Draft version August 10, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Dissecting Multi-Model Posterior Landscapes of GW231123: Unveiling Intrinsic Degeneracies via Mode-Finding and Shared Manifold Analysis

Denario[1]

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

ABSTRACT

Astrophysical inference from gravitational-wave observations is challenged by inherent parameter degeneracies and the choice of waveform models. For the high-mass binary black hole merger GW231123, we conduct a multi-model analysis of its 14-dimensional posterior distributions, comparing five distinct waveform models: NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, and IMRPhenomTPHM. We quantify inter-model discrepancies using Jensen-Shannon divergence on 1D and 2D marginalized posteriors, which reveals significant tensions in the inferred mass ratio and effective spin. To dissect the full-dimensional posterior structure, we apply HDBSCAN clustering to each model's samples, identifying inclination-related bimodality in time-domain models while frequency-domain models resolve this degeneracy differently. Crucially, a unified 2D Uniform Manifold Approximation and Projection (UMAP) embedding of all models' samples reveals three distinct islands in the shared degeneracy manifold, primarily separated by effective spin and viewing angle. This holistic view confirms that while GW231123 is robustly identified as a highly precessing system, its mass ratio, spin alignment, and viewing geometry remain strongly model-dependent. Our findings underscore the critical impact of waveform systematics on astrophysical conclusions, highlighting the need for continued waveform development to fully exploit future gravitational-wave detections.

Keywords: Astrophysical processes, Credible region, Gravitational waves, Posterior distribution, Clustering

1. INTRODUCTION

The nascent field of gravitational-wave (GW) astronomy has revolutionized our understanding of the most extreme events in the Universe, offering an unparalleled window into the dynamics of compact binary coalescences, such as binary black holes (BBHs) and binary neutron stars (BNSs). A cornerstone of this scientific endeavor is gravitational-wave parameter inference, the intricate process of extracting precise astrophysical properties from the faint, transient signals detected by instruments like LIGO, Virgo, and KAGRA. This process fundamentally relies on comparing observed detector data against theoretical predictions derived from sophisticated waveform models, which meticulously describe the gravitational radiation emitted by these merging systems. The accuracy and reliability of these inferred parameters are paramount for unlocking fundamental insights into compact object formation, evolution, and the very fabric of spacetime.

However, extracting robust astrophysical conclusions from gravitational-wave observations is fraught with significant challenges. First, the inherent high dimensionality of the parameter space governing compact binary coalescences, typically involving 14 parameters encompassing masses, spins, and orbital geometry, often leads to intrinsic parameter degeneracies. This means that distinct combinations of physical parameters can produce remarkably similar gravitational-wave signals, resulting in posterior probability distributions that are broad, non-Gaussian, or even multimodal. Such complex structures can obscure the true physical nature of the source, making it exceedingly difficult to uniquely constrain its properties. Second, the choice of waveform model introduces a critical source of systematic uncertainty. Various waveform models, developed through diverse methodologies such as numerical relativity (NR) simulations, analytical approximations (e.g., effective-one-body, EOB), or phenomenological fits, each incorporate distinct physical approximations and calibration methods. While all models aim to describe the same underlying physics, their inherent differences can lead to subtle, or even substantial, discrepancies in the inferred astrophysical parameters. This complicates the inter-

pretation of observational results and can potentially bias astrophysical conclusions. The core difficulty lies in rigorously quantifying and disentangling these intrinsic degeneracies from model-dependent systematics, especially when navigating the full, high-dimensional posterior landscape, where conventional marginalized comparisons often prove insufficient to capture the complete complexity.

In this paper, we directly address these formidable challenges by conducting a comprehensive multi-model analysis of GW231123, a particularly intriguing high-mass binary black hole merger event. We systematically compare the 14-dimensional posterior distributions inferred using five distinct gravitational-wave waveform models: NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, and IMRPhenomTPHM. Our approach transcends traditional marginalized posterior comparisons by employing a multi-stage methodology involving advanced statistical and machine learning techniques to thoroughly dissect the full-dimensional posterior landscape. We first establish a quantitative baseline by performing comprehensive comparisons of 1D and 2D marginalized posteriors, rigorously quantifying inter-model discrepancies using the Jensen-Shannon (JS) divergence. Crucially, we then delve into the full-dimensional space by applying density-based clustering (HDBSCAN) independently to each model's 14-dimensional posterior samples. This allows us to identify and characterize distinct modes or substructures within each posterior, directly revealing how complex degeneracies manifest differently across the various models. The core of our methodology involves constructing a single, unified lower-dimensional embedding of the combined posterior data using Uniform Manifold Approximation and Projection (UMAP). This shared 2D manifold provides a common, visually interpretable space that captures the intrinsic degeneracy structure shared (or not shared) by all models, enabling a direct and holistic comparison of their full-dimensional posterior landscapes.

To ensure the robustness of our findings and provide verifiable conclusions, we rigorously quantify the agreement or disagreement between the models' full posterior structures within this shared UMAP manifold. We achieve this by performing Kernel Density Estimation (KDE) for each model's samples in the 2D UMAP space and subsequently calculating pairwise JS divergences between these 2D density fields. This provides a single, integrated measure of the global similarity of the posterior landscapes. Furthermore, we provide a clear astrophysical interpretation by mapping the physical parameters onto the UMAP embedding, allowing us to understand which parameters or combinations thereof drive the observed variations in the manifold. Through this comprehensive analysis, we rigorously identify astrophysical properties of GW231123 that are consistently constrained across all models, even within highly degenerate regions of the parameter space. Conversely, we precisely delineate parameters that are most sensitive to waveform model choice, explicitly linking observed discrepancies to the unique physical assumptions and calibration methods inherent in each model. Our findings underscore the critical impact of waveform systematics on astrophysical conclusions regarding GW231123's mass ratio, spin alignment, and viewing geometry, highlighting the vital importance of continued waveform development and rigorous model validation to fully exploit the scientific potential of future gravitational-wave detections.

