

Skeptical review: Unveiling Structural Discrepancies: A Manifold and Information-Theoretic Comparison of Gravitational Waveform Posteriors for GW231123

Summary

The manuscript compares high-dimensional posterior structure for GW231123 across five waveform models (NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, IMRPhenomTPHM). Using a 9-parameter source-frame set (Sec. 2.1), the authors (i) standardize samples and apply PCA to infer intrinsic dimensionality and leading degeneracy directions (Secs. 2.3, 3.2), and (ii) treat (standardized) covariance matrices as points on the SPD manifold and compute affine-invariant Riemannian distances to quantify inter-model differences, supplemented by element-wise covariance-difference heatmaps to localize drivers (Secs. 2.4, 3.3–3.4). The paper finds overall qualitative agreement that the system is high-mass and strongly precessing, but substantial waveform-model dependence in parameters such as χ_{eff} , component masses, redshift/distance, and especially orientation-related correlations; time-domain models cluster more closely, while IMRPhenomXO4a appears most structurally distinct (Secs. 3.1–3.5). The approach is timely and potentially valuable as a complement to $1\backslash\text{rm}\{D\}/2\backslash\text{rm}\{D\}$ marginals, but several core clarifications and robustness checks are needed—particularly regarding the meaning of per-model standardization (correlation vs covariance in physical units), non-Gaussian/angular posterior structure, sampling uncertainty, and reproducibility of the underlying PE setup—so that the reported “structural discrepancies” can be interpreted as genuine waveform systematics rather than artifacts of preprocessing, topology, or Monte Carlo noise.

Strengths

- Clear motivation for moving beyond 1D/2D marginals toward quantitative comparison of full multivariate posterior structure (Introduction, Sec. 1).
- Interesting and potentially useful combination of tools: PCA-based degeneracy characterization plus affine-invariant Riemannian geometry on the SPD manifold to compare posterior second-moment structure (Secs. 2.3–2.4).
- Thoughtful attempt to choose a non-redundant core parameter set (nine parameters) and keep derived parameters for interpretation rather than as analysis coordinates (Sec. 2.1).
- Coherent narrative connecting marginal disagreements to PCA loadings/alignment, then to covariance-manifold distances and element-wise covariance differences to identify key drivers (Secs. 3.1–3.4).
- The main qualitative result (time-domain models closer; IMRPhenomXO4a most discrepant; orientation-related correlations important) is plausible and, if made robust, would be valuable to the waveform-systematics community (Secs. 3.3–3.5).
- Correct handling of PCA sign ambiguity via absolute dot products for alignment, and a standard statement of the affine-invariant SPD distance formula (Secs. 2.3.4, 2.4.2).
- Figures are conceptually well-chosen (marginals \rightarrow PCA \rightarrow distances $\rightarrow \Delta C$ attribution), and several plots effectively support the manuscript’s narrative when readable.

Major issues

1. **The manuscript’s core comparisons are covariance-based (and PCA is linear), but it does not establish that the relevant posteriors are close enough to unimodal/elliptical in the chosen coordinates for second-moment summaries to be representative.** For a high-mass, strongly precessing event, posteriors can be skewed, heavy-tailed, bounded, curved, or multimodal—especially in orientation variables—so two posteriors may share similar covariances while differing substantially in shape/topology, or vice versa (Secs. 2.3–2.4, 3.2–3.4).

Recommendation: In Sec. 2.3 and Secs. 3.2–3.4, add explicit posterior-shape diagnostics per model for the most-discussed parameters (masses/redshift or distance; $\cos\theta_{jn}$; ϕ_{jl} ; tilts): e.g., 1D skewness/kurtosis (or robust alternatives), and a small set of representative 2D projections that show whether degeneracies are approximately linear/elliptical or visibly curved/multimodal. Then, either (i) clearly scope claims as “second-moment/correlation-structure comparisons,” or (ii) add one complementary sample-based multivariate discrepancy check for at least one key pair (e.g., NRSur7dq4 vs IMRPhenomXO4a): sliced Wasserstein, energy distance/MMD, or a classifier two-sample test on the same 9D space, to demonstrate that the headline conclusions (clustering; orientation-driven mismatch) persist beyond covariance summaries.

