
Latent-Space Forecasting of URANS Wake Flows Across Reynolds Numbers and LLM-Assisted Transition Regime Identification

Research assistance: GPT-5.2, EXAONE-4.01-32B
Inference models: gpt-4.1-mini, gemini-2.5-flash, claude-haiku-4-5

Soo Min Oh
Department of Data
Science, Inha University
Incheon 22212,
Republic of Korea
soominoh.phys@gmail.com

Subin Jeong
Department of Mechanical
Engineering, Texas A&M
University
College Station, TX 77843.
United States
sooibing@tamu.edu

Sangyeong Park
Future and Fusion Lab of
Architectural, Civil and
Environmental Engineering,
Korea University
Seoul, 02841,
Republic of Korea
kjk00095@korea.ac.kr

Abstract

Accurate prediction of unsteady wake flows across Reynolds numbers remains challenging due to high-dimensional spatiotemporal dynamics and regime-dependent behavior. This study presents a compact, data-driven framework built on time-resolved unsteady Reynolds-averaged Navier–Stokes (URANS) simulations for $\text{Re} \in \{1000, 2000, 3000, 4000, 5000, 7000, 9000, 11000, 22000\}$. Two-component velocity snapshots (u, v) are compressed using principal component analysis (PCA), yielding a low-dimensional latent representation with near-lossless reconstruction at $L = 50$. Using the $\text{Re} = 22000$ dataset, we formulate short-horizon latent forecasting from a finite history window and compare a linear extrapolation baseline with regularized linear predictors; Elastic Net provides the most consistent improvement and approaches the PCA reconstruction bound after decoding. Finally, as an auxiliary analysis, lightweight LLMs are queried as black-box similarity judges between endpoint references, and the entropy of repeated outputs is used as a heuristic uncertainty indicator for regime variation. Together, the results demonstrate an efficient pipeline for latent-space forecasting and a complementary, uncertainty-oriented view of transitional behavior from compressed URANS data. Notably, the LLM-based uncertainty trends consistently interpret the PCA-derived flow representations and provide an effective signal for transition-regime screening across Reynolds numbers.

1 Introduction

Turbulent flows exhibit high-dimensional, multiscale, and strongly nonlinear dynamics, which makes reliable prediction over long horizons challenging. While the Navier–Stokes equations provide a complete description, direct numerical simulation becomes prohibitively expensive at high Reynolds numbers (1). This motivates reduced-order representations that retain dominant flow structures while enabling efficient time evolution (2). A common approach is to project flow fields onto a low-dimensional subspace and model the reduced dynamics. Methods based on proper orthogonal decomposition (POD) and related operator-based formulations can compactly represent energetic

structures (3), but predictive performance may degrade in strongly nonlinear regimes due to truncation, instability, and model-form limitations (4). Data-driven approaches can learn reduced representations and temporal evolution from data, yet robust generalization across operating conditions remains difficult in chaotic flow systems (2).

In this study, we investigate reduced-order prediction of unsteady wake flows using time-resolved unsteady Reynolds-averaged Navier–Stokes (URANS) simulations. Two-component velocity fields are compressed via principal component analysis (PCA) into a low-dimensional latent space, where short-horizon forecasting is posed as a multivariate time-series problem and decoded back to the physical space for error evaluation (5; 6). The analysis spans multiple Reynolds numbers to examine how compression efficiency and near-term predictability vary as wake dynamics and turbulence-closure responses change (6; 7). Finally, lightweight large language models (LLMs) are used as black-box similarity judges on the reduced representations to provide an auxiliary, uncertainty-oriented view of regime variation, without claiming physical interpretability or replacing turbulence modeling.

2 Problem Formulation

We consider a two-dimensional URANS flow over a spatial domain $\Omega \subset \mathbb{R}^2$, with $\mathbf{x} \in \Omega$. After spatial discretization on N_p grid points, each snapshot is represented by the two-component velocity field $(u(\mathbf{x}, t), v(\mathbf{x}, t))$. Here, u and v correspond to the streamwise and transverse velocity components, respectively (equivalently $u \equiv u_1$ and $v \equiv u_2$ in index notation). By stacking the two components, the high-dimensional state vector is written as

$$\mathbf{u}(t) = [u(\mathbf{x}_1, t), \dots, u(\mathbf{x}_{N_p}, t), v(\mathbf{x}_1, t), \dots, v(\mathbf{x}_{N_p}, t)]^\top \in \mathbb{R}^{2N_p}.$$

Here, $\mathbf{u}(t)$ denotes the resolved mean-flow variables obtained from URANS, while the effects of turbulent fluctuations are modeled through a turbulence closure. Although URANS significantly reduces the computational cost compared to direct numerical simulation, the resulting system remains high-dimensional and nonlinear, and its temporal evolution can exhibit complex dynamics depending on the flow configuration and Reynolds number.

To enable predictive modeling in a reduced space, we assume that the high-dimensional URANS state vectors $\mathbf{u}(t)$ admit an accurate approximation in a low-dimensional subspace of \mathbb{R}^{2N_p} . Accordingly, we introduce a generic encoding-decoding mapping,

$$\mathbf{z}(t) = \mathcal{E}(\mathbf{u}(t)), \quad \hat{\mathbf{u}}(t) = \mathcal{D}(\mathbf{z}(t)), \quad (1)$$

where $\mathcal{E} : \mathbb{R}^{2N_p} \rightarrow \mathbb{R}^r$ and $\mathcal{D} : \mathbb{R}^r \rightarrow \mathbb{R}^{2N_p}$ denote an encoding and decoding operator, respectively, and $r \ll 2N_p$ is the latent dimension. This formulation is agnostic to the specific choice of reduced representation and is introduced here to define the general latent-space prediction framework.

