

Skeptical review: Challenges in Data-Driven Equation Discovery: A Case Study of a 3D Fluid System with Limited Temporal Resolution

Summary

This manuscript presents a negative-results case study on data-driven PDE/equation discovery for a 3D periodic, fluid-like dataset where only 10 temporal snapshots are available on a 128^3 grid for four fields (ρ , u_x , u_y , u_z) (Sec. 2.1, Sec. 3.1). The authors construct a fixed library of 34 candidate features spanning base variables, spatial derivatives (∇ , $\nabla \cdot$, $\nabla \times$, Δ), advective terms ($u \cdot \nabla \rho$ and componentwise $u \cdot \nabla u$), and polynomial/inverse combinations (Sec. 2.2.1), approximate time derivatives by finite differences with nominal $\Delta t = 1$ (Sec. 2.2.2), and fit separate sparse linear models with LassoCV (Sec. 2.3). The discovered equations are dominated by algebraic terms, with advective/diffusive operators largely absent or assigned very small coefficients (Sec. 3.3), and achieve very low predictive quality ($R^2 \leq 0.11$) with predicted derivative fields that are overly smooth compared to the reference derivatives (Sec. 3.4). The paper argues the failures stem primarily from extreme temporal undersampling/aliasing (large effective CFL), compounded by missing pressure-gradient information, feature collinearity, and the fluctuation-like nature of ρ (Sec. 3.5, Conclusions). The case study is potentially valuable as a cautionary example, but the current version does not yet isolate which failure mechanisms are empirically responsible (undersampling vs. library/model mismatch vs. evaluation/protocol artifacts), and several key methodological/reproducibility details are underspecified (Sec. 2–3).

Strengths

- Clear, coherent narrative centered on a practically important regime for PDE discovery: extremely sparse temporal sampling in a 3D flow-like system (Introduction, Sec. 3.5, Conclusions).
- Transparent reporting of negative results with explicit learned equations, quantitative R^2 values, and qualitative derivative-field comparisons that make the failure modes easy to see (Sec. 3.3–3.4).
- Reasonably rich, physically motivated feature library including spatial derivatives and advective terms (Sec. 2.2.1).
- Exploratory data analysis identifies a strong z-directed mean flow and interprets other components largely as fluctuations, helping contextualize why simple regression may struggle (Sec. 3.1).
- The discussion identifies several plausible failure mechanisms (temporal undersampling, missing pressure/projection, collinearity), which—if better substantiated—could yield broadly useful guidance (Sec. 3.5).

Major issues

1. **The data-generating physical/simulation model is insufficiently specified, preventing readers from knowing what equations/terms are actually “recoverable” (Sec. 2.1, Sec. 3.1).** The manuscript describes a “three-dimensional periodic fluid system” but does not clearly state the governing PDEs, whether the velocity is incompressible, the role/definition of ρ (density vs signed fluctuation vs tracer), boundary conditions, forcing, parameters (e.g., Reynolds/Péclet), numerical scheme, and (critically) the actual simulation time step and output cadence. Without this, it is difficult to evaluate claims about missing pressure gradients, expected advective/diffusive balance, or the severity of temporal undersampling.

Recommendation: Expand Sec. 2.1 (or add a dedicated subsection) to document the dataset provenance and the data-generating model: explicit PDEs (even if only “believed/assumed”), boundary conditions (periodic), forcing/initial conditions (as known), nondimensional parameters, and the true simulation time step Δt_{sim} and snapshot interval Δt_{snap} . If the dataset is external and details are unknown, state what is unknown explicitly, cite the source, and reframe subsequent claims (pressure/CFL) as conditional hypotheses rather than established facts.

2. **The core claim—temporal undersampling/aliasing dominates failure—is not quantitatively demonstrated, and the “ $\Delta t = 1$ ” assumption is treated too literally (Sec. 2.2.2, Sec. 3.5, Conclusions).** With only 10 snapshots, finite-difference ∂_t estimates can be dominated by truncation error and/or noise; moreover, if the physical time axis is unknown, scaling directly affects the regression targets and regularization path. The $\text{CFL} \approx 128$ argument is currently presented without a tight link to the true sampling interval and velocity scale.