2. **Per-model standardization (StandardScaler applied separately to each waveform posterior) fundamentally changes what is being compared: it removes cross-model mean shifts and forces each marginal variance to ~ 1 , so the SPD-manifold analysis and ΔC largely probe correlation structure rather than covariance in physical units.** Several interpretations (e.g., near-zero diagonals of ΔC , statements about “complete local quadratic approximation”) read as empirical findings rather than consequences of the preprocessing choice (Secs. 2.3.1, 2.4.1, 3.4; also wording around p.4 and p.9).

Recommendation: Make the comparison target explicit in Sec. 2.3.1 and Sec. 2.4.1: if variables are standardized per model, refer to the resulting matrices as (approximately) correlation matrices and state that diagonal variance differences are not identifiable by construction. Revise the “complete local quadratic approximation” language to reflect ‘shape up to axis rescaling’ rather than full quadratic form in original coordinates. If scientific conclusions involve differences in absolute uncertainties or mean shifts, add a parallel analysis:

(i) compute SPD distances on unstandardized covariances in consistent units/coordinates (or using a shared/pool scaler), and/or (ii) separately compare mean vectors and marginal variances across models (e.g., a table/plot of σ differences before standardization). Clearly separate results that come from correlation-structure differences versus scale/location differences.

3. **Angular/topological variables are treated with Euclidean linear tools without sufficient specification of conventions and wrap handling, even though the main discrepancies are attributed to orientation correlations (notably $\cos \theta_{jn} - \phi_{ji}$).** For periodic angles (ϕ_{ji}/ϕ_{j1}) and bounded variables ($\cos \theta_{jn}$, $\cos \theta_{jn} - \phi_{ji}$), naive covariances can be dominated by the branch cut/wrapping choice, leading to artificial correlations or model-dependent artifacts (Secs. 2.1, 2.3.4, 3.2.2, 3.4). There is also a notation inconsistency (ϕ_{ji}) vs ϕ_{j1} .

Recommendation: In Sec. 2.1, provide explicit definitions, ranges, and reference-frame conventions for $\cos \theta_{jn}$, ϕ_{ji} (choose one symbol consistently), and describe how samples from all pipelines are transformed into one common convention before PCA/covariances. In Sec. 3.4, demonstrate correlation findings under a wrap-safe representation: e.g., replace ϕ_{ji} (ΔC entries and model-distance ordering persist. If the representation changes dimensionality, state this clearly and (if needed) present the angular-robust check as a focused appendix/supplement.) with $\sin \phi_{j1}$, $\cos \phi_{j1}$.

4. **Key parameter-estimation (PE) configuration details are missing, limiting reproducibility and making it difficult to attribute differences to waveform physics rather than run-to-run analysis choices (priors, PSD estimation, calibration marginalization, $f_{\text{low}}/f_{\text{high}}$, reference frequency, sampler settings, reweighting, etc.) (Secs. 2.1–2.2).** This also weakens the discussion that attributes differences to frequency-domain approximations (Secs. 3.5.2, 4).

Recommendation: Add a dedicated PE-setup subsection (e.g., Sec. 2.1.1) listing: detectors and data segment (GPS, duration), PSD estimation method, calibration-uncertainty treatment, frequency bounds and reference frequency, priors for the nine core parameters (and any fixed cosmology used to map distance \leftrightarrow redshift, if applicable), sampler/inference engine and settings, any reweighting, and convergence diagnostics. Explicitly state that all waveform-model runs used identical settings except for the waveform model (or enumerate differences and assess their likely impact). In Secs. 3.5.2 and 4, temper causal claims about specific waveform approximations unless supported by targeted controls; otherwise frame them as hypotheses.

5. **Uncertainty/robustness of the quantitative metrics is not assessed.** Covariances, PCA directions, and SPD distances can be sensitive to finite effective sample size (ESS), autocorrelation, and near-singular covariance estimation in 9D; without uncertainty bands, it is unclear whether distance differences (e.g., 4.17 vs 3.85) or alignment differences are meaningful (Secs. 2.3–2.4, 3.2–3.3). The SPD requirement is assumed; sample covariance is only guaranteed PSD, not necessarily strictly PD (Sec. 2.4.2).