The temporal evolution of the flow is modeled in the latent space as a discrete-time dynamical system. Let $\{\mathbf{z}_n\}_{n=1}^T$ denote the latent states sampled at uniform time intervals Δt . We define a latent dynamics operator \mathcal{F} such that

$$\mathbf{z}_{n+1} = \mathcal{F}(\mathbf{z}_{n-k+1}, \dots, \mathbf{z}_n), \quad (2)$$

where $k \geq 1$ denotes the length of the temporal history used to predict the next state. This formulation allows the latent dynamics to incorporate memory effects and nonlinear temporal dependencies that arise from the coupled mean-flow evolution and turbulence modeling, without assuming Markovian behavior in the reduced space.

Given an initial sequence of latent states $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$, the model generates a predicted latent trajectory $\{\hat{\mathbf{z}}_{k+1}, \hat{\mathbf{z}}_{k+2}, \dots\}$ by iteratively applying \mathcal{F} . The corresponding high-dimensional flow prediction at time step n is then obtained via the decoder as

$$\hat{\mathbf{u}}_n = \mathcal{D}(\hat{\mathbf{z}}_n). \quad (3)$$

In this study, the encoder-decoder mapping is instantiated using PCA, yielding a linear and interpretable latent representation of the URANS flow fields. The latent dynamics operator \mathcal{F} is implemented using regularized linear regression models, with Elastic Net (8) adopted as the primary

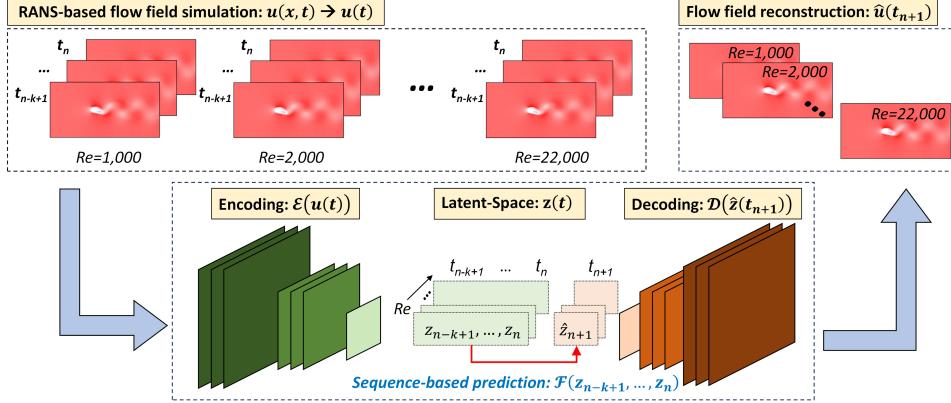


Figure 1: Overall workflow of the latent-space dynamical prediction framework.

predictor and Ridge (9) and Lasso (10) regressions employed as reference baselines for comparative evaluation.

Finally, we introduce a transition-regime screening task based on PCA-derived latent representations. Each Reynolds-number case is summarized as a fixed latent flow image, and intermediate cases are compared against low- and high-Re endpoint references using a black-box similarity judge that returns a soft preference. By repeating the comparison and tracking the consistency of the returned preference, we obtain an uncertainty signal. Reynolds numbers exhibiting higher uncertainty are treated as candidates for transitional behavior.

3 Methodology

In the present study, the encoding–decoding operators \mathcal{E} and \mathcal{D} are instantiated using PCA, yielding a linear and interpretable latent representation. Figure 1 provides an overview of the proposed methodology for turbulent flow prediction based on latent-space dynamical learning. High-dimensional flow fields are first obtained from time-dependent URANS simulations and used as input data. These flow snapshots are mapped onto a low-dimensional latent representation through an encoding process. The temporal evolution of the latent variables is then modeled as a sequence-based prediction problem, where future latent states are inferred from a finite history of past states. Finally, the predicted latent states are decoded back into the physical space to reconstruct the flow field at a target time, enabling assessment of both in-distribution prediction accuracy and generalization across Reynolds numbers. The image-based regime identification using large language models is introduced separately in Section 3.4 as a complementary analysis.

3.1 URANS-Based Flow Simulation and Data Preparation

Time-dependent flow fields were generated using two-dimensional URANS simulations, which describe the conservation of mass and momentum for the resolved mean flow,

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0, \quad (4)$$

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = -\frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_i} + \nu \frac{\partial^2 \bar{u}_i}{\partial x_j^2} - \frac{\partial}{\partial x_j} \overline{u'_i u'_j}, \quad (5)$$

where indices $i, j \in \{1, 2\}$ denote the streamwise and transverse directions, respectively, and the Einstein summation convention is implied (11). Here, \bar{u}_i and \bar{p} are the Reynolds-averaged velocity components and pressure, and $\overline{u'_i u'_j}$ is the Reynolds stress tensor, which is closed using a turbulence model. In the present two-dimensional configuration, the mean velocity field is written as $\bar{\mathbf{u}} = (\bar{u}, \bar{v})$.

URANS simulations were performed for a selected set of Reynolds numbers, $Re \in \{1000, 3000, 5000, 7000, 9000, 11000, 13000, 15000, 22000\}$, covering a broad range from low to high Reynolds numbers. This set provides sufficient variation in wake dynamics while remaining

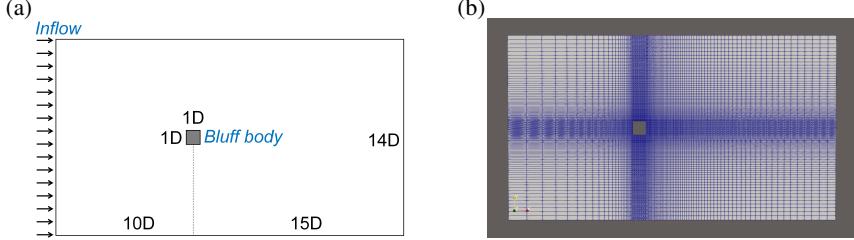


Figure 2: Schematic illustration of (a) geometrical domain and (b) computational mesh configuration

computationally tractable for long, time-resolved simulations. Across this range, the temporal evolution of the mean flow is explicitly resolved, while the effects of turbulent fluctuations are modeled through a turbulence closure (12).