Recommendation: In Sec. 2.2.2 and Sec. 3.5: (i) clearly distinguish “index time step” from physical/nondimensional time; if physical Δt is unknown, present results in a scale-free way and treat Δt as an unknown factor; (ii) provide a derivation of the effective CFL (or a dimensionless sampling ratio $\Delta t_{\text{snap}}/\tau_{\text{char}}$) using reported Δx , a characteristic velocity (mean/max $|u|$), and Δt_{snap} ; and (iii) add quantitative derivative-diagnostic evidence, e.g., compare central vs one-sided vs higher-order differences, temporal smoothing (Savitzky–Golay/local polynomial in time), or regularized differentiation, and report how the distribution/variance of $\partial_t f$ and downstream R^2 change. Even a small sensitivity table would substantially strengthen the “undersampling is primary” conclusion.

3. **Causal attribution is confounded: library inadequacy (pressure/projection), regression/model-class limitations, and derivative error are discussed together but not empirically disentangled (Sec. 2.2.1, Sec. 3.5).** In particular, the manuscript highlights “missing pressure gradient” as a major limitation for momentum equations, yet does not test whether adding pressure-like information (or

using a formulation that removes pressure) changes the outcome; similarly, it is unclear whether incompressibility holds ($\nabla \cdot \mathbf{u} \approx 0$) and whether a vorticity/projection-based equation discovery would be more appropriate.

Recommendation: Re-structure Sec. 3.5 to separate hypotheses and support each with at least one diagnostic/ablation: (a) check and report incompressibility/divergence statistics (distribution of $\nabla \cdot \mathbf{u}$, relative magnitude vs $|\nabla \mathbf{u}|$); (b) if pressure is available, add ∇p to the library and rerun; if not available, explicitly label the pressure discussion as a hypothesis and add a proxy experiment: regress the vorticity equation (pressure-free) or apply a solenoidal projection/Helmholtz decomposition in Fourier space (natural for periodic domains) before regression; (c) for ρ , explicitly test a passive-scalar advection–diffusion-only library ($\mathbf{u} \cdot \nabla \rho$, $\Delta \rho$, optional source) to see whether “algebraic dominance” persists.

4. **The experimental protocol and evaluation are under-specified and may be optimistic/unstable given strong spatiotemporal correlations (Sec. 2.3, Sec. 3.4).** The manuscript does not clearly state whether R^2 is computed in-sample, on held-out points, or via cross-validation; how train/test splitting is performed given only 10 time slices; and whether results are stable across random subsamples/seeds. Random pointwise splits can leak information due to spatial correlation and may not reflect generalization across time.

Recommendation: In Sec. 2.3.3 and Sec. 3.4: (i) specify the exact split protocol (by time slice holdout is strongly preferred here; alternatively spatial-block holdout), and report both train and test R^2 ; (ii) report variability across multiple random seeds/subsamples (mean \pm std of selected terms and R^2); and (iii) add at least one complementary diagnostic beyond pointwise R^2 , such as RMSE/MAE and a spectral comparison (power spectrum of predicted vs true ∂_t fields) to quantify the observed over-smoothing. If feasible, attempt a short-horizon forward integration/rollout from one snapshot (even if it fails), and report what fails (instability vs drift vs loss of small scales).

5. **Methodological breadth is too narrow to support broad conclusions about “data-driven equation discovery” under sparse temporal sampling (Sec. 2.3, Sec. 3.2–3.4).** The negative result is based essentially on one pipeline (pointwise feature library + LassoCV). It remains unclear whether the failure is intrinsic to the data regime or specific to LassoCV, scaling choices, or the pointwise strong-form formulation.

Recommendation: Add one controlled comparison (minimal but informative): e.g., STLSQ/SINDy thresholding, ridge + pruning, or a weak-form/WSINDy variant that reduces derivative sensitivity; and/or a Fourier/spectral formulation leveraging periodicity. Report whether the same qualitative pathology (algebraic-term dominance; near-zero advection/diffusion; low generalization) persists. This can be done as a small ablation in Sec. 3.2–3.4 without reframing the entire paper.

