The Astrophysical Journal, 978:153 (12pp), 2025 January 10 https://doi.org/10.3847/1538-4357/ad9253

No Need to Know: Toward Astrophysics-free Gravitational-wave Cosmology

Amanda M. Farah[1], Thomas A. Callister[2], Jose María Ezquiaga[3], Michael Zevin[4], and Daniel E. Holz[1,2,5,6]
[1] Department of Physics, The University of Chicago, Chicago, IL 60637, USA; afarah@uchicago.edu
[2] Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, USA
[3] Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
[4] The Adler Planetarium, 1300 South DuSable Lake Shore Drive, Chicago, IL 60605, USA
[5] Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA
[6] Enrico Fermi Institute, The University of Chicago, Chicago, IL 60637, USA
Received 2024 April 4; revised 2024 October 21; accepted 2024 October 23; published 2025 January 8

Abstract

Gravitational waves (GWs) from merging compact objects encode direct information about the luminosity distance to the binary. When paired with a redshift measurement, this enables standard-siren cosmology: a Hubble diagram can be constructed to directly probe the Universe's expansion. This can be done in the absence of electromagnetic measurements, as features in the mass distribution of GW sources provide self-calibrating redshift measurements without the need for a definite or probabilistic host galaxy association. This "spectral siren" technique has thus far only been applied with simple parametric representations of the mass distribution, and theoretical predictions for features in the mass distribution are commonly presumed to be fundamental to the measurement. However, the use of an inaccurate representation leads to biases in the cosmological inference, an acute problem given the current uncertainties in true source population. Furthermore, it is commonly presumed that the form of the mass distribution must be known a priori to obtain unbiased measurements of cosmological parameters in this fashion. Here, we demonstrate that spectral sirens can accurately infer cosmological parameters without such prior assumptions. We apply a flexible, nonparametric model for the mass distribution of compact binaries to a simulated catalog of 1000 GW signals, consistent with expectations for the next LIGO–Virgo–KAGRA observing run. We find that, despite our model's flexibility, both the source mass model and cosmological parameters are correctly reconstructed. We predict a 11.2% measurement of H_0, keeping all other cosmological parameters fixed, and a 6.4% measurement of $H(z=0.9)$ when fitting for multiple cosmological parameters (1σ uncertainties). This astrophysically agnostic spectral siren technique will be essential to arrive at precise and unbiased cosmological constraints from GW source populations.

Unified Astronomy Thesaurus concepts: Gravitational wave astronomy (675); Bayesian statistics (1900); Hubble constant (758); Compact binary stars (283)

1. Introduction

Like light, gravitational waves (GWs) are redshifted as they propagate across the Universe, thereby bearing imprints of the Universe's cosmic expansion history. Unlike light, however, the form of GW signals is known from first principles, directly from the theory of general relativity. Furthermore, because GWs propagate across the Universe without attenuation from intervening matter, and because the properties of GW detectors are well characterized, GW selection effects are extremely well understood. This allows for a precise estimate of each GW catalog's completeness and an unbiased measurement of the true GW source population (R. Abbott et al. 2023a, 2023b; R. Essick 2023). Additionally, the GW signals observed by the LIGO, Virgo, and KAGRA detectors (F. Acernese et al. 2015; J. Aasi et al. 2015; T. Akutsu et al. 2021) provide direct measurements of the distance to their sources. This makes them "standard sirens": direct probes of cosmological parameters that circumvent the need for a cosmological distance ladder (B. F. Schutz 1986; D. E. Holz & S. A. Hughes 2005).

A well-known demonstration of standard siren cosmology was the multimessenger event GW170817 (B. P. Abbott et al. 2017a; D. A. Coulter et al. 2017; N. R. Tanvir et al. 2017; S. Valenti et al. 2017), whose clear association with a host galaxy provided a precise redshift measurement and allowed for a direct "bright siren" measurement of the Hubble constant, H_0 (B. P. Abbott et al. 2017b). External redshift information can also come from galaxy catalogs, which provide an ensemble of possible redshifts for each GW signal, allowing for a probabilistic "dark siren" measurement of H_0 when multiple GW detections are combined (W. Del Pozzo 2012; H.-Y. Chen et al. 2018; M. Fishbach et al. 2019; M. Soares-Santos et al. 2019; R. Gray et al. 2020; B. P. Abbott et al. 2021a; R. Abbott et al. 2023c; R. Gray et al. 2022; J. R. Gair et al. 2023; R. Gray et al. 2023; S. Mastrogiovanni et al. 2023a).

Electromagnetic information about GW source redshifts need not be available in order to use them as standard sirens, however. GW signals provide direct measurements of each source's luminosity distance, D_L, and redshifted (detector-frame) masses, $m_{\rm det} = m_{\rm source}(1+z)$ (e.g., X. Chen et al. 2019). Therefore, if the source-frame mass is known, each GW signal provides a direct mapping between luminosity distance and redshift, allowing for a measurement of the expansion of the Universe at the time the GW signal was emitted, $H(z)$.

In practice, the source-frame masses of individual GW signals are not known (unless tidal information is available; e.g., C. Messenger & J. Read 2012; D. Chatterjee et al. 2021). It is, however, possible to consider the *population* of compact binaries at large and use known features in their source-frame mass distribution to obtain self-calibrated redshift estimates.

