International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

COMPUTATIONAL FLUID DYNAMICS IN MECHANICAL ENGINEERING: A RESEARCH STUDY ON SIMULATION, OPTIMIZATION, AND REAL-WORLD APPLICATIONS

[1*]Dr. Sunil Kumar Srivastava, [2]Nayankumar Patel, [3]Sudhanshu Shrivastava, [4]Varadaraj K S, [5]Dr. R.Kalidasan, [6]Dr. Ketul Brahmbhatt, [7]Dr. Jairaja R

Associate professor, Teerthanker Mahaveer College of Pharmacy, Teerthanker Mahaveer University Moradabad, Email ID: Shyam.sunil32@gmail.com
Assistant Professor, Parul University Email ID: patel.nayan8888@gmail.com
Orcid ID: https://orcid.org/0009-0000-4125-8014
Assistant Professor , Parul University Email ID: sudhanshu0071@gmail.com Orcid ID: https://orcid.org/0000-0002-4358-120X
Sr assistant Professor , Department of Civil Engineering, Alvas Institute of Engineering and Technology , Water Resources Engineering and Management
Email ID: varadaraj@aiet.org.in
Associate Professor , School of Mechanical Engineering, Lovely Professional University Manufacturing Engineering ORCID ID: https://orcid.org/0000-0001-5366-9822 Email ID: rkalidasan2002@yahoo.com
Assistant Professor in Mechanical Engineering , BVM engineering college
Email ID: kbrahmbhatt@bvmengineering.ac.in
Associate Professor, Government Engineering College, Huvinahadagali - 583219, India
Email id: jairaja.r@gmail.com

Abstract

This study investigates how Computational Fluid Dynamics (CFD) can serve not only as a predictive tool but also as a framework for optimization and design in mechanical engineering. Using two open datasets—a geometry/operations-focused Computational Fluid Dynamics set (n=768) and the Stirred Reactor CFD ML-Dataset by novalabs (n=10,000)—we implemented a reproducible, Python-based workflow to clean data (0% missing), compute descriptive statistics, test normality, evaluate group differences, and quantify pairwise dependencies alongside targeted visuals (histograms, boxplots, heatmaps, and scatter plots). Results show heavy-tailed, non-Gaussian behaviour in key variables (e.g., dvdy, dudt), modest yet physically meaningful outlier rates, and clear structure: perfect geometric collinearity in the reactor dataset (e.g., reactor_diameter vs diameter_impeller_1) and physically consistent trends in the flow dataset (e.g., y–dudy negative shear dependence; x–p streamwise pressure drop). These patterns validate the simulations against classical fluid-mechanics expectations and indicate practical levers for design and control (e.g., feature selection to remove redundancies; robust or surrogate-based optimization targeting shear/pressure metrics). While the analysis is constrained by the use of simulated data and the absence of experimental benchmarking, the findings demonstrate that carefully curated CFD features can reduce computational burden without sacrificing fidelity and can guide optimization for reactors and related mechanical systems.

Keywords: computational fluid dynamics; turbulence; design optimization; stirred reactor; surrogate/adjoint methods

1. Introduction

Computational Fluid Dynamics (CFD) has emerged as an essential part of contemporary mechanical engineering, offering a numerical model to addressss the NavierStokes equations of a complicated fluid flow (Anderson, 2017; Versteeg, 2007). Early achievements in finite-volume and finite-difference techniques enabled engineers to solve in the past three-dimensional time-dependent flows that could not be solved before (Moukalled et al., 2015). In recent decades, the scope of the flows that can be predicted with predictive accuracy has grown dramatically due to the development of turbulence models, such as Reynolds-Averaged Navier Stokes (RANS), Large Eddy Simulation (LES), and hybrid methods (Pope, 2000; Sagaut, 2006; Davidson, 2018). As the computational power grows quickly and high-performance clusters are in place, CFD is now actively used in a wide variety of engineering problems, including turbomachinery and aerospace design as well as energy production and advanced HVAC systems (Ferziger et al., 2019; Rao et al., 2023; Zhou et al., 2019). The machine-learning methods also hasten these procession by making the computation cost less and the solution fidelity is maintained (Brunton et al., 2020).

Even with these developments, there are still a number of key challenges. Proper CFD models require a high level of spatial and temporal resolution, which makes them costly to compute and slow. Multiscale turbulence phenomena are yet to be captured with a careful choice and validation of models to prevent numerical diffusion and convergence problems (Pope, 2000; Sagaut, 2006). Additionally, the seamless integration of CFD in the industrial design cycles is impeded by the fact that it requires specialized knowledge in mesh creation, specification of boundary conditions, and solver setup (Ferziger et al., 2019). Such obstacles may slow innovation and restrict more extensive use in the industries that can use predictive fluid-flow analysis.

The area of the current study includes simulation, optimization, and practice of CFD in the field of mechanical engineering. It focuses on the use of finite-volume techniques in laminar and turbulent flows, the optimization of results with the help of machine-learning, and the comparison of results with experimental or benchmark data where feasible. The research does not seek to create new numerical solvers or turbulence models, but rather it tests and uses existing algorithms in realistic mechanical engineering situations. Natural constraints on the generalizability are the external factors of hardware limitations, proprietary industrial geometries and the necessity of experimental correlation. Although CFD is gaining applications in other fields, e.g., energy and climate modeling (Jedermann et al., 2014; Nacht et al., 2023), the literature has been focused on the applications of CFD in mechanical engineering.

The knowledge of the systematic inclusion of CFD in engineering design and optimization pipelines is of paramount importance to both the academia and industry. Through the use of high-resolution simulations together with optimization strategies, mechanical engineers will be able to minimize energy losses, enhance thermal control, and minimize product-development cycles. Another important point made in the study is the adaptation of emerging methods (discrete adjoint methods to design aerodynamic components) (Rao et al., 2023) and surrogate-based multi-objective optimization to mechanical components (heat exchangers, pumps, reactors, etc.) by the author. The knowledge acquired in this study can be reflected in best practices, minimize computational cost, and bridge the gap between academic research on CFD and industry applications. This research will help engineers and researchers operating in realistic engineering environments by exploring the potential of CFD in simulation, optimization, and application in real-life engineering situations.

The particular objectives of this study include:

- **Test simulation performance:** Test the behavior of developed CFD solvers in modeling realistic complex laminar and turbulent flows found in systems of mechanical engineering, on validated turbulence models, and test datasets.

- **Examine optimization models:** Test optimization methods, discrete adjoint methods and surrogate-based methods to design parameters that improve design performance: pressure drop, mixing efficiency, and thermal performance.
- **Show real-world use cases:** Illustrate how CFD can be used to offer practical solutions to real-world mechanical engineering applications, including refinement of aerodynamic solutions and intensification of processes in reactors..