6. **A “positive control” is missing, so it is hard to know whether the pipeline is correctly implemented and whether the observed failure threshold is specific to this dataset (Sec. 3.5, Conclusions).** As written, the paper implicitly generalizes from a single case with unknown governing equations and potentially missing key fields (pressure).

Recommendation: Include a compact validation experiment: generate (or cite) a known PDE system on a periodic grid (e.g., 3D/2D advection–diffusion for a scalar, or a simpler fluid surrogate), run the same feature construction and regression, and then downsample in time to 10 snapshots. Show how identification quality degrades as snapshots are reduced, ideally isolating the “10 snapshots” regime. This would directly substantiate the central message while keeping the current dataset as the main case study.

7. **Feature scaling, nondimensionalization, and collinearity are not analyzed rigorously, yet they are central to interpreting the learned equations (Sec. 2.2.1, Sec. 2.3, Sec. 3.3–3.5).** The manuscript notes standardization and “unstandardization,” but does not provide the exact mapping (including intercept handling), and dimensional consistency of mixed terms (constants, u , u^2 , derivative terms) is not verifiable. Collinearity is argued qualitatively (e.g., u_{mag} vs component squares) without quantitative diagnostics.

Recommendation: In Sec. 2.3 and Sec. 3.5: (i) state clearly whether variables/features are nondimensional and/or normalized before feature construction; (ii) provide the exact coefficient unstandardization formula (including intercept adjustment); (iii) quantify collinearity (feature correlation matrix for key groups, condition number/VIF), and test a small decorrelation ablation (e.g., remove u_{mag} if u_x^2, u_y^2, u_z^2 are present; or separate mean + fluctuations $u = \bar{u} + u'$ and build features on fluctuations). Report whether term selection and R^2 change materially.

8. **Reproducibility is limited by missing details on data access, preprocessing, derivative computation, and LassoCV configuration (Sec. 2.1–2.3).** Given the paper’s value as a cautionary benchmark, the community benefit depends heavily on being able to replicate and extend the study.

Recommendation: Strengthen Sec. 2.1–2.3 with: dataset availability/source link (or explicit constraints), preprocessing steps (precision, periodic derivative handling, any filtering), exact subsampling strategy for the 100,000 points (uniform/stratified over time), random seeds, and full LassoCV settings (folds, alpha grid, max_iter, tol, normalization). If possible, release code and either the dataset or a small subset plus scripts to rebuild derivatives/features.

Minor issues

1. The feature library definition is not fully explicit, making the reported “34 features” hard to verify and reproduce (Sec. 2.2.1, Sec. 3.2). Several groups are described with “such as ...” rather than enumerated, and inverse-feature definitions are ambiguous.

Recommendation: Provide a table (main text or appendix) listing all 34 features exactly as implemented, grouped by type. For inverse terms, state the exact formula (e.g., $1/(\rho + \epsilon)$ vs $1/(\rho^2 + \epsilon)$, whether $|\rho|$ is used, and where ϵ is inserted). Ensure the total count matches 34.

2. The pointwise strong-form regression formulation (flattening all space-time points as i.i.d. samples) may overweight spatial variation relative to temporal information, especially with only 10 times and a strong mean flow (Sec. 2.3, Sec. 3.1, Sec. 3.4). This interacts with the interpretation that the method “learns algebraic surrogates.”

Recommendation: Add a brief diagnostic in Sec. 3.4 quantifying how much variance in $\partial_t f$ is across space vs across time (e.g., per-time-slice means/variances), and discuss how this affects regression. Consider (even as a small supplementary test) a time-slice-wise regression or averaging/weak-form approach that reduces sensitivity to pointwise noise.

3. Spatial derivative computation choices are not well motivated given the periodic domain (Sec. 2.2.1). For periodic grids, spectral derivatives can be far more accurate than finite differences, and derivative accuracy matters when comparing advection/diffusion terms.

Recommendation: State whether derivatives are finite-difference or spectral, and why. If finite differences are used, report the stencil order and confirm periodic wrap-around. If feasible, add a small comparison (FD vs spectral) for derivative statistics and whether it changes selected terms/ R^2 .