Recommendation: Report per-model posterior sample counts and (at minimum) an ESS estimate (Sec. 2.1). Add bootstrap/jackknife (or repeated subsampling) to quantify uncertainties on: explained-variance curves/intrinsic dimensionality (Fig. 2), leading PC loadings/alignment (Fig. 4), and pairwise Riemannian distances (Fig. 5/related). State the condition for SPD and what you do if matrices are ill-conditioned (e.g., shrinkage regularization $C \rightarrow C + \epsilon I$; monitor condition numbers) (Sec. 2.4.2). Present distance values with uncertainty (e.g., mean \pm sd across bootstrap) and comment on stability of the model clustering and the “most distant” model conclusion (Sec. 3.3).

6. **The PCA alignment methodology may be misleading because it matches the k -th PC across models by index and uses $|\text{PC}_k^A \cdot \text{PC}_k^B|$, but if eigenvalues are near-degenerate, directions can rotate within the dominant subspace and the notion of a uniquely defined ‘PC1 vs PC1’ comparison breaks down (Secs. 2.3.4, 3.2.3).**

Recommendation: In Sec. 3.2 (and/or Fig. 2), show eigenvalue spectra (not only cumulative variance) to indicate separation/degeneracy. Complement the current alignment plot with a subspace-based comparison: principal angles between the top- k subspaces, or Procrustes alignment within the top- k space. When discussing misalignment, explicitly note possible near-degeneracy and interpret results at the subspace level when appropriate.

7. **Some astrophysical interpretations and generalizations overreach the presented evidence, especially formation-channel implications tied to χ_{eff} when χ_{eff} is shown to be waveform-model dependent for this event; additionally, results are based on a single event, so ‘population-level’ implications are not directly supported (Secs. 3.1, 3.5.1–3.5.2, 4).**

Recommendation: In Secs. 3.5 and 4, clearly separate (i) methodological conclusions likely to generalize (the multivariate comparison toolkit) from (ii) event-specific numerical patterns (e.g., IMRPhenomXO4a being the most discrepant). Soften or qualify formation-channel language and emphasize limitations from waveform dependence, priors, and selection effects. If feasible, add a brief roadmap for applying the framework to multiple events (even as future work) and cite relevant population/systematics studies for context.

Minor issues

1. Intrinsic dimensionality is defined via a single cumulative explained-variance threshold (95%), and sensitivity to this choice is not explored; similarly, PCA’s linear nature and its limitations for curved degeneracies are not emphasized enough (Secs. 2.3.3, 3.2.1–3.2.2).

Recommendation: Justify the 95% choice in Sec. 2.3.3 and add a short sensitivity check (e.g., 90%/99%) in Sec. 3.2.1 (small table or appendix). Add a short explicit note that PCA captures linear correlations around the mean and may not represent curved degeneracies well; briefly mention possible nonlinear extensions (kernel PCA, diffusion maps) as future work (Sec. 2.3 or 3.2).

2. Interpretation of the affine-invariant Riemannian distance magnitude is not intuitive: readers may not know what a distance of ~ 2 vs ~ 4 means in terms of covariance/correlation differences (Sec. 3.3).

Recommendation: Add a short interpretation aid in Sec. 3.3: relate the distance to the log-eigenvalues of $C_A^{-1}C_B$ (e.g., typical multiplicative discrepancy in principal variances/correlations), or provide a simple toy example mapping distance values to eigenvalue-ratio ranges.

3. Metric robustness to alternative choices is not shown; it would strengthen confidence to compare the SPD-manifold distances with simpler baselines (e.g., Frobenius norm of covariance/correlation differences) (Secs. 2.4.2, 3.3).

Recommendation: Add a short comparison (main text or appendix) showing whether the qualitative ranking/clustering is consistent under at least one simpler metric (e.g., Frobenius norm on correlation matrices) and state what changes, if any.

4. Parameter-basis dependence is insufficiently explored. Using $(m_{1,\text{source}}, m_{2,\text{source}}, z)$ may obscure measurement-adapted degeneracies better expressed in (chirp mass, mass ratio, distance or z), and source-frame vs detector-frame choices can change linear structure (Secs. 2.1, 2.3.4, 3.2).

Recommendation: Add a robustness check repeating the PCA/correlation-structure analysis under at least one alternative non-redundant parameterization (e.g., chirp mass, symmetric mass ratio, z or D_L ; keep the remaining spin/orientation parameters consistent). Summarize whether the key conclusions (time-domain clustering; orientation-driven discrepancies) persist (Sec. 3.2; appendix acceptable).

