

Skeptical review: Characterizing the Multi-Scale and Geometric Structure of PINN Latent Space via Wavelets and Ricci Scalar

Summary

The paper proposes a dual analysis pipeline to characterize internal representations of a Physics-Informed Neural Network (PINN) trained on Burgers' equation by treating each of 10 latent components $L_i(\mathbf{x}, t)$ (extracted on a 100×100 (\mathbf{x}, t) grid; Sec. 2.1–2.3) as a 2D scalar field. It applies a 2D discrete wavelet transform (DWT) to quantify multiscale structure (Sec. 2.4, Sec. 3.1), reporting strong fine-scale energy concentration, heavy-tailed coefficient distributions (high kurtosis), and approximate power-law energy decay across DWT levels (Sec. 3.1.1–3.1.4). It then computes Gaussian curvature and the associated Ricci scalar $R = 2K$ of the graph surfaces $z = L_i(\mathbf{x}, t)$ via numerical derivatives (Sec. 2.5, Sec. 3.2), finding spatially structured positive/negative curvature patterns with near-zero means and component-dependent variances (Sec. 3.2.2–3.2.3). The idea—combining multiscale signal tools with geometric diagnostics to probe PINN internals—is interesting and potentially useful for interpretability. However, key elements required to interpret and trust the findings are currently missing or inconsistent: the PINN/PDE setup and provenance are under-specified (Sec. 1, Sec. 2.1–2.2), the wavelet configuration and even the wavelet family and energy definition are inconsistent between Methods and Results (Sec. 2.4 vs Sec. 3.1), scaling/fractality and “shock encoding” claims are not sufficiently validated given the limited scale range and lack of fit diagnostics (Sec. 3.1.4, Sec. 3.3), and curvature estimates are potentially sensitive to second-derivative noise/boundary handling (Sec. 3.2.1). Finally, there are no baselines or links to solution accuracy/residuals to establish what is PINN-specific or physically meaningful (Sec. 3.3–4). Addressing these points would substantially strengthen the paper’s credibility and impact.

Strengths

- Clear objective: focuses on characterizing a PINN latent representation using explicit mathematical diagnostics rather than proposing a new architecture (Sec. 1–2).
- Novel and conceptually coherent combination of multiscale (2D DWT) and geometric (Gaussian curvature/Ricci scalar) analyses for latent component fields (Sec. 2.4–2.5, Sec. 3.1–3.2).
- Provides quantitative summaries (kurtosis, scaling exponents, derivative-test errors, Ricci statistics) rather than only qualitative visual inspection (Sec. 3.1.2–3.1.4, Sec. 3.2.1–3.2.3).
- Curvature formulas for graph surfaces are stated in a standard form and applied consistently as $R_i(\mathbf{x}, t) = 2K_i(\mathbf{x}, t)$ (Sec. 2.5.2–2.5.3).

- Figures (especially the multi-panel presentations) help communicate cross-component heterogeneity in energy decay and curvature patterns (Fig. 1, Fig. 3), even though presentation improvements are needed.

Major issues

1. **Insufficient specification of the PINN, PDE setup, and latent extraction makes the results hard to interpret and impossible to reproduce (Sec. 1, Sec. 2.1–2.2).** The manuscript does not state the explicit Burgers equation form (including viscosity/parameters), domain, IC/BC, and whether “2D” refers to one space + time (x, t) or two spatial dimensions. The PINN architecture and training details are also missing (layer sizes, activations, normalization, where the 10D latent is taken—pre/post nonlinearity, loss terms/weights, collocation strategy, optimizer/schedule, seeds/runs). The provenance of the provided NumPy array $(100, 100, 12)$ is unclear (who trained it, which checkpoint, whether multiple runs were analyzed).

Recommendation: Expand Sec. 2.1–2.2 (or add a dedicated subsection) to fully specify: (i) the exact PDE and parameters, plus IC/BC and domain, explicitly clarifying that the grid is spatio-temporal (x, t) if that is the case; (ii) the PINN architecture (depth/width, activations, parameter count, where the 10D latent is extracted and whether it is before/after activation); (iii) the training objective (data terms, PDE residual, IC/BC terms) and weights, sampling strategy, optimizer and schedule, epochs/batches, and number of seeds; (iv) basic performance metrics versus a reference solution (e.g., L^2 /NRMSE of $u(x, t)$ and/or PDE residual statistics). Add a short provenance/reproducibility statement describing how the $(100, 100, 12)$ tensor was produced and whether code/data will be released (Sec. 4).

