

# *Skeptical review: Data-Driven Discovery of Governing Equations for a 3D Fluid System: Addressing Feature Collinearity in Sparse Regression*

---

## Summary

The manuscript presents an end-to-end, SINDy-style sparse-regression workflow to discover PDE representations from 3D periodic fluid data consisting of three velocity components and a density field on a  $128^3$  grid over 10 time snapshots (Sec. 1, Sec. 2.1–2.6). The pipeline includes exploratory data analysis, spectral low-pass filtering, spectral spatial differentiation, finite-difference temporal differentiation, construction of a large candidate feature library (including both composite operators such as advection/Laplacian/divergence/curl and their constituent primitive terms), iterative thresholding sparse regression fit independently per field, and validation via (i) held-out local derivative prediction and (ii) one-step forward integration with RK4 and adaptive substepping (Sec. 2.3–2.6, Sec. 3.1–3.3). A central theme is that severe feature collinearity—partly induced by the library design—creates strong non-identifiability, allowing “null-space exploitation” where large, opposite-signed coefficients cancel, yielding non-parsimonious and hard-to-interpret equations while still giving strong short-horizon predictive performance (Sec. 3.2–3.2.1). The study is a useful cautionary case: it convincingly demonstrates high one-step predictive accuracy, but currently underspecifies key experimental/reproducibility details and overreaches in places on claims about learning a unique/true governing PDE and long-horizon dynamical fidelity given the demonstrated non-identifiability and the limited temporal data window (Sec. 3.3.2, Sec. 4).

## Strengths

- Clear, well-organized exposition of a full PDE-discovery workflow from raw 3D space-time data through preprocessing, feature construction, sparse regression, and validation (Sec. 2–3).
- Useful and timely case study highlighting identifiability failure under redundant PDE libraries; the null-space framing ( $\xi = \xi_{\text{true}} + \xi_{\text{null}}$ ,  $\Theta\xi_{\text{null}} = 0$ ) is conceptually correct and practically important (Sec. 3.2–3.2.1).
- Appropriate use of spectral methods for periodic domains: low-pass filtering and Fourier differentiation are well-motivated for stabilizing derivative estimation (Sec. 2.3–2.4, Sec. 3.1).
- Validation includes both local derivative prediction and forward integration, with quantitative metrics and qualitative field comparisons that demonstrate strong short-horizon predictive capability (Sec. 3.3).
- The manuscript explicitly discusses trade-offs between parsimony, interpretability, and predictive accuracy—an important message for practitioners deploying sparse regression for PDE discovery (Sec. 4).

1. **Identifiability vs. prediction is not cleanly separated in the claims and narrative.** The paper correctly demonstrates severe non-uniqueness under collinearity (null-space exploitation) (Sec. 3.2–3.2.1), but still frequently refers to “the discovered PDEs,” “true dynamical operator,” and implies physical identification. Given redundancy and scaling/centering effects, the output is better described as learning a predictive operator  $f(u) \approx \partial_t u$  on the data manifold, not a uniquely identified physical PDE decomposition (Sec. 3.3.2, Sec. 4).

*Recommendation:* Reframe key statements in Sec. 3.3.2 and Sec. 4 to explicitly distinguish (i) learning a predictive right-hand-side representation  $\Theta\xi$  that matches  $\partial_t u$  on the sampled data, from (ii) identifying a unique, physically interpretable PDE form. Add an explicit statement of an equivalence class: many  $\xi$  yield nearly identical  $\Theta\xi$  due to collinearity; therefore, individual coefficients/term selections are not identifiable without additional constraints. Moderate wording accordingly throughout (Abstract, Sec. 3.3.2, Sec. 4).

2. **The provenance and physical nature of the dataset are insufficiently specified (Sec. 1, Sec. 2.1, Sec. 3.1).** Without knowing whether the data come from DNS of a known PDE (e.g., compressible/incompressible Navier–Stokes, Boussinesq, etc.) and which parameters/forcing are used, it is hard to evaluate physical plausibility of the learned operators, interpret the density dynamics, or understand what “should” be recovered.

*Recommendation:* Augment Sec. 2.1 (and briefly Sec. 1) with a concise description of the data source: the generating equations (if known), solver type, forcing, viscosity/diffusivity, nondimensionalization/units, and boundary conditions beyond periodicity. If the generating PDE is unknown/proprietary, state this explicitly and enumerate what is known. In Sec. 3.2–3.3, briefly compare discovered dominant operator families against what one would expect for the stated physical system.