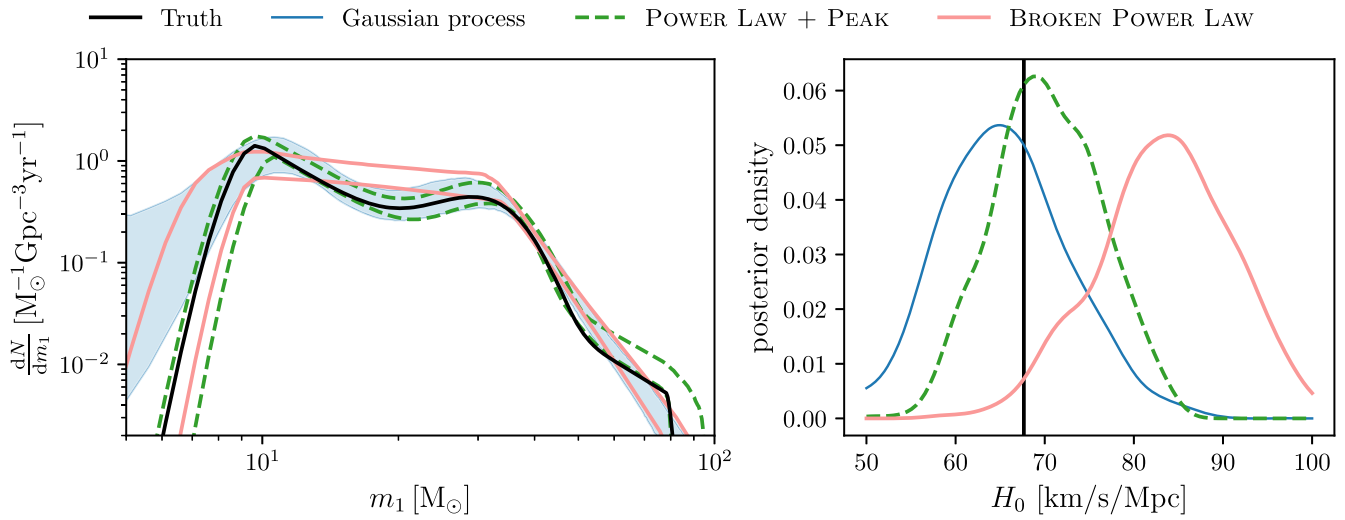

Figure 1. Spectral siren measurement for a simulated catalog with the correct parametric model (POWER LAW + PEAK, green dashed), a deliberately incorrect parametric model (BROKEN POWER LAW, pink solid) and the nonparametric model presented in this work (Gaussian process, blue filled). The left panel shows the recovered source-frame primary mass distribution for each model, and the simultaneously inferred posteriors on H_0 are shown in the right panel. The bands in the left panel represent the 90% credible intervals. The mass distribution and H_0 value used to generate the data are shown by a solid black line in each panel. The deliberately incorrect parametric model fails to recover the true mass distribution and therefore produces an estimate of H_0 that is offset from the true value, whereas both the correct and nonparametric models recover the mass distribution and H_0. As the true mass distribution is unknown for real observations, using a nonparametric model mitigates systematic uncertainty that would otherwise arise from mismodeling the compact binary coalescence (CBC) population.

The full mass distribution therefore acts analogously to an electromagnetic spectrum, in which the apparent locations of spectral features relative to their rest-frame locations provide a redshift measurement. The method of using the mass distribution of GW sources to measure cosmological parameters has therefore been coined "spectral sirens" (J. M. Ezquiaga & D. E. Holz 2022). Spectral sirens were first demonstrated to be a feasible method to measure the Hubble constant by D. F. Chernoff & L. S. Finn (1993) and S. R. Taylor et al. (2012) using the binary neutron star mass distribution, and extended to the binary black hole (BBH) mass distribution by W. M. Farr et al. (2019). Spectral siren analyses have since been implemented by the LIGO–Virgo–KAGRA Collaborations using the latest GW catalog (R. Abbott et al. 2023c).

Central to the spectral siren methodology is knowledge of the compact binary mass distribution. However, first principles models for mass distributions of merging compact binaries are not available: significant theoretical uncertainties exist about the shapes, locations, and very existence of predicted features in the mass distribution (e.g., M. Zevin et al. 2017; M. Mapelli 2020; M. Zevin et al. 2021; P. Marchant & J. Bodensteiner 2024). This includes uncertainties about potentially dominant features, such as the existence of a "pileup" due to pulsational-pair instability, a pair-instability-driven upper mass gap (R. Farmer et al. 2019), the maximum neutron star mass (C. L. Fryer & V. Kalogera 2001; J. Alsing et al. 2018), and the existence of a putative lower mass gap between neutron stars and black holes (F. Özel et al. 2010; W. M. Farr et al. 2011; A. Farah et al. 2022). To this end, spectral siren cosmology relies on *simultaneously* measuring a source-frame mass spectrum alongside cosmological parameters. This is typically accomplished by adopting a phenomenological, parametric model for the mass distribution, usually composed of power laws and Gaussians (e.g., R. Abbott et al. 2023c; S. Mastrogiovanni et al. 2023b).

Such parametric modeling of the compact binary mass distribution raises its own set of dangers. It is well known that different parametric models can generically yield very different constraints on cosmological parameters (R. Abbott et al. 2021b, 2023a). This is problematic: because the measured mass distribution serves as the template by which to extract redshifts, a mismodeled mass distribution would introduce systematic errors in inferred redshifts and, in turn, systematically bias any resulting cosmological inference (S. Mastrogiovanni et al. 2021; J. M. Ezquiaga & D. E. Holz 2022; S. Mukherjee 2022; G. Pierra et al. 2024).

