$. (b) "Hybrid Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

the recirculation vortex aligning much more closely with the ground truth compared to the baseline model. Despite these improvements, the solution retains artifacts indicative of its data-driven nature. A secondary recirculation zone appears in the LA adjacent to the orifice wall, a flow feature absent in the CFD reference. Furthermore, the characteristic streaking remains visible, and the velocity gradient along the shear layer lacks the sharpness observed in the CFD solution. A specific artifact also persists along the vertical line extending from the orifice ($x^* = 0$), where the magnitude is locally underpredicted. These observations indicate that while test-time adaptation effectively steers the global solution towards the correct regime, the model may still produce local physical inconsistencies or "hallucinations" in regions lacking direct supervision. Notably, the streaking artifacts persist within the observational window where direct supervision is applied. This behavior stands in contrast to the pure PINN approach, which achieved negligible error magnitudes in the supervised region. The persistence of these artifacts stems from the hybrid training strategy. The model continues to train on the full distribution of training cases simultaneously with the fine-tuning loss. Consequently, the network balances the adaptation to the specific sparse observations against the general operator learning derived from the broader dataset.

The improvement in predictive fidelity extends beyond the specific case targeted during the test-time adaptation. As evidenced by the aggregate validation metrics presented in Figure 3.2, both the "Hybrid Extra" and "Hybrid Extra2" models demonstrate a notable reduction in the total NRMSE across the validation set compared to the baseline "Conti" model. This indicates that fine-tuning the operator on a single out-of-distribution case positively influences the inference quality for neighboring cases in the parameter space.
A noteworthy observation is the improvement of the ν_t-field, despite the absence of additional data supervision for this variable during the adaptation step. This suggests that the shared parameters of the DeepONet adjust in a manner that implicitly pulls the turbulent viscosity prediction towards a more accurate state as well. Conversely, the "Hybrid Inter" model yields negligible improvements in the overall validation metrics. This outcome is expected, as this interpolation case is located in a region densely populated by training samples, where the baseline prediction is already highly accurate. While the adaptation step largely mitigates the streaking artifacts discussed previously, it does not eliminate them entirely in the interpolation case either. Consequently, the fundamental topology and magnitude of the solution remain largely unchanged and are not shown separately.

The previous Figure 3.7 presents the "Extra2" case ($h = 25\,\text{mm},\ \Delta p = 24\,\text{kPa}$). To further assess the hybrid approach, the "Extra" case ($h = 20\,\text{mm},\ \Delta p = 14\,\text{kPa}$) is also examined, shown in Figure 3.8. The NRMSE distribution map for this model is provided in the Appendix Figure C.12. When compared to the baseline distribution, the error field is noticeably more homogeneous, and the u-field's accuracy is increased around the adapted case. Given a baseline "Conti" accuracy of 40.8% for the velocity magnitude (Figure C.13), the fine-tuned "Hybrid Conti Extra" model shows a $\times 5.48$ lower NRMSE of 7.45%. This improvement highlights that a more accurate starting point from the pre-trained operator facilitates an even higher fidelity in the final fine-tuned result. The flow topology matches the reference closely. The center of the recirculation vortex exhibits only a minor misalignment in the x-direction, while the streamlines align well throughout the rest of the domain. The secondary recirculation artifact observed in the "Extra2" case is absent here. Additional figures for the following field evaluations are provided in the Appendix. The v-velocity component exhibits an NRMSE reduction of $45.7 \rightarrow 29.4\%$ relative to the "Conti" baseline. The provided Figure C.14 shows the low and localized magnitude of this field, which amplifies the relative error metric despite low absolute deviations. The error map reveals a persistent vertical line of underprediction along the orifice at $x^* = 0$. The pressure field achieves an improved NRMSE of $40.9 \rightarrow 32.2\%$. The prediction captures the

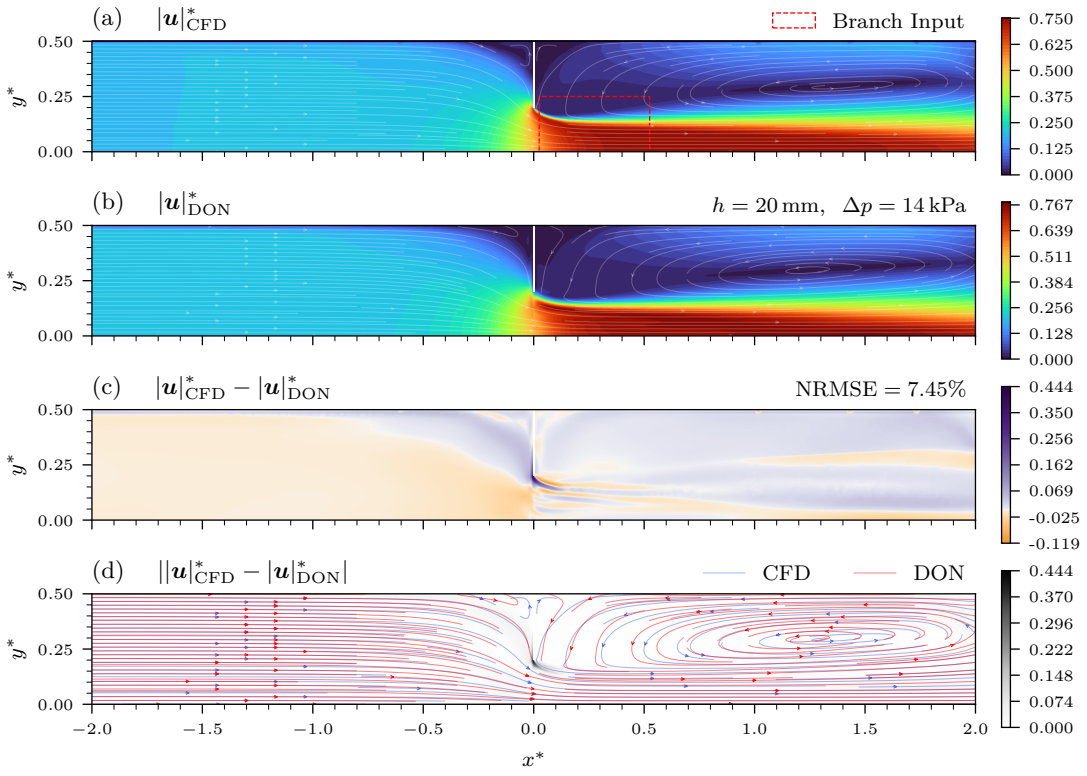

Figure 3.8.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 20\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$. (b) "Hybrid Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

steep gradient across the orifice and the homogeneous distribution elsewhere effectively, although the magnitude within the LV remains slightly lower than the target (Figure C.15). The training setup was optimized for velocity field performance, so improved pressure predictions and stronger adaptation to pressure measurements remain feasible with targeted adjustments. Finally, the turbulent viscosity field error drops from 19.4 → 13.3% (Figure C.16). The model refines this additional quantity based on the updated operator state, even without its inclusion in the sparse adaptation dataset.

Two alternative DeepONet architectures introduced in the Methodology Chapter, the Fourier-feature network and the stream-function-based "Exact" model, were also evaluated to determine if addressing spectral bias or automatically satisfying continuity could yield further improvements. While the Fourier model achieves low training error (Figure 3.2), it exhibits signs of overfitting. The predictions are prone to high-frequency, jagged artifacts, indicating that the standard MLP architecture is more robust and better suited for representing the overall smooth field characteristics of the RANS solutions. Hybrid adaptation of the "Fourier" model results in nonphysical flow patterns significantly worse than the standard "Hybrid Conti" result. In general, a baseline model candidate can outperform another baseline model candidate both qualitatively and quantitatively, yet a fine-tuned hybrid from the better baseline is not necessarily superior to the hybrid based on the other, which complicates development.

Similarly, the Exact model demonstrated that satisfying the continuity equation by definition does not guarantee a solution close to the ground truth. The basis model for this architecture suffers from degradation in the flow convergence zone upstream of the orifice and increased streaking artifacts. This performance drop could be attributed to the optimization bottleneck of deriving two velocity components from a single scalar output (ψ),

which restricts the network's expressivity compared to predicting u, v directly. Although the fine-tuned "Hybrid Exact" model achieves comparable error metrics to the "Hybrid Conti" approach, the visual quality of the streamlines and flow topology remains consistently inferior. Furthermore, a "Hybrid RANS" configuration incorporating momentum residuals was tested but yielded inferior results, possibly driving the solution toward the same inaccurate state observed in the PINN approach analysis. While both ideas were briefly re-evaluated during the development of 2D-Transient models due to their theoretical benefits, they echoed the findings of 2D-Steady and are therefore not included in this work, except the "Exact" model, which is revisited for 3D-Transient, where the formulation restores the same number of output variables.

3.1.3. Interim Conclusion

In summary, the investigation of the stationary 2D channel flow establishes the Hybrid DeepONet as the superior framework for reconstructing flow fields from sparse observations. The pure PINN approach, despite utilizing physics constraints, fails to converge to the correct flow topology or pressure gradients, seemingly unable to resolve the ill-posed problem given the limited boundary data and complex orifice geometry. Conversely, purely data-driven DeepONets demonstrate high effectiveness for interpolation within the training distribution, particularly when the training dataset size is increased or when soft physical constraints are included in the loss function. However, these basis models consistently struggle to generalize to parameter regimes outside the training set, yielding high errors and artifacts in extrapolation cases. The hybrid training strategy successfully bridges this gap. By fine-tuning the pre-trained "Conti" operator with the sparse measurements specific to the validation case, the model achieves a five-fold reduction in error for extrapolation cases without compromising the already high accuracy of interpolation predictions. The standard MLP-based Hybrid DeepONet successfully combines the generalization power of a pre-trained model with the specificity of sparse observations, overcoming the individual limitations of purely data-driven or physics-informed approaches. The practical ineffectiveness of including the full momentum equations, evidenced by the failure of the turbulence model in the PINN and the degradation of the "Hybrid RANS" results, suggests that for this specific setup, simpler constraints like continuity provide a more robust regularization.

3.2. 2D-Transient

The 2D-Transient dataset introduces the complexity of time-varying, periodic flow to the reconstruction task. This section evaluates the models' ability to capture transient phenomena, such as the formation and dissipation of the regurgitant jet over the cardiac cycle, based on the setups defined in Table 2.4.

3.2.1. PINN

The performance of the PINN is evaluated first to establish a baseline for a purely physics-driven surrogate. The "PINN RANS" model is trained on a single case ($h = 17.5\,\text{mm}$, $\Delta p = 14\,\text{kPa}$). Figure 3.9 displays the velocity magnitude prediction at $t^* = 0.25$. Consistent with the 2D-Steady results, the prediction error remains minimal within the region directly supervised by the sparse measurement data. The reconstructed regurgitant jet propagates downstream past this supervised window, although it exhibits a reduced length and diminished velocity magnitude relative to the CFD ground truth. While the primary recirculation vortex center aligns with the reference, the flow topology diverges in the upstream LV, where the network underpredicts the velocity magnitude and the streamlines diverge near the upper wall. At the later timestep $t^* = 0.75$, the reconstruction in the LA remains accurate. However, the model again fails to resolve the flow dynamics within the LV. The jet appears as a truncated feature localized near the orifice. This structure demonstrates an incorrect temporal evolution, retracting towards the opening, while the reference flow extends into the ventricle. This failure suggests that the RANS loss was

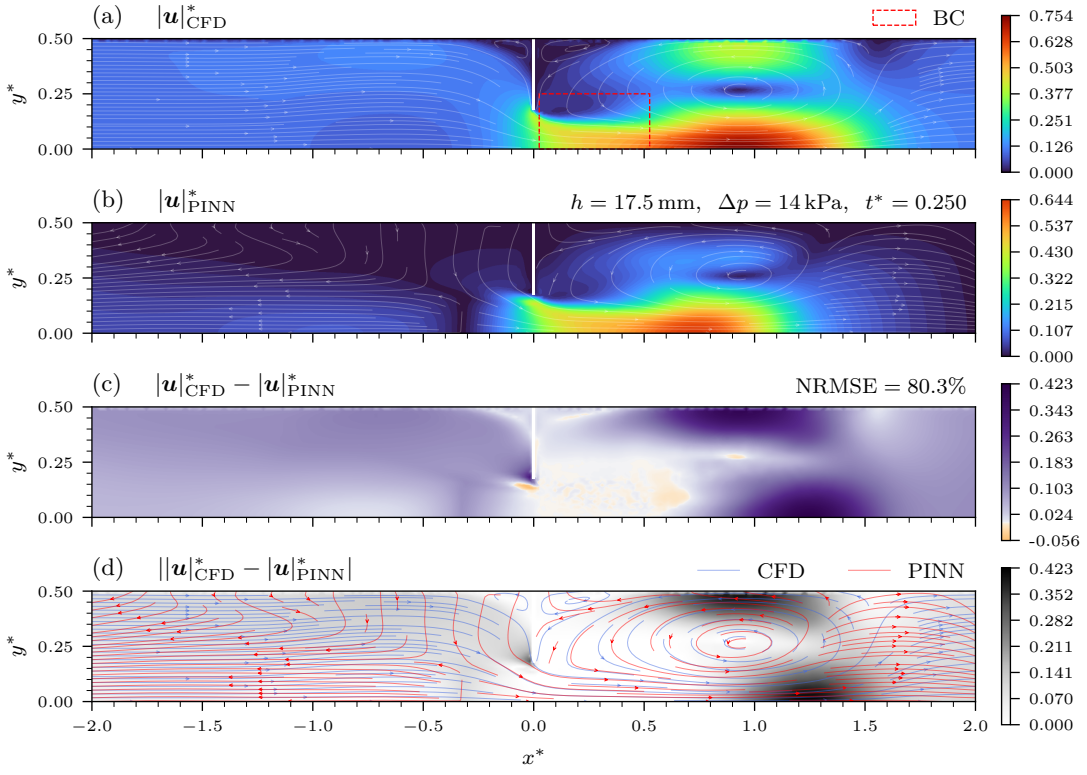

Figure 3.9.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\text{mm}$, $\Delta p = 14\,\text{kPa}$, $t^* = 0.25$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

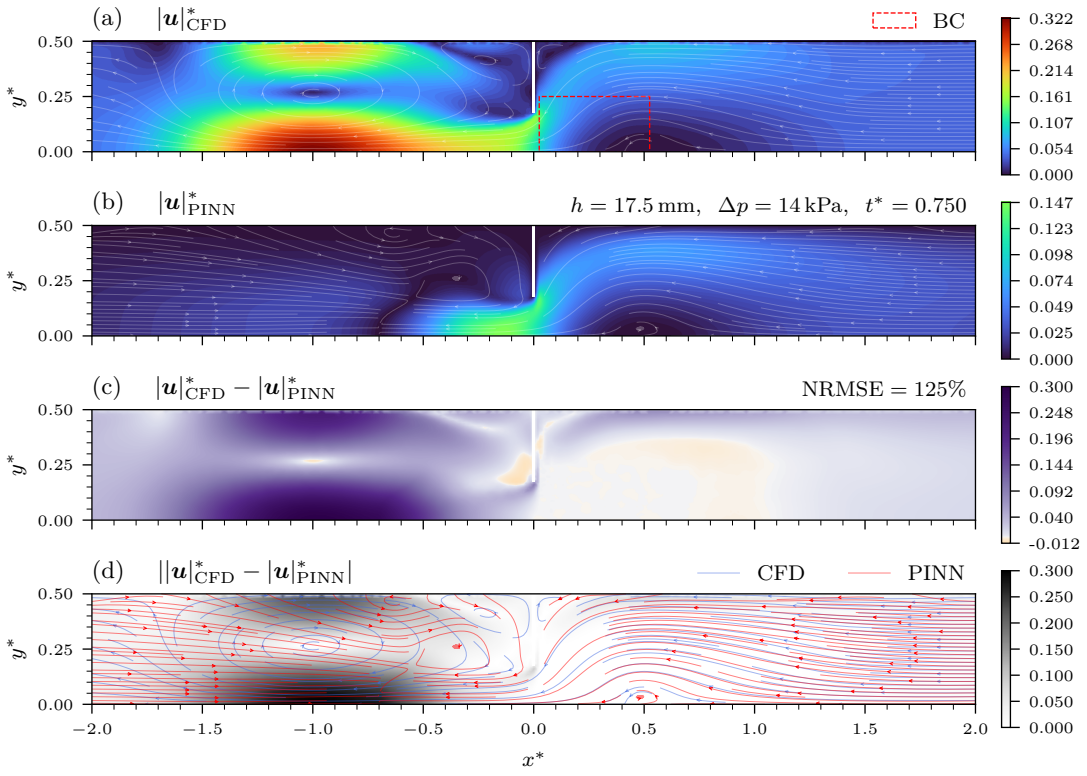

Figure 3.10.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\text{mm}$, $\Delta p = 14\,\text{kPa}$, $t^* = 0.75$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

unable to sustain this specific flow feature without more explicit data supervision.
The baseline "PINN DD" model (Appendix Figures C.17 and C.18) exhibits higher error but retains structural similarities to the physics-informed result, such as the jet extension beyond the supervision window during systole and better flow capture in the LA. This confirms that sparse data independently orients the solution, while the RANS equations are essential for substantial improvements in fidelity. The comparative performance of all modeling approaches is summarized by the temporal evolution of the NRMSE in Figure 3.11. The trajectories confirm that the inclusion of RANS equations in the PINN yields a persistent improvement over the data-driven ANN. The turbulent viscosity output avoids the trivial zero-solution observed in 2D-Steady but remains physically inconsistent without any explicit supervision. The pressure prediction usually fails to recover correct fields, and because boundary information is very sparse, it only approximates them briefly at specific time-steps (e.g. Appendix Figure C.19). Additional p-sensor locations might improve this field's prediction, and coupled through the momentum equation, the velocity fields' too. This is a change that can be implemented in an experimental setup as well. The relatively low global v-error reflects good agreement inside the PIV window; low ground-truth magnitudes outside that window keep the global error small.

3.2.2. DeepONet

All DeepONet configurations significantly outperform the instance-specific PINN models across the entire cardiac cycle and for every flow variable, showing roughly a $\times 2$ reduction in NRMSE. The horizontal lines in Figure 3.11 summarize this finding. These substantially lower errors demonstrate that the operator learning framework, by leveraging a distribution

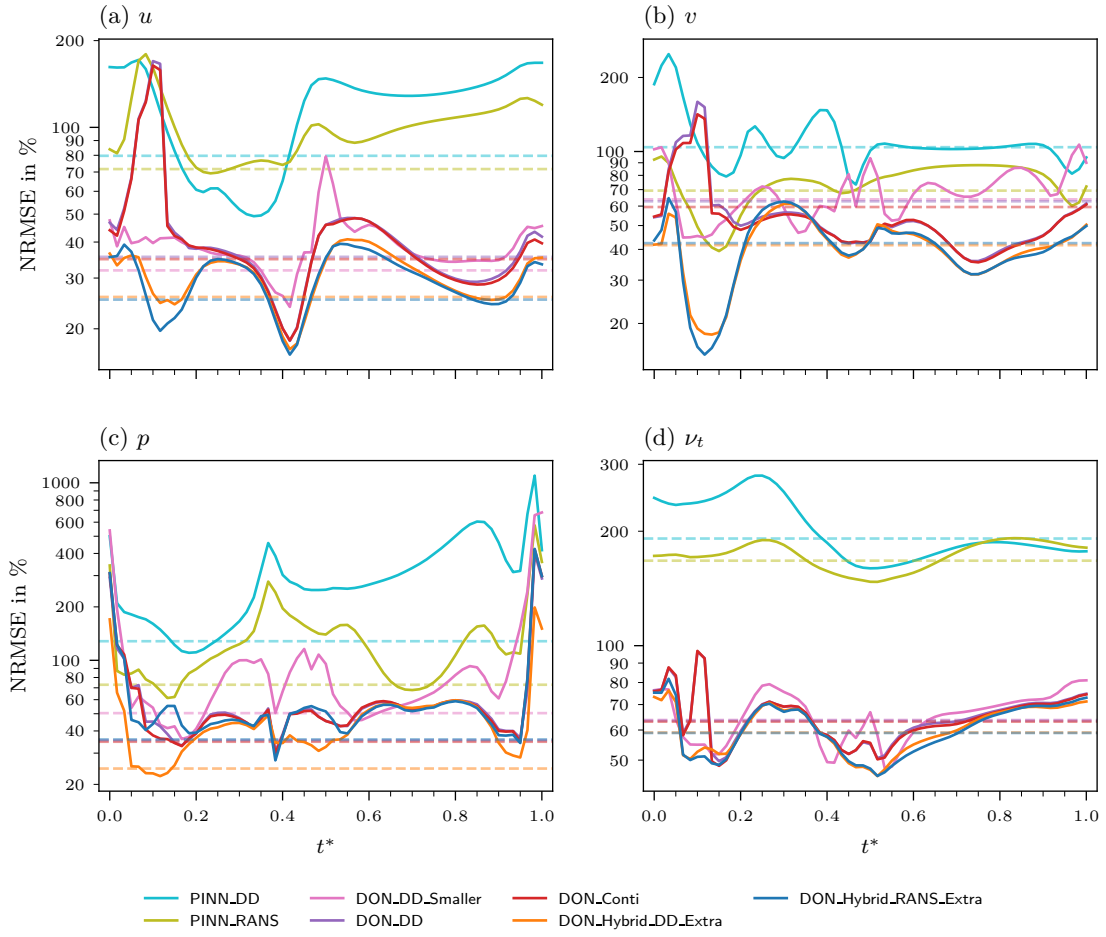

Figure 3.11.: Temporal evolution of the NRMSE for all fields over the cardiac cycle for the 2D-Transient $h = 17.5\,\text{mm}$, $\Delta p = 14\,\text{kPa}$ case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

of training cases, provides a strong foundation for the reconstruction task. Consequently, the remainder of this section focuses on evaluating the DeepONet architectures.
It is important to note that this comparison reflects a specific validation case. While the DeepONet leverages the computational investment of the pre-calculated dataset to achieve superior accuracy here, the PINN remains a viable candidate for further development. Although the current physics-driven performance leaves room for improvement, optimizing the formulation could yield a robust solver for scenarios where generating training data is impossible. Conversely, the DeepONet's reliance on the training distribution highlights a different path for optimization: its robustness can be directly enhanced by strategically expanding the CFD dataset to cover broader parameter regimes. A scenario likely exists where a baseline DeepONet performs worse than the PINN due to extreme extrapolation; whether this disparity is best resolved through hybrid adaptation, improved PINN constraints, or a denser DeepONet training distribution remains uncertain.

The aggregate performance of the DeepONet configurations is compared in Figure 3.12. Note that the validation metrics (b) for 2D-Transient span a narrower parameter distribution than in 2D-Steady. The average distance between a validation case and the nearest training sample is smaller, and extreme cases are no longer included.

The influence of network size is assessed by comparing the "DD Smaller" (6×256) against the "DD" configuration (8×512). Preliminary tests with even smaller architectures yielded

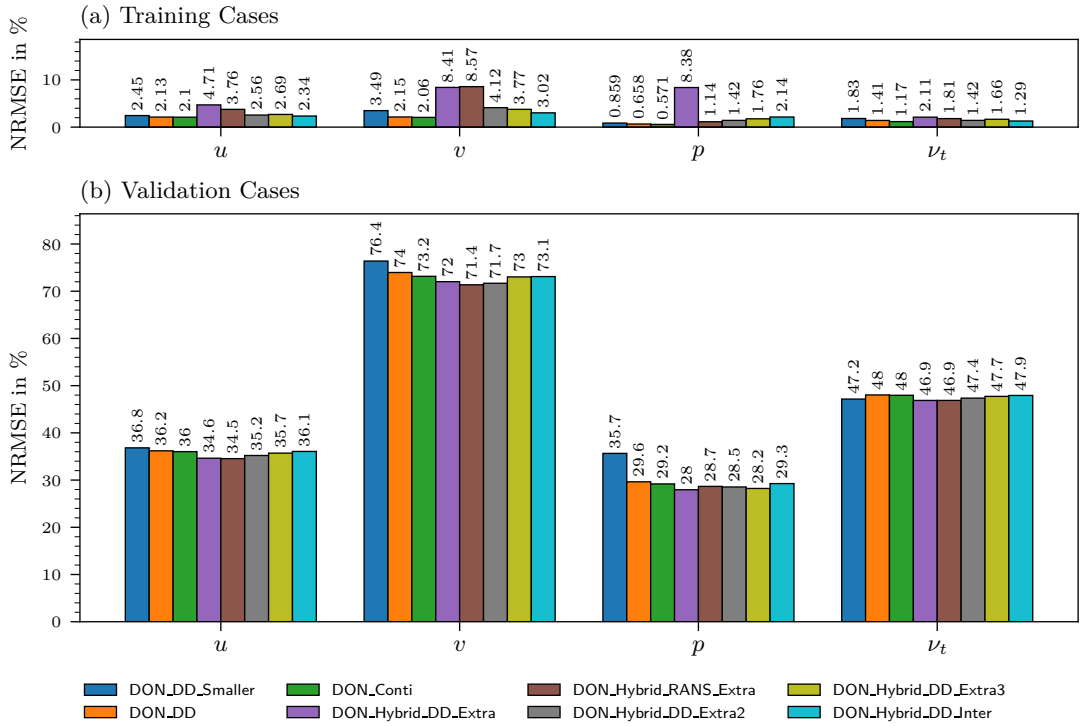

Figure 3.12.: Comparison of NRMSE for each 2D-Transient ANN output across various trained DeepONets. The charts display error percentages aggregated over the cardiac cycle for (a) training cases and (b) validation cases.

consistently inferior results. Despite the large parameter count, the network does not overfit at this size and instead achieves superior generalization. This improvement is most pronounced in the pressure field, where the error drops from 35.7% to 29.0%. However, this gain is not strictly uniform across all metrics; for example, ν_t has a slightly higher NRMSE. A recurring challenge in evaluating these models is that global scalar metrics can obscure significant qualitative differences, as fine-scale topological improvements or localized artifacts are absorbed into the aggregate value. The benefit of increased expressivity becomes evident when examining specific flow features. The larger model renders the diastolic jet within the LV with significantly higher physical consistency. This is reflected in the temporal evolution of the v-component error for the validation case (Figure 3.11,b), where the error during diastole drops from approximately 70% in the smaller model to 40%. Conversely, the increased complexity introduces a transient instability. The larger models exhibit a momentary artifact in the regurgitation jet around $t^* = 0.1$, visible as a localized error spike in (a). This artifact momentarily drives the error higher than even the PINN baseline, skewing the global NRMSE. A snapshot of this is appended in Figure C.20. However, because this specific instability is readily corrected during the subsequent hybrid fine-tuning step, the larger architecture is selected as the superior basis for the hybrid framework. The subsequent addition of the continuity constraint in the "Conti" model yields further improvements. While the gains are more subtle than in the 2D-Steady case, the constraint appears to act as a beneficial regularizer, slightly smoothing the prediction without conflicting with the data. Appendix Figures C.21 and C.22 display the resulting flow fields at $t^* = 0.25$ and $t^* = 0.75$, which surpass the PINN baseline accuracy. A streak of underpredicted velocity magnitude remains downstream of the orifice, caused by the model extrapolating to an orifice height beyond the maximum observed in training data. Compared to the 2D-Steady models, the finer streaks below this one are no longer present due to the many observed transient flow states. A RANS-informed baseline is

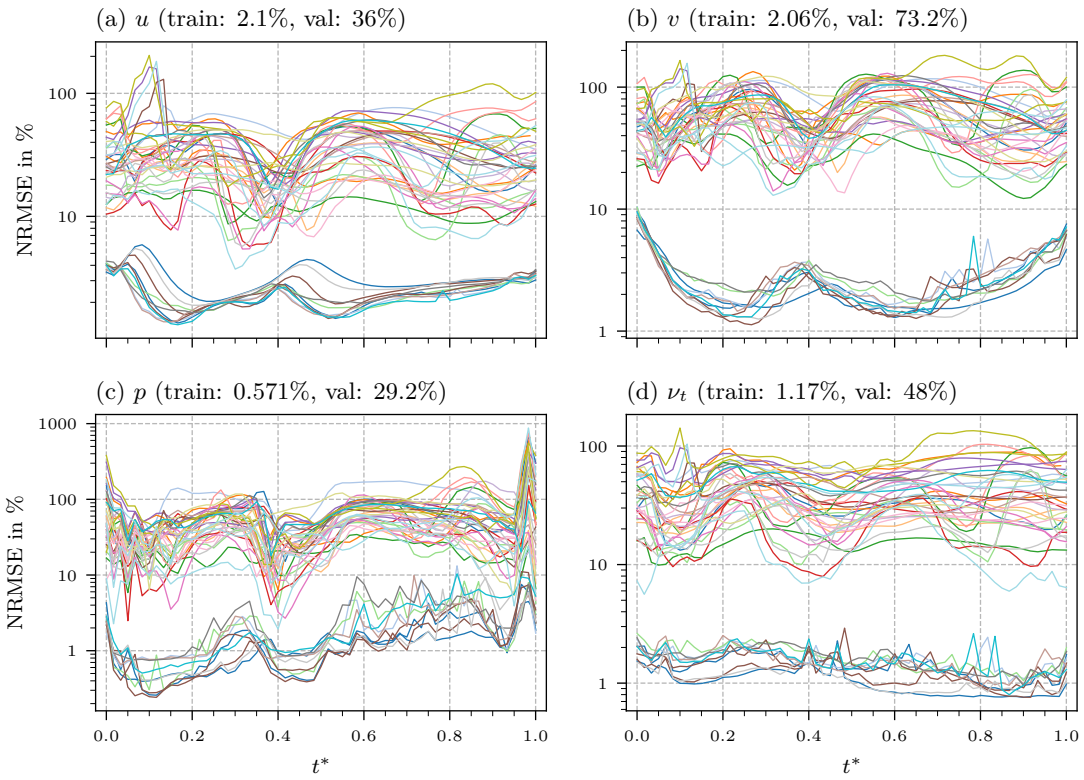

Figure 3.13.: Temporal evolution of NRMSE for all fields over the cardiac cycle for the 2D-Transient "Conti" model. The 45 lines represent all cases from the BC distribution.

excluded from this comparison. Following the negligible or negative impact observed in 2D-Steady and confirmed by preliminary tests on this dataset, the momentum equations seem redundant. The training process involves full-field supervision from CFD data that was solved to satisfy these physics; the additional loss term appears to offer little value beyond what the network learns directly from the data.

Figure 3.13 expands the analysis from a single instance to the entire dataset, illustrating the temporal NRMSE evolution for all 45 cases using the "Conti" model. It indicates that the transient artifact at $t^* = 0.1$ affects four cases. Distinctly clustered are the training cases, which maintain a consistent low-error baseline, and the validation cases, which exhibit higher variance. Regions of lower and higher accuracy seem to occasionally correlate with flow strength, but many trajectories are unique. A critical observation concerns the pressure field, where the error rises sharply towards the end of the cycle, affecting even the training set. The logarithmic scaling obfuscates the difference, with a maximum of 876% compared to the global validation NRMSE of 29.2%. This deviation stems from the cyclic nature of the ground truth data. As detailed in the methodology (Figure 2.9), the pressure waveform quickly reverts to positive values in the final few frames (approx. $0.9 < t^* < 1.0$) to close the loop. The network lacks specific treatment for periodicity and tends to preserve the negative pressure state established during diastole. Furthermore, the ground truth pressure magnitude approaches zero at the cycle's end; in this regime, faulty predictions create substantial relative errors. Strategies such as shifting the phase of the pressure BC to align the sharp BC gradient with the start of the simulation or employing temporal padding by duplicating initial and final snapshots could help the network better resolve this rapid transient without architectural changes.

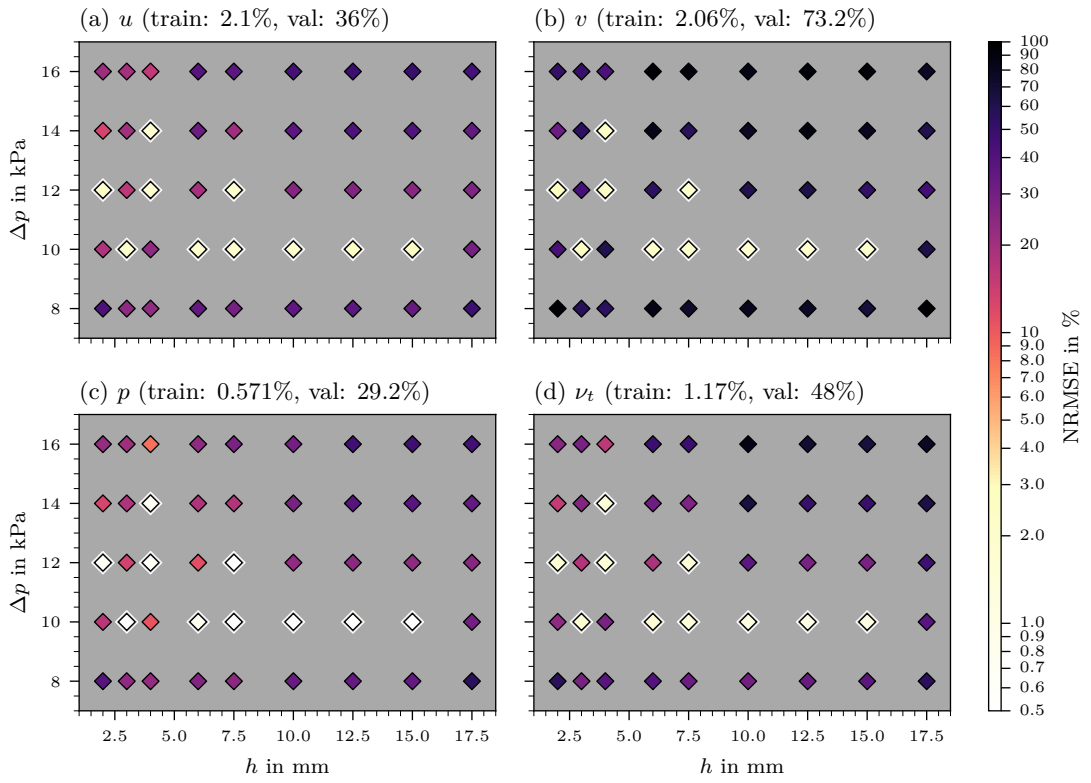

Figure 3.14.: Distribution of NRMSE for 2D-Transient "Conti" for all fields across the BC parameter space defined by orifice height and pressure difference. The color of each marker indicates the error magnitude over the cardiac cycle, with the 10 training cases distinguished by a white border. Values higher than 100% are clipped.

Figure 3.14 condenses the temporal performance into a spatial distribution, displaying the cycle-aggregated NRMSE for each case across the h-Δp parameter space. Compared to the 2D-Steady results, the error distribution appears significantly more heterogeneous, with even central interpolation cases exhibiting error magnitudes previously associated with extrapolation. While this pattern suggests potential overfitting, the metric is heavily influenced by the transient nature of the problem. Unlike the steady-state case, where topological alignment drives the score, the transient evaluation penalizes even minor temporal misalignments. A slight phase shift in the jet formation results in a high mathematical error, even if the flow structure is physically sound. Consequently, even the "interpolation" regime in this dynamic context presents a substantially more complex challenge for the operator than the steady-state equivalent.

This sensitivity to phase alignment is exemplified by the hybrid interpolation result shown in Figure 3.15. Visually, the reconstruction of the velocity magnitude and streamlines appears nearly flawless, recovering the flow topology with high fidelity. However, the quantitative error remains elevated at 29%. The signed error map in Figure 3.15(c) reveals the source of this discrepancy: a distinct dipole pattern of positive and negative error regions emerging due to the predicted regurgitant jet pulse, which is slightly behind the ground truth location. While not the sole source of error, this temporal shift may occur across all flow fields and significantly elevates the error metrics whenever it does. Therefore, a high error metric does not necessarily signify a meaningful failure to reconstruct the flow.