3. **Library specification is qualitative rather than explicit, preventing replication and weakening the collinearity diagnosis (Sec. 2.5).** The core claim (redundancy between composite operators and primitives) depends on exactly which terms (cross-products, derivative orders, mixed derivatives, component-wise curls/divergences, etc.) are included and how they are scaled.

*Recommendation:* In Sec. 2.5, report the total number of features (columns of  $\Theta$ ) and counts per category (constant, linear, quadratic, derivatives by order, products-with-derivatives, composite operators). Provide an appendix table listing every feature with a symbolic definition, including which cross-terms and mixed derivatives are includ-

ed/excluded. For each learned equation, report the number of nonzero terms and group them into interpretable operator families (e.g., net advection, net diffusion) where possible.

4. **The paper’s central collinearity analysis is largely descriptive and limited to a single highly redundant library and a single sparse regression scheme; quantitative diagnostics and mitigation/ablation experiments are missing (Sec. 3.2–3.2.1). As written, the paper strongly demonstrates the problem but only partially “addresses” it.**

*Recommendation:* Strengthen Sec. 3.2–3.2.1 with (i) quantitative collinearity diagnostics: numerical rank, singular value spectrum of  $\Theta$ , condition number (or spectrum) of  $\Theta^\top \Theta$ , and representative correlations between composite operators and their constituent primitives; and (ii) at least one ablation/mitigation experiment, e.g.: remove composite operators (primitives-only) vs. composite-only, or apply ridge/elastic-net / sequentially-thresholded ridge / group sparsity with physically linked groups (e.g., the 3 advection components). Report impacts on sparsity, coefficient stability/magnitude, and predictive metrics. If additional runs are infeasible, narrow claims in Sec. 4 to “demonstrates” rather than “addresses,” and clearly state limitations.

5. **Validation is too narrow to support claims about dynamical fidelity or “extended time horizons.” Forward integration is only shown for one step ( $t = 4 \rightarrow 5$ ) within a dataset of only 10 snapshots (Sec. 3.3.2, Sec. 4). High one-step  $R^2$  can occur even when long-horizon dynamics drift, especially when the learned operator is non-identifiable and tuned on the same data manifold.**

*Recommendation:* In Sec. 3.3.2, add multi-step rollouts over as many steps as the dataset permits (e.g.,  $t = 4 \rightarrow 6 \rightarrow 7 \rightarrow 8 \rightarrow 9$ ), reporting error growth ( $R^2$ /RMSE vs horizon). Track basic physical diagnostics during rollout (e.g., mean density, kinetic energy, and divergence statistics) to assess drift. If you cannot add experiments, revise Sec. 3.3.2 and Sec. 4 to explicitly limit conclusions to short-horizon (one-step) predictive accuracy.

6. **Potential train/test leakage and temporal-derivative construction are under-specified given only 10 time slices (Sec. 2.4–2.6, Sec. 3.3.1). Central differencing uses neighboring times; if evaluation points share neighbors with training points, derivative targets can couple train/test. Also, spatiotemporal correlation makes random point splits overly optimistic.**

*Recommendation:* Clarify in Sec. 2.4–2.6 and Sec. 3.3.1 exactly which time indices contribute to  $\partial_t$  targets (central difference implies only 8 usable times) and how train/test splits are formed (spatial only, temporal blocks, or both). Prefer blocked temporal evaluation (e.g., train on a subset of time indices and test on held-out times

whose  $\partial_t$  targets do not use training-time neighbors). Report the number of sampled space–time points used for regression and evaluation and justify independence assumptions.

- 7. Reproducibility-critical algorithmic details are missing or incomplete: spectral filter definition/cutoffs, FFT conventions and dealiasing, temporal-difference stencil specifics, sparse regression hyperparameters and stopping criteria, and integration/CFL settings (Sec. 2.3–2.6, Sec. 3.3.2).**

*Recommendation:* Expand Sec. 2.3–2.6 with concrete specifications: (i) filter type (sharp/Gaussian/exponential), cutoff wavenumber(s), and whether filtering is per-field; (ii) FFT normalization and k-grid definition (including  $2\pi/L$ ); (iii) aliasing/dealiasing treatment; (iv) temporal stencil used and how endpoints are handled; (v) sparse regression pseudocode (initial solve, threshold rule, refit procedure, stopping criteria, any ridge term); (vi) hyperparameter selection protocol (search range, validation criterion) and the chosen values per equation; and (vii) RK4/CFL details (CFL number, max  $dt$ , typical substeps).