The flow configuration consists of a canonical two-dimensional bluff-body wake (Fig. 2). A square prism of size $1D \times 1D$ is placed in a rectangular domain normalized by the obstacle diameter D , with the inlet located $10D$ upstream and the outlet $15D$ downstream of the prism. The domain height is $14D$. A uniform inflow velocity U_∞ is prescribed at the inlet, with standard outflow and lateral boundary conditions applied to minimize confinement effects.

All simulations were conducted using the finite-volume solver OpenFOAM with a transient URANS formulation and a $k-\omega$ turbulence model (13). A consistent solver and turbulence-closure setup was used for all Reynolds numbers to ensure comparability across cases. After initial transients decayed and a statistically stationary state was reached, snapshots of the Reynolds-averaged velocity field (\bar{u}, \bar{v}) were sampled at uniform temporal intervals. For brevity, we denote (\bar{u}, \bar{v}) by (u, v) in the remainder of the paper. Each snapshot was vectorized to form the high-dimensional state representation used for the subsequent latent-space analysis. At lower Reynolds numbers, the modeled eddy-viscosity contribution becomes small and the URANS solutions approach laminar-like behavior (14).

3.2 Latent Space Construction and Temporal Dynamics Prediction

Following the general formulation introduced in Section 2, the encoder-decoder mapping is instantiated using PCA, yielding a linear and interpretable latent representation of the URANS velocity fields. Each flow snapshot, consisting of the two-component velocity field (u, v) , is projected onto a low-dimensional latent vector $\mathbf{z}(t) \in \mathbb{R}^r$, with $r \ll N$, while retaining the dominant spatiotemporal wake structures. The temporal evolution of the flow is modeled directly in the latent space. Latent states sampled at uniform time intervals are treated as a multivariate time series, and short-horizon prediction is formulated using a finite history window of length k . The next latent state is predicted from the previous k states using a regression-based latent dynamics model. We consider multiple regression models for this latent predictor and compare their short-horizon performance. This separation between spatial compression and temporal prediction enables efficient reduced-order modeling of unsteady wake dynamics and facilitates systematic evaluation of predictive accuracy and generalization behavior across Reynolds numbers.

3.3 Flow Field Reconstruction and Evaluation Metrics

Following the latent-space prediction described in Section 2, the predicted latent states are decoded back into the physical space to reconstruct the corresponding velocity fields. This reconstruction enables direct comparison between predicted flow fields and the reference URANS solutions, allowing quantitative assessment of prediction accuracy in the original high-dimensional state space. Reconstruction accuracy is first evaluated using a global normalized ℓ_2 error defined at each time step as $\varepsilon(t) = \|\hat{\mathbf{u}}(t) - \mathbf{u}(t)\|_2 / \|\mathbf{u}(t)\|_2$ which provides a compact measure of instantaneous prediction fidelity. In addition, spatial distributions of reconstruction error are examined to identify localized discrepancies associated with wake structures, shear layers, and regions of strong velocity gradients.

Beyond instantaneous errors, statistical flow characteristics are used to assess the long-term predictive behavior of the latent-space model. Time-averaged velocity fields and selected second-order statistics are computed from the reconstructed trajectories and compared against reference URANS statistics.

These comparisons assess whether the latent-space predictions preserve the dominant mean-flow organization and variability of the wake. All evaluation metrics are applied to both in-distribution cases, where the Reynolds number is included in the training data, and out-of-distribution cases, where predictions are performed at Reynolds numbers not seen during training. This distinction enables systematic analysis of prediction accuracy, error accumulation, and generalization behavior as the flow regime varies.

3.4 Identification of transition regime via Image-Based Inference

In addition to latent-space dynamical prediction, this study explores the use of large language models (LLMs) as a novel inference tool for identifying transitional flow regimes from reduced flow-field images. Figure 3 illustrates the overall workflow of the proposed LLM-based image inference procedure, from reduced flow-field representations at reference Reynolds numbers to entropy-based identification of transition regimes. Unlike conventional classifiers trained on fixed labels or regression targets, LLMs are used here as black-box visual similarity judges: given endpoint reference images, they assign soft preferences to intermediate cases. We summarize the variability of these preferences via repeated queries and entropy to obtain an uncertainty proxy for regime change. To reduce model-specific bias and assess robustness, multiple lightweight LLMs are employed in this study, including gpt-4.1-mini, gemini-2.5-flash, and claude-haiku-4-5.

Reduced flow-field images corresponding to two reference Reynolds numbers, $Re = 1000$ and $Re = 22000$, are first provided to the LLMs as endpoint examples representing low- and high-Reynolds-number wake regimes, respectively. Given an input image at an intermediate Reynolds number, each LLM is prompted to infer its association with either reference regime based solely on the visual characteristics of the reduced flow representation. The inference outcome is encoded as a binary preference vector, reflecting the model's relative association with each endpoint regime.

To ensure robustness and mitigate stochastic variability inherent to LLM-based inference, the classification process is repeated over multiple independent realizations. For each Reynolds number, ten independent inference trials are conducted, and the resulting one-hot vectors (15; 16) are averaged to obtain a probabilistic measure of regime association. The entropy of the averaged output distribution is then computed and used as a quantitative indicator of classification uncertainty. High entropy values indicate Reynolds numbers for which the LLMs exhibit difficulty in decisively assigning the flow to either reference regime, suggesting the presence of a transitional or transient flow regime.

The use of LLM-based image inference in this context departs from conventional turbulence analysis and data-driven classification approaches. The LLM outputs are treated strictly as heuristic judgments of image-level pattern similarity, without any claim of physical interpretability or causal modeling. The analysis focuses on uncertainty trends induced by repeated inference across Reynolds numbers, enabling identification of Reynolds-number ranges associated with increased regime ambiguity. This perspective provides a complementary, uncertainty-oriented view of flow regime variation that augments the latent-space dynamical framework.

