Physics-Informed Discrepancy Decomposition and Robust Astrophysical Inference for GW231123

Denario[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

Robust astrophysical interpretations from gravitational-wave parameter inference critically depend on understanding model-dependent biases. We introduce a novel physics-informed framework to systematically decompose and attribute discrepancies among five gravitational-wave waveform models (NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, IMRPhenomTPHM) for the GW231123 event. Our methodology involves extensive exploratory data analysis using Jensen-Shannon Divergence and Wasserstein distance, high-dimensional degeneracy analysis via Uniform Manifold Approximation and Projection (UMAP), and a core Physics-Informed Discrepancy Decomposition. This decomposition quantifies multi-dimensional divergences within physically motivated parameter subspaces (mass and distance, effective spin, individual spin and orientation, remnant properties), enabling us to link model differences to specific physical approximations. Our analysis reveals significant disagreements in inferred parameters, notably for component masses, effective spin, and redshift, with UMAP embedding clearly separating models into distinct clusters in the high-dimensional parameter space. The physics-informed decomposition attributes these discrepancies: the individual spin and orientation subspace exhibits the most severe model dependence, directly linked to differing treatments of spin precession, while remnant properties are sensitive to merger-ringdown modeling. Crucially, we find that no key astrophysical parameter for GW231123 is robustly constrained across all five models, demonstrating that systematic waveform model uncertainties often exceed statistical uncertainties. This work underscores that for high-mass, precessing binary black hole mergers, waveform model choice is a dominant factor, precluding firm astrophysical conclusions without accounting for these model-dependent biases.

1. INTRODUCTION

The advent of gravitational-wave (GW) astronomy, initiated by the direct detection of binary black hole (BBH) mergers by the LIGO-Virgo-KAGRA (LVK) collaboration, has fundamentally reshaped our understanding of the most energetic astrophysical events. Through the meticulous analysis of these GW signals, we can infer fundamental properties of black holes, explore their formation mechanisms, and probe the nature of spacetime under extreme conditions. However, the reliability and precision of these astrophysical interpretations are critically dependent on the theoretical waveform models employed to describe the intricate dynamics of merging compact objects. These models, which are computationally efficient approximations of Einstein's field equations, inherently incorporate varying levels of physical fidelity, ranging from highly accurate but computation-ally expensive numerical relativity (NR) simulations to more efficient analytical and phenomenological approximations.

The inherent complexity of BBH dynamics, particularly for systems characterized by high masses, significant spins, and orbital precession, necessitates the use of such approximate models. While NR simulations serve as invaluable benchmarks, their immense computational cost prohibits their routine application in large-scale parameter inference campaigns. Consequently, a diverse suite of semi-analytical and phenomenological models has been developed, each presenting unique trade-offs in terms of computational efficiency, the inclusion of higher-order waveform modes, and the treatment of spin precession.

This inherent diversity among waveform models inevitably introduces model-dependent biases into the inferred astrophysical parameters. This poses a substan-

tial challenge to deriving robust scientific conclusions, particularly for events like GW231123, which exhibits characteristics indicative of a high-mass and potentially highly precessing system. For such events, the systematic uncertainties arising from waveform modeling can become a dominant factor, potentially outweighing the statistical uncertainties inherent in the measurement process itself. The core problem extends beyond merely observing discrepancies between models; it demands an understanding of *why* these models yield different results, and crucially, which specific physical aspects of the source are most sensitive to these model approximations. Without a clear and systematic understanding of these dependencies, our capacity to confidently constrain astrophysical properties is severely hampered.

In this paper, we introduce and implement a novel, physics-informed framework designed to systematically decompose and attribute discrepancies among multiple gravitational-wave waveform models for the GW231123 event (Li et al. 2025). We rigorously investigate five distinct waveform models: NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, and IMRPhenomTPHM (Li et al. 2025).

Our central innovation, termed "Physics-Informed Discrepancy Decomposition," transcends simple global comparisons of posterior distributions. Instead, we define a set of physically motivated parameter subspaces, specifically those related to the binary's mass and distance, the effective spin, the individual spin components and orbital orientation, and the properties of the final remnant black hole. By meticulously quantifying multi-dimensional divergences within these targeted subspaces, our aim is to establish direct correlations between observed discrepancies in inferred parameters and known differences in the underlying physical approximations of the waveform models, such as their treatment of spin precession or the inclusion of higher-order waveform modes (Li et al. 2025).

To verify the efficacy and provide comprehensive insights into our approach, we first perform extensive exploratory data analysis. This includes computing one-dimensional marginal posterior comparisons using advanced statistical metrics such as the Jensen-Shannon Divergence (JSD) and the 1-Wasserstein distance, establishing a baseline of model agreement and disagreement. We then employ Uniform Manifold Approximation and Projection (UMAP) to visualize and analyze the complex high-dimensional degeneracies and discrepancies across the full parameter space, revealing how different models occupy and cluster within this space.

The subsequent physics-informed decomposition provides quantitative metrics of disagreement within each of our predefined physical subspaces. By systematically correlating these subspace-specific discrepancies with the known physical characteristics and approximation schemes of the waveform models, we robustly identify which astrophysical properties of GW231123 are consistently constrained across all models and which remain highly sensitive to specific waveform model approximations. This comprehensive methodology allows us to provide clear, interpretable insights into the robustness of astrophysical inferences and to derive a more reliable set of consensus constraints for GW231123, thereby advancing our ability to extract confident scientific knowledge from complex gravitational-wave signals.

2. METHODS

2.1. *Data Acquisition and Pre-processing*

Our analysis initiates with the acquisition and meticulous pre-processing of posterior samples derived from the gravitational-wave event GW231123. These samples, representing the probability distributions of various source parameters, were generated using five distinct gravitational-wave waveform models: NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, and IMRPhenomTPHM. Each model's posterior samples were provided as individual CSV files, specifically located at '/mnt/ceph/users/fvillaescusa/AstroPilot/GW/Iteration1/data/GW '/mnt/ceph/users/fvillaescusa/AstroPilot/GW/Iteration1/data/GW '/mnt/ceph/users/fvillaescusa/AstroPilot/GW/Iteration1/data/GW '/mnt/ceph/users/fvillaescusa/AstroPilot/GW/Iteration1/data/GW and '/mnt/ceph/users/fvillaescusa/AstroPilot/GW/Iteration1/data

Upon loading, each CSV file was parsed into a separate pandas DataFrame (Shanbhag & Chimalakonda 2022). To facilitate unified analysis while preserving model attribution, a 'model' column was appended to each DataFrame, explicitly identifying the waveform model from which the samples originated. These individual DataFrames were then consolidated into a single, master Python dictionary, with model names serving as keys, providing a structured and accessible representation of the entire dataset.