2. **Wavelet analysis is currently internally inconsistent and under-specified in ways that can materially change all reported DWT results (Sec. 2.4 vs Sec. 3.1).** (a) Wavelet family mismatch: Methods say ‘db1’ (Haar) (Sec. 2.4.1, p.3) while Results state ‘sym2’ (Sec. 3.1, p.5). (b) Energy definition mismatch: Methods include detail energies plus coarsest approximation energy (Sec. 2.4.2) while Results define per-level energy as detail-only sums (Sec. 3.1.2). (c) A 100×100 grid is non-dyadic, so DWT coefficient statistics depend on padding/extension mode (e.g., symmetric/periodization/zero padding) and the feasible maximum level; these choices are not documented. (d) It is unclear how subbands (H/V/D) are aggregated, whether energies are normalized by coefficient count per level, and what exactly the reported reconstruction NRMSE signifies (perfect reconstruction is expected for standard DWT implementations and is not, by itself, evidence of representation quality).

Recommendation: In Sec. 2.4.1–2.4.2, explicitly specify: the exact wavelet used in all reported figures/tables (or separate results by wavelet if multiple were tried), the Py-Wavelets (or equivalent) extension mode, the exact number of levels used and how it was determined for 100×100 , and the precise definition of per-level energy (detail-

only vs detail+approx, and how the approximation term is handled). State how H/V/D subbands are combined and whether energies are normalized (e.g., by coefficient counts) before cross-level or cross-component comparisons. If feasible, add a sensitivity check by repeating key metrics (energy fractions, kurtosis, fitted exponents) on a dyadic-cropped grid (e.g., 96×96) and/or with periodization to assess padding/boundary effects. Reframe or remove the reconstruction NRMSE as an interpretability result; if kept, clarify it is a numerical sanity check only (Sec. 3.1.1).

3. **Power-law/self-affinity/fractality claims are overstated relative to the evidence and the limited scale range, and the fitting procedure lacks diagnostics (Sec. 3.1.4, Sec. 3.3, Conclusion).** Exponents are inferred from linear fits across only ~ 5 levels, without reporting goodness-of-fit (R^2), residuals, standard errors/confidence intervals, or sensitivity to the chosen fit range (coarsest/finest levels are often contaminated by finite-size and discretization effects). In addition, the slope-to-exponent conversion depends on the logarithm base, but the manuscript does not specify whether ‘log’ is \ln or \log_{10} (Sec. 3.1.4).

Recommendation: In Sec. 3.1.4, report for each component: the exact fit range (which levels), R^2 (or another fit metric), residual plots or a summary of residual structure, and uncertainty estimates (standard errors or confidence intervals) for the fitted slopes/exponents. Explicitly state the log base used in plots/fits and use the matching conversion $\alpha_i = m_i / \log_b(2)$ (Sec. 3.1.4). Add a short limitation statement that with 5 levels the results indicate at most “approximate scaling over a narrow range,” not definitive fractality; adjust language in Sec. 3.3/Conclusion accordingly.

4. **Geometric analysis risks conceptual over-interpretation and may be numerically fragile due to second-derivative noise and boundary handling (Sec. 2.5, Sec. 3.2).** The computed “Ricci scalar” is the intrinsic curvature of the graph surface $z = L_i(x, t) \subset \mathbb{R}^3$, not curvature of a learned 10D latent manifold with a model-induced metric; the current phrasing can mislead readers about what is being measured (Sec. 2.5.3, Sec. 3.2). Numerically, curvature depends directly on L_{xx}, L_{tt}, L_{xt} , yet the derivative validation reports $\sim 15.7\%$ relative error for f_{tt} on a smooth test function (Sec. 3.2.1), and boundary treatment (cropping vs one-sided differences vs keeping edges) is not described. These factors can substantially affect Ricci magnitude and possibly spatial patterns (Sec. 3.2.2–3.2.3).

Recommendation: Clarify terminology throughout Sec. 2.5 and Sec. 3.2: describe results as “Gaussian curvature / Ricci scalar of the graph surfaces $z = L_i(x, t)$ ” unless a latent-space metric is explicitly introduced. Strengthen numerical robustness in Sec. 2.5.1 and Sec. 3.2.1 by: (i) reporting errors not only for derivatives but also for K and R on the analytic test surface; (ii) specifying and justifying boundary handling and which points enter Table 2 statistics; (iii) adding a sensitivity analysis (e.g., higher-order finite differences, mild smoothing prior to differentiation, and/or varying grid

spacing if available) and reporting how Table 2 statistics and representative Ricci maps change. If possible, compute derivatives via automatic differentiation from the original PINN rather than finite differences on a saved grid, and compare.