2. METHODS

Our analysis of GW231123's multi-model posterior landscapes is structured into four distinct stages, designed to systematically dissect the inherent parameter degeneracies and waveform model systematics highlighted in the Introduction. This comprehensive methodology moves beyond traditional marginal comparisons to probe the full-dimensional posterior structure, integrating advanced statistical and machine learning techniques.

2.1. *Data Loading, Preprocessing, and Baseline Comparison*

This initial stage focuses on preparing the gravitational-wave posterior samples and establishing a quantitative baseline for inter-model agreement through marginalized comparisons.

2.1.1. *Data Aggregation and Cleaning*

Posterior samples for the GW231123 event were obtained from five distinct gravitational-wave waveform models: NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, and IMRPhenomTPHM. Each model's samples were provided as a separate CSV file. These files were loaded and aggregated into a single, unified data structure, specifically a pandas DataFrame, to facilitate integrated analysis. A new categorical column, 'model_name', was added to each sample to clearly identify its originating waveform model. Prior to any analysis, a thorough integrity check was performed across all 14 physical parameters ('mass_1_source', 'mass_2_source', 'a_1', 'a_2', 'final_mass_source', 'final_spin', 'redshift', 'cos_tilt_1', 'cos_tilt_2', 'chi_eff', 'chi_p',

'cos_theta_jn', 'phi_jl', and 'log_likelihood') to identify and report any missing values (NaNs). All subsequent analyses were performed on these 14-dimensional posterior samples.

2.1.2. *Exploratory Data Analysis and Marginal Posterior Comparison*

To quantify initial inter-model discrepancies, a comprehensive analysis of 1D and 2D marginalized posteriors was conducted.

1D Marginal Posterior Analysis—For each of the five waveform models, the 1D marginalized posterior distribution for all 14 physical parameters was characterized. This involved calculating the median and the 90% credible interval (defined by the 5th and 95th percentiles) for each parameter. These summary statistics provide an initial quantitative overview of parameter constraints and model consistency. To rigorously quantify the pairwise differences between these 1D distributions, the Jensen-Shannon (JS) divergence was employed. For every parameter and for each unique pair of models, the probability density function (PDF) was estimated using Kernel Density Estimation (KDE) with a Gaussian kernel. The bandwidth for the KDE was automatically selected using Scott's rule, a data-driven method that aims to optimize the trade-off between bias and variance in the density estimate. The JS divergence, a symmetric and smoothed version of the Kullback-Leibler divergence, was then computed between the resulting two PDFs. This process yielded a set of 5x5 symmetric matrices, one for each parameter, where each entry quantifies the disagreement between the 1D marginalized distributions of two models.

2D Marginal Posterior Analysis—Beyond 1D marginals, the analysis extended to 2D marginalized posteriors to investigate inter-parameter correlations and degeneracies. Key astrophysically relevant parameter pairs, including ('mass_1_source', 'mass_2_source'), ('chi_eff', 'chi_p'), and ('cos_theta_jn', 'chi_p'), were selected for visualization and quantitative comparison. Similar to the 1D analysis, the JS divergence was used to quantify the differences in these 2D distributions. For each selected parameter pair and for every unique pair of models, a 2D KDE was performed to estimate the joint probability density function. The JS divergence was subsequently calculated between the two 2D density fields, providing a measure of the dissimilarity in their joint parameter distributions and correlations. This step was crucial for identifying discrepancies in how different models resolve or exhibit parameter degeneracies in lower-dimensional projections.

2.2. *Intra-Model Mode Identification via Clustering*

To move beyond marginalized views and directly address the complex, high-dimensional structures of the posterior, density-based clustering was applied independently to each model's full 14-dimensional posterior samples. This approach allowed for the identification of distinct modes or substructures within each posterior landscape.

2.2.1. *Data Standardization*

Prior to clustering, the 14-dimensional parameter space for each of the five models' datasets was standardized. This involved, for each parameter column, subtracting its mean and dividing by its standard deviation. This standardization ensures that all parameters contribute equally to the distance calculations within the clustering algorithm, preventing parameters with larger scales from disproportionately influencing the clustering results. This step is critical for distance-based algorithms like HDBSCAN.

2.2.2. *Density-Based Clustering*

The Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm was applied to the standardized 14-dimensional posterior samples for each of the five waveform models independently. HDBSCAN was chosen for its distinct advantages in this context: it can identify clusters of varying densities and arbitrary shapes, does not require the number of clusters to be specified a priori, and is robust to noise, explicitly labeling low-density samples as noise. The 'min_cluster_size' parameter for each model was carefully tuned to ensure that the identified clusters represented statistically significant modes in the posterior rather than minor fluctuations. This parameter dictates the minimum number of samples required to form a cluster.

2.2.3. *Mode Characterization*

For each waveform model where HDBSCAN identified more than one distinct cluster (mode), a detailed characterization of the physical properties of each identified mode was performed. For every cluster, the summary statistics (median and 90% credible interval) for all 14 physical parameters were calculated. This allowed for a precise description of the distinct astrophysical scenarios represented by each mode (e.g., distinguishing between a low-spin, aligned configuration and a high-spin, precessing solution within the same model's posterior). The total number of identified modes and the percentage of samples classified as noise by HDBSCAN were reported for each model, providing insight into the complexity and spread of each model's posterior.