A critical step in pre-processing involved thorough data cleaning and verification (Khan et al. 2019; Lima et al. 2024). This included confirming the consistency of column names across all files, checking for the presence of any 'NaN' or missing values (none were found, as expected for posterior samples of this nature) (Khan et al. 2019), and verifying the sensible range of *log_likelihood* values.

2.2. *Exploratory Data Analysis and Baseline Comparison*

Prior to undertaking advanced discrepancy decomposition, we performed an extensive exploratory data analysis to establish a baseline understanding of the agreements and disagreements among the five waveform models. This phase provided initial quantitative insights into the parameter inferences for GW231123 (Cuceu et al. 2025; Collaboration et al. 2025).

2.2.1. *Summary Statistics*

For each of the five waveform models, we computed key summary statistics for the following astrophysical parameters: *mass_1_source* (primary component mass), *mass_2_source* (secondary component mass), *chi_eff* (effective inspiral spin parameter), *chi_p* (precessing spin parameter), *redshift*, *final_mass_source* (remnant black hole mass), and *final_spin* (remnant black hole spin). Specifically, we calculated the median and the 90% credible interval (defined by the 5th and 95th percentiles) for the 1D marginal posterior distribution of each parameter. These statistics were compiled into a comprehensive table, offering an immediate, quantitative overview of the central tendencies and uncertainties predicted by each model.

2.2.2. *Pairwise Statistical Divergence*

To rigorously quantify the disagreement between the 1D marginal posterior distributions of each parameter across all model pairs, we employed two robust statistical divergence metrics: the Jensen-Shannon Divergence (JSD) (Nielsen 2022; Hoyos-Osorio & Sanchez-Giraldo 2024) and the 1-Wasserstein distance. For each parameter, and for every unique pair of waveform models, the following procedure was applied:

1. The 1D marginal posterior samples for the given parameter from each model were extracted.

2. A Kernel Density Estimator (KDE) was used to estimate the Probability Density Function (PDF) for each set of samples. A common, optimized bandwidth (e.g., determined by Scott's rule or Silverman's rule) was applied across all PDFs for a given parameter to ensure consistent smoothing.

3. The JSD was calculated between the estimated PDFs of the two models. The JSD is a symmetric and finite measure of the similarity between two probability distributions, ranging from 0 (identical distributions) to 1 (maximally divergent distributions), and is based on the Kullback-Leibler divergence.

4. The 1-Wasserstein distance (also known as Earth Mover's Distance) was computed between the empirical distributions of the two models. This metric quantifies the minimum cost of transforming one distribution into the other, effectively measuring the "distance" between probability distributions.

This process yielded two 5x5 symmetric matrices for each key astrophysical parameter, one for JSD values and one for 1-Wasserstein distances. These matrices served as a quantitative baseline for understanding the degree of agreement or disagreement between models on a parameter-by-parameter basis, highlighting where significant univariate discrepancies first emerge.

2.3. *High-Dimensional Degeneracy and Discrepancy Analysis*

To gain a holistic understanding of how the different waveform models populate the high-dimensional parameter space and to visualize complex degeneracies, we employed Uniform Manifold Approximation and Projection (UMAP) (Ghojogh et al. 2021; Vela et al. 2024).

2.3.1. *Data Preparation for UMAP*

All posterior samples from the five waveform models were combined into a single, large DataFrame. This consolidated dataset included all 13 physical parameters typically inferred for binary black hole mergers. To ensure that parameters with differing scales did not disproportionately influence the dimensionality reduction, all parameter columns were standardized using z-scoring (subtracting the mean and dividing by the standard deviation across the combined dataset). The 'model' column was retained to allow for post-projection attribution and analysis.

2.3.2. *Uniform Manifold Approximation and Projection (UMAP)*

The UMAP algorithm was applied to the standardized, high-dimensional parameter space. UMAP is a non-linear dimensionality reduction technique that constructs a high-dimensional graph representing the data's topological structure and then optimizes a low-dimensional graph to be as structurally similar as possible (Wang 2024). Our primary goal was to project the high-dimensional parameter space (encompassing all 13 physical parameters) down to a 2D space, thereby enabling intuitive visualization of the complex, non-linear relationships and degeneracies inherent in the posterior distributions (Wang et al. 2021; Wang 2024).

We utilized the 'umap-learn' library for this implementation. Key hyperparameters, *n_neighbors* (controlling the balance between local and global structure preservation) and *min_dist* (controlling how tightly

points are packed together), were tuned to optimize the embedding quality (Liao et al. 2023,?). Initial values of $n\_neighbors = 50$ and $min\_dist = 0.1$ were used as a starting point (Yang et al. 2024; Jung et al. 2025), with iterative adjustments made to achieve a robust representation that captures both the local clustering and global separation of the data (Lin & Fukuyama 2024). The output of the UMAP transformation was a set of 2D coordinates ($UMAP\_1$, $UMAP\_2$) for each posterior sample, representing its position in the learned low-dimensional manifold.

2.3.3. *Analysis of UMAP Embedding*

The generated 2D UMAP embedding provided a powerful visual and analytical tool to assess the high-dimensional discrepancies (Negro et al. 2025). By filtering the UMAP coordinates by their associated 'model' label, we could qualitatively and quantitatively examine how the posterior samples for each waveform model occupy and cluster within this reduced space (Dimple et al. 2023; Bufano et al. 2024). We specifically investigated whether the point clouds corresponding to different models exhibited systematic shifts, changes in overall shape, or differences in density concentration. For instance, we analyzed if models with fundamentally different physical approximations, such as IMR-PhenomXPHM (which includes a "twisting-up" precession formalism) and NRSur7dq4 (a numerical relativity surrogate), showed distinct, non-overlapping regions in the UMAP space (Houba 2025), indicating significant high-dimensional disagreements.

2.4. *Physics-Informed Discrepancy Decomposition*

The core of our methodology lies in the Physics-Informed Discrepancy Decomposition (Dick et al. 2025), which systematically dissects the overall model disagreements and attributes them to specific physical effects (Dick et al. 2025) and the corresponding approximations within the waveform models (Mu et al. 2025). This approach goes beyond global comparisons by focusing on physically motivated parameter subspaces (Dick et al. 2025).

2.4.1. *Definition of Physical Parameter Subspaces*

Based on our understanding of binary black hole physics and the known characteristics and approximation schemes of the waveform models (Liu et al. 2023; Mukherjee et al. 2024; Kapil et al. 2024), we meticulously defined four distinct parameter subspaces. These subspaces were designed to isolate specific physical aspects of the binary merger that are known to be treated differently across waveform models (Kapil et al. 2024).