5. **Lack of baselines and limited scope (single model, single layer, single PDE setting) prevents determining what is PINN-specific versus generic, and limits the paper’s broader interpretability claims (Sec. 2.1–2.3, Sec. 3, Sec. 4).** All findings are based on one 10D latent layer from one pre-trained network; no comparisons are provided to (i) a supervised network trained on the same $u(x, t)$ data, (ii) an untrained network, (iii) alternative PINN settings (loss weights, viscosity), or (iv) other layers/latent dimensions.

Recommendation: Add at least one baseline in Sec. 3 (new subsection): run the same wavelet + curvature pipeline on (i) a comparable supervised MLP trained on solution snapshots (no PDE residual) and/or (ii) an untrained network with the same architecture. Compare energy distributions, kurtosis, scaling exponents, and Ricci statistics. If feasible, analyze at least one additional layer (earlier vs later) or one additional PDE configuration (e.g., different viscosity) to assess stability. If new experiments are not possible, explicitly reframe the work in Sec. 4 as exploratory and avoid PINN-specific generalizations.

6. **Connections to physics, performance, and the paper’s key interpretations (“shock encoding”, “precisely the right properties”) are asserted but not directly tested (Sec. 3.3, Conclusion), and the hypothesized wavelet–curvature linkage is not quantified.** The manuscript does not relate latent fine-scale energy or high $|R|$ regions to the actual Burgers solution $u(x, t)$, its gradients (e.g., $|u_x|$), or PDE residual/error maps, nor does it quantify whether components with more negative α_i also have larger curvature variance (claims in Sec. 3.3 appear only qualitative and potentially inconsistent with Table 2 ordering).

Recommendation: In Sec. 3.3, add direct, testable linkages: (i) plot $u(x, t)$ and a shock/steepness proxy such as $|\partial u / \partial x|$ (or PDE residual) alongside representative L_i , fine-scale wavelet magnitude maps, and $|R_i|$; (ii) compute spatial correlations/overlap metrics between high-gradient regions and high fine-scale energy / high $|R|$ regions; (iii) compute across-component correlations such as $\text{corr}(\alpha_i, \text{Var}(R_i))$ and report uncertainty given $n = 10$. If these additions are out of scope, soften claims in Sec. 3.3 and the Conclusion to “consistent with” and present shock/turbulence interpretations as hypotheses for future work.

Minor issues

1. Decomposition-level selection heuristic is arithmetically inconsistent with the stated grid size (Sec. 2.4.1, p.3). The paper states $J = \lfloor \log_2(\min(N_x, N_t)) \rfloor - c$ with $c \approx 2$ or 3 , and claims this yields $\sim 5\text{--}6$ levels for 100×100 ; but $\lfloor \log_2(100) \rfloor = 6$, so subtracting $2\text{--}3$ gives $3\text{--}4$.

Recommendation: Correct the heuristic arithmetic and/or replace it with the actual rule used by the implementation (including dependence on wavelet filter length and extension mode). State explicitly the number of levels used for all reported results.

2. Interpretation of extremely high kurtosis as “sparsity” is plausible but not uniquely diagnostic and may be affected by padding/boundary artifacts, nonstationarity, or estimator conventions (Sec. 3.1.3). The text mentions Fisher kurtosis (normal = 0), but does not clarify estimator details or sample-size effects at each level.

Recommendation: In Sec. 3.1.3, clarify kurtosis convention and estimator. Add at least one complementary sparsity/heavy-tail metric (e.g., fraction of energy in top 1% coefficients, Gini coefficient, $|c|_1/|c|_2$) and show representative coefficient histograms/QQ-plots for a couple of components/levels to support the “sparse localized features” interpretation.

3. Exploratory data analysis (EDA) in Sec. 2.3 (descriptive statistics, correlations) is not integrated into the Results and currently feels disconnected from the main narrative.

Recommendation: Either (i) add a brief results paragraph/figure in Sec. 3 summarizing the correlation matrix and key distributional stats of L_i (motivating why per-component multiscale/geometric analysis is needed), or (ii) shorten Sec. 2.3 and explicitly state that EDA was preliminary and not central to later conclusions.

4. Figures need improvements for readability and to avoid ambiguity in cross-panel comparisons, especially given many small panels (Fig. 1, Fig. 3). Fig. 1’s size makes ticks/labels hard to read and does not clearly state log base or normalization; Fig. 3 does not clearly confirm a shared symmetric, zero-centered color scale for signed curvature and lacks axis labels/units/domain extents.

Recommendation: Increase Fig. 1 and Fig. 3 sizes (or split across rows/figures), use shared axes where appropriate, and export as vector PDF/SVG or ≥ 300 dpi. For Fig. 1: label y-axis as $\log_{10}(E)$ or $\ln(E)$, enforce a common y-range, and consider plotting normalized energy fractions per level. For Fig. 3: use a single global symmetric diverging colormap centered at 0 with a shared colorbar, specify vmin/vmax in the caption, add x/t axis ranges, and (optionally) overlay an $R = 0$ contour.