- 8. Feature scaling / intercept handling is internally unclear and affects both null-space claims and coefficient interpretation (Sec. 2.5, Sec. 3.2). The text states standard-scaling  $\Theta$  (mean 0, std 1) while including a constant feature (1). A constant column has zero variance and cannot be standard-scaled without special handling; additionally, exact linear identities among raw features generally become affine relations after centering unless an intercept is handled consistently.**

*Recommendation:* Explicitly state how the constant/intercept is handled: whether the constant column is excluded from scaling, whether an intercept is fit separately, and whether targets ( $\partial_t$  fields) are centered/scaled. Provide the exact forward transform to scaled  $\Theta$  and the back-transform for coefficients (including any intercept correction). Clarify whether null-space relations are claimed for the raw library or the scaled design matrix actually used in regression, and show the transformed dependence (linear vs affine) accordingly.

- 9. Some numerical “cancellation” examples in the null-space discussion appear arithmetically inconsistent with the stated near-zero residual interpretation (Sec. 3.2–3.2.1). Reported residuals such as  $-291.5015$ ,  $-171.66317$ ,  $-45.06$ , and  $+38.545$  are not “near zero” under typical tolerances, suggesting transcription/sign/grouping errors or a mismatch between what is being summed and what is being claimed.**

*Recommendation:* Audit the cancellation examples in Sec. 3.2–3.2.1: confirm the exact grouping (e.g., composite operator coefficient vs sum of constituent coefficients), ensure consistent sign conventions, and update the numbers and/or the explanatory text

so the arithmetic matches the intended point. If the cancellation is approximate rather than near-exact, quantify it (e.g., relative residual) and explain why it is not closer to zero (e.g., scaling/centering effects).

## Minor issues

1. Sparsity–accuracy trade-off is asserted but not quantified. Despite a sparsity-promoting method, the learned equations retain **83–93** active terms per variable (Sec. 3.2–3.2.1), and it remains unclear whether much sparser models with similar predictive power exist.

*Recommendation:* In Sec. 3.2–3.2.1, sweep the sparsity threshold/regularization and report curves of #active terms vs local-derivative  $R^2$  (and optionally one-step rollout error). Consider a second-stage pruning/grouping pass: remove nearly canceling redundant groups and refit coefficients to quantify how much complexity is truly needed for performance.

2. Density equation interpretation is underdeveloped given its lower local-derivative  $R^2$  ( $\approx 0.36$ ) (Sec. 3.1, Sec. 3.3.1, Sec. 4). Because  $\partial_t \rho$  variance may be tiny,  $R^2$  can be misleading; the model may be fitting noise or numerical differentiation error rather than dynamics.

*Recommendation:* Add diagnostics for the density target: variance/std of  $\partial_t \rho$ , filtered vs unfiltered comparisons, and normalized error metrics (e.g., nRMSE). Compare against simple baselines ( $\partial_t \rho = 0$ ; advection–diffusion with few terms) and discuss whether learning  $\rho$  is meaningful for this dataset or whether a constrained/regularized form is more appropriate.

3. “Incompressible-like” claims are mostly qualitative and based on density statistics rather than direct divergence measurements (Sec. 3.1). Without reporting  $\nabla \cdot \mathbf{v}$  statistics, it is hard to interpret the role of density and whether continuity-like structure is present.

*Recommendation:* In Sec. 3.1, report summary statistics/histograms of  $\nabla \cdot \mathbf{v}$  over space and time (on filtered data if used for regression), and compare its scale to typical velocity gradients. State whether incompressibility was enforced in the generating simulation. Use this to support or revise the incompressibility analogy and to motivate constraints/groupings in the library if applicable.

4. Evaluation metrics focus on global  $R^2$ /RMSE; there is limited assessment of scale-dependent fidelity important for 3D flows (Sec. 3.3). Spatially averaged metrics can hide spectral/structural errors even when one-step prediction is strong.

*Recommendation:* Add one or two complementary diagnostics in Sec. 3.3: e.g., kinetic energy spectra of predicted vs true velocity for the rollout step(s), scale-dependent error (in Fourier shells), or error conditioned on magnitude/vorticity. If adding plots is

infeasible, explicitly note this limitation and outline it as future work in Sec. 4.