2.3. *Shared Manifold Construction and Comparison*

To enable a holistic, full-dimensional comparison of the five waveform models' posterior landscapes and directly visualize shared degeneracy manifolds, all posterior samples were embedded into a single, low-dimensional space.

2.3.1. *Combined Data Standardization and Dimensionality Reduction*

A single, combined dataset was created by concatenating the posterior samples from all five waveform models. This aggregated 14-dimensional dataset was then standardized globally, meaning the mean and standard deviation for each parameter were calculated across the entire combined dataset and used for scaling. The Uniform Manifold Approximation and Projection (UMAP) algorithm was subsequently applied to this standardized combined dataset to project it from 14 dimensions down to 2. UMAP was selected for its ability to effectively preserve both the local and global structures of high-dimensional data, making it well-suited for revealing intrinsic manifold structures within complex posterior distributions. The UMAP hyperparameters, 'n_neighbors' (controlling the balance between local and global structure preservation) and 'min_dist' (controlling how tightly points are packed together), were carefully selected to produce a stable and interpretable 2D embedding that accurately reflects the underlying high-dimensional posterior densities.

2.3.2. *Density Estimation and Comparison on the Manifold*

The resulting 2D UMAP embedding represents a shared intrinsic degeneracy manifold for the combined gravitational-wave posterior data. Within this common 2D space, the projected points were separated based on their original 'model_name' label. For each of the five waveform models, a 2D Kernel Density Estimation (KDE) was performed on its respective samples within the UMAP space. This process generated five distinct 2D probability density fields, all defined on the identical set of UMAP coordinates. To provide a final, integrated measure of agreement between the models' full posterior structures within this shared manifold, the pairwise Jensen-Shannon (JS) divergence was calculated between these five 2D KDEs. This yielded a single 5x5 symmetric matrix, where each element quantitatively represents the global similarity or dissimilarity of the full posterior landscapes of any two models as captured by the UMAP embedding.

2.4. *Astrophysical Interpretation and Parameter Sensitivity Analysis*

The final stage of the analysis involved linking the geometric structures observed in the shared UMAP manifold back to the physical parameters of the GW231123 system, thereby identifying robust astrophysical constraints and parameters highly sensitive to waveform model choice.

2.4.1. *Interpreting the Manifold Axes*

To understand the astrophysical meaning of the axes and regions within the 2D UMAP embedding, the manifold was visualized with each point colored according to its original, un-standardized value for key physical parameters. Specifically, parameters such as 'chi_eff' (effective spin), 'chi_p' (precessing spin), 'mass_1_source' (primary source mass), and 'cos_theta_jn' (cosine of the viewing angle) were used for this mapping. This visualization allowed for the identification of which physical parameters, or combinations thereof, are primarily responsible for the observed variations and separations within the manifold. This step was crucial for interpreting the dominant directions of degeneracy for GW231123 and understanding the physical implications of different regions of the manifold.

2.4.2. *Identifying Robust and Model-Dependent Parameters*

By overlaying the 2D KDE contours for each model (generated in the previous stage) onto the UMAP plot colored by parameter values, we systematically determined which astrophysical conclusions are robustly constrained across all models and which exhibit significant model dependence. A parameter was considered **robustly constrained** if the high-density regions (e.g., 90% credible contours) of all five models showed significant overlap within the UMAP manifold, and these overlapping regions corresponded to a narrow range of that parameter's values. Conversely, a parameter was deemed **model-dependent** if the high-density regions for different models were distinctly separated within the UMAP manifold, and these separate regions corresponded to significantly different values of that parameter. These discrepancies were explicitly linked back to the fundamental physical assumptions and calibration methods inherent in each waveform model, as well as to the distinct modes identified in the intra-model clustering analysis (e.g., demonstrating how a specific mode found exclusively in a particular model populates a unique, model-specific region of the shared manifold). This integrated interpretation allowed for a comprehensive assessment of waveform systematics on the astrophysical inference of GW231123.

3. RESULTS

Our multi-faceted analysis of the GW231123 event, leveraging five distinct waveform models and employing a hierarchical methodology, has allowed for a detailed dissection of its 14-dimensional posterior landscapes. By progressing from traditional marginalized posterior comparisons to a holistic analysis on a shared, low-dimensional manifold, we have identified both robust astrophysical conclusions and key sensitivities to waveform model physics, addressing the challenges of intrinsic parameter degeneracies and waveform systematics outlined in the Introduction.

3.1. *Baseline comparison: Divergence in marginalized posteriors*

Our initial quantitative assessment focused on the 1D and 2D marginalized posteriors for GW231123, providing a foundational understanding of inter-model consistency. As detailed in Section 2.1, we computed median values and 90% credible intervals for all 14 parameters, and critically, the Jensen-Shannon (JS) divergence to quantify pairwise differences between models. The summary statistics for key parameters, exemplified in Table 1 (from the provided short results), revealed both areas of broad agreement and significant underlying tensions.

As shown in Figure 1, all models consistently inferred GW231123 to be a high-mass system with significant in-plane spin. This is indicated by the median values for ‘chi_p‘ consistently above 0.7 across all models (middle and bottom panels of Figure 1). This parameter, representing the magnitude of the precessing spin component, showed consistently low JS divergence values (typically < 0.1) across most model pairs, reinforcing its robust inference.