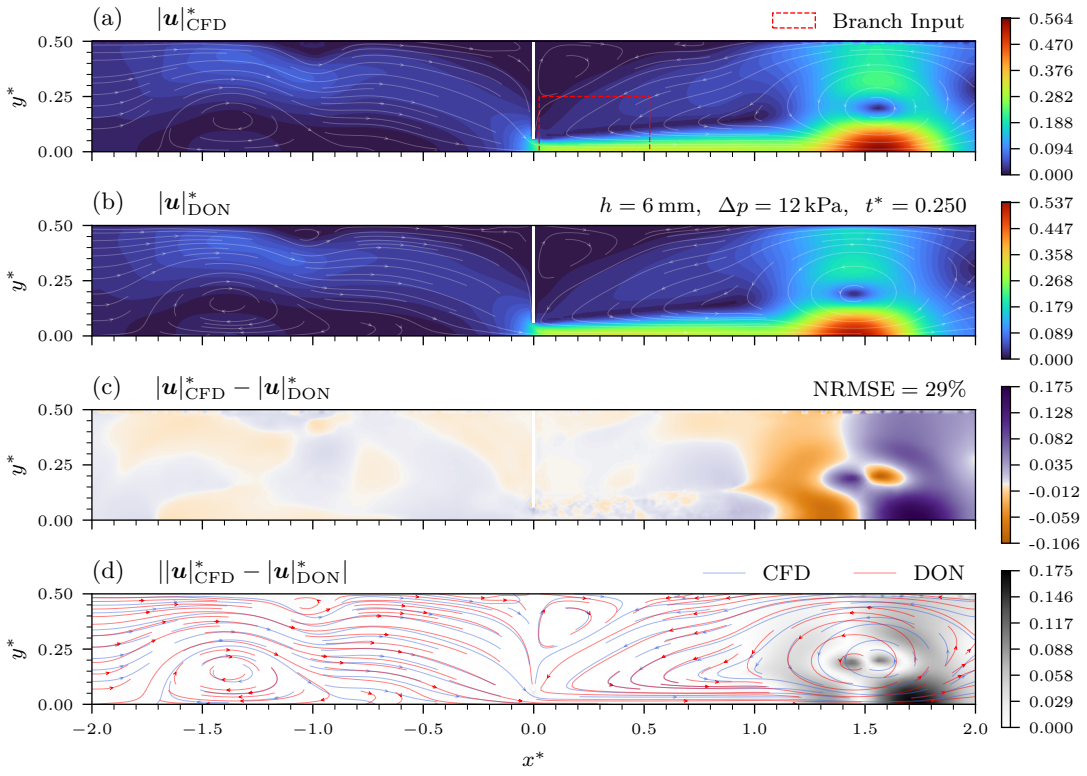

Figure 3.15.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 6\,\text{mm}$, $\Delta p = 12\,\text{kPa}$, $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

Figures 3.16 and 3.17 present the predictions of the "Hybrid DD" model for the same case previously analyzed with the PINN. Further comparisons of the v-velocity, pressure, and turbulent viscosity fields against the CFD ground truth are provided in Appendix Figures C.25 through C.27. At $t^* = 0.25$, the flow field within the LV ($x^* < 0$) is reconstructed with significantly higher fidelity, displaying low magnitude error and accurate streamline alignment. The error in the LA is similarly reduced, although the regurgitant jet shape remains slightly shorter than the CFD reference, resembling the PINN prediction in this specific feature. At $t^* = 0.75$, the LV flow again closely matches the ground truth; the streamlines indicate a correctly aligned vortex center, and the velocity magnitude is largely recovered. Crucially, the diastolic jet develops realistically, extending into the ventricle rather than retracting. Conversely, the streamlines in the LA exhibit some dissimilarity at this time step, particularly in the region around $x^* = 1$. The temporal evolution of the error, displayed in Figure 3.11, confirms that the hybrid fine-tuning step effectively resolves the transient artifact characteristic of the data-driven baseline models. The most significant improvement over the "Conti" model occurs before $t^* = 0.2$, coinciding with the appearance of the regurgitant jet within the supervised observation window. In this phase, the high flow magnitudes lead to large absolute errors in the baseline, which are subsequently corrected by the data supervision. For the remainder of the cycle, the NRMSE remains below that of the "Conti" model. However, the magnitude of this reduction is considerably smaller than the five-fold improvement observed in the 2D-Steady case and still mostly due to the better accuracy within the constrained window.
A comparison with the baseline "Conti" model predictions provided in Appendix Figures C.21 and C.22 reveals that the hybrid fine-tuning induces only very subtle adjustments to the jet magnitudes and streamlines outside the supervised window. These changes remain

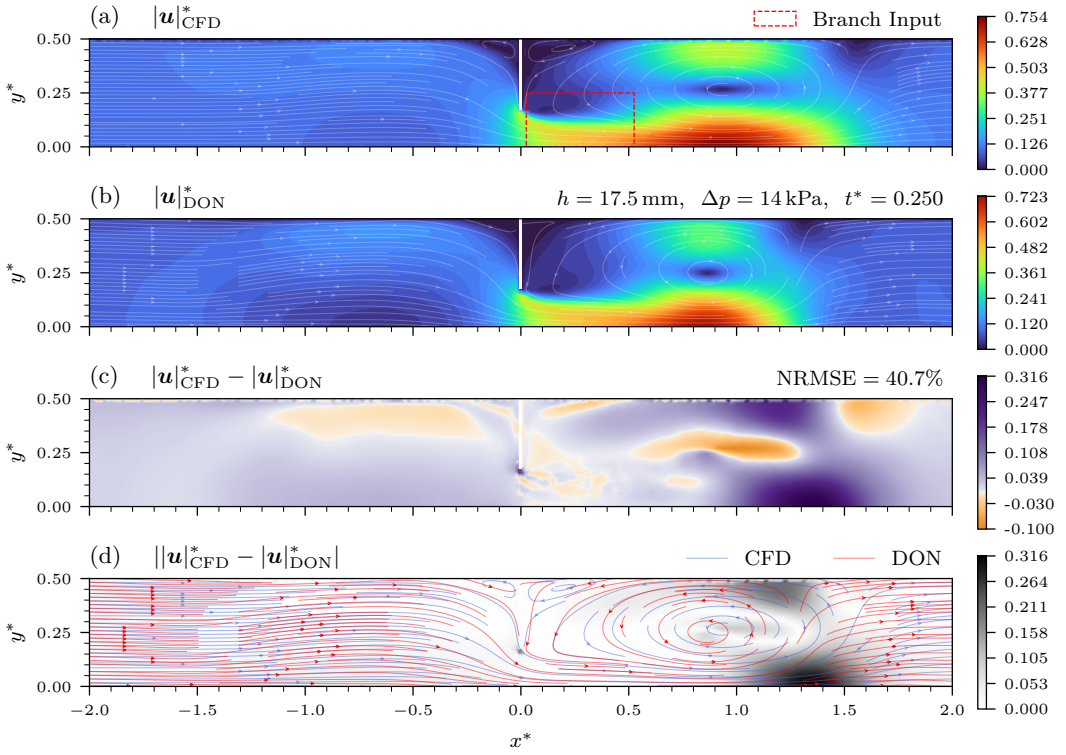

Figure 3.16.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

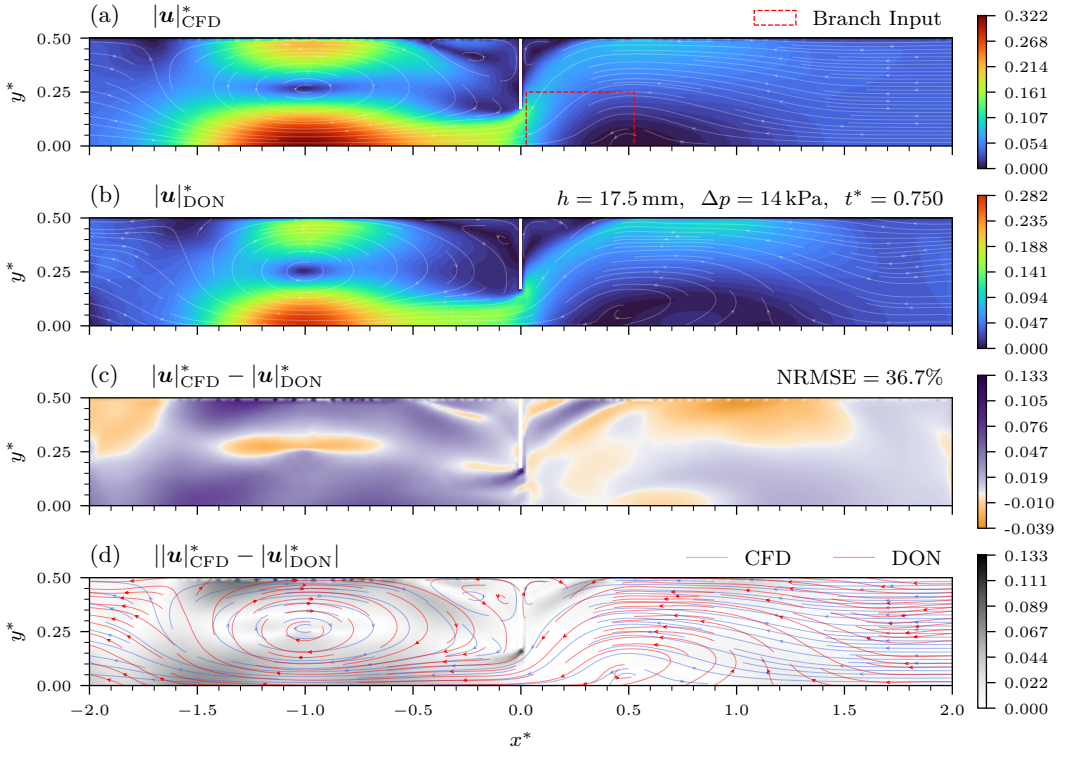

Figure 3.17.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.75$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

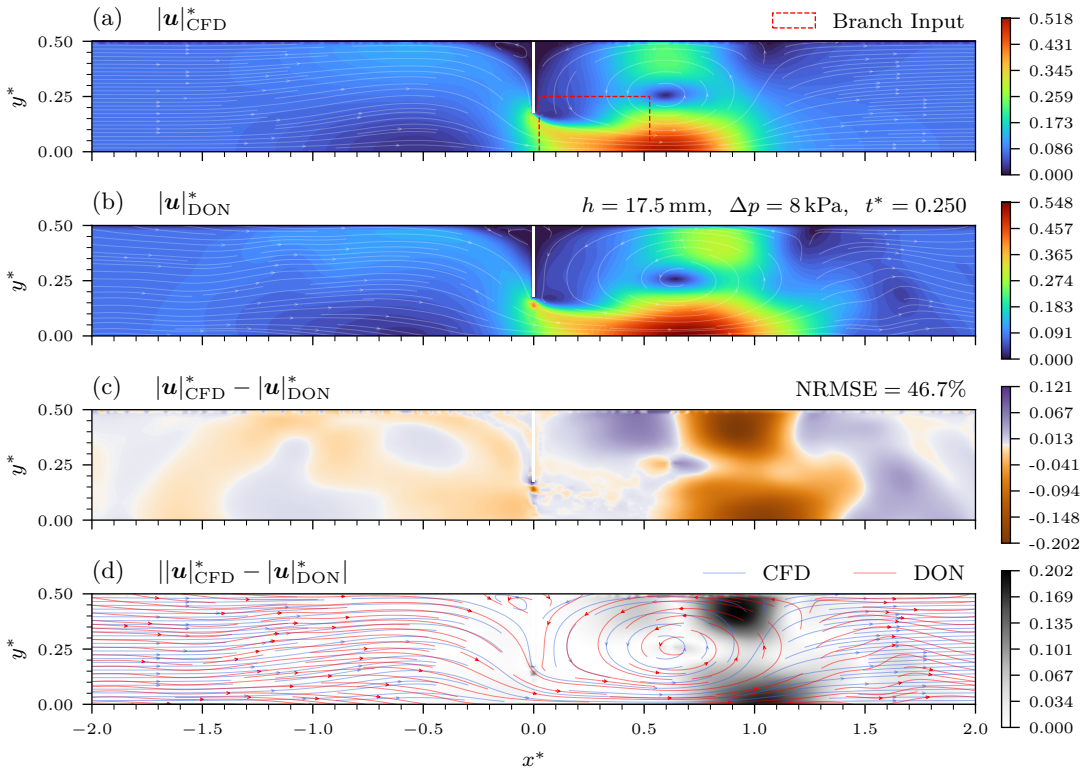

Figure 3.18.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 8\,\mathrm{kPa}$, $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

barely perceptible at a glance and likely exert minimal influence on the global error metrics. The error reduction within the window itself holds limited value, as this region allows for direct replacement by the measured input data during post-processing. Consequently, the distinct global improvement observed in the 2D-Steady case is not replicated to the same extent here. Nevertheless, the DeepONet predictions are decently accurate and exhibit strong temporal coherence, ultimately achieving a higher fidelity than the instance-specific PINN.

The "Extra2" and "Extra3" configurations, the cases with the lowest accuracy in the "Conti" baseline evaluation, serve to demonstrate the application of the hybrid framework to the most challenging extrapolation scenarios. While the temporal evolution of the flow generally remains physically plausible, specific transient intervals exhibit substantial deviations that drive up the aggregate error metrics. Figures 3.18 and 3.19 illustrate worst-case snapshots to highlight limitations of the current approach. Figure 3.18 displays the low-pressure "Extra2" case ($h = 17.5\,\mathrm{mm}$, $\Delta p = 8\,\mathrm{kPa}$), where streamlines unexpectedly diverge downstream of the regurgitation jet. Additionally, the jet appears elongated and is surrounded by erratic overpredictions of velocity magnitude. Supplementary figures in the Appendix (C.28, C.29) reveal that the magnitude of the early diastolic jet is significantly overpredicted, while the local pressure peak during diastole similarly exhibits excessive magnitude and temporal misalignment. The high-pressure "Extra3" case ($h = 10\,\mathrm{mm}$, $\Delta p = 16\,\mathrm{kPa}$) presented in Figure 3.19 reveals a visibly elongated vortex during early diastole, effectively appearing like two distinct flows overlaid. This feature subsequently collapses into the shorter form and remains spatially lagged behind the ground truth jet. The temporal evolution of the NRMSE for these additional configurations is presented in Appendix Figures C.30 through C.32. The interpolation case achieves a cycle-aggregated

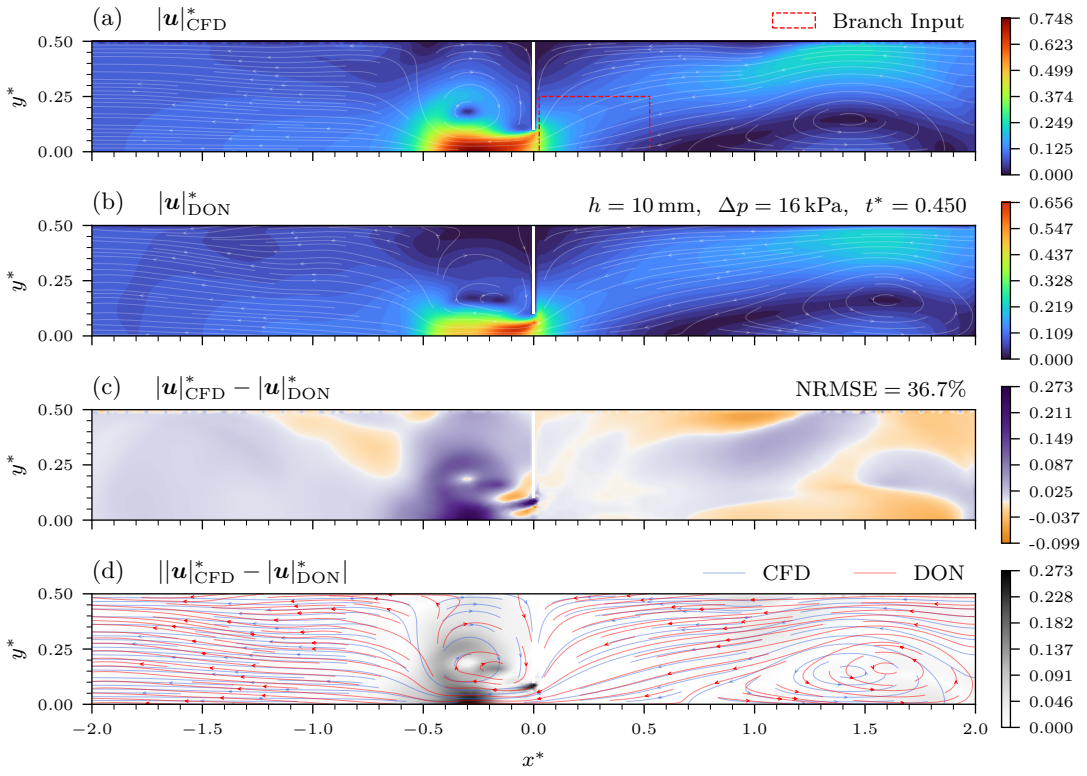

Figure 3.19.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 10\,\mathrm{mm}$, $\Delta p = 16\,\mathrm{kPa}$, $t^* = 0.45$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

NRMSE for the u-velocity of 18.7%, outperforming the 25.8% of the primary validation case, while the more challenging extrapolation scenarios yield higher errors of 39.0% and 41.2%. The trajectories echo the behavior observed in the previous temporal analysis, where the hybrid fine-tuning yields minimal improvement to the instantaneous performance outside the specific intervals where the PIV window captures the regurgitant jet; the distinct artifacts identified in the worst-case snapshots remain unaddressed by the adaptation step.

In addition to the data-driven hybrid models, the "Hybrid RANS" configuration, which incorporates momentum residuals during the fine-tuning phase, was evaluated. As indicated in Figure 3.11, this model achieved lower errors for u and ν_t, but the difference compared to the "Hybrid DD" model proved negligible. Therefore, the simpler data-driven configuration was selected as the primary framework. The inclusion of the RANS equations did not induce significant changes. This suggests that the operator's background training ($\mathcal{L}^{\mathrm{train}}_{\mathrm{data}}$) dominates the solution or that the RANS formulation faces similar convergence challenges to those observed in the pure PINN baseline. While extended training might eventually allow the physics losses to exert greater influence, the current termination strategy prioritizes optimal validation performance, which occurs rapidly in both setups. This represents a simplification compared to the optimal 2D-Steady strategy, which successfully leveraged additional no-slip, symmetry, and continuity terms during the adaptation step. Further optimization using these auxiliary constraints remains a potential avenue for enhancement.

3.2.3. Interim Conclusion

In summary, the investigation of the transient 2D channel flow establishes the DeepONet framework as the superior approach for reconstructing pulsatile flows from sparse data,

significantly outperforming the instance-specific PINN. While the PINN demonstrates the theoretical capability to model the flow without prior training data, it struggles to maintain physical consistency in the global flow topology and fails to capture the correct temporal evolution of the jet. Conversely, the DeepONet leverages the training distribution to provide a generally stable prediction foundation. Evaluation reveals a key distinction from the 2D-Steady results: in the transient domain, the hybrid adaptation acts primarily as a local constraint. Corrections are largely confined to the supervised window, where they are least needed, rather than propagating global improvements to the flow physics in extrapolation regions. Consequently, random artifacts and temporal misalignments persist, underscoring the data-driven model's inherent dependence on the training distribution's coverage.

Several avenues for enhancing predictive fidelity remain. The spatial resolution of the interpolated training data and the size of the branch input vector could be increased to match the density used in 2D-Steady. Some information in the training data and branch input may be lost as a result of the interpolation to a grid. Furthermore, optimizing the RANS implementation to overcome the practical ineffectiveness observed in both 2D-Steady and 2D-Transient could provide better results than the simple continuity constraint currently employed. Moreover, analogous to how the MLP architecture flattens spatial dimensions into a vector, the snapshot-based framework processes time instants in isolation and without special consideration of the periodic nature of the cardiac cycle, effectively discarding temporal information; both of these could be addressed. Alternatively, the DeepONet's robustness could be directly improved by expanding the training distribution. Despite these potential optimizations, the "Hybrid DD" DeepONet approach currently demonstrates promising predictive fidelity and is evaluated for 4D hemodynamics of the simulator in the following section.

3.3. 3D-Transient

The final validation section addresses the 3D-Transient flow within the hemodynamic simulator. This setup introduces significant flow complexity compared to the 2D channels. The single-plane PIV measurement at $z^* = 0$ represents an imperfect observation, as multiple 3D flow configurations could theoretically produce similar 2D footprints. This ambiguity makes the reconstruction task highly ill-posed.

3.3.1. PINN

The evaluation begins with the "PINN RANS" model, trained on the DropXL case. Figure 3.20 presents the velocity magnitude prediction at $t^* = 0.2$ during systole. At first glance, the reconstructed regurgitant jet appears decent, largely due to the data fitting area being relatively large. However, a closer inspection reveals several significant issues with the prediction.

Firstly, the streamlines lack detail, appearing more homogeneous and parallel than the reference, particularly within the LV. Similarly, the velocity magnitude distribution misses fine-scale features observed in the CFD ground truth. Most crucially, the model largely ignores the presence of the walls. The jet clearly penetrates the geometry, whereas the ground truth demonstrates how the flow should roll off along the wall in all directions. This failure to respect the wall BC is further evidenced in other regions, such as around the aorta, where the velocity magnitude clearly does not approach zero. The geometry is only discernible in the visualization because it was manually cut out during post-processing. The limitations extend to the temporal evolution, as shown in Appendix Figure C.33 for the diastolic phase. Similar to the 2D-Transient results, the reconstruction of the jet remains inaccurate during this phase. The velocity magnitude rises and falls immediately adjacent to the MROP, yet the model fails to form a detached jet with significant magnitude that travels towards the left ventricular wall. Consequently, the LA, which contains the supervised window, exhibits better performance than the LV. However, the prediction quality for both regions is inferior to the 2D-Transient results. It is worth noting that the 2D-Transient domain lacked a wall at the end of the jet path, so whether that model would have respected the no-slip condition in a bounded domain remains a matter of speculation. The most critical limitation of the current PINN implementation is illustrated in Appendix Figure C.34 The solution fails to propagate into the surrounding domain along the z-direction. A sharp discontinuity appears at $z^* = 0$, corresponding to the supervised data plane. While the flow features exist within this slice, the velocity magnitude drops to effectively zero immediately outside of it.

Investigations into the loss function weighting suggest this balance acts as an obstacle. The no-slip condition becomes visible only after substantially reducing the weights for the sparse velocity observations. However, this adjustment leads to a considerable deterioration in data-fitting accuracy; the network struggles to satisfy both constraints simultaneously. Furthermore, introducing a second parallel supervision plane results in negligible improvements and fails to extend the solution volumetrically. Overall, the PINN approach fails to produce satisfying results here.

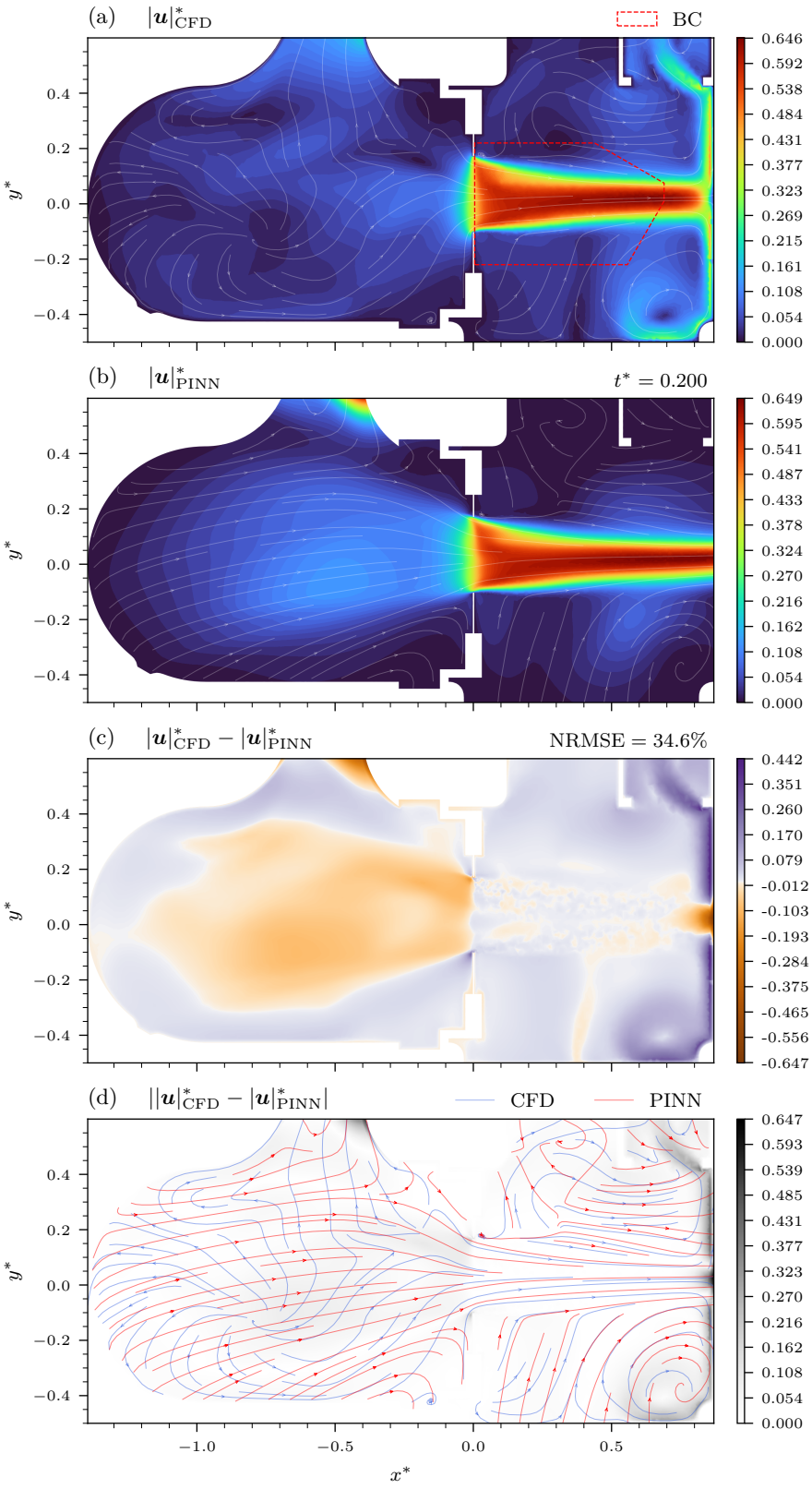

Figure 3.20.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.2$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

3.3.2. DeepONet

The aggregate performance of the DeepONet configurations is summarized in Figure 3.21. A comparison with previous results reveals that the training case errors (a) are noticeably higher than those observed in 2D-Steady and 2D-Transient, reflecting the increased difficulty of fitting the complex 3D transient data. Conversely, the pressure field retains a low fitting error. As detailed in the methodology, the experimental and CFD setups were calibrated to match physiological pressure targets, resulting in fields that are similar while also being largely homogeneous within the LV and LA chambers. The validation case metrics are no longer combined into one, since only two remain. Their NRMSE values are comparable to the more challenging 2D-Transient "Extra2" and "Extra3" cases. In addition to the potential for temporal misalignment, this dataset uses CFD data points directly, which are spatially biased towards areas of interest, those that have mesh refinements to accommodate higher velocities and gradients. Consequently, larger absolute deviations appear more often in the metrics calculation, which should be considered when comparing to 2D-Steady and 2D-Transient.

The "DD Smaller" model, which shares the 8×512 size of the successful 2D-Transient

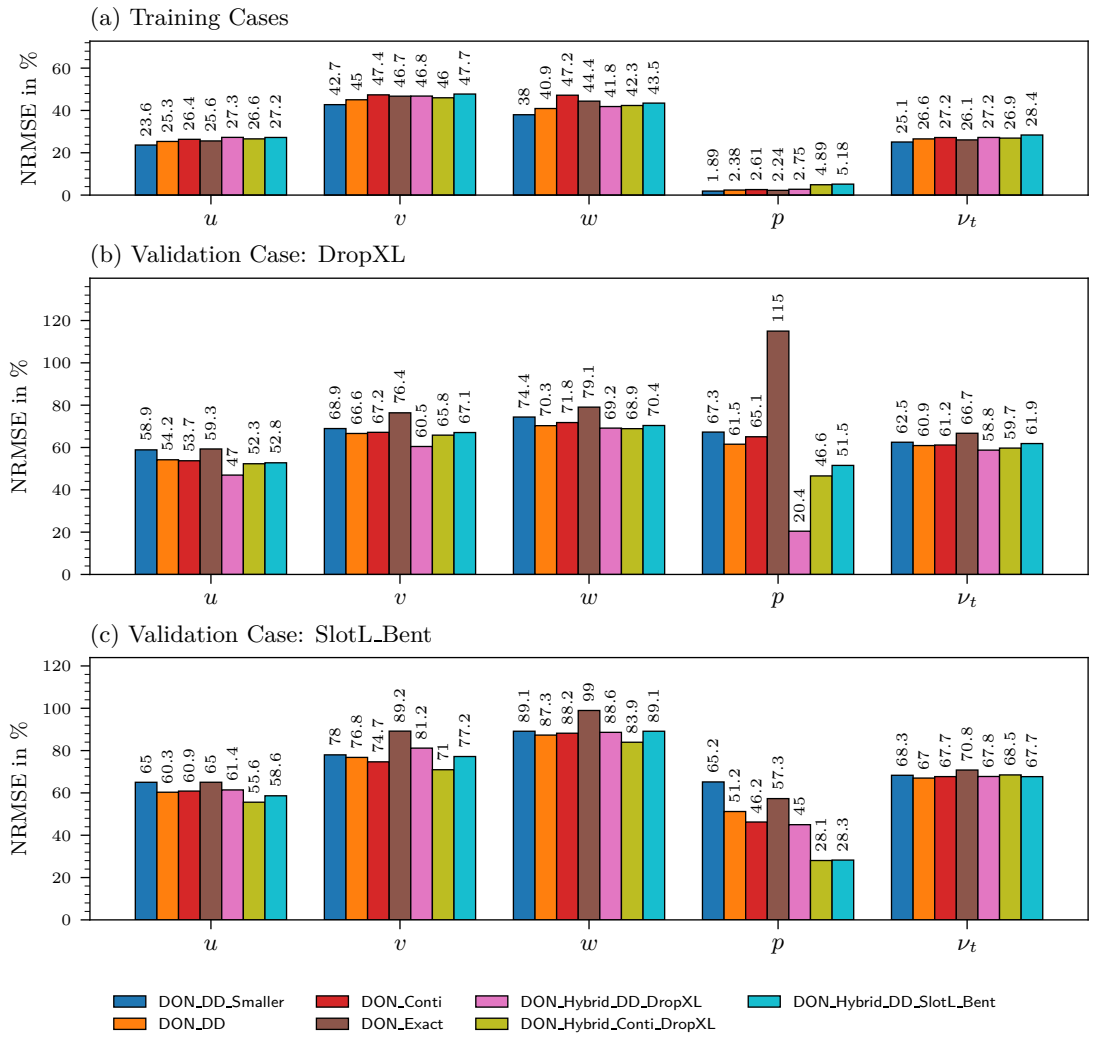

Figure 3.21.: Comparison of NRMSE for each 3D-Transient ANN output across various trained DeepONets. The charts display error percentages aggregated over the cardiac cycle for (a) training cases and the two (b, c) validation cases.

DeepONet, already yields results superior to the PINN. Increasing the network capacity to the 9×640 "DD" configuration provides further gains. This change improved the visual fidelity of the jet and is reflected in the reduced error metrics shown in the barplot. The inclusion of physical constraints is investigated through the "Conti" and "Exact" configurations. The "Conti" model yields NRMSE values very similar to the data-driven baseline. While tests with larger batch sizes or effective weights induced a smoothing effect on the fields, confirming the loss term's impact, further tuning was constrained by computational limits. The memory demand scales with the number of collocation points evaluated per branch input snapshot. Overcoming this bottleneck would necessitate either increased hardware resources or a more efficient code implementation, both of which fell outside the scope of this study. The "Exact" architecture, which utilizes a vector potential formulation to intrinsically satisfy the continuity equation, similarly fails to provide a benefit. Visually, the jet reconstruction degrades, and the error metrics in Figure 3.21 are noticeably higher than "DD". This performance drop could stem from the additional gradient pass required to calculate velocity from the potential, which complicates the optimization landscape. The impact of these approaches on the physics compliance is visualized in the PDE residual analysis provided in Appendix Figure C.35. While the "Exact" model

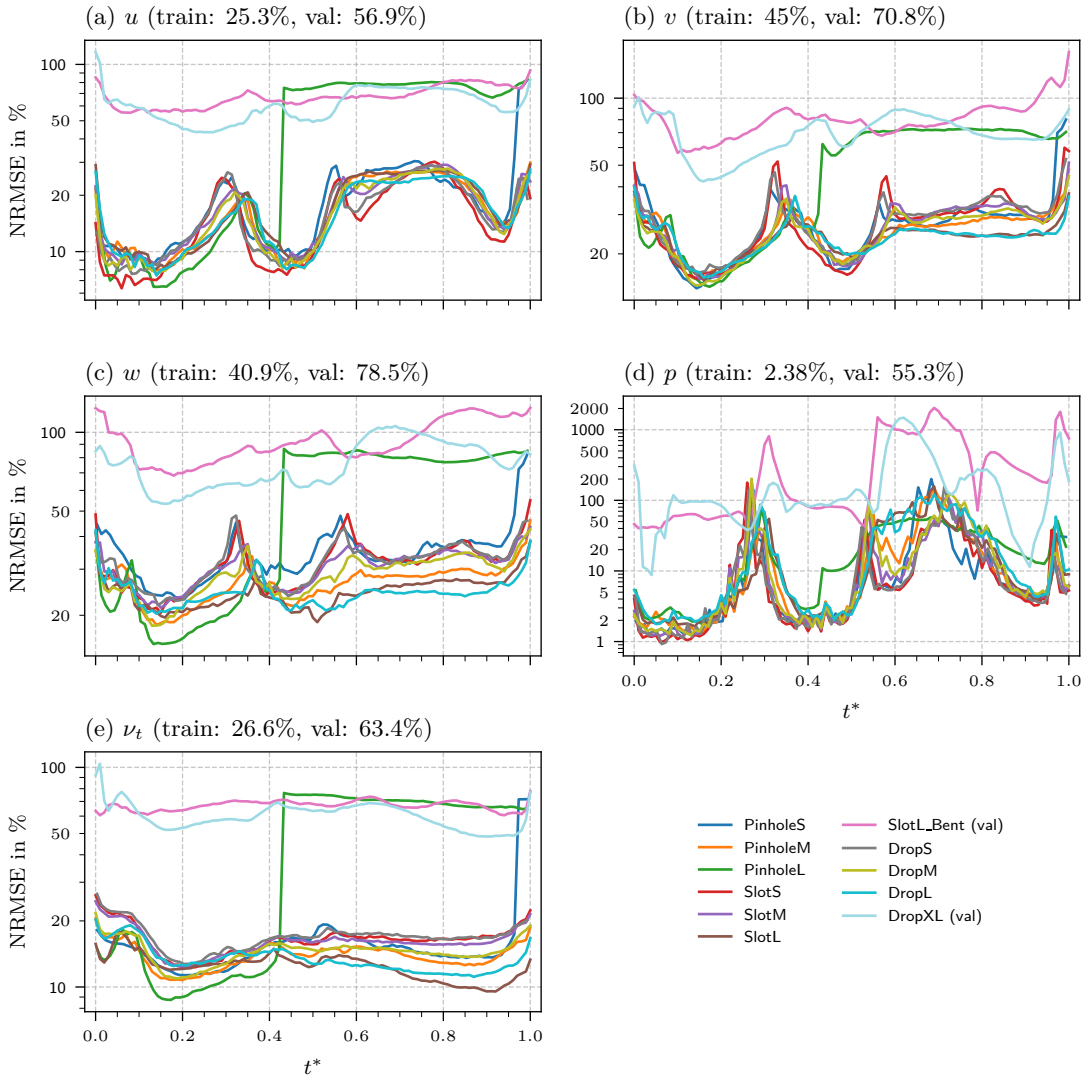

Figure 3.22.: Temporal evolution of NRMSE for all fields over the cardiac cycle for the 3D-Transient "DD" model. Each line is a different CFD case.

achieves near-perfect satisfaction of the continuity equation, barring numerical noise, this mathematical rigor does not translate into a tangible improvement in matching the ground truth flow. This PDE residual bar plot further confirms that the "Conti" configuration reduces the continuity residual compared to the "DD" baseline. Furthermore, the standard "DD" model consistently yields lower residuals than the "DD Smaller" variant, aligning with the improved accuracy observed in the direct field comparisons. The x-momentum residual is significantly higher than that of the y and z components; the flow is dominated by the high-velocity u-component of the jet. Finally, the disparity between training and validation residual percentages is notably smaller here than the gap observed in the NRMSE metrics. While a correlation between residual magnitude and predictive fidelity exists, this reduced separation makes it difficult to gauge the absolute quality of a validation result based solely on the PDE residuals, which becomes important for the evaluation of experimental results in Chapter 4.

Figure 3.22 provides the temporal context for the aggregate values presented in the bar plot, illustrating how the NRMSE evolves over the cardiac cycle for the "DD" model. The separation between the training distribution and the unseen validation cases remains clearly defined throughout the period. A distinct anomaly appears in the PinholeL training case, where the error spikes suddenly after $t^* = 0.4$. While not investigated further, this behavior resembles the transient artifacts encountered in the 2D-Transient analysis. Generally, the cases track closely together, with the error magnitudes oscillating in response to the changing flow physics. These fluctuations stem from the interplay between the flow conditions and the error metric itself. Phases characterized by low magnitudes can artificially inflate the relative error, while high-velocity intervals may lead to larger absolute deviations. Additionally, specific parts of the cardiac cycle involve inherently more complex flow dynamics, presenting a greater challenge for the operator to resolve compared to quiescent phases. Analogous to the NRMSE bar plot, a version of this figure that instead concerns the PDE residuals is included in the Appendix Figure C.36. Here, the distinct separation between training and validation sets disappears. The validation cases blend into the cluster of training trajectories, occasionally exhibiting even lower residuals than some supervised instances.

As the previous bar plot 3.21 shows, the fine-tuned "Hybrid" models slightly improve the aggregate NRMSE over the baseline "DD" model for their respective cases. Like with 2D-Transient, it is a modest difference. Figure 3.23 provides a direct temporal comparison for the DropXL validation case. The most substantial error reduction appears during the systolic phase ($t^* < 0.3$), coinciding with the development of the regurgitant jet within the supervised region. While the numerical performance tracks the baseline closely for several intermediate time steps, a distinct improvement emerges again during the later diastolic phase, starting around $t^* = 0.5$. These reductions in global error likely stem from the enhanced fitting accuracy within the observation window.

The "Hybrid Conti" model was evaluated to determine if including the continuity PDE loss term during the adaptation phase improves the solution. The aggregate error metrics in Figure 3.21(b) indicate a performance regression compared to the "Hybrid DD" approach. Figure 3.23 details this behavior over the cardiac cycle, showing consistently higher errors for the physics-informed model. The pressure trace highlights a numerical characteristic of the NRMSE formulation (Appendix 5.1). The instantaneous error spikes to extreme values during phases of low or uniform values. In these intervals, the standard deviation in the denominator approaches zero, causing even negligible absolute errors to produce massive relative values. The aggregate metric remains low because it is not calculated as the average of these instantaneous ratios but is instead normalized against the total variance of the entire sequence, which is dominated by the high-magnitude phases. The PDE residual evolution is shown in Appendix Figure C.37. While the continuity adherence

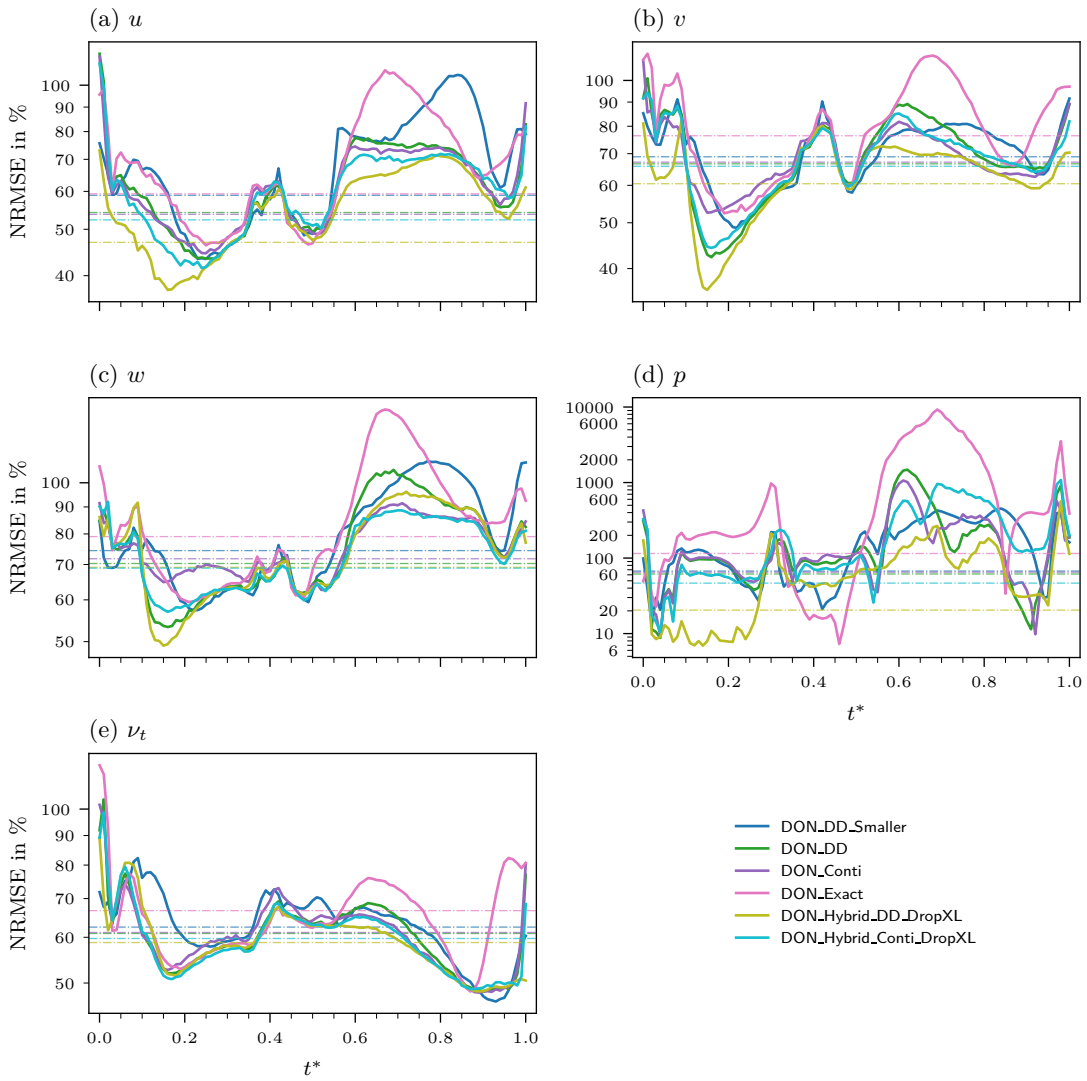

Figure 3.23.: Temporal evolution of the NRMSE for all fields over the cardiac cycle for the 3D-Transient DropXL case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

improves, the x-momentum residuals increase, indicating a deterioration of the global flow physics. This effect is visually confirmed in the snapshot provided in Appendix Figure C.47. The constraint imposes a strong smoothing effect that diminishes the jet magnitude and hinders the model from accurately fitting the data even within the supervised window. Although further hyperparameter tuning might reduce this smoothing to a beneficial effect, the potential gains are expected to be moderate based on previous 2D-Steady and 2D-Transient results. Furthermore, training a full "Hybrid RANS" model proved computationally prohibitive, as the memory requirements exceeded the presently available GPU capacity even at minimum batch sizes. Consequently, the purely data-driven "Hybrid DD" framework is selected as the optimal approach.