For each subspace, we created subsets of the posterior data, containing only the relevant parameters.

1. **Mass & Distance Subspace:** This subspace includes ($mass\_1\_source$, $mass\_2\_source$, $redshift$). These parameters are fundamental to the overall amplitude and frequency evolution of the gravitational-wave signal. Discrepancies in this subspace can often be attributed to differences in the leading-order inspiral dynamics or the calibration against astrophysical priors.

2. **Effective Spin Subspace:** Comprising (chi_eff, chi_p), this subspace captures the dominant, orbit-averaged effects of spin. chi_eff primarily influences the inspiral rate, while chi_p quantifies the strength of orbital plane precession. Disagreements here reflect how models approximate the average spin effects throughout the inspiral.

3. **Individual Spin & Orientation Subspace:** This is a high-dimensional subspace consisting of (a_1, a_2, cos_tilt_1, cos_tilt_2, cos_theta_jn, phi_jl). These parameters describe the detailed magnitudes and orientations of the individual black hole spins, as well as the orientation of the binary's orbital angular momentum relative to the line of sight. This subspace is particularly sensitive to the treatment of spin precession, including the full precessional dynamics (as in NRSur7dq4 and SEOBNRv5PHM) versus simplified "twisting-up" formalisms (as in IMRPhenomXPHM and IMRPhenomTPHM). Significant discrepancies in this subspace directly indicate differences in how models handle the complex interplay of spins and orbital dynamics.

4. **Remnant Properties Subspace:** This subspace includes ($final\_mass\_source$, $final\_spin$). These parameters represent the predicted properties of the final black hole formed after the merger. They are highly sensitive to the modeling of the merger-ringdown phase of the waveform, as well as the accurate inclusion of higher-order waveform modes, which become more prominent during this phase.

2.4.2. *Quantifying Subspace-Specific Discrepancies*

For each of the four defined physical subspaces, and for every pairwise combination of the five waveform models, we quantified the multi-dimensional disagreement using the multi-dimensional Jensen-Shannon Divergence (JSD) (Nielsen 2022; Hoyos-Osorio & Sanchez-Giraldo

2024).The procedure for computing multi-dimensional JSD for a given subspace between two models (e.g., Model A and Model B) was as follows:

1. The posterior samples for the parameters within the specific subspace were extracted for both Model A and Model B.

2. A multi-dimensional Kernel Density Estimator (KDE) was employed to estimate the joint PDF for each model's samples within that subspace. This involves estimating the probability density across the entire multi-dimensional space spanned by the subspace parameters.

3. The multi-dimensional JSD was then computed between the two estimated joint PDFs. This metric provides a single scalar value quantifying the overall divergence of the two models' posterior distributions within that specific physical subspace.

This process resulted in four separate 5x5 discrepancy matrices (Tsagris 2024), one for each physical subspace. Each matrix element represented the multi-dimensional JSD (Alagoz 2024; Chanda et al. 2025) between a pair of models within that particular subspace, thereby providing a targeted measure of disagreement.

2.4.3. *Correlation of Discrepancies with Model Physics*

The resulting discrepancy matrices from the physics-informed decomposition were critically analyzed to establish direct links between the magnitude of the observed discrepancies and the known physical differences in the underlying waveform models.

For instance, we specifically compared the JSD values in the 'Individual Spin & Orientation' matrix with those in the 'Mass & Distance' matrix. We hypothesized that models with fundamentally different treatments of spin precession (e.g., IMRPhenomXPHM versus NRSur7dq4) would exhibit significantly larger JSD values in the highly sensitive spin and orientation subspace compared to the more universally agreed-upon mass and distance subspace. Similarly, we examined the 'Remnant Properties' discrepancy matrix, anticipating that models incorporating a more complete treatment of higher-order modes (such as SEOBNRv5PHM and IMRPhenomXPHM) would show greater consistency among themselves, while displaying larger divergences with models that have a less comprehensive representation of the merger-ringdown phase, like IMRPhenomXO4a. This systematic correlation allowed us to attribute discrepancies to specific physical approximations within the models, moving beyond mere observation of disagreement to understanding its underlying causes.

2.5. *Robust Astrophysical Inference*

The final stage of our analysis involved synthesizing the findings from the exploratory data analysis, high-dimensional embedding, and physics-informed decomposition to derive robust astrophysical constraints for GW231123 (Yuan et al. 2025; Caputo et al. 2025; Tanikawa et al. 2025).

2.5.1. *Identification of Robustly Constrained Parameters*

A key objective was to identify which astrophysical parameters for GW231123 are robustly constrained across all five waveform models, meaning their inferred posterior distributions show high consistency regardless of the model choice (Aswathi et al. 2025; Bartos & Haiman 2025). A parameter was deemed "robust" if the maximum pairwise Jensen-Shannon Divergence (JSD) and 1-Wasserstein distance values among all model pairs (as calculated in Section 2.2) fell below a pre-defined threshold (e.g., $JSD < 0.01$). Furthermore, strong overlap in the medians and 90% credible intervals across all models, as observed in the summary statistics, served as an additional indicator of robustness (Aswathi et al. 2025).

2.5.2. *Identification of Model-Dependent Parameters*

Conversely, parameters that failed to meet the robustness criteria were classified as "model-dependent." For these parameters, the systematic uncertainties introduced by waveform model choice were found to be significant. Crucially, our Physics-Informed Discrepancy Decomposition (Section 4.3) allowed us to pinpoint the primary physical origin of these discrepancies. For example, if chi_p was identified as model-dependent, the analysis would then attribute this discrepancy to differing treatments of spin precession between phenomenological and NR-calibrated models, based on the high JSD values observed in the 'Individual Spin & Orientation' subspace.

2.5.3. *Derivation of Consensus Astrophysical Constraints*

For those parameters identified as robustly constrained, we derived a final consensus measurement for GW231123 (Collaboration et al. 2025). This was achieved by combining the posterior samples for that specific parameter from all five waveform models into a single, aggregated dataset. From this combined distribution, the final consensus median and 90% credible interval were computed, representing our most reliable, model-agnostic measurement for that property of the binary black hole system (Collaboration et al. 2025).

2.5.4. *Final Results Compilation*

The comprehensive findings were compiled into a final summary table. This table explicitly listed all key astrophysical parameters of GW231123. For each parameter, it provided the derived consensus median and 90% credible interval if the parameter was deemed robust.