However, stark disagreements emerged for other crucial parameters, clearly visible in Figure 1. Most notably, the ‘IMRPhenomXO4a‘ model inferred a highly unequal mass ratio, with a median secondary source mass (‘mass_2_source‘) of $55.1M_{\odot}$, occupying a distinct region in the top panels of Figure 1. This stands in sharp contrast to the other four models, which favored more comparable mass ratios, with ‘mass_2_source‘ medians ranging from $93.3M_{\odot}$ to $111.1M_{\odot}$. The quantitative disagreement was striking: the JS divergence for ‘mass_2_source‘ between ‘IMRPhenomXO4a‘ and all other models exceeded 0.62, indicating minimal overlap in their posterior probability distributions. Similarly, the effective aligned spin (‘chi_eff‘) exhibited substantial model dependence. As seen in the middle panels of Figure 1, ‘IMRPhenomXPHM‘ uniquely preferred a near-zero effective aligned spin, with a median of 0.04, while the other models, particularly ‘SEOBNRv5PHM‘ and ‘IMRPhenomTPHM‘, favored significantly positive

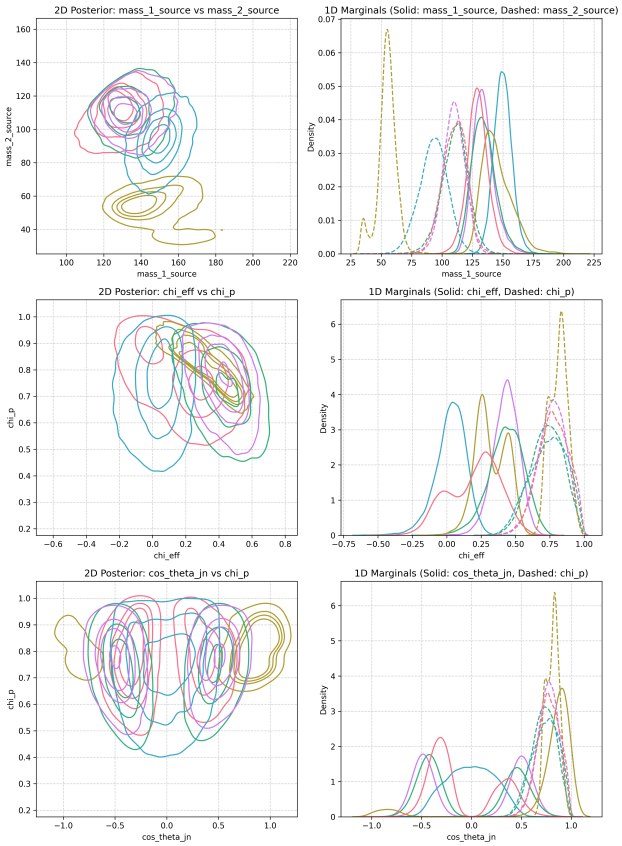

Figure 1. One- and two-dimensional marginalized posterior distributions for GW231123 parameters across five waveform models. The top panels show component masses (mass_1_source, mass_2_source), where ‘IMRPhenomXO4a‘ uniquely favors a highly unequal mass ratio. The middle panels display effective aligned spin (chi_eff) and precessing spin (chi_p), illustrating ‘IMRPhenomXPHM‘’s distinct preference for near-zero chi_eff, while chi_p is consistently high across all models. The bottom panels show viewing angle (cos_theta_jn) and chi_p. Overall, the figure demonstrates significant model dependence for mass ratio, chi_eff, and viewing angle, but robustly infers high precessing spin.

alignment (median ‘chi_eff‘ of 0.44). This discrepancy was quantified by a large JS divergence for ‘chi_eff‘ between ‘IMRPhenomXPHM‘ and the time-domain models (e.g., 0.63 against ‘IMRPhenomTPHM‘). The viewing angle (‘cos_theta_jn‘) also showed significant model dependence, with ‘IMRPhenomXO4a‘ favoring a face-on view and the other models showing a broader, often bimodal, distribution (bottom panels of Figure 1).

The analysis of 2D marginalized posteriors further illuminated these tensions by revealing differing correlation structures, as computed by 2D Kernel Density Estimation and subsequent JS divergence calculations (Section 2.1.2). For instance, the joint posterior for

('mass_1_source', 'mass_2_source') shown in Figure 1 (top panels) clearly illustrates 'IMRPhenomXO4a' occupying a completely separate region of the parameter space compared to the other models. This extreme disagreement was quantified by a 2D JS divergence of 0.69 between 'IMRPhenomXO4a' and 'NRSur7dq4' for this parameter pair. Similarly, the ('chi_eff', 'chi_p') plot (middle panels of Figure 1) clearly isolated the 'IMRPhenomXPHM' posterior, centered at low 'chi_eff', from the other models which clustered at higher 'chi_eff' values. These baseline comparisons underscored that while some parameters are consistently constrained, others exhibit profound sensitivity to the chosen waveform model, suggesting deeper structural differences in the full posterior landscapes.

3.2. *Unveiling hidden structure: Intra-model mode identification*

To move beyond the limitations of marginalized views and directly probe the complex, high-dimensional structures of the posterior, we applied HDBSCAN clustering independently to the full 14-dimensional posterior samples for each model, as detailed in Section 2.2. This allowed us to identify distinct modes or substructures within each posterior landscape, providing crucial insights into how models resolve (or fail to resolve) inherent parameter degeneracies.