The spatial reconstruction capabilities of the "Hybrid DD" model are visualized in Figure 3.24 for the systolic phase at $t^* = 0.25$. The model reconstructs the high-velocity regurgitant jet with high fidelity, matching both the magnitude and the spatial extent of the CFD reference. In strong contrast to the PINN results, the prediction respects the complex boundaries; the jet interacts naturally with the atrial walls, rolling off the geometry rather

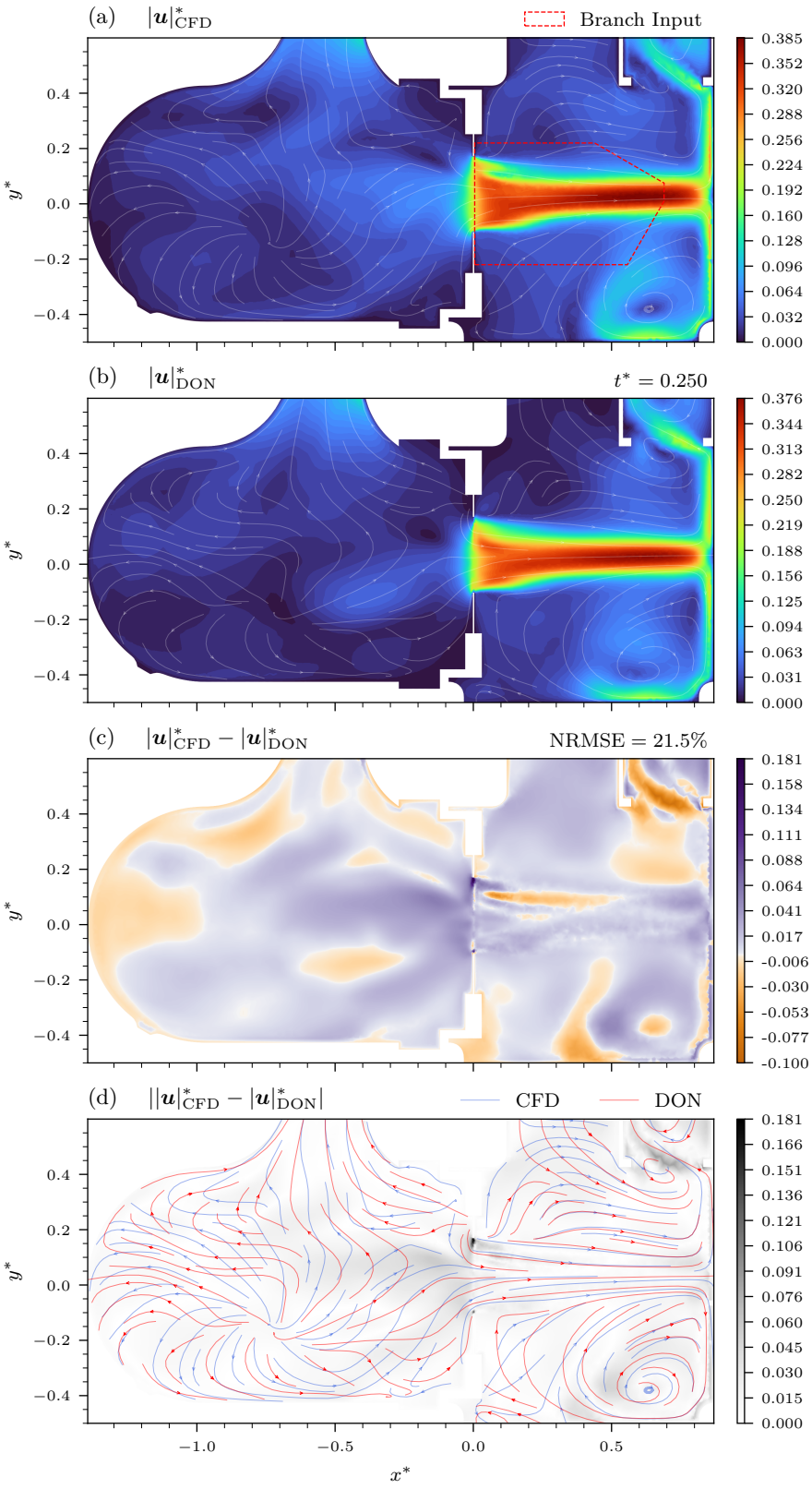

Figure 3.24.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

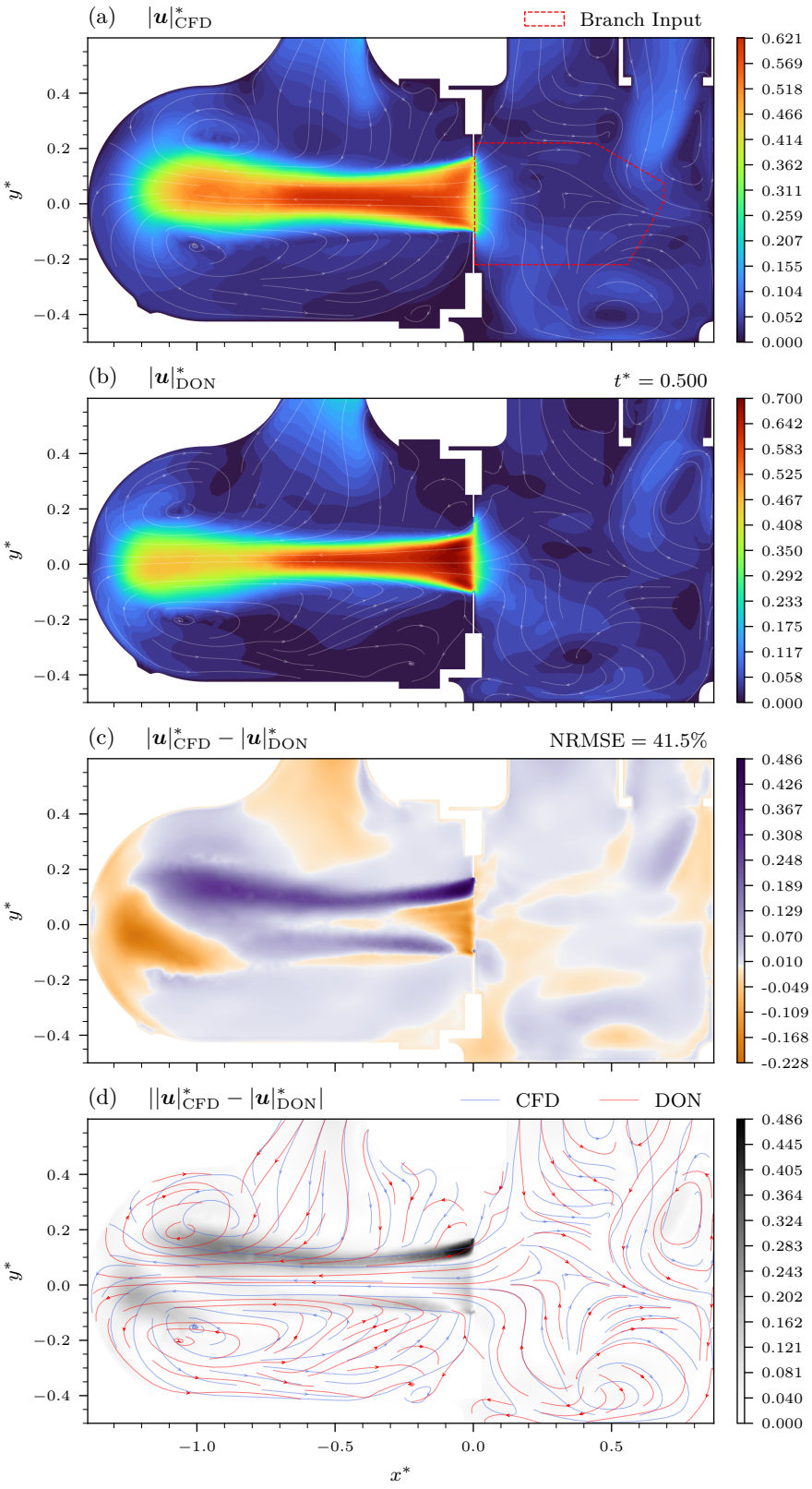

Figure 3.25.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.5$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

than penetrating it, resulting in low error magnitudes along the surface. Despite this high overall accuracy, exact streamline matching across the entire complex 3D domain proves more challenging than in the simplified 2D setups. Deviations are visible in secondary flow features, such as the upward path connecting the LV to the reservoir. A comparison with the baseline "DD" prediction provided in Appendix Figure C.38 confirms that the hybrid fine-tuning induces minimal structural changes outside the supervised PIV window. This reinforces the observation that the accurate global topology is primarily derived from the pre-trained operator, while the adaptation step locally refines the solution to match the specific experimental observation. To demonstrate the general reconstruction quality of the remaining flow variables, Appendix Figures C.43 through C.46 display the predictions for v, w, p and ν_t using the current snapshot. The diastolic jet is visualized in Figure 3.25 at $t^* = 0.5$. The operator successfully captures the overall topology, with streamlines largely matching the ground truth. A distinct region of underprediction persists as a thick streak emerging from the orifice. This error stems from the extrapolation to a larger orifice size than encountered during training, affecting the LV side where the sparse supervision provides no direct correction. A comparison with the baseline "DD" prediction in Appendix Figure C.39 reveals that the hybrid adaptation improves the jet magnitude and streamline alignment outside the PIV window in this phase in a slightly more noticeable but still minor way.

The ability of the framework to infer the full 3D flow field from 2D data is assessed in Figure 3.26, which displays the velocity magnitude on the y-plane at $t^* = 0.05$. The model successfully generates a volumetric solution that extends into the depth of the domain. However, a comparison with the baseline "DD" prediction in Appendix Figure C.40 reveals that the fine-tuned result appears less physically cohesive despite achieving a lower numerical error. This stems from the localized nature of the hybrid update. The baseline prediction positions the jet pulse ahead of the ground truth. The fine-tuning step accurately retracts the jet within the supervised $z^* = 0$ plane to match the observations. Crucially, this correction does not propagate uniformly across the full z-width of the jet. The outer fluid layers remain in their original advanced position, creating a disjointed appearance in the cross-section. Furthermore, the underprediction streak caused by the orifice size extrapolation manifests in the LA, as the single-plane supervision leaves these off-center regions unconstrained. In other phases of the cardiac cycle, the model achieves significantly higher volumetric fidelity. Appendix Figures C.41 and C.42 illustrate the y-plane predictions during systole and diastole, respectively. In these instances, the velocity magnitude distribution is more accurately recovered throughout the cross-section. The primary residual deviation remains the streaking artifact downstream of the orifice, resulting from the extrapolation of the gap height. These results indicate that the framework produces qualitatively robust 3D reconstructions, yet the restriction to a single input and fine-tuning plane at $z^* = 0$ imposes an inherent limit on the off-plane accuracy in specific transient states.

The generalization capability is further tested on the SlotL-Bent case. This geometry differs from the training set due to the bent orifice. Figure 3.27 visualizes the results for the systolic phase. The branch network input for this case is extracted from a location shifted in the positive x-direction. This adjustment accounts for the physical displacement of the orifice opening caused by the bent geometry. However, the supervision remains anchored to the original coordinate system to align with the training distribution, where the jet origin is invariant. This positioning ensures that the DeepONet infers the jet start in a location consistent with its prior learning. The experimental PIV data supervision is subsequently applied in the same manner in Chapter 4. The impact of this inherent geometric misalignment appears clearly in the error maps in Figure 3.27(c). A localized error spike exists at the jet origin because the predicted flow initiation point is shifted

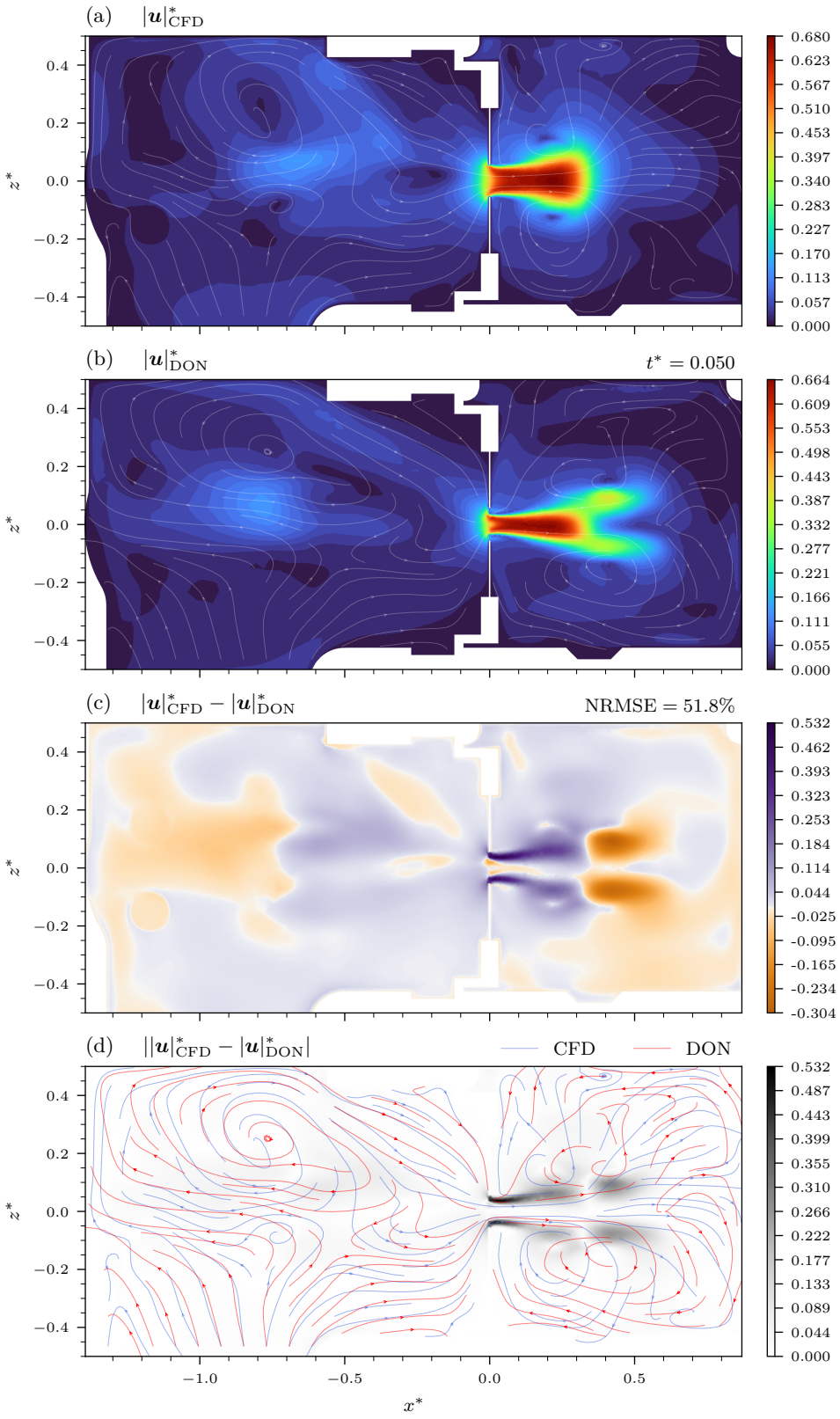

Figure 3.26.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for y-plane at $t^* = 0.05$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

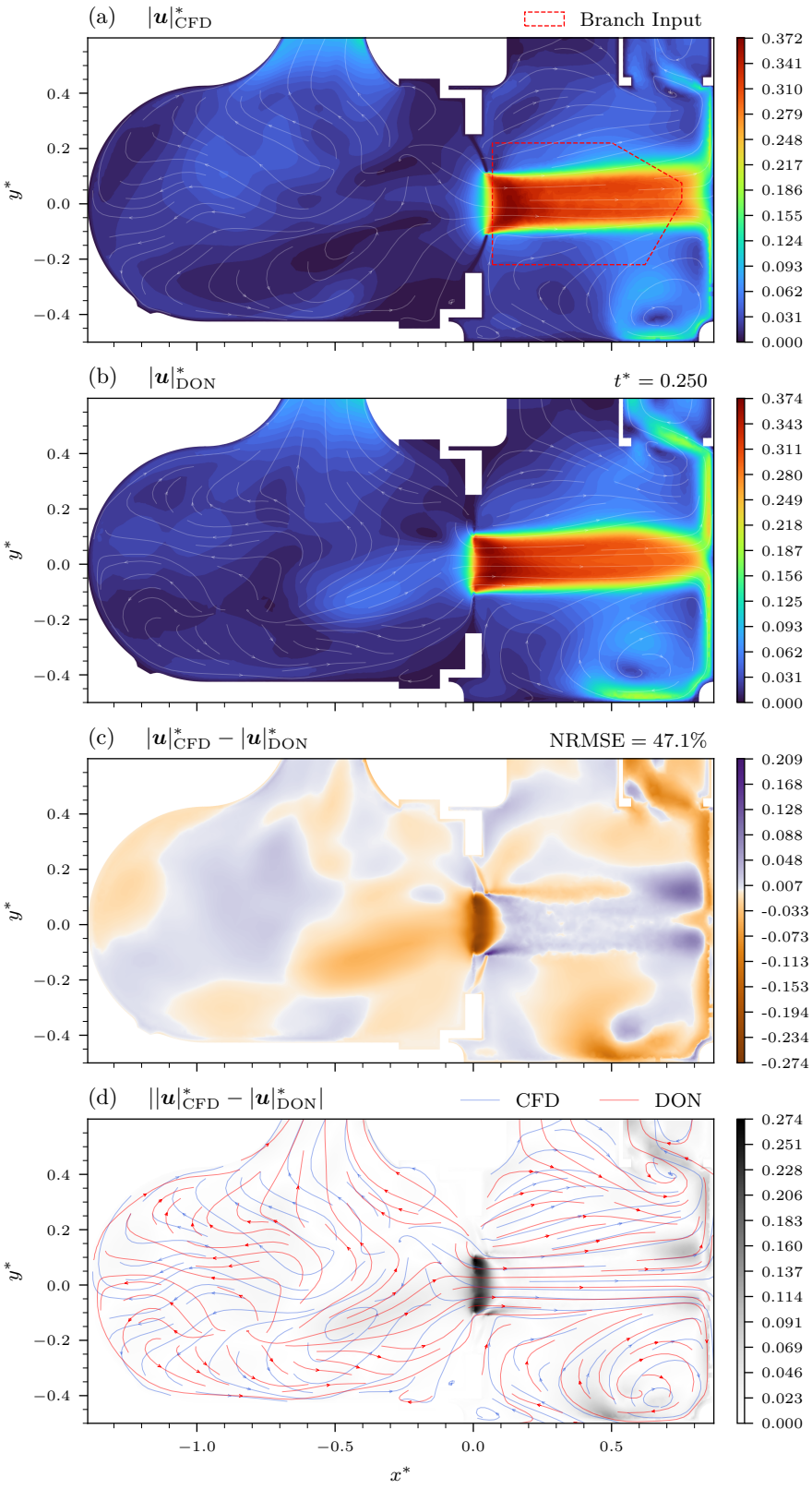

Figure 3.27.: (a) Reference 3D-Transient SlotL-Bent CFD velocity magnitude for z-plane at $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

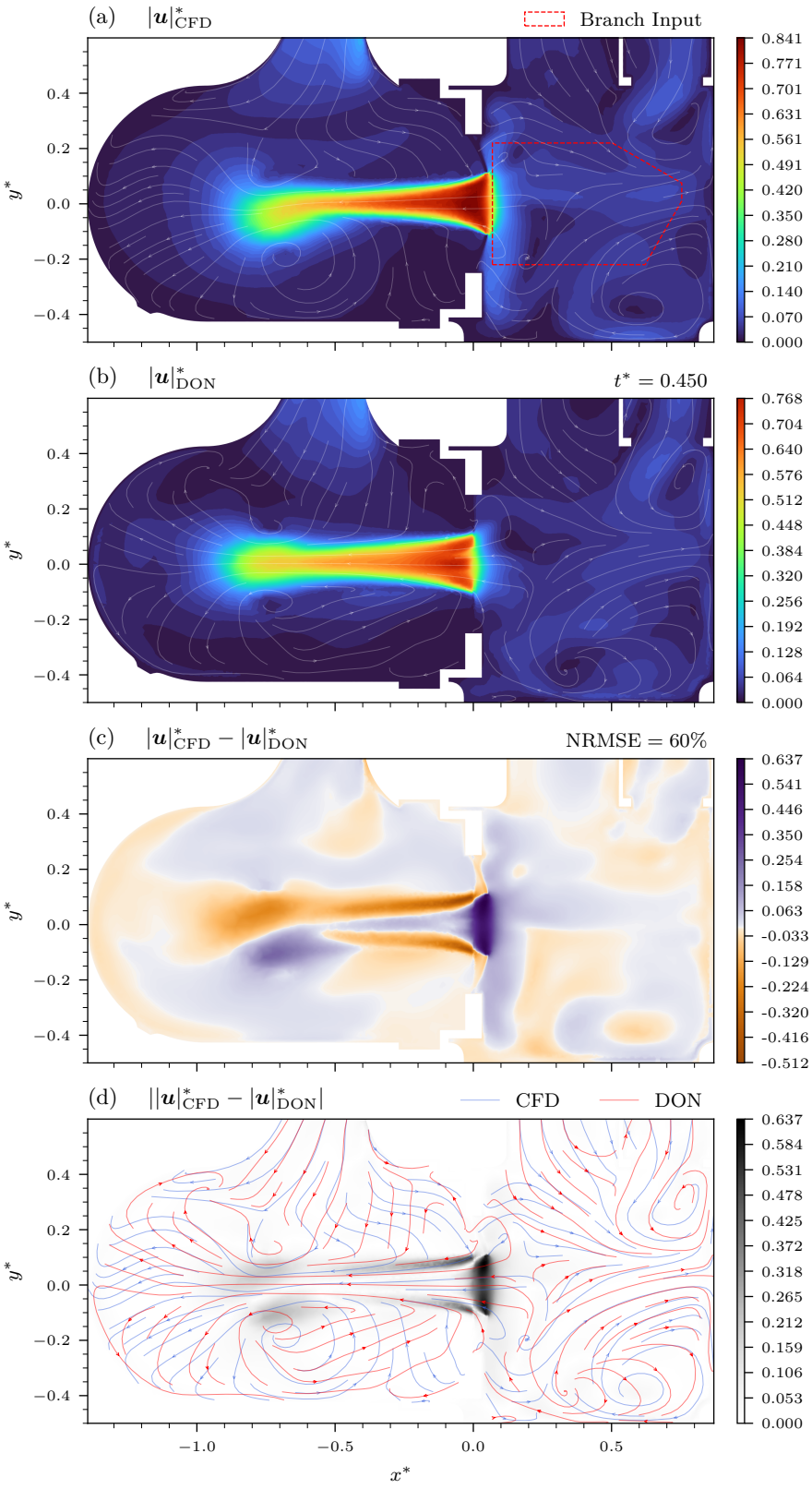

Figure 3.28.: (a) Reference 3D-Transient SlotL-Bent CFD velocity magnitude for z-plane at $t^* = 0.45$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

relative to the ground truth. Outside this specific region, the prediction quality remains robust. Figure 3.28 confirms this behavior for the diastolic phase. Here, the maximum velocity magnitude is underpredicted. This intensity drop actually lessens the NRMSE impact of the misalignment; a full-magnitude jet in the shifted position would produce a larger deviation. Distinct streaking artifacts are also visible. Unlike in the 2D-Steady, 2D-Transient, or DropXL cases, where such features stem from extrapolating unknown orifice heights, the flat SlotL geometry is included in the training distribution. Consequently, the streaking here is a result of the novel bent shape and the input coordinate shift. The temporal evolution of the NRMSE for this case is detailed in Appendix Figure C.48. The "DD" baseline clearly outperforms the "Exact" and "DD Smaller" architectures. The hybrid fine-tuning step yields minimal improvement here, and for some fields, the aggregate metrics actually exceed those of the baseline, a trend reflected in the bar plot 3.21(c). Regarding physical consistency, the temporal evolution of the residuals in Appendix Figure C.49 and the residual bar plot in Figure C.35(c) indicate that the metrics deteriorate following the hybrid adaptation. This increase in residuals is similarly observed for the DropXL case (C.37 and C.35,c) and likely stems from the disjointed 3D jet structure observed in the snapshot analysis.

3.3.3. Interim Conclusion

The validation on the 3D hemodynamic simulator dataset reinforces the conclusions drawn from the 2D cases while highlighting the unique challenges of volumetric reconstruction. The instance-specific PINN fails to solve the ill-posed problem of inferring a 3D flow field from a single 2D slice. Lacking the prior knowledge of the flow topology, the network overfits to the dense data window and fails to propagate the solution physically into the depth of the domain or respect the boundary conditions.

In contrast, the DeepONet framework successfully leverages the training distribution to reconstruct plausible 3D flow fields from sparse 2D inputs. The "Hybrid DD" approach proves to be the more effective strategy. While the inclusion of physical constraints like continuity or vector potential formulations offer theoretical benefits, they induce over-smoothing or optimization difficulties in practice without yielding tangible accuracy gains. Consequently, the purely data-driven architecture was selected as the optimal surrogate. The validation revealed that while the hybrid fine-tuning effectively corrects the solution within the supervised plane, it can introduce disjointed artifacts in the third dimension, as the correction does not always propagate uniformly across the volumetric jet width. However, the analysis of the SlotL-Bent case demonstrates the model's robustness to geometric variations and coordinate misalignments. This capability is useful for the transition to real-world application, where exact geometric matching is challenging. Having established the "Hybrid DD" DeepONet as a capable and robust tool on synthetic ground truth data, the following chapter applies this framework to the experimental PIV data to reconstruct the full 4D hemodynamics of the simulator.

4. Application to Experimental Data

Having established the superiority of the "Hybrid DD" DeepONet architecture in the validation chapter, this section applies the final framework to real experimental data [43] from the hemodynamic simulator. The primary objective is to assess the framework's capability to perform 2D-to-3D flow field reconstruction on real-world transient cases. Unlike in the validation phase, a complete 3D "ground truth" velocity field is unavailable for these experimental cases. Consequently, the evaluation focuses on the qualitative plausibility of the volumetric reconstruction and its physical consistency, which is assessed through the calculation of PDE residuals.

The application covers four distinct experimental configurations: PinholeL, SlotL, DropXL, and EccJet. For each case, the pre-trained DeepONet is adapted using the specific sparse PIV observations. A quantitative comparison against available measurements is performed for the EccJet case, where additional PIV planes at different z-depths were recorded and serve as a validation set. It is important to note that the experimental datasets are temporally sparser than the CFD simulations, containing between 25 and 50 irregularly spaced time steps instead of 300. Consequently, fewer snapshots are provided as branch input and utilized for supervision during the fine-tuning process. This chapter assesses whether the DeepONet can produce physically plausible and experimentally consistent 3D flow fields from these limited real-world observations.

4.1. Results

The physical consistency of the experimental reconstructions is evaluated using the PDE residuals, summarized in Figure 4.1. The experimental datasets are limited to up to 50 irregularly spaced snapshots; these available time steps predominantly capture the active regurgitation phase. As observed in the validation of 3D-Transient, residuals during these active phases are typically higher than during quiescent periods. Therefore, a direct like-for-like comparison with the cycle-aggregated CFD metrics requires caution.
Despite this bias towards challenging flow states, the residual magnitudes align closely with

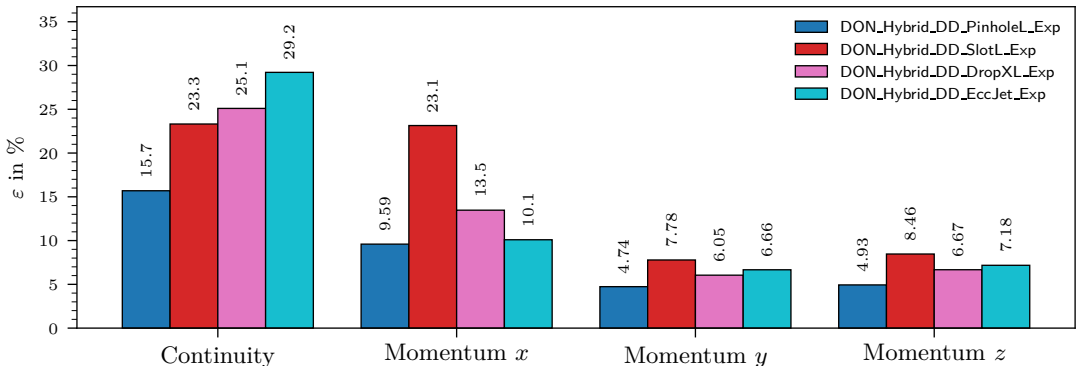

Figure 4.1.: Comparison of PDE residuals for the RANS equations across various trained 3D-Transient DeepONets. The chart displays residual percentages aggregated over the cardiac cycle for the fine-tuned experimental cases.

the validation results. For context, the CFD-based DropXL validation case yielded aggregate residual percentages of 22.8 (Continuity), 9.97 (x-Momentum), 4.21 (y-Momentum), and 4.77 (z-Momentum). The SlotL-Bent case exhibited values of 22.4, 32.4, 5.05, and 5.57, respectively (Figure C.35b,c). In comparison, the experimental applications in Figure 4.1 show ranges of 15.7–29.2, 9.59–23.1, 4.74–6.66, and 4.93–7.18. These values indicate that the PIV-based reconstructions achieve a level of physical adherence comparable to, and in some metrics even surpassing, the CFD validation cases, suggesting that the model maintains some level of consistency when driven by sparse real-world data. Figure 4.2 breaks down the aggregate metrics, illustrating how the PDE residuals evolve over the available time steps for each experimental case. This temporal resolution offers nuance to the favorable aggregate comparison observed in the bar plot. When compared to the equivalent evolution plots from the 3D-Transient validation (e.g., Figure C.36), the residual peaks in the experimental application tend to reach the upper bound of, or exceed, the highest values recorded for the CFD cases in the early systolic phases. This indicates that while the average physical inconsistency is comparable, the model struggles more significantly to resolve the governing equations during specific, highly dynamic instants when driven by the experimental input.

The visual analysis of the velocity magnitude reconstruction begins with the PinholeL case. As noted in Section 2.1.1.2, this configuration demonstrated the strongest agreement between the URANS validation and the experimental measurements. Figure 4.3 presents three snapshots of the predicted field. Subfigure 4.3(a) displays the z-plane at $t^* = 0.26$. This snapshot depicts a coherent and plausible flow topology. The streamlines are well-formed, and the regurgitant jet extends naturally to the right LA wall, where it rolls off

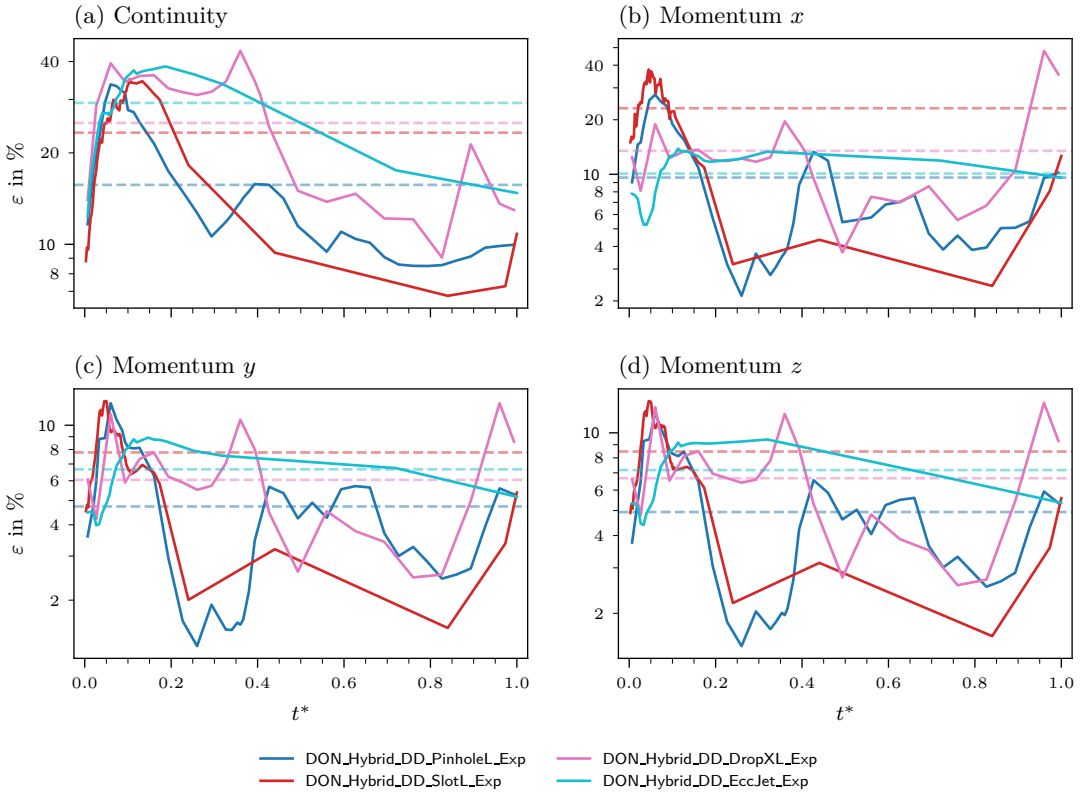

Figure 4.2.: Temporal evolution of the PDE residuals for the RANS equations over the cardiac cycle for the 3D-Transient "Hybrid DD" model. Each line is a different fine-tuned experimental case.

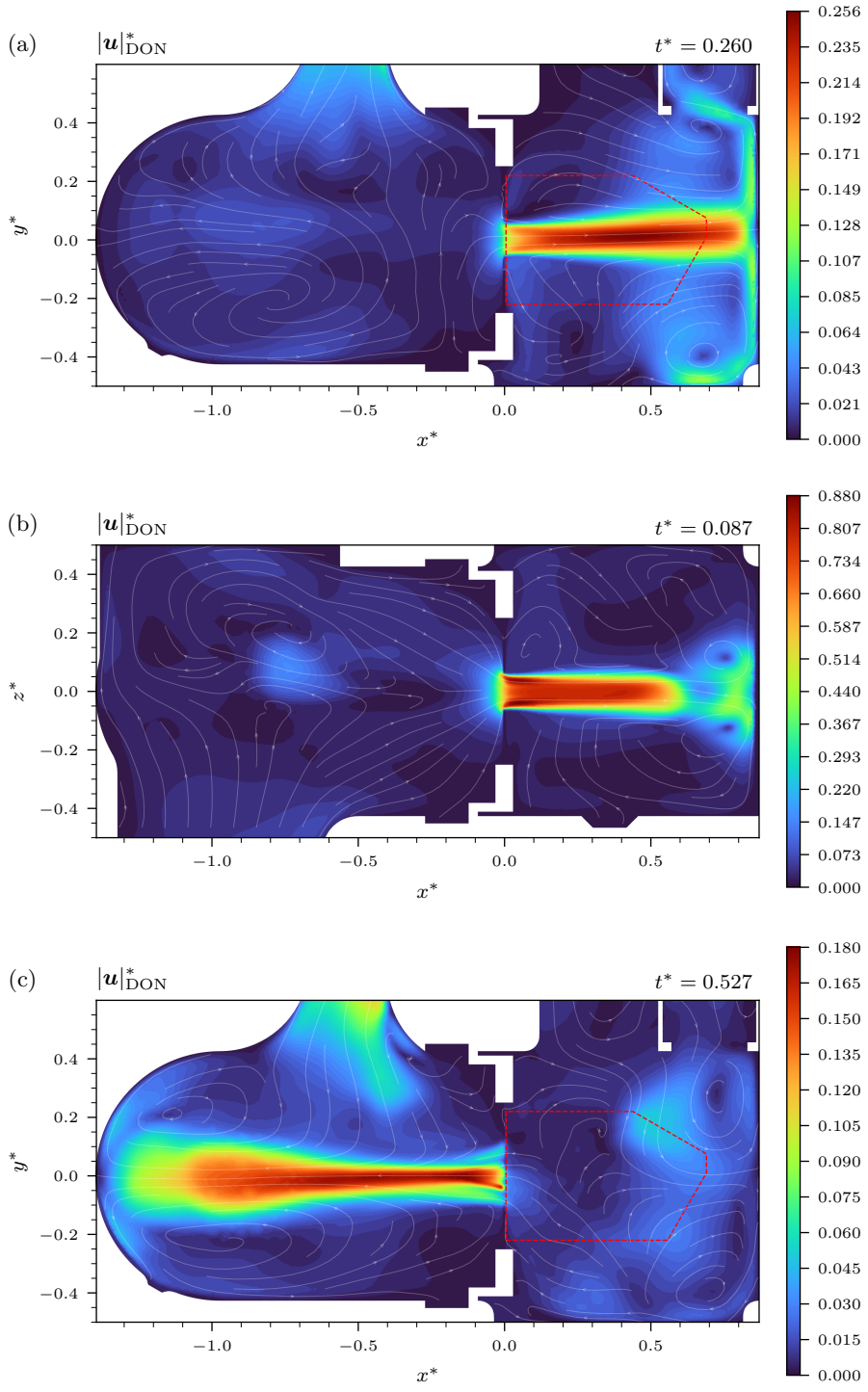

Figure 4.3.: "Hybrid DD" velocity magnitude predictions for 3D-Transient PinholeL Exp, (a) z-plane for $t^* = 0.26$, (b) y-plane for $t^* = 0.087$ and (c) z-plane for $t^* = 0.527$.

the surface as expected physically. However, the phases preceding this stable state exhibit noticeable artifacts not captured in this single frame. Specifically, the jet connects to the wall prematurely in a speckled, disjointed manner. While the jet tip remains clearly visible within the supervised area, high velocity magnitudes appear simultaneously to the right of this region. Consequently, the vertical structures along the wall, which correspond to the circular spreading of the flow on the flat surface (visible as vertical roll-off in 2D), emerge earlier than the main jet body arrives.
This phenomenon is explicitly captured in subfigure 4.3(b), which shows the y-plane at $t^* = 0.087$. Here, the premature connection to the wall is clearly visible. Furthermore, the flow vertices appear near the wall, significantly ahead of the actual jet tip. This artifact mirrors the specific limitation identified during the validation of 3D-Transient: the hybrid fine-tuning effectively corrects the jet propagation within the supervised $z^* = 0$ plane, but this correction does not propagate laterally to adjust the full z-width of the jet. As a result, the outer layers of the inferred 3D jet remain in their original, uncorrected positions, shown in Appendix Figure C.50, creating a disjointed topology. These structural inconsistencies coincide temporally with the spikes observed in the PDE residual evolution (Figure 4.2).
Subfigure 4.3(c) illustrates the diastolic phase at $t^* = 0.527$. Here, a degradation in prediction quality is evident compared to the CFD validation. While streaking artifacts are present in the 3D-Transient validation, the issue manifests differently and more severely in current context. The snapshot displays a strong central jet core surrounded by a diffuse region of weaker velocity magnitude. It is important to distinguish this artifact from those observed in the validation chapter; it does not stem from geometric factors such as a bent nozzle or orifice height extrapolation, as the PinholeL geometry is represented in the training set. Rather, this streaking appears to be a specific byproduct of using the experimental data. This specific degradation likely points to a challenge associated with the domain shift between the training distribution and the experimental input. The baseline DeepONet is trained exclusively on idealized synthetic data derived from URANS simulations. In contrast, the PIV measurements are subject to inherent experimental noise and statistical errors arising from the phase-averaging acquisition process. As the training phase does not employ augmentation strategies to prepare for these imperfections, the network appears sensitive to the discrepancy in signal characteristics.

The analysis of the SlotL experimental case begins with the physical consistency metrics. The PDE residuals presented in the bar plot (Figure 4.1) and the temporal evolution (Figure 4.2) indicate a performance degradation compared to the PinholeL case. Notably, the x-momentum error is the highest among all experimental configurations. This case's data finely captures the complex initial development of the jet. The experimental flow features a double vortex ring formation [43], a complex topological structure that is not present in the initial training distribution and is now driving the model's output through the branch input layer. Figure 4.4 visualizes the velocity magnitude reconstruction, presenting a pattern of artifacts similar to those observed in the PinholeL analysis. Subfigure (a) displays the z-plane prediction, which appears qualitatively reasonable; the jet is well-formed and connects to the atrial wall, albeit too narrow near the junction. However, the temporal progression reveals that this connection to the LA wall occurs prematurely again. Subfigure (b), showing the y-plane, confirms this early impact. Again, the structure is disjointed in z-direction, and the vortices are ahead of the jet due to the test-time adaptation correcting the baseline model's prediction insufficiently.

The PDE residual analysis for the DropXL case, shown in Figure 4.1, reveals an increase in the continuity error compared to the PinholeL and SlotL results. The momentum residuals, however, remain within the intermediate range of the experimental cohort, indicating a middling performance in terms of dynamic consistency relative to the other cases.
Figure 4.5 presents inferred velocity fields. Subfigure (a) depicts the regurgitant jet in the z-

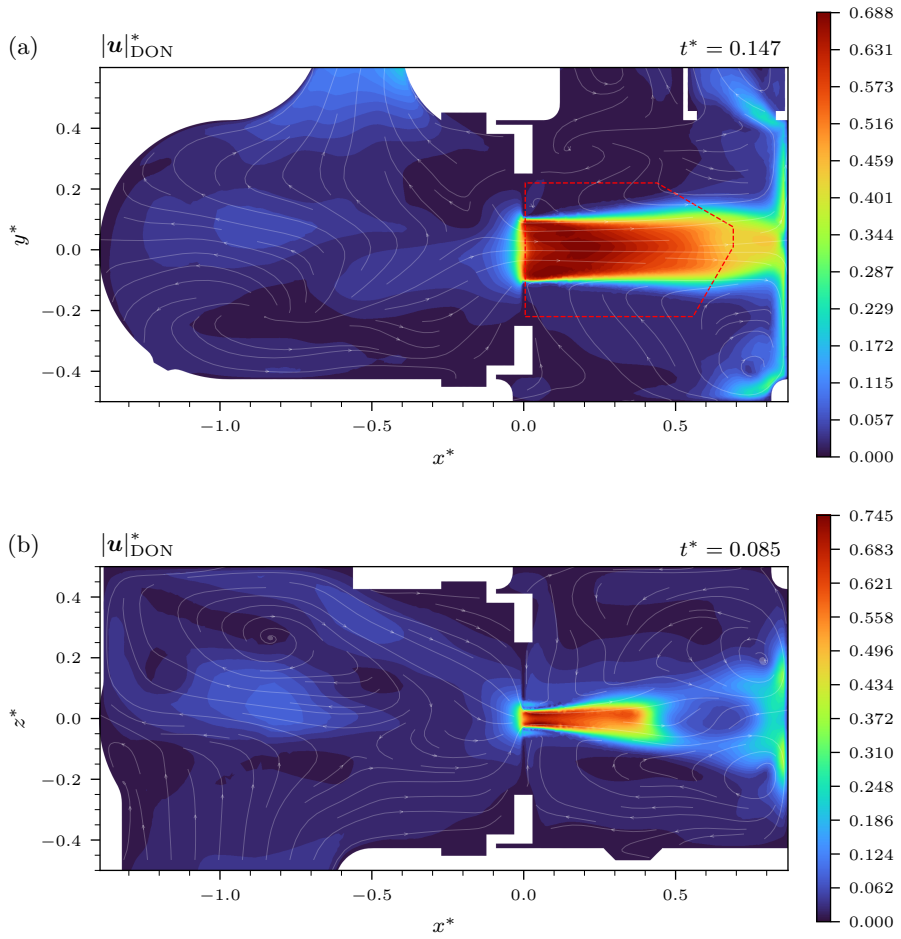

Figure 4.4.: "Hybrid DD" velocity magnitude predictions for 3D-Transient SlotL Exp, (a) z-plane for $t^* = 0.147$ and (b) y-plane for $t^* = 0.085$.

plane. Once established, the flow topology appears coherent, with the jet clearly extending to the LA wall. However, the temporal evolution again indicates a premature connection. A distinct feature in this case is the intensity parallel to the wall; the flow exhibits high velocity magnitudes relative to the main jet body, appearing disproportionately energetic compared to observations in the previous cases. Subfigure (b) presents the y-plane slice during the fully connected phase. While many figures in this chapter are selected to illustrate specific reconstruction challenges, this snapshot serves as a counterpoint, demonstrating the model's capacity to generate results that appear reasonable. In contrast, the onset of the diastolic phase shown in subfigure (c) highlights significant reconstruction failures. The field is marred by speckled artifacts and a prominent streak of high-velocity magnitude. A similar flaw is observed in the y-plane view for these time steps.