5. Figures (especially Fig. 1) are hard to read due to dense overplotting, small fonts, and limited accessibility; some captions/labels mix code-style names and lack units/frame annotations (Figs. 1, 2, 4, 6; Sec. 3.1).

Recommendation: Improve readability: increase size or split Fig. 1; use colorblind-safe palettes plus line styles; move legends outside; label axes with symbols/units and frame (source vs detector); overlay medians/credible intervals. For matrix plots, show only one triangle, annotate key values, and ensure parameter ordering is stated in the caption (Figs. 2/4/6).

6. Element-wise ΔC plots are not accompanied by uncertainty or by the underlying matrices, making it hard to judge significance and interpretability; standardization reference is also easy to miss (Sec. 3.4; Fig. 6).

Recommendation: Add uncertainty estimates for ΔC entries via bootstrap (e.g., mark statistically stable largest-magnitude entries), and consider providing the underlying correlation matrices (or a representative subset) in an appendix/supplement. In captions, state explicitly whether matrices are computed after per-model standardization (dimensionless correlation structure).

7. Related-work context for multivariate posterior-comparison tools (beyond standard corner-plot comparisons) and for waveform-systematics quantification is thin, making novelty harder to place (Introduction, Sec. 2).

Recommendation: Add a short related-work paragraph/subsection (between Sec. 1 and 2 or within Sec. 1) citing prior PCA/degeneracy analyses in GW PE, posterior-comparison metrics used in LVC/systematics contexts (e.g., KL-based or other divergences), and any prior use of SPD/covariance geometry in similar settings. Clearly state what is new in this paper’s combination and application.

8. Reproducibility of the presented numeric claims is limited: the manuscript does not clearly state whether posterior samples, derived covariance/correlation matrices, PCA eigenvectors, and distance matrices will be made available (Sec. 2; throughout).

Recommendation: Provide links/DOIs to posterior samples (or at least the standardized correlation matrices, PCA loadings/eigenvalues, distance matrices, and ΔC matrices) and code (with versions). If sharing full samples is not possible, provide machine-readable derived products sufficient to reproduce Figs. 2–6 and key numeric statements.

Very minor issues

1. Notation and naming inconsistencies (e.g., IMRPhenomXO4a vs IMRPhenomX04a; SEOBNRv5PHM spelling; ϕ_{jI} vs ϕ_{j1} ; Frobenius norm notation) and minor typesetting artifacts (split words, inconsistent percent formatting) distract from an otherwise technical presentation (Introduction; Secs. 2.1–2.4, 3.1–3.5, 4).

Recommendation: Proofread and standardize model names and symbols throughout; fix split-word artifacts; harmonize percent/math formatting; and use a single Frobenius norm notation with a clearly parenthesized matrix-log argument (Sec. 2.4.2).

2. Some long sentences—especially in the Introduction and the waveform-approximation discussion—reduce readability for non-specialists (Introduction; Sec. 3.5.2).

Recommendation: Edit a small number of multi-clause sentences into shorter statements; explicitly separate demonstrated results from hypotheses about waveform-physics causes.

3. Captions sometimes omit key methodological specifics (e.g., KDE settings; whether samples are weighted; standardization choice; parameter ordering in heatmaps), which makes standalone interpretation harder (Figs. 1, 2, 4, 6).

Recommendation: Augment captions with: sample weighting (if any), KDE bandwidth choice (if relevant), standardization/scaling reference, parameter ordering for matrices, and whether plotted values are correlations or covariances.

4. The Conclusions (Sec. 4) partly repeats detailed numeric results already covered in Sec. 3.5, which dilutes the highest-level takeaways.

Recommendation: Streamline Sec. 4 to emphasize: (i) what the method adds, (ii) the most robust empirical findings, and (iii) limitations/future work; refer back to Sec. 3 for detailed numbers.