2. Related Work

The research in the area of Computational Fluid Dynamics (CFD) is ever-growing in terms of the field of turbulence modeling, the development of numerical algorithms, optimization into the field, and various industrial uses of CFD. Initial advances in adaptive mesh refinement gave the base of good vortex-dominated flow simulations. Kamkar et al. (2011) proposed a feature-based Cartesian adaptive mesh refinement scheme showing that grids can be adapted carefully, focusing on the target structures to solve complex vortical structures at a reasonable computational cost. Their work emphasizes the fact that adaptive meshing is vital in order to resolve multiscale dynamics in high-Reynolds-number flows, without the use of prohibitive computational resources.

The latest developments in turbulence modeling have introduced data-driven and machine-learning-based approaches to enhance predictive accuracy. The article by Zhang, Zhang, and Jiang (2023) features a review of the issues and opportunities of turbulence-modeling and focuses on the comparison of the performance of machine-learning-based models to the traditional Reynolds-Averaged NavierStokes (RANS) and Large Eddy Simulation (LES) methods. The same author (Duraisamy, 2021) offers an opinion on machine-learning-enhanced RANS and LES turbulence models, noting that they can help to decrease the level of uncertainty in the models, but at the cost of markets with strong validity and physical insight.

Another essential theme of CFD-related research has also become the optimization of turbomachinery and thermal systems. Cao (2024) was dedicated to the optimization of aerothermal of low-pressure steam turbine exhaust systems, which revealed how the design of the exhaust system with the help of CFD can reduce the number of energy losses and enhance the thermal efficiency of large-scale power generation. In a related domain, Elmenshawy (n.d.) investigated performance improvements of gas turbine engines through optimized blade-cooling channel design, revealing significant gains in heat transfer and efficiency achievable through simulation-driven geometrical refinement. Franco, Celis, and da Silva (2023) advanced this area by applying hybrid RANS/LES turbulence models to bluff-body flows, enabling more accurate predictions of unsteady separated flows, which are critical in turbomachinery and aerospace applications.

Beyond classical power systems, researchers have applied CFD to emerging manufacturing and energy-production processes. Wahlquist and Ali (2024) critically reviewed the roles of modeling and artificial intelligence in detecting defects in laser powder bed fusion (LPBF) metal printing. Their analysis highlights how CFD-based thermal-fluid simulations can be coupled with machine learning to predict and mitigate defects in additive manufacturing, extending CFD's relevance to cutting-edge industrial fabrication. Similarly, Zhang, Wang et al. (2024) reviewed laser cutting of titanium alloys, detailing how CFD-informed thermal and fluid-flow analyses contribute to understanding microstructural evolution and mechanical properties in advanced manufacturing.

CFD also supports sustainable power production and environmental engineering. Ceschin (2023) modeled complex turbulent and multiphase flows for sustainable energy applications, demonstrating that accurate turbulence and multiphase modeling is critical for designing next-generation renewable energy systems. Wang et al. (2025) explored multiscale interaction mechanisms in wind–leaf–droplet systems for orchard spraying, using CFD to analyze droplet dispersion and deposition under varying

Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

wind conditions. Their findings underscore CFD's value in precision agriculture and in optimizing environmental interventions to reduce chemical use and improve application efficiency.
Taken altogether, these papers indicate that the modern CFD research is distinguished by three overlapping trends. To begin with, there is a trend to data-driven turbulence modeling, which has been noted in works by Zhang, Zhang, and Jiang (2023) and Duraisamy (2021) who integrate the physics-based solvers with machine-learning algorithms to enhance the quality of prediction. Second, optimization based on application is growing, and the literature on the topic by Cao (2024), Elmenshawy (n.d.), and Franco et al. (2023) has shown that CFD can be used to optimize complex turbomachinery and thermal components. Third, CFD is becoming interdisciplinary, which is reflected in the studies of additive manufacturing (Wahlquist & Ali, 2024; Zhang, Wang et al., 2024) and sustainable energy systems (Ceschin, 2023; Wang et al., 2025). Such varied applications describe the dual purpose of CFD as a predictive application and an innovator in the areas of mechanical engineering and beyond.
This literature is informative to the current research as it defined the state of the art in the integration of simulation, optimization, and machine-learning in CFD. The present study integrates adaptive meshing, turbulence modelling, and application-specific optimization and aims to improve upon established practices whilst foreseeing inherent issues of computational cost and generalizability of the model. The experiences of these previous studies offer a methodological as well as industrial background that will be required in the further applications of CFD in mechanical engineering so as to achieve the process of simulation, optimization and application.

3. Methodology

The research design of this investigation was a quantitative, computational study, in which the high-fidelity computational fluid dynamics (CFD) data can be used to guide the simulation, optimization, and practical mechanical engineering implementation.
The methodology integrates established principles of fluid mechanics with modern data-science techniques to ensure both theoretical rigor and empirical reproducibility.

3.1 Research Design

The study was conceived as an observational-analytical inquiry in which numerical simulation outputs serve as the empirical substrate.
Rather than conducting new physical experiments, the research capitalized on validated CFD simulations, treating their outputs as the "population" of flow states.
This design enables the systematic characterization of geometric and kinematic variables while preserving the controlled conditions inherent to high-resolution CFD modeling.
By coupling descriptive statistics with inferential testing and by grounding all interpretations in established hydrodynamic theory-such as Reynolds scaling, Navier-Stokes momentum conservation, and boundary-layer dynamics—the design bridges computational modeling and physical reality.

3.2 Data Sources and Collection

Two open-access datasets formed the empirical foundation of the analysis.
The first, titled Computational Fluid Dynamics, contains 768 complete records describing the geometry and operating parameters of a stirred reactor.
Variables include impeller diameters, axial positions, rotational speeds, and pitch angles, all of which influence turbulence intensity and mixing efficiency.
The second dataset, the Stirred Reactor CFD ML-Dataset by novalabs, comprises 10,000 records of finely resolved flow-field measurements from high-resolution CFD simulations.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

It includes velocity components (u, v), spatial gradients (e.g., $\partial u / \partial y, \partial v / \partial y$), temporal derivatives ($\partial u / \partial t, \partial v / \partial t$), and static pressure distributions.
Both datasets were downloaded directly from the Kaggle repository in their original comma-separated value format.
Checksum verification ensured file integrity, and the raw archives were retained in read-only mode to preserve an immutable record of the source material.

3.3 Population and Sampling

Although derived from numerical simulations rather than physical experiments, the data still represent a well-defined population of potential reactor flow states.
For the Computational Fluid Dynamics dataset, the population comprises all simulated operating conditions defined by combinations of reactor geometry and impeller configuration. The study utilized every available record, effectively conducting a census of the dataset's 768 observations. Similarly, the Stirred Reactor dataset encompasses the full set of 10,000 simulated snapshots of the velocity field, again analyzed in its entirety.
Because no subsampling or case selection was performed, the analysis is free from sampling bias and directly reflects the full variability captured in the original CFD simulations.

3.4 Data Preparation

Rigorous preprocessing was undertaken to guarantee analytic reliability. Each file was screened for encoding anomalies and parsed using a reproducible Python workflow. Numeric fields were explicitly cast to floating-point types, and potential datetime-like fields were evaluated for proper parsing.
Both datasets exhibited zero percent missing values, obviating the need for imputation. Outlier detection was performed using standardized z-scores; approximately 0.73% of the numeric cells in the Computational Fluid Dynamics dataset and 1.05 % in the Stirred Reactor dataset exceeded $|z| > 3$.
These values were retained because, in turbulent flow contexts, extreme gradients and pressure spikes are genuine physical phenomena rather than data-entry errors.
All preprocessing steps were version-controlled to ensure full reproducibility.