4. Sample size and sampling strategy are not analyzed for sensitivity (Sec. 2.3). Using 100,000 points out of $\sim 10 \times 128^3$ can be fine, but results may depend on stratification across the 10 times and on spatial correlation.

Recommendation: Report whether sampling is uniform across time slices (recommended) and add a small sensitivity check: e.g., 50k/100k/200k points, with mean \pm std of R^2 and selected terms across seeds.

5. Figure 3 (and related qualitative comparisons) is central to the “overly smooth derivative” claim, but presentation choices reduce interpretability: sequential colormap around zero, unclear shared color limits across time, and limited panel metadata (Sec. 3.1–3.4, Fig. 3).

Recommendation: Use a diverging colormap centered at $\mathbf{0}$ for fluctuation fields (ρ , u_x , u_y ; and for u_z consider plotting $u_z - \bar{u}_z$). State whether colorbar limits are shared across time per variable and annotate slice plane, time ordering, and units/nondimensionalization. If space is tight, split Fig. 3 by variable or provide a higher-resolution supplement.

6. Terminology intermittently blurs “symbolic regression” and “sparse regression over a fixed library” (Sec. 1, Sec. 2.3). This matters for positioning relative to prior work and for reader expectations.

Recommendation: Tighten terminology: describe the method consistently as sparse linear regression / SINDy-style library regression (with LassoCV), and reserve “symbolic regression” for methods that search expressions beyond a fixed library.

7. Scope of conclusions could be stated more carefully given this is one dataset with unknown/partially specified physics and potentially missing key variables (pressure) (Sec. 3.5, Conclusions).

Recommendation: In Sec. 3.5 and Conclusions, explicitly separate (i) empirically demonstrated findings in this dataset (low R^2 , algebraic dominance) from (ii) hypothesized causes (pressure, CFL/aliasing), and clarify which lessons are expected to generalize vs which are specific to this configuration.

Very minor issues

1. Endpoint temporal-derivative formulas are referenced but not written explicitly (forward at t_0 , backward at t_9), making target construction harder to verify (Sec. 2.2.2).

Recommendation: Add the explicit one-sided difference formulas used at the first/last snapshots (including order) and clarify how endpoint rows are stacked/aligned with features.

2. The manuscript contains small typography/notation inconsistencies (e.g., 128^3 vs 128^3 , occasional LaTeX glitches in velocity subscripts, stray heading symbols) (Sec. 1, Sec. 2.1, Sec. 3.1, figure captions).

Recommendation: Proofread for consistent math typesetting (always 128^3 or $128 \times 128 \times 128$; consistent u_x, u_y, u_z, ρ), remove stray symbols in headings, and fix minor spacing/line-break artifacts.

3. In Sec. 3.3, equations use ellipses (“+ ...”) without a precise statement of omission criteria, which can obscure what was actually selected at nonzero weight.

Recommendation: State what “...” represents (e.g., terms below a coefficient threshold) and provide a complete coefficient table for each learned equation in an appendix or supplement.

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper contains definitions of engineered PDE-inspired features (gradients, divergence, curl, Laplacians, advective terms), a central-difference temporal derivative formula, and several discovered sparse linear equations. There are no multi-step symbolic derivations; most mathematics is definitional. Core checks focus on operator definitions, finite-difference correctness, symbol consistency, and whether equations have consistent dimensions given the stated feature types.

Checked items

1. ✓ **Data tensor and variable mapping** (Sec. 2.1, p.2)
 - **Claim:** Data has shape $(10, 4, 128, 128, 128)$ with variables (ρ, u_x, u_y, u_z) on a periodic cube of side $L = 1$.
 - **Checks:** symbol/definition consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Indices correspond to (time, variable, x, y, z) ordering as stated.
 - **Notes:** Variable naming and dimensional description are consistent internally.
2. ✓ **Divergence definition** (Sec. 2.2.1, p.3)
 - **Claim:** $\nabla \cdot u = \partial u_x / \partial x + \partial u_y / \partial y + \partial u_z / \partial z$.
 - **Checks:** algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** $u = (u_x, u_y, u_z)$.
 - **Notes:** Standard divergence definition; consistent with later use of u components.
3. ✓ **Curl components definition** (Sec. 2.2.1, p.3)
 - **Claim:** $(\nabla \times u)_x = \partial u_z / \partial y - \partial u_y / \partial z$, $(\nabla \times u)_y = \partial u_x / \partial z - \partial u_z / \partial x$, $(\nabla \times u)_z = \partial u_y / \partial x - \partial u_x / \partial y$.
 - **Checks:** algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Right-handed (x, y, z) coordinate system.
 - **Notes:** Component formulas are internally consistent and correctly patterned.
4. ✓ **Laplacian definition for scalar** (Sec. 2.2.1, p.3)