The situation is demonstrated in Figure 1, in which we perform spectral siren cosmology on a simulated population of binary black holes. We reconstruct the mass distribution using two parametric models, one that contains the true simulated mass distribution and one that does not. While the former yields a measurement of H_0 consistent with the true underlying value, the latter does not. Such systematic biases may already be relevant, as cosmological measurements by the LIGO–Virgo–KAGRA are known to depend on the choice of mass model used (R. Abbott et al. 2023c). Furthermore, these biases may become a dominant source of uncertainty in the near future (G. Pierra et al. 2024). The prospects of such a dominant systematic uncertainty is troubling. If prior knowledge of the mass distribution's morphology (whether an exact theoretical prediction or knowledge of the correct parametric family of models) is a prerequisite for the spectral siren method, the effectiveness of such a technique would be significantly hampered.

In this work, we explicitly demonstrate that no prior knowledge of the shape of the compact binary coalescence (CBC) mass spectrum is necessary to use the spectral siren methodology. We do this by inferring $H(z)$ with a flexible, nonparametric model for the mass distribution of CBCs (blue shaded band in Figure 1). This model makes minimal prior assumptions about the shape of the mass distribution, enabling it to accurately infer a wide range of morphologies and remain agnostic to the astrophysical processes that give rise to features in the mass distribution. Despite its flexibility, our approach is able to consistently obtain unbiased measurements of cosmological parameters, showing that nonparametric methods are

not only sufficient for a spectral siren measurement, they can also mitigate systematic effects in the measurement caused by model misspecification.

The nonparametric mass model's ability to recover injected cosmological parameters demonstrates that the information in the spectral siren measurement does not come from the enforcement of specific features in the mass distribution. Rather, it is provided by the assumption that either all CBCs follow a common mass distribution, or that any evolution of the mass distribution with redshift does not exactly mimic cosmology (J. M. Ezquiaga & D. E. Holz 2022).

We find that our nonparametric model allows for a 11.2% measurement of H_0 and a 6.4% measurement of $H(z=0.9)$ during the fifth LIGO–Virgo–KAGRA observing run (O5), when the detectors will reach their design sensitivity. We highlight measurements of H_0 within a ΛCDM universe in order to benchmark the accuracy and precision of our nonparametric method, as well as explore the role of spectral sirens in elucidating the Hubble tension. However, a primary utility of spectral siren measurements will be in constraining $H(z)$ under different cosmological models, and at redshifts that are relatively inaccessible by electromagnetic observations, especially with next-generation gravitational-wave detectors (J. M. Ezquiaga & D. E. Holz 2021; Z.-Q. You et al. 2021; J. M. Ezquiaga & D. E. Holz 2022; H.-Y. Chen et al. 2024).[7]

This paper is organized as follows. Section 2.1 describes the simulated data set. Section 2.2 introduces the spectral siren method, demonstrating how cosmological parameters are inferred from the mass distribution of GW sources. Section 2.3 describes the nonparametric mass distribution we develop for use within the spectral siren method. In Section 3, we present the results of using parametric and nonparametric mass distributions, as well as projections for future constraints on $H(z)$. We discuss the implications of our results and outline future work in Section 4.

This study was carried out using the reproducibility software show your work! (R. Luger et al. 2021), which leverages continuous integration to programmatically download the data from Zenodo, create the figures, and compile the manuscript. The icons next to each figure caption are hyperlinks that lead to the code used to make that figure (pencil icons), and to the data behind the figure (down-arrow icons). Some programmatically generated numbers in the text also have these icons, which function in the same way. We encourage readers to click these hyperlinks to verify or reproduce the results and methods of this work. The git repository associated to this study allows anyone to rebuild the entire manuscript, and is publicly available;[8] a copy of version 0.2 has been deposited to Zenodo at doi:10.5281/zenodo.13363159. Inference results are also available on Zenodo at doi:10.5281/zenodo.13363131.

2. Cosmology with an Astrophysically Agnostic Mass Model

The spectral siren method functions by identifying the relationship between luminosity distance and redshift that causes all source-frame masses to follow a distribution that smoothly varies as a function of time. This allows for the simultaneous inference of both a mass distribution and cosmological parameters, even if the form of the source-frame mass distribution is not known in advance. As described above, though, strongly parameterized models for the mass distribution yield biased measurements of cosmological parameters if they poorly approximate the true mass distribution of compact binaries. We aim to circumvent such biases and instead model the population of GW sources in a flexible and astrophysically agnostic way.

There exist several nonparametric methods developed for this purpose (I. Mandel et al. 2019; V. Tiwari 2021; B. Edelman et al. 2022; S. Rinaldi & W. Del Pozzo 2022; J. Sadiq et al. 2022; T. A. Callister & W. M. Farr 2024; B. Edelman et al. 2023; A. Ray et al. 2023). While well suited to infer the GW source population with a fixed cosmology, several of these methods employ fixed features in source-frame mass, such as bin edges (I. Mandel et al. 2019; A. Ray et al. 2023) or spline nodes (B. Edelman et al. 2022). Since these locations were chosen with a fixed cosmology, they risk causing the inference to prefer the cosmological parameters assumed when choosing the feature locations. Indeed, S. Mastrogiovanni et al. (2021) show how using fixed features can significantly bias cosmological inference within the spectral siren methodology. We therefore opt for a model of the source-frame mass distribution that forgoes the need to define such features.