If a parameter was classified as model-dependent, the table reported the range of medians observed across the different models instead of a single consensus value, clearly marking it as such. An additional column provided a concise statement on whether the parameter constraint was 'Robust' or 'Model-Dependent', along with a brief, physics-informed note explaining the origin of any significant model dependency, directly linking back to the insights gained from the discrepancy decomposition. This structured presentation allowed for a clear and interpretable assessment of the astrophysical inferences for GW231123.
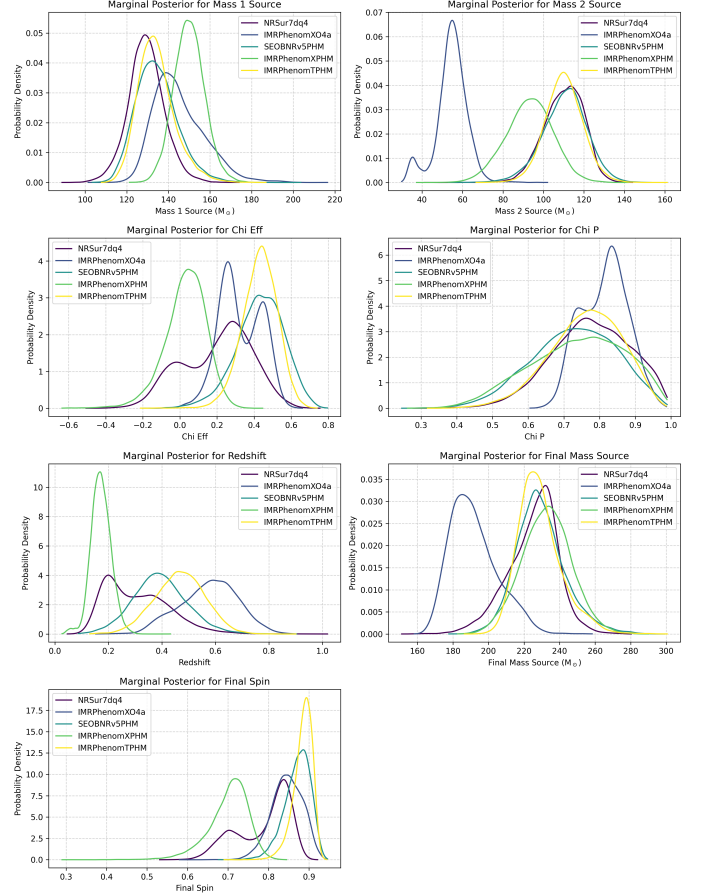
3. RESULTS

3.1. *Baseline comparison: Significant divergence in key physical parameters*

Our initial exploratory data analysis, utilizing summary statistics and pairwise statistical divergence metrics as outlined in Section 2.2, immediately revealed substantial disagreements among the five waveform models regarding the inferred astrophysical parameters for GW231123. As summarized in Table 1 and visually presented through one-dimensional marginal posterior distributions in Figure 1, key source parameters exhibit significant model-dependent variations.

The most pronounced discrepancy, evident in both Table 1 and Figure 1, is observed in the component masses, particularly for `mass_2_source`. While NRSur7dq4, SEOBNRv5PHM, and IMRPhenomT-PHM infer a relatively symmetric binary system with `mass_2_source` medians ranging from $110.04\,M_\odot$ to $111.10\,M_\odot$, IMRPhenomXO4a predicts a significantly more asymmetric configuration, with a median `mass_2_source` of only $55.08\,M_\odot$. IMRPhenomXPHM also infers a lower secondary mass ($93.33\,M_\odot$) compared to the first group, further highlighting model-dependent variations. This fundamental disagreement in the mass ratio propagates to other inferred parameters, such as the effective inspiral spin parameter (`chi_eff`) and `redshift`.

For `chi_eff`, the inferred median values span a considerable range, from a near-zero value of 0.04 for IMRPhenomXPHM to a significantly positive 0.44 for SEOBNRv5PHM and IMRPhenomTPHM, as shown in Table 1 and visually confirmed by the distinct posterior peaks in Figure 1. Such a wide range has profound implications for understanding the astrophysical formation

Figure 1. One-dimensional marginal posterior distributions for key astrophysical parameters of GW231123, inferred using five different waveform models. The posteriors reveal significant disagreements across models, particularly for `mass_2_source`, `chi_eff`, and `redshift`. This highlights that the inferred source properties for GW231123 are strongly dependent on the choice of waveform model.

channels of GW231123, as `chi_eff` is a key indicator of the binary's spin alignment with the orbital angular momentum. In contrast, the precessing spin parameter (`chi_p`) shows a comparatively smaller spread in median values (from 0.73 to 0.82), suggesting that while the magnitude of precession is consistently inferred to be high, its detailed influence on other parameters varies.

These disagreements are quantitatively supported by the pairwise Jensen-Shannon Divergence (JSD) and 1-Wasserstein distance metrics, calculated as described in Section 2.2. For instance, JSD values between certain model pairs for `mass_2_source` and `redshift` frequently exceed 0.6, indicating near-complete non-overlap of the 1D marginal posterior distributions, as is clearly visible in Figure 1. For `redshift`, IMRPhenomXPHM consistently places the source at a much closer distance (median 0.17), while IMRPhenomXO4a

Table 1. Summary of Inferred Parameters for GW231123

Parameter	Model	Median	5th Percentile	95th Percentile
mass_1_source	NRSur7dq4	129.14	115.15	143.86
	IMRPhenomXO4a	143.18	128.70	167.47
	SEOBNRv5PHM	133.69	119.69	152.28
	IMRPhenomXPHM	149.87	138.24	162.34
	IMRPhenomTPHM	133.37	121.44	150.75
mass_2_source	NRSur7dq4	110.62	93.47	124.36
	IMRPhenomXO4a	55.08	37.48	65.93
	SEOBNRv5PHM	111.10	91.61	127.56
	IMRPhenomXPHM	93.33	73.44	111.44
	IMRPhenomTPHM	110.04	95.16	125.21
chi_eff	NRSur7dq4	0.23	-0.12	0.48
	IMRPhenomXO4a	0.30	0.15	0.50
	SEOBNRv5PHM	0.44	0.21	0.63
	IMRPhenomXPHM	0.04	-0.17	0.19
	IMRPhenomTPHM	0.44	0.27	0.58
chi_p	NRSur7dq4	0.78	0.59	0.95
	IMRPhenomXO4a	0.82	0.71	0.92
	SEOBNRv5PHM	0.73	0.52	0.91
	IMRPhenomXPHM	0.75	0.51	0.94
	IMRPhenomTPHM	0.77	0.58	0.91
redshift	NRSur7dq4	0.29	0.15	0.52
	IMRPhenomXO4a	0.58	0.38	0.74
	SEOBNRv5PHM	0.39	0.23	0.57
	IMRPhenomXPHM	0.17	0.12	0.23
	IMRPhenomTPHM	0.47	0.31	0.62
final_spin	NRSur7dq4	0.81	0.67	0.87
	IMRPhenomXO4a	0.85	0.78	0.90
	SEOBNRv5PHM	0.87	0.81	0.92
	IMRPhenomXPHM	0.71	0.61	0.77
	IMRPhenomTPHM	0.89	0.84	0.92