- **Claim:** $\nabla^2 \rho = \partial^2 \rho / \partial x^2 + \partial^2 \rho / \partial y^2 + \partial^2 \rho / \partial z^2$ (and analogously for velocity components).
 - **Checks:** algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Cartesian coordinates.
 - **Notes:** Standard Laplacian; consistent with described computation.
5. ✓ **Advective term $u \cdot \nabla \rho$** (Sec. 2.2.1, p.3)
- **Claim:** $u \cdot \nabla \rho = u_x \partial \rho / \partial x + u_y \partial \rho / \partial y + u_z \partial \rho / \partial z$.
 - **Checks:** algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Dot product between velocity and gradient.
 - **Notes:** Correct expansion.
6. ✓ **Componentwise advective acceleration $u \cdot \nabla u_x$ (example)** (Sec. 2.2.1, p.3)
- **Claim:** For $u \cdot \nabla u$, a component example is $u_x \partial u_x / \partial x + u_y \partial u_x / \partial y + u_z \partial u_x / \partial z$.
 - **Checks:** algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** $u \cdot \nabla$ acts as directional derivative along u .
 - **Notes:** Correct directional derivative for the x-component.
7. ✓ **Velocity magnitude definition** (Sec. 2.2.1, p.3)
- **Claim:** $u_{\text{mag}} = \sqrt{u_x^2 + u_y^2 + u_z^2}$.
 - **Checks:** algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Euclidean norm.
 - **Notes:** Consistent with later references to u_{mag} .
8. ✓ **Central-difference temporal derivative** (Sec. 2.2.2, p.3)
- **Claim:** For interior times: $(\partial f / \partial t)(t_i) \approx (f(t_{i+1}) - f(t_{i-1})) / (2\Delta t)$.
 - **Checks:** algebra, definition consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Uniform time step Δt .
 - **Notes:** Correct second-order centered finite difference for first derivative.
9. ⚠ **Endpoint temporal derivatives (forward/backward) unspecified** (Sec. 2.2.2, p.3)
- **Claim:** Uses forward difference at t_0 and backward difference at t_9 .
 - **Checks:** derivation completeness
 - **Verdict:** UNCERTAIN; confidence: high; impact: minor

- **Assumptions/inputs:** One-sided schemes are applied at endpoints.
 - **Notes:** The exact formulas (order, denominators, and whether higher-order one-sided stencils are used) are not stated, so symbolic verification at endpoints is not possible.
10. \triangle **Use of nominal $\Delta t = 1$** (Sec. 2.2.2, p.3 and Sec. 3.2, p.7)
- **Claim:** Assumes a nominal time step $\Delta t = 1$ for temporal derivative calculations.
 - **Checks:** definition consistency, dimensional/units sanity-check
 - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Time is in arbitrary units or already nondimensional.
 - **Notes:** Internally consistent as a computational convention, but without stated nondimensionalization/units it is unclear whether mixing $\Delta t = 1$ with spatial scaling ($L = 1, N = 128$) yields meaningful dimensional consistency for PDE discovery.
11. \triangle **Inverse-feature definition with epsilon** (Sec. 2.2.1, p.3)
- **Claim:** Includes $1/\rho$ and $1/\rho^2$ with $\epsilon = 1 \times 10^{-6}$ added to the denominator to prevent division by zero.
 - **Checks:** definition consistency
 - **Verdict:** UNCERTAIN; confidence: high; impact: minor
 - **Assumptions/inputs:** A specific analytic form is implemented.
 - **Notes:** Ambiguous whether features are $1/(\rho + \epsilon)$, $1/(\rho^2 + \epsilon)$, $1/(|\rho| + \epsilon)$, etc. Exact analytic forms matter for symbolic interpretability.
12. \checkmark **Feature matrix/target vector construction** (Sec. 2.3.1, p.4)
- **Claim:** Flatten spatial grids per feature and concatenate across time into X ; similarly stack temporal derivatives into Y .
 - **Checks:** symbol/definition consistency
 - **Verdict:** PASS; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Row alignment between X and Y is consistent per spatio-temporal point.
 - **Notes:** Construction is described consistently at a high level; however, exact indexing/alignment is not shown.
13. \triangle **Standardization then 'unstandardization' of coefficients** (Sec. 2.3.1, p.4 and Sec. 3.3, p.7)
- **Claim:** Features are standardized to zero mean/unit variance before training; reported coefficients are unstandardized to original feature space.
 - **Checks:** derivation completeness, notation consistency