For this purpose, we construct a model with a Gaussian process (GP), a common tool for nonparametric inference. GPs define a random space of functions in which any subset of function values are jointly Gaussian distributed (C. E. Rasmussen & C. K. I. Williams 2006). Their smoothness properties make them widely useful in GW data analysis for regression problems, such as modeling time-domain waveforms (Z. Doctor et al. 2017; E. A. Huerta et al. 2018) and the neutron star equation of state (P. Landry & R. Essick 2019), density estimation problems, such as estimating posterior densities of single-event parameters from parameter estimation samples (V. D'Emilio et al. 2021), and as a prior on histogram bin heights for population inference (I. Mandel et al. 2017; Y.-J. Li et al. 2021; A. Ray et al. 2023).

Our use case is slightly different from previous analyses: we utilize a GP as a prior on the functions that describe the primary mass distribution of CBCs. This choice encodes very little prior information about the shape of the mass distribution, besides enforcing that it must be smooth.

2.1. Simulated Data

To demonstrate the effectiveness of our GP-based mass distribution in agnostically inferring both cosmological parameters and the population properties of GW sources, we apply our methodology to a simulated data set. By generating a catalog of GW sources from a known population and cosmological model, we are able to quantify the accuracy of our method. The use of simulated data also enables us to make projections for future data sets and safely ignore dimensions such as spin that do not impact cosmological measurements but would otherwise be important to simultaneously fit to avoid biases in population inference of real data, as real data exhibit correlations between distance, mass, and spin measurements (S. Biscoveanu et al. 2022).

We design our simulated catalog to match the characteristics of the data expected from one year of observation in O5. The BBHs in this catalog are drawn from an underlying population described by the POWER LAW + PEAK mass distribution presented in C. Talbot & E. Thrane (2018) and used in

[7] See example spectral siren cosmological inference using parametric mass spectrum models: https://github.com/ezquiaga/spectral_sirens.

[8] https://github.com/afarah18/spectral-sirens-with-GPs

B. P. Abbott et al. (2019) and R. Abbott et al. (2023a), and follow the redshift distribution presented in T. Callister et al. (2020), with hyperparameters consistent with those found in R. Abbott et al. (2023a). We assume the mass distribution does not evolve across the redshift range to which the O5 detectors will be sensitive. This assumption is consistent with current data (M. Fishbach et al. 2021; L. A. C. van Son et al. 2022; R. Abbott et al. 2023a), in which no redshift evolution of the black hole mass function is detected. At the same time, a redshift-dependent mass function is a generic astrophysical prediction, due either to changing evolutionary environments or evolving mixture fractions between distinct compact binary formation channels (C. J. Neijssel et al. 2019; L. A. C. van Son et al. 2022; S. Torniamenti et al. 2024; C. S. Ye & M. Fishbach 2024). We discuss this possibility further in Section 4, but leave it primarily for future work.

We use the `GWMockCat` (A. M. Farah et al. 2023) package to apply O5-like selection effects to the drawn BBHs, generate realistic measurement uncertainty, and produce sensitivity estimates that are consistent with the simulated GW signals. We will use the term "event" to refer to GW signals that pass the criteria for detection. This process results in a catalog of $N_{\rm ev} = 591$ GW signals that pass the criteria for detection, hereon called events. Additional details of the data simulation, including the form of the injected population, are described in Appendix A.

2.2. The Spectral Siren Method

To simultaneously infer cosmological parameters and the population of GW sources, we employ a hierarchical Bayesian analysis. This allows us to undo the selection effects of GW detectors to obtain a true, astrophysical population and constrain the cosmic expansion history.

Given a source population and background cosmology described by hyperparameters Λ, the likelihood of observing data $\{d\}$ that contain $N_{\rm ev}$ detected GW signals, each with parameters θ_i, is (T. Loredo 2009; S. R. Taylor et al. 2012; I. Mandel et al. 2019; S. Vitale et al. 2020)

$$p(\{d\}, \{\theta\}|\Lambda) \propto e^{-N_{\rm exp}(\Lambda)} \prod_i^{N_{\rm ev}} p(d_i|\theta_i) \frac{dN}{dt_{\rm det} d\theta}(\theta_i; \Lambda). \quad (1)$$

Here, $\frac{dN}{dt_{\rm det} d\theta}(\theta; \Lambda)$ is the detector-frame merger rate density of BBHs, conditioned on hyperparameters Λ. Following T. A. Callister & W. M. Farr (2024), we use a semicolon to explicitly indicate that this is a function of Λ, not a density over Λ. $N_{\rm exp}(\Lambda)$ is the expected number of detections given Λ and the GW detector sensitivity, and is calculated using a Monte Carlo sum over $N_{\rm inj}$ found signals injected into the data stream (see R. Essick & M. Fishbach 2021; R. Essick & W. Farr 2022 for a detailed explanation of this process).

In this work, we have restricted our analysis to the BBH primary mass distribution. However, the method can be trivially extended to the full mass distribution of CBCs (e.g., M. Fishbach et al. 2020; J. M. Ezquiaga & D. E. Holz 2022). Additionally, Appendix B demonstrates that including the mass ratio distribution in the fit does not change the results of the analysis presented here, so long as the mass ratio distribution is accurately modeled.