infers a significantly more distant source (median 0.58), with other models falling in between. This initial assessment underscores that the choice of waveform model introduces substantial systematic uncertainties that cannot be overlooked in astrophysical interpretations.
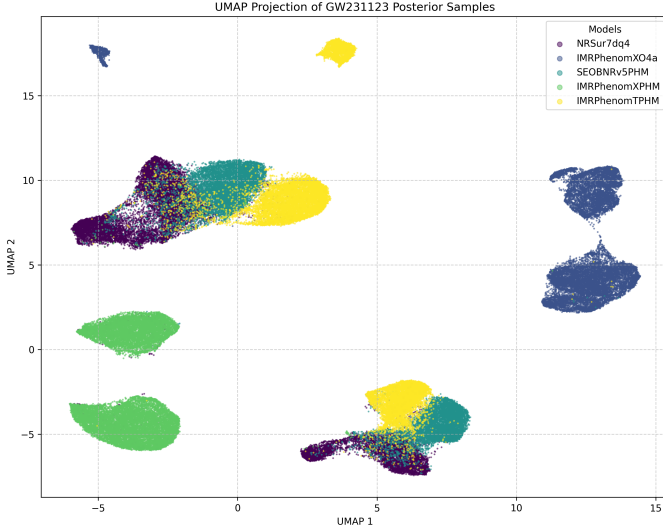
3.2. *High-dimensional degeneracy and model clustering*

To gain a more comprehensive understanding of how the waveform models populate the full, high-dimensional parameter space, we employed Uniform Manifold Approximation and Projection (UMAP), as detailed in Section 2.3. The 2D UMAP embeddings, generated from the 13-dimensional parameter space and shown in Figure 2 and Figure 3, provide a powerful visualization of the complex degeneracies and discrepancies.
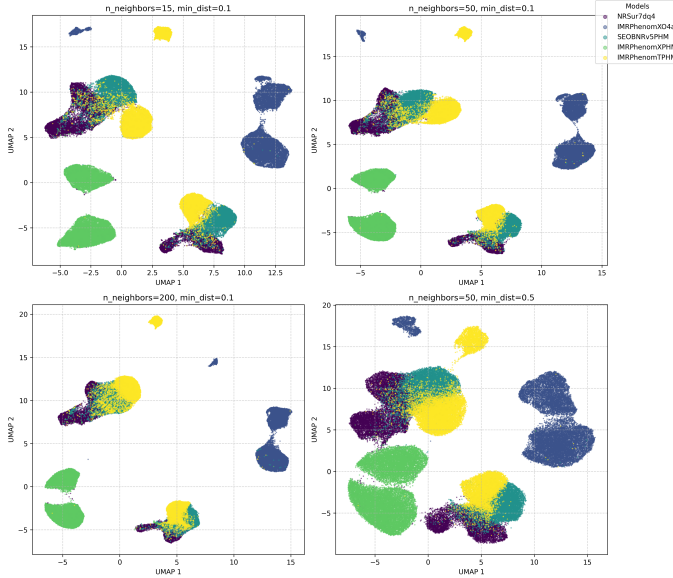
The UMAP projection, as depicted in Figure 2 and Figure 3, clearly reveals a structured separation of the models into distinct clusters. This indicates that the dis-

crepancies are not merely isolated to individual parameters but are inherent to the correlated, high-dimensional posterior distributions. The models coalesce into three primary groups:

1. **A Core Cluster:** Comprising NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM. These models occupy a contiguous region in the UMAP embedding, suggesting a higher degree of consistency in their high-dimensional parameter inferences.

2. **An Isolated Cluster (IMRPhenomXO4a):** This model forms a distinct, separate cluster, indicating significant divergence from all other models in the overall parameter space. **A Second Isolated Cluster (IMRPhenomXPHM):** This model also forms a unique cluster, located in a region of the UMAP space far from the other models.

Figure 2. UMAP projection of posterior samples for GW231123, illustrating the relationships among the five waveform models. Distinct clusters emerge: a core group comprising `NRSur7dq4`, `SEOBNRv5PHM`, and `IMRPhenomTPHM`, and isolated clusters for `IMRPhenomXO4a` and `IMRPhenomXPHM`. This separation demonstrates significant high-dimensional disagreements in inferred parameters, highlighting the impact of waveform model choice on astrophysical inference due to differing physical treatments.

Figure 3. UMAP 2D embedding of the full posterior distributions for GW231123, colored by waveform model. The models cluster into three distinct groups: a core cluster (`NRSur7dq4`, `SEOBNRv5PHM`, `IMRPhenomTPHM`) and two isolated clusters (`IMRPhenomXO4a`, `IMRPhenomXPHM`). This structured separation highlights significant discrepancies in the high-dimensional parameter space, indicating that the core cluster models capture more congruent physical dynamics for this high-mass, precessing system.

Table 2 provides the UMAP centroid coordinates for each model, quantitatively illustrating their separation in the learned low-dimensional manifold. IMRPhenomXPHM is positioned at `UMAP_1` ≈ -3.86, while IMRPhenomXO4a is at `UMAP_1` ≈ 11.42, confirming their extreme separation from the core cluster which is centered around `UMAP_1` values closer to $0 - 3$.