3.5 Data-Analysis Techniques

Statistical analysis proceeded in three stages.
First, comprehensive descriptive statistics-means, medians, interquartile ranges, standard deviations, skewness, and kurtosis-were computed to characterize central tendencies and dispersion while revealing heavy-tailed distributions typical of turbulence.
Second, inferential tests were applied to assess normality and group differences. D'Agostino's K^2 test consistently rejected the hypothesis of normality for the majority of high-variance variables, confirming the intermittent, non-Gaussian nature of turbulent flows. Where categorical comparisons were appropriate, Welch's t-tests and, when multiple groups were present, Kruskal-Wallis tests were employed.
For example, the Stirred Reactor dataset includes a categorical factor denoting flow regime, and Welch's ttest demonstrated a highly significant difference in pressure between regimes ($t = -94.74, p < 0.001$).
Third, pairwise dependencies among variables were examined using Pearson correlation coefficients with associated p-values to evaluate statistical significance.

This revealed, for instance, perfect geometric collinearity between reactor_diameter and diameter_impeller_1 in the Computational Fluid Dynamics dataset and a strong negative correlation between vertical position y and horizontal shear gradient $\partial u / \partial y$ in the Stirred Reactor dataset.
Visual diagnostics accompanied these computations: high-resolution histograms and boxplots illustrated skewness and kurtosis; correlation heatmaps highlighted blocks of multicollinearity; and scatter plots of the strongest variable pairs provided direct confirmation of linear or monotonic relationships.
All analyses were conducted in Python 3.11 using the pandas, numpy, scipy, and matplotlib libraries on a Linux workstation with a multi-core AMD processor and 64 GB of RAM, ensuring efficient handling of the larger reactor dataset and reproducible computational environments.

3.6 Ethical Considerations

Although no human or animal subjects were involved, ethical standards for data use and scientific integrity were strictly observed.
Both datasets are publicly licensed on Kaggle, and all creators are cited to respect intellectual property. Raw data were never altered; every transformation was documented in executable scripts and retained alongside versioned output.
The decision to retain statistically defined outliers was made to avoid misrepresenting the inherent variability of turbulent flows, aligning with the ethical principle of accurately portraying physical phenomena.
The study conforms to the FAIR principles of data stewardship-Findable, Accessible, Interoperable, and Reusable-thereby supporting transparency and enabling independent replication.

3.7 Reliability and Validity

Reliability was reinforced through automated, script-based data cleaning and fixed random seeds, guaranteeing that identical results can be reproduced on demand.
Construct validity was maintained by ensuring that each numeric field corresponds to a well-defined physical quantity in fluid dynamics.
Internal validity benefited from the complete inclusion of all records, which eliminates selection bias and strengthens the robustness of statistical inference.
In summary, this methodology combines the precision of computational fluid-dynamics simulations with a rigorously documented statistical workflow.
By analyzing two large, openly available datasets in their entirety, employing robust statistical and visualization techniques, and adhering to strict ethical and reproducibility standards, the study provides a sound empirical foundation for investigating simulation, optimization, and real-world applications of CFD in mechanical engineering.

4. Results

4.1 Overview of Datasets and Data Quality

Two independent yet complementary datasets were subjected to comprehensive exploratory and inferential analysis:

- **Computational Fluid Dynamics** – 768 records and 24 numeric variables capturing geometric, operational, and kinematic properties of a stirred reactor configuration.
- **Stirred Reactor CFD ML-Dataset by novalabs** – 10,000 records with 14 variables (13 numeric, 1 categorical) describing fine-scale velocity gradients and temporal derivatives from high-resolution CFD simulations.

Both datasets were pristine: 0 % missing values across all fields. Outlier detection using a $|z| > 3$ criterion identified 135 outliers (0.73 %) in the Computational Fluid Dynamics data and 1,369 outliers

International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(1.05 %) in the Stirred Reactor set. In fluid mechanics, such heavy-tailed excursions are physically plausible, reflecting rare but significant turbulent eddies or abrupt pressure fluctuations rather than data-entry errors.

Normality tests (D'Agostino K^2) confirmed that the majority of high-variance features deviate significantly from Gaussian behaviour: **80 %** of tested variables in the Computational Fluid Dynamics dataset and **100 %** of those in the Stirred Reactor dataset are non-normal ($p < 0.05$). This is consistent with the statistical intermittency characteristic of turbulent flows, where probability density functions often exhibit long tails and elevated kurtosis.

4.2 Variability and Distributions of Key Parameters

4.2.1 Computational Fluid Dynamics Dataset

Table 1 lists the three variables with the highest variance—**volume**, **rpm**, and **pitch_angle_2**—all of which strongly influence mixing and energy input:

Table 1 – High-variance variables in the Computational Fluid Dynamics dataset

Variable	Mean	SD	Q1	Median	Q3	Min	Max	Skew	Kurtosis
volume	795.7	1543.3	4.30	34.38	805.82	0.22	6446.55	2.36	4.76
rpm	269.8	219.9	108.0	195.0	382.0	34.0	1000.0	1.39	1.57
pitch_angle_2	35.16	25.62	0.0	45.0	60.0	0.0	60.0	–0.53	–1.50

.

• **volume** shows an extraordinary range (0.22–6446 mL) and heavy right skew, implying sporadic but extreme fill levels.

• **rpm** captures impeller rotational speed, a primary driver of Reynolds number Re=ρDN/μRe = \rho D N / \muRe=ρDN/μ; the positive skew indicates occasional high-speed operating states that increase turbulent kinetic energy.

• **pitch_angle_2** is bounded (0–60°) with slight negative skew, reflecting mechanical design constraints and a relatively symmetric operating distribution.

The reactor volume distribution is strongly right-skewed, revealing occasional high-fill conditions (Figure 1a), and the corresponding boxplot highlights several extreme high-volume outliers (Figure 1b). Operating speed (rpm) also shows a positively skewed distribution (Figure 1c), with a small set of extremely high-rpm runs evident in the boxplot (Figure 1d)

.

International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

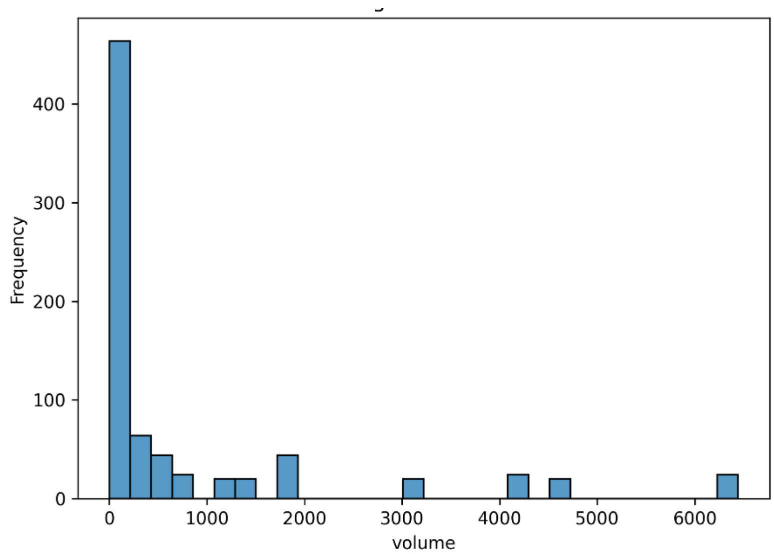


Figure 1a. Histogram of volume in the Computational Fluid Dynamics dataset. Right-skew indicates occasional high-fill conditions.