Since parameters of individual events are not perfectly measured, we marginalize over the possible properties of each event. Practically, this is done by a Monte Carlo average over the posterior samples $\{\theta_j\}_i$ of each event i and dividing out the prior used when inferring those posterior samples, $\pi_{\rm PE}(\theta)$:

$$\begin{aligned} p(\{d\}|\Lambda) &\propto e^{-N_{\rm exp}(\Lambda)} \prod_i^{N_{\rm ev}} \int d\theta_i \, p(d_i|\theta_i) \frac{dN}{dt_{\rm det} d\theta}(\theta_i; \Lambda) \\ &\approx e^{-N_{\rm exp}(\Lambda)} \prod_i^{N_{\rm ev}} \frac{1}{N_{\rm samps}} \sum_{j=1}^{N_{\rm samps}} \frac{\frac{dN}{dt_{\rm det} d\theta}(\theta_{j,i}; \Lambda)}{\pi_{\rm PE}(\theta_{j,i})}. \end{aligned} \quad (2)$$

When combined with a prior $p(\Lambda)$ (to be discussed below) on the population and cosmological parameters, the result is a posterior on both the compact binary population and the background cosmology. The full set of hyperparameters Λ therefore includes the shape of the mass distribution as well as all cosmological parameters that dictate the D_L–z relation: the local expansion rate H_0, the present fractional energy densities of dark matter Ω_M, dark energy Ω_Λ, and radiation Ω_r, and the equation of state of dark energy w. In this work, we fix $\Omega_\Lambda = 1 - \Omega_M$, $\Omega_r = 0$, and $w = -1$ and use uniform priors on H_0 and Ω_M, corresponding to a flat ΛCDM cosmology.

The process by which we sample the likelihood in Equation (1) is outlined in Appendix C.

2.3. Gaussian Process–based Mass Distribution

In this Section, we give an overview of the nonparametric mass model developed for this work. Further details on this model, including an introduction to GPs and a discussion of their properties is given in Appendix C.

With the GP approach, the hyperparameters describing the mass distribution are the rate at each event-level posterior sample's source-frame mass, and the rate at each found injection's source-frame mass. The GP *is* $p(\Lambda)$, the prior on population parameters (except in the case of cosmological parameters, which all have uniform priors). This is demonstrated in Figure 2, where the left panel shows draws from a GP, which are prior draws for the population inference. The smooth appearance of individual draws from the population prior, as well as the absence of overdensities at specific source-frame mass values in the full prior distribution shown in Figure 2 demonstrate that we have successfully fulfilled our goal to construct a model without predefined features in source-frame mass. Combined with the population likelihood in Equation (1), the prior illustrated in the left panel of Figure 2 gives the population posterior in the right panel.

The lack of data just below the minimum black hole mass and just above the maximum black hole mass, combined with the fact that GW detectors are sensitive to objects at those masses, causes the GP to learn a relatively low merger rate at the edges of the mass distribution. That said, there is both a lack of data and little detector sensitivity at masses above $\sim 100\, M_\odot$ and below $\sim 5\, M_\odot$. The mass distribution is therefore uninformed in this region and the GP reverts to its prior distribution, which resembles random scatter around the mean differential merger rate. Similar effects can be seen in other nonparametric methods (T. A. Callister & W. M. Farr 2024; B. Edelman et al. 2023). The combination of these two effects results in what appears as an uptick in the merger rate below

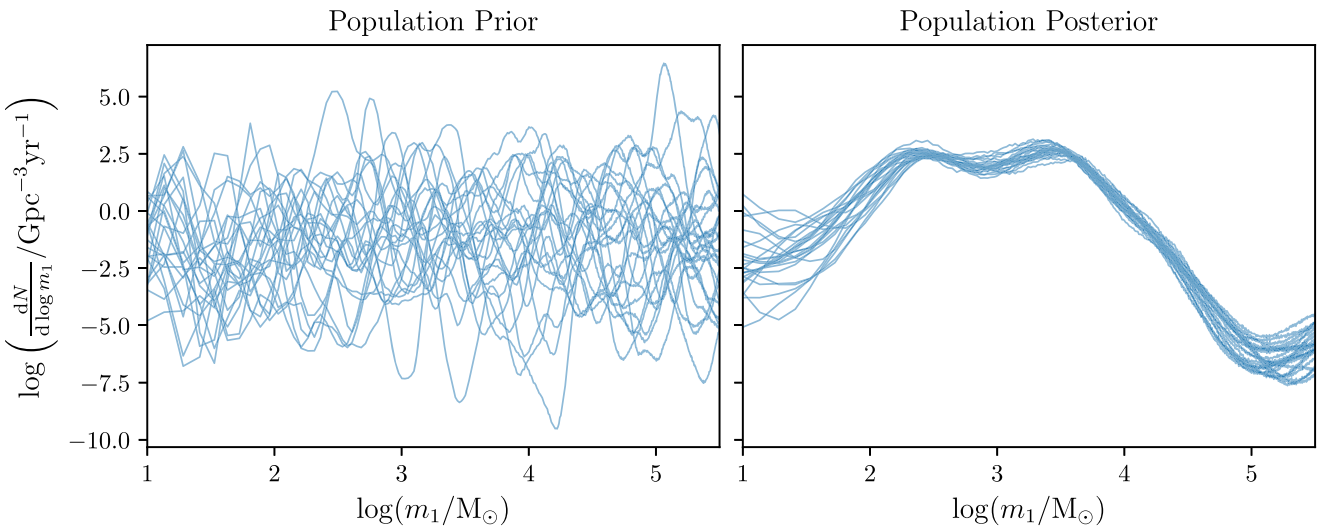

Figure 2. Draws from the Gaussian process (GP) used to model the mass distribution. The left panel shows prior draws from the GP, and the right panel shows posterior draws once the population inference is performed on the simulated data. The posterior draws in the right panel are a subset of those used to create the 90% credible intervals in Figure 1.