A potential explanation for the specific challenges observed in this case is provided in Appendix Figure A.4: The experimental setup for this run operated under a different boundary condition than the standard training set due to a hardware limitation; the cardiac pump reached its power limit, resulting in a "clipped" input waveform. This deviation alters the driving pressure dynamics significantly compared to the smooth profiles the model is trained on, forcing the operator to reconstruct a flow field governed by unfamiliar constraints.

The final application case, EccJet, presents the most significant deviation from the training distribution due to its highly eccentric jet angle. This is reflected in the physical consistency

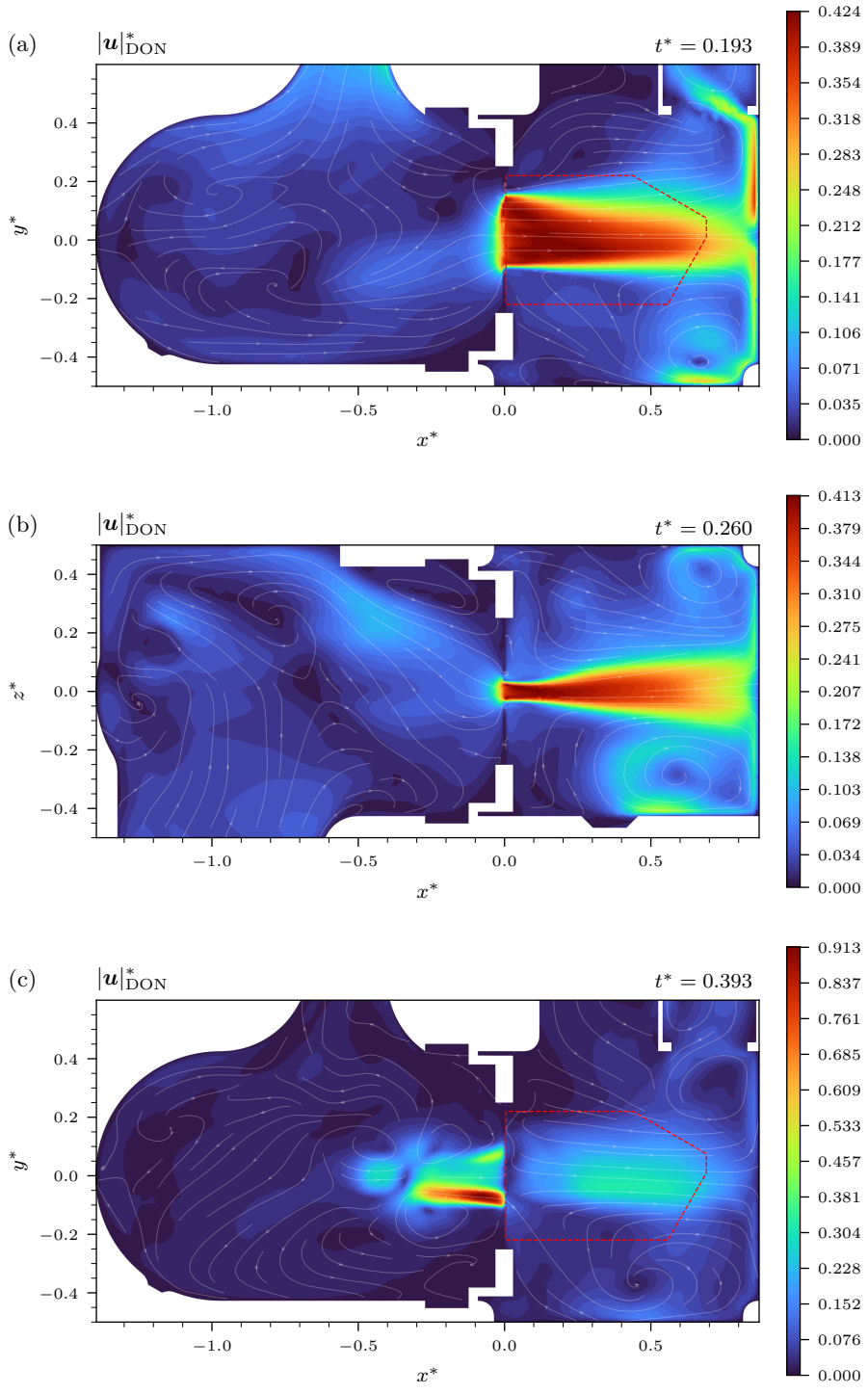

Figure 4.5.: "Hybrid DD" velocity magnitude predictions for 3D-Transient DropXL Exp, (a) z-plane for $t^* = 0.193$, (b) y-plane for $t^* = 0.260$ and (c) z-plane for $t^* = 0.393$.

metrics; the continuity residual, shown in Figure 4.1, is the highest among all experimental configurations. The challenge is evident when examining the baseline prediction. Appendix Figure C.51 illustrates that the standard "DD" model fails to resolve the novel flow trajectory, incorrectly predicting a jet that connects to the right atrial wall. The efficacy of the hybrid adaptation in correcting this behavior is visualized in Figure 4.6. The fine-tuned model successfully reorients the jet direction to match the eccentric angle observed in the PIV data. The reconstructed jet extends beyond the immediate bounds of the supervised area. However, the physical validity of this extension remains ambiguous; it bears a resemblance to the extrapolation observed in the 2D-Transient data-driven PINN, suggesting it may be a numerical byproduct of the MLP architecture rather than a true propagation of flow physics. Furthermore, temporal analysis reveals that the jet does not progress significantly beyond this snapshot and fails to reach the bottom wall of the atrium.

Despite the directional correction, remnants of the training distribution persist. Faint regions of elevated velocity magnitude appear along the right atrial wall, the location where jets typically impinged in the training cases. Additionally, the streamline quality in the LV is notably degraded. The flow patterns appear overly uniform and lack the nuanced features observed in previous reconstructions, indicating that the global flow prediction degrades due to the adjustment to the novel jet.

Given the limitations established throughout the validation chapter and the preceding experimental analyses, specifically the recurring artifacts and the insufficient propagation of the hybrid fine-tuning, the poor qualitative results observed in the EccJet snapshots are an expected outcome. However, despite this predictable performance deficit, the EccJet case offers a unique analytical value: The experimental dataset includes time-resolved PIV measurements for adjacent z-planes. This allows for at least a partial numerical judgment of the experimental application, quantifying the reconstruction accuracy along the z-direction, even though a full validation of the entire volumetric domain remains impossible.

The quantitative analysis is visualized in Figure 4.7, which compares the reference PIV data and the DeepONet prediction across different z-planes. The left half of the figure shows the plane at $z^* = -0.02$, corresponding to a dimensional depth of 2 mm. Subfigure (a) presents the PIV reference, revealing fine-grained velocity magnitude details. Topologically, this slice is characterized by a large upper vortex and a smaller secondary vortex located near

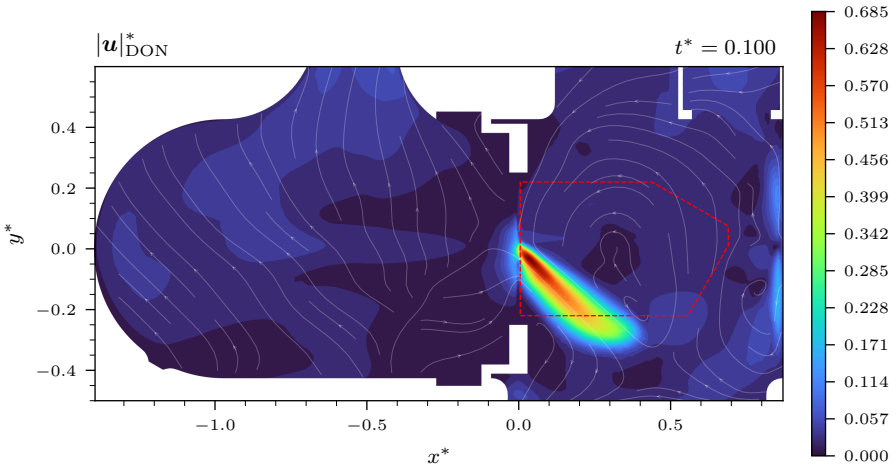

Figure 4.6.: "Hybrid DD" velocity magnitude prediction for 3D-Transient EccJet Exp, z-plane at $t^* = 0.1$.

NRMSE in %	$t^* = 0.040$	$t^* = 0.053$	$t^* = 0.067$	$t^* = 0.080$	$t^* = 0.093$
$z^* = -0.02$	28.2	26.5	24.6	23.9	21.0
$z^* = -0.04$	62.0	52.9	43.1	41.6	35.1
$z^* = -0.06$	96.3	72.1	52.6	49.8	43.9

Table 4.1.: NRMSE values for the "Hybrid DD" velocity magnitude prediction compared against PIV measurements for the EccJet case across available time-steps and z-planes.

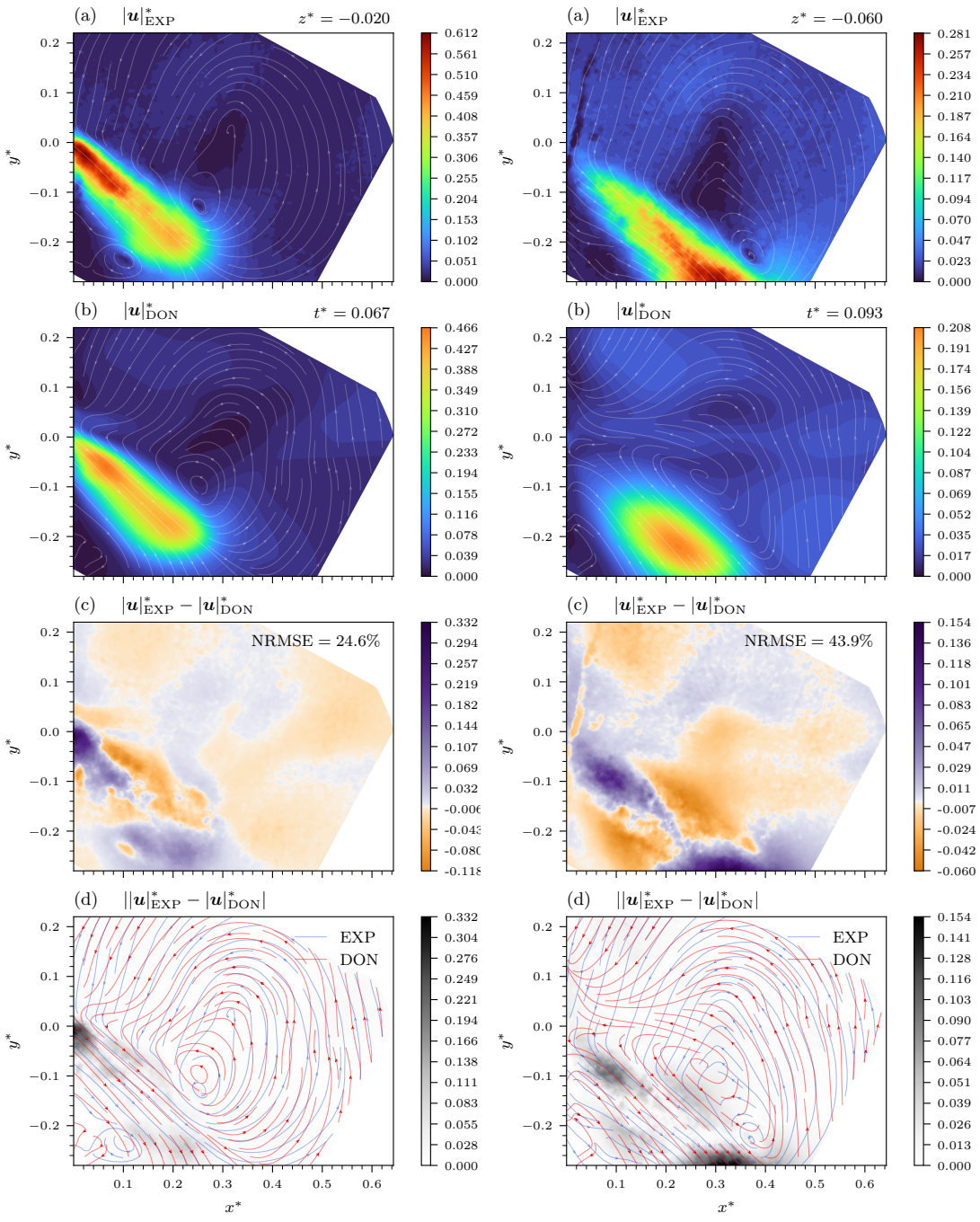

Figure 4.7.: (a) Reference 3D-Transient EccJet Exp velocity magnitude. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines. The left half shows $z^* = -0.02$ at $t^* = 0.067$, the right half shows $z^* = -0.06$ at $t^* = 0.093$.

the jet tip. The corresponding DeepONet prediction in subfigure (b) successfully captures the general macroscopic features, with the jet direction and length roughly matching the experimental observation. The streamlines in the prediction also reproduce the large upper vortex structure. However, a closer inspection reveals localized discrepancies: the streamlines momentarily invert direction just before reaching the upper edge of the jet, and the smaller lower vortex is not properly resolved. These inaccuracies result in an NRMSE of 24.6% for this plane and time-step.

The entire error quantification for the five available PIV time-steps is summarized in Table 4.1, which reveals a distinct spatial dependency: while prediction accuracy remains moderate near the supervised center slice, the error metrics degrade as the distance from the $z^* = 0$ plane increases. For example, at the time-step shown in the left subfigure ($t^* = 0.067$), the NRMSE doubles from 24.6% at $z^* = -0.02$ to 52.6% at $z^* = -0.06$. This quantitative evidence reinforces the qualitative observations regarding the limited lateral propagation of the hybrid corrections.

The right half of Figure 4.7 visualizes the flow at a deeper plane ($z^* = -0.06$) during a later time-step ($t^* = 0.093$), where the jet has propagated to the edge of the recorded domain. The absolute velocity magnitudes here are lower, approximately half those observed in the previous subfigure; the colormaps are scaled separately. The DeepONet prediction (b) exhibits a marked loss of definition, appearing even smoother and more featureless than the shallower slice. The structural agreement also degrades considerably; even the large vortex structure fails to match the experimental reference. While the model predicts some influence of the jet penetrating into this layer, the resulting feature remains undefined, confirming the rapid decline in reconstruction quality away from the supervised center plane. Appendix Figure C.52 displays the reconstruction for the intermediate plane at $z^* = -0.04$ and $t^* = 0.080$, illustrating a prediction quality that naturally falls between the extremes of the shallower and deeper slices discussed above.

4.2. Interim Conclusion

The transition from synthetic validation to experimental application demonstrates that the DeepONet framework maintains a baseline level of robustness when driven by real-world measurements. Despite the domain shift, characterized by measurement noise, temporal sparsity, and unseen boundary conditions, the model generated volumetric flow fields with aggregate physical consistency comparable to the CFD benchmarks. The “Hybrid DD” approach successfully adapted the baseline model to match specific experimental observations.

However, this chapter confirms the limitation in the framework’s ability to ensure volumetric coherence from planar supervision. The detailed snapshot analysis reveals that while the hybrid fine-tuning effectively constrains the flow within the supervised observation plane, this correction fails to propagate laterally through the depth of the domain. This results in disjointed 3D topologies where the outer flow layers and vortex rings remain misaligned with the corrected central jet. Furthermore, the model exhibits sensitivity to experimental signal characteristics; the absence of augmentation in the training phase likely contributes to the severe artifacts observed during the diastolic phases. This performance degradation is further compounded by the systemic discrepancies identified in the URANS Validation Section 2.1.1.2. The CFD training dataset, characterized by overpredicted jet magnitudes and idealized orifice geometries, conflicts with the physical reality of the flexible orifice foil in the experiment, creating a fundamental domain gap that may drive these elevated errors.

The quantitative assessment of the EccJet z-planes empirically confirms these qualitative findings. The reconstruction accuracy degrades sharply as the distance from the supervised slice increases, proving that the current single-plane fine-tuning strategy is insufficient to constrain the full volumetric solution reliably. Consequently, while the DeepONet functions as a flexible surrogate capable of assimilating sparse real-world data, achieving accurate 4D hemodynamics requires further work.

5. Conclusion and Outlook

Accurate characterization of hemodynamic flows is fundamental to the diagnosis and treatment planning of cardiovascular pathologies. While clinical imaging modalities like 4D Flow MRI offer volumetric insight, they remain constrained by limited spatiotemporal resolution. Computational Fluid Dynamics (CFD) provides high-fidelity resolution but entails high computational costs and relies on boundary condition assumptions. This thesis addressed the gap between these domains by developing a machine learning framework capable of reconstructing time-resolved volumetric (4D) flow fields. The investigation was grounded in a physiological hemodynamic simulator designed to replicate left heart dynamics and mitral valve regurgitation. This experimental setup defined the scope of experimentally observable inputs used to drive the reconstruction: time-resolved pressure boundary conditions acquired from sensors and planar velocity slices obtained via two-component Particle Image Velocimetry (2D2C PIV).

The core objective was to establish a robust methodology for solving the ill-posed problem of inferring full 4D hemodynamics from these limited observations. To achieve this, the research evaluated two contrasting architectural strategies. The first approach utilized an instance-specific Multilayer Perceptron (MLP) trained via the Physics-Informed Neural Network (PINN) method, optimizing a single network for a specific flow configuration. The second approach employed (Physics-Informed) Deep Operator Networks (PI-DeepONets), designed to learn the mapping between a discretized input function and the resulting volumetric flow field through multiple training cases, thereby acting as a generalizable surrogate. The input function was chosen as a spatially restricted subset of the full velocity field, serving as an analogue to the planar slice captured by the experimental PIV setup.

To enable this investigation, synthetic ground truth data was created using Unsteady Reynolds-averaged Navier-Stokes (URANS) simulations, which replicated the operating conditions, geometry of the hemodynamic simulator, and the specific mitral regurgitation orifice phantoms (MROPs) employed. This dataset has been validated against experimental measurements. Building upon this setup, two simplified dataset tiers were derived to facilitate a hierarchical development strategy. The 2D-Steady setup established a foundational environment, isolating the spatial reconstruction challenge from temporal complexities by modeling steady-state flows through varying orifice sizes. The 2D-Transient setup reintroduced temporal dynamics, employing pulsatile pressure boundary conditions to mimic the cardiac cycle while maintaining a reduced computational cost compared to the full volumetric domain. Using these datasets, the reconstruction frameworks were evaluated using three training modalities: purely data-driven models, physics-augmented strategies incorporating non-dimensionalized RANS loss-functions, and a hybrid "test-time adaptation" approach. The latter strategy leverages a pre-trained DeepONet as a generalized base model, which is then fine-tuned to maximize consistency with the specific sparse target observations. The final framework was applied to real experimental PIV data acquired from the hemodynamic simulator, assessing the model's capability to generalize to real-world measurements.

The validation of these models was conducted using extensive snapshot analysis, flow field comparisons, and aggregate error metrics. Through this hierarchical evaluation, the operator learning framework consistently demonstrated its superiority over the instance-specific

PINN approach. Across all complexity levels, the PINN failed to satisfactorily solve the ill-posed reconstruction problem given the sparsity of the boundary data. In the 2D-Steady case, the PINN failed to converge to the correct flow topology or pressure gradients. Moving to the 2D-Transient environment, the PINN was unable to capture the correct temporal evolution of flow features, resulting in a retracting jet artifact during diastole rather than the physical extension into the ventricle. In the 3D-Transient setup, the model overfitted to the data window and failed to propagate the solution physically into the depth of the domain or respect the boundary conditions. Similar to the PINN, alternative DeepONet architectures, including Fourier-feature networks and "exact" continuity formulations utilizing stream functions or vector potentials, yielded results inferior to the standard MLP despite their theoretical merit. Consequently, the standard MLP-based DeepONet was selected as the primary architecture for further optimization.

The investigation of steady-state 2D channel flow established the "Hybrid" training strategy as a critical innovation for handling extrapolation cases. The performance of the DeepONet proved reliant on the training dataset distribution, where increasing the number of training cases significantly improved generalization capabilities. The test-time adaptation strategy demonstrated high effectiveness; by continuing the background training of the operator while simultaneously minimizing the error on the sparse validation observations, the Normalized Root-Mean-Square Error (NRMSE) for out-of-distribution cases was reduced by a factor of 5.53, dropping from 84.0% to 15.2%. While the addition of physics-informed loss-terms such as the continuity equation improved generalization, these constraints required careful balancing even while utilizing automatic Neural Tangent Kernel (NTK) weighting. This balancing act became progressively more difficult in the subsequent transient and 3D domains. The hybrid fine-tuning also largely corrected "streaking" artifacts caused by the appearance of common shapes in the training data and extrapolating to larger orifice sizes, although some remnants persisted.

Introducing temporal dynamics with 2D-Transient revealed that while the DeepONet remained superior to the PINN, the effectiveness of the hybrid adaptation became spatially localized. The DeepONet provided generally stable predictions across the cardiac cycle, achieving approximately a two-fold improvement in error metrics compared to instance-specific PINNs. Unlike in the steady-state case, the improvements from hybrid fine-tuning were largely contained within the supervision window, where they are least needed due to the availability of measurement data, and failed to propagate global improvements significantly. Consequently, the streaking artifacts observed in 2D-Steady were no longer fully corrected in 2D-Transient due to the presence of the diastolic jet flowing in the opposite direction and the increased temporal complexity.

The 4D reconstruction in 3D-Transient highlighted the geometric limitations of inferring volumes from planar data while simultaneously demonstrating a significant computational efficiency potential: a single URANS simulation for 3D-Transient required approximately 120 hours of computation time, whereas the baseline DeepONet trained in roughly 12 hours, and the test-time adaptation for a new case was completed in as little as 15 minutes. While not a direct comparison due to the prerequisite of the CFD training set, this highlights the rapid inference capability of the surrogate. The final optimal strategy, "Hybrid DD", relied purely on data-driven losses; physics constraints were abandoned as they induced over-smoothing and increased computational cost without yielding accuracy gains in the tested configurations. Similar to 2D-Transient, the hybrid correction was effective within the supervised $z = 0$ slice but failed to propagate outside of it. This resulted in disjointed 3D jet structures, where the corrected central slice misaligned with the uncorrected outer flow layers.

Applying the framework to real experimental PIV data added the challenge of domain

shift between the idealized synthetic training data and the noisy real-world measurements. The model proved sensitive to this discrepancy in signal characteristics, which destabilized the operator's inference, leading to more severe artifacts, especially during the diastolic phases. While the diastolic jet is a byproduct of using fixed MROPs rather than functional mitral valves and lies outside the primary research focus on systolic regurgitation jets, the model's struggle to render this challenging feature accurately serves as a valuable marker of robustness. The model was evaluated qualitatively and through PDE residuals across four distinct cases. The training distribution did not cover the specific geometry of the EccJet case, leading the baseline operator to predict a standard jet angle rather than the highly eccentric angle observed in the experiment. The hybrid adaptation successfully reoriented the jet direction in this case to match the observations. However, quantitative assessment confirmed that reconstruction accuracy degrades sharply as the distance from the supervised slice increases.

The investigation across all complexity levels concludes that while hybrid fine-tuning offers a mechanism to correct the inevitable artifacts arising from the DeepONet's dependence on training data for out-of-distribution predictions, this adaptation presently remains insufficient in transient applications. Improvements fail to propagate globally and remain largely confined to the additionally supervised region. Addressing these limitations requires either enhancing the baseline prediction accuracy to reduce the initial error or amplifying the global impact of the fine-tuning step, possibly achievable through the architectural and methodological advancements proposed below.

For enhancing spatial predictive fidelity, the current reliance on MLPs necessitates flattening the PIV input data, which inevitably discards spatial relationships. Replacing the branch network with Convolutional Neural Networks (CNNs) could preserve this structural information and potentially improve the encoding of the velocity fields. Furthermore, the framework might benefit from multi-branch architectures capable of fusing heterogeneous data sources, such as pressure readings and geometric shape parameters, directly into the operator. Utilizing geometric parameterization of the orifice phantoms instead of velocity snapshots implies a trade-off, as it would eliminate the dependency on instantaneous flow recordings while potentially reducing the immediate state information available to the network. To address the challenge of volumetric coherence, incorporating multi-plane supervision using orthogonal recordings appears necessary to anchor the solution in the third dimension and facilitate lateral propagation during the fine-tuning process. Moreover, reducing the output vector size by excluding variables such as pressure or turbulent viscosity when not strictly necessary could reduce the complexity of the mapping task and improve performance, provided that a purely data-driven framework remains the optimal choice.

The temporal consistency of the reconstruction requires attention as well. The current snapshot-based methodology processes time instants in isolation and neglects the flow history. Integrating temporal learning structures, such as Temporal Convolutional Networks (TCNs), would allow the model to process sequences and maybe assist in resolving the transient artifacts observed in this study. Similarly, architectures that explicitly account for the periodic nature of the cardiac cycle could further stabilize the predictions. Otherwise incorporating information from adjacent time steps or employing temporal padding strategies might also help the network resolve the rapid transients and boundary condition shifts that occur at the end of the cycle. However, utilizing multiple snapshots simultaneously introduces a new constraint: the architecture must be capable of handling irregular time intervals, or the input data must either be regularly spaced, a requirement not currently met by the available experimental PIV datasets.

Enhancements to the physics-informed components and data strategies also warrant investigation. Optimizing the formulation of the Reynolds-averaged Navier-Stokes equations

used for the physics losses could help overcome the practical limitations observed in the current setup. Specifically, the incorporation of established two-equation turbulence models might offer superior closure compared to the current approach. Refining the weighting of these physical loss terms remains a critical step for balancing generalization with data fidelity. Parallel to model improvements, the robustness of the operator is fundamentally tied to the training distribution. Strategically expanding the CFD dataset to cover broader parameter regimes would likely reduce extrapolation errors by converting difficult test cases into easier interpolation tasks. To bridge the domain gap between idealized simulations and real-world measurements, implementing data augmentation strategies such as the addition of artificial noise during training could improve the operator's resilience to experimental signal characteristics. Improving the accuracy of the underlying URANS simulations represents another pathway to reduce the domain shift impact, though transitioning to scale-resolving methods such as Large Eddy Simulations (LES) or Direct Numerical Simulations (DNS) would likely incur prohibitive computational costs. Furthermore, higher fidelity data would not address the fundamental limitation where fine-tuning corrections fail to propagate volumetrically in transient flows, as this behavior persisted even during validation against synthetic data.

Finally, it must be noted that while extensive testing was conducted, the exploration of hyperparameter tuning and computational efficiency optimizations was not exhaustive, leaving room for potential performance gains through systematic refinement. In terms of experimental design, future setups could accommodate additional pressure sensor locations. This would not only refine the pressure field prediction but, through the coupling of PDE losses, potentially improve the overall velocity field solutions. Regarding the validation process, the current sparse experimental dataset is sufficient to characterize the existing performance limitations. However, as the reconstruction fidelity improves, acquiring more comprehensive volumetric measurements will become important to provide a clearer assessment of the model's capabilities. Ultimately, these methodological advancements aim to pave the way for applying this framework with reduced-cost inference to more complex mitral valve geometries and enabling the detailed analysis of advanced flow features such as wall shear stress and vorticity.

Bibliography

[1] C. Anitescu, B. İsmail Ateş, and T. Rabczuk. „Physics-Informed Neural Networks: Theory and Applications“. en. In: *Machine Learning in Modeling and Simulation: Methods and Applications*. Ed. by T. Rabczuk and K.-J. Bathe. Cham: Springer International Publishing, 2023, pp. 179–218. ISBN: 978-3-031-36644-4. DOI: 10.1007/978-3-031-36644-4_5.

[2] H. Bai et al. *Data-driven modeling of unsteady flow based on deep operator network*. arXiv:2404.06791 [physics] version: 1. Apr. 2024. DOI: 10.48550/arXiv.2404.06791.

[3] C. M. Bishop and H. Bishop. *Deep Learning: Foundations and Concepts*. en. Cham: Springer International Publishing, 2024. ISBN: 978-3-031-45467-7 978-3-031-45468-4. DOI: 10.1007/978-3-031-45468-4.

[4] S. Cai et al. „Physics-informed neural networks (PINNs) for fluid mechanics: a review“. en. In: *Acta Mechanica Sinica* 37.12 (Dec. 2021), pp. 1727–1738. ISSN: 1614-3116. DOI: 10.1007/s10409-021-01148-1.

[5] P. Callmer et al. *Deep learning for temporal super-resolution 4D Flow MRI*. Jan. 2025. DOI: 10.48550/arXiv.2501.08780.

[6] O. L. Cruz-González, V. Deplano, and B. Ghattas. *Enhanced Vascular Flow Simulations in Aortic Aneurysm via Physics-Informed Neural Networks and Deep Operator Networks*. arXiv:2503.17402 [cs]. Mar. 2025. DOI: 10.48550/arXiv.2503.17402.

[7] H. Csala et al. „A comparison of machine learning methods for recovering noisy and missing 4D flow MRI data“. en. In: *International Journal for Numerical Methods in Biomedical Engineering* 40.11 (2024), e3858. ISSN: 2040-7947. DOI: 10.1002/cnm.3858.

[8] L. Davidson. *Fluid mechanics, turbulent flow and turbulence modeling*. Sept. 2025. URL: https://www.tfd.chalmers.se/~lada/postscript_files/solids-and-fluids_turbulent-flow_turbulence-modelling.pdf.

[9] M. Dreisbach et al. *PINNs4Drops: Convolutional feature-enhanced physics-informed neural networks for reconstructing two-phase flows*. arXiv:2411.15949 [physics]. Nov. 2024. DOI: 10.48550/arXiv.2411.15949.

[10] H. Eivazi et al. „Physics-informed neural networks for solving Reynolds-averaged Navier–Stokes equations“. In: *Physics of Fluids* 34.7 (July 2022), p. 075117. ISSN: 1070-6631. DOI: 10.1063/5.0095270.

[11] S. Elfwing, E. Uchibe, and K. Doya. „Sigmoid-weighted linear units for neural network function approximation in reinforcement learning“. In: *Neural Networks*. Special issue on deep reinforcement learning 107 (Nov. 2018), pp. 3–11. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2017.12.012.

[12] M. Enriquez-Sarano, C. W. Akins, and A. Vahanian. „Mitral regurgitation“. In: *The Lancet* 373.9672 (Apr. 2009). Publisher: Elsevier, pp. 1382–1394. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(09)60692-9.

[13] J. H. Ferziger, M. Perić, and R. L. Street. *Computational Methods for Fluid Dynamics*. en. Cham: Springer International Publishing, 2020. ISBN: 978-3-319-99691-2 978-3-319-99693-6. DOI: 10.1007/978-3-319-99693-6.

[14] P. Garg et al. „Cardiovascular magnetic resonance imaging in mitral valve disease“. In: *European Heart Journal* 46.7 (Feb. 2025), pp. 606–619. ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehae801.

[15] M. Gholampour et al. „Parameterized physics-informed neural networks for a transient thermal problem: A pure physics-driven approach“. In: *International Communications in Heat and Mass Transfer* 159 (Dec. 2024), p. 108330. ISSN: 0735-1933. DOI: 10.1016/j.icheatmasstransfer.2024.108330.

[16] S. Ghosh et al. *RANS-PINN based Simulation Surrogates for Predicting Turbulent Flows*. arXiv:2306.06034 [cs]. Aug. 2023. DOI: 10.48550/arXiv.2306.06034.

[17] S. Goswami et al. „Deep transfer operator learning for partial differential equations under conditional shift“. en. In: *Nature Machine Intelligence* 4.12 (Dec. 2022). Publisher: Nature Publishing Group, pp. 1155–1164. ISSN: 2522-5839. DOI: 10.1038/s42256-022-00569-2.

[18] S. Goswami et al. „Physics-Informed Deep Neural Operator Networks“. en. In: *Machine Learning in Modeling and Simulation: Methods and Applications*. Ed. by T. Rabczuk and K.-J. Bathe. Cham: Springer International Publishing, 2023, pp. 219–254. ISBN: 978-3-031-36644-4. DOI: 10.1007/978-3-031-36644-4_6.

[19] J. H. Harmening, F. Pioch, and D. Schramm. „Physics Informed Neural Networks as Multidimensional Surrogate Models of CFD Simulations“. In: *Proceedings of the Machine Learning und Artificial Intelligence in Strömungsmechanik und Strukturanalyse, Wiesbaden, Germany* (2022), pp. 16–17. URL: https://www.researchgate.net/publication/368364374_Physics_Informed_Neural_Networks_as_Multidimensional_Surrogate_Models_of_CFD_Simulations (visited on 01/27/2025).

[20] J. Hong et al. „Estimating Aortic Pressure Waveform in a 1D Hemodynamic Model of the Human Arterial System using DeepONet“. In: *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. ISSN: 2694-0604. July 2024, pp. 1–5. DOI: 10.1109/EMBC53108.2024.10781510.

[21] V. L. Inc. *Pulse Duplicator System User Manual*. en. 2015.

[22] V. L. Inc. *SuperPump System User Manual*. en. 2014.

[23] V. L. Inc. *SuperPump System User Manual*. en. 2018.

[24] B. Iung and A. Vahanian. „Epidemiology of valvular heart disease in the adult“. en. In: *Nature Reviews Cardiology* 8.3 (Mar. 2011). Publisher: Nature Publishing Group, pp. 162–172. ISSN: 1759-5010. DOI: 10.1038/nrcardio.2010.202.

[25] A. D. Jagtap, K. Kawaguchi, and G. E. Karniadakis. „Adaptive activation functions accelerate convergence in deep and physics-informed neural networks“. In: *Journal of Computational Physics* 404 (Mar. 2020), p. 109136. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2019.109136.

[26] G. James et al. *An Introduction to Statistical Learning: with Applications in Python*. en. Springer Texts in Statistics. Cham: Springer International Publishing, 2023. ISBN: 978-3-031-38746-3 978-3-031-38747-0. DOI: 10.1007/978-3-031-38747-0.

[27] R. Karl. „Patient-specific hemodynamic simulators for cardiovascular therapies“. Dissertation. Ruprecht Karl University of Heidelberg, 2024. DOI: 10.11588/heidok.00035976.

[28] R. Karl et al. „An ex-vivo and in-vitro dynamic simulator for surgical and transcatheter mitral valve interventions“. en. In: *International Journal of Computer Assisted Radiology and Surgery* 19.3 (Mar. 2024), pp. 411–421. ISSN: 1861-6429. DOI: 10.1007/s11548-023-03036-4.

[29] R. Leister et al. „Investigating the Shortcomings of the Flow Convergence Method for Quantification of Mitral Regurgitation in a Pulsatile In-Vitro Environment and with Computational Fluid Dynamics“. en. In: *Cardiovascular Engineering and Technology* (Jan. 2025). ISSN: 1869-4098. DOI: 10.1007/s13239-024-00763-w.

[30] H. Li et al. „An architectural analysis of DeepOnet and a general extension of the physics-informed DeepOnet model on solving nonlinear parametric partial differential equations“. In: *Neurocomputing* 611 (Jan. 2025), p. 128675. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2024.128675.

[31] L. Li, X.-C. Tai, and R. H.-F. Chan. „A new method to compute the blood flow equations using the physics-informed neural operator“. In: *Journal of Computational Physics* 519 (Dec. 2024), p. 113380. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2024.113380.

[32] Z. Li et al. „Physics-Informed Neural Operator for Learning Partial Differential Equations“. In: *ACM / IMS J. Data Sci.* 1.3 (May 2024), 9:1–9:27. DOI: 10.1145/3648506.

[33] J. Liang, R. He, and T. Tan. „A Comprehensive Survey on Test-Time Adaptation Under Distribution Shifts“. en. In: *International Journal of Computer Vision* 133.1 (Jan. 2025), pp. 31–64. ISSN: 1573-1405. DOI: 10.1007/s11263-024-02181-w.

[34] H. J. Ling et al. „Physics-Guided Neural Networks for Intraventricular Vector Flow Mapping“. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 71.11 (Nov. 2024), pp. 1377–1388. ISSN: 1525-8955. DOI: 10.1109/TUFFC.2024.3411718.

[35] A. Long et al. „Deep Learning for Echo Analysis, Tracking, and Evaluation of Mitral Regurgitation (DELINEATE-MR)“. In: *Circulation* 150.12 (Sept. 2024). Publisher: American Heart Association, pp. 911–922. DOI: 10.1161/CIRCULATIONAHA.124.068996.

[36] I. Loshchilov and F. Hutter. *Decoupled Weight Decay Regularization*. arXiv:1711.05101 [cs]. Jan. 2019. DOI: 10.48550/arXiv.1711.05101.

[37] L. Lu et al. „Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators“. en. In: *Nature Machine Intelligence* 3.3 (Mar. 2021). Publisher: Nature Publishing Group, pp. 218–229. ISSN: 2522-5839. DOI: 10.1038/s42256-021-00302-5.

[38] B. Maidu et al. „Super-resolution Left Ventricular Flow and Pressure Mapping by Navier-Stokes-Informed Neural Networks“. In: *bioRxiv* (Apr. 2024), p. 2024.04.12.589319. ISSN: 2692-8205. DOI: 10.1101/2024.04.12.589319.

[39] N. Malekjani, A. Kharaghani, and E. Tsotsas. „A comparative study of dimensional and non-dimensional inputs in physics-informed and data-driven neural networks for single-droplet evaporation“. In: *Chemical Engineering Science* 306 (Mar. 2025), p. 121214. ISSN: 0009-2509. DOI: 10.1016/j.ces.2025.121214.

[40] A. Maznyczka et al. „Artificial Intelligence in Valvular Heart Disease: Innovations and Future Directions“. In: *JACC: Cardiovascular Interventions* 18.20 (Oct. 2025), pp. 2439–2457. ISSN: 1936-8798. DOI: 10.1016/j.jcin.2025.08.031.

[41] Y. Mei et al. „Fully Convolutional Network-Enhanced DeepONet-Based Surrogate of Predicting the Travel-Time Fields“. In: *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), pp. 1–12. ISSN: 1558-0644. DOI: 10.1109/TGRS.2024.3401196.

[42] F. R. Menter. „Two-equation eddy-viscosity turbulence models for engineering applications“. en. In: *AIAA Journal* 32.8 (Aug. 1994), pp. 1598–1605. ISSN: 0001-1452, 1533-385X. DOI: 10.2514/3.12149.

[43] L. Neff. „Characterization of Flow Dynamics in a Heart Simulator by means of PIV“. Master's Thesis. Karlsruhe Institute of Technology (KIT), 2023. DOI: 10.5445/IR/1000164471.

[44] *NVIDIA PhysicsNeMo Symbolic: Algorithms and utilities to explicitly physics inform the training of PhysicsNeMo models.* Feb. 2023. URL: https://github.com/NVIDIA/physicsnemo-sym.

[45] *NVIDIA PhysicsNeMo: An open-source framework for physics-based deep learning in science and engineering.* Feb. 2023. URL: https://github.com/NVIDIA/physicsnemo.

[46] E. Pajaziti et al. „Shape-driven deep neural networks for fast acquisition of aortic 3D pressure and velocity flow fields“. en. In: *PLOS Computational Biology* 19.4 (Apr. 2023). Publisher: Public Library of Science, e1011055. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011055.