Key statements and references

- • For the time-domain models NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM, the first principal component (PC1) of the standardized 9-parameter posterior is dominated by a strong anti-correlation between the component source-frame masses ($m_{1,\text{source}}, m_{2,\text{source}}$) and the redshift, reflecting the well-known degeneracy whereby a heavier, more distant binary can produce a waveform similar to a lighter, closer system due to amplitude and frequency evolution of the signal [Figure 3].
- *Reference(s)*: Figure 3
- • In contrast to the time-domain models, the frequency-domain phenomenological model IMRPhenomXPHM has a PC1 that is primarily governed by a $m_{1,\text{source}}$ –redshift anti-correlation with comparatively weak contribution from $m_{2,\text{source}}$, while IMRPhenomXO4a’s PC1 includes strong contributions from spin parameters ($a_1, \cos\text{tilt}_1$) in addition to mass and redshift, indicating that its dominant degeneracy involves a coupled mass–distance–spin structure rather than a simple mass–redshift trade-off [Figure 3].
- *Reference(s)*: Figure 3
- • Quantitative comparison of principal-component orientations shows that the PC1 directions of the time-domain models NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM are highly aligned, with absolute dot products exceeding 0.81, whereas the alignment between the two frequency-domain models IMRPhenomXO4a and IMRPhenomXPHM is only 0.17, implying that their leading degeneracy directions are nearly orthogonal and structurally distinct [Figure 4].
- *Reference(s)*: Figure 4
- • Using the affine-invariant Riemannian metric $d(C_A, C_B) = \left| \log \left(C_A^{-1/2} C_B C_A^{-1/2} \right) \right|_F$ on the 9×9 covariance matrices, the largest pairwise distance found is 4.17 between NRSur7dq4 and IMRPhenomXO4a, while NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM form a relatively coherent cluster with mutual distances in the range 2.30–3.51, and IMRPhenomXPHM is most similar to IMRPhenomTPHM with a distance of 2.22 [Figure 5].
- *Reference(s)*: Figure 5
- • Element-wise comparison of the standardized covariance matrices for NRSur7dq4 and IMRPhenomXO4a shows that the largest discrepancy is in the covariance between $\cos\theta_{j1}$ (inclination) and ϕ_{j1} (spin azimuth), with a difference of -0.86 , and additional substantial differences in covariances such as $m_{2,\text{source}}-\cos\theta_{j1}$ (-0.37), $m_{1,\text{source}}-m_{2,\text{source}}$ (0.34), and $\text{redshift}-\phi_{j1}$ (-0.42), demonstrating that the dominant structural disagreements arise from how these models encode correlations between orientation, masses, and distance [Figure 6].
- *Reference(s)*: Figure 6

Mathematical consistency audit

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

Maths relevance: substantial

The paper’s core methodology is mathematical: PCA on a 9D standardized parameter space, pairwise alignment of principal components via absolute dot products, and a Riemannian-manifold (affine-invariant) distance between (standardized) covariance matrices using a matrix logarithm and Frobenius norm. The main internal-consistency concern is that per-model standardization makes the matrices effectively correlation matrices (diagonal fixed), yet the text interprets diagonal differences and “full covariance/shape” claims as if variances in original units were retained.

Checked items

- ✓ **Affine-invariant Riemannian distance formula** (Sec. 2.4.2, p.4 (also stated in Intro, p.2))
 - **Claim:** Defines distance between two SPD covariance matrices as $d(C_A, C_B) = \left| \log \left(C_A^{-1/2} C_B C_A^{-1/2} \right) \right|_F$.
 - **Checks:** algebra/definition consistency, notation consistency, domain/assumption check (SPD requirement)
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** C_A and C_B are symmetric positive definite (SPD), $\log(\cdot)$ denotes the matrix logarithm, $|\cdot|_F$ is the Frobenius norm
 - **Notes:** The formula is internally consistent and correctly defined given SPD matrices. Symmetry and nonnegativity follow from properties of the matrix log and Frobenius norm, but SPD-ness is assumed rather than ensured.
- ✓ **Frobenius norm definition** (Sec. 2.4.2, p.4)
 - **Claim:** Defines $\|M\|_{\text{Fas}} = \sqrt{\sum_{ij} |M_{ij}|^2}$
 - **Checks:** definition correctness, notation/clarity
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Matrix entries may be real; absolute value is harmless, Summation is over all indices
 - **Notes:** Correct definition; typesetting is slightly ambiguous but interpretable.
- ✓ **Standardization step (zero mean, unit variance)** (Sec. 2.3.1, p.3)