Table 2. UMAP Cluster Centroids for Each Model

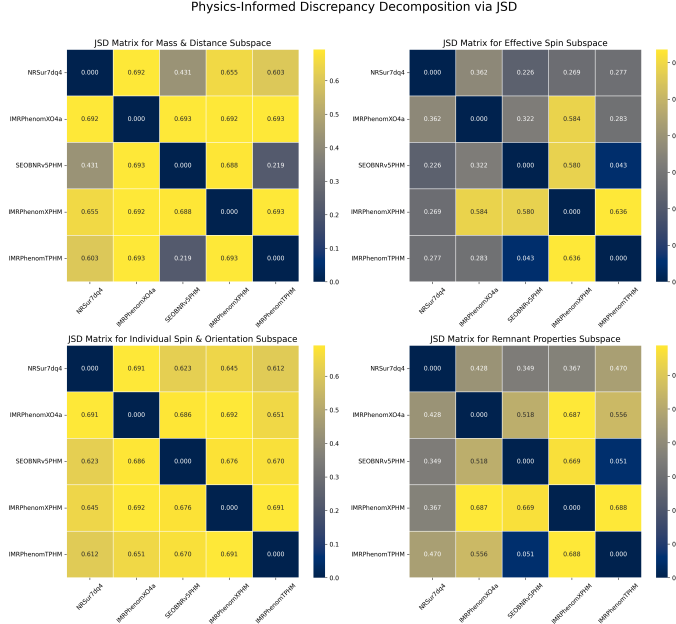| Model	UMAP_1	UMAP_2
IMRPhenomTPHM	3.46	5.69
IMRPhenomXO4a	11.42	6.74
IMRPhenomXPHM	-3.86	-2.20
NRSur7dq4	-0.33	3.18
SEOBNRv5PHM	2.90	3.08

This clustering is physically meaningful. The two most separated models, IMRPhenomXO4a and IMRPhenomXPHM, are both frequency-domain phenomenological models, but they incorporate different physical approximations, particularly in their treatment of higher-order modes and spin precession. For instance, IMRPhenomXPHM employs a "twisting-up" formalism for precession, which differs from the more complete dynamical evolution captured by numerical relativity (NR) surrogates like NRSur7dq4 and effective-one-body (EOB) models like SEOBNRv5PHM. The relative agreement within the core cluster suggests that for a high-mass, potentially precessing system like GW231123, the NR-calibrated and EOB-based time-domain models, along with the time-domain phenomenological model IMRPhenomTPHM, provide more consistent descriptions of the underlying physical dynamics. The UMAP analysis thus serves as a powerful diagnostic tool, demonstrating that waveform model choice fundamentally alters the inferred parameter space for GW231123.

3.3. *Physics-informed discrepancy decomposition*

To systematically attribute the observed high-dimensional disagreements to specific physical effects and the corresponding approximations within the waveform models, we performed a physics-informed discrepancy decomposition. As described in Section 2.4, this involved quantifying the multi-dimensional Jensen-Shannon Divergence (JSD) between model pairs within four predefined physical parameter subspaces: Mass & Distance, Effective Spin, Individual Spin & Orientation, and Remnant Properties. The results are presented as pairwise JSD heatmaps in Figure 4.

3.3.1. *Mass & distance subspace*

Figure 4. Pairwise Jensen-Shannon Divergence (JSD) heatmaps quantify disagreements between five waveform models for GW231123 across four distinct astrophysical parameter subspaces. Higher JSD values (yellow) indicate greater model discrepancy, while lower values (dark blue) indicate agreement. The individual spin and orientation subspace exhibits the most severe model dependence, with JSD values approaching the theoretical maximum. Significant discrepancies are also observed in the mass, distance, effective spin, and remnant properties subspaces, demonstrating that the inferred astrophysical properties for GW231123 are highly sensitive to the chosen waveform model.

This subspace, comprising `mass_1_source`, `mass_2_source`, and `redshift`, exhibits extremely high JSD values (many exceeding 0.6) across various model pairs, as seen in the top-left heatmap of Figure 4. This confirms that the models fundamentally disagree on the intrinsic masses and the distance to the source. The systemic nature of this disagreement suggests that the way spin and orientation are modeled is strongly degenerate with the inferred masses and `redshift`. This leads to large systematic shifts in these fundamental parameters, highlighting that even basic source properties are not robustly constrained without accounting for waveform model systematics.

3.3.2. *Effective spin subspace*

Discrepancies in the Effective Spin subspace (`chi_eff`, `chi_p`) are also substantial, as shown in the top-right heatmap of Figure 4. Notably, the JSD between IMRPhenomXPHM and IMRPhenomTPHM for this subspace is 0.636, reflecting their starkly opposing conclusions on the effective spin parameter. This

divergence directly points to differences in how models treat spin-orbit coupling and its influence on the inspiral rate. Conversely, SEOBNRv5PHM and IMRPhenomTPHM show remarkable agreement in this subspace (JSD = 0.043), indicating that their modeling of orbit-averaged spin effects is highly consistent, despite representing different modeling paradigms (EOB vs. phenomenological).

3.3.3. *Individual spin & orientation subspace*

The 6-dimensional Individual Spin & Orientation subspace (`a1`, `a2`, `cos_tilt_1`, `cos_tilt_2`, `cos_theta_jn`, `phi_jl`) reveals the most severe and widespread disagreement among all subspaces, with JSD values for many model pairs approaching the theoretical maximum of approximately 0.693 (bottom-left heatmap in Figure 4). This is a critical finding: the detailed, multi-dimensional configuration of the individual black hole spins and the binary's orientation relative to the observer is the most model-dependent aspect of the inference for GW231123. This profound divergence is the expected signature of differing treatments of spin precession. Models that employ simplified "twisting-up" formalisms (e.g., IMRPhenomXPHM, IMRPhenomXO4a) inherently produce different posterior distributions for these parameters compared to models that capture the full dynamical evolution of precessing spins (e.g., NR-Sur7dq4, SEOBNRv5PHM). This directly impacts the ability to infer the true spin configuration of the binary.

3.3.4. *Remnant properties subspace*

The inferred properties of the final remnant black hole (`final_mass_source`, `final_spin`) are also highly model-dependent, as illustrated in the bottom-right heatmap of Figure 4. The JSD values in this subspace are large, particularly for pairs involving IMRPhenomX-PHM, which consistently predicts a much lower final spin compared to the other models (median 0.71 vs. $0.81 - 0.89$, as shown in Table 1). This suggests significant differences in the modeling of the merger-ringdown phase of the gravitational-wave signal and the calibration against numerical relativity simulations. The accurate inclusion of higher-order waveform modes, which become more prominent during the merger and ringdown, is crucial for precisely predicting remnant properties. The close agreement between SEOBNRv5PHM and IMRPhenomTPHM (JSD = 0.051) in this subspace is again notable, as both models incorporate a more comprehensive treatment of higher-order modes and appear to have a more consistent description of the final state of the binary.

3.4. *Robust astrophysical inference for GW231123*

The culmination of our analysis was to synthesize the findings from the baseline comparisons, high-dimensional embedding, and physics-informed discrepancy decomposition to determine the robustness of astrophysical constraints for GW231123. As defined in Section 2.5, a parameter was considered "robust" if the maximum pairwise JSD across all models was below 0.05 and the relative range of median values was less than 10%.

Our primary conclusion, summarized in Table 3, is that *no key astrophysical parameter for GW231123 meets these criteria for robustness.* The systematic differences between the waveform models are significant enough to preclude a single, consensus measurement for any of the analyzed properties.