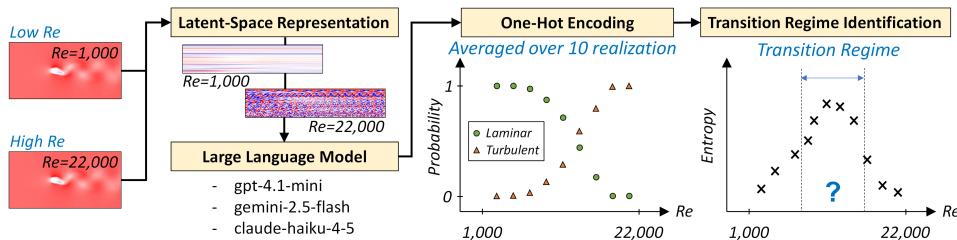


Figure 3: LLM-based identification of transitional flow regimes using reduced flow-field images.

4 Results

4.1 PCA-Based Compression of Time-Resolved Velocity Fields

We first construct a compact representation of the time-resolved velocity fields using PCA. For each $Re \in \{1000, 2000, 3000, 4000, 5000, 7000, 9000, 11000, 22000\}$, we collect the two-component

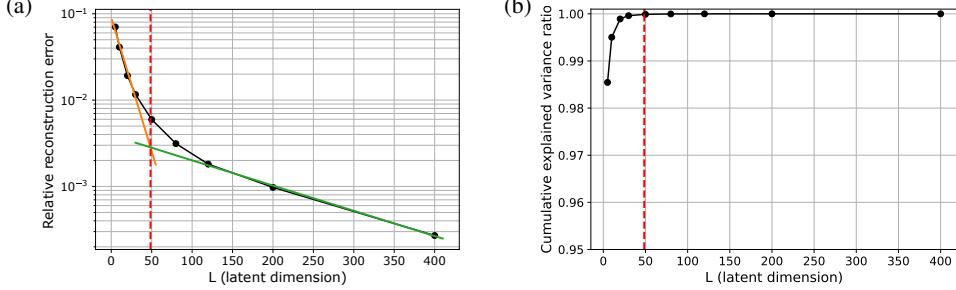


Figure 4: PCA compression performance versus latent dimension L . (a) Relative reconstruction error (normalized ℓ_2 error) showing a two-regime decay; the fitted trends intersect at $L^* \approx 48.56$ (vertical dashed line). (b) Cumulative explained variance ratio, approaching unity near $L \approx 50$.

velocity snapshots (u, v) at a time resolution of 0.01 from $t = 15$ s to $t = 20$ s, yielding 501 time steps per case. Each snapshot is defined on a grid with 15000 pixels, so the raw data tensor has shape $(N_{\text{Re}}, N_t, N_p, N_f) = (9, 501, 15000, 2)$. We then vary the PCA latent dimension $L \in \{5, 10, 20, 30, 50, 80, 120, 200, 400\}$ and evaluate reconstruction quality using the relative reconstruction error (normalized ℓ_2 norm), together with the cumulative explained variance ratio (Fig. 4). The reconstruction error exhibits a clear two-regime decay: it decreases rapidly for small L and more slowly for large L . To quantify this behavior, we fit two trends in the small- L and large- L ranges and define their intersection as the optimal compression dimension. The intersection occurs at $L^* \approx 48.5569$, where the reconstruction error is 0.0028343 and the cumulative variance ratio is 0.999875, indicating near-lossless reconstruction. Based on this result, we select $L = 50$ as the practical latent dimension that guarantees a similarly high reconstruction fidelity. Consequently, the PCA-compressed dataset has shape $(9, 501, 50)$, meaning that each (u, v) field of dimension 15000×2 is reduced to a 50-dimensional latent vector. Finally, we apply z -score normalization over the full dataset. This representation can be used directly as a time series in latent space, or equivalently interpreted as nine images of size $(501, 50)$, one per Re (Fig. 5).

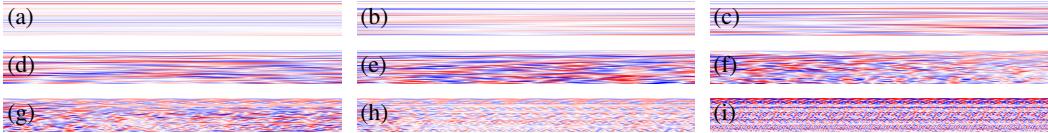


Figure 5: z -score normalized PCA latent images of size $(501, 50)$ for each Reynolds number: (a) $\text{Re} = 1000$, (b) 2000, (c) 3000, (d) 4000, (e) 5000, (f) 7000, (g) 9000, (h) 11000, and (i) 22000.

4.2 Short-Horizon Forecasting in the Turbulent Regime

Next, we focus on a reference turbulent case ($\text{Re} = 22000$) and investigate whether the reduced-order representation enables accurate short-horizon prediction. We use the PCA-compressed and normalized data in the $(501, 50)$ form with a 0.01 time interval. Our goal is to predict the next latent vector from a short history window of length 20, i.e., using the previous 20 time steps as input to estimate the immediately following step. We construct the dataset by sliding a one-step window across time, producing 481 input–output pairs. Among them, the first 400 pairs are used for training and the remaining 81 pairs (corresponding to the later time range, roughly $t = 19.2$ s to $t = 20$ s) are used for testing. When validation is required, we reserve 10% of the training set as a validation split. We compare four models: (i) a one-step linear extrapolation baseline (used as a reference), (ii) Ridge regression, (iii) Lasso regression, and (iv) Elastic Net regression. All models are trained in the latent space and then evaluated by predicting the test interval. The results (Fig. 6) show that the regression-based models (ii–iv) substantially outperform the baseline (i). Among them, Elastic Net provides the most consistent performance. Importantly, after decoding the predicted latent vectors back to the original (u, v) fields, Elastic Net achieves a reconstruction quality that is nearly indistinguishable from the intrinsic PCA decoding fidelity (Fig. 7). This indicates that the dominant prediction error is limited by the PCA reconstruction bound rather than the forecasting model itself.