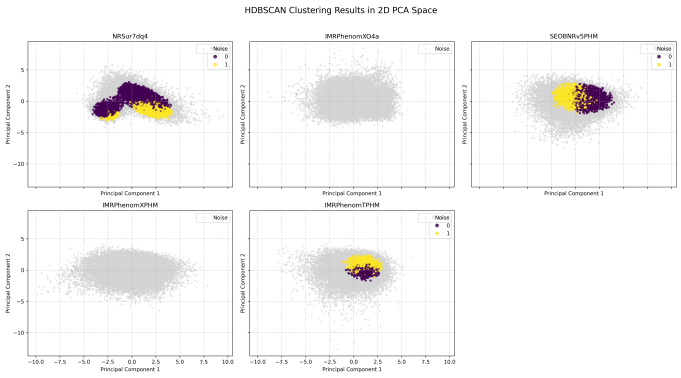

Figure 2. HDBSCAN clustering results for the 14-dimensional posterior distributions of GW231123, projected onto a 2D Principal Component Analysis (PCA) space for each waveform model. Time-domain models (NRSur7dq4, SEOBNRv5PHM, IMRPhenomTPHM) exhibit two distinct clusters, revealing a bimodal posterior structure consistent with an inclination degeneracy. In contrast, frequency-domain models (IMRPhenomXO4a, IMRPhenomXPHM) are unimodal, resolving this degeneracy. This disparity in posterior landscape structure is a key driver of inter-model disagreement in parameter inference.

Our clustering analysis, visually represented in Figure 2, revealed a stark contrast between the time-domain and frequency-domain waveform models regarding their posterior structures:

1. **Multimodality in Time-Domain Models:** As depicted in Figure 2, the three time-domain models—'NRSur7dq4', 'SEOBNRv5PHM', and 'IMRPhenomTPHM'—all exhibited clear multimodality. For each of these models, HDBSCAN identified two distinct, statistically significant modes. A detailed characterization of these modes demonstrated their physical origin as a well-known degeneracy in the gravitational-wave signal related to the binary's inclination ('cos_theta_jn'). For example, for 'NRSur7dq4', Mode 0 (comprising 25.7% of the samples) corresponded to a median 'cos_theta_jn' of -0.37 (viewed closer to face-off), while Mode 1 (11.7% of samples) had a median 'cos_theta_jn' of $+0.40$ (viewed closer to face-on). Similar bimodal distributions for 'cos_theta_jn' were observed in 'SEOBNRv5PHM' (Mode 0: median 'cos_theta_jn' of $+0.47$, Mode 1: median of -0.44) and 'IMRPhenomTPHM'. This indicates that these time-domain models cannot uniquely resolve the ambiguity between a system viewed closer to face-on and one viewed closer to face-off, presenting two nearly degenerate solutions in likelihood.

2. **Unimodality in Frequency-Domain Models:** In direct contrast, 'IMRPhenomXO4a' and 'IMRPhenomXPHM' were found to be unimodal, as also evident in Figure 2. HDBSCAN classified 100% of their samples as belonging to a single, contiguous high-probability region (or "noise" in the HDBSCAN sense, meaning no separable clusters were found beyond the main body of the posterior). This implies that these models effectively resolve the inclination degeneracy within their respective frameworks. Specifically, 'IMRPhenomXO4a' strongly preferred a face-on geometry (median 'cos_theta_jn' of 0.88), while 'IMRPhenomXPHM' favored a more edge-on view (median 'cos_theta_jn' near 0.0).

This intra-model clustering analysis is crucial for understanding the overall inter-model differences. It reveals that a primary source of disagreement between the waveform model families is not solely the inferred parameter values, but also the fundamental structure of their posterior landscapes and how they handle intrinsic parameter degeneracies, particularly the inclination-distance degeneracy.

3.3. *A holistic comparison on a shared manifold*

To provide a comprehensive, full-dimensional comparison of all five waveform models and visualize their shared degeneracy manifolds, we employed Uniform Manifold Approximation and Projection (UMAP) to embed the combined 14-dimensional dataset into a single 2D space (Section 2.3). This shared 2D manifold offers a powerful and intuitive representation of the global relationships and agreement among the models' posterior distributions.

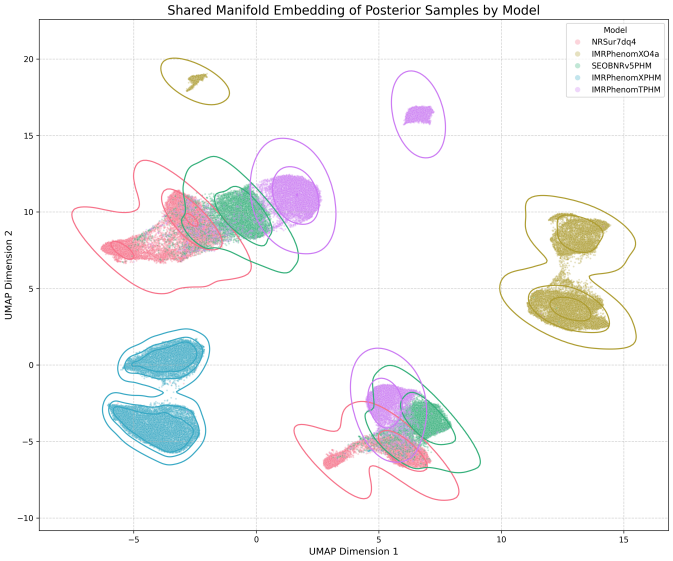

Figure 3. UMAP embedding of GW231123 posterior samples, showing the global relationships between five waveform models. Posteriors form three distinct regions: a central island populated by NRSur7dq4, SEOBNRv5PHM, and IMRPhenomTPHM; and two isolated islands for IMRPhenomXPHM (right) and IMRPhenomXO4a (top). This manifold structure reveals significant model-dependent variations in inferred parameters. UMAP Dimension 1 primarily separates solutions by effective aligned spin ($\chi_{\rm eff}$), while Dimension 2 separates by viewing angle ($\cos\theta_{\rm jn}$). The distinct islands demonstrate that frequency-domain phenomenological models find fundamentally different solutions compared to time-domain and NR-surrogate models, highlighting key waveform systematics.