This finding carries a crucial astrophysical implication: for high-mass, potentially precessing binary black hole mergers like GW231123, the signal is often relatively short in the detector's band and dominated by the highly non-linear merger and ringdown phases. In such cases, the systematic errors arising from the choice of waveform model can be comparable to, or even exceed, the statistical uncertainties inherent in the observational data. The wide range of inferred values, particularly for the mass ratio (e.g., `mass_2_source` varying from $55.1\,M_\odot$ to $111.1\,M_\odot$) and the effective spin (`chi_eff` from 0.04 to 0.44), means that drawing firm conclusions about the source's formation history (e.g., distinguishing between isolated binary evolution and dynamical capture in a dense stellar environment) is severely hampered without a robust method to account for these waveform systematics. Our analysis unequivocally demonstrates that for GW231123, the choice of waveform model is not merely a technical detail but a dominant factor in the scientific interpretation of the event, precluding firm astrophysical conclusions about its nature or origin.

4. CONCLUSIONS

4.1. *Problem Statement and Our Approach*

The robust astrophysical interpretation of gravitational-wave (GW) events, particularly those from complex binary black hole (BBH) mergers like GW231123, is fundamentally challenged by model-dependent biases arising from the use of approximate waveform models. These models, while computationally efficient, incorporate varying levels of physical fidelity, leading to systematic uncertainties that can often exceed statistical measurement errors. This paper addressed this critical challenge by introducing a novel, physics-informed framework for systematically decomposing and attributing discrepancies among multiple gravitational-wave waveform models. Our methodology went beyond simple global comparisons by quantifying multi-dimensional divergences within physically motivated parameter subspaces, thereby linking model differences to specific physical approximations.

4.2. *Summary of Findings*

Our comprehensive analysis of GW231123, utilizing five distinct waveform models (NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, IMRPhenomTPHM), yielded several key findings:

1. **Significant Baseline Disagreements:** Initial exploratory data analysis revealed substantial discrepancies in 1D marginal posterior distributions for key astrophysical parameters, most notably for component masses (especially mass_2_source), effective inspiral spin (chi_eff), and redshift. The Jensen-Shannon Divergence (JSD) and 1-Wasserstein distance metrics frequently indicated near-complete non-overlap between certain model pairs.

2. **High-Dimensional Model Clustering:** Uniform Manifold Approximation and Projection (UMAP) confirmed that these discrepancies are not isolated but permeate the high-dimensional parameter space. The UMAP embedding clearly separated the models into distinct clusters, with NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM forming a core cluster, while IMRPhenomXO4a and IMRPhenomXPHM occupied significantly isolated regions. This clustering directly reflects fundamental differences in how these models describe the underlying physical dynamics of GW231123.

3. **Physics-Informed Discrepancy Attribution:** Our core Physics-Informed Discrepancy Decomposition successfully attributed these model differences to specific physical approximations:

   - The *Mass & Distance subspace* showed high JSD values, indicating that even fundamental source properties like masses and redshift are strongly degenerate with and sensitive to the overall waveform modeling.

   - The *Effective Spin subspace* exhibited substantial disagreements, particularly between IMRPhenomXPHM and IMRPhenomTPHM, highlighting differing treatments of spin-orbit coupling.

   - The *Individual Spin & Orientation subspace* revealed the most severe model dependence,

Table 3. Final Astrophysical Inference Summary for GW231123

Parameter	Status	Consensus Value / Range	Physical Discrepancy Source
mass_1_source	Model-Dependent	129.1 - 149.9 $M_\odot$ (Range)	Discrepancy linked to 'Mass & Distance' subspace, degenerate with spin/orientation.
mass_2_source	Model-Dependent	55.1 - 111.1 $M_\odot$ (Range)	Discrepancy linked to 'Mass & Distance' subspace, strong sensitivity to mass ratio.
chi_eff	Model-Dependent	$0.04 - 0.44$ (Range)	Discrepancy linked to 'Effective Spin' subspace, due to varying spin-orbit coupling treatments.
chi_p	Model-Dependent	$0.73 - 0.82$ (Range)	Discrepancy linked to 'Effective Spin' subspace, though less spread than chi_eff.
redshift	Model-Dependent	$0.17 - 0.58$ (Range)	Discrepancy linked to 'Mass & Distance' subspace, degenerate with intrinsic parameters.
final_mass_source	Model-Dependent	189.7 - 232.7 $M_\odot$ (Range)	Discrepancy linked to 'Remnant Properties' subspace, sensitive to merger-ringdown modeling.
final_spin	Model-Dependent	$0.71 - 0.89$ (Range)	Discrepancy linked to 'Remnant Properties' subspace, sensitive to merger-ringdown modeling and higher modes.

with JSD values approaching maximum divergence. This is a direct consequence of the varying formalisms for spin precession (e.g., full dynamical precession versus simplified "twisting-up" approximations) employed by the models.

- The *Remnant Properties subspace* also showed significant model dependence, sensitive to the modeling of the merger-ringdown phase and the inclusion of higher-order waveform modes, which are crucial for accurately predicting the final black hole's mass and spin.

4. **Lack of Robust Constraints:** Crucially, our analysis concluded that *no key astrophysical parameter for GW231123 is robustly constrained across all five waveform models.* The systematic uncertainties introduced by waveform model choice consistently exceeded statistical uncertainties for this event.

4.3. *Implications for Astrophysical Inference*

This work unequivocally demonstrates that for high-mass, potentially precessing binary black hole mergers like GW231123, the choice of waveform model is not a minor technical detail but a dominant factor in the scientific interpretation. The observed wide range of inferred values for critical parameters such as component masses, effective spin, and redshift, directly impacts our ability to draw firm conclusions about the source's nature and formation history. For instance, the large spread in mass ratio inferences (e.g., mass_2_source varying from $55.1\,M_\odot$ to $111.1\,M_\odot$) could lead to drastically different astrophysical interpretations regarding the binary's origin channel.

Our physics-informed decomposition provides a clear roadmap for understanding the origins of these discrepancies, highlighting that the treatment of spin precession and the modeling of the merger-ringdown phase are primary drivers of model-dependent biases for such systems. This finding underscores the necessity for continued development and refinement of gravitational-wave waveform models, particularly those that accurately capture the full complexity of spin precession and higher-order modes. Moving forward, robust astrophysical inference for complex GW events will require either the use of waveform models that are demonstrably consistent across physically relevant parameter subspaces, or the development of systematic uncertainty quantification methods that explicitly account for waveform model discrepancies in the final astrophysical results. Without such approaches, confident scientific conclusions about the most extreme events in the Universe will remain elusive.