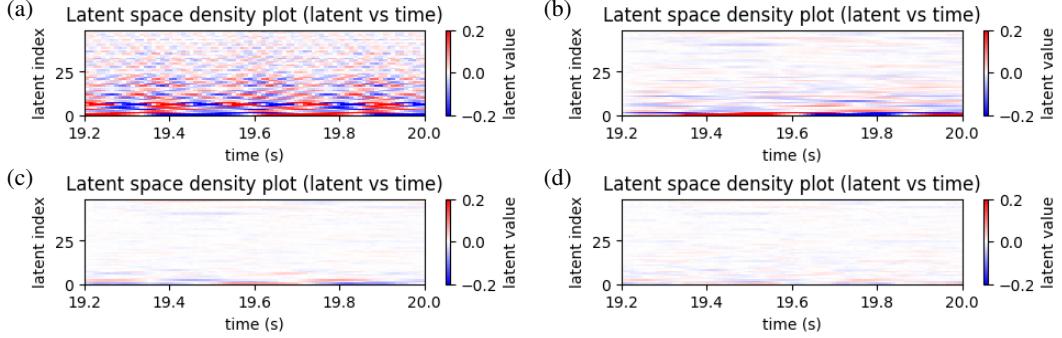


Figure 6: Test-set latent-space error maps (model prediction minus PCA-decoded ground truth) at $\text{Re} = 22000$ for (a) linear extrapolation, (b) Ridge, (c) Lasso, and (d) Elastic Net.

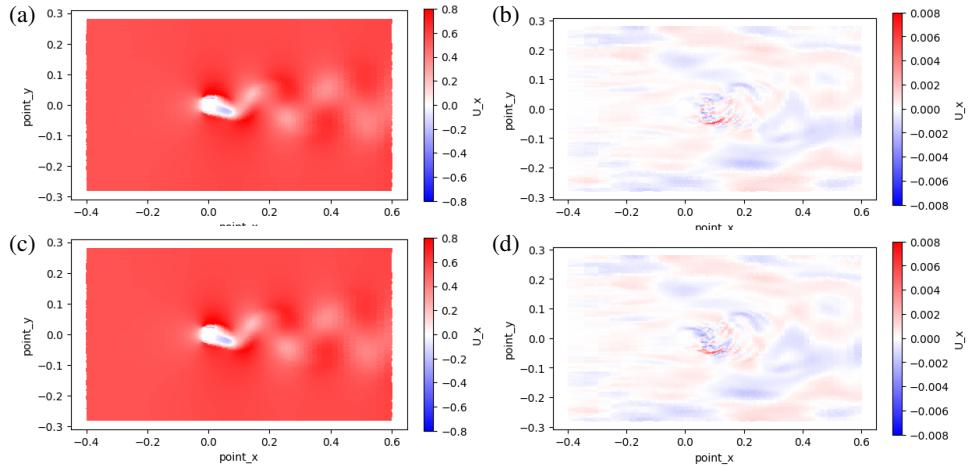


Figure 7: Comparison of inverse-PCA reconstructions and Elastic Net short-horizon prediction for the streamwise velocity field u_x at $\text{Re} = 22000$. (a) u_x reconstructed by decoding the PCA-compressed representation (inverse PCA). (b) Reconstruction error, $u_x^{\text{PCA decode}} - u_x^{\text{orig}}$. (c) Elastic Net one-step prediction in latent space decoded back to the physical field (inverse PCA). (d) Prediction error, $u_x^{\text{EN pred}} - u_x^{\text{orig}}$, showing that the Elastic Net forecast closely matches the original field.

Furthermore, the same latent forecasting pipeline was also applied to the remaining eight Re cases, and we verified that the prediction performance remains consistently strong across all Re .

4.3 LLM-Based Identification of the transition regime

Finally, we propose an LLM-based transition-regime indicator using the PCA flow images described in Fig. 5. The goal is to extract a heuristic uncertainty trend across Reynolds numbers from reduced URANS representations. We consider three lightweight models, gpt-4.1-mini, gemini-2.5-flash, and claude-haiku-4-5. Each model is provided with two endpoint reference images via a consistent JSON prompt: a low-Re reference at $\text{Re} = 1000$ with $(y_1, y_2) = (1, 0)$ and a high-Re reference at $\text{Re} = 22000$ with $(y_1, y_2) = (0, 1)$. Here, y_1 and y_2 denote the model's soft preference for the low- and high-Re references, respectively. At $\text{Re} = 1000$, the modeled eddy-viscosity contribution is small and the URANS solution is laminar-like; at $\text{Re} = 22000$, the wake is fully unsteady and turbulence-model effects are significant.

We then query the remaining Reynolds-number images one by one. For each Re , we run $K = 10$ realizations per LLM (temperature 0.2) and average the predicted probabilities (y_1, y_2) , along with the corresponding binary entropy $H(y_1, y_2)$ to quantify uncertainty. Figure 8(a–c) reports the mean and standard deviation of $y_1(\text{Re})$ and $y_2(\text{Re})$, and we define the maximally confusing intersection Re^* as the point where $y_2(\text{Re}) = 0.5$. Figure 8(d–f) shows $H(\text{Re})$, from which we infer the transition