- **Claim:** Each parameter is transformed to have mean 0 and variance 1 per model prior to PCA and covariance analysis.
 - **Checks:** definition consistency, implications for later covariance computations
 - **Verdict:** PASS; confidence: high; impact: critical
 - **Assumptions/inputs:** StandardScaler is applied independently for each model, Scaling uses each model’s own sample mean and sample standard deviation
 - **Notes:** Mathematically consistent; however, this implies subsequent covariances are computed in standardized units and will have unit diagonals (up to estimator conventions).
4. ✓ **Covariance of standardized data equals correlation matrix** (Sec. 2.4.1, p.4 (implied by Sec. 2.3.1, p.3))
- **Claim:** Computes 9×9 sample covariance matrices from standardized samples for each model.
 - **Checks:** dimensional/unit consistency, logical implication check
 - **Verdict:** PASS; confidence: high; impact: critical
 - **Assumptions/inputs:** Each variable is standardized to unit variance per model, Covariance is computed on these standardized variables
 - **Notes:** Given per-model unit-variance scaling, these matrices are (approximately) correlation matrices; they no longer encode original marginal variances. Later interpretation should reflect this.
5. ✘ **Interpretation of ΔC diagonal near zero after standardization** (Sec. 3.4, p.9)
- **Claim:** Because diagonal elements of ΔC are near zero, this indicates relative parameter spreads are broadly similar across models after standardization, and major discrepancies lie in off-diagonals.
 - **Checks:** logical consistency with earlier definitions
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** $\Delta C = C_A - C_B$ where C_A and C_B are covariances of standardized variables computed separately per model
 - **Notes:** With per-model standardization, $\text{diag}(C_A) \approx \text{diag}(C_B) \approx 1$ by construction, hence $\text{diag}(\Delta C) \approx 0$ is essentially forced and cannot be used to infer similarity of relative spreads across models. The conclusion that ‘differences therefore lie in off-diagonals’ is an artifact of the preprocessing choice, not necessarily a property of the original posteriors.
6. ✘ **Claim that standardized covariances represent “complete local quadratic approximation”** (Sec. 2.4.1, p.4)
- **Claim:** States that the computed covariance matrices represent the complete local quadratic approximation of the posterior’s shape.
 - **Checks:** conceptual-mathematical consistency (coordinate dependence)
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** Covariances are computed after per-model standardization (Sec. 2.3.1)
 - **Notes:** After per-model standardization, the matrices cannot represent the full quadratic form in the original coordinates because all marginal scales are removed. They represent correlation structure (shape/orientation up to axis rescaling), not the full local metric in physical parameter units.
7. ✓ **PCA eigenvector/eigenvalue relationship** (Sec. 2.3.2, p.3)
- **Claim:** Principal components are eigenvectors of the covariance matrix of standardized data; explained variances are the corresponding eigenvalues.
 - **Checks:** linear algebra definition check
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Covariance matrix is computed in the standardized space, PCA uses orthonormal eigenvectors
 - **Notes:** Consistent with PCA on centered data; standardization ensures variance comparability across parameters.
8. ✓ **Intrinsic dimensionality definition via cumulative explained variance threshold** (Sec. 2.3.3, p.3 and Sec. 3.2.1, p.6)
- **Claim:** Defines intrinsic dimensionality as the minimum number of PCs needed to reach 95% cumulative explained variance.
 - **Checks:** definition consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Explained variance ratios are computed from PCA eigenvalues, Cumulative sum is taken in descending eigenvalue order
 - **Notes:** Definition is clear and consistent; it is a chosen operational definition (not a theorem).
9. ✓ **PC alignment metric via absolute dot product** (Sec. 2.3.4, p.4)
- **Claim:** Quantifies alignment of k -th PCs between models as $|\text{PC}_k^A \cdot \text{PC}_k^B|$, with values near 1 indicating similar directions, near 0 indicating orthogonality.
 - **Checks:** geometry/algebra check, assumption check (normalization)
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** PC vectors are unit-norm, Sign ambiguity is handled by absolute value
 - **Notes:** Correct provided PCs are normalized (typical in PCA). Absolute value properly removes arbitrary sign flips.
10. △ **Stability of k -th PC matching across models** (Sec. 2.3.4, p.4 (method) and Sec. 3.2.3, p.7-8 (interpretation))