- **Verdict:** UNCERTAIN; confidence: high; impact: minor
 - **Assumptions/inputs:** A standard linear rescaling and intercept adjustment is applied.
 - **Notes:** No explicit formula is given for mapping standardized coefficients/intercept back to original units, so the reported equations cannot be symbolically verified against the stated preprocessing.
14. ✓ **Discovered equation notation consistency (example: density equation)** (Sec. 3.3, p.7)
- **Claim:** One learned model is $\partial\rho/\partial t = -0.017 + 0.091 u_{\text{mag}} - 0.074 u_z^2 - 0.043 u_x^2 - 0.042 u_y^2 + 0.018 \rho u_x - 0.008 (u \cdot \nabla u_z) + \dots$
 - **Checks:** symbol/definition consistency
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** u_{mag} and $u \cdot \nabla u_z$ are among the engineered features as defined earlier.
 - **Notes:** All symbols appearing ($u_{\text{mag}}, u_x, u_y, u_z, \rho, u \cdot \nabla u_z$) were defined in the feature library section.
15. △ **Dimensional homogeneity of discovered equations** (Sec. 3.3, p.7)
- **Claim:** The RHS combines constants, velocities, velocity squares, products ρu_x , and derivative-based features to model $\partial\rho/\partial t$ or $\partial u/\partial t$.
 - **Checks:** dimensional/units sanity-check
 - **Verdict:** UNCERTAIN; confidence: high; impact: critical
 - **Assumptions/inputs:** Variables might be nondimensionalized, but this is not stated.
 - **Notes:** Without an explicit nondimensionalization/scaling statement, terms like u_z^2 and $u \cdot \nabla u_z$ generally carry different dimensions than $\partial\rho/\partial t$ or $\partial u/\partial t$, so analytic dimensional consistency cannot be confirmed.
16. ✓ **CFL number estimate** (Sec. 3.5 item 1, p.8 (and Sec. 2.1, p.2; Sec. 3.1.1, p.4-5))
- **Claim:** Effective CFL number is approximately 128.
 - **Checks:** algebra, dimensional/units sanity-check
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** $\Delta x = L/128$ with $L = 1$, Characteristic speed ≈ 1 (from u_z mean ≈ 1), $\Delta t = 1$
 - **Notes:** $\text{CFL} \approx |u|\Delta t/\Delta x \approx 1 \times 1/(1/128) = 128$, consistent with the stated grid and nominal timestep.

Limitations

- Audit is based on the provided PDF text content; many equations are not numbered and some model equations are truncated with ellipses, limiting what can be checked symbolically.

- The paper does not provide an explicit complete list of the 34 features nor the explicit coefficient unstandardization mapping, preventing full analytic verification of the regression model expressions.
- No nondimensionalization/units specification is given for variables and derived features, blocking a definitive dimensional-homogeneity check for the discovered governing equations.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

Arithmetic and logical consistency checks across dataset sizing, feature-library bookkeeping, derivative indexing, reported R^2 values, summary-statistic range containment, and selected equation coefficient relationships all passed. Several additional numeric claims (e.g., CFL estimate and some distributional/figure-based assertions) could not be verified from the provided text-only numerals.