5. Figures are visually helpful but often lack actionable context: axis labels/units, slice orientation, domain scale, and consistent color limits across panels (Figures 1–6, especially 1, 2, 3, 6). This impairs interpretability and reproducibility.

*Recommendation:* Update figure captions and panels to include axis labels/units, slice-plane/orientation (e.g.,  $z = \text{const}$ ), domain extent, and fixed stated color limits per variable across comparable panels. Standardize panel labels (a–l) and add per-panel annotations (variable/time).

6. Connections to prior PDE-discovery literature and to known approaches for collinearity/identifiability (weak-form SINDy, group sparsity, Bayesian approaches) are too brief (Sec. 1, Sec. 2.5).

*Recommendation:* Add a short Related Work subsection (end of Sec. 1 or before Sec. 2) that cites core SINDy/PDE-FIND and representative methods addressing derivative noise and collinearity (weak form, ridge/elastic net, group sparse, Bayesian). Clearly state how your 3D case study and null-space demonstration differs from or complements these.

## Very minor issues

1. Fourier differentiation and wavenumber conventions are not fully defined (missing explicit  $k$ -grid definition and  $2\pi/L$  factors), which can create ambiguity even if the implementation is correct (Sec. 2.4).

*Recommendation:* Define the Fourier transform convention, domain lengths, and the discrete wavenumber grid explicitly (e.g.,  $k_x = 2\pi n/L_x$  with  $n$  ordering). State FFT normalization used in code to ensure reproducibility.

2. The dataset description uses “1283 periodic grid,” which is ambiguous (Sec. 2.1).

*Recommendation:* Replace “1283” with “128<sup>3</sup>” or “128 × 128 × 128” consistently throughout.

3. Notation/typography inconsistencies and minor formatting issues reduce polish (e.g., split word “math

ematical”; inconsistent  $R^2$  rendering; inconsistent vector/operator formatting; inconsistent section heading styles) (Sec. 1, Sec. 2.4–2.5, Sec. 3–4).

```
*Recommendation:* Proofread to standardize typography: use one  $R^2$  notation, consistent boldface for vectors, consistent  $\nabla \cdot \mathbf{v}$  /  $\nabla \times \mathbf{v}$  formatting, and consistent section numbering/capitalization; fix the line-break artifact in Sec. 1.
```

4. The  $\leftrightarrow$  notation used for Fourier-space correspondence is not explicitly defined (Sec. 2.4).

*Recommendation:* Add a short clarification sentence defining  $\leftrightarrow$  (e.g., “corresponds under the Fourier transform”) and whether it denotes an identity, an implementation rule, or both.

## Key statements and references

- • **A common and effective strategy in data-driven equation discovery involves constructing a comprehensive library of candidate mathematical terms (e.g., linear, non-linear, and derivative terms) and subsequently employing sparse regression techniques to identify the minimal set of terms that best describe the observed system dynamics, but when the candidate library contains highly correlated or linearly dependent terms (feature collinearity), sparse regression can yield unstable and non-unique coefficient estimates that hinder physical interpretability [11].**
  - *Reference(s):* [11]
- • **Sparse regression methods for discovering governing equations, such as SINDy-like algorithms that use iterative least squares with thresholding, are designed to find a sparse coefficient vector  $\xi_k$  satisfying  $\partial_t u_k \approx \Theta \xi_k$  from a rich feature library, and have been successfully applied in prior work to identify parsimonious models from data [11].**
  - *Reference(s):* [11]
- • **The use of spectral methods based on Fast Fourier Transforms (FFTs) to compute spatial derivatives in periodic domains—by transforming a field  $u$  to Fourier space  $\hat{u}$ , multiplying by  $ik_x$  for first derivatives or  $-k_x^2$  for second derivatives, and transforming back—follows established numerical analysis practices for accurate derivative estimation in such settings [11].**
  - *Reference(s):* [11]

## Mathematical consistency audit

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

**Maths relevance:** light

The paper is primarily methodological/empirical; the explicit mathematics is limited to standard spectral differentiation identities, a central-difference time derivative, the SINDy-style linear model  $\partial_t u \approx \Theta \xi$ , and qualitative null-space/collinearity arguments based on composite operators (advection/divergence/curl/Laplacian) being linear combinations of constituent deriva-

tive-product terms. Core symbolic concern: the stated standard-scaling of  $\Theta$  conflicts with inclusion of a constant feature and complicates/undermines claims of exact linear dependence in the matrix actually used for regression unless additional preprocessing details hold.