5. Related work and positioning are skewed and do not sufficiently situate the contribution within core PINN literature and interpretability of neural PDE solvers (Sec. 1, References).

Recommendation: Expand Sec. 1 with a short, focused related-work paragraph covering: (i) PINNs and known training/pathology issues relevant to internal representations, (ii) interpretability/representation analysis in scientific ML (including any prior wavelet/spectral analyses of learned PDE solvers), and (iii) clarify how the present diagnostics complement existing tools. Update references accordingly.

6. Curvature magnitudes depend on coordinate scaling; although \mathbf{x} and t are normalized to $[0, 1]$, the manuscript does not discuss commensurability or how rescaling would affect Ricci statistics (Sec. 2.1, Sec. 2.5).

Recommendation: Add a short note in Sec. 2.5 stating that curvature of $z = L_i(\mathbf{x}, t)$ depends on the scaling of \mathbf{x} and t , and justify the chosen normalization (or report how results change under simple rescalings).

Very minor issues

1. Typos/formatting and placeholder artifacts reduce professionalism and readability (Sec. 1–3). Examples include broken words/line breaks, inconsistent “Burgers/Burgers’/Burgers’”, unformatted code snippets, placeholder labels like “(LABEL:2DDiscreteWaveletTransform)”, and inconsistent naming such as `numpy_gradient` vs `numpy.gradient`.

Recommendation: Proofread the manuscript end-to-end, standardize “Burgers’ equation”, format code in monospace blocks, remove placeholder labels, and ensure consistent function/wavelet names throughout.

2. Notation for indices and scales is somewhat inconsistent (Sec. 2.4, Sec. 3.1.4): j/k used interchangeably for DWT level; i/k reused across component/level indices; mapping between “scale” and “level” is presented in multiple forms without a single explicit definition.

Recommendation: Define and adhere to a single convention (e.g., component index i ; DWT level j ; scale $s_j = 2^j$). Add a one-line mapping in Sec. 3.1.4 showing $\log E(j) = \alpha j \log 2 + c$ and state explicitly what is regressed against what.

3. References appear to contain malformed/duplicated entries and inconsistent bibliographic fields (References; also impacts Sec. 1–2 citations).

Recommendation: Audit all references: remove duplicates, fix author/year/title/venue fields, and ensure every in-text citation has one correct bibliography entry.

4. Some captions/text are redundant and some figure/table references are inconsistent in capitalization/style (Sec. 3).

Recommendation: Lightly edit captions to avoid repeating main-text descriptions, and standardize references to “Figure 1”, “Table 2”, and “DWT level j ” throughout.

Key statements and references

- **✘** The `raw_data` array used in this study has shape $(100, 100, 12)$, where the first two dimensions correspond to a spatio-temporal grid previously described in astronomical spatio-temporal analyses, and the third dimension contains 12 features, from which the spatial mesh `x_mesh =`

`raw_data[:, :, 0]`, temporal mesh `t_mesh = raw_data[:, :, 1]`, and latent space components `L_components = raw_data[:, :, 2:]` are extracted following established practices for handling such simulation outputs.

- *Reference(s)*: Schad, 2018, Brodiano et al., 2021, Morosin et al., 2022
- *Justification*: None of the references describe or use a `raw_data` array with shape $(100, 100, 12)$, a 12-feature axis, or derived variables \mathbf{x}_{mesh} , t_{mesh} , and latent components. Schad (2018) analyzes 2D images and 3D (x, y, t) data cubes with the Rolling Hough Transform but no such array structure. Brodiano et al. (2021) compute spatio-temporal spectra $E(\mathbf{k}, \omega)$ from 3D CMHD simulations, not $100 \times 100 \times 12$ feature arrays. Morosin et al. (2022) perform NLTE inversions and neural-network prediction of atmospheric parameters, without defining a $100 \times 100 \times 12$ data tensor or ‘latent space components’.
- ✘ **The unique spatial and temporal coordinates are obtained as `x_coords = x_mesh[:, 0]` and `t_coords = t_mesh[0, :]`, and the grid is verified to be uniformly spaced in both dimensions, with spacings Δx and Δt computed as differences between consecutive coordinates, a procedure consistent with prior work on curvilinear and nested grids in astrophysical simulations and Poisson solvers.**
- *Reference(s)*: Grimm-Strele et al., 2013, Munoz et al., 2015, Matsumoto and Hanawa, 2002
- *Justification*: Neither paper describes extracting unique coordinates from mesh arrays (e.g., `x_coords = x_mesh[:, 0]`, `t_coords = t_mesh[0, :]`) or verifying uniform spacing by taking differences of consecutive coordinates, nor do they discuss temporal grids or Δt . Grimm-Strele et al., 2013 state that the computational space for mapped curvilinear grids is equidistant and Cartesian $(\delta\xi, \delta\eta)$, while the physical grid is generally nonuniform (Sections 2–2.1), and do not present such a verification procedure. Matsumoto and Hanawa, 2002 define a nested grid composed of uniform grids with different spacings per level (Eqs. 1–7; abstract), not a single uniformly spaced grid, and they do not treat time discretization. Thus the claim about uniform spacing in both dimensions and its consistency with prior curvilinear/nested-grid work is unsupported.
- ✘ **For exploratory data analysis of the latent space components `L_i(x,t)`, standard descriptive statistics (mean, variance, minimum, maximum, median, skewness, kurtosis) are computed over the 100×100 grid for each component, in line with recent applications of such statistics to astronomical data and cosmological summary statistics for simulation-based inference.**
- *Reference(s)*: Nikolic et al., 2017, 2018, Lehman et al., 2024, Sui et al., 2023
- *Justification*: None of the cited works describe computing standard descriptive statistics (mean, variance, min, max, median, skewness, kurtosis) over a 100×100 grid for latent space components $L_i(x, t)$. Nikolic et al., 2017, 2018 focus on memoization for