Checked items

1. ✓ **C1_dataset_shape_product** (Sec. 2.1 Dataset description (page 2); also Abstract/Intro mentions 10 time slices and 128^3 grid)
 - **Claim:** Dataset dimensions are $(10, 4, 128, 128, 128)$, representing 10 time steps, 4 variables, and a $128 \times 128 \times 128$ spatial grid.
 - **Checks:** shape_consistency_and_count
 - **Verdict:** PASS
 - **Notes:** Recomputed $\text{spatial_points}=128^3 = 2,097,152$ and $\text{total_elements}=10 \times 4 \times 128^3 = 83,886,080$ from the stated shape; consistent with the provided dimensions.
2. ✓ **C2_spatiotemporal_points_total** (Sec. 2.3.1 Feature matrix construction and scaling (page 4))
 - **Claim:** Large dataset described as (10×128^3) spatio-temporal points).
 - **Checks:** integer_recomputation
 - **Verdict:** PASS
 - **Notes:** Recomputed $128^3 = 2,097,152$ and $10 \times (128^3) = 20,971,520$ spatio-temporal points; matches the implied total.
3. ✓ **C3_subsample_fraction** (Sec. 2.3.1 Feature matrix construction and scaling (page 4))
 - **Claim:** A random subsample of 100,000 spatio-temporal points was extracted for model training from the large dataset (10×128^3) points).
 - **Checks:** ratio_recomputation
 - **Verdict:** PASS

- **Notes:** Computed subsample fraction = $100,000/20,971,520 = 0.0047683716$ (0.476837%). No stated target fraction to verify.
4. ✓ **C4_feature_count_sum_to_34** (Sec. 2.2.1 Spatial derivatives and derived quantities (page 3))
- **Claim:** In total, a library of 34 candidate features was constructed, consisting of listed groups (base variables, gradients, divergence, curl, Laplacians, advective terms, algebraic combinations, inverse terms).
 - **Checks:** component_count_recomputation
 - **Verdict:** PASS
 - **Notes:** Summed explicitly enumerated group counts: $4 + 3 + 1 + 3 + 4 + 2 = 17$; remainder to reach 34 is 17 (nonnegative), consistent with additional unspecified feature groups accounting for the balance.
5. ✓ **C5_eps_value** (Sec. 2.2.1 Spatial derivatives and derived quantities (page 3))
- **Claim:** Inverse terms use a small epsilon ($\epsilon = 10^{-6}$) added to the denominator to prevent division by zero.
 - **Checks:** scientific_notation_parse
 - **Verdict:** PASS
 - **Notes:** Parsed 10^{-6} as 1×10^{-6} ; matches expected epsilon value.
6. ✓ **C6_time_indexing_and_dt** (Sec. 2.2.2 Temporal derivatives (page 3))
- **Claim:** For ten time points $t_0..t_9$, interior points use central difference; endpoints use forward/backward difference; nominal $\Delta t = 1$.
 - **Checks:** indexing_consistency
 - **Verdict:** PASS
 - **Notes:** Indexing implies $n = 9 - 0 + 1 = 10$ time slices; interior points = $n - 2 = 8$; consistent with the stated setup. Δt parsed as 1.
7. ✓ **C7_R2_all_below_0_11** (Abstract (page 1) and Sec. 3.4 (page 8) and Conclusions (page 9))
- **Claim:** R^2 scores are consistently below 0.11 for all variables; listed R^2 values are 0.068, 0.106, 0.107, 0.019.
 - **Checks:** threshold_check
 - **Verdict:** PASS
 - **Notes:** Computed max $R^2 = 0.107$, which is strictly less than 0.11 as claimed.
8. ✓ **C8_R2_summary_stats** (Sec. 3.4 (page 8))
- **Claim:** Four R^2 scores are reported: 0.068, 0.106, 0.107, 0.019.
 - **Checks:** cheap_aggregate_recompute
 - **Verdict:** PASS