[47] F. Pioch et al. „Turbulence Modeling for Physics-Informed Neural Networks: Comparison of Different RANS Models for the Backward-Facing Step Flow“. en. In: *Fluids* 8.2 (Feb. 2023). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 43. ISSN: 2311-5521. DOI: 10.3390/fluids8020043.

[48] S. B. Pope. *Turbulent Flows.* Cambridge University Press, 2013. ISBN: 978-0-521-59125-6. URL: https://www.cambridge.org/universitypress/subjects/physics/nonlinear-science-and-fluid-dynamics/turbulent-flows.

[49] T. Rabczuk and K.-J. Bathe, eds. *Machine Learning in Modeling and Simulation: Methods and Applications.* en. Computational Methods in Engineering & the Sciences. Cham: Springer International Publishing, 2023. ISBN: 978-3-031-36643-7 978-3-031-36644-4. DOI: 10.1007/978-3-031-36644-4.

[50] M. Raissi, P. Perdikaris, and G. E. Karniadakis. „Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations“. In: *Journal of Computational Physics* 378 (Feb. 2019), pp. 686–707. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2018.10.045.

[51] P. Schober, C. Boer, and L. A. Schwarte. „Correlation Coefficients: Appropriate Use and Interpretation“. eng. In: *Anesthesia and Analgesia* 126.5 (May 2018), pp. 1763–1768. ISSN: 1526-7598. DOI: 10.1213/ANE.0000000000002864.

[52] S. D. I. Software. *Simcenter STAR-CCM+ Documentation Version 2406.* 2024.

[53] M. Tancik et al. *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains.* Version Number: 1. 2020. DOI: 10.48550/ARXIV.2006.10739.

[54] M. A. Ur Rehman et al. „Fluid–structure interaction analysis of pulsatile flow in arterial aneurysms with physics-informed neural networks and computational fluid dynamics“. In: *Physics of Fluids* 37.3 (Mar. 2025), p. 031913. ISSN: 1070-6631. DOI: 10.1063/5.0259296.

[55] A. Velikorodny et al. „Deep operator learning for blood flow modelling in stenosed vessels“. en. In: *npj Artificial Intelligence* 1.1 (Nov. 2025). Publisher: Nature Publishing Group, p. 35. ISSN: 3005-1460. DOI: 10.1038/s44387-025-00035-5.

[56] H. Wang et al. *Recent Advances on Machine Learning for Computational Fluid Dynamics: A Survey.* arXiv:2408.12171 [cs]. Aug. 2024. DOI: 10.48550/arXiv.2408.12171.

[57] S. Wang, H. Wang, and P. Perdikaris. „Learning the solution operator of parametric partial differential equations with physics-informed DeepONets“. In: *Science Advances* 7.40 (Sept. 2021). Publisher: American Association for the Advancement of Science, eabi8605. DOI: 10.1126/sciadv.abi8605.

[58] S. Wang, X. Yu, and P. Perdikaris. „When and why PINNs fail to train: A neural tangent kernel perspective“. In: *Journal of Computational Physics* 449 (Jan. 2022), p. 110768. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2021.110768.

[59] C. Westbrook. *Calculator to estimate the density and viscosity of glycerine / water mixtures*. Apr. 2018. URL: https://www.met.reading.ac.uk/~sws04cdw/viscosity_calc.html (visited on 11/10/2025).

[60] H. S. Wong et al. „3D velocity and pressure field reconstruction in the cardiac left ventricle via physics informed neural network from echocardiography guided by 3D color Doppler“. In: *Computer Methods and Programs in Biomedicine* 263 (May 2025), p. 108671. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2025.108671.

[61] S. Wu. *Fine-Tuning DeepONets to Enhance Physics-informed Neural Networks for solving Partial Differential Equations*. arXiv:2410.14134 [math]. Oct. 2024. DOI: 10.48550/arXiv.2410.14134.

[62] H. Xu, W. Zhang, and Y. Wang. „Explore missing flow dynamics by physics-informed deep learning: The parameterized governing systems“. In: *Physics of Fluids* 33.9 (Sept. 2021), p. 095116. ISSN: 1070-6631. DOI: 10.1063/5.0062377.

[63] D. L. Young, C. H. Tsai, and C. S. Wu. „A novel vector potential formulation of 3D Navier–Stokes equations with through-flow boundaries by a local meshless method“. In: *Journal of Computational Physics* 300 (Nov. 2015), pp. 219–240. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2015.07.040.

[64] X. Zeng et al. „Echocardiography of the Mitral Valve“. In: *Progress in Cardiovascular Diseases*. Echocardiography in Clinical Practice and Research 57.1 (July 2014), pp. 55–73. ISSN: 0033-0620. DOI: 10.1016/j.pcad.2014.05.010.

[65] Z. Zhang et al. *DeepONet as a Multi-Operator Extrapolation Model: Distributed Pretraining with Physics-Informed Fine-Tuning*. arXiv:2411.07239 [cs]. Nov. 2024. DOI: 10.48550/arXiv.2411.07239.

[66] M. Zhu et al. „Reliable extrapolation of deep neural operators informed by physics or sparse observations“. In: *Computer Methods in Applied Mechanics and Engineering* 412 (July 2023), p. 116064. ISSN: 0045-7825. DOI: 10.1016/j.cma.2023.116064.

Symbols and Abbreviations

Symbol	Meaning of Latin symbol
b_i	Branch network output components
e_i	Element-wise product of branch and trunk outputs
Eu	Euler number
h	(Half for 2D-Steady/Transient) Orifice height
k	Turbulent kinetic energy
l	Length
$\mathcal{L}$	Loss term
$\dot{m}$	Mass flow
m, n	Grid dimensions for branch network input discretization
N	Number of discrete points/samples
p	Pressure
Re	Reynolds number
St	Strouhal number
t	Time
t_i	Trunk network output components
$u/v/w$	Velocity components
$\boldsymbol{u},\ u_i$	Velocity vector
w_i	Loss term weights
$\boldsymbol{x},\ x_i$	Coordinate vector
x	Regurgitation jet direction
y	Vertical direction
y^+	Wall distance in non-dimensional viscous length scaling
z	Spanwise depth direction

Symbol	Meaning of Greek symbol
δ_f	Foil or orifice wall thickness
ε	L1 integral norms of PDE residuals
θ	Artificial neural network parameters
μ	Dynamic viscosity
μ_{eff}	Effective viscosity used for turbulence modeling
ν	Kinematic viscosity
ν_t	Turbulent viscosity
ρ	Density
ψ	Stream function
ψ_i	Vector potential
ω	Specific dissipation rate

Symbol	Meaning of operator or index
$\partial\square$	Partial derivative
$\Delta\square$	Interval or difference
$\square_{i/j}$	Cartesian tensor components, Einstein notation implies summation
$\overline{\square}$	(Time-)averaged value
$\lvert\square\rvert$	Absolute value or vector magnitude
$\square^*$	Non-dimensional quantity
$\square_c$	Characteristic value used for non-dimensionalization

Symbol	Meaning of abbreviation
2/3D-Steady/Transient	Terms describing the three setups/datasets of CFD simulations
4D	Four-dimensional (3D space + time)
ANN	Artificial neural network
BC	Boundary condition
BPM	Beats per minute heart rate
CAD	Computer-aided design
CFD	Computational fluid dynamics
(PI-)DeepONet	(Physics-informed) Deep operator network, also DON
CFL	Courant-Friedrichs-Lewy condition
CNN	Convolutional neural network
Conti	Continuity equation
DD	Data-driven
DNS	Direct numerical simulation
Exp	Experiment
Extra/Inter	Extrapolation/Interpolation case
FOV	Field of view
IIS	Institute of Industrial Science
ISTM	Institute of Fluid Mechanics
KIT	Karlsruhe Institute of Technology
LA	Left atrium
LV	Left ventricle
ML	Machine learning
MLP	Multilayer perceptron
Mom,i	Momentum equation for i-direction
MR	Mitral regurgitation
MRI	Magnetic resonance imaging
MROP	Mitral regurgitation orifice phantom
MV	Mitral valve
NRMSE	Normalized root-mean-square error
NSE	Navier-Stokes equations
NTK	Neural Tangent Kernel
PDE	Partial differential equation
PhysicsNeMo Sym	NVIDIA PhysicsNeMo Symbolic, formerly Modulus Symbolic
PINN	Physics-informed neural network
PIV	Particle image velocimetry
(U)RANS	(Unsteady) Reynolds-averaged Navier-Stokes
SSE	Sum of squared errors
Sym	Symmetry
Train	Training dataset
UTokyo	The University of Tokyo
Val	Validation dataset or single application case

List of Figures

List of Tables

Appendix

A. Computational Fluid Dynamics

A.1. 3D-Transient Mesh Independence

For the PinholeL MROP, four meshes of different base cell sizes and slightly different relative refinement sizes were investigated. Table A.1 lists the resulting cell counts. Figure A.1 shows that the mass flow through the orifice is almost identical. After 0.06 s, the two finer and coarser meshes separate more clearly regarding the maximum velocity magnitude. This value appears in the tip of the regurgitant jet, which is relatively small compared to the entire structure. Due to computational cost, the Mesh2 settings are chosen. The development of the ML framework itself is not contingent upon the source data being maximally accurate; improving the training data at a later stage is possible. For the PinholeL case specifically, Mesh3 is chosen, since the data is already available.

mesh	Mesh1	Mesh2	Mesh3	Mesh4
cell count	146,000	426,000	1,042,000	1,993,000

Table A.1.: PinholeL mesh cell counts rounded to nearest 1,000

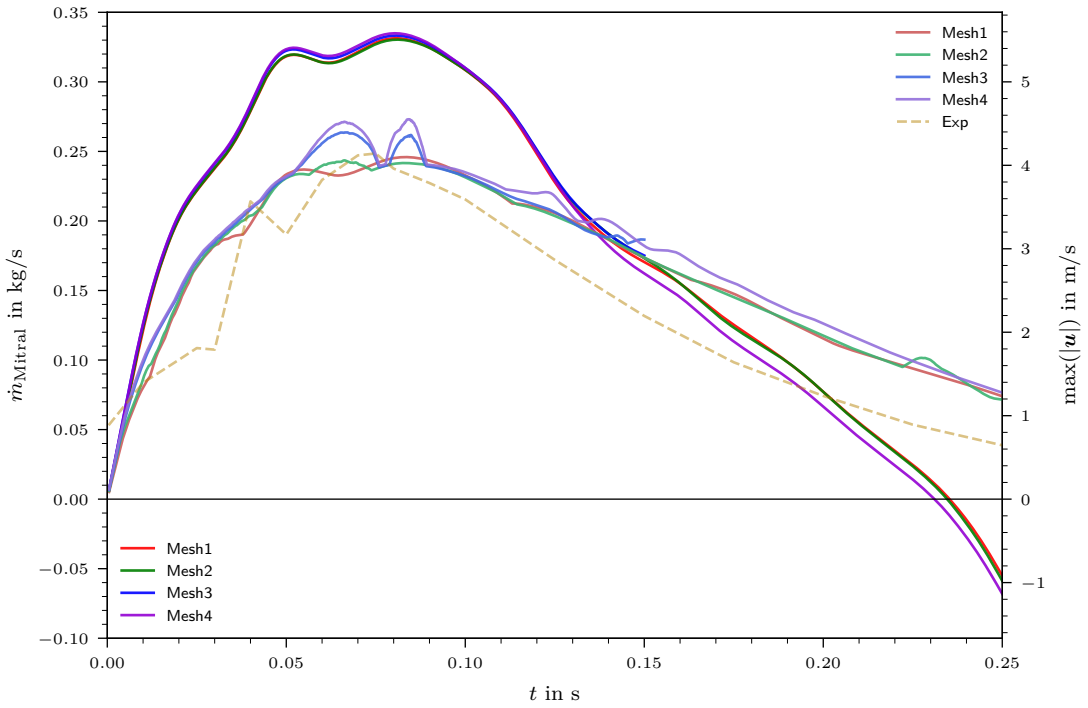

Figure A.1.: Mass flow through the orifice and global maximum velocity magnitudes in the beginning of the simulations for 4 different PinholeL meshes. Experimental values derived from PIV recording

A.2. 3D-Transient Validation

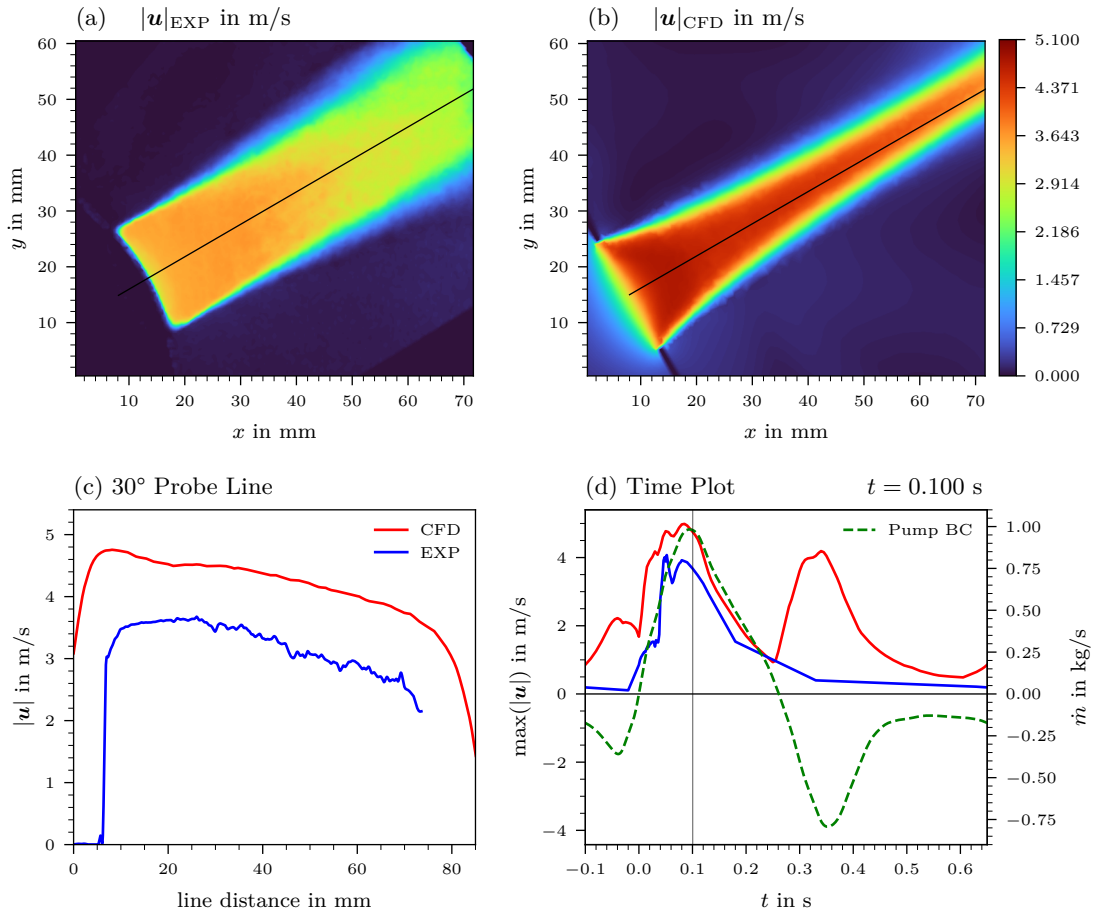

Figure A.2.: Comparison of PIV and CFD results for SlotL at $t = 0.1\,\mathrm{s}$. Magnitude overestimated and jet shape inaccurately rendered by CFD.

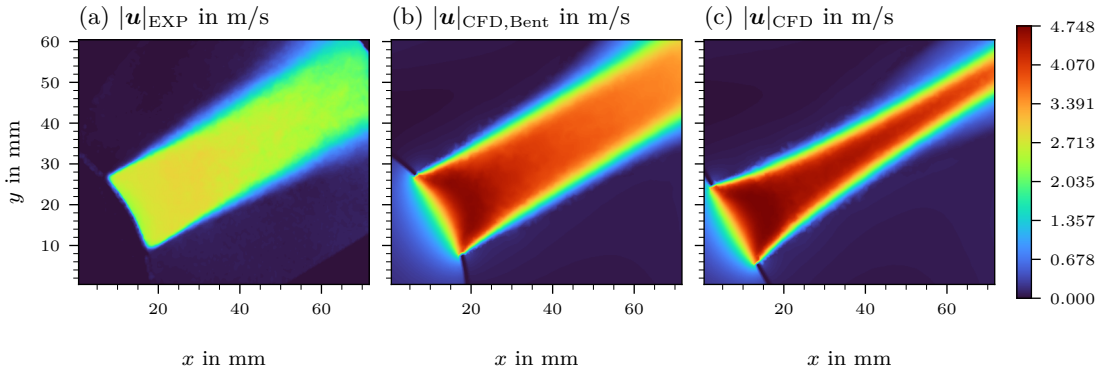

Figure A.3.: Comparison of PIV and CFD result for SlotL with deformed (b) and flat (c) rigid foil walls. While the magnitude remains inaccurate, the shape of the jet is more closely matched in (b).

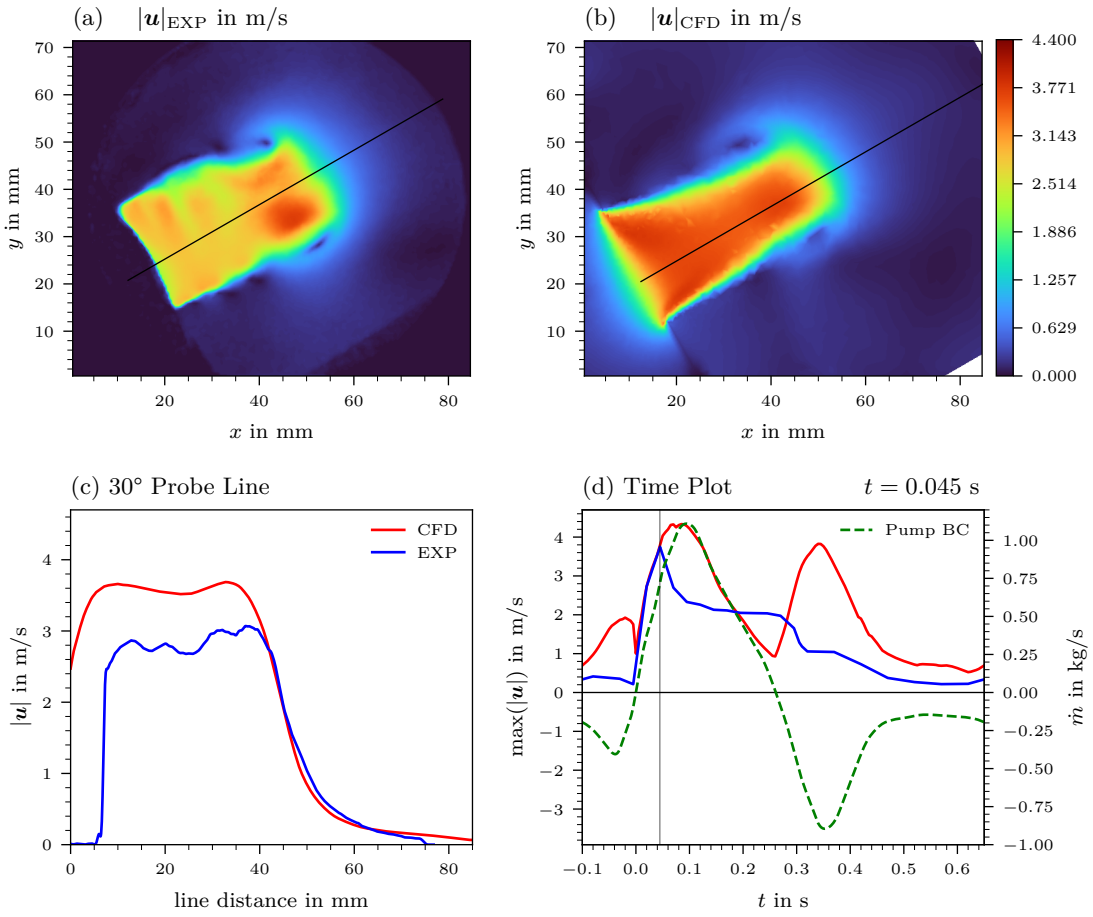

Figure A.4.: Comparison of PIV and CFD results for DropXL at $t = 0.045\,\mathrm{s}$. After this time-step, the cardiac pump reached its power-limit and "clipped" the waveform. The effect of this is demonstrated by the blue line in (d); the velocity magnitude suddenly drops off and maintains similar values for an uncharacteristically long time. This is not modeled in CFD, which uses the green BC line. A clipped (and inverted) pump BC trace is displayed in Neff [43] Figure 4.3 (a). The effect of this clipping is clearly visible in the pressure and aorta mass flow recordings too; not included in this work.

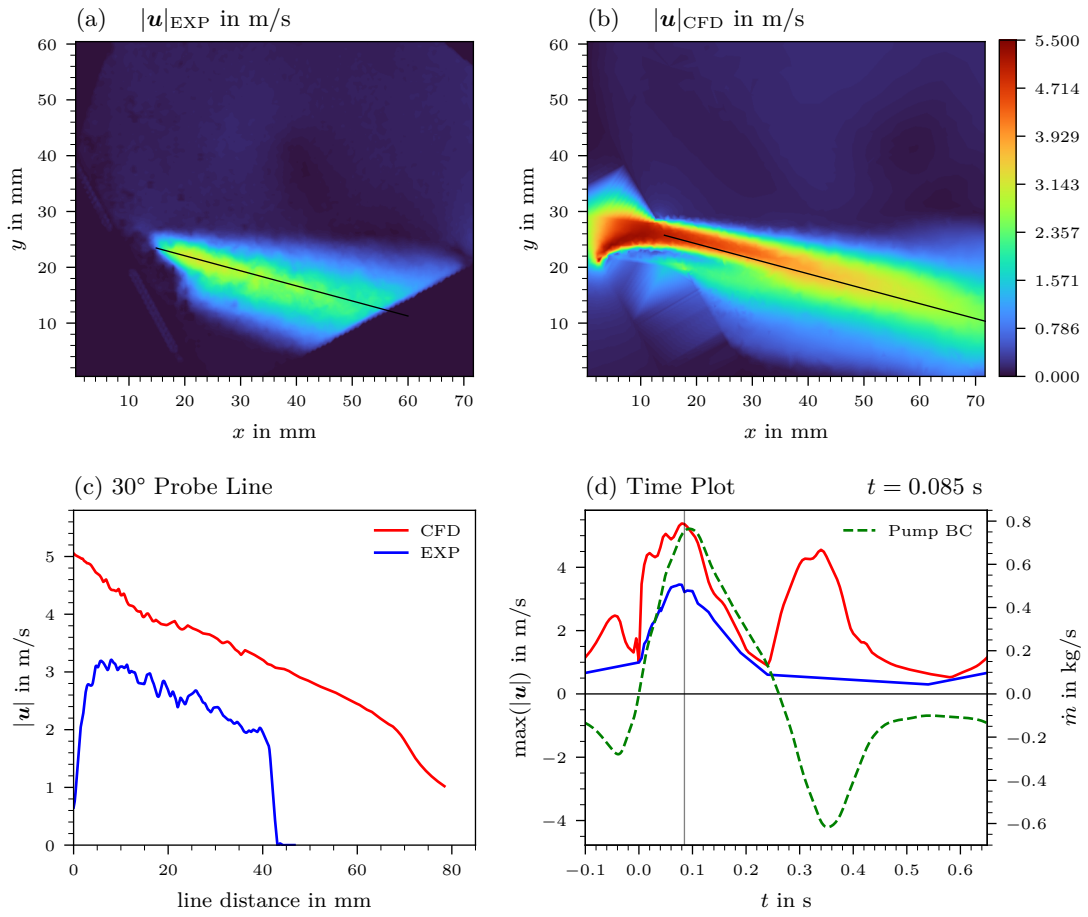

Figure A.5.: Comparison of PIV and CFD results for EccJet at $t = 0.085\,\mathrm{s}$. Plotting artifacts inside the 12 mm thick MROP. Magnitude overestimated by CFD.

B. Physics-Informed Machine Learning

B.1. Error Metrics

B.1.1. NRMSE

The normalized root-mean-square error (NRMSE) is the primary error metric chosen for evaluations of network prediction $\hat{y}$ against ground truth data y of N discrete values. This is also the default PhysicsNeMo Sym evaluation metric. The normalization happens via the standard deviation σ, where $\overline{y}$ represents the mean of the ground truth data:

$$\text{NRMSE} = \frac{\text{RMSE}}{\sigma} = \frac{\sqrt{\frac{1}{N}\sum_k \left(\hat{y}_k - y_k\right)^2}}{\sqrt{\frac{1}{N}\sum_k \left(y_k - \overline{y}\right)^2}} \quad \text{with} \quad \overline{y} = \frac{1}{N}\sum_k y_k \tag{5.1}$$

If a trivial model simply predicted the mean value of the dataset for every single point, the RMSE would equal the standard deviation. Therefore, a value of 1.0 indicates the model has no predictive power beyond knowing the average value of the target variable. Values lower than 1.0 indicate the error is smaller than the natural variance of the data. This metric is mathematically related to the R^2 score commonly used in regression. While R^2 measures the variance explained, NRMSE measures the variance unexplained [51]:

$$R^2 = 1 - \text{NRMSE}^2 \tag{5.2}$$

For example, an NRMSE of 0.9 corresponds to $R^2 = (1 - 0.9^2) = 0.19$. This suggests that only 19% of the variance in the target field is "explained" or captured by the model's prediction. Conversely, 81% of the variance remains unexplained, indicating that there are relevant factors or dynamics that the model has failed to capture. Similarly, an NRMSE of 0.5 already captures 75% of the variance.

B.1.2. PDE Residuals

To evaluate the satisfaction of the governing equations, the NSE are now considered in their compressible form. Because the model outputs might not strictly satisfy the continuity equation, the simplifications from a divergence free velocity field no longer apply to the momentum equations. The compressible form implemented in PhysicsNeMo Sym [44, 45] is modified in the same way as Equation (2.6) with non-dimensional numbers and the effective viscosity turbulence model. All terms are moved to one side and named in the following way to build the residuals:

$$\begin{aligned} r_{\text{mom},i} = & \underbrace{\text{St}\,\frac{\partial u_i^*}{\partial t^*}}_{\text{Transient } T_i} + \underbrace{u_j^*\frac{\partial u_i^*}{\partial x_j^*}}_{\text{Convection } C_i} + \underbrace{\text{Eu}\,\frac{\partial p^*}{\partial x_i^*}}_{\text{Pressure } P_i} \\ & \underbrace{-\frac{1}{\text{Re}}\frac{\partial}{\partial x_j^*}\left(\mu_{\text{eff}}\frac{\partial u_i^*}{\partial x_j^*}\right)}_{\text{Shear } S_i} \underbrace{-\frac{1}{\text{Re}}\left[\frac{\partial}{\partial x_i^*}\left(-\frac{2}{3}\mu_{\text{eff}}\,\Theta\right) + \mu_{\text{eff}}\frac{\partial \Theta}{\partial x_i^*}\right]}_{\text{Dilatational } D_i} \underbrace{-\frac{1}{\text{Re}}\left(\frac{\partial \mu_{\text{eff}}}{\partial x_j^*}\frac{\partial u_j^*}{\partial x_i^*}\right)}_{\text{Coupling } V_i} \\ & \text{and} \quad r_{\text{conti}} = \frac{\partial u_j^*}{\partial x_j^*} =: \Theta \end{aligned} \tag{5.3}$$

where the viscous term is split up into S_i, D_i and V_i. The dilatational term D_i is exactly zero when the velocity field is divergence free. Removing this part under the assumption

of incompressibility and restructuring S_i and D_i is what leads to Equation (2.6). The viscosity–gradient coupling term C_i is zero if μ_{eff} is constant and large if it has steep gradients due to the turbulent viscosity field.

The PDE satisfaction of models is quantified using the L1 integral norms of these PDE residuals, which are calculated in the following way: The instantaneous spatial error approximates the spatial integral of the absolute residual (L1 norm) over the domain volume Ω at a specific time instance t. This is computed by weighing the absolute residual value at each discrete sampling point by a uniform spatial weight, defined as the total non-dimensional domain volume, 1.5421 for 3D-Transient, divided by the number of sampled points:

$$\varepsilon_{\text{space}}(t) = \int_{\Omega} |r(t)| \, d\Omega \approx \sum_{k} w_{\text{space}} \, |r_k(t)| \quad \text{where} \quad w_{\text{space}} = \frac{|\Omega|^*}{N_{\text{points}}} \tag{5.4}$$

The time-integrated total error approximates the integral of the error over both space and the time duration. Subsequently, the total spatiotemporal error is obtained by aggregating the instantaneous spatial errors over the simulation duration using a temporal weight, calculated as the total time span divided by the number of analyzed snapshots:

$$\varepsilon_{\text{total}} = \int_{t} \int_{\Omega} |r(t)| \, d\Omega \, dt \approx \sum_{n} w_{\text{time}} \varepsilon_{\text{space}}(t_n) \quad \text{where} \quad w_{\text{time}} = \frac{t_{\max} - t_{\min}}{N_{\text{snapshots}}} \tag{5.5}$$

These weights are necessary to normalize the discrete summation, ensuring the resulting error metric remains invariant to changes in sampling density across space and time. To isolate specific sources of physical error, the constituent terms of $r_{\text{mom},i}$ (T_i, C_i, ...) are evaluated independently before their summation into the final aggregate errors.

B.2. PINN Verification

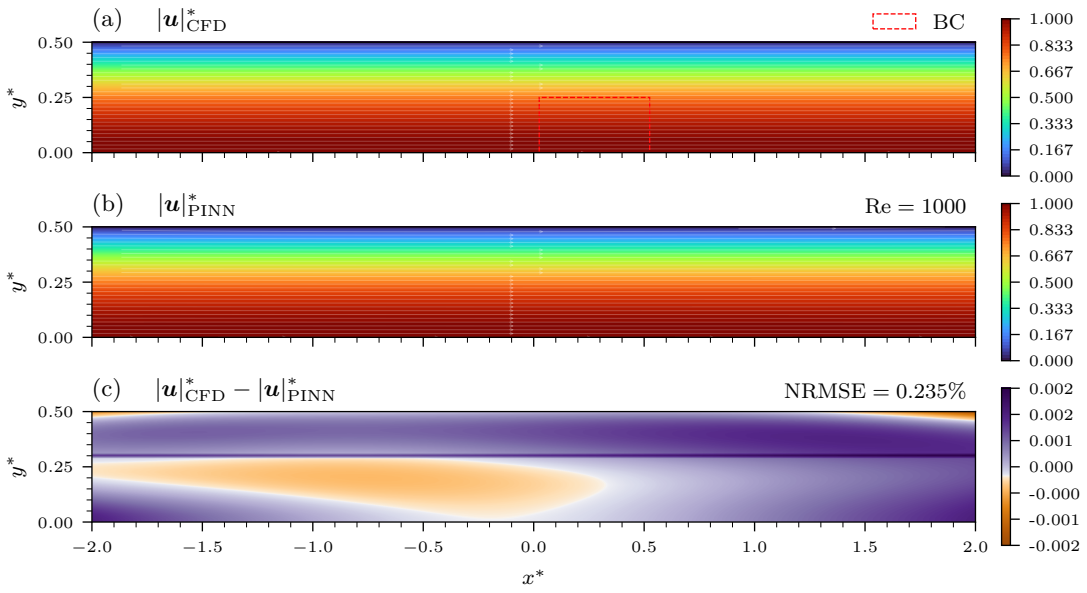

Figure B.6.: Comparison of (a) CFD ground truth and (b) PINN velocity magnitudes for laminar 2D channel flow at Re = 1000. Velocities u, v inside the red rectangle and pressures p at top left and right corner supplied as data supervision. No-slip and symmetry BCs included. 5 hidden layers at 512 neurons each. Results in exceptionally low error (c).

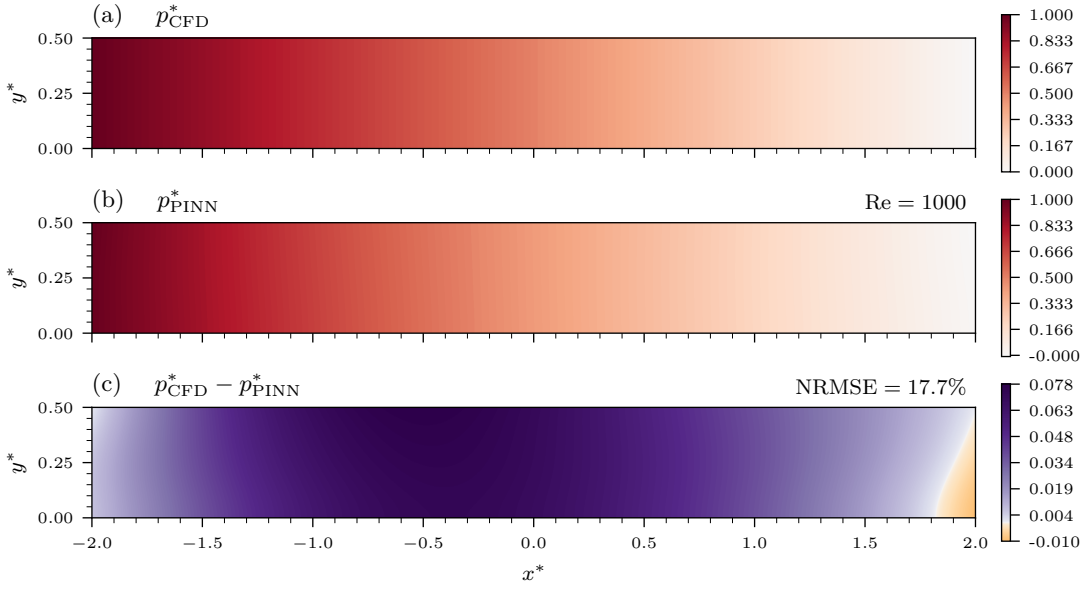

Figure B.7.: Comparison of (a) CFD ground truth and (b) PINN pressure fields for same setup as Figure B.6. Error (c) worse but would have further improved with more training steps.

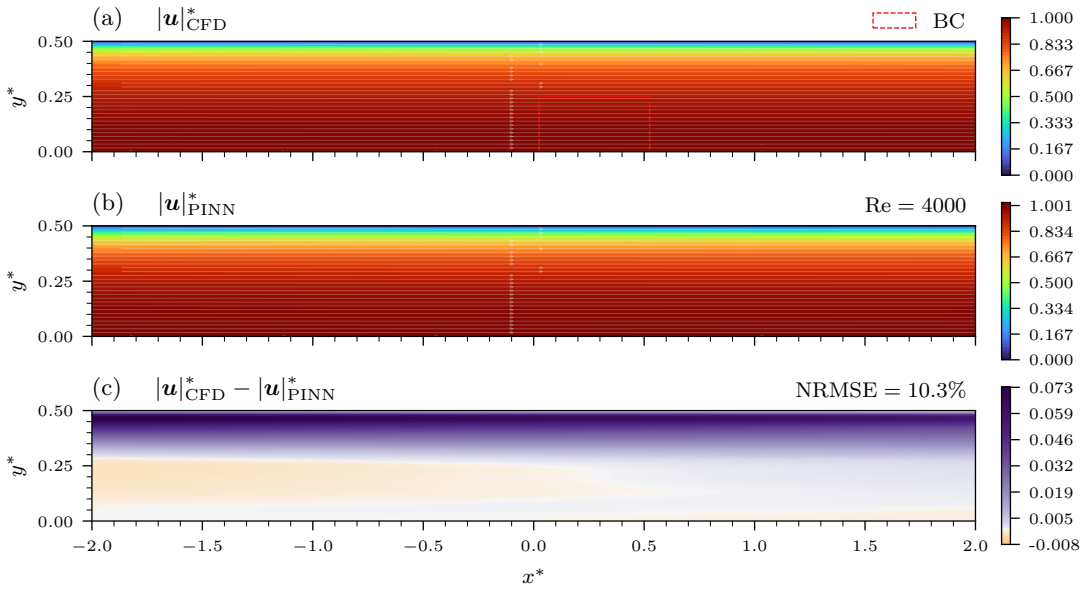

Figure B.8.: Comparison of (a) CFD ground truth and (b) PINN velocity magnitudes for turbulent 2D channel flow at Re = 4000. Velocities u, v inside the red rectangle and pressures p at top left and right corner supplied as data supervision. No-slip and symmetry BCs included. Additional full supervision of turbulent viscosity ν_t over entire domain to test turbulence model, as it struggled to learn this field without any labeled data. 5 hidden layers at 512 neurons each. Error (c) generally low, but increases with distance from velocity data, suggesting too early training termination or potentially unsuccessful implementation of ν_t turbulence model.

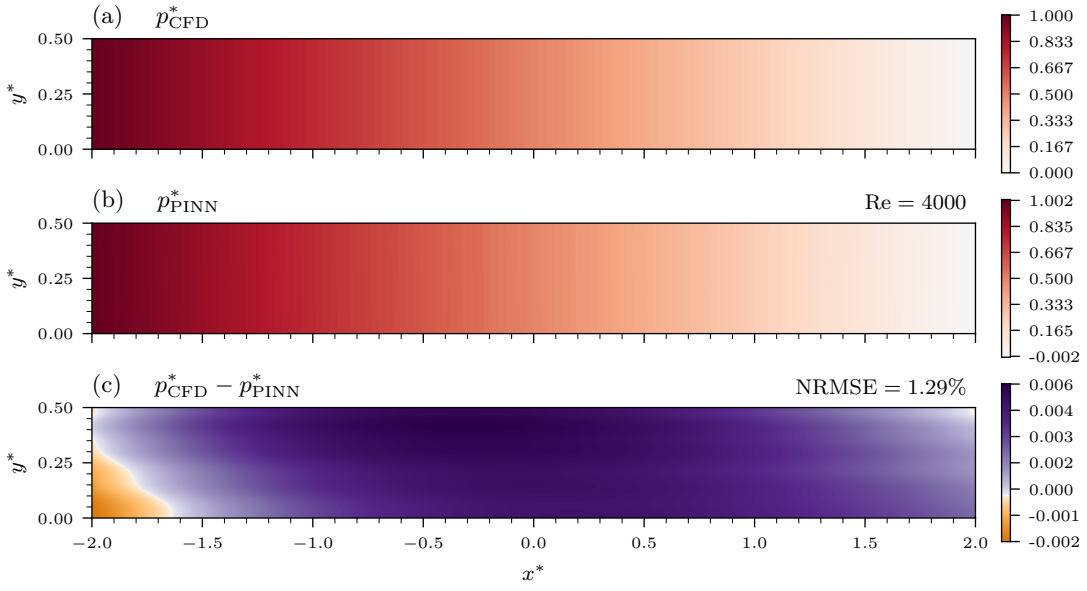

Figure B.9.: Comparison of (a) CFD ground truth and (b) PINN pressure fields for same setup as Figure B.8. Results in exceptionally low error (c).

C. Results

C.1. 2D-Steady

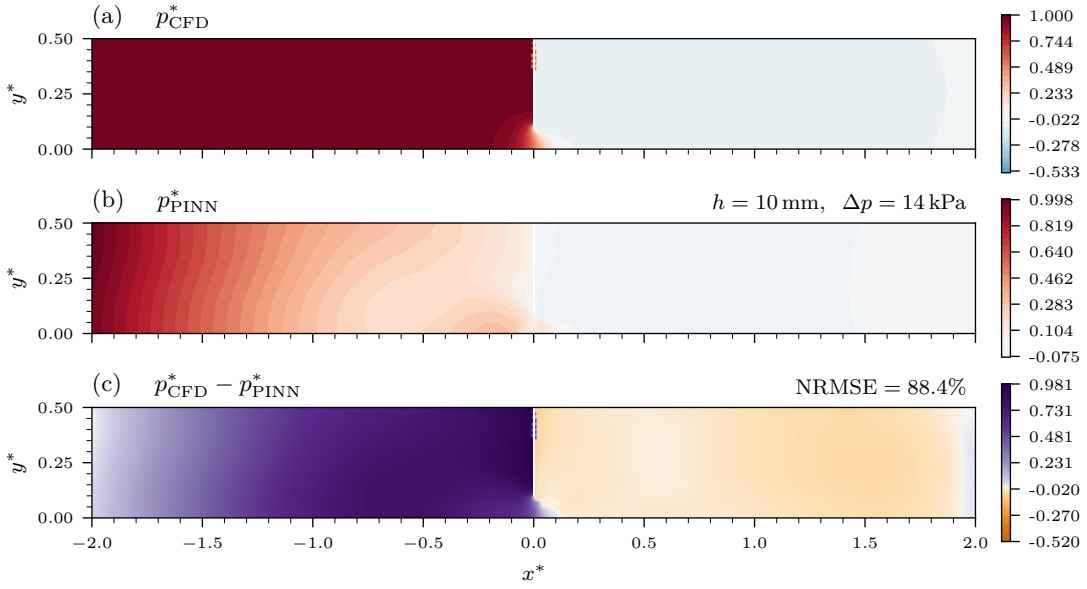

Figure C.10.: (a) Reference 2D-Steady CFD pressure for $h = 10\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$. (b) "PINN RANS" prediction. (c) Signed non-linear error.