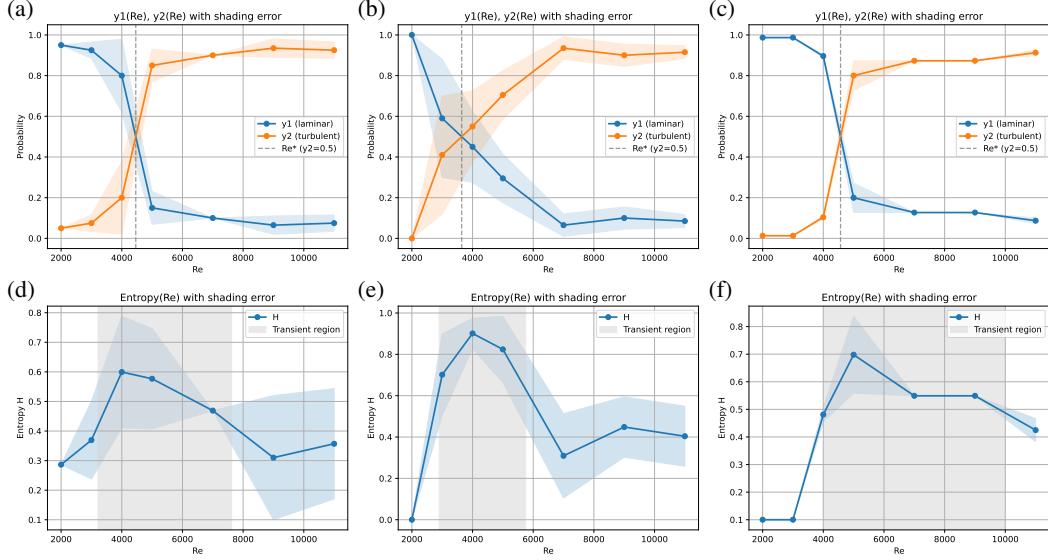


Figure 8: LLM-based identification of the transition regime in Re . Panels (a–c) show one-hot probabilities $y_1(\text{Re})$ (laminar) and $y_2(\text{Re})$ (turbulent), with the maximally confusing intersection Re^* given by $y_2 = 0.5$ (dashed line). Panels (d–f) show the entropy $H(\text{Re})$, used to define the transition regime (shaded). Models: (a,d) gpt-4.1-mini, (b,e) gemini-2.5-flash, (c,f) claude-haiku-4-5.

regime ($\text{Re}_{\min}, \text{Re}_{\max}$) by selecting the entropy maximum and including neighboring Reynolds numbers until $H(\text{Re})$ falls below 70% of its peak value.

As an external physical reference, we also analyze macroscopic aerodynamic indicators, namely C_L and C_D , as functions of Re (Fig. 9). From these trends, transitional behavior is observed over approximately $\text{Re} = 3000$ to 8000 . Table 1 summarizes the LLM-inferred Re^* and $(\text{Re}_{\min}, \text{Re}_{\max})$, and additionally reports overlap metrics (IoU, precision, recall) computed relative to the baseline interval $\text{Re} = 3000$ – 8000 . To quantify agreement, IoU measures the fractional overlap between the predicted and baseline intervals, precision measures the fraction of the predicted interval that lies within the baseline, and recall measures the fraction of the baseline interval covered by the prediction. Among the three models, gpt-4.1-mini shows the strongest agreement with the baseline (IoU = 0.881) and perfect precision (1.000), and high recall (0.881), indicating that its inferred transition range is fully contained within $\text{Re} = 3000$ – 8000 while covering 88.1% of that baseline range. In contrast, gemini-2.5-flash yields a lower IoU (0.539) with high precision (0.965) but modest recall (0.550), suggesting a conservative (narrow) transition estimate that largely stays within the baseline but captures only about half of it. Finally, claude-haiku-4-5 attains IoU = 0.569

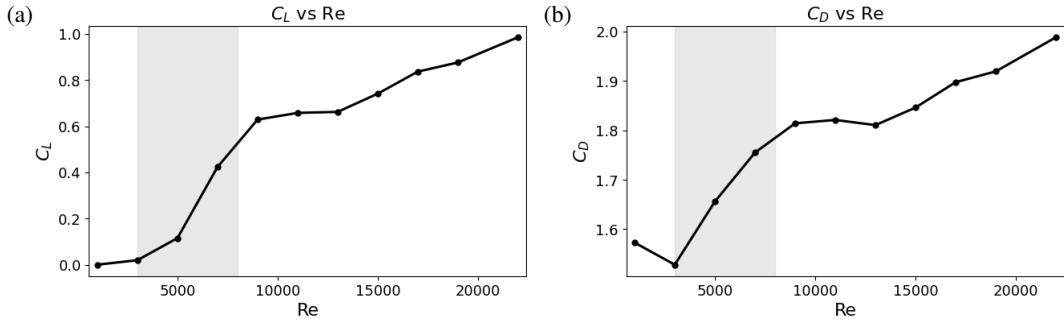


Figure 9: Macroscopic aerodynamic coefficients used as a physical baseline for validating the LLM-based transitional-regime estimate. Shown are (a) lift coefficient C_L and (b) drag coefficient C_D computed from the original (u, v) fields as functions of Re . The observed trend change suggests a transitional range of approximately $\text{Re} = 3000$ – 8000 (shaded).

Table 1: Summary of the LLM-inferred intersection Re^* and transition regime ($\text{Re}_{\min}, \text{Re}_{\max}$), along with overlap metrics (IoU, precision, recall) relative to the baseline $\text{Re} = 3000\text{--}8000$.

Model	Re^*	transition regime			IoU	Precision	Recall
		Re_{\min}	Re_{\max}				
gpt-4.1-mini	4462	3218	7621	0.881	1.000	0.881	
gemini-2.5-flash	3643	2899	5750	0.539	0.965	0.550	
claude-haiku-4-5	4569	4035	9974	0.569	0.668	0.793	
Baseline (from C_L/C_D analysis)	-	3000	8000	-	-	-	-

with recall 0.793 but reduced precision (0.668), reflecting a broader transition estimate that covers most of the baseline while extending beyond it toward higher Re . Overall, the inferred ranges are broadly consistent with the aerodynamic baseline, indicating that lightweight LLMs can provide a complementary uncertainty-oriented signal for identifying the transition regime from reduced URANS flow representations.