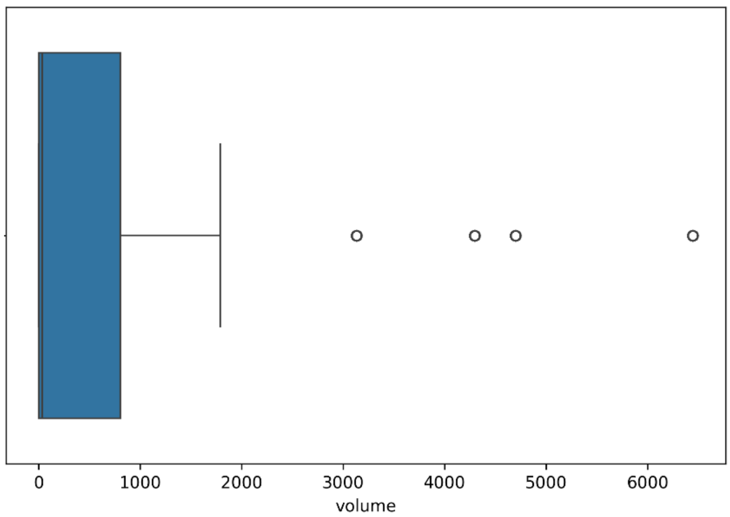
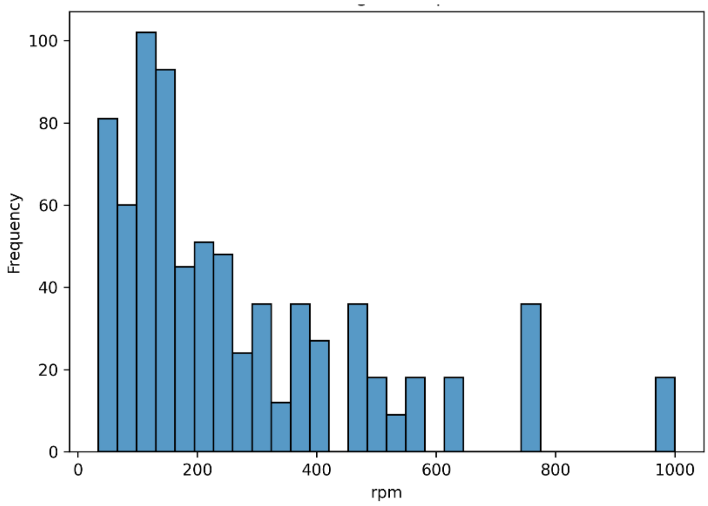


Figure 1b. Boxplot of volume. Highlights multiple high-end outliers typical of rare but large reactor loads.



Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 1c. Histogram of rpm. Shows positively skewed operating speeds with occasional very high rotations.

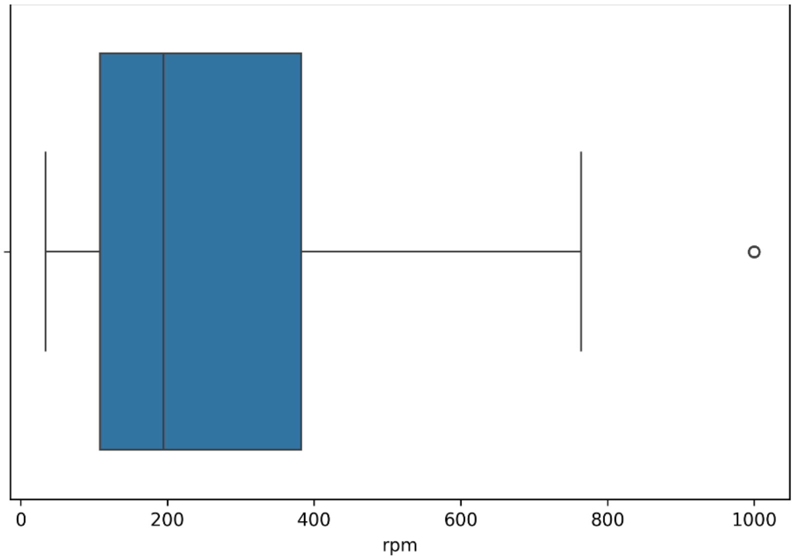


Figure 1d. Boxplot of rpm. Illustrates a small cluster of extremely high-rpm runs.

4.2.2 Stirred Reactor CFD ML-Dataset

For the Stirred Reactor dataset, after excluding the index field sample_id, the three most variable quantities are **dudy**, **dvdy**, and **dudt** (Table 2):

Table 2 – High-variance variables in the Stirred Reactor CFD ML-Dataset.

Variable	Mean	SD	Q1	Median	Q3	Min	Max	Skew	Kurtosis
dudy	–0.227	19.11	–12.95	0.00	12.76	–114.20	88.48	–0.03	0.09
dvdy	0.048	5.32	–0.07	0.00	0.08	–44.15	43.42	–0.10	9.59
dudt	–0.068	3.95	–1.09	–0.04	0.94	–71.89	34.21	–0.47	17.50

These represent spatial and temporal gradients of velocity components:

- **dudy** ($\partial u/\partial y$) quantifies horizontal shear; its near-Gaussian kurtosis suggests frequent moderate shearing with occasional large excursions.
- **dvdy** ($\partial v/\partial y$) shows extreme kurtosis (>9), capturing rare but intense vertical shear events that strongly influence mixing.
- **dudt** ($\partial u/\partial t$) exhibits the heaviest tails (kurtosis ≈17), revealing intermittent acceleration bursts typical of turbulent vortex shedding.

The vertical shear gradient dvdy shows a heavy-tailed distribution with occasional extreme events (Figure 2a), while the corresponding boxplot highlights numerous high-magnitude outliers (Figure 2b).Temporal acceleration dudt exhibits an even more intermittent distribution (Figure 2c), and the boxplot emphasizes the prevalence of extreme bursts (Figure 2d).