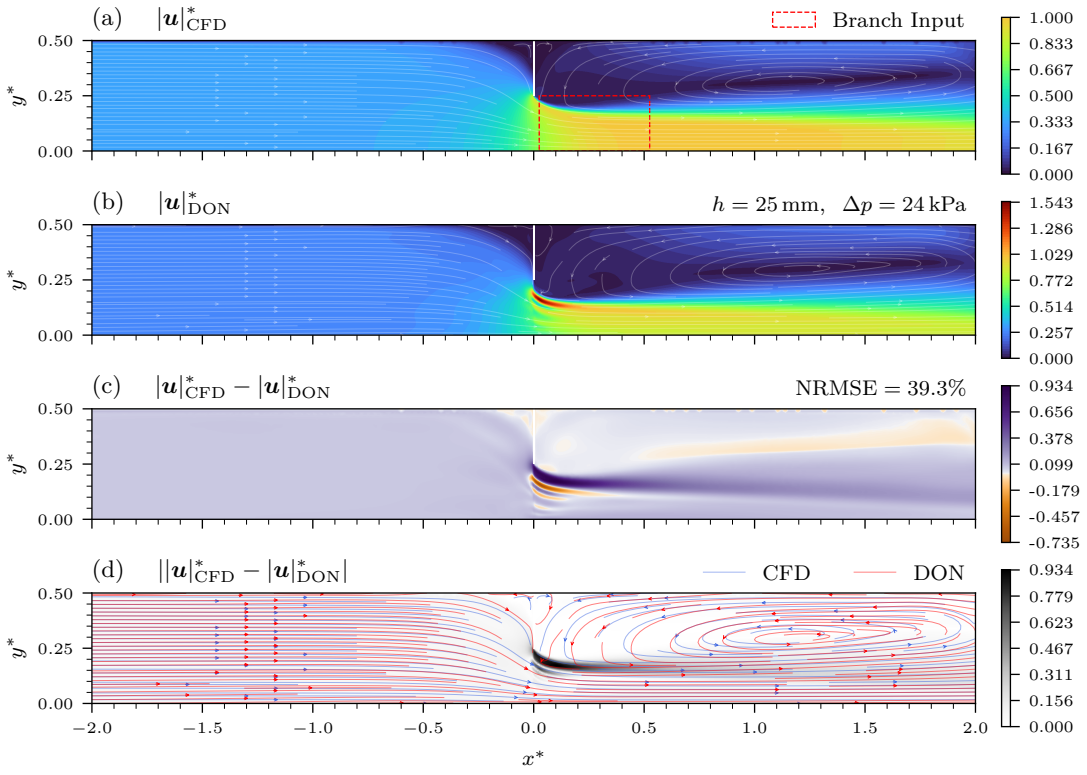

Figure C.11.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 25\,\mathrm{mm}$, $\Delta p = 24\,\mathrm{kPa}$. (b) "Dist20" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

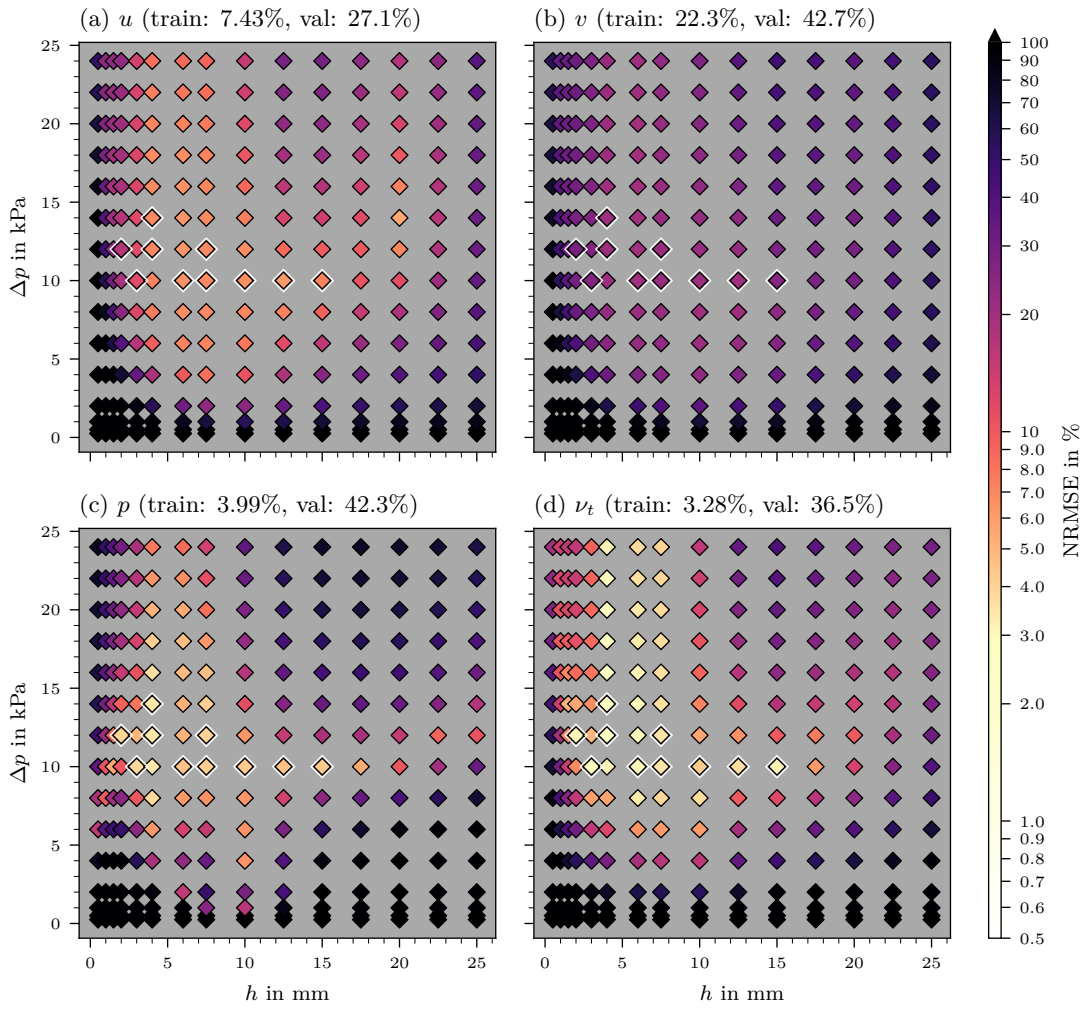

Figure C.12.: Distribution of NRMSE for 2D-Steady "Hybrid Conti Extra" for all fields across the BC parameter space defined by orifice height and pressure difference. The color of each marker indicates the error magnitude, with the 10 training cases distinguished by a white border. Values higher than 100% are clipped.

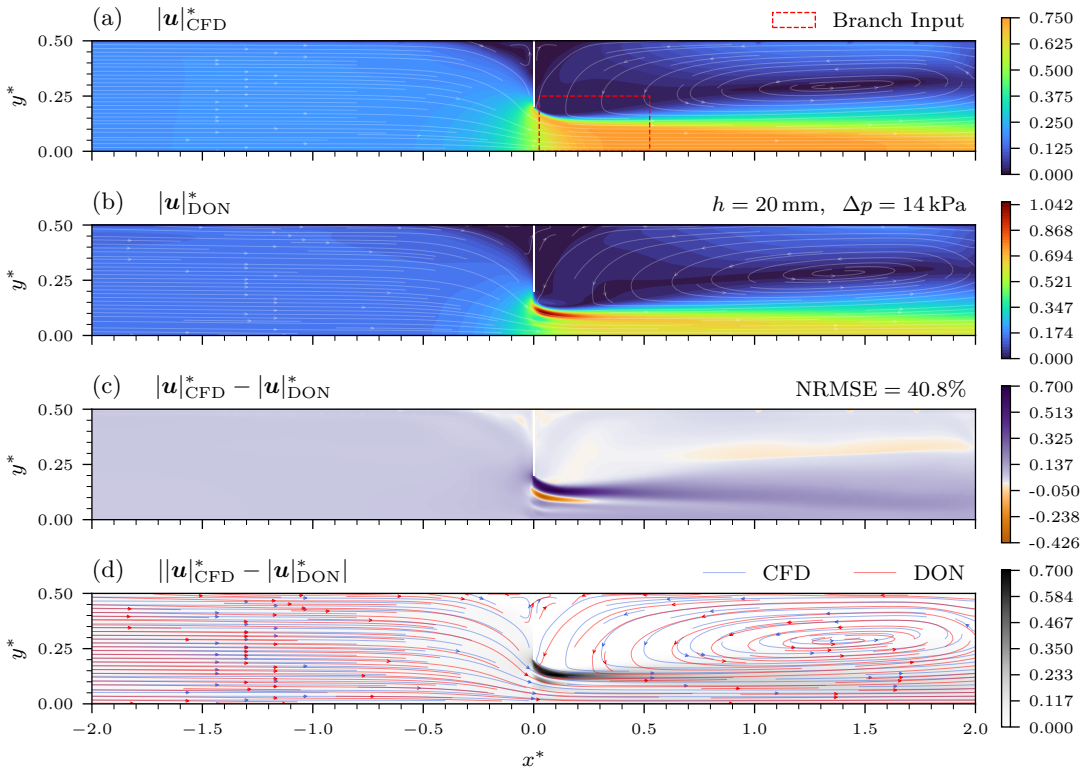

Figure C.13.: (a) Reference 2D-Steady CFD velocity magnitude for $h = 20\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$. (b) "Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

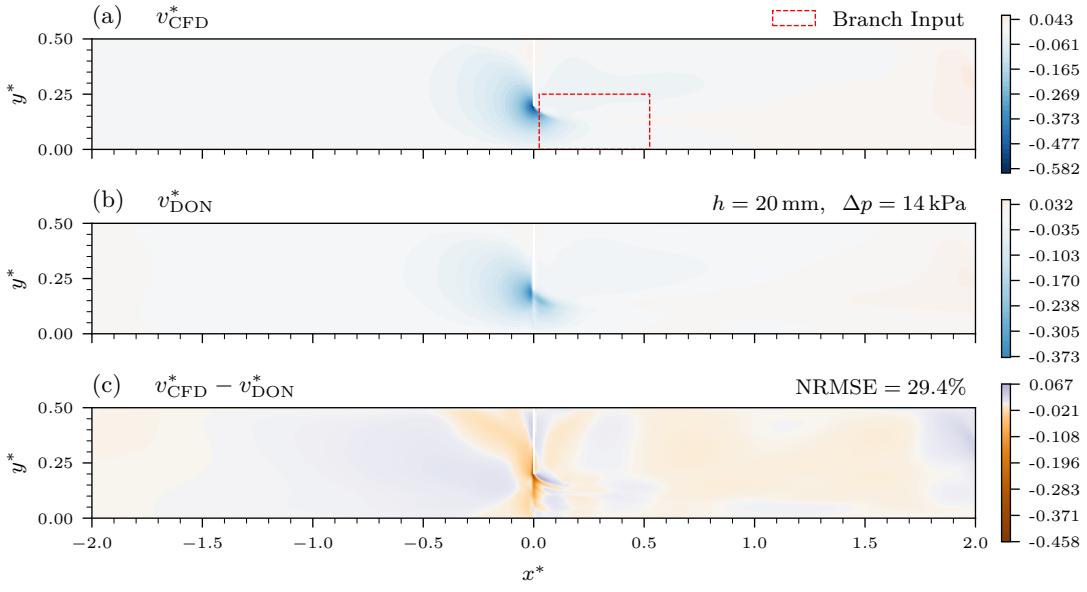

Figure C.14.: (a) Reference 2D-Steady CFD v-velocity for $h = 20\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$. (b) "Hybrid Conti" prediction. (c) Signed non-linear error.

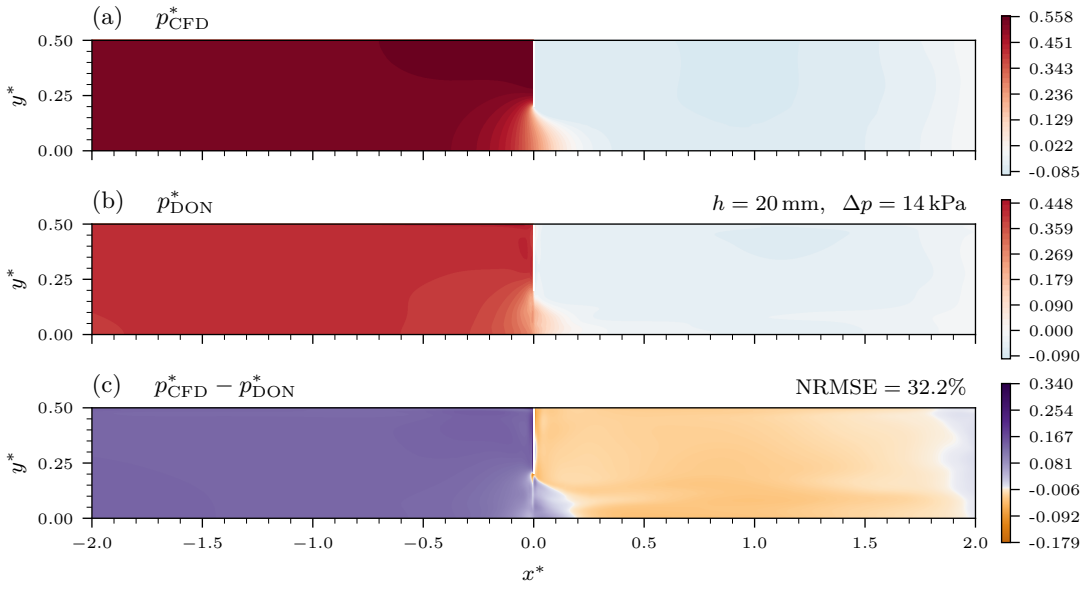

Figure C.15.: (a) Reference 2D-Steady CFD pressure for $h = 20\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$. (b) "Hybrid Conti" prediction. (c) Signed non-linear error.

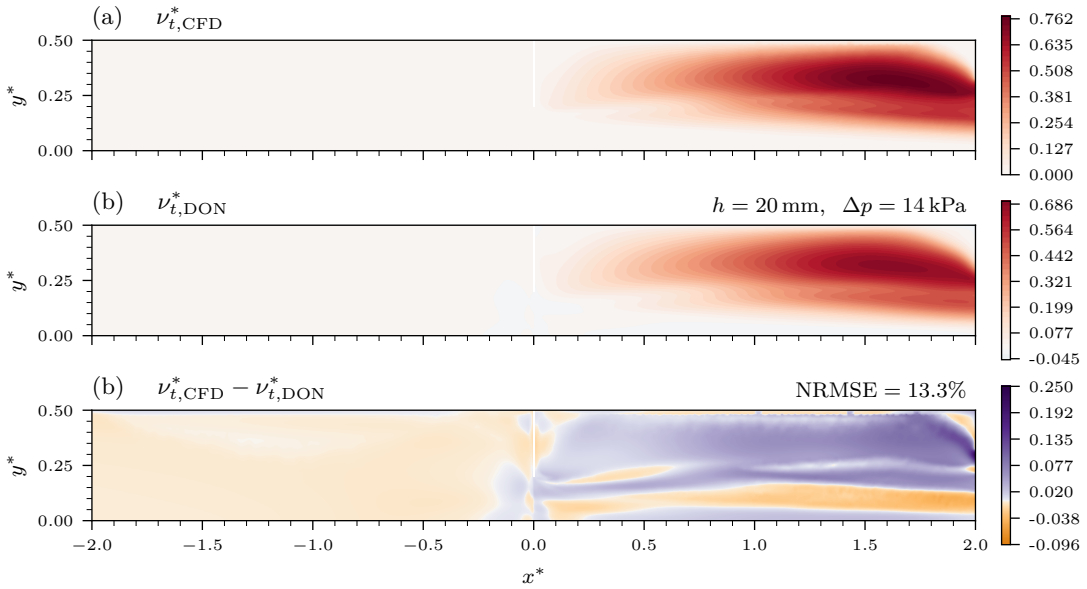

Figure C.16.: (a) Reference 2D-Steady CFD turbulent viscosity for $h = 20\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$. (b) "Hybrid Conti" prediction. (c) Signed non-linear error.

C.2. 2D-Transient

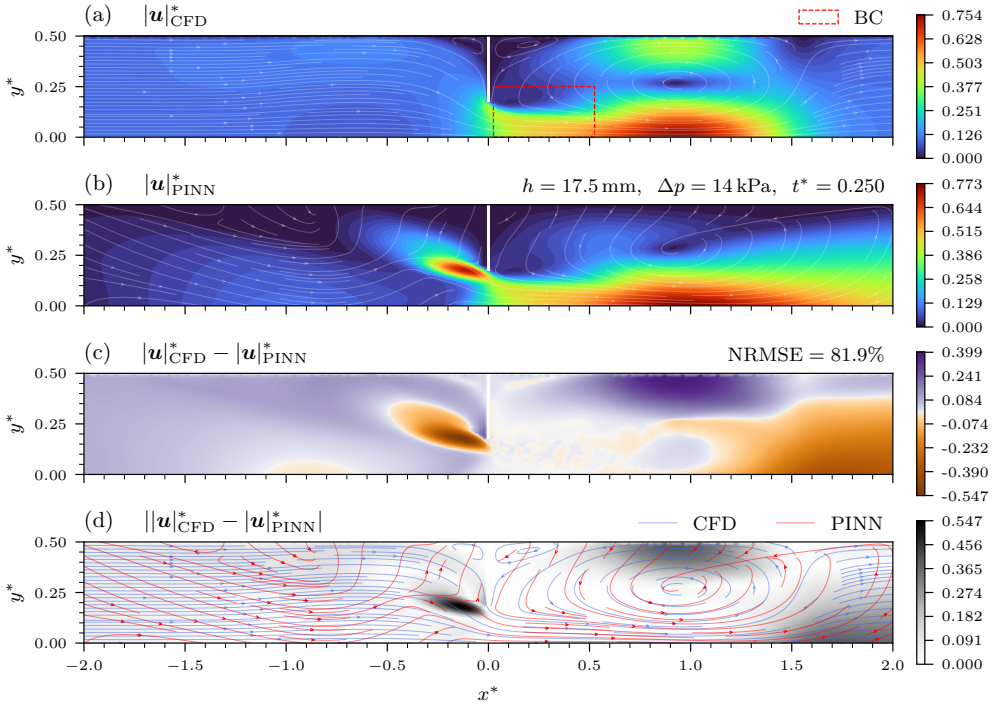

Figure C.17.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.25$. (b) "PINN DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

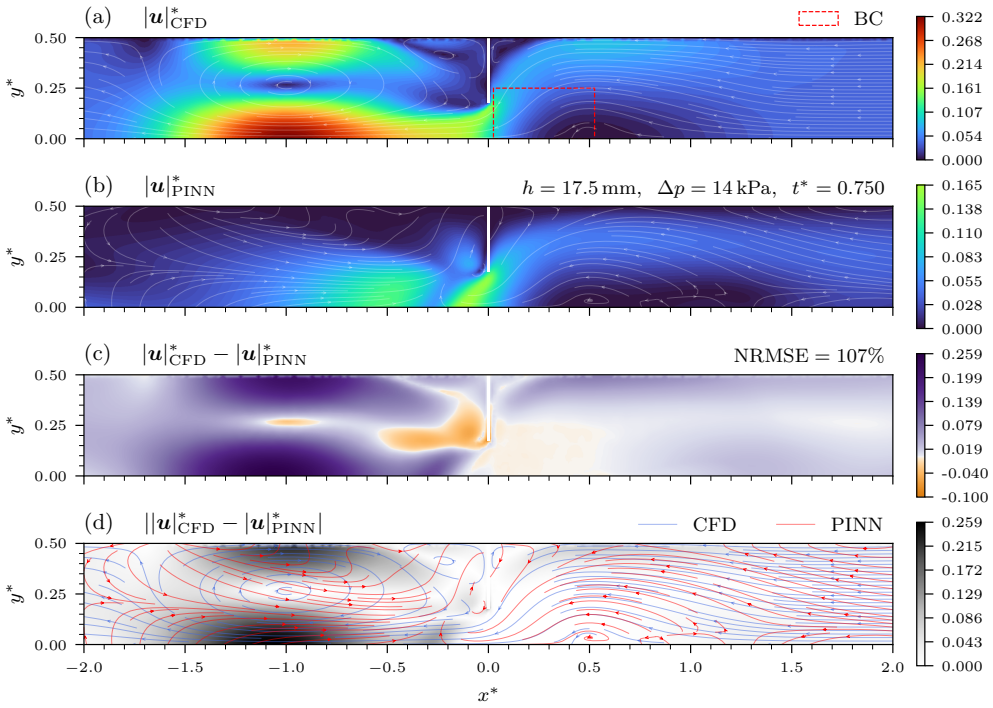

Figure C.18.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.75$. (b) "PINN DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamline comparison.

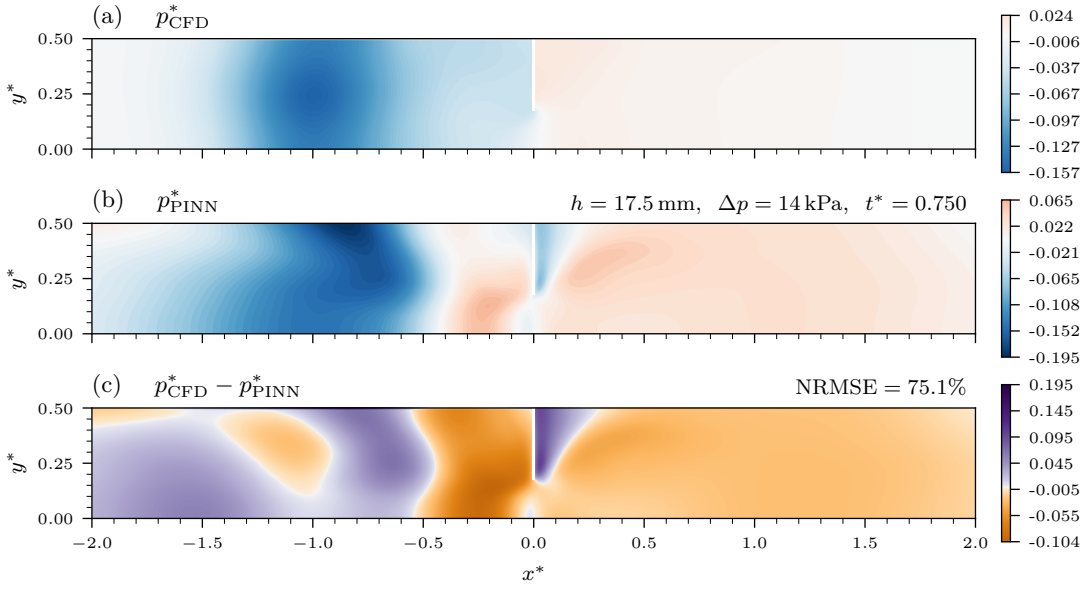

Figure C.19.: (a) Reference 2D-Transient CFD pressure for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.75$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

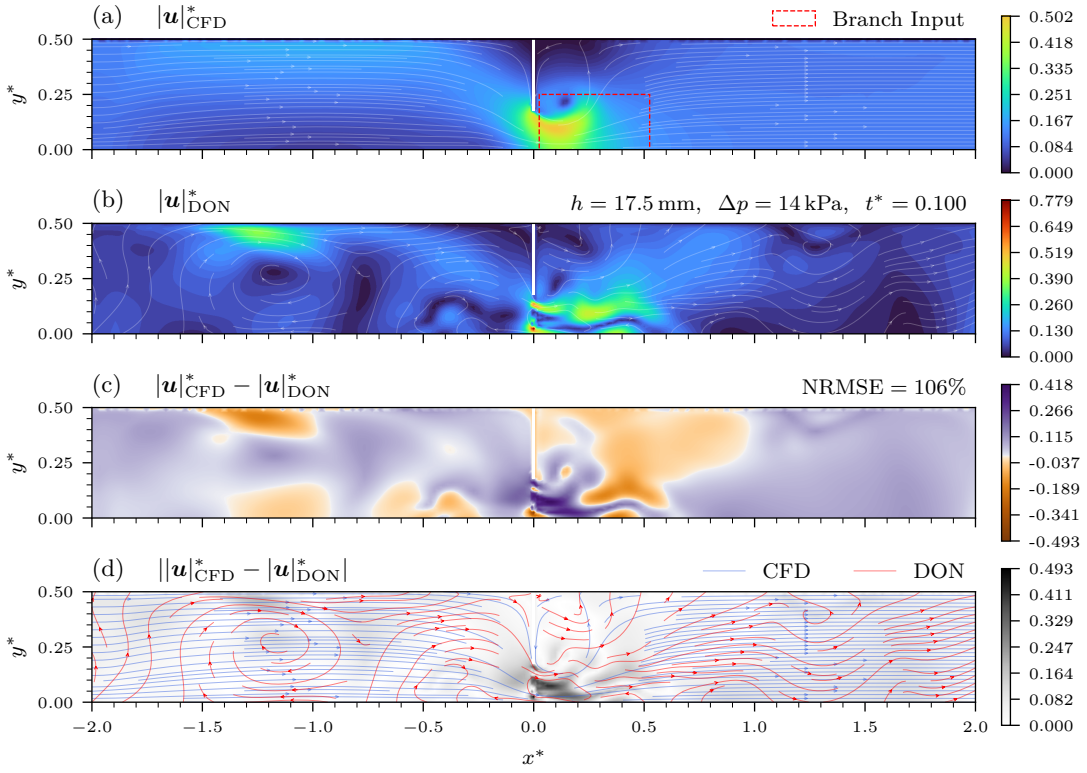

Figure C.20.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.1$. (b) "Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

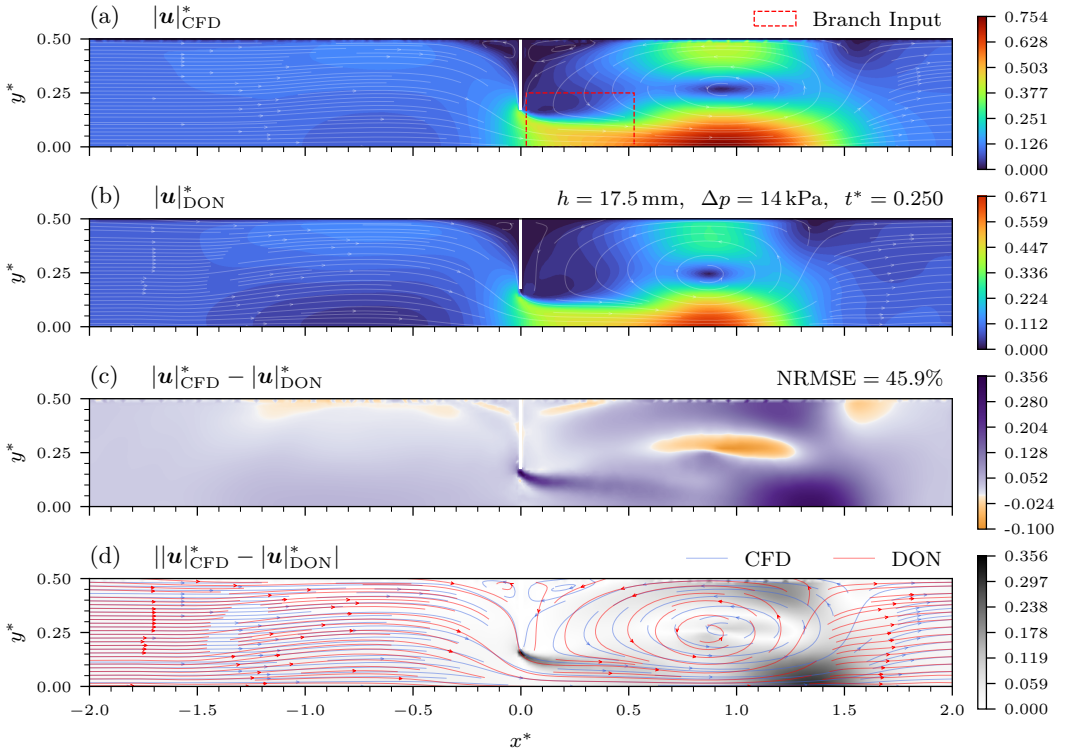

Figure C.21.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.25$. (b) "Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

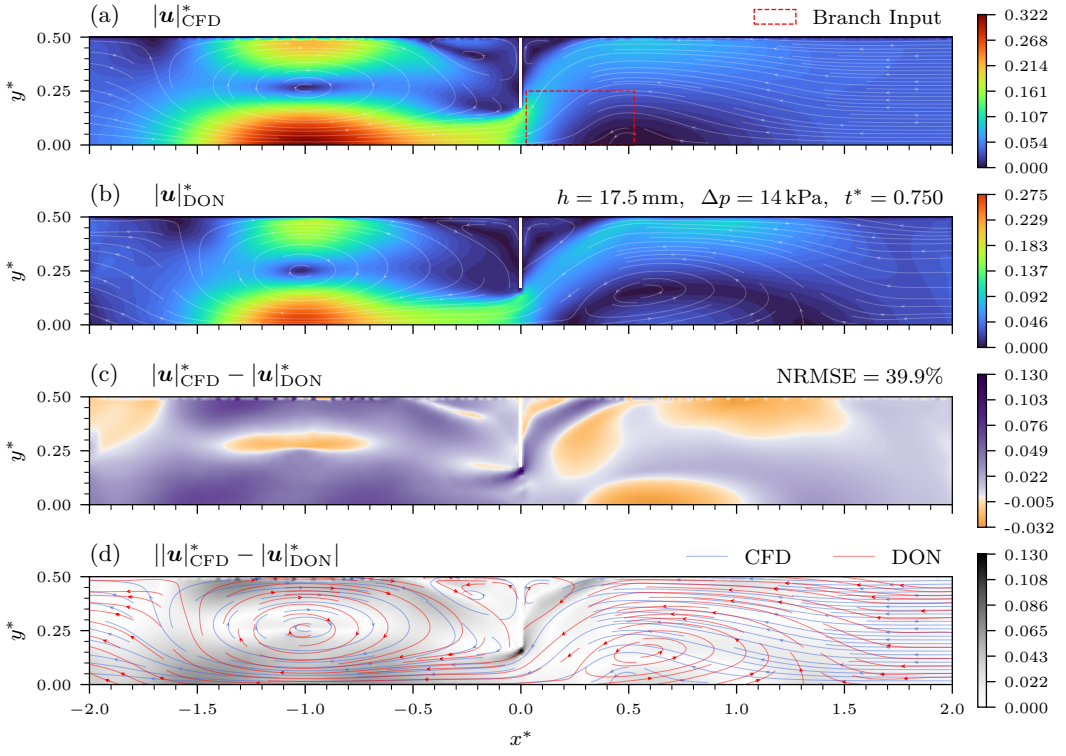

Figure C.22.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.75$. (b) "Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

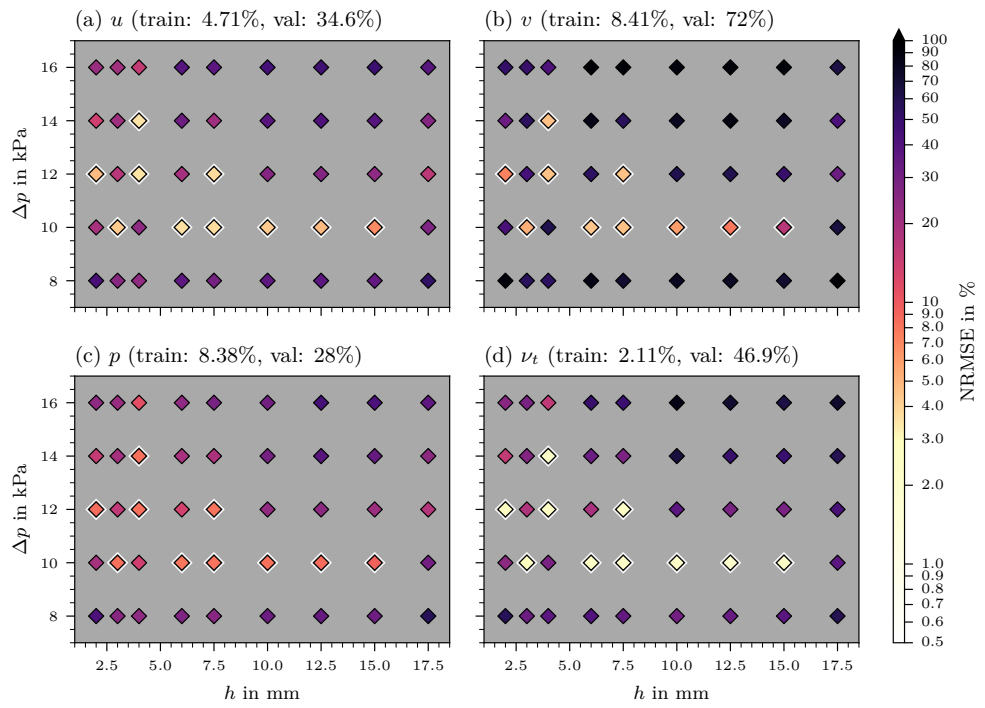

Figure C.23.: Distribution of NRMSE for 2D-Transient "Hybrid DD Extra" for all fields across the BC parameter space defined by orifice height and pressure difference. The color of each marker indicates the error magnitude over the cardiac cycle, with the 10 training cases distinguished by a white border. Values higher than 100% are clipped.

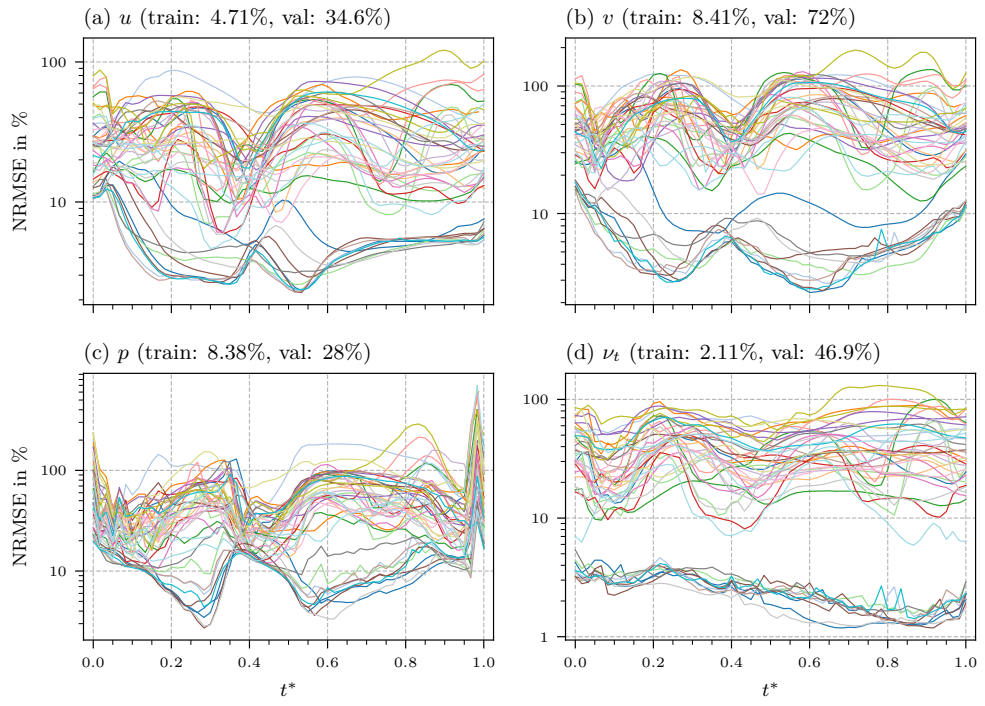

Figure C.24.: Temporal evolution of NRMSE for all fields over the cardiac cycle for the 2D-Transient "Hybrid DD Extra" model. The 45 lines represent all cases from the BC distribution.