### Checked items

1. ✓ **Grid spacing from  $L = 1$  and  $N = 128$**  (Sec. 2.1 Dataset, p.2)
  - **Claim:** With box size  $L = 1$  and grid size 128 in each dimension, the spacing is  $\Delta x = \Delta y = \Delta z = 1/128$ .
  - **Checks:** algebra, definition consistency
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Domain length in each dimension is exactly 1, There are 128 grid points per dimension
  - **Notes:** If the domain is  $[0,1)$  with 128 points,  $\Delta = 1/128$  is consistent. (Minor textual ambiguity remains with “1283” vs  $128^3$ .)
  
2. ✓ **Spectral first-derivative identity** (Sec. 2.4 Derivative estimation (Spatial derivatives), p.3)
  - **Claim:** In Fourier space,  $\partial u / \partial x$  corresponds to multiplication of  $\hat{u}$  by  $ik_x$ .
  - **Checks:** symbolic correctness, notation/definition consistency
  - **Verdict:** PASS; confidence: medium; impact: minor
  - **Assumptions/inputs:** Periodic domain,  $k_x$  denotes the appropriate spatial wavenumber under the chosen Fourier convention
  - **Notes:** Correct up to Fourier-convention constants;  $k_x$  is not explicitly defined (could absorb  $2\pi/L$ ).
  
3. ✓ **Spectral second-derivative identity** (Sec. 2.4 Derivative estimation (Spatial derivatives), p.3)
  - **Claim:** In Fourier space,  $\partial^2 u / \partial x^2$  corresponds to multiplication of  $\hat{u}$  by  $-k_x^2$ .
  - **Checks:** symbolic correctness
  - **Verdict:** PASS; confidence: medium; impact: minor
  - **Assumptions/inputs:** Same Fourier convention as for first derivative
  - **Notes:** This follows from applying the first-derivative rule twice:  $(ik_x)^2 = -k_x^2$ .
  
4. ✓ **Central-difference temporal derivative** (Sec. 2.4 Derivative estimation (Temporal derivatives), p.3)
  - **Claim:**  $\partial u / \partial t \approx [u(t + \Delta t) - u(t - \Delta t)] / (2\Delta t)$  for interior points; end-points excluded.
  - **Checks:** algebra, approximation form consistency
  - **Verdict:** PASS; confidence: high; impact: minor

- **Assumptions/inputs:** Uniform time step  $\Delta t$
  - **Notes:** Formula and endpoint exclusion are consistent with the central-difference scheme.
5. ✓ **Convective term expansion** (Sec. 2.5 Feature library construction, p.3)
- **Claim:**  $(\mathbf{v} \cdot \nabla)v_x = v_x \partial_x v_x + v_y \partial_y v_x + v_z \partial_z v_x$  (and similarly for other components).
  - **Checks:** algebra, vector calculus identity
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:**  $\mathbf{v} = (v_x, v_y, v_z)$ ,  $\nabla = (\partial_x, \partial_y, \partial_z)$
  - **Notes:** Standard directional derivative identity; internally consistent with notation.
6. ✓ **Sparse regression model form** (Sec. 2.5 Sparse regression, p.3)
- **Claim:** For each variable  $u_k$ , the model is  $\partial_t u_k \approx \Theta \xi_k$ .
  - **Checks:** dimensional/shape consistency, notation consistency
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:**  $\Theta$  columns are candidate features evaluated at sample points,  $\xi_k$  is a coefficient vector
  - **Notes:** As a linear-in-features regression statement, this is consistent. No contradictory dimensions are stated.
7. ✗ **Standard scaling vs constant feature** (Sec. 2.5 Feature library construction, p.3)
- **Claim:**  $\Theta$  includes a constant term (1) and is standard-scaled to mean 0 and std 1.
  - **Checks:** algebra, preprocessing consistency
  - **Verdict:** FAIL; confidence: high; impact: critical
  - **Assumptions/inputs:** Standard scaling is applied column-wise, Constant term is literally constant across samples
  - **Notes:** A constant column has standard deviation 0 and cannot be transformed to unit variance via standard scaling as stated. The paper must specify special handling (exclude constant from scaling, or omit centering/scaling, or treat intercept separately). Without this, subsequent claims about null-space behavior in the scaled matrix and coefficient unscaling are not fully verifiable.
8. ✓ **Laplacian decomposition used in cancellation argument** (Sec. 3.2.1 Null-space phenomenon, p.5)
- **Claim:**  $\nabla^2 \rho$  equals the sum of diagonal second derivatives ( $\partial_{xx} \rho + \partial_{yy} \rho + \partial_{zz} \rho$ ), so including both creates a linear dependency that can cancel with opposite coefficients.