astronomical data processing, not such EDA. Lehman et al., 2024 learns summary statistics with GNNs and analyzes them via PCA, correlations, Isomap, and t-SNE, without grid-based descriptive statistics. Sui et al., 2023 evaluates cosmological summaries via mutual information (e.g., power spectrum, bispectrum, scattering transform), not by computing these descriptive statistics or using a 100×100 grid. Therefore, the statement is not supported.

- **✗ Inter-component relationships among the 10 latent dimensions are quantified by computing the Pearson correlation coefficient matrix $C_{ij} = \text{corr}(\text{flatten}(L_i), \text{flatten}(L_j))$ after flattening each 100×100 latent field into a 10,000-element vector, following correlation-based analyses used in recent cosmological and large-scale structure studies.**
- *Reference(s):* Krywonos et al., 2024, Atkins et al., 2024
- *Justification:* Neither Krywonos et al., 2024 nor Atkins et al., 2024 describe latent dimensions or computing Pearson correlations of flattened 100×100 latent fields. Krywonos et al., 2024 focuses on angular power spectra, Fisher matrices, and photo-z systematics; Atkins et al., 2024 develops covariance/correlation matrices for binned CMB power spectra. While correlation matrices of power spectrum bins are shown (e.g., Atkins et al., 2024), no method matching $C_{ij} = \text{corr}(\text{flatten}(L_i), \text{flatten}(L_j))$ for latent fields is presented.
- **✗ The numerical estimation of first and second derivatives of each latent component $L_i(x,t)$ with respect to x and t is performed using `numpy.gradient` with central differences in the interior and one-sided differences at the boundaries, a finite-difference strategy that has been adopted in recent astrophysical and latent-variable modeling work to enable subsequent curvature-based geometric analyses.**
- *Reference(s):* Grassi et al., 2021, Horta et al., 2025, Ferreira and Bejger, 2025
- *Justification:* Neither paper mentions using `numpy.gradient` or finite-difference schemes (central interior, one-sided boundaries) to compute first/second derivatives of latent components. Horta et al., 2025 compute derivatives analytically via Jacobians (e.g., $A \cdot B+$), not finite differences, and do not discuss curvature-based geometric analyses. Ferreira and Bejger, 2025 do not compute such derivatives and focus on cVAE training/inference. Thus the statement is not supported by the attached papers.

Mathematical consistency audit

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

Maths relevance: substantial

The paper’s core analytic content consists of: (i) defining a 2D DWT decomposition of latent fields $L_i(x, t)$, summarizing coefficient energies across levels, and positing/estimating a power-law energy–scale relation with exponent α_i via log-linear fits; and (ii) treating each $L_i(x, t)$ as a graph surface (x, t, L_i) and computing Gaussian curvature from first/second derivatives, then converting to a Ricci scalar via $R = 2K$. There are notable internal inconsistencies in the definitions used for the wavelet analysis (wavelet family and energy definition), while the differential-geometry formulas are largely consistent as stated.