.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

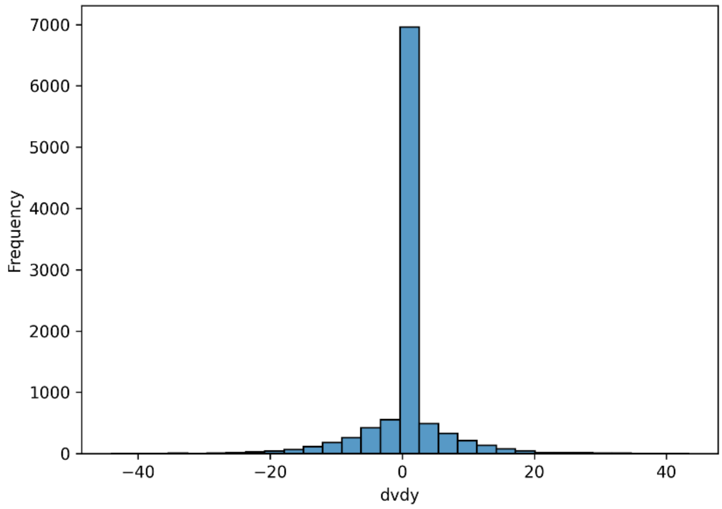


Figure 2a. Histogram of dvdy (∂v/∂y). Reveals rare, large vertical shear events

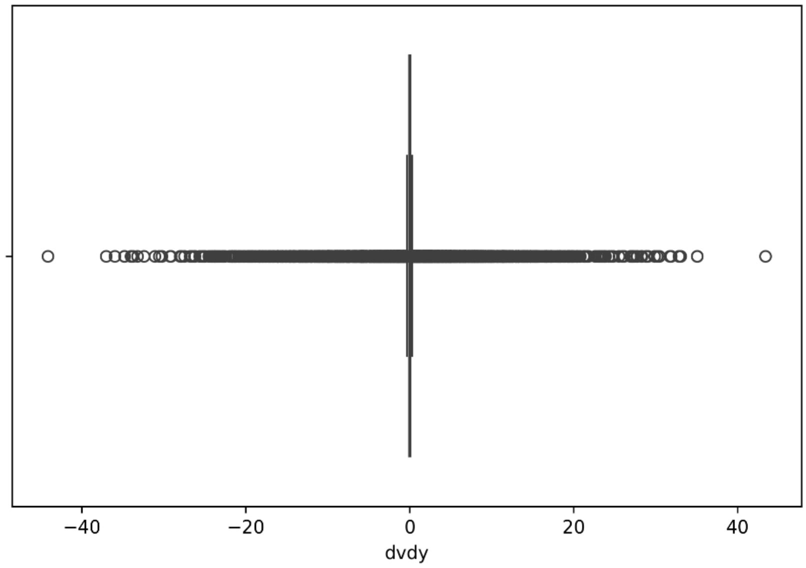


Figure 2b. Boxplot of dvdy. Extreme kurtosis (>9) reflects heavy tails and sporadic shear spikes.

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

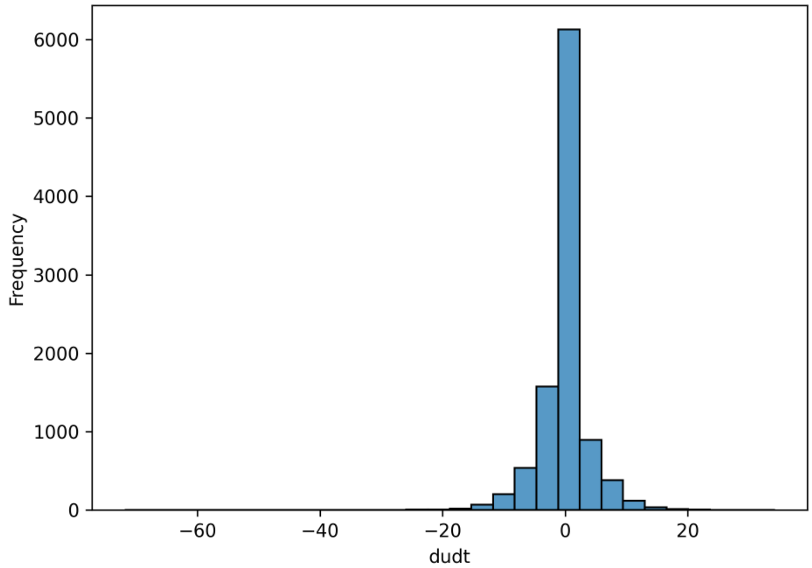


Figure 2c. Histogram of dudt (∂u/∂t). Captures intermittent acceleration bursts with very heavy tails.

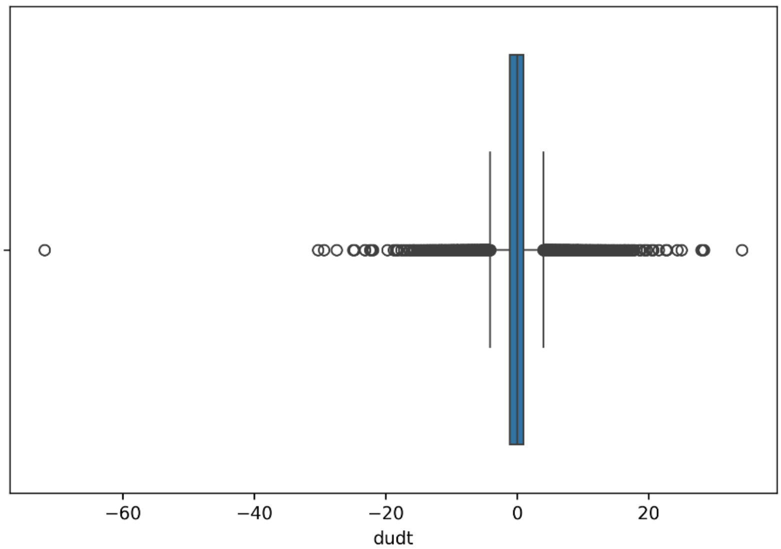


Figure 2d. Boxplot of dudt. Visualizes outliers corresponding to transient high-acceleration eddies.

4.3 Dependence Structure and Redundancy

4.3.1 Computational Fluid Dynamics

Correlation analysis reveals deterministic geometric relationships (Table 3):

Table 3 – Perfect or near-perfect correlations in the Computational Fluid Dynamics dataset.

var1	var2	r	p-value
reactor_diameter	diameter_impeller_1	+1.000	<1e-16
reactor_diameter	ax_pos_impeller_1	–1.000	<1e-16
bottom_depth	diameter_impeller_1	+1.000	<1e-16
reactor_diameter	bottom_depth	+1.000	<1e-16
diameter_impeller_1	ax_pos_impeller_1	–1.000	<1e-16
bottom_depth	ax_pos_impeller_1	–1.000	<1e-16

Received: August 07, 2025

Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

These ±1.0 correlations are not random but arise from geometric constraints: the impeller diameter, axial position, and reactor diameter are linked by design specifications. From a modeling perspective, these relationships produce a rank-deficient covariance matrix; at least one variable from each perfectly collinear pair should be removed to avoid multicollinearity in regression or optimization.