- **Notes:** Computed summary stats for reference (no stated aggregate to verify): min= 0.019, max= 0.107, mean= 0.075, median= 0.087 ($n = 4$).
9. ✓ **C9_uz_range_contains_mean** (Sec. 3.1.1 Statistical moments and distributions (page 5))
- **Claim:** u_z mean is 1.000 and ranges from 0.983 to 1.007.
 - **Checks:** range_contains_value
 - **Verdict:** PASS
 - **Notes:** Verified $0.983 \leq 1.000 \leq 1.007$.
10. ✓ **C10_uz_std_vs_range_width** (Sec. 3.1.1 Statistical moments and distributions (page 5))
- **Claim:** u_z has std dev 0.002 and ranges from 0.983 to 1.007.
 - **Checks:** range_width_vs_std_sanity
 - **Verdict:** PASS
 - **Notes:** Heuristic sanity metrics: range width= 0.0240, half-range= 0.0120, max |dev from mean| = 0.0170; ratios: half_range/std \approx 6.0 and max_abs_dev/std \approx 8.5 (std does not exceed range).
11. ✓ **C11_ux_uy_means_near_zero_compare** (Sec. 3.1.1 Statistical moments and distributions (page 5))
- **Claim:** u_x mean is 9.89×10^{-6} and u_y mean is 3.70×10^{-5} (near zero).
 - **Checks:** scientific_notation_parse_and_ratio
 - **Verdict:** PASS
 - **Notes:** Parsed means as $u_x = 9.89 \times 10^{-6}$ and $u_y = 3.70 \times 10^{-5}$; computed $|u_y|/|u_x| \approx 3.7412$ (reference only; no exact claimed ratio).
12. ✓ **C12_rho_range_contains_mean** (Sec. 3.1.1 Statistical moments and distributions (page 5))
- **Claim:** ρ mean is -3.74×10^{-5} and ranges from -0.773 to 0.752.
 - **Checks:** range_contains_value
 - **Verdict:** PASS
 - **Notes:** Verified $-0.773 \leq -3.74 \times 10^{-5} \leq 0.752$.
13. ✓ **C13_rho_range_width** (Sec. 3.1.1 Statistical moments and distributions (page 5))
- **Claim:** ρ ranges from -0.773 to 0.752.
 - **Checks:** difference_recompute
 - **Verdict:** PASS
 - **Notes:** Recomputed range width: $0.752 - (-0.773) = 1.525$; matches the stated computation.

14. ✓ **C14_equation_duy_constant_matches_coeff_u2z** (Sec. 3.3 Discovered governing equations (page 7))
- **Claim:** For $\partial u_y / \partial t$: $-0.352 + 0.352 u_z^2 + \dots$ (same magnitude coefficients).
 - **Checks:** coefficient_equality
 - **Verdict:** PASS
 - **Notes:** Confirmed $\text{coeff}(u_z^2) = 0.352$ equals $-(\text{constant term}) = 0.352$.
15. ✓ **C15_equation_duz_constant_matches_coeff_u2z** (Sec. 3.3 Discovered governing equations (page 7))
- **Claim:** For $\partial u_z / \partial t$: $0.002 - 0.002 u_z^2 + \dots$ (same magnitude coefficients).
 - **Checks:** coefficient_equality
 - **Verdict:** PASS
 - **Notes:** Confirmed $\text{coeff}(u_z^2) = -0.002$ equals $-(\text{constant term}) = -0.002$.
16. ✓ **C16_equation_duz_scientific_coeff_parse** (Sec. 3.3 Discovered governing equations (page 7))
- **Claim:** In $\partial u_z / \partial t$ equation, coefficient -6.4×10^{-5} multiplies u_y^2 .
 - **Checks:** scientific_notation_parse
 - **Verdict:** PASS
 - **Notes:** Parsed coefficient -6.4×10^{-5} as -6.4×10^{-5} ; consistent.

Limitations

- Checks are restricted to arithmetic/logical consistency among numerals explicitly stated in the provided PDF text; underlying NumPy data and trained model artifacts are not available.
- No quantitative extraction from figures/heatmaps/scatter plots is performed (image-based value reading is out of scope).
- Many scientific claims (e.g., CFL estimate, turbulence interpretation, regression failure causes) depend on missing inputs or computations not fully specified in the text and are therefore listed as unverified.