5 Discussion

We presented a data-driven framework for turbulent-flow modeling that unifies three capabilities in a common reduced representation: compact encoding of time-dependent URANS velocity fields, short-horizon forecasting in latent space, and uncertainty-oriented identification of the transition regime. The key observation is that PCA yields an information-preserving and standardized spatiotemporal descriptor of the wake dynamics, enabling both learning and cross- Re comparison at a fraction of the original dimensional cost. In this latent space, simple regularized predictors already capture the dominant short-time evolution, and the remaining discrepancy after decoding is largely constrained by the intrinsic compression–reconstruction fidelity rather than the forecasting model itself. This suggests that, for near-term prediction, representation quality is the primary bottleneck. Finally, we showed that lightweight LLMs can meaningfully interpret PCA-derived flow images to produce a coherent uncertainty signal across Reynolds numbers. By anchoring laminar and turbulent endpoints and tracking maximum confusion and entropy peaks, the inferred transition regime aligns with macroscopic aerodynamic trends in C_L and C_D (17), indicating that compact spatiotemporal flow descriptors can serve as an effective interface between reduced-order CFD outputs and foundation-model-based regime screening. The results presented here should be interpreted in the context of the modeling choices and scope adopted in this study. The latent representation is based on PCA, which is linear and may not fully capture nonlinear structures in more complex or higher-dimensional flows, despite achieving near-lossless reconstruction for the present wake configurations. The latent dynamics are evaluated in a short-horizon, one-step prediction setting; long-horizon stability and error accumulation are not addressed. In addition, all results are based on URANS simulations, and the learned dynamics may inherit biases associated with the turbulence-closure model, particularly in transitional regimes. The LLM-based regime identification is intentionally heuristic and relies on uncertainty trends from image-level similarity rather than physical classification. Future work will focus on extending the framework to nonlinear encoders, longer-horizon latent forecasting, and higher-fidelity or three-dimensional flow data. The LLM-based analysis could also be complemented by physics-informed or data-driven uncertainty measures.

6 Conclusion

We introduced a compact, data-driven workflow for analyzing time-dependent wake flows across Reynolds numbers by unifying three components: linear compression, latent-space forecasting, and LLM-based regime inference. By representing high-dimensional (u, v) fields as PCA-derived spatiotemporal descriptors, we obtain a standardized, information-preserving input that can be directly consumed by lightweight predictive models and, importantly, by modern LLMs. Our results demonstrate that LLMs can meaningfully interpret these reduced flow representations to produce an uncertainty-oriented signal that supports transition-regime localization, complementing conventional aerodynamic indicators. This direction suggests a scalable pathway toward automated, physics-aware screening and monitoring of complex flow behavior, with potential impact on high-throughput CFD analysis, surrogate modeling, and digital-twin pipelines where rapid decision support is required.

Acknowledgements

We acknowledge the use of large language models (LLMs) as research assistants throughout the full lifecycle of this study. In particular, GPT-5.2 (<https://platform.openai.com/docs/models>) and EXAONE-4.01-32B (<https://huggingface.co/LGAI-EXAONE/EXAONE-4.0.1-32B>) were used interactively via prompting to support research ideation and refinement, including the design of the overall methodology, the compression strategy for URANS (u, v) fields, the latent-space forecasting setup, and the selection and positioning of lightweight LLMs as a core component of the proposed novelty. In addition, three lightweight LLMs were used directly as experimental inference models that are explicitly incorporated into the proposed regime-identification pipeline, where they infer the transition regime from URANS (u, v) fields across Reynolds numbers via official APIs: gpt-4.1-mini (<https://platform.openai.com/docs/models/gpt-4.1-mini>), gemini-2.5-flash (<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-flash>), and claude-haiku-4-5 (<https://www.anthropic.com/news/claude-haiku-4-5>). All model usage followed the respective providers' access policies and terms, and no proprietary model weights were redistributed.

References

- [1] Javier Jiménez. Computing high-reynolds-number turbulence: will simulations ever replace experiments? *Journal of Turbulence*, 4(22), 2003. doi: 10.1088/1468-5248/4/1/022.
- [2] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52(1):477–508, 2020. doi: 10.1146/annurev-fluid-010719-060214.
- [3] Kunihiko Taira, Steven L. Brunton, Scott T. M. Dawson, Clarence W. Rowley, Tim Colonius, Beverley J. McKeon, Oliver T. Schmidt, Stanislav Gordeyev, Vassilios Theofilis, and Lawrence S. Ukeiley. Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12):4013–4041, 2017. doi: 10.2514/1.J056060.
- [4] Irina Kalashnikova and Matthew F. Barone. On the stability and convergence of a galerkin reduced order model (rom) of compressible flow with solid wall and far-field boundary treatment. *International Journal for Numerical Methods in Engineering*, 83(10):1345–1375, 2010. doi: 10.1002/nme.2867.
- [5] Kazuki Hasegawa, Kai Fukami, Takumi Murata, and Koji Fukagata. Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes. *Theoretical and Computational Fluid Dynamics*, 34:367–383, 2020. doi: 10.1007/s00162-020-00528-w.
- [6] Alberto Racca, Nguyen Anh Khoa Doan, and Luca Magri. Predicting turbulent dynamics with the convolutional autoencoder echo state network. *Journal of Fluid Mechanics*, 975, 2023. doi: 10.1017/jfm.2023.716.
- [7] Alberto Solera-Rico, Esteban Ferrer, Charles Moulinec, et al. β -variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nature Communications*, 15(1), 2024. doi: 10.1038/s41467-024-45578-4.
- [8] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.
- [9] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [10] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [11] Pritanshu Ranjan and Anupam Dewan. Partially averaged navier stokes simulation of turbulent heat transfer from a square cylinder. *International Journal of Heat and Mass Transfer*, 89: 251–266, 2015.

- [12] Florian R Menter. Improved two-equation k-omega turbulence models for aerodynamic flows. *Moffett Field, California, USA*, 1992.
- [13] David C Wilcox et al. *Turbulence modeling for CFD*, volume 2. DCW industries La Canada, CA, 1998.
- [14] W Rodi. Comparison of les and rans calculations of the flow around bluff bodies. *Journal of wind engineering and industrial aerodynamics*, 69:55–75, 1997.
- [15] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [16] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2 edition, 2009.
- [17] Jeppe Johansen and Jens N Sørensen. Prediction of laminar/turbulent transition in airfoil flows. *Journal of aircraft*, 36(4):731–734, 1999.