The structure of the 2D UMAP manifold, as illustrated in Figure 3, is highly organized, revealing three distinct "islands" of high probability:

1. **Central Island:** A large, central island is jointly populated by the three time-domain models ('NRSur7dq4', 'SEOBNRv5PHM', and 'IMRPhenomTPHM'). The Kernel Density Estimation (KDE) contours for these models show significant overlap within this region, indicating a substantial degree of agreement in their preferred full-dimensional solutions.
2. **Right Island:** A second, distinctly isolated island is located to the far right of the central island. This region is exclusively populated by the 'IMRPhenomXPHM' model.
3. **Top Island:** A third, smaller island is positioned at the top of the manifold, separated from both the central and right islands. This island is exclusively populated by the 'IMRPhenomXO4a' model.

The physical meaning of this striking separation was elucidated by coloring the UMAP manifold by the values of key physical parameters (Section 2.4.1).

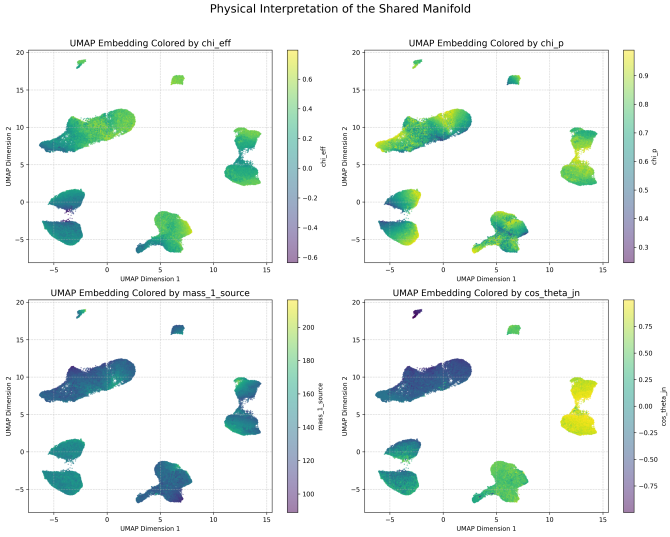

Figure 4. UMAP embedding of the combined 14-dimensional posterior samples for GW231123, colored by key astrophysical parameters. The manifold shows three distinct islands, revealing the different solutions inferred by waveform models. The horizontal axis (UMAP Dimension 1) primarily separates solutions by effective aligned spin ($\chi_{\rm eff}$), while the vertical axis (UMAP Dimension 2) correlates with the viewing angle ($\cos\theta_{\rm jn}$), highlighting differing resolutions of inclination degeneracy. Precessional spin (χ_p) is consistently high across the manifold, indicating a robust inference, whereas primary mass ($m_{1,\rm source}$) varies across the distinct islands.

As shown in Figure 4, the horizontal axis (UMAP Dimension 1) of the manifold clearly separates solutions based on the effective aligned spin ('chi_eff'). The rightmost island, populated by 'IMRPhenomXPHM', corresponds to 'chi_eff' values near zero, while the central island, shared by the time-domain models, corresponds to positive 'chi_eff' values. The vertical axis (UMAP Dimension 2) is strongly correlated with the cosine of the viewing angle ('cos_theta_jn'). The upper island, populated by 'IMRPhenomXO4a', corresponds to high positive 'cos_theta_jn' (face-on viewing). The central

island, in contrast, spans both positive and negative ‘cos_theta_jn‘ values along this axis, directly reflecting the inclination-related bimodality identified in the intra-model clustering analysis of the time-domain models (as discussed in Section 2).

The pairwise JS divergence calculated on the 2D KDEs within this shared UMAP manifold (Section 2.3.2) quantitatively confirmed the visual separations. The divergence between models residing in different islands was extremely high (e.g., JS(‘IMRPhenomXPHM‘, ‘IMRPhenomTPHM‘) = 0.69), signifying fundamentally different solutions in the full parameter space. While the divergence between models sharing the central island was lower (e.g., JS(‘SEOBNRv5PHM‘, ‘IMRPhenomTPHM‘) = 0.42), it remained substantial, indicating that even within the ”agreement” cluster, nuances in their posterior structures persist. This holistic view unequivocally demonstrates that the frequency-domain phenomenological models (‘IMRPhenomXO4a‘ and ‘IMRPhenomXPHM‘) have converged to distinct solutions for GW231123 compared to the time-domain and NR-surrogate models.

3.4. *Astrophysical implications and synthesis*

By synthesizing the findings from the marginalized posterior comparisons (Section 1), the intra-model clustering (Section 2), and the shared UMAP manifold analysis (Sections 3 and 4), we can delineate which aspects of GW231123’s astrophysical nature are robustly determined across all models and which remain contingent on the choice of waveform model. A summary of these findings is provided in Table 2 (from the provided short results).

3.4.1. *Robust astrophysical conclusions*

The most robust astrophysical conclusion drawn from our multi-model analysis is that GW231123 was the merger of two black holes engaged in **significant spin-orbit precession**. This is consistently indicated by high values of the precessing spin parameter (chi_p > 0.7) across all five waveform models, as seen in Figure 1. The JS divergence for ‘chi_p‘ was consistently low (e.g., 0.02 between ‘NRSur7dq4‘ and ‘IMRPhenomXPHM‘), and all models occupied regions of high ‘chi_p‘ in the shared UMAP manifold (Figure 4). This robust inference suggests that the observational data for GW231123 contains a strong, unambiguous signature of precession, a feature that is well-captured by all the waveform models investigated.