$\sim 5\,M_\odot$ and above $\sim 100\,M_\odot$. However, this reversion to the prior is uninformative for the H_0 constraint and does not affect inference on cosmological parameters. Additionally, the posterior on H_0 is distinct from its prior distribution (uniform in the range [30 $\mathrm{km}^{-1}\,\mathrm{s}^{-1}$ Mpc, 120 $\mathrm{km}^{-1}\,\mathrm{s}^{-1}$ Mpc], indicating that the data are informative despite the flexibility of the population model.

We note that other nonparametric methods may be adapted to avoid predefined features, such as fitting for the locations of their features simultaneously with the rest of the inference (e.g., V. Tiwari 2021) or, in the case of splines, by using a smoothing function that allows for features to occur at arbitrary locations appropriate smoothing (e.g., B. Edelman et al. 2023).

The smoothness of a given GP is determined by its kernel, which is a function that defines the covariance between input points in the GP (in our case, two source-frame mass values). It defines the notion of similarity between adjacent points and thereby encodes our assumptions about the smoothness of the source-frame mass distribution (C. E. Rasmussen & C. K. I. Williams 2006). Kernels themselves have parameters that determine their properties. In our use case, these are one level further removed from hyperparameters, so we adopt the terminology used in T. A. Callister & W. M. Farr (2024) and call them "hyper-hyperparameters." We fit these hyper-hyperparameters along with the hyperparameters Λ to minimize prior assumptions about the form of the mass distribution.

3. Results

In this Section we show that fitting an incorrect functional form to the mass distribution of CBCs biases the inference of cosmological parameters when using the spectral siren methodology. We then demonstrate that our flexible model alleviates this bias without the need to know the morphology of the mass distribution a priori. We illustrate this explicitly by using three different models for the source-frame mass distribution to infer the cosmic expansion rate from the simulated catalog described in Section 2.1 and Appendix A. The three models are as follows:

1. POWER LAW + PEAK, which includes the true mass distribution within its hyperprior,
2. the BROKEN POWER LAW model presented in R. Abbott et al. (2021b), as we do not employ a high-mass truncation, which does not include the true mass distribution within its hyperprior, and
3. the flexible, GP-based model described in Section 2.3, which is able to closely approximate the morphology of the true mass distribution, along with many other morphologies.

For all models considered in this work, we assume the form of the redshift distribution used to generate the data, described in Equation (A3). We have performed the analysis both with a fixed redshift distribution and while simultaneously fitting for the redshift distribution and find no qualitative differences in our conclusions: fitting for the redshift distribution broadens the inferred posteriors on H_0 equally for all mass models, but does not affect their mean values.

The results of fitting each model to the same data set are shown in Figure 1. The left panel shows the inferred source-frame mass distribution for each of the considered models, and the right panel shows the corresponding posteriors on H_0. We indicate the true underlying source mass distribution and H_0 value with solid black lines in each panel.

The fits presented in Figure 1 are representative results from a single simulated catalog. These provide insight into the full statistical results presented below. In particular, it can be seen that the BROKEN POWER LAW (orange curve) is inconsistent with the true value of H_0: in the run shown in Figure 1, the true value of H_0 is offset from the mean of the posterior by 2.7σ. By contrast, the POWER LAW + PEAK and GP-based models (green and blue curves) are consistent with the underlying truth. These models recover mean values of H_0 that are offset from the true value at 0.8σ and 0.3σ, respectively. Additionally, the mass distribution inferred with the GP-based model closely resembles the true, simulated distribution. This indicates that using models that cannot accurately approximate the true mass distribution will lead to a noticeable systematic bias in the estimation of cosmological parameters.

This bias is not due to the need to know the morphology of the mass distribution a priori, as the GP-based model recovers the correct value of H_0 despite making minimal assumptions about the mass distribution. In reality, we do not know the true functional form of the mass distribution, so it may be desirable to use a nonparametric approach to avoid potential systematic errors introduced by choosing a parametric model that likely does not contain the true mass distribution within its hyperprior.

To obtain a quantitative measure of the systematic bias introduced by mismodeling the mass distribution, we repeat the parametric analyses with 50 separate simulated catalogs of ∼1000 events each. We find that the BROKEN POWER LAW model produces an over- or underestimate of H_0 at greater than 1σ 90%✎ of the time, and the POWER LAW + PEAK model reaches the same level of bias only 26%✎ of the time, meaning that the BROKEN POWER LAW model produces a bias more than 3 times as often as the POWER LAW + PEAK model. Additionally, we show in Appendix D that the BROKEN POWER LAW model typically overestimates H_0, whereas the POWER LAW + PEAK model produces a roughly equal number of over- and underestimates of H_0. This demonstrates that mismodeling the mass distribution can introduce statistically significant systematic biases into measurements of cosmological parameters.

Collectively, our results indicate that prior knowledge of the shape of the mass distribution is not required to perform an unbiased spectral siren measurement, so long as strong assumptions about the shape of the mass distribution are not made.