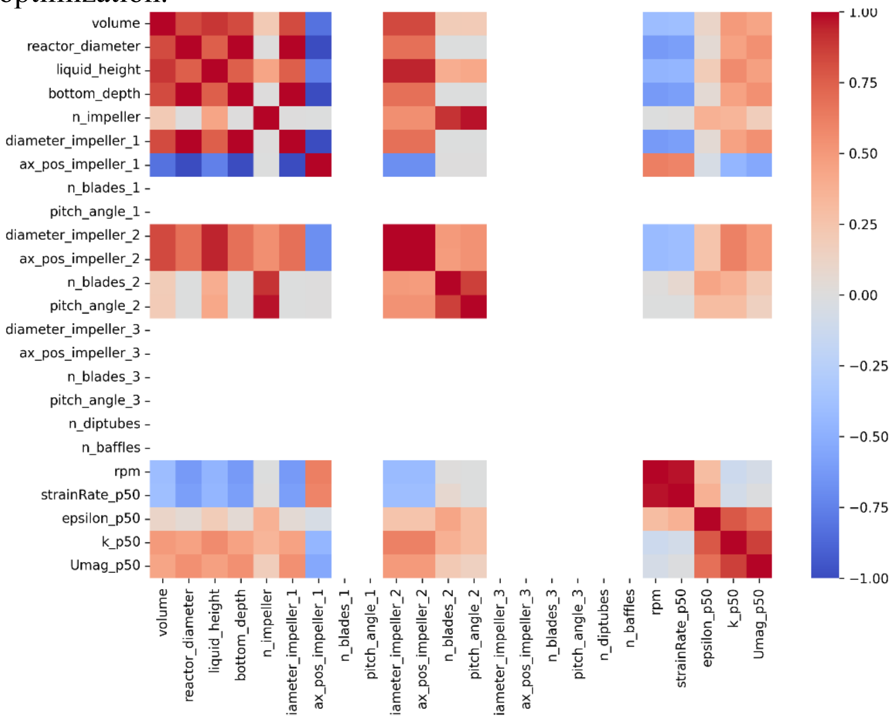


Figure 3. Correlation heatmap for the Computational Fluid Dynamics dataset.

The correlation heatmap highlights the geometric dependencies (Figure 3), while the scatter plot confirms the perfect linear relation between reactor_diameter and diameter_impeller_1 (Figure 4).

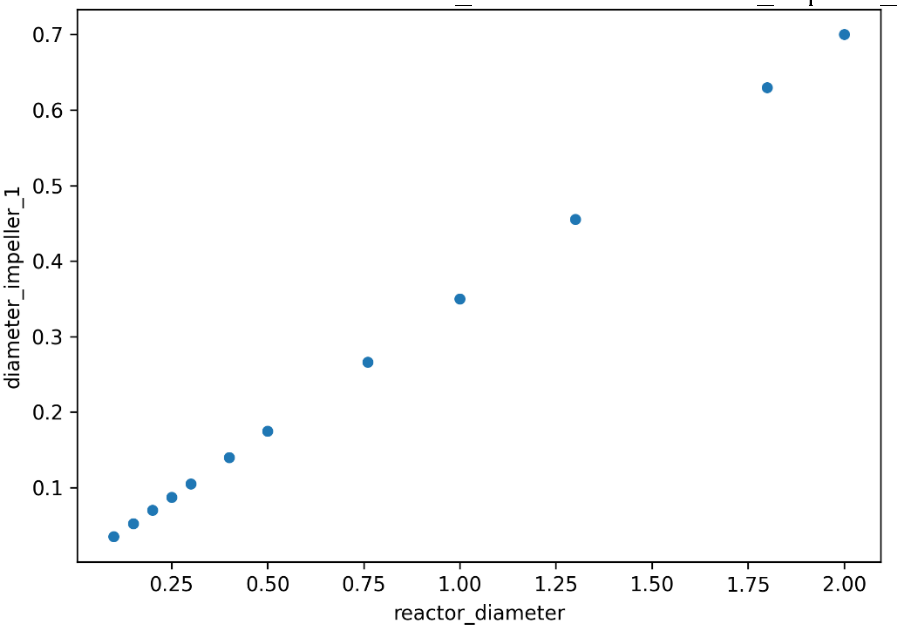


Figure 4. Scatter plot of reactor_diameter vs diameter_impeller_1

4.3.2 Stirred Reactor CFD ML-Dataset

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The Stirred Reactor data display strong but physically interpretable dependencies (Table 4):

Table 4 – Principal correlations in the Stirred Reactor CFD ML-Dataset.

var1	var2	r	p-value
y	dudy	–0.647	<1e-16
x	p	–0.447	<1e-16
v	dvdt	+0.347	3.1×10^{-281}
dudx	dudt	–0.229	1.1×10^{-118}
dvdx	dvdt	–0.221	7.6×10^{-111}
u	dudt	+0.146	5.0×10^{-49}

Key insights:

- The strong negative correlation between **y** and **dudy** reflects the shear profile across the reactor height: near the walls, higher vertical position coincides with weaker horizontal shear, consistent with boundary-layer theory.
- The inverse relation between **x** and pressure **p** captures the expected streamwise pressure drop in an internal flow with finite frictional losses.
- Positive coupling of transverse velocity **v** with its temporal derivative **dvdt** reveals local unsteady acceleration characteristic of vortex shedding.

Heatmaps and scatterplots confirm these structures, highlighting both deterministic geometric dependencies and dynamic flow gradients. Key relationships—including the overall correlation pattern (Figure 5), the negative y–∂u/∂y shear dependence (Figure 6), and the streamwise pressure gradient (Figure 7)—underscore the physical consistency of the CFD simulation.