- **Claim:** Directly compares k -th PC across models as if it denotes the same degeneracy mode.
 - **Checks:** well-posedness/identifiability check
 - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Eigenvalues are well-separated so PCs are uniquely defined, No near-degenerate eigenspaces causing rotations/mode swapping
 - **Notes:** If eigenvalues are close, the k -th PC direction is not stable/unique; comparisons should then be made at the subspace level. The paper does not show eigenvalue gaps, so the validity of one-to-one PC matching cannot be verified from the PDF.
11. **△ SPD assumption for covariance matrices used in manifold distance** (Sec. 2.4.2, p.4 and Sec. 3.3, p.8)
- **Claim:** Treats sample covariance matrices as points on the SPD manifold for computing affine-invariant distance.
 - **Checks:** domain/assumption check
 - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
 - **Assumptions/inputs:** Each sample covariance is strictly positive definite, Matrix inverse square root and matrix logarithm are well-defined
 - **Notes:** Sample covariance is guaranteed PSD, not necessarily SPD; SPD requires full rank. The paper states parameter choice avoids linear dependencies, but does not state sample-rank/regularization handling. Without that, the mathematical preconditions for $C_A^{-1/2}$ and $\log(\cdot)$ cannot be confirmed.
12. **✓ Element-wise covariance difference interpretation (off-diagonals)** (Sec. 2.4.3, p.4 and Sec. 3.4, p.9)
- **Claim:** Off-diagonal entries of ΔC indicate differences in covariances/correlations and can be used to attribute discrepancies to specific parameter-pair relationships.
 - **Checks:** definition consistency, unit/scale consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Matrices compared are computed in the same coordinate system, Here, variables are standardized per model
 - **Notes:** Given per-model standardization, off-diagonals correspond to differences in correlations (not raw covariances). Attribution to differing correlation structure is mathematically consistent; the paper should avoid interpreting these as covariance-in-units differences.

Limitations

- Audit is restricted to the provided PDF text and figures; no underlying code, sample sizes, estimator conventions (population vs unbiased covariance), or preprocessing details beyond narrative are available.
- Figures contain numerical values (e.g., distances, heatmap entries) that were not numerically validated per the scope; only the analytic meaning/consistency of the formulas and interpretations was assessed.
- No formal derivations are provided in the paper beyond definitions; where methodological steps rely on unstated conditions (e.g., SPD-ness, eigenvalue separation), those are marked UNCERTAIN.

Numerical results audit

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

12 candidate numerical checks were identified (PCA dimensionality counts; PC1/PC2 alignment thresholds/values; Riemannian distance matrix properties and specific extrema/ranges; standardized covariance-difference entries). All 12 checks are UNCERTAIN because required posterior-sample-derived quantities were not available to compute the comparisons.

Checked items

1. **△ C1_intrinsic_dimensionality_counts** (Page 6, Sec. 3.2.1 (Intrinsic dimensionality) and Fig. 2 caption)
 - **Claim:** NRSur7dq4, IMRPhenomXO4a, and SEOBNRv5PHM each require 7 PCs to explain 95% variance; IMRPhenomXPHM and IMRPhenomTPHM require 8 PCs.
 - **Checks:** threshold_count_from_cumulative_variance
 - **Verdict:** UNCERTAIN
 - **Notes:** Insufficient inputs to compute PCA cumulative explained variance thresholds; requires posterior samples or PCA eigenvalues/eigenvectors per model.
2. **△ C2_alignment_range_claim_time_domain_PC1** (Page 8, Sec. 3.2.3)
 - **Claim:** For PC1, the time-domain models (NRSur7dq4, SEOBNRv5PHM, IMRPhenomTPHM) are highly aligned, with dot products exceeding 0.81.
 - **Checks:** pairwise_dot_product_threshold
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute PC1 vectors and their pairwise absolute dot products without posterior samples or precomputed PCA components.
3. **△ C3_alignment_value_XO4a_vs_XPHM_PC1** (Page 8, Sec. 3.2.3)