- **Checks:** algebra, definition consistency
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:** Laplacian is defined as  $\partial_{xx} + \partial_{yy} + \partial_{zz}$ , Features are constructed consistently
  - **Notes:** Identity is correct. Exact dependency in the regression design matrix depends on preprocessing (centering/scaling) details not fully specified.
9. ✓ **Divergence decomposition used in cancellation argument** (Sec. 3.2.1 Null-space phenomenon, p.5)
- **Claim:**  $\nabla \cdot \mathbf{v}$  equals  $\partial_x v_x + \partial_y v_y + \partial_z v_z$ , enabling cancellation if both composite and constituent terms are present.
  - **Checks:** algebra, definition consistency
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:** Divergence defined in the usual way, Features computed consistently
  - **Notes:** Identity is correct; same preprocessing caveat as above.
10. ✓ **Curl x-component decomposition used in cancellation argument** (Sec. 3.2.1 Null-space phenomenon, p.5)
- **Claim:**  $\text{curl}_x = \partial_y v_z - \partial_z v_y$ , so including  $\text{curl}_x$  and its two derivative terms creates a linear dependency and permits near-cancellation.
  - **Checks:** algebra, definition consistency
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Curl component is defined as stated
  - **Notes:** Definition is internally consistent with the stated decomposition; cancellation logic is plausible given opposite coefficients.
11. △ **Null-space solution decomposition claim** (Sec. 3.2.1 Null-space phenomenon, p.5)
- **Claim:** With collinear features, the regression can converge to  $\xi = \xi_{\text{true}} + \xi_{\text{null}}$  where  $\Theta \xi_{\text{null}} = \mathbf{0}$ , leaving predictions  $\Theta \xi$  unchanged.
  - **Checks:** linear algebra consistency
  - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
  - **Assumptions/inputs:**  $\Theta$  has a nontrivial null space (i.e., columns linearly dependent in the design matrix used for regression)
  - **Notes:** The linear-algebra statement is correct if  $\Theta$  (the matrix actually used in regression) has exact linear dependence. However, the paper also states  $\Theta$  is mean-centered and variance-scaled, and it does not specify how constant/intercept handling is done. Centering can turn exact linear equalities into affine relations; whether  $\Theta$  used in regression truly has a null space is unclear without explicit preprocessing definitions.

## Limitations

- Audit is based only on the provided 8-page parsed PDF text and embedded figure text; the explicit discovered PDEs and full coefficient tables are not present, so their internal algebra cannot be checked.
- Because the paper omits the precise Fourier-transform convention and wavenumber definitions, spectral-derivative identities can only be verified up to convention-dependent constants.
- Because preprocessing/unscaled steps are not fully specified (especially regarding the constant term), claims about exact null-space structure in the regression matrix cannot be definitively verified.

## Numerical results audit

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

Arithmetic and consistency checks for dataset dimensions, spacing, time indexing, and several narrative metrics largely pass. However, 4 of 5 null-space “cancellation” coefficient-sum demonstrations fail the stated near-zero behavior by large margins, suggesting sign/transcription/grouping errors in those specific illustrative coefficient lists.

### Checked items

1. ✓ **C1\_grid\_points\_cube** (Page 1 Abstract; Page 2 §2.1 Dataset)
  - **Claim:** Dataset is on a "128<sup>3</sup> periodic grid" / "grid of size 128<sup>3</sup> points".
  - **Checks:** integer\_power\_consistency
  - **Verdict:** PASS
  - **Notes:** 128<sup>3</sup> = 2,097,152 exactly.
2. ✓ **C2\_spatial\_resolution\_from\_L\_over\_N** (Page 2 §2.1 Dataset)
  - **Claim:** With box size  $L = 1$  and grid size 128<sup>3</sup>, spatial resolution is stated as  $\Delta x = \Delta y = \Delta z = 1/128$ .
  - **Checks:** unit\_consistent\_recompute
  - **Verdict:** PASS
  - **Notes:**  $L/N = 1/128 = 0.0078125$ .
3. ✓ **C3\_dataset\_shape\_product** (Page 2 §2.1 Dataset)
  - **Claim:** Full dataset shape is (10, 4, 128, 128, 128).
  - **Checks:** shape\_size\_recompute
  - **Verdict:** PASS
  - **Notes:** Total scalar values =  $10 \times 4 \times 128 \times 128 \times 128 = 83,886,080$ .
4. ✓ **C4\_time\_step\_and\_slices\_implied\_span** (Page 2 §2.1 Dataset; Page 4 Figure 1 caption ( $t = 0, 4, 9$ ))