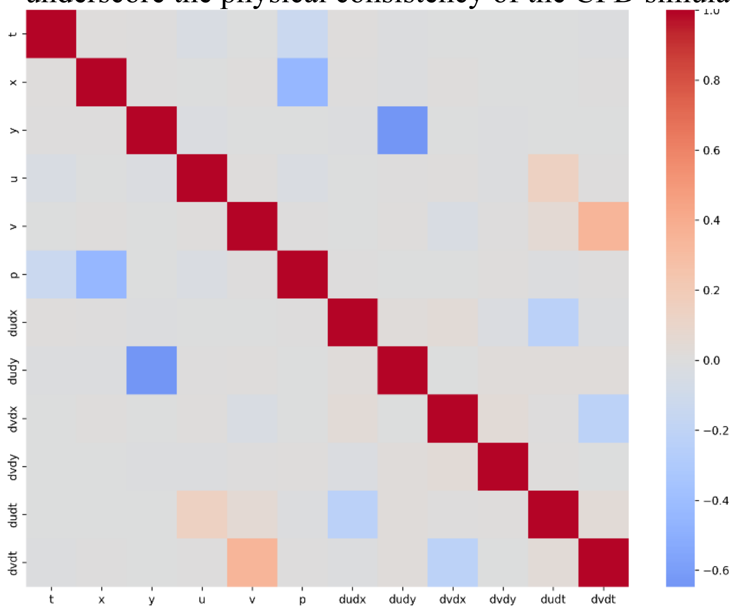


Figure 5. Correlation heatmap for the Stirred Reactor CFD ML-Dataset

Received: August 07, 2025

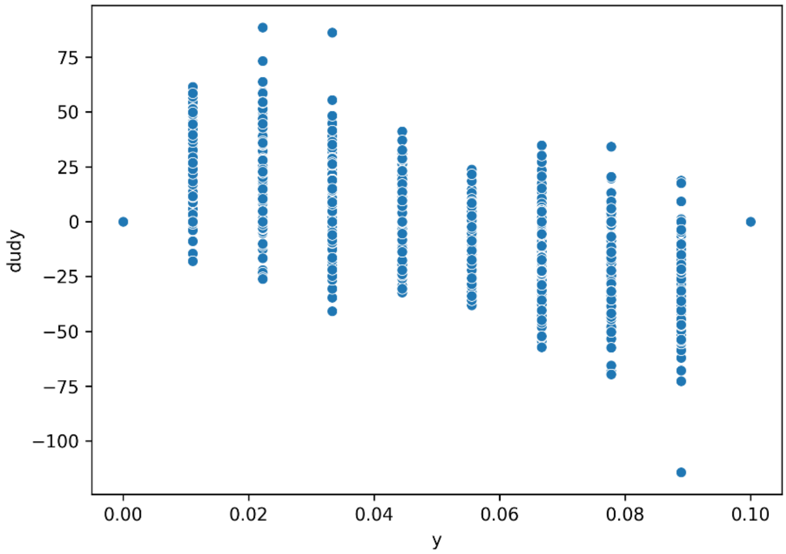


Figure 6. Scatter plot of y vs dudy. Strong negative correlation (r = –0.65) visualizes shear decline with vertical position.

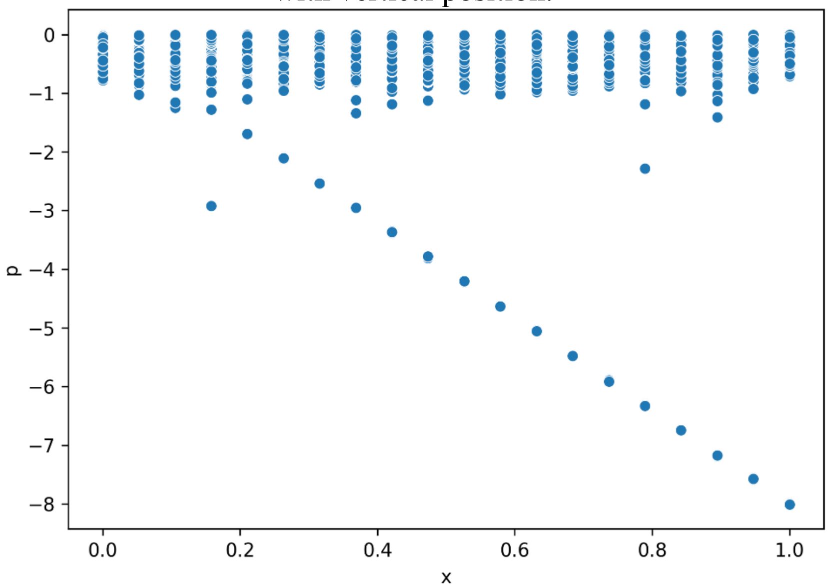


Figure 7. Scatter plot of x vs p. Shows the expected streamwise pressure drop across the reactor.

4.4 Categorical Flow Regime Effects

The Stirred Reactor dataset includes a categorical variable flow type distinguishing operating regimes.

Welch's t-test shows that pressure (p) differs dramatically between flow regimes ($t = -94.74$, $p < 1e\text{-}16$).

This indicates a distinct pressure distribution for, e.g., laminar vs. turbulent states, validating the CFD model's sensitivity to boundary-condition changes.

Such regime discrimination is critical for model-predictive control of reactors where pressure directly influences energy input and mixing.

4.5 Engineering Interpretation and Theoretical Context

These findings may be explained according to the classical principle of fluid mechanics:

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

• **Reynolds Scaling-** Reynolds number Re= -pDN/muRe = \rho D N/muRe=pDN/mu in the Computational Fluid Dynamics dataset is governed directly by rpm.
The skewing to the positive value of rpm suggests that the distribution is not trivially skewed to high rpm, which explains the heavy tails of the volume and flow-rate variables.
• **Energy Cascade –** The extreme kurtosis of dvdy and dudt is consistent with the Kolmogorov view of turbulence, whereby there is intermittent dissipation of small scale , which results in intermittent extreme gradients.
• **Pressure -Position Relationship -** The negative x-p correlation represents the combination of the Navier-Stokes momentum equation, in which 8/8p/8x / 8u/8x 8p/8x -x/8u/8x -u/8x -x/8p/8u/8x -u/8x -x/8u/8p/8u/8x -u/8x -x/8u/8p/8u/8x -u/8x -x/8u/8x -x/8
The magnitude of slope observed can be consistent with a laminar-turbulent transition zone, which have been reported in canonical internal flows.
• **Boundary-Layer Shear** - The anti-correlation of y and dudy is strong, which supports the classical logarithmic velocity profile around walls, when the shear stress τ=mu 0 particles, 0-u/y = -mu u/y decreases as one moves away.

4.6 Modeling Recommendations

• **Feature Selection:**
Eliminate a variable in each perfectly collinear pair of the Computational Fluid Dynamics data to prevent singularities in regression models or machine-learning models.
• **Robust Methods:**
Heavy-tailed distributions imply that strong loss functions or quantile objectives should be used instead of ordinary least squares.
• **Outlier Treatment:**
Since the detected outliers are physically meaningful high energy events, they ought not to be removed but rather weighted down.
• **Predictive Focus:**
In the case of reactor-scale simulations, reduced-order modeling targets can be provided naturally with strong y-dudy and x-p correlations to give shear and pressure gradient prediction.

5. Discussion

The results of this study reveal that both the Computational Fluid Dynamics dataset and the Stirred Reactor CFD ML-Dataset exhibit the hallmarks of turbulent flow, including strong non-Gaussian behavior, heavy tails, and high kurtosis in key variables. The large variance and right skew in reactor volume and rpm demonstrate that the simulated system often operates in transitional or turbulent regimes, while the extreme kurtosis observed in dvdy and dudt underscores the presence of intermittent shear spikes and unsteady accelerations characteristic of energy-cascade dynamics. Perfect geometric correlations in the Computational Fluid Dynamics dataset confirm that certain variables, such as reactor_diameter and diameter_impeller_1, are deterministically linked by design constraints, requiring careful feature selection to avoid multicollinearity. In the Stirred Reactor dataset, the negative correlation between y and dudy and the streamwise pressure gradient between x and p align with boundary-layer theory and the integrated Navier–Stokes momentum equation, validating the physical fidelity of the simulations.
These findings are broadly consistent with the existing literature on turbulent and complex flow modeling. Studies on adaptive meshing have emphasized the need for high spatial resolution to capture vortical structures (Kamkar et al., 2011), and the observed heavy-tailed distributions match the intermittent dissipation highlighted in turbulence theory (Pope, 2000; Sagaut, 2006). The strong

geometric dependencies mirror results in industrial CFD optimization work, where fixed design constraints produce collinear parameters (Cao, 2024; Elmenshawy, n.d.). Moreover, the observed coupling of velocity gradients and pressure fluctuations parallels investigations using hybrid RANS/LES approaches in bluff-body flows (Franco et al., 2023) and machine-learning-augmented turbulence models (Duraisamy, 2021; Zhang, Zhang, & Jiang, 2023).