Checked items

1. ✓ Latent-space extraction and indexing (Sec. 2.2, pp.2–3)

- **Claim:** `raw_data` has shape $(100, 100, 12)$ with $x_{\text{mesh}} = \text{raw_data}[:, :, 0]$, $t_{\text{mesh}} = \text{raw_data}[:, :, 1]$, and latent components $L_{\text{components}} = \text{raw_data}[:, :, 2 :]$ giving 10 fields $L_i(x, t)$, $i \in 0, \dots, 9$.
- **Checks:** definition consistency, index/shape consistency
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** Third axis ordering is exactly as described (x, t then latent channels)., Array axes 0 and 1 correspond to x and t respectively.
- **Notes:** Slicing `2 :` yields 10 features from 12 total, consistent with $i = 0 \dots 9$ and later references to L_0 – L_9 .

2. ✓ Grid spacing definitions (Sec. 2.2, p.3)

- **Claim:** Uniform grid spacings are defined as $\Delta x = x_{\text{coords}}[1] - x_{\text{coords}}[0]$, $\Delta t = t_{\text{coords}}[1] - t_{\text{coords}}[0]$.
- **Checks:** definition consistency, dimensional/units sanity
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** x_{coords} and t_{coords} are monotone and uniformly spaced.
- **Notes:** Definitions are standard finite-difference spacings; no algebra issues.

3. ✓ Derivative notation and mapping to axes (Sec. 2.5.1, p.4)

- **Claim:** $L_{i,x}$ and $L_{i,t}$ are computed along axes 0 and 1 using spacings Δx and Δt ; second derivatives $L_{i,xx}$, $L_{i,tt}$, $L_{i,xt}$ are computed by differentiating first derivatives.
- **Checks:** notation consistency, operator consistency
- **Verdict:** PASS; confidence: medium; impact: moderate
- **Assumptions/inputs:** Axis 0 corresponds to x and axis 1 corresponds to t (consistent with earlier `raw_data` shape statement).
- **Notes:** Symbol definitions match later curvature formulas. Numerical scheme details are out of scope; analytically the mapping is consistent if axis conventions hold.

4. ✖ **Wavelet selection (Methods)** (Sec. 2.4.1, p.3)

- **Claim:** The 2D DWT uses the 'db1' (Haar) mother wavelet.
- **Checks:** definition consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** This wavelet choice is used throughout subsequent wavelet results.
- **Notes:** Contradicted by Sec. 3.1 (p.5), which states the DWT used 'sym2'. Both cannot simultaneously be the sole basis for the reported Tables/Figures without additional clarification.

5. ✖ **Wavelet selection (Results)** (Sec. 3.1, p.5)

- **Claim:** Results are obtained using the 2D DWT with the 'sym2' mother wavelet up to 5 levels.
- **Checks:** definition consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** This matches the wavelet defined in Methods.
- **Notes:** Contradicts the Methods section, which specifies 'db1'. The inconsistency undermines interpretability of coefficient statistics and scaling exponents as mathematical properties of a specific transform.

6. ✖ **Decomposition level heuristic J** (Sec. 2.4.1, p.3)

- **Claim:** Typical decomposition depth is $J = \lfloor \log_2(\min(N_x, N_t)) \rfloor - c$ with $c \approx 2$ or 3 ; for 100×100 , this allows $\sim 5-6$ levels.
- **Checks:** algebra/arithmetic consistency
- **Verdict:** FAIL; confidence: high; impact: moderate
- **Assumptions/inputs:** $N_x = N_t = 100$ as stated.
- **Notes:** With $\min(N_x, N_t) = 100$, $\lfloor \log_2(100) \rfloor = 6$, so $J = 6 - c$ gives $3-4$ for $c = 2-3$, not $5-6$. The stated heuristic and the stated implication are inconsistent.

7. ⚠ **Energy across scales definition (Methods)** (Sec. 2.4.2, p.3)

- **Claim:** Energy at each level is sum of squared detail coefficients at that level, plus energy of approximation coefficients at the coarsest level.
- **Checks:** definition consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: critical
- **Assumptions/inputs:** This energy definition is the one used in later scaling fits.
- **Notes:** Later (Sec. 3.1.2) energy is described as detail-only per level, with no mention of approximation energy. Without clarification, $E_i(\mathbf{k})$ used in Fig. 1/2 and Table 1 cannot be verified as a consistently defined quantity.

8. ✖ **Energy across scales definition (Results)** (Sec. 3.1.2, p.5)

- **Claim:** Energy at level k is the sum of squared horizontal/vertical/diagonal detail coefficients at level k .
- **Checks:** definition consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** Matches Methods definition of $E_i(k)$.
- **Notes:** Directly conflicts with Sec. 2.4.2, which adds approximation energy at the coarsest level. This affects any analytic statement about $E_i(k)$ and the fitted scaling exponent.