3.1. Projections for Future Measurements

Figure 1 demonstrates an expected 11.2%✎ (1σ uncertainties) measurement of H_0 after one year of O5 using the GP-based spectral siren method, and an 8%✎ measurement with parametric spectral sirens, demonstrating comparable statistical uncertainties. However, we note that the precision reached in O5 may lessen depending on the actual level of measurement uncertainty in individual events, and the existence (or lack thereof) of a maximum mass feature, which we have assumed to be present in our simulated data set. These numbers are estimated from a fit to a single simulated catalog, but we find similar levels of statistical uncertainty from fits to different catalog realizations. By the time of O5, the GW detector network is projected to detect BBHs up to redshift ≈3, with most sources lying near redshift ≃1.2 (H.-Y. Chen et al. 2021). Additionally, next-generation detectors will be sensitive to sources up to redshift ∼100. This means that future GW observations will be more sensitive to $H(z \gtrsim 1)$ than to H_0, and can therefore constrain several additional cosmological parameters (H.-Y. Chen et al. 2024). We demonstrate this by repeating the same GP-based spectral siren analysis while also simultaneously fitting for the local matter density, Ω_M. The result is shown in Figure 3. We emphasize that these precise measurements over a wide range in redshift enable precision estimation of additional cosmological parameters governing $H(z)$.

We find O5 observations to be most sensitive to $H(z = 0.9)$✎, which is measured at 6.4%✎. The left panel of Figure 3 demonstrates a strong anticorrelation between the Ω_M and H_0 posteriors, resulting in similarly informative constraints on the two parameters. This is because H_0 controls the y-intercept of the $H(z)$ curves on the right panel, while Ω_M informs the slope of those curves; the same measurement of $H(z \neq 0)$ can be obtained by increasing the slope while decreasing the y-intercept, and vice versa. Similar behavior can be observed in current measurements of the BBH redshift distribution, which exhibits a tightening of the posterior at $z \sim 0.2$ with current observations (R. Abbott et al. 2023a; T. A. Callister & W. M. Farr 2024).

Next-generation detectors will be sensitive to a larger range of redshifts (ET steering committee et al. 2020; M. Evans et al. 2021) and will therefore break the degeneracy between cosmological parameters and allow for tighter constraints on both Ω_M and H_0. However, the small cosmological volume (and thus low number of mergers) at low redshift will generally limit the constraining power of spectral sirens at $z = 0$, potentially making this method more sensitive to cosmological parameters that affect higher redshifts. Combining spectral sirens with other methods that are sensitive to the local expansion rate, such as those that employ electromagnetic counterparts, may increase the precision of GW standard sirens at all redshifts (e.g., H.-Y. Chen et al. 2024).

4. Discussion

GWs are unique cosmic messengers in that they carry both redshift and distance information, making them remarkably clean probes of the Universe's expansion history calibrated directly by the theory of general relativity. However, the current method of determining GW redshifts via the distribution of their source masses (i.e., spectral sirens) employs an assumption of the shape of their population, typically encapsulated by simplified parametric functions. This choice, often presumed to be necessary or fundamental to the method, may introduce a systematic bias to a measurement of the cosmological parameters that is otherwise appealing for its unique elegance and simplicity.

In this work we show that a specific choice of mass distribution is unnecessary to arrive at informative and accurate posteriors on H_0 and Ω_M. We do this by accurately measuring these cosmological parameters with a highly flexible model for the mass distribution. This reinforces the notion that the information in the spectral siren measurement comes from the assumption that all GW sources come from the same population, a far less stringent statement than the assumption that we understand the exact astrophysical processes that give rise to that population (i.e., the physics governing compact binary formation and evolution).

Spectral sirens are particularly useful in the context of the current Hubble tension: a disagreement between multiple methods of measuring the local expansion rate of the Universe (W. L. Freedman 2021). As this tension can only be explained by nonstandard physics or yet-unknown systematic uncertainties in either the cosmic distance ladder or cosmic microwave background measurements, direct and independent probes of the local expansion—such as those presented here—may help determine whether the current discrepancy represents a fundamental crack in our standard model of physics and/or cosmology.

Neither the mechanism from which compact binaries were formed nor the physical processes within the mechanism have been conclusively determined. The true functional form of the mass distribution therefore remains elusive, meaning that systematic uncertainties arising from an incorrect choice for