- **Claim:** There are 10 time slices uniformly spaced with  $\Delta t = 1$ ; representative times include  $t = 0, 4, 9$ .
  - **Checks:** sequence\_consistency
  - **Verdict:** PASS
  - **Notes:** Span  $(10 - 1) \times 1 = 9$  matches  $t_{\max} - t_{\min}$ ; 4 is within  $[0, 9]$ .
5. ✓ **C5\_central\_difference\_excluded\_points\_count** (Page 3 §2.4 Temporal derivatives)
- **Claim:** Temporal derivatives use central difference; "The first and last time points were excluded".
  - **Checks:** count\_recompute
  - **Verdict:** PASS
  - **Notes:** Timepoints with  $\partial_t$  defined:  $10 - 2 = 8$ .
6. ✓ **C6\_total\_training\_targets\_points** (Page 3 §2.4 Temporal derivatives; Page 2 §2.1 Dataset)
- **Claim:** Each candidate term evaluated at every spatial-temporal point where  $\partial_t u$  was computed; with  $128^3$  grid and excluding first/last times.
  - **Checks:** rows\_count\_recompute
  - **Verdict:** PASS
  - **Notes:** Rows per variable:  $2,097,152 \times 8 = 16,777,216$ ; all 4 vars: 67,108,864.
7. ✗ **C7\_advection\_cancellation\_sum** (Page 5 §3.2.1 The null-space phenomenon (Advection Cancellation bullet))
- **Claim:** Advection Cancellation:  $\text{adv}_{v_z}$  coefficient +145.7538 and parts  $-145.7541(v_x \partial_x v_z)$ ,  $-145.7541(v_y \partial_y v_z)$ ,  $-145.7471(v_z \partial_z v_z)$  "resulting in a sum near zero".
  - **Checks:** linear\_combination\_near\_zero
  - **Verdict:** FAIL
  - **Notes:** Residual =  $145.7538 - 145.7541 - 145.7541 - 145.7471 = -291.5015$  (not near 0).
8. ✗ **C8\_laplacian\_cancellation\_sum** (Page 5 §3.2.1 The null-space phenomenon (Laplacian Cancellation bullet))
- **Claim:** Laplacian Cancellation:  $\nabla^2 \rho$  coefficient +85.83159 and diagonal second derivatives  $-85.83160(\partial_x^2 \rho)$ ,  $-85.83158(\partial_y^2 \rho)$ ,  $-85.83158(\partial_z^2 \rho)$  "perfectly canceling out".
  - **Checks:** linear\_combination\_near\_zero
  - **Verdict:** FAIL