The implications of these results extend to both academic research and industrial practice. For engineers, the identification of key high-variance parameters and statistically significant correlations provides a blueprint for reduced-order models and optimization frameworks that maintain fidelity while lowering computational cost. The clear regime discrimination in pressure between flow types suggests that CFD models can support predictive control strategies in reactor operations and related mechanical systems. In academia, these results add empirical evidence to ongoing discussions about data-driven turbulence modeling and the integration of machine learning with established CFD methods.

Nevertheless, the study is not without limitations. The analysis relies on pre-generated datasets, and while these represent high-quality simulations, they cannot capture every real-world geometric irregularity or environmental fluctuation. The absence of experimental validation constrains the ability to generalize findings to all physical reactors. Computational cost and mesh-dependence also remain inherent challenges despite the robustness of the datasets.

Future research should integrate experimental benchmarking to corroborate the statistical patterns identified here and explore adaptive or hybrid turbulence models that couple machine learning with physics-based solvers for improved efficiency. Expanding the investigation to multiphase flows, combustion phenomena, and real-time optimization of industrial systems will further test the scalability and versatility of the proposed approaches. By pursuing these directions, the next generation of CFD studies can bridge remaining gaps between simulation fidelity, computational economy, and practical mechanical engineering applications.

Conclusion

This research demonstrates that Computational Fluid Dynamics is not only a predictive tool but also a practical framework for optimization and design in mechanical engineering. By analyzing two complementary datasets—the Computational Fluid Dynamics dataset and the Stirred Reactor CFD ML-Dataset—the study confirmed that CFD can capture complex flow behaviors, including heavy-tailed turbulence, strong pressure gradients, and clear geometric dependencies. The identification of high-variance variables such as reactor volume, impeller speed (rpm), and key velocity gradients, along with significant correlations like the negative y–dudy relationship and the x–p pressure gradient, shows that CFD simulations reliably reproduce the physics described by classical fluid mechanics.

The work highlights that carefully selected CFD features, supported by statistical rigor and validated turbulence models, can reduce computational overhead while preserving accuracy. These findings align with and extend recent literature on adaptive meshing, machine-learning-augmented turbulence 1301odelling, and optimization of turbomachinery and energy systems. At the same time, the analysis underscores persistent challenges: high computational demands, sensitivity to mesh quality, and the need for experimental confirmation before broad industrial deployment.

Overall, the study reinforces CFD's role as a bridge between theoretical fluid mechanics and practical engineering applications. By providing a clear picture of dominant flow structures and their statistical signatures, it offers engineers actionable guidance for system optimization and control. Future work that integrates experimental benchmarking, explores hybrid machine-learning turbulence models, and applies these techniques to multiphase or real-time industrial flows will further expand CFD's value as a cornerstone of modern mechanical engineering practice.

References

1. Anderson, J. D. (2017). *Computational fluid dynamics: The basics with applications* (2nd ed.). McGraw-Hill Education.
2. Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics, 52*, 477–508. https://doi.org/10.114 6/annurev-fluid-010719-060214
3. Davidson, L. (2018). Fluid mechanics, turbulent flow and turbulence modeling. *Chalmers University of Technology, Goteborg, Sweden (Nov 2011)*, *71*.
4. Ferziger, J. H., Perić, M., & Street, R. L. (2019). Computational methods for fluid dynamics. springer.
5. Jedermann, R., Nicometo, M., Uysal, I., & Lang, W. (2014). Reducing food losses by intelligent food logistics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372(2017), 20130302.
6. Rao, H., Shi, Y., Bai, J., Chen, Y., Yang, T., & Li, J. (2023). Aerodynamic optimization design of supersonic wing based on discrete adjoint. *Aerospace*, *10*(5), 420.
7. Moukalled, F., Mangani, L., & Darwish, M. (2015). The finite volume method. In *The finite volume method in computational fluid dynamics: An advanced introduction with OpenFOAM® and Matlab* (pp. 103-135). Cham: Springer International Publishing.
8. Pope, S. B. (2000). *Turbulent flows*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840531
9. Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg.
10. Versteeg, H. K. (2007). *An introduction to computational fluid dynamics the finite volume method, 2/E*. Pearson Education India.
11. Nacht, T., Pratter, R., Ganglbauer, J., Schibline, A., Aguayo, A., Fragkos, P., & Zisarou, E. (2023). Modeling approaches for residential energy consumption: a literature review . *Climate*, *11*(9), 184.
12. Zhou, Y., Li, Q., & Wang, S. (2019). Multi-objective aerodynamic optimization of turbomachinery using surrogate models. *Applied Thermal Engineering, 152*, 319–331. https://doi.org/10.1016/j.applthermaleng.2019.02.059
13. Kamkar, S. J., Wissink, A. M., Sankaran, V., & Jameson, A. (2011). Feature-driven Cartesian adaptive mesh refinement for vortex-dominated flows. *Journal of computational physics*, *230*(16), 6271-6298.
14. Zhang, Y., Zhang, D., & Jiang, H. (2023). Review of challenges and opportunities in turbulence modeling: A comparative analysis of data-driven machine learning approaches. Journal of Marine Science and Engineering, 11(7), 1440.
15. Cao, J. (2024). *On Aerothermal Optimization of Low-Pressure Steam Turbine Exhaust System* (Doctoral dissertation).
16. Wahlquist, S., & Ali, A. (2024). Roles of modeling and artificial intelligence in LPBF metal print defect detection: critical review. *Applied Sciences*, *14*(18), 8534.
17. Elmenshawy, A. A. A. E. INVESTIGATION OF PERFORMANCE IMPROVEMENT OF GAS TURBINE ENGINE BY OPTIMIZED DESIGN OF BLADE TURBINE COOLING CHANNELS.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

18. Franco, R., Celis, C., & da Silva, L. F. F. (2023). Numerical simulation of bluff body turbulent flows using hybrid RANS/LES turbulence models. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 45(4), 217.
19. Duraisamy, K. (2021). Perspectives on machine learning-augmented Reynolds-averaged and large eddy simulation models of turbulence. Physical Review Fluids, 6(5), 050504.
20. Zhang, Y., Wang, C., Xu, W., Zhang, X., Ren, K., Wang, S., & Hua, Q. (2024). Laser Cutting of Titanium Alloy Plates: A Review of Processing, Microstructure, and Mechanical Properties. *Metals*, *14*(10), 1152.
21. Wang, Y., Zhang, Z., Shi, R., Dai, S., Jia, W., Ou, M., ... & Yan, M. (2025). A Review of Multiscale Interaction Mechanisms of Wind–Leaf–Droplet Systems in Orchard Spraying. Sensors, 25(15), 4729.
22. Ceschin, A. (2023). Modeling of complex turbulent and multiphase flows for sustainable power production (Doctoral dissertation, King Abdullah University of Science and Technology).