- **Claim:** The alignment between IMRPhenomXO4a and IMRPhenomXPHM for PC1 is 0.17.
 - **Checks:** pairwise_dot_product_value
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute $\|\text{PC1}\|_{\text{rm } XO4a} \cdot \|\text{PC1}\|$ without posterior samples or PCA eigenvectors.
4. \triangle **C4_alignment_value_NRSur_vs_TPHM_PC2** (Page 8, Sec. 3.2.3)
- **Claim:** NRSur7dq4 and IMRPhenomTPHM show exceptional alignment for PC2 (0.91).
 - **Checks:** pairwise_dot_product_value
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute $\|\text{PC2}\|_{\text{rm } NRSur} \cdot \|\text{PC2}\|$ without posterior samples or PCA eigenvectors.
5. \triangle **C5_alignment_value_SEOBNRv5PHM_with_group_PC2** (Page 8, Sec. 3.2.3)
- **Claim:** SEOBNRv5PHM is also closely aligned (0.81) for PC2 with the time-domain group.
 - **Checks:** pairwise_dot_product_value_or_min
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute PC2 dot products for (SEOBNR, NRSur) and (SEOBNR, TPHM) without posterior samples or PCA eigenvectors.
6. \triangle **C6_poor_alignment_threshold_PC2_frequency_vs_time** (Page 8, Sec. 3.2.3)
- **Claim:** Frequency-domain models show poor PC2 alignment (< 0.3) with the time-domain group and with each other.
 - **Checks:** pairwise_dot_product_upper_bound
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute the required set of PC2 absolute dot products without posterior samples or PCA eigenvectors.
7. \triangle **C7_riemann_distance_symmetry_and_diagonal** (Page 4, Sec. 2.4.2 and Page 8, Sec. 3.3)
- **Claim:** Pairwise distances were compiled into a 5×5 symmetric distance matrix, where diagonal entries are zero.
 - **Checks:** matrix_property_check
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute the 5×5 affine-invariant Riemannian distance matrix without standardized covariance matrices (or posterior samples to derive them).
8. \triangle **C8_riemann_max_distance_value** (Page 8, Sec. 3.3 and Fig. 5 caption)
- **Claim:** The largest distance observed (4.17) is between NRSur7dq4 and IMRPhenomXO4a.
 - **Checks:** argmax_and_value_check
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot identify the maximum off-diagonal distance or verify the 4.17 value without computed pairwise distances.
9. \triangle **C9_riemann_cluster_range_time_domain** (Page 8, Sec. 3.3)
- **Claim:** NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM have pairwise distances ranging from 2.30 to 3.51.
 - **Checks:** min_max_range_check
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute the three pairwise distances among the time-domain trio to verify the stated min/max range.
10. \triangle **C10_riemann_XPHM_most_similar_to_TPHM** (Page 8, Sec. 3.3 and Fig. 5 caption)
- **Claim:** IMRPhenomXPHM is most similar to IMRPhenomTPHM (distance 2.22).
 - **Checks:** rowwise_argmin_and_value_check
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot determine XPHM’s closest model by distance or verify the 2.22 value without computed distances.
11. \triangle **C11_covdiff_major_entry_cosTheta_phiJL** (Page 9, Sec. 3.4 and Fig. 6 caption)
- **Claim:** The most significant discrepancy is the covariance between $\cos\theta_{jn}$ and ϕ_{jl} (-0.86) in $\Delta C = C_{\text{NRSur}} - C_{\text{IMRPhenomXO4a}}$ (standardized space).
 - **Checks:** covariance_difference_entry_value_and_extremum
 - **Verdict:** UNCERTAIN
 - **Notes:** Cannot compute standardized covariance matrices or ΔC entries (and cannot check whether this is the largest-magnitude off-diagonal) without posterior samples or the matrices.
12. \triangle **C12_covdiff_other_listed_entries** (Page 9, Sec. 3.4 bullet list)
- **Claim:** Other significant covariance differences in $\Delta C = C_{\text{NRSur}} - C_{\text{IMRPhenomXO4a}}$ include: $m_{2,\text{source}} - \cos\theta_{jn}$ (-0.37), $m_{1,\text{source}} - m_{2,\text{source}}$ (0.34), redshift- ϕ_{jl} (-0.42).
 - **Checks:** covariance_difference_entry_values
 - **Verdict:** UNCERTAIN

- **Notes:** Cannot compute or extract the cited ΔC off-diagonal entries without posterior samples or the covariance-difference matrix.

Limitations

- Only the provided PDF text/images were used; several referenced numeric tables (e.g., Table 1) are not available in the extracted content, preventing extraction of explicit values.
- Figure-based values not explicitly stated in the text (heatmap cell annotations, curve values) were not extracted because pixel/plot-value reading is out of scope per instructions.
- Many checks require access to the underlying posterior samples (CSV files) described in Methods; without those files, the proposed FAST checks can be written but not executed/validated here.