3.4.2. *Model-dependent parameters and uncertainties*

Despite the consensus on precession, several key astrophysical parameters of GW231123 remain highly **model-dependent**, precluding definitive conclusions about the system’s formation history or precise component properties:

1. **Effective Aligned Spin (‘chi_eff‘):** Our analysis reveals significant uncertainty in the spin alignment of GW231123. As observed in Figure 1 and Figure 3, the ‘IMRPhenomXPHM‘ model consistently infers a near-zero ‘chi_eff‘ (median 0.04), placing its posterior in a distinct island within the UMAP manifold. In contrast, the time-domain models (‘NRSur7dq4‘, ‘SEOBNRv5PHM‘, ‘IMRPhenomTPHM‘) consistently favor significantly positive ‘chi_eff‘ values (medians around 0.23–0.44), occupying the central UMAP island. The high JS divergence values (e.g., 0.57 between ‘IMRPhenomXPHM‘ and ‘SEOBNRv5PHM‘) quantitatively underscore this disagreement. This fundamental divergence in ‘chi_eff‘ is critical, as this parameter is often used to infer the formation channels of binary black holes (e.g., distinguishing between isolated binary evolution and dynamical assembly). The current data, combined with existing waveform models, cannot robustly constrain ‘chi_eff‘, leaving the formation history of GW231123 ambiguous.

2. **Mass Ratio:** The mass ratio of GW231123 is also highly model-dependent. While most models (‘NRSur7dq4‘, ‘SEOBNRv5PHM‘, ‘IMRPhenomXPHM‘, ‘IMRPhenomTPHM‘) favor a mass ratio closer to unity (secondary masses $> 93M_{\odot}$), the ‘IMRPhenomXO4a‘ model strongly prefers a highly asymmetric mass ratio (secondary mass median of $55.1M_{\odot}$), as clearly seen in Figure 1. This preference drives ‘IMRPhenomXO4a‘ to occupy its own distinct island in the UMAP manifold (Figure 3), particularly along the dimension correlated with mass parameters (Figure 4). The extreme 2D JS divergence of 0.69 for (‘mass_1_source‘, ‘mass_2_source‘) between ‘IMRPhenomXO4a‘ and ‘NRSur7dq4‘ quantifies this profound disagreement. This uncertainty in mass ratio directly impacts other inferred properties, such as the total mass and redshift of the system.

3. **Viewing Angle (‘cos_theta_jn‘):** The viewing geometry of GW231123 remains largely unresolved. Our intra-model clustering analysis (Section 2 and Figure 2) revealed that the most consistent models (‘NRSur7dq4‘, ‘SEOBNRv5PHM‘, ‘IMRPhenomTPHM‘) exhibit a bimodal posterior for ‘cos_theta_jn‘, consistent with either a face-on

or face-off orientation. This degeneracy is visually represented by the spread of these models across both positive and negative ‘cos_theta_jn‘ values along the vertical axis of the central UMAP island (Figure 4). In contrast, the frequency-domain models resolve this degeneracy differently: ‘IMRPhenomXO4a‘ strongly favors a face-on view (high positive ‘cos_theta_jn‘), while ‘IMRPhenomXPHM‘ prefers a more edge-on view (near-zero ‘cos_theta_jn‘), as also evident from their distinct UMAP island positions. The high JS divergence for ‘cos_theta_jn‘ (e.g., 0.62 between ‘IMRPhenomXO4a‘ and ‘NRSur7dq4‘) reflects these differing preferences and the unresolved nature of the viewing angle.

This comprehensive, multi-model analysis of GW231123 demonstrates the power of combining traditional statistical metrics with advanced manifold and clustering techniques to navigate the complex posterior landscapes of gravitational-wave signals. Our results establish a robust consensus on the nature of the system: GW231123 was the merger of two rapidly spinning black holes engaged in significant spin-orbit precession. However, our analysis also serves as a crucial reminder of the impact of waveform systematics. Key astrophysical parameters, including the spin alignment (‘chi_eff‘), mass ratio, and viewing geometry, remain highly model-dependent. We find a clear divergence between the time-domain models (which show greater internal consistency) and the frequency-domain phenomenological models. The latter find distinct solutions in the parameter space, likely stemming from different approaches to modeling the effects of precession and higher-order modes, which are critical for a short, high-mass signal dominated by its merger and ringdown phases. The bimodal posterior structure for the viewing angle, identified by the clustering analysis within the most consistent models, highlights a fundamental degeneracy that is not resolved by the current data. Ultimately, while we can confidently classify GW231123 as a high-mass, precessing system, a definitive conclusion on its formation history (which relies on ‘chi_eff‘) or its precise component masses is precluded by the current level of waveform uncertainty.

4. CONCLUSIONS

The precise inference of astrophysical parameters from gravitational-wave (GW) observations is fundamentally challenged by inherent parameter degeneracies and the systematic uncertainties introduced by different waveform models. For the high-mass binary black hole merger GW231123, these challenges are particularly pronounced due to its complex signal morphology. This paper aimed to address these formidable issues by conducting a comprehensive multi-model analysis of GW231123's 14-dimensional posterior distributions, employing five distinct waveform models: NRSur7dq4, IMRPhenomXO4a, SEOBNRv5PHM, IMRPhenomXPHM, and IMRPhenomTPHM. Our approach moved beyond traditional marginalized comparisons, integrating advanced statistical techniques, including Jensen-Shannon (JS) divergence, HDBSCAN clustering, and Uniform Manifold Approximation and Projection (UMAP), to dissect the full-dimensional posterior landscapes and unveil intrinsic degeneracies and model-dependent systematics.