9. ✔ **Power-law scaling statement in terms of 2^j** (Sec. 2.4.2, p.4)

- **Claim:** Self-affinity is suggested by $E_i(j) \propto (2^j)^\alpha$; tested via $\log(E_i(j))$ vs $\log(2^j)$.
- **Checks:** algebraic transformation, notation consistency
- **Verdict:** PASS; confidence: high; impact: moderate
- **Assumptions/inputs:** Scale variable is $s_j = 2^j$ (or proportional).
- **Notes:** If $E_i \propto (2^j)^\alpha$, then $\log E_i = \alpha \log(2^j) + \text{const} = \alpha j \log 2 + \text{const}$, which is linear in either $\log(2^j)$ or j .

10. △ **Exponent estimation: α from slope m** (Sec. 3.1.4, p.6)

- **Claim:** Fitting $\log(E_i(k)) = m_i k + c_i$ implies $\alpha_i = m_i / \log(2)$.
- **Checks:** algebraic transformation, definition/base consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** The logarithm base used in $\log(E)$ and in $\log(2)$ is the same.
- **Notes:** The relation holds if \log denotes a consistent base b : $m = \alpha \log_b(2)$, hence $\alpha = m / \log_b(2)$. The paper does not specify whether \log is \ln or \log_{10} , so the conversion as written is not fully verifiable.

11. ✔ **Gaussian curvature formula for a graph surface** (Sec. 2.5.2, p.4)

- **Claim:** For $z = f(x, y)$, $K = \frac{f_{xx}f_{yy} - (f_{xy})^2}{(1 + f_x^2 + f_y^2)^2}$.
- **Checks:** formula structure sanity, symbol consistency
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Surface is the graph of a scalar function over a Euclidean base plane.
- **Notes:** Expression is algebraically consistent (quadratic numerator in second derivatives, quartic-like denominator). No internal contradictions in its presentation.

12. ✔ **Application of Gaussian curvature to $L_i(x, t)$** (Sec. 2.5.2, p.4)

- **Claim:** $K_i(x, t) = \frac{L_{i,xx}L_{i,tt} - (L_{i,xt})^2}{(1 + L_{i,x}^2 + L_{i,t}^2)^2}$ when treating t as the second coordinate.
- **Checks:** symbol substitution, notation consistency
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** t is used analogously to y as a coordinate in the parameter domain.
- **Notes:** Correctly mirrors the stated $z = f(x, y)$ formula under the substitution $y \rightarrow t, f \rightarrow L_i$.

13. ✓ **Gaussian curvature restatement in Results** (Sec. 3.2.1, p.6)

- **Claim:** $K_i = \frac{L_{i,xx}L_{i,tt} - L_{i,xt}^2}{(1 + L_{i,x}^2 + L_{i,t}^2)^2}$.
- **Checks:** notation consistency, cross-check with earlier definition
- **Verdict:** PASS; confidence: medium; impact: moderate
- **Assumptions/inputs:** $L_{i,xt}^2$ denotes $(L_{i,xt})^2$ (not $L_{i,(xt)}^2$).
- **Notes:** Matches Sec. 2.5.2 up to notational style. The only ambiguity is typographic: $L_{i,xt}^2$ is interpreted as $(L_{i,xt})^2$, consistent with surrounding context.

14. ✓ **Ricci scalar equals twice Gaussian curvature in 2D** (Sec. 2.5.3, p.4; reiterated Sec. 3.2, p.6)

- **Claim:** For a 2D surface, $R_i(x, t) = 2K_i(x, t)$.
- **Checks:** definition consistency
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Ricci scalar refers to the intrinsic scalar curvature of the induced 2D metric on the surface.
- **Notes:** Used consistently: compute K then multiply by 2 to obtain R .

15. ✓ **Interpretation of sign of curvature** (Sec. 2.5.3, p.4; Sec. 3.2.2, p.7)

- **Claim:** Positive R corresponds to locally elliptic (bowl-shaped) regions; negative R to hyperbolic (saddle-shaped) regions; near-zero to flat/parabolic-like regions.
- **Checks:** logical consistency
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** R and K share sign since $R = 2K$.
- **Notes:** Because $R = 2K$, sign-based interpretations are internally consistent within the paper's own definitions.

Limitations

- The provided PDF text contains few explicit, numbered equations and omits detailed derivations; many claims (e.g., NRMSE definition, exact energy computation, fitting procedure details) cannot be verified symbolically beyond checking stated formulas for consistency.

- Figures are referenced for scaling linearity and map structure; the audit does not validate graphical content or numeric outputs and cannot cross-check whether plotted quantities match the stated definitions.
- Any numerical differentiation accuracy discussions are out of scope except insofar as they affect symbol/definition consistency; discretization-specific identities (e.g., equality of mixed partials) are not audited.