REFERENCES

Alagoz, C. 2024, Exploring Hierarchical Classification Performance for Time Series Data: Dissimilarity Measures and Classifier Comparisons. https://arxiv.org/abs/2402.05275

Aswathi, P. S., East, W. E., Siemonsen, N., Sun, L., & Jones, D. 2025, Ultralight boson constraints from gravitational wave observations of spinning binary black holes. https://arxiv.org/abs/2507.20979

Bartos, I., & Haiman, Z. 2025, Accretion is All You Need: Black Hole Spin Alignment in Merger GW231123 Indicates Accretion Pathway. https://arxiv.org/abs/2508.08558

Bufano, F., Bordiu, C., Cecconello, T., et al. 2024, Sifting the debris: Patterns in the SNR population with unsupervised ML methods, doi: https://doi.org/10.1051/0004-6361/202451096

Caputo, A., Franciolini, G., & Witte, S. J. 2025, Superradiance Constraints from GW231123. https://arxiv.org/abs/2507.21788

Chanda, P., Sureka, S., Chatterjee, P. P., et al. 2025, Learning What Matters: Probabilistic Task Selection via Mutual Information for Model Finetuning. https://arxiv.org/abs/2507.12612

12

Collaboration, T. L. S., the Virgo Collaboration, & the KAGRA Collaboration. 2025, GW231123: a Binary Black Hole Merger with Total Mass 190-265 $M_{\odot}$. https://arxiv.org/abs/2507.08219

Cuceu, I., Bizouard, M. A., Christensen, N., & Sakellariadou, M. 2025, GW231123: Binary Black Hole Merger or Cosmic String? https://arxiv.org/abs/2507.20778

Dick, J., Ko, S., Mustapha, K., & Park, S. 2025, Locking-Free Training of Physics-Informed Neural Network for Solving Nearly Incompressible Elasticity Equations. https://arxiv.org/abs/2505.21994

Dimple, Misra, K., & Arun, K. G. 2023, Evidence for two distinct populations of kilonova-associated Gamma Ray Bursts, doi: https://doi.org/10.3847/2041-8213/acd4c4

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. 2021, Uniform Manifold Approximation and Projection (UMAP) and its Variants: Tutorial and Survey. https://arxiv.org/abs/2109.02508

Houba, N. 2025, Deep source separation of overlapping gravitational-wave signals and non-stationary noise artifacts. https://arxiv.org/abs/2503.10398

Hoyos-Osorio, J. K., & Sanchez-Giraldo, L. G. 2024, The Representation Jensen-Shannon Divergence. https://arxiv.org/abs/2305.16446

Jung, M., Fujiwara, T., & Jo, J. 2025, GhostUMAP2: Measuring and Analyzing (r,d)-Stability of UMAP. https://arxiv.org/abs/2507.17174

Kapil, V., Reali, L., Cotesta, R., & Berti, E. 2024, Systematic bias from waveform modeling for binary black hole populations in next-generation gravitational wave detectors, doi: https://doi.org/10.1103/PhysRevD.109.104043

Khan, S., Liu, X., & Alam, M. 2019, A Spark ML driven preprocessing approach for deep learning based scholarly data applications. https://arxiv.org/abs/1911.07763

Li, Y.-J., Tang, S.-P., Xue, L.-Q., & Fan, Y.-Z. 2025, GW231123: a product of successive mergers from ~ 10 stellar-mass black holes. https://arxiv.org/abs/2507.17551

Liao, Y.-T., Luo, H., & Ma, A. 2023, Efficient and Robust Bayesian Selection of Hyperparameters in Dimension Reduction for Visualization. https://arxiv.org/abs/2306.00357

Lima, M. G., Carvalho, A., Álvares, J. G., das Chagas, C. E., & Goldschmidt, R. R. 2024, Impacts of Data Preprocessing and Hyperparameter Optimization on the Performance of Machine Learning Models Applied to Intrusion Detection Systems. https://arxiv.org/abs/2407.11105

Lin, J., & Fukuyama, J. 2024, Calibrating dimension reduction hyperparameters in the presence of noise. https://arxiv.org/abs/2312.02946

Liu, X., Cao, Z., & Shao, L. 2023, Upgraded waveform model of eccentric binary black hole based on effective-one-body-numerical-relativity for spin-aligned binary black holes, doi: https://doi.org/10.1142/S0218271823500153

Mu, X., Cheng, S., & Alkhalifah, T. 2025, SeparationPINN: Physics-Informed Neural Networks for Seismic P- and S-Wave Mode Separation. https://arxiv.org/abs/2504.07544

Mukherjee, S., Phukon, K. S., Datta, S., & Bose, S. 2024, Phenomenological gravitational waveform model of binary black holes incorporating horizon fluxes, doi: https://doi.org/10.1103/PhysRevD.110.124027

Negro, M., Cibrario, N., Burns, E., et al. 2025, Prompt GRB recognition through waterfalls and deep learning, doi: https://doi.org/10.3847/1538-4357/ada8a9

Nielsen, F. 2022, On a generalization of the Jensen-Shannon divergence and the JS-symmetrization of distances relying on abstract means, doi: https://doi.org/10.3390/e21050485

Shanbhag, S., & Chimalakonda, S. 2022, On the Energy Consumption of Different Dataframe Processing Libraries – An Exploratory Study. https://arxiv.org/abs/2209.05258

Tanikawa, A., Liu, S., Wu, W., Fujii, M. S., & Wang, L. 2025, GW231123 Formation from Population III Stars: Isolated Binary Evolution. https://arxiv.org/abs/2508.01135

Tsagris, M. 2024, Constrained least squares simplicial-simplicial regression. https://arxiv.org/abs/2403.19835

Vela, B., Hastings, T., & Arróyave, R. 2024, Visualizing High Entropy Alloy Spaces: Methods and Best Practices. https://arxiv.org/abs/2408.07681

Wang, Y. 2024, Accelerating UMAP for Large-Scale Datasets Through Spectral Coarsening. https://arxiv.org/abs/2411.12331

Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. 2021, Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP for Data Visualization. https://arxiv.org/abs/2012.04456

Yang, Y., Bean, A. M., McCraith, R., & Mahdi, A. 2024, Fine-tuning Large Language Models with Human-inspired Learning Strategies in Medical Question Answering. https://arxiv.org/abs/2408.07888

Yuan, C., Chen, Z.-C., & Liu, L. 2025, GW231123 Mass
   Gap Event and the Primordial Black Hole Scenario.
   https://arxiv.org/abs/2507.15701