AI Co-Scientist Challenge Korea Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: The abstract and Introduction accurately describe the paper's scope and three main contributions—compact encoding of time-dependent URANS velocity fields, latent-space forecasting, and LLM-based transition regime inference—without overstating generality or guarantees.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: Section 6 explicitly discusses key limitations, including the linear PCA representation, short-horizon forecasting focus, URANS turbulence-closure bias near transition, and the heuristic nature of the LLM-based inference.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper does not present formal theoretical results requiring proofs; equations are used for problem formulation and metric definitions (Sections 2–4).

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [No]

Justification: Due to page-length constraints, not all low-level solver parameters and implementation details are exhaustively reported. However, the URANS simulations are based on standard, widely used models and numerical settings, and the flow configuration, Reynolds numbers, and data-processing pipeline are described in sufficient detail to allow partial reproduction of the results at a qualitative and reduced-order level. Furthermore, the paper specifies the full data-processing and evaluation pipeline used in the main claims. PCA is used as a linear encoder-decoder to map each (u, v) snapshot to a latent vector (with a fixed latent dimension and z -score normalization), and all Reynolds-number cases and time discretization used to form the latent “flow images” are reported. For short-horizon prediction, we describe the supervised one-step-ahead setup in latent space (history-window length, train/test split, and the family of regularized linear models considered), enabling reproduction of the forecasting protocol. For transition-regime identification, we report the three lightweight LLM APIs used and the endpoint-anchored prompting protocol (laminar and turbulent reference images, repeated sampling, and entropy-based uncertainty scoring). Due to page-length constraints, we do not include the full verbatim JSON prompts and low-level API/solver parameters; however, the prompt structure and all decision criteria (intersection/entropy-peak rules and thresholds) are fully described, providing a clear path to reproduce the main experimental trends and conclusions.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While AI Co-Scientist Challenge Korea does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: **[No]**

Justification: The raw URANS simulation outputs and derived datasets are prohibitively large in size, making open release impractical. Moreover, the data are generated using standard CFD solvers and post-processing pipelines rather than a bespoke benchmark. Instead, we provide detailed descriptions of the flow configuration, Reynolds numbers, numerical setup, and data-processing steps, which allow interested researchers to reproduce analogous datasets or verify the reduced-order analysis at a conceptual and methodological level. In addition, while we do not publicly release our implementation at submission time, the main analysis is driven by a small set of self-contained scripts (PCA fitting/encoding–decoding, latent time-series construction, one-step forecasting with regularized linear regression, and the LLM querying/entropy scoring routine). These scripts follow standard Python scientific-stack dependencies and can be made available upon request or in a post-acceptance release, subject to contest policy and data-sharing constraints. We also provide sufficient algorithmic detail (inputs/outputs, preprocessing, train/test protocol, and decision rules) for independent re-implementation without access to our code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The experimental settings for the learning and evaluation tasks are explicitly defined in the manuscript. The construction of the latent-space time series, the training–test split, the temporal history length, normalization procedure, and evaluation metrics are clearly specified. The learning models employed are regularized linear predictors (Elastic Net, Ridge, and Lasso), which are deterministic and require only a small number of transparent hyperparameters. This avoids ambiguity associated with stochastic optimization or complex training protocols and ensures that the reported results can be readily interpreted and reproduced from the provided description.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The study does not rely on stochastic learning procedures or repeated randomized experiments for which classical statistical significance testing would be appropriate. The URANS simulations, PCA compression, and latent-space prediction models (Elastic Net, Ridge, and Lasso) are deterministic given the input data and model parameters. As a result, repeated trials yield identical outcomes, and variability-based statistical tests are not meaningful in this context. Instead, model performance is assessed through physically interpretable error metrics, reconstruction bounds, and systematic comparisons across Reynolds numbers, which directly reflect predictive fidelity and generalization behavior. In contrast, the LLM-based regime identification exhibits sampling variability across repeated API calls; thus, for Fig. 8 we report error bars as $\pm 1\sigma$ computed from $K = 10$ realizations per Reynolds number for both the one-hot encoding probabilities and the corresponding entropy values.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).

- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [No]

Justification: Justification: The manuscript does not report compute resources (hardware, runtime, and memory) for URANS simulations, preprocessing, regression training, and LLM inference, preventing readers from estimating reproduction cost.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://nips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The study uses simulated CFD data and standard regression/LLM inference without human-subject experiments or sensitive personal data, and it does not involve deceptive or harmful data collection.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The paper focuses on methodological contributions for reduced-order CFD analysis and does not explicitly discuss potential positive and negative societal impacts or mitigation strategies.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The work does not release a high-risk pretrained model or scraped dataset; it uses simulation data and commercial LLM endpoints for inference, so release-time safeguards are not applicable.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The URANS flow fields are generated using OpenFOAM, which is open-source software released under the GNU General Public License v3 (GPLv3), and is used in compliance with its license (<https://openfoam.org/licence/>). We use standard open-source Python packages for data processing and modeling (e.g., NumPy, SciPy, scikit-learn, and Matplotlib) and comply with their respective licenses. For LLM-assisted regime identification, we access closed-source models only through the providers' official APIs and comply with their applicable terms of use and usage policies; we do not redistribute any proprietary model weights or restricted assets. Specifically, the experimental inference models used via API are gpt-4.1-mini (<https://platform.openai.com/docs/models/gpt-4.1-mini>), gemini-2.5-flash (<https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-flash>), and claude-haiku-4-5 (<https://www.anthropic.com/news/claude-haiku-4-5>). In addition, GPT-5.2 and EXAONE-4.01-32B were used for research assistance (ideation/coding/writing) under their respective access terms, without redistributing weights or restricted content. Experiments and analyses were executed in Google Colab, following Colab's applicable terms (<https://colab.research.google.com/terms>).

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New Assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: While a URANS dataset is generated for this study, no new assets are released alongside the submission; therefore, documentation bundled with released assets is not applicable.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and Research with Human Subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No human-subject research is conducted; therefore, participant risk disclosure and IRB approval are not applicable.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.