Numerical results audit

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

Nine internal arithmetic/dimensional and table-to-text consistency checks were executed and all passed within stated tolerances. Several additional quantitative claims (NRMSE, wavelet energy distribution, kurtosis, power-law fit slope-to-exponent conversion, and derivative-validation error rates) could not be recomputed from the provided text/tables.

Checked items

1. ✓ **C1_dataset_shape_features_split** (Page 2, Sec. 2.1–2.2 (Dataset; Data Loading and Preprocessing))
 - **Claim:** Raw data has dimensions $[N_x, N_t, N_{\text{features}}]$ with $N_x = 100$, $N_t = 100$, $N_{\text{features}} = 12$; $L_{\text{components}} = \text{raw_data}[:, :, 2 :]$ has dimensions $(100, 100, 10)$.
 - **Checks:** shape_consistency_from_slicing
 - **Verdict:** PASS
 - **Notes:** Computed sliced feature dimension equals $N_{\text{features}} - \text{slice_start} = 12 - 2 = 10$.
2. ✓ **C2_grid_point_count_flatten_length** (Page 3, Sec. 2.3.2 (Inter-Component Correlation))
 - **Claim:** Each 100×100 latent component matrix is flattened into a 10000-element vector.
 - **Checks:** multiplication_total
 - **Verdict:** PASS
 - **Notes:** $100 * 100 = 10000$.
3. ✓ **C3_correlation_matrix_shape** (Page 3, Sec. 2.3.2 (Inter-Component Correlation))
 - **Claim:** There are 10 latent components; the correlation matrix C is 10×10 .
 - **Checks:** dimension_consistency
 - **Verdict:** PASS

- **Notes:** For $n = 10$ components, a pairwise correlation matrix is $n \times n = 10 \times 10$.
4. ✓ **C4_dwt_level_count_feasibility** (Page 5, Sec. 3.1 (Wavelet decomposition performed up to 5 levels on a 100×100 grid))
- **Claim:** DWT decomposition was performed up to 5 levels on a 100×100 grid.
 - **Checks:** log2_level_feasibility
 - **Verdict:** PASS
 - **Notes:** Feasibility check: $2^5 = 32 \leq 100$ (ignoring any padding/boundary conventions).
5. ✓ **C5_powerlaw_alpha_range_from_table1** (Page 6, Table 1; also referenced in Abstract and Conclusions)
- **Claim:** Scaling exponents α_i range from approximately -3.13 to -2.56 (Table 1 values).
 - **Checks:** min_max_from_list
 - **Verdict:** PASS
 - **Notes:** Parsed Table 1 values give min= -3.13 (L4) and max= -2.56 (L7), matching the stated range.
6. ✓ **C6_table1_count_matches_10_components** (Page 6, Table 1)
- **Claim:** Table 1 lists exponents for components L_0 through L_9 (10 entries).
 - **Checks:** count_entries
 - **Verdict:** PASS
 - **Notes:** Unique labels found: L_0 – L_9 (10 entries), consistent with 10 latent components.
7. ✓ **C7_ricci_variance_range_from_table2** (Page 7–8, Sec. 3.2.3 and Table 2)
- **Claim:** Ricci scalar variance ranges from approximately 0.093 (L3) to 0.963 (L6).
 - **Checks:** min_max_from_list
 - **Verdict:** PASS
 - **Notes:** From Table 2: min= 0.09319 at L3 and max= 0.9628 at L6; rounding to three decimals matches 0.093 and 0.963 .
8. ✓ **C8_table2_minmax_example_L1** (Page 7, Sec. 3.2.3 narrative; Page 8, Table 2)
- **Claim:** Example range for L1 is -14.51 to 4.42 .
 - **Checks:** table_lookup_consistency
 - **Verdict:** PASS

- **Notes:** Table 2 shows L1 min= -14.51 and max= 4.418 ; 4.418 rounds to 4.42 (2 decimals).
9. ✓ **C9_table2_minmax_example_L6** (Page 7, Sec. 3.2.3 narrative; Page 8, Table 2)
- **Claim:** Example range for L6 is -14.36 to 24.02 .
 - **Checks:** table_lookup_consistency
 - **Verdict:** PASS
 - **Notes:** Narrative and Table 2 values match at two-decimal precision for both min and max.

Limitations

- Only parsed text from the PDF was available; numeric series underlying figures (wavelet energies, coefficient stats) are not present, so figure-based claims cannot be recomputed without reading plot pixels (disallowed).
- Several quantitative claims (NRMSE, kurtosis, fitted slopes, derivative errors) depend on access to the latent-space arrays and intermediate wavelet/derivative outputs, which are not included in the PDF text/tables.
- Checks here focus on internal arithmetic/dimensional consistency and table-to-text consistency using explicitly stated numbers.