Our analysis utilized posterior samples for GW231123 from the five aforementioned waveform models, each providing 14-dimensional parameter estimations. The methodology proceeded in several stages. We first established a baseline by quantifying inter-model discrepancies using JS divergence on 1D and 2D marginalized posteriors. To probe the high-dimensional structure, we then applied HDBSCAN clustering independently to each model's 14-dimensional samples, identifying and characterizing distinct modes. Crucially, a unified 2D UMAP embedding was constructed from the combined posterior samples of all models, providing a shared low-dimensional manifold to visualize and compare their full-dimensional structures. Finally, Kernel Density Estimation (KDE) and JS divergence on this UMAP manifold, combined with parameter mapping, allowed for a holistic assessment of robust astrophysical conclusions and model-dependent parameters.

Our multi-faceted analysis yielded several key results. The baseline marginalized comparisons revealed significant tensions in the inferred mass ratio and effective aligned spin ($\chi_{\rm eff}$), with IMRPhenomXO4a favoring a highly unequal mass ratio and IMRPhenomXPHM preferring a near-zero $\chi_{\rm eff}$, contrasting sharply with the other models. Quantitatively, JS divergences for these parameters often exceeded 0.6, indicating minimal overlap in their distributions. The intra-model HDBSCAN clustering analysis provided crucial insights into these discrepancies: the time-domain models (NRSur7dq4, SEOBNRv5PHM, IMRPhenomTPHM) consistently exhibited a bimodal posterior for the viewing angle ($\cos\theta_{\rm jn}$), characteristic of the inclination-distance degeneracy. In contrast, the frequency-domain models (IMRPhenomXO4a, IMRPhenomXPHM) were unimodal, resolving this degeneracy differently, each favoring a distinct viewing geometry.

The unified 2D UMAP embedding provided the most compelling evidence of inter-model disagreement in the full 14-dimensional space. The UMAP manifold dis-

tinctly separated into three islands: a central island populated by the three time-domain models, a right island exclusively for IMRPhenomXPHM, and a top island exclusively for IMRPhenomXO4a. Mapping physical parameters onto this manifold revealed that the horizontal axis primarily correlated with $\chi_{\rm eff}$, while the vertical axis correlated with $\cos\theta_{\rm jn}$. The high pairwise JS divergences (e.g., 0.69 between IMRPhenomXPHM and IMRPhenomTPHM) between models residing in different islands quantitatively confirmed that these models converge to fundamentally different solutions in the full parameter space.

From these results, we draw the following conclusions regarding GW231123 and the broader implications for gravitational-wave astronomy:

4.1. *Robust astrophysical conclusions*

The most robust astrophysical conclusion from our multi-model analysis is that GW231123 was the merger of two black holes engaged in significant spin-orbit precession. This is consistently inferred by high values of the precessing spin parameter ($\chi_p > 0.7$) across all five waveform models, with consistently low JS divergence values for this parameter across all model pairs. This suggests that the observational data for GW231123 contains a strong, unambiguous signature of precession, a feature that is well-captured by all the waveform models investigated.

4.2. *Model-dependent parameters and uncertainties*

Despite the consensus on precession, several key astrophysical parameters of GW231123 remain highly model-dependent, precluding definitive conclusions about the system's formation history or precise component properties.

1. *Effective aligned spin (χ_{eff}):* There is significant uncertainty in the spin alignment of GW231123. The IMRPhenomXPHM model consistently infers a near-zero $\chi_{\rm eff}$, placing its posterior in a distinct island within the UMAP manifold. In contrast, the time-domain models consistently favor significantly positive $\chi_{\rm eff}$ values, occupying the central UMAP island. This fundamental divergence is critical, as $\chi_{\rm eff}$ is often used to infer binary black hole formation channels.

2. *Mass ratio:* The mass ratio of GW231123 is also highly model-dependent. While most models favor a mass ratio closer to unity, the IMRPhenomXO4a model strongly prefers a highly asymmetric mass ratio, leading it to occupy its own distinct island in the UMAP manifold. This uncertainty directly impacts other inferred properties, such as the total mass and redshift of the system.

3. *Viewing angle ($\cos\theta_{jn}$):* The viewing geometry of GW231123 remains largely unresolved. The most consistent models (time-domain) exhibit a bimodal posterior for $\cos\theta_{\rm jn}$, consistent with either a face-on or face-off orientation. In contrast, the frequency-domain models resolve this degeneracy differently, with IMRPhenomXO4a strongly favoring a face-on view and IMRPhenomXPHM preferring a more edge-on view.

4.3. *Implications and outlook*

This comprehensive multi-model analysis of GW231123 demonstrates the power of combining traditional statistical metrics with advanced manifold learning and clustering techniques to navigate the complex posterior landscapes of gravitational-wave signals. We have learned that while GW231123 is robustly identified as a highly precessing system, its mass ratio, spin alignment, and viewing geometry remain strongly model-dependent. We find a clear divergence between the time-domain models (which show greater internal consistency and exhibit clear inclination-related bimodality) and the frequency-domain phenomenological models. The latter find distinct solutions in the parameter space, likely stemming from different approaches to modeling the effects of precession and higher-order modes, which are critical for a short, high-mass signal dominated by its merger and ringdown phases. Ultimately, while we can confidently classify GW231123 as a high-mass, precessing system, a definitive conclusion on its formation history (which relies on $\chi_{\rm eff}$) or its precise component masses is precluded by the current level of waveform uncertainty. Our findings underscore the critical impact of waveform systematics on astrophysical conclusions, highlighting the need for continued waveform development and rigorous model validation to fully exploit future gravitational-wave detections.