- **Notes:** Residual =  $85.83159 - 85.83160 - 85.83158 - 85.83158 = -171.66317$  (not near 0).
9. ✘ **C9\_curl\_cancellation\_sum** (Page 5 §3.2.1 The null-space phenomenon (Curl Cancellation bullet))
- **Claim:** Curl Cancellation:  $\text{curl}_x$  weighted  $-45.127$ ; individual derivatives  $+45.167(\partial_y v_z)$  and  $-45.100(\partial_z v_y)$  "net contribution approximately zero".
  - **Checks:** linear\_combination\_near\_zero
  - **Verdict:** FAIL
  - **Notes:** Residual =  $-45.127 + 45.167 - 45.100 = -45.06$  (not near 0).
10. ✘ **C10\_divergence\_cancellation\_sum** (Page 5 §3.2.1 The null-space phenomenon (Divergence Cancellation bullet))
- **Claim:** Divergence Cancellation:  $\nabla \cdot \mathbf{v}$  received  $-19.468$ , offset by  $+19.339(\partial_x v_x)$ ,  $+19.327(\partial_y v_y)$ ,  $+19.347(\partial_z v_z)$ .
  - **Checks:** linear\_combination\_near\_zero
  - **Verdict:** FAIL
  - **Notes:** Residual =  $-19.468 + 19.339 + 19.327 + 19.347 = +38.545$  (not near 0).
11. ✔ **C11\_density\_constant\_cancellation\_sum** (Page 5 §3.2.1 The null-space phenomenon (Density Constant Cancellation bullet))
- **Claim:** Density Constant Cancellation:  $95.53(\rho) - 47.88(\rho^2) - 47.65(\text{const}) \approx 0$ .
  - **Checks:** scalar\_balance\_at\_rho\_approx\_1
  - **Verdict:** PASS
  - **Notes:** At  $\rho = 1$ :  $95.53 - 47.88 - 47.65 = 0.00$ .
12. ✔ **C12\_R2\_range\_matches\_listed\_values\_velocity** (Page 6 §3.3.1 Local derivative accuracy; Figure 5 caption; Page 1 Abstract; Page 2 end of Introduction)
- **Claim:**  $R^2$  values between  $0.593$  and  $0.732$  for velocity components; reported individual  $R^2$  are  $0.678 (v_x)$ ,  $0.732 (v_y)$ ,  $0.593 (v_z)$ .
  - **Checks:** range\_consistency
  - **Verdict:** PASS
  - **Notes:** min/max of  $0.678, 0.732, 0.593$  equals  $0.593, 0.732$ .
13. ✔ **C13\_velocity\_rmse\_order\_of\_magnitude\_bounds** (Page 6 §3.3.1 Local derivative accuracy; Figure 5 caption)
- **Claim:** Velocity derivative RMSE stated to be on the order of  $5 \times 10^{-3}$  to  $7 \times 10^{-3}$ .
  - **Checks:** interval\_parse\_check
  - **Verdict:** PASS

- **Notes:** Parsed bounds  $0.005 < 0.007$ ; both are  $O(1 \times 10^{-3})$ .
14. ✓ **C14\_density\_rmse\_scientific\_notation\_parse** (Page 6 §3.3.1 Local derivative accuracy; Figure 5 caption)
- **Claim:** Density derivative RMSE reported as  $1.7 \times 10^{-4}$ .
  - **Checks:** scientific\_notation\_parse
  - **Verdict:** PASS
  - **Notes:**  $1.7 \times 10^{-4} = 0.00017$ .
15. ✓ **C15\_forward\_rmse\_vs\_std\_percentage** (Page 6 §3.3.2 Forward predictive modeling)
- **Claim:** Forward prediction: RMSE for predicted velocity states  $\approx 5 \times 10^{-3}$ , representing roughly 2% of the standard deviation of the velocity fields; earlier std dev stated around 0.23 to 0.25.
  - **Checks:** percentage\_recompute
  - **Verdict:** PASS
  - **Notes:**  $0.005/0.23 = 0.02174$  (2.174%);  $0.005/0.25 = 0.02$  (2.0%), consistent with “roughly 2%”.
16. ✓ **C16\_forward\_R2\_thresholds\_consistency** (Page 6 §3.3.2 Forward predictive modeling; Page 1 Abstract; Page 2 end of Introduction; Page 7 Conclusions)
- **Claim:** Forward-time integration yields  $R^2$  values exceeding 0.999 for velocity fields and 0.992 for density over a subsequent time step.
  - **Checks:** logical\_threshold\_check
  - **Verdict:** PASS
  - **Notes:** Both values lie in  $[0, 1]$ ;  $0.992 < 0.999$  is permissible given the claim structure.
17. ✓ **C17\_active\_terms\_range\_width** (Page 5 §3.2 Equation discovery; Page 7 Conclusions)
- **Claim:** Number of active terms ranges from 83 to 93 terms per variable.
  - **Checks:** range\_width\_recompute
  - **Verdict:** PASS
  - **Notes:** Range width =  $93 - 83 = 10$ .

## Limitations

- Only parsed text content was available; numeric values embedded solely in plots/graphics (Figures 2, 4–6) cannot be extracted reliably without image/pixel data.
- Many performance metrics ( $R^2$ /RMSE) and statistical summaries are stated narratively; recomputation would require access to the underlying dataset and model outputs, which are not included in the PDF text.

- Null-space/cancellation checks can only verify arithmetic of reported coefficients, not whether the corresponding features are exactly linearly dependent in the constructed feature matrix.