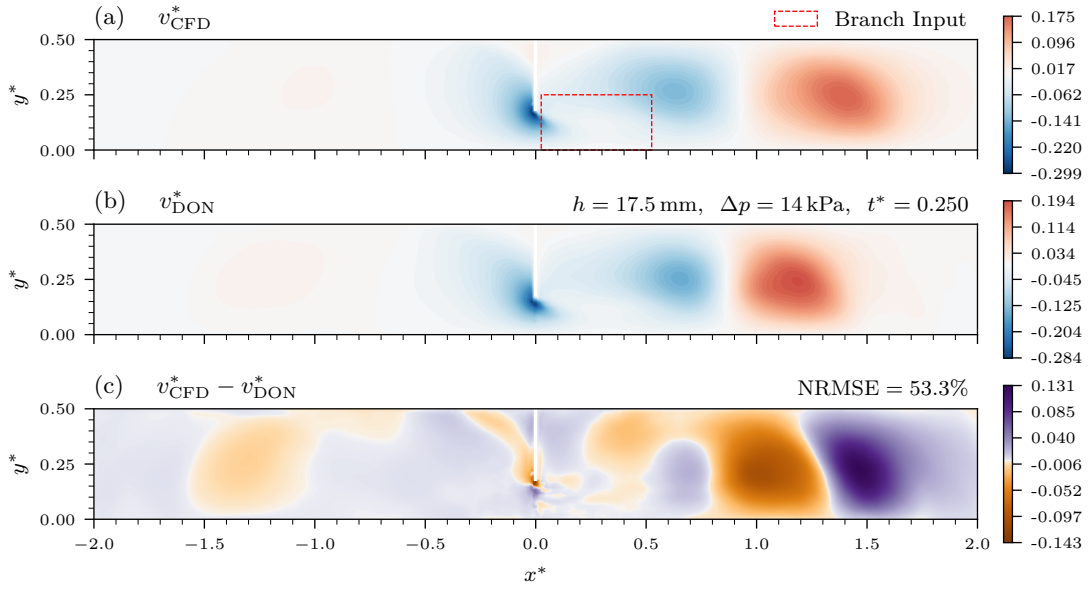

Figure C.25.: (a) Reference 2D-Transient CFD v-velocity for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

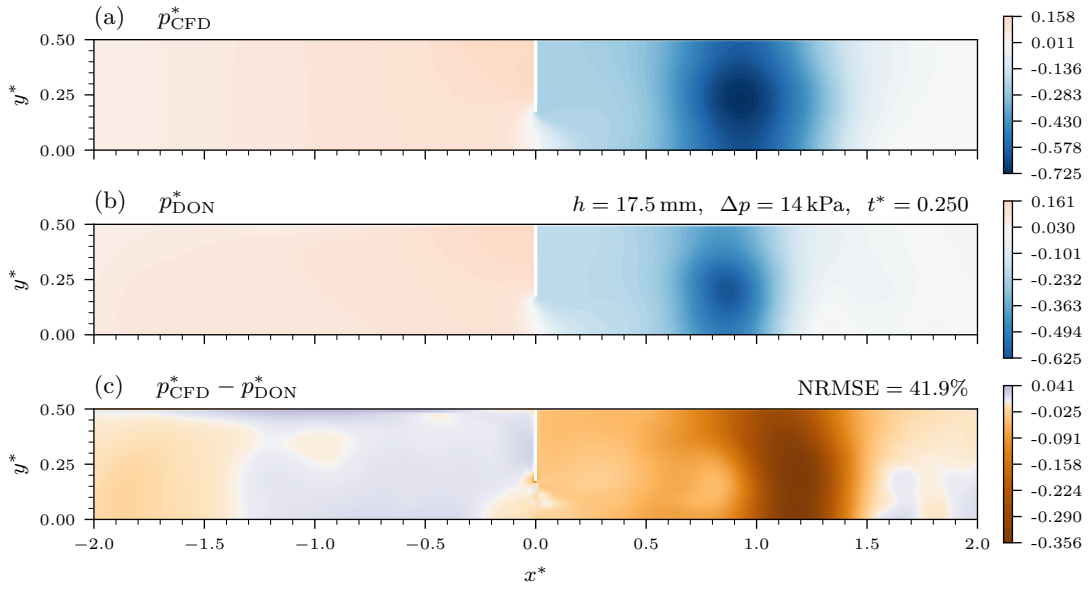

Figure C.26.: (a) Reference 2D-Transient CFD pressure for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

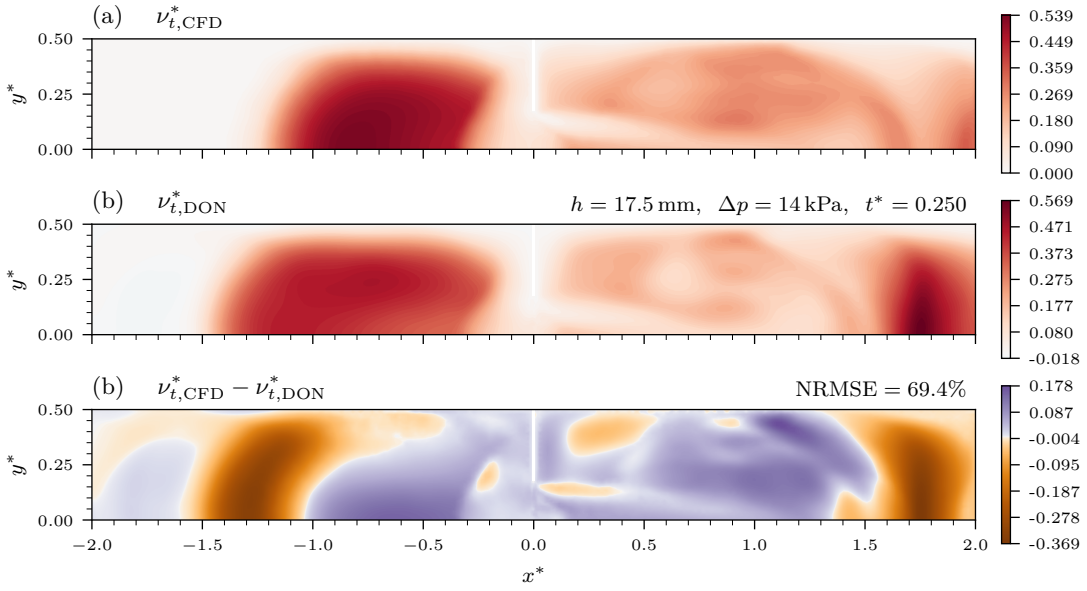

Figure C.27.: (a) Reference 2D-Transient CFD turbulent viscosity for $h = 17.5\,\mathrm{mm}$, $\Delta p = 14\,\mathrm{kPa}$, $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

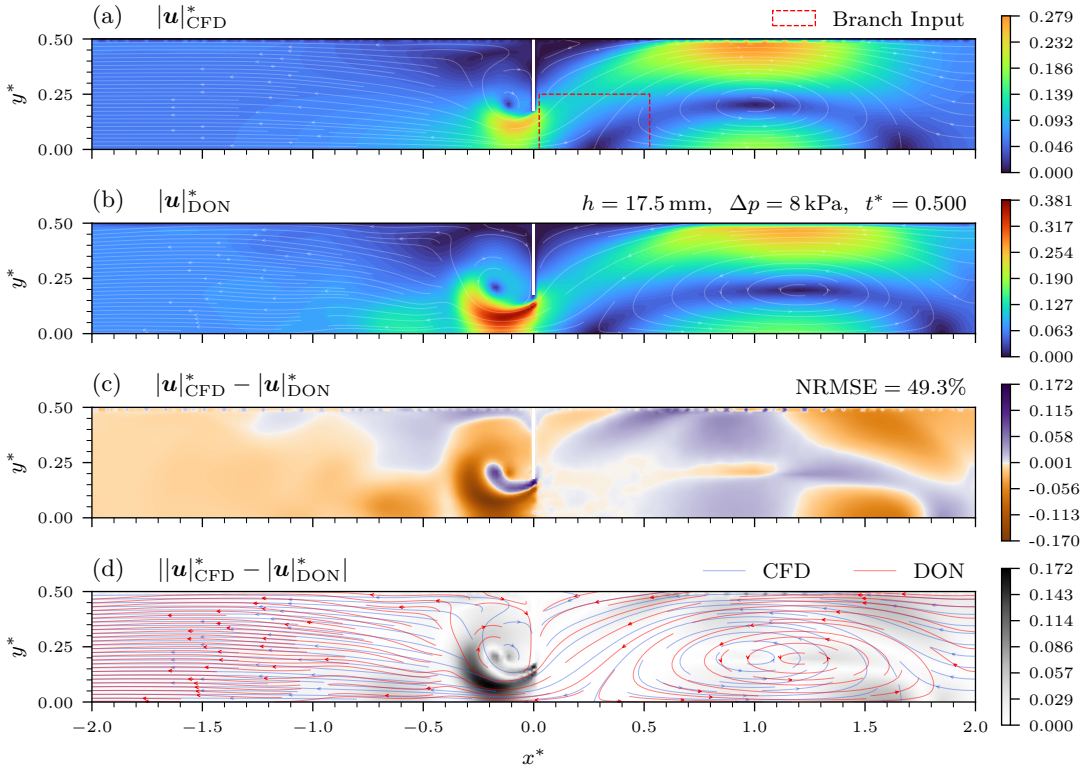

Figure C.28.: (a) Reference 2D-Transient CFD velocity magnitude for $h = 17.5\,\mathrm{mm}$, $\Delta p = 8\,\mathrm{kPa}$, $t^* = 0.5$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

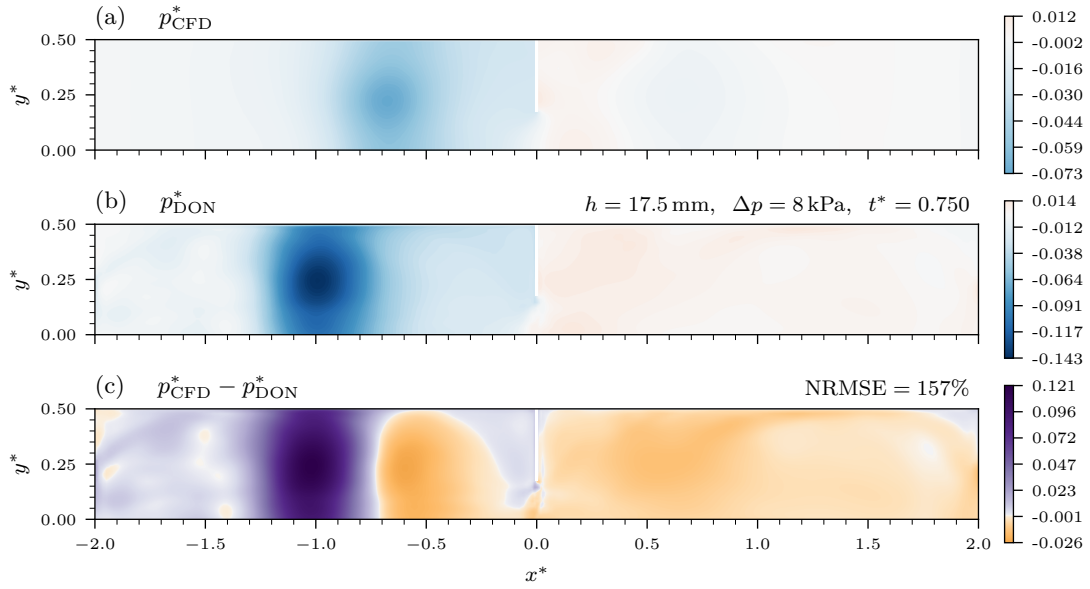

Figure C.29.: (a) Reference 2D-Transient CFD pressure for $h = 17.5\,\mathrm{mm}$, $\Delta p = 8\,\mathrm{kPa}$, $t^* = 0.75$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

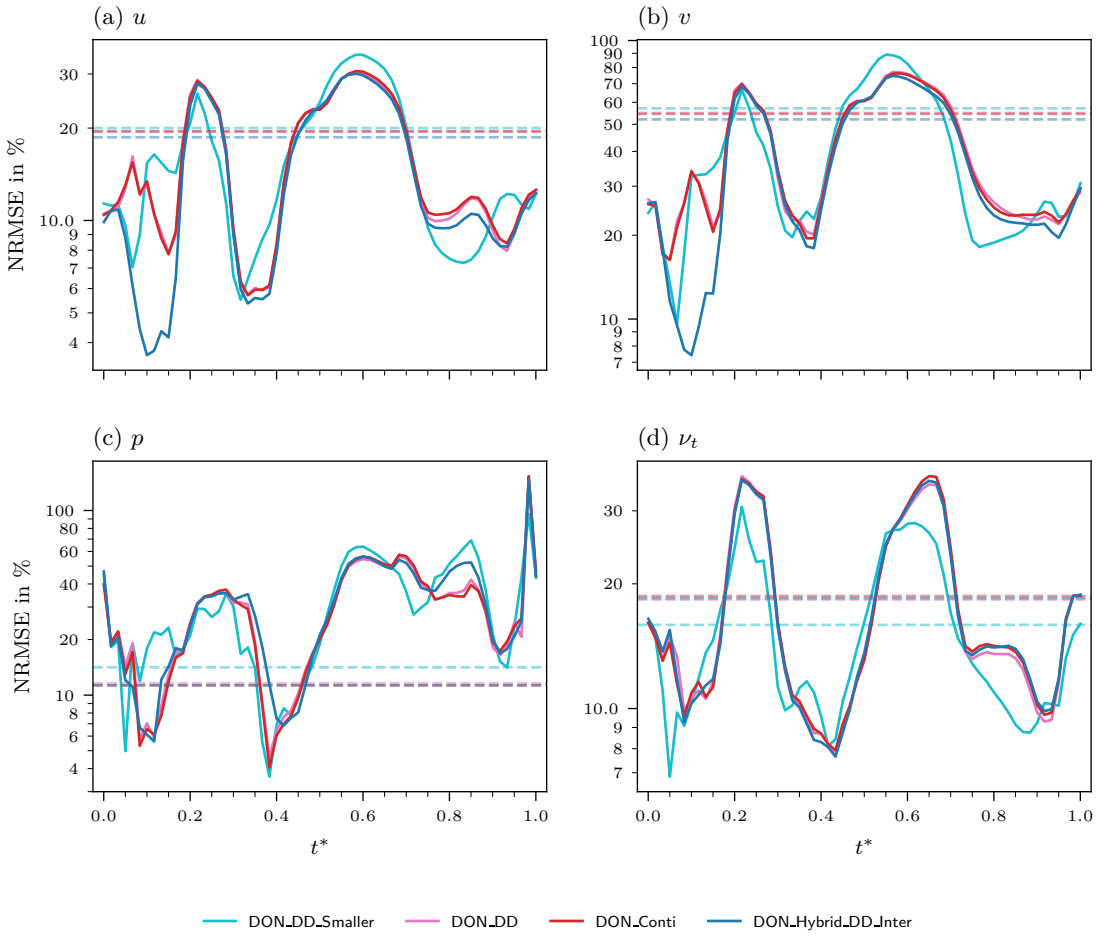

Figure C.30.: Temporal evolution of the NRMSE for all fields over the cardiac cycle for the 2D-Transient $h = 6\,\mathrm{mm}$, $\Delta p = 12\,\mathrm{kPa}$ case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

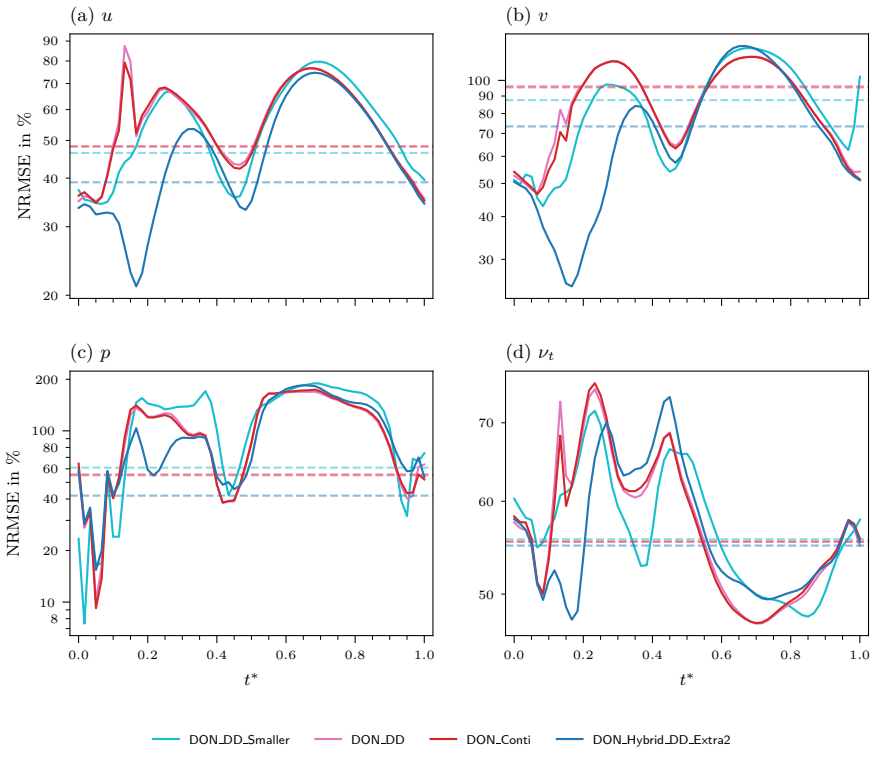

Figure C.31.: Temporal evolution of the NRMSE for all fields over the cardiac cycle for the 2D-Transient $h = 17.5\,\text{mm}$, $\Delta p = 8\,\text{kPa}$ case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

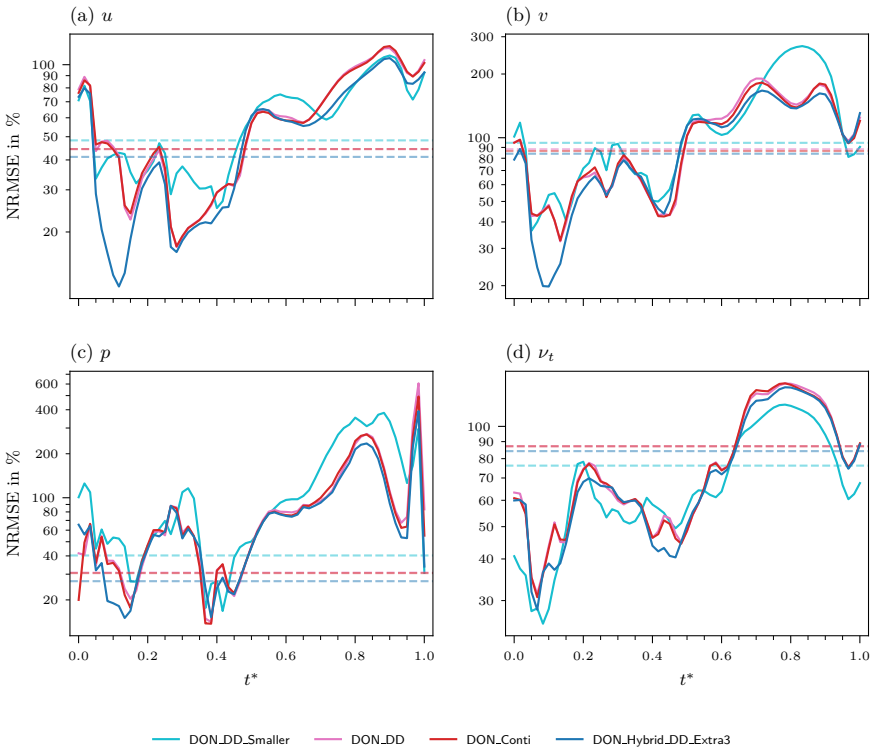

Figure C.32.: Temporal evolution of the NRMSE for all fields over the cardiac cycle for the 2D-Transient $h = 10\,\text{mm}$, $\Delta p = 16\,\text{kPa}$ case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

C.3. 3D-Transient

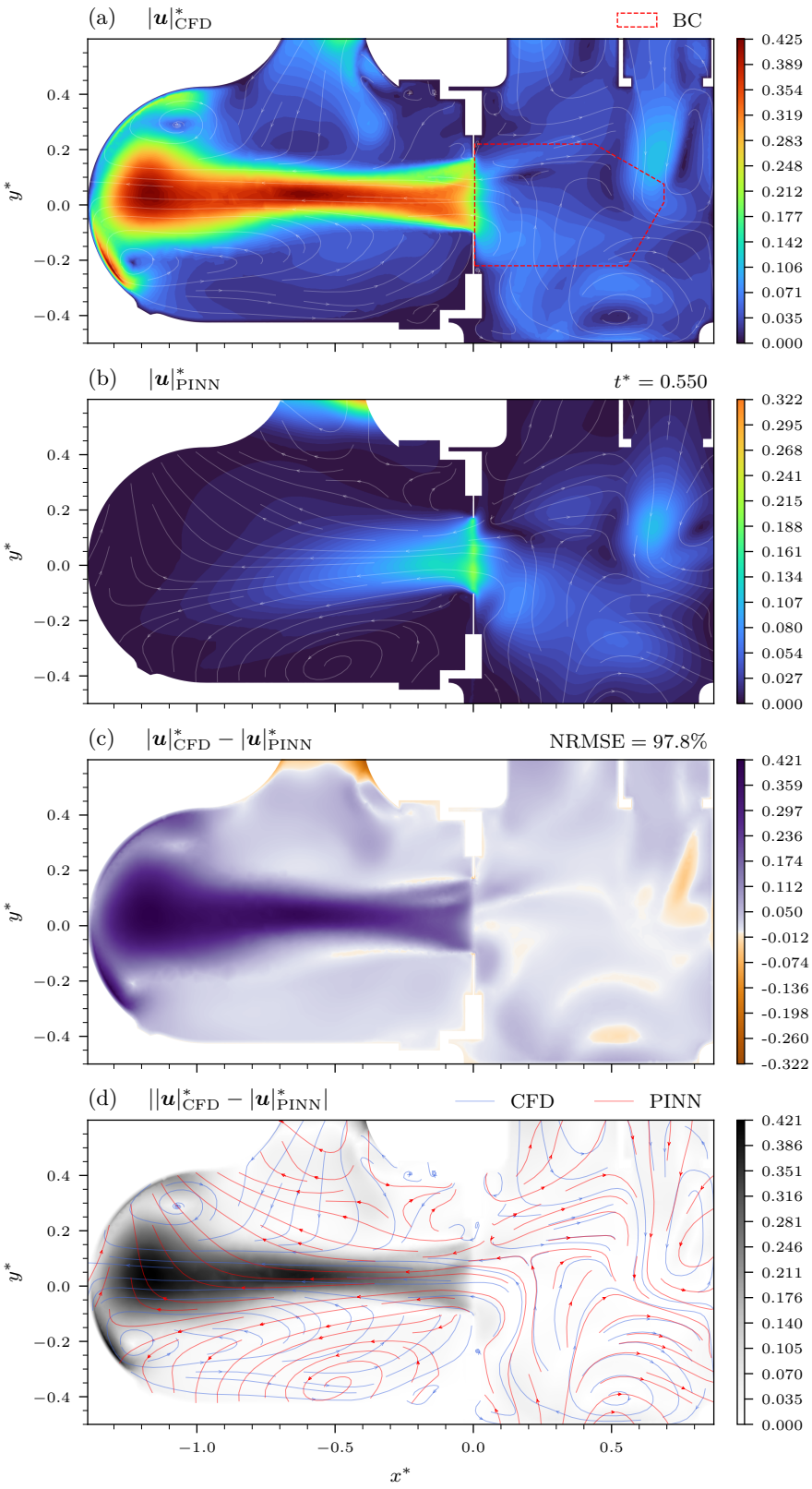

Figure C.33.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.55$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

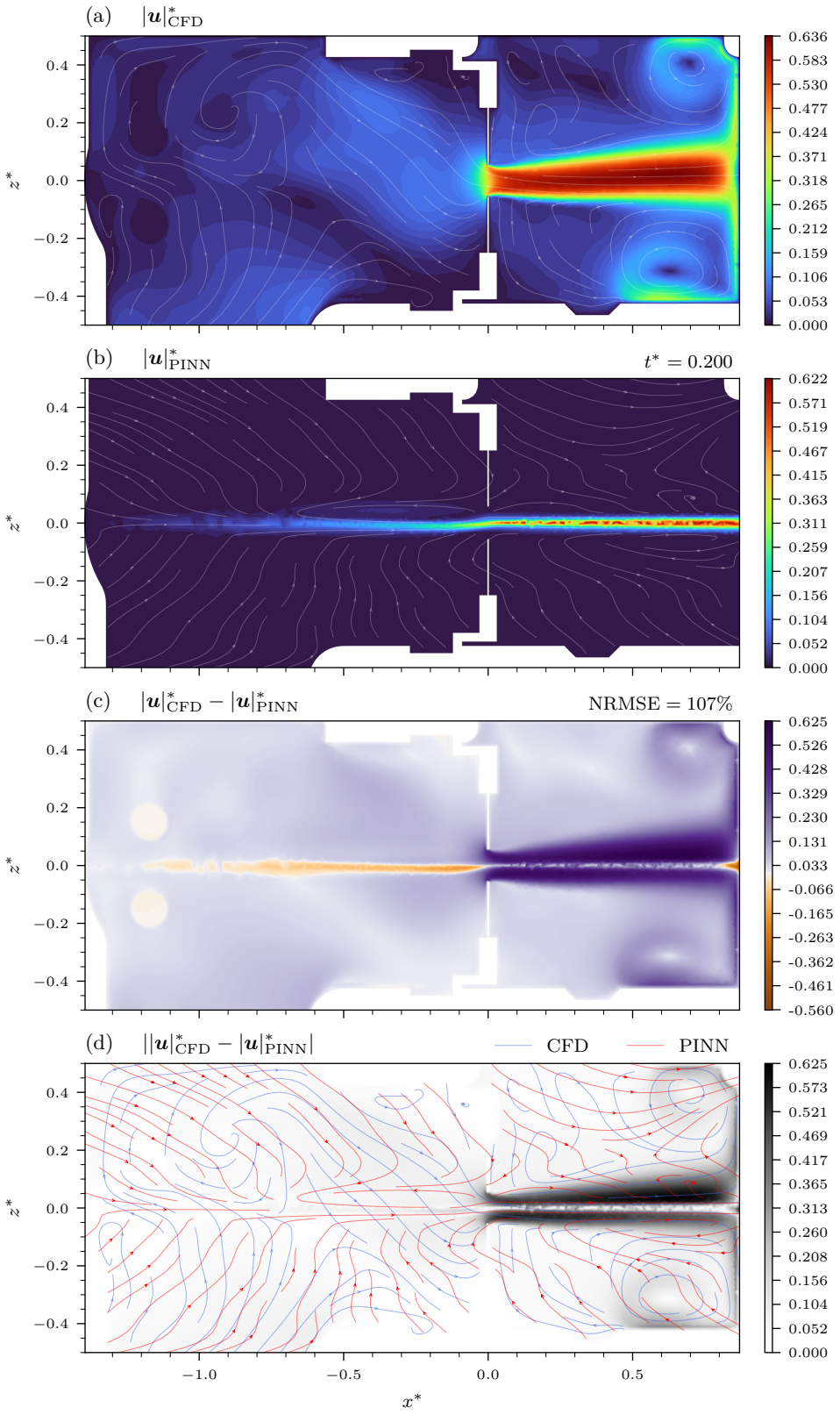

Figure C.34.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for y-plane at $t^* = 0.2$. (b) "PINN RANS" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

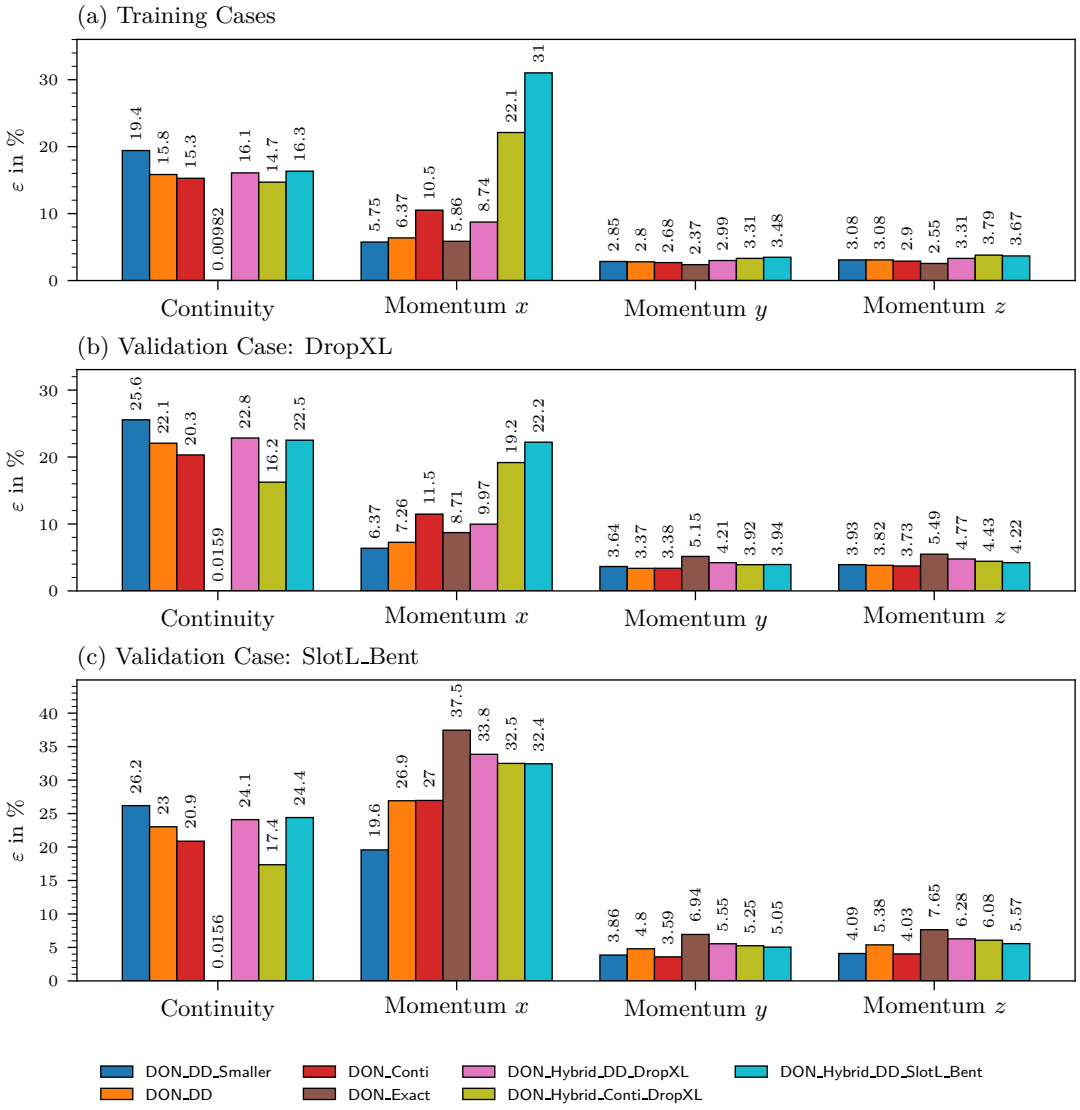

Figure C.35.: Comparison of 3D-Transient PDE residuals for the RANS equations across various trained DeepONets. The charts display residual percentages aggregated over the cardiac cycle for (a) training cases and the two (b, c) validation cases.

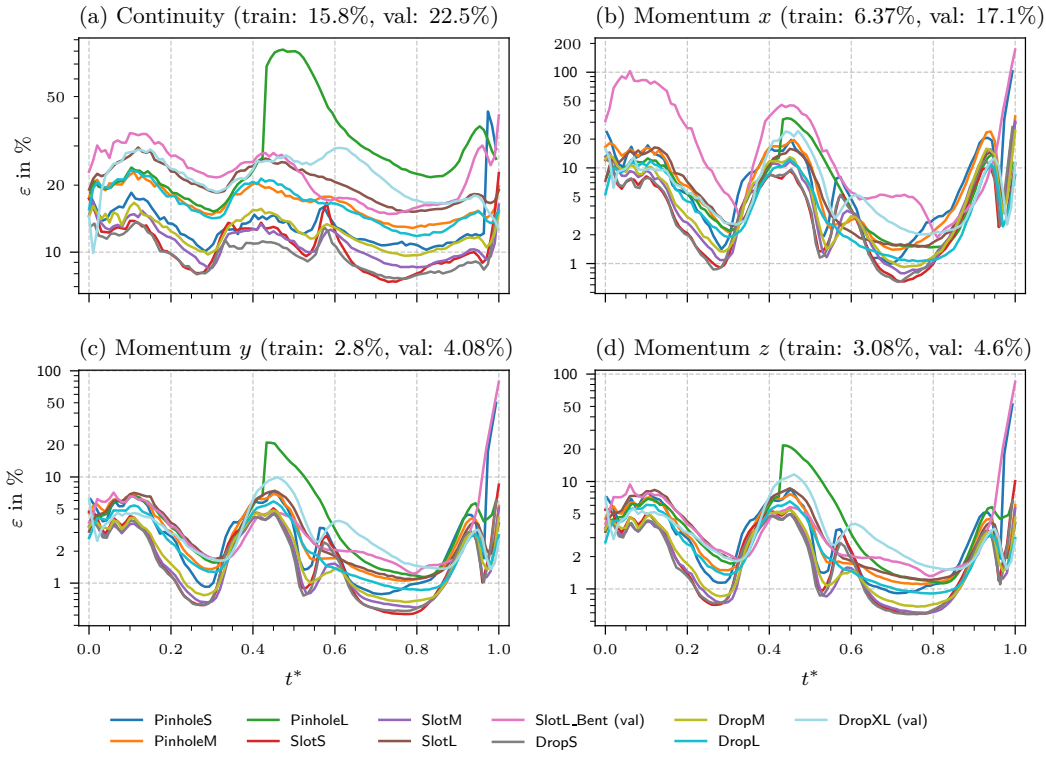

Figure C.36.: Temporal evolution of the PDE residuals for the RANS equations over the cardiac cycle for the 3D-Transient "DD" model. Each line is a different CFD case.

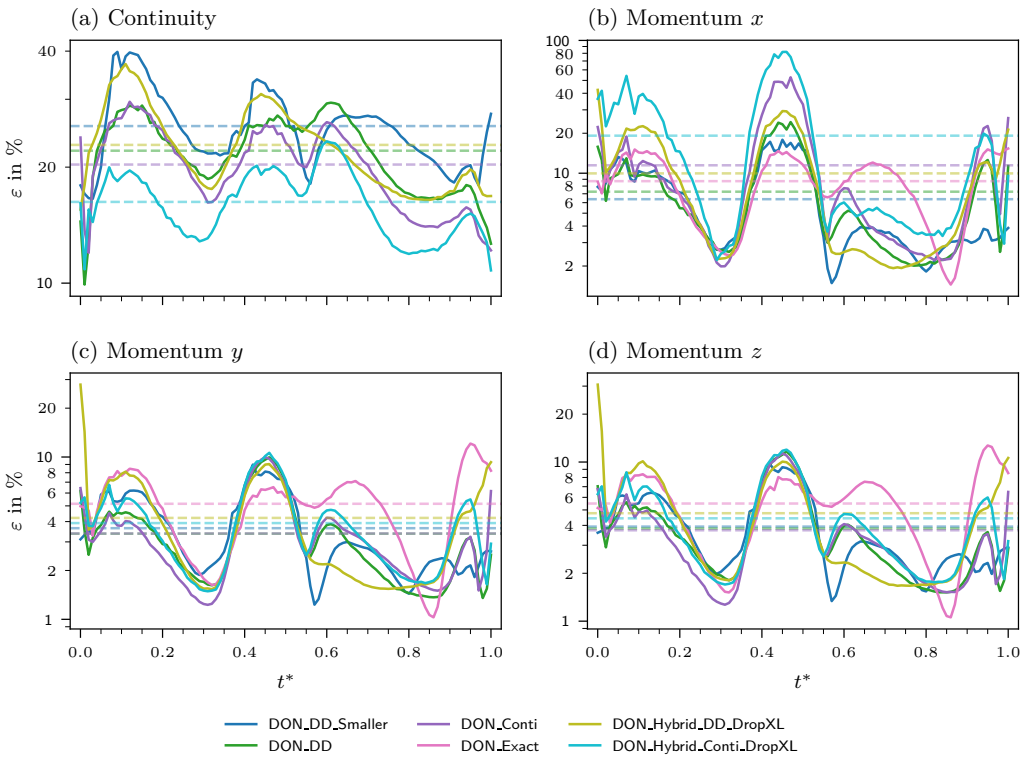

Figure C.37.: Temporal evolution of the PDE residuals for the RANS equations over the cardiac cycle for the 3D-Transient DropXL case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

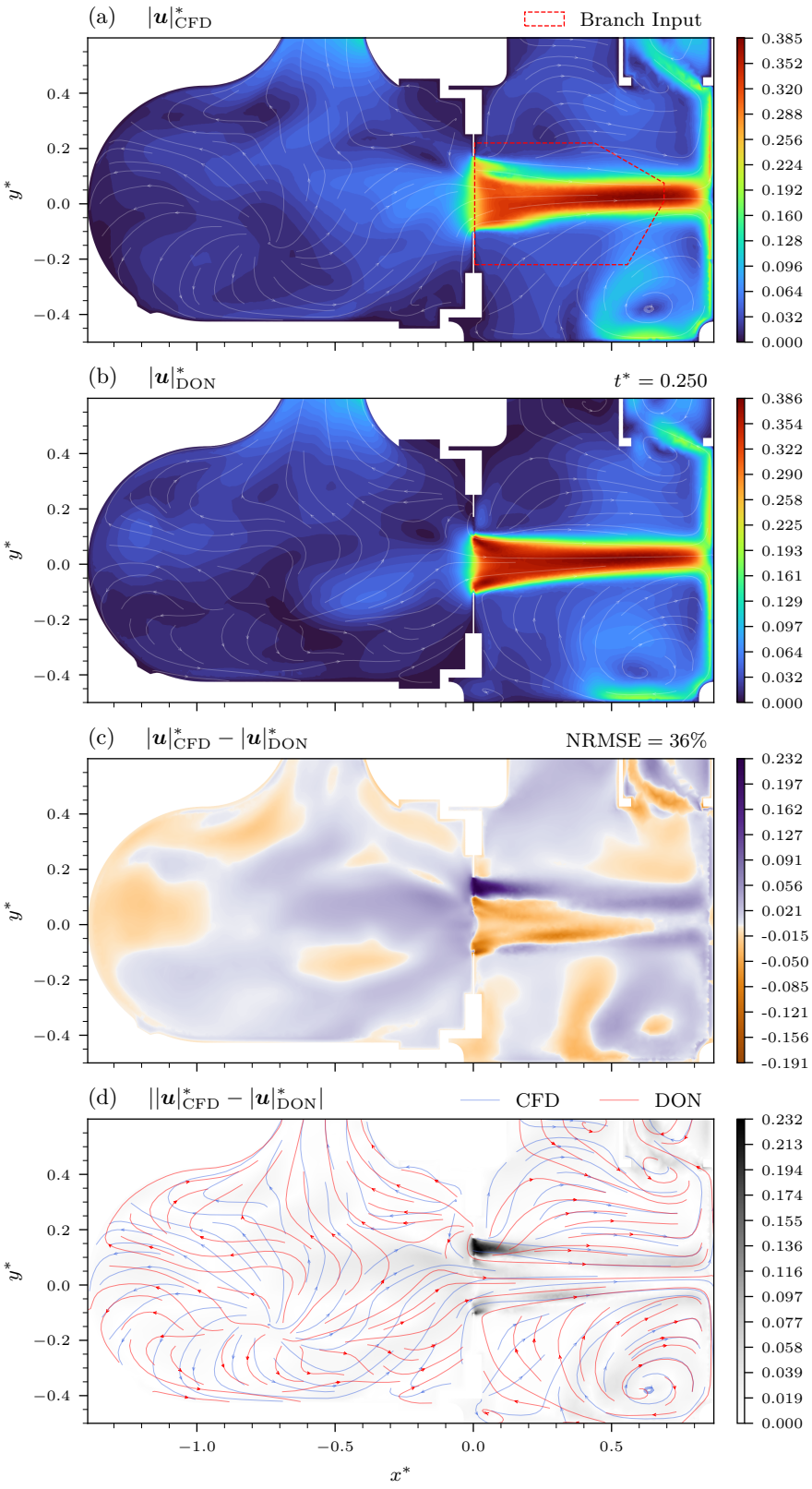

Figure C.38.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.25$. (b) "DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

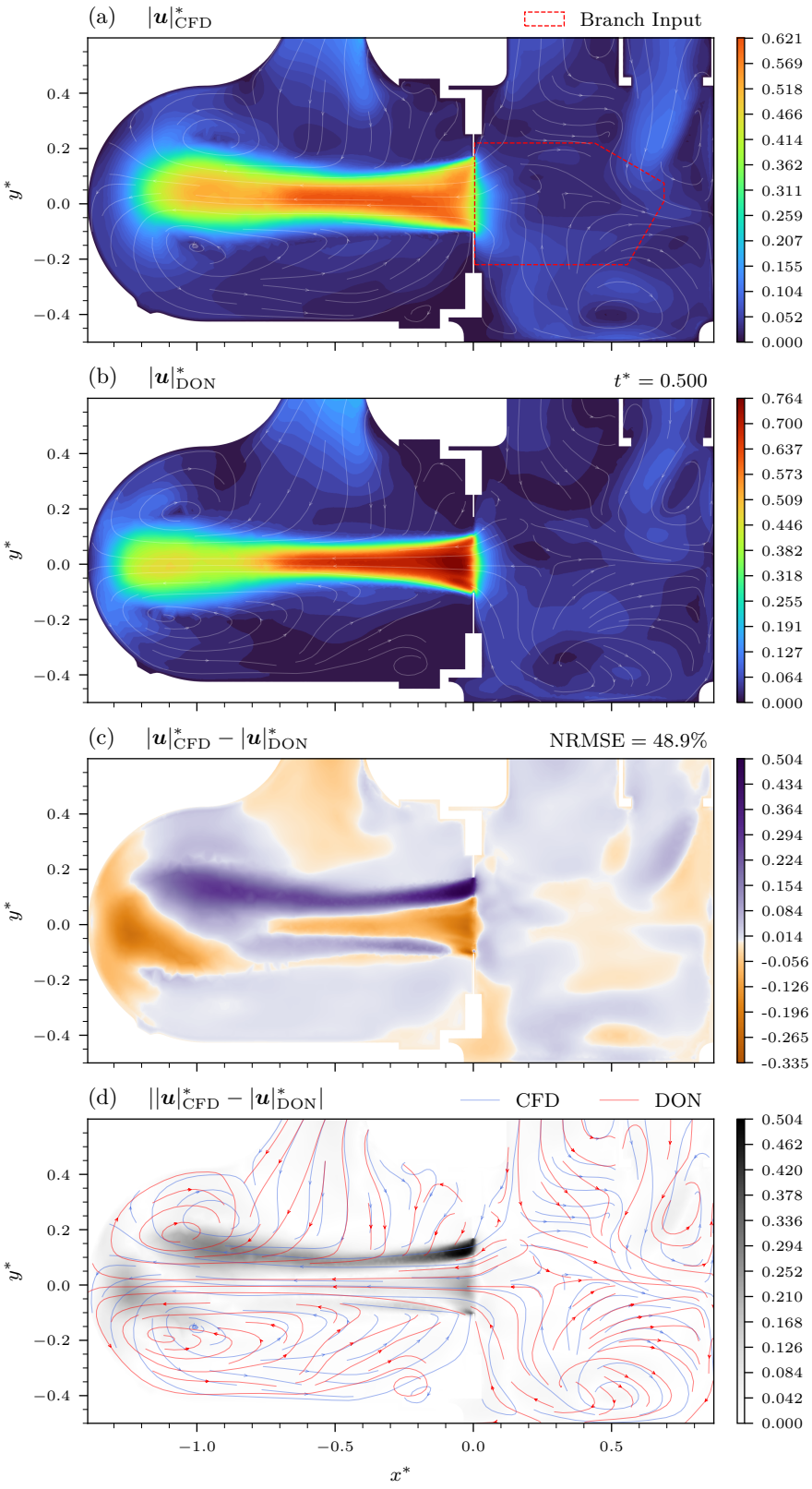

Figure C.39.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.5$. (b) "DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

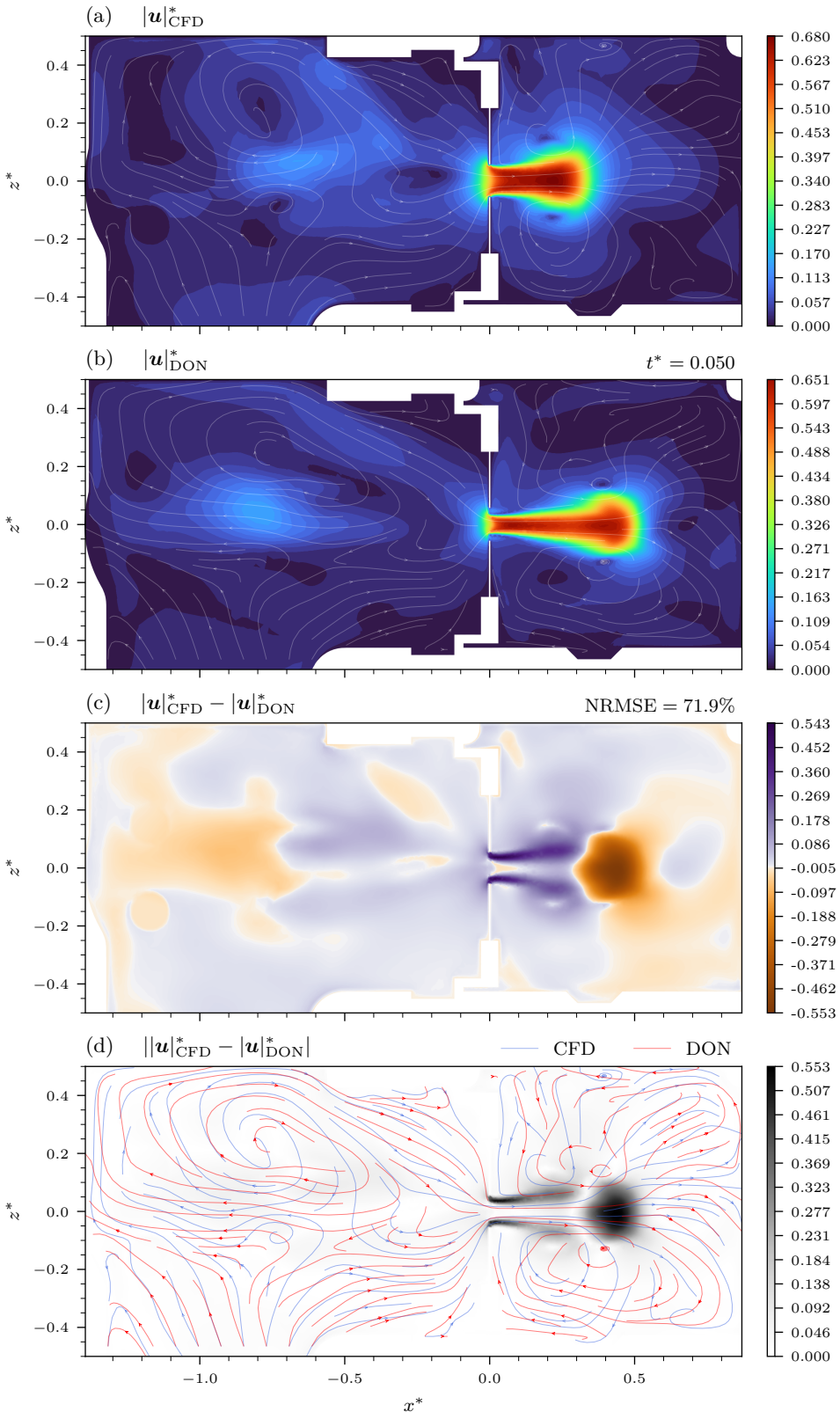

Figure C.40.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for y-plane at $t^* = 0.05$. (b) "DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

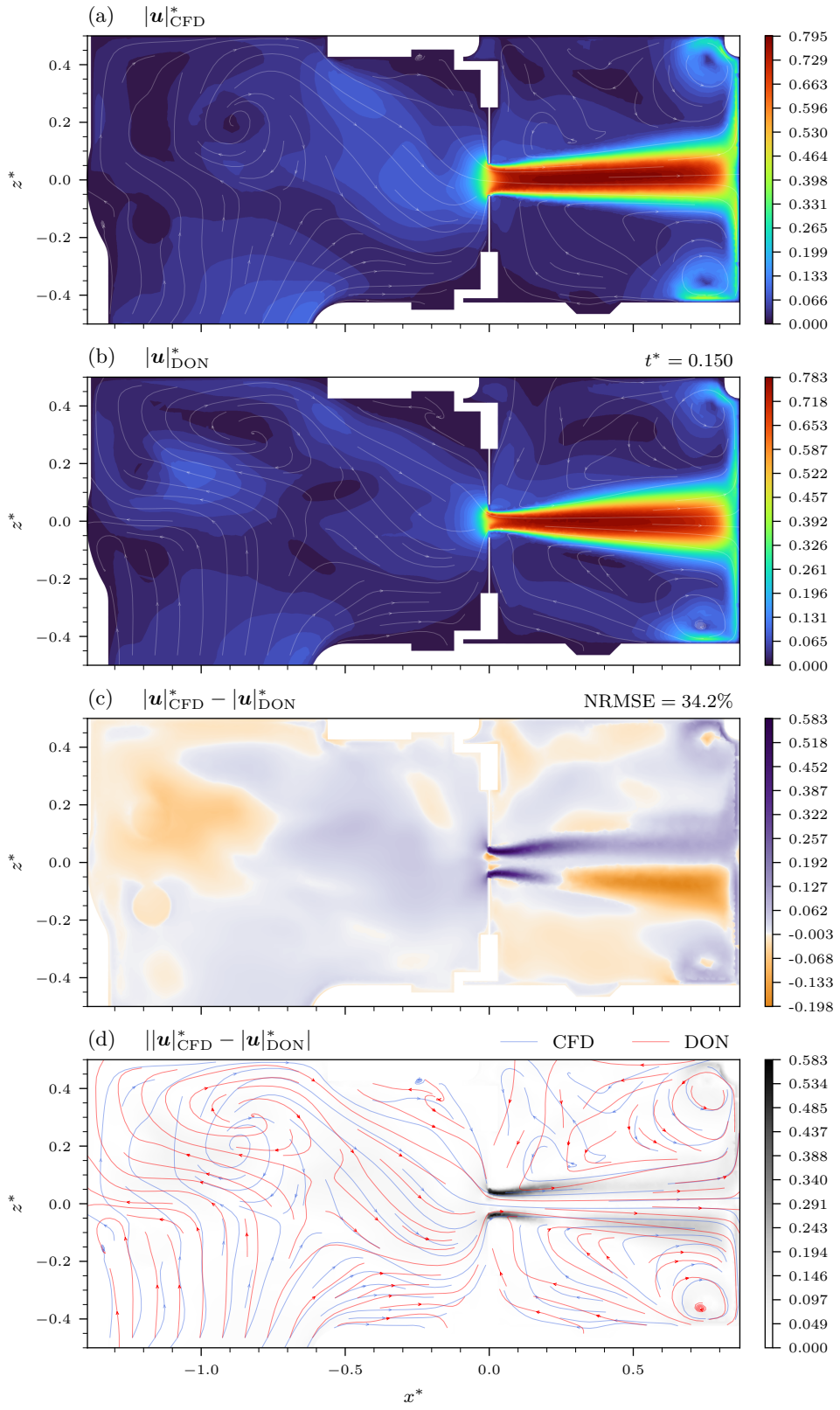

Figure C.41.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for y-plane at $t^* = 0.15$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

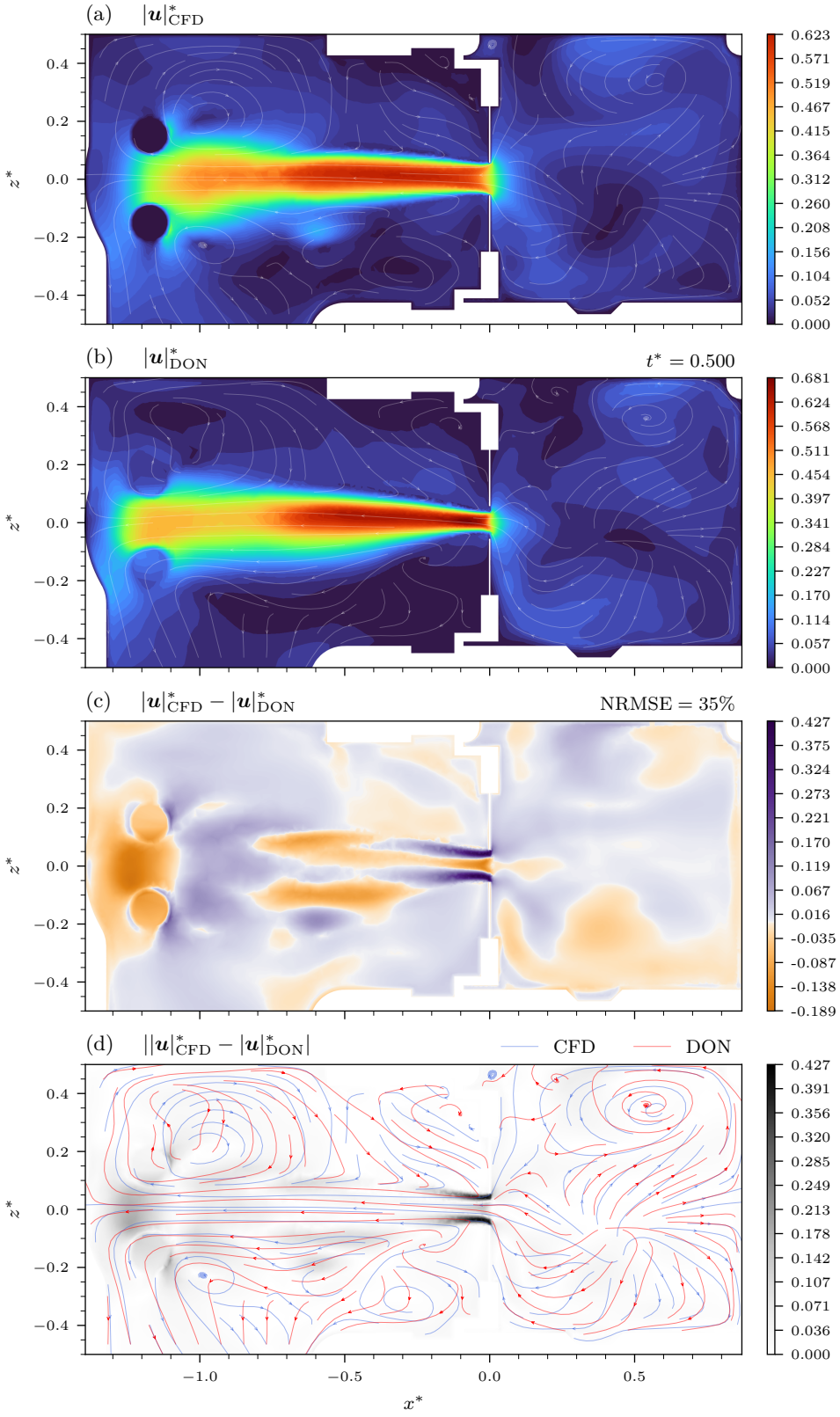

Figure C.42.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for y-plane at $t^* = 0.5$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

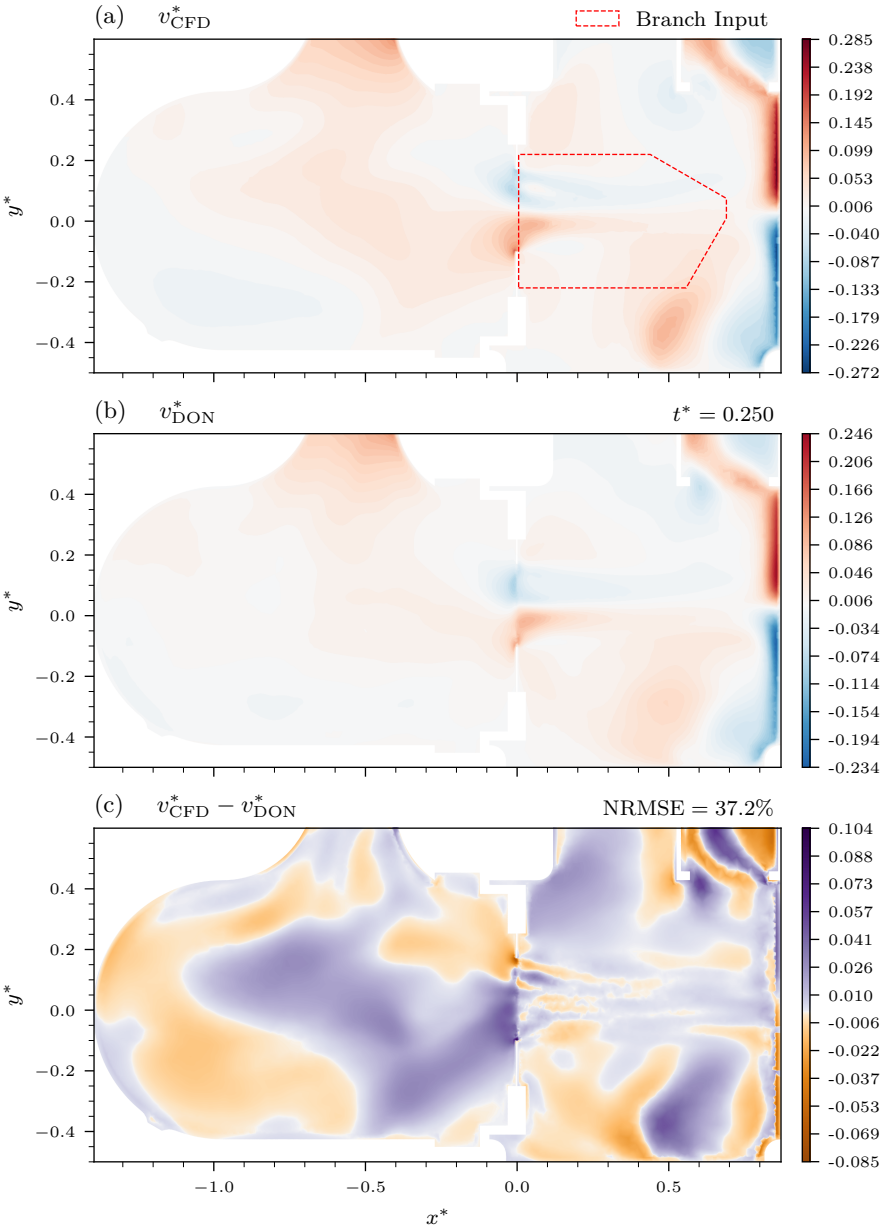

Figure C.43.: (a) Reference 3D-Transient DropXL CFD v-velocity for z-plane at $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

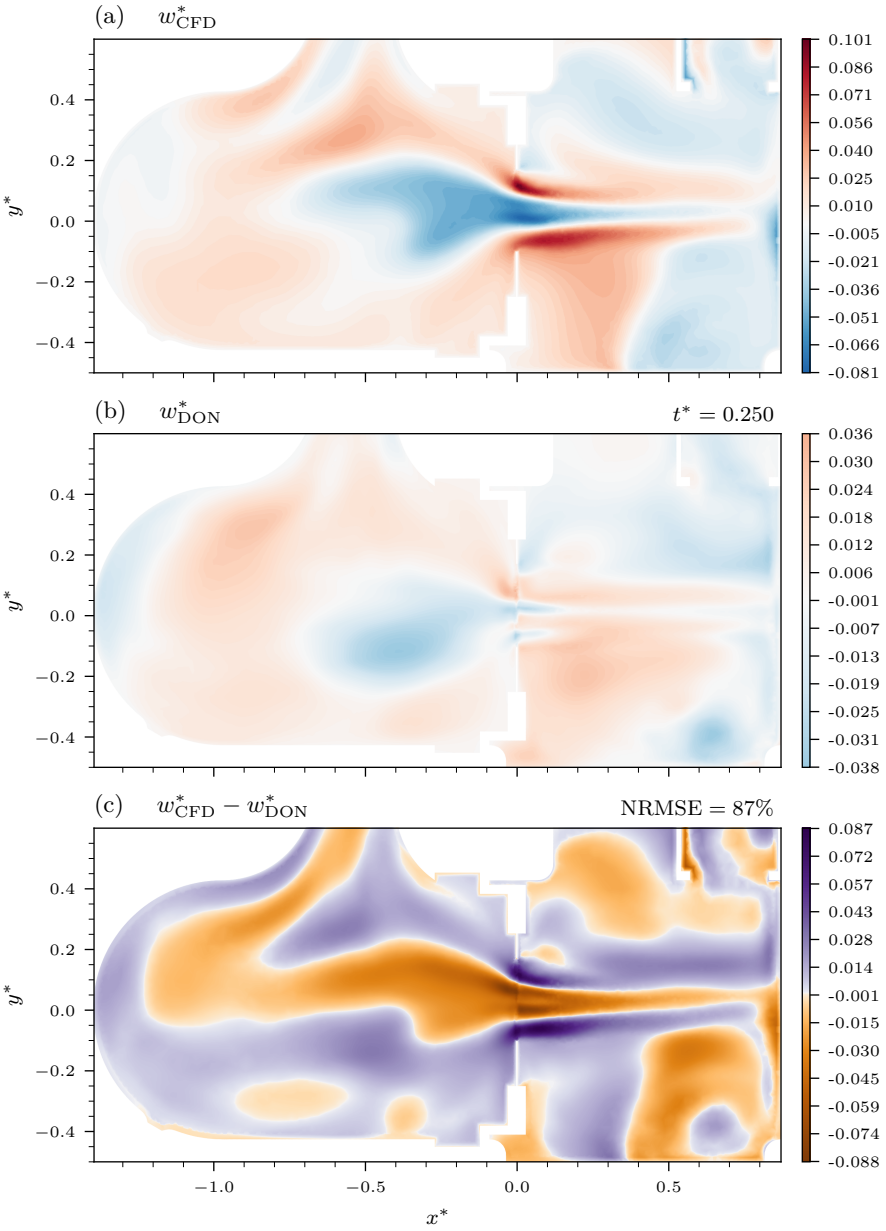

Figure C.44.: (a) Reference 3D-Transient DropXL CFD w-velocity for z-plane at $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

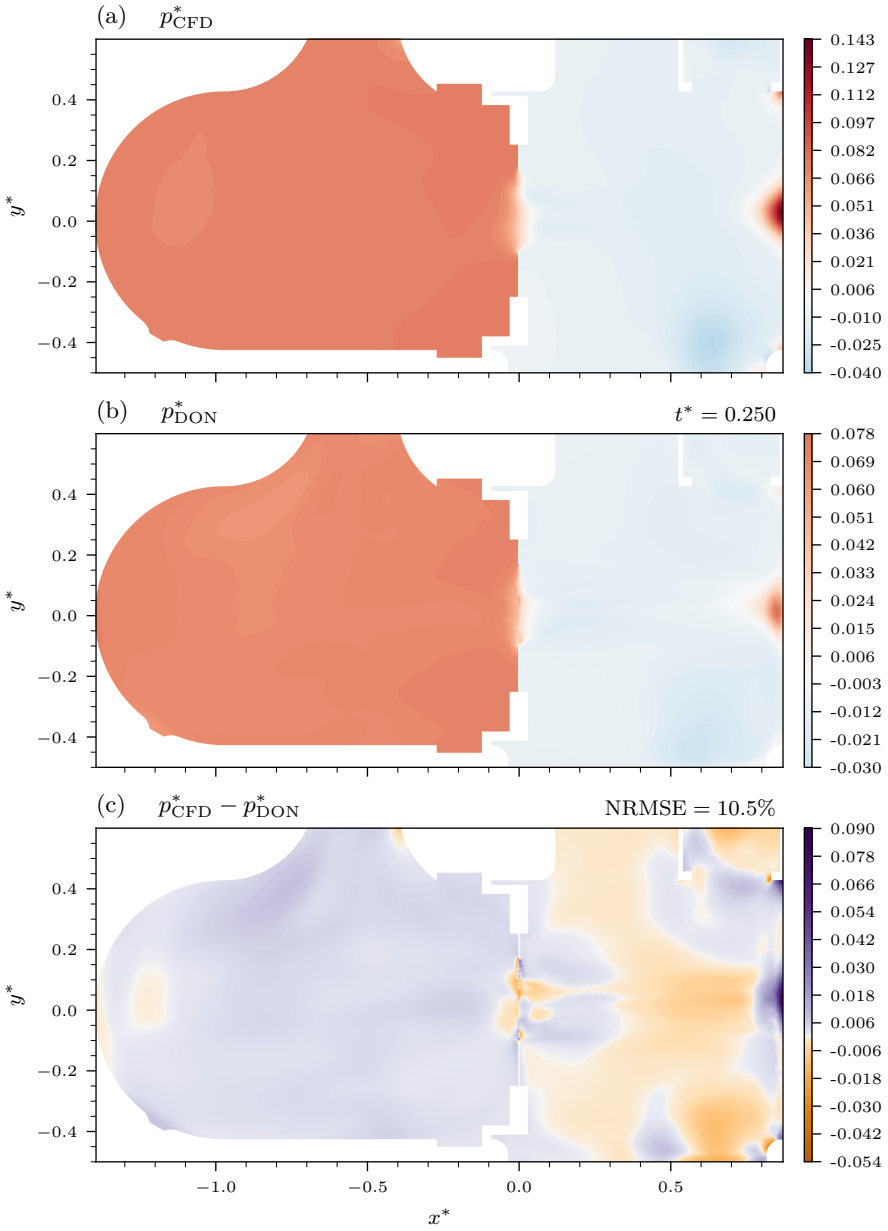

Figure C.45.: (a) Reference 3D-Transient DropXL CFD pressure for z-plane at $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

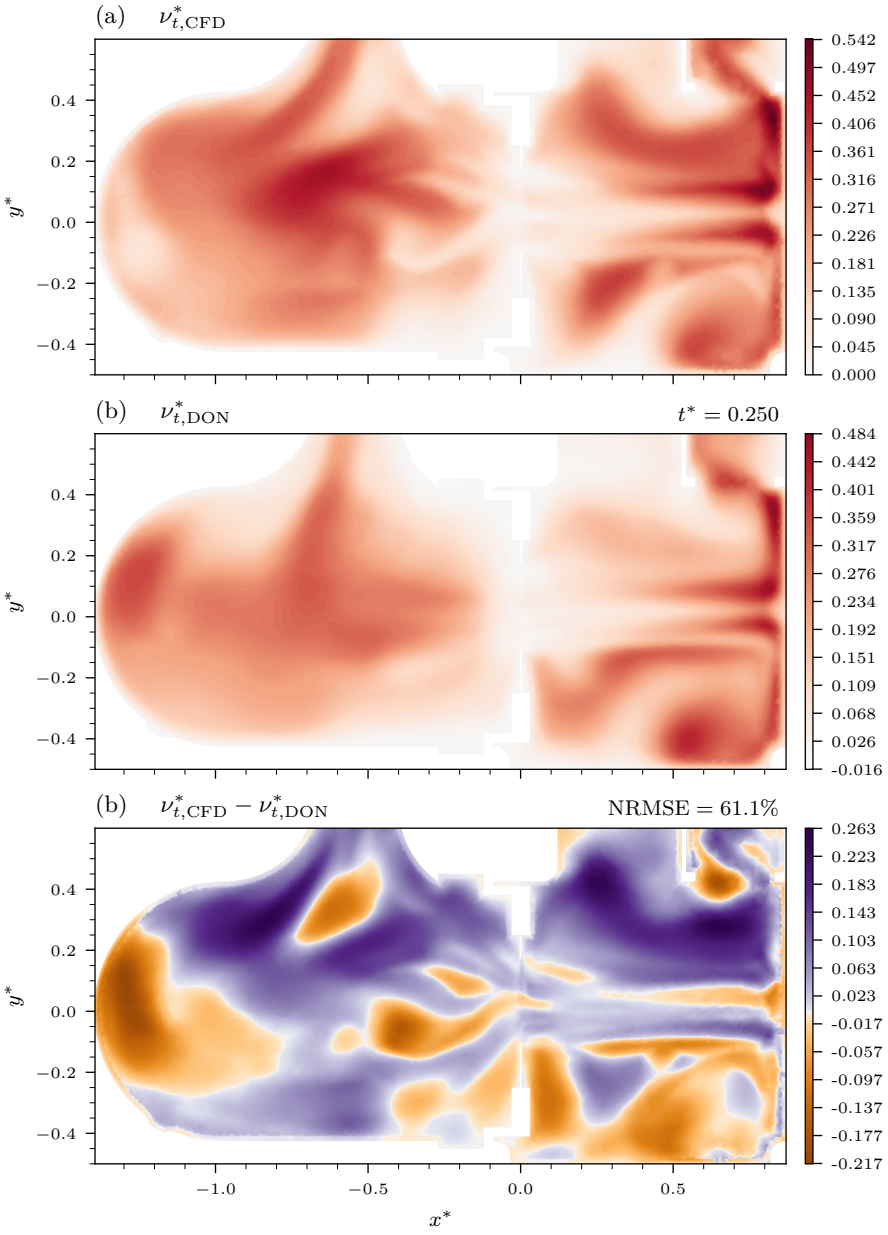

Figure C.46.: (a) Reference 3D-Transient DropXL CFD turbulent viscosity for z-plane at $t^* = 0.25$. (b) "Hybrid DD" prediction. (c) Signed non-linear error.

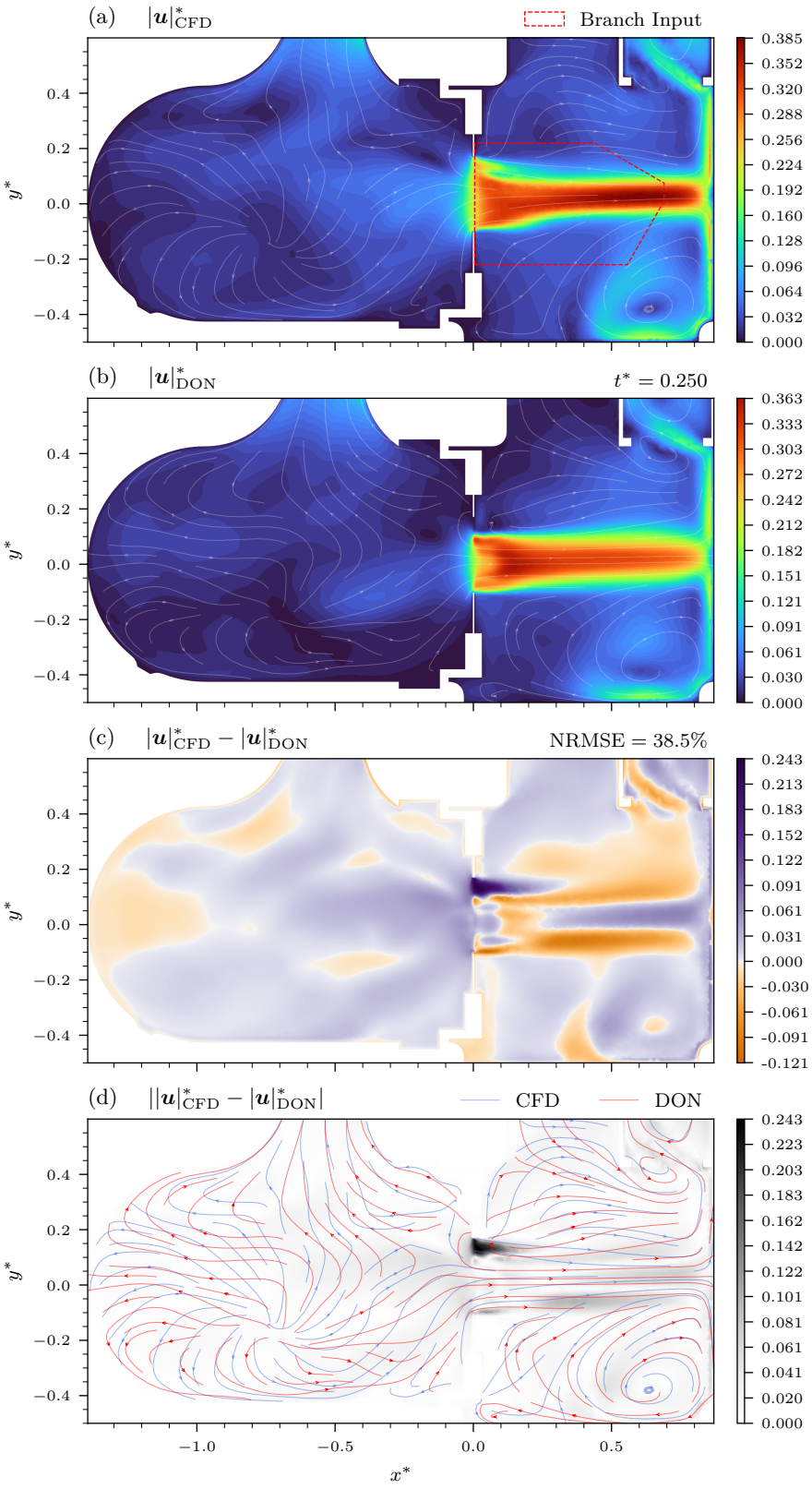

Figure C.47.: (a) Reference 3D-Transient DropXL CFD velocity magnitude for z-plane at $t^* = 0.25$. (b) "Hybrid Conti" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.

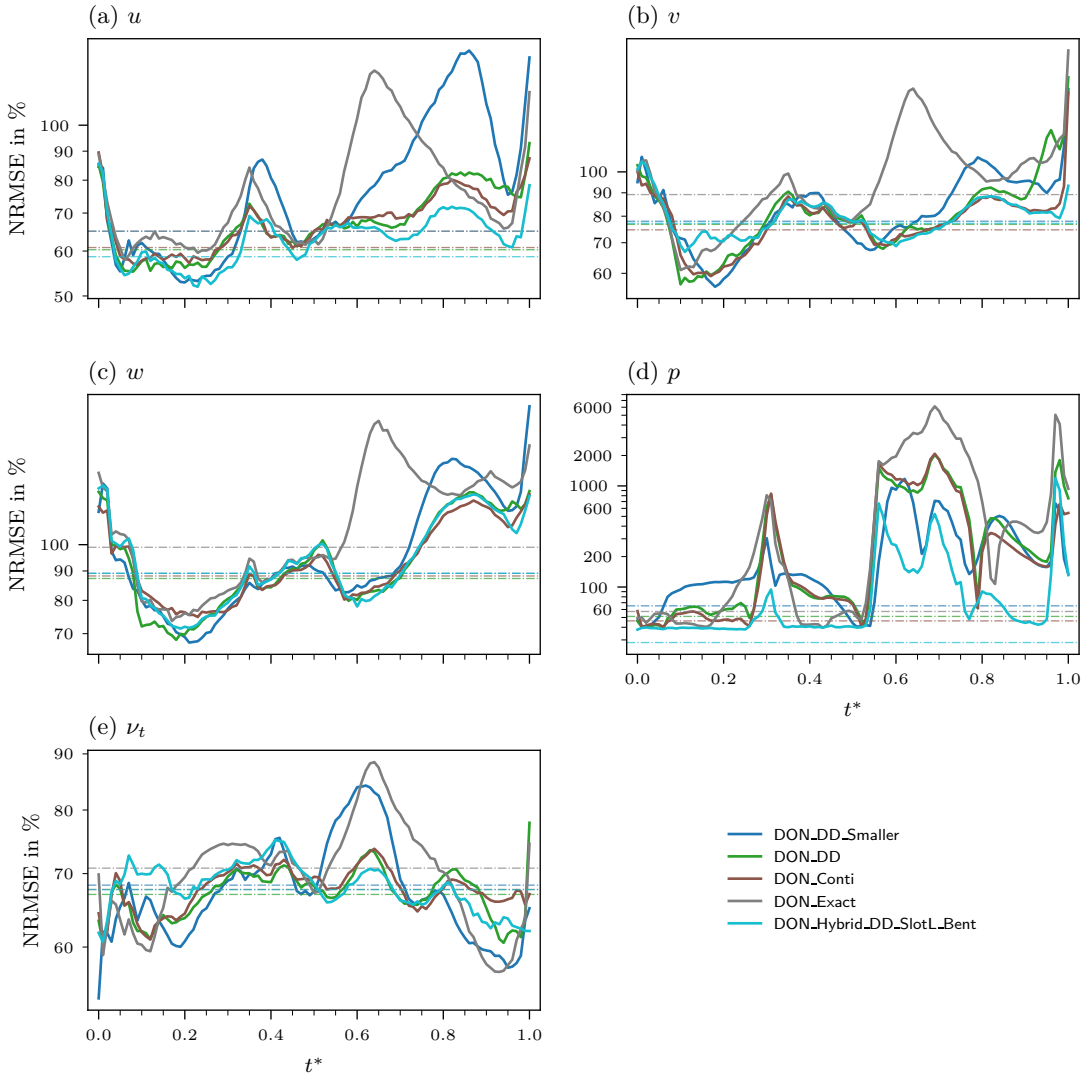

Figure C.48.: Temporal evolution of the NRMSE for all fields over the cardiac cycle for the 3D-Transient SlotL-Bent case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

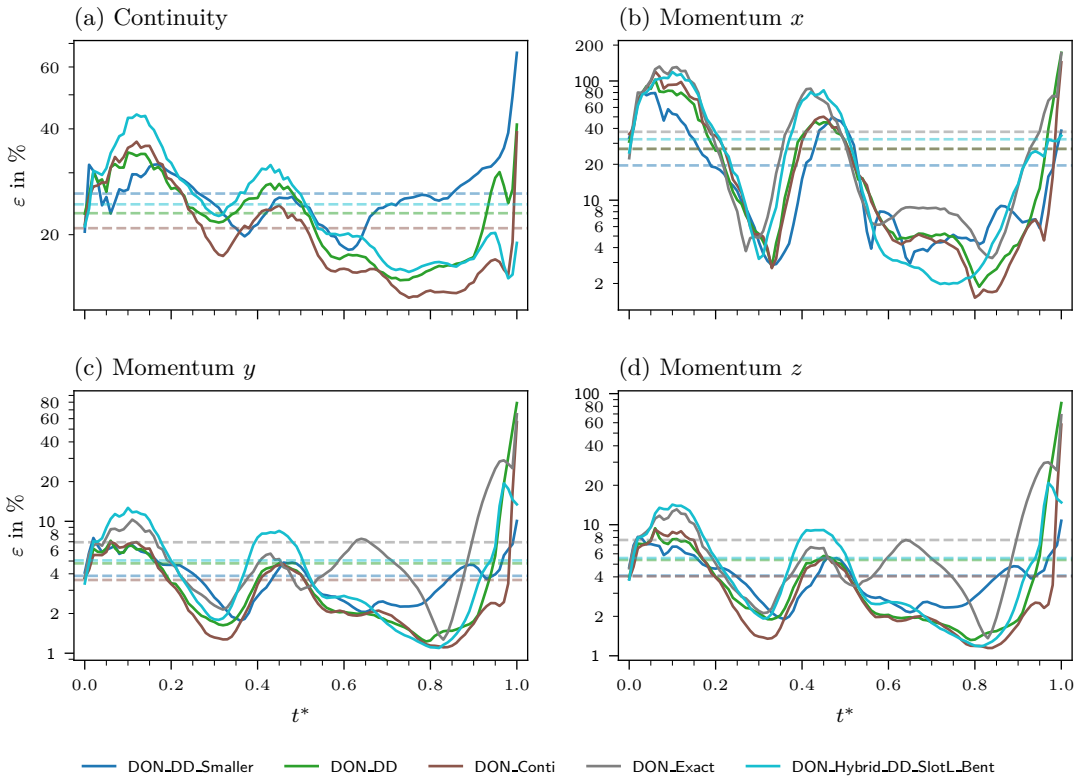

Figure C.49.: Temporal evolution of the PDE residuals for the RANS equations over the cardiac cycle for the 3D-Transient SlotL-Bent case. Horizontal dashed lines indicate the respective aggregate values computed over the entire period.

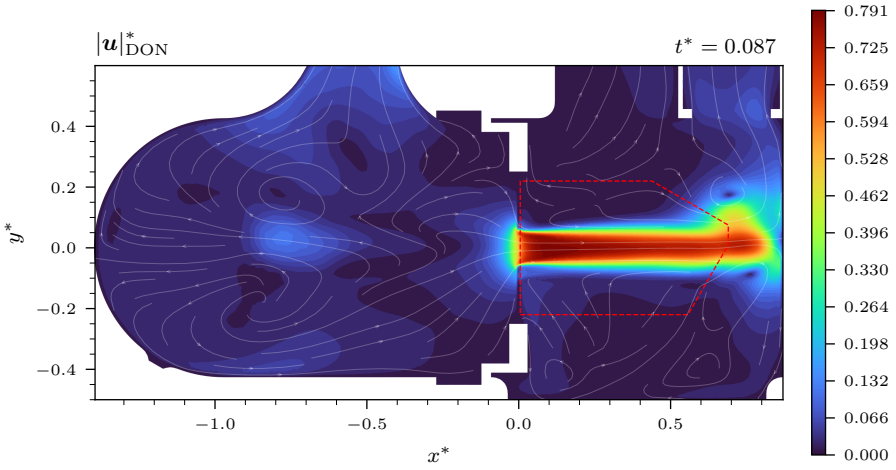

Figure C.50.: "DD" velocity magnitude for 3D-Transient PinholeL Exp, z-plane at $t^* = 0.087$

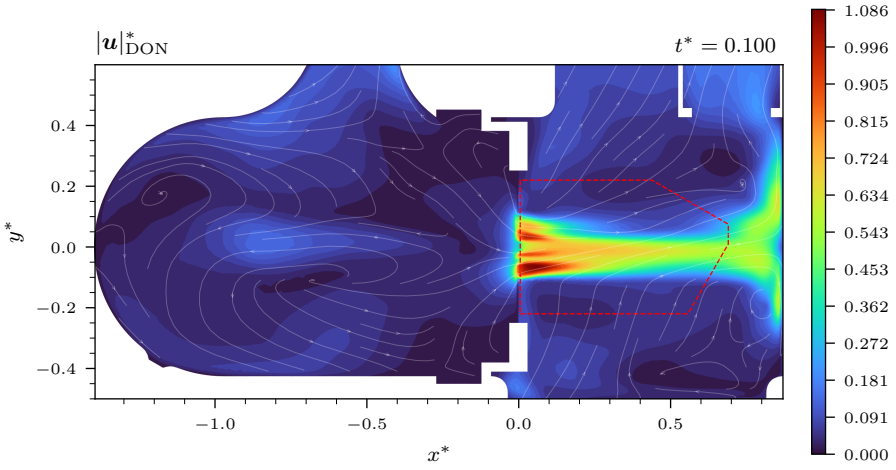

Figure C.51.: "DD" velocity magnitude for 3D-Transient EccJet Exp, z-plane at $t^* = 0.1$

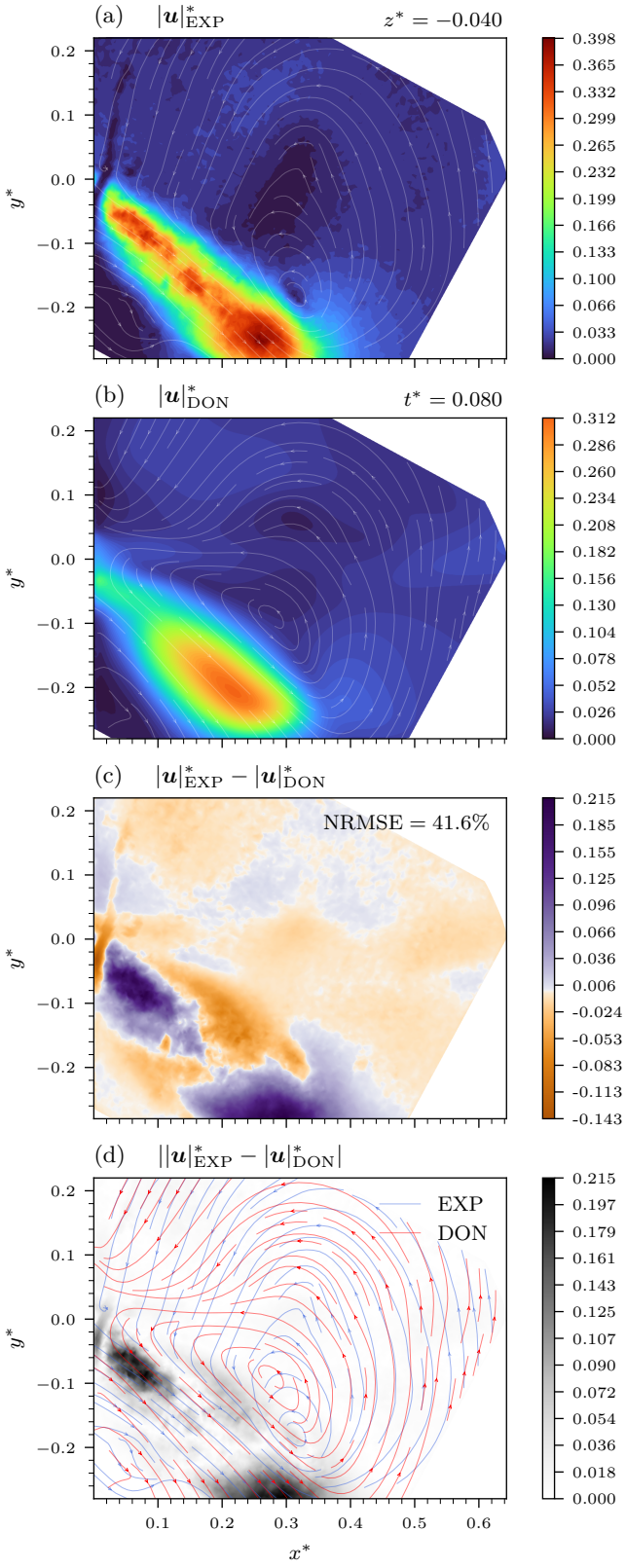

Figure C.52.: (a) Reference 3D-Transient EccJet Exp velocity magnitude for $z^* = -0.04$ at $t^* = 0.08$. (b) "Hybrid DD" prediction. (c) Signed non-linear and (d) absolute error magnitude with streamlines.