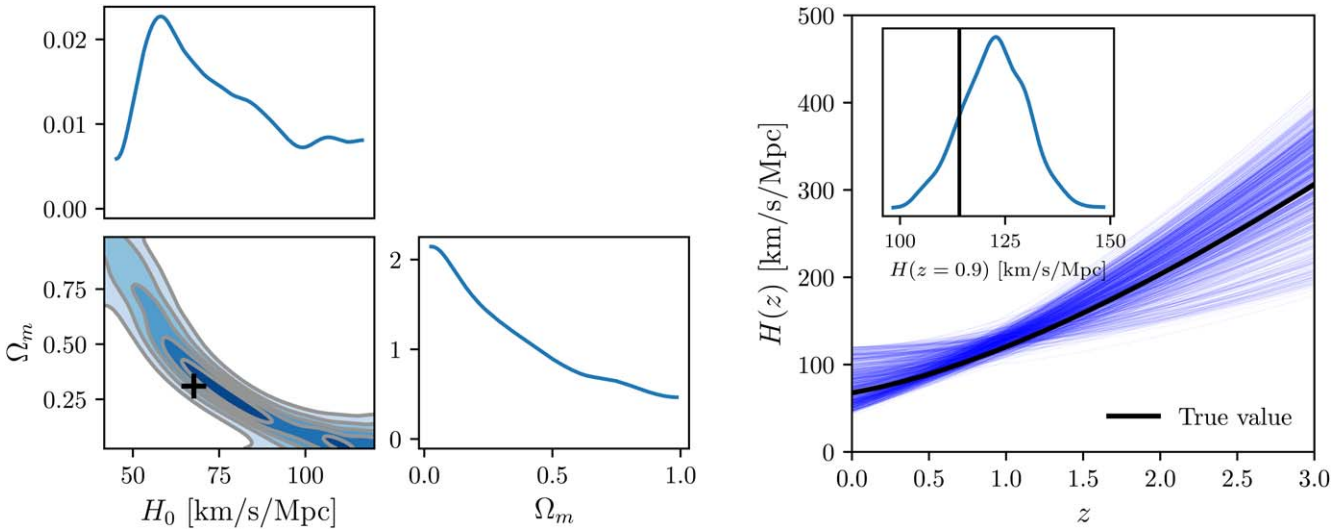

Figure 3. Projected constraints on multiple cosmological parameters after one year of observing at the LIGO–Virgo–KAGRA's design sensitivity, using the GP-based spectral siren method. The right panel shows the inferred expansion history of the Universe, $H(z)$. It will be measured most precisely at $z = 0.9$, as can be seen by the narrowing of the inferred $H(z)$ curves there. The inset shows the posterior on $H(z = 0.9)$. Black solid lines indicate the true value of $H(z)$ in both the inset and main panel. The left panel shows the two-dimensional posterior on H_0 and Ω_M, with the true value indicated by a black "+." The two parameters are strongly degenerate because of the multiple ways of measuring $z = 0.9$. Spectral sirens are particularly well suited for measuring cosmological parameters that affect the Universe at $z \gtrsim 0.2$.

the form of the mass distribution are inevitable. With current observations, these systematic effects are likely smaller than statistical uncertainties. However, next-generation detectors will herald sufficient GW observations to substantially decrease statistical uncertainty in these measurements; for example, G. Pierra et al. (2024) show that incorrect assumptions about the shape of the mass distribution can lead to $\sim 3\sigma$ systematic biases in H_0 with catalogs of 2000 events, although this bias may be an overestimate as it does not include measurement uncertainty of the GW parameters. Thus, nonparametric approaches may be preferable to avoid the systematic errors associated with choosing a parametric model.

In parallel with this work, I. Magaña Hernandez & A. Ray (2024) performed a spectral siren analysis on public LIGO–Virgo–KAGRA data using histogram bins defined at fixed locations in source-frame mass to flexibly measure the BBH mass distribution (as originally presented in I. Mandel et al. 2017; A. Ray et al. 2023). Both our method and theirs employ GPs. In our approach, the black hole mass function is *itself* described by a Gaussian process, with no predefined features. In I. Magaña Hernandez & A. Ray (2024), by contrast, the mass distribution is fundamentally a binned piecewise constant function, with a GP prior governing the relative heights of these bins. While this binned model is flexible, sharp bin edges at predefined locations nevertheless constitute unphysical features. These features may yield overly optimistic constraints on cosmological parameters, possibly explaining the increase in precision that I. Magaña Hernandez & A. Ray (2024) find relative to the parametric analysis on the exact same data set performed in R. Abbott et al. (2023c). I. Magaña Hernandez & A. Ray (2024) also include a binned reconstruction of the BBH merger rate as a function of redshift, which we fix or model parametrically in our own work. We stress, though, that this is not an intrinsic limitation of our method; one could straightforwardly adopt a second Gaussian process to govern the overall scaling of the merger rate with redshift.

More interesting is the possibility that the black hole mass distribution itself evolves with redshift. As demonstrated in J. M. Ezquiaga & D. E. Holz (2022), the degeneracy between an evolving mass distribution and the expansion of the Universe can be broken provided that there are multiple features present in the mass distribution, and that we do not live in a fine-tuned Universe where the evolutionary effects governing the CBC mass distribution perfectly mimic the effects of cosmological redshift. The first condition (multiple features) is known to be met in current data (R. Abbott et al. 2023a), with three robust features (A. M. Farah et al. 2023): a maximum black hole mass and overdensities at $\sim 10\,M_\odot$ and $\sim 35\,M_\odot$. It is also predicted by many population synthesis studies, (e.g., N. Giacobbo et al. 2018; M. Mapelli & N. Giacobbo 2018; L. A. C. van Son et al. 2023). The latter scenario —with features identically and monotonically shifting to higher masses with increasing redshift—would be astrophysically unlikely as the locations of features in the mass distribution are each thought to be governed by fundamentally different physical processes (M. Mapelli 2020). Extensions to the nonparametric methods presented here will allow for arbitrary redshift evolution. Additionally, since Gaussian processes naturally scale to multiple data dimensions (C. E. Rasmussen & C. K. I. Williams 2006), the method presented here can easily be generalized to fit the redshift dependence of the mass distribution. Future work will extend the method developed here to mass distributions that are allowed to evolve with redshift. When applying an evolving mass distribution model to a data set that exhibits evolution, we expect errors on cosmological parameters to broaden. Noticeable redshift evolution in the mass distribution is expected to occur above $z \simeq 1.5$ (L. A. C. van Son et al. 2022), which is not yet relevant in current data or in expected O5 data, but will certainly be visible in next-generation detectors.

Because GW observations are the only data input to spectral sirens, they are sensitive to the expansion history of the Universe over a wide range of cosmological redshift, rather than just the local expansion H_0. We have shown this by simultaneously measuring Ω_M and H_0; the method can be trivially expanded to constrain additional cosmological

parameters that govern $H(z)$, such as the dark energy equation of state parameter, w. It is also possible to use a non-ΛCDM cosmological model for $H(z)$, with different parameters entirely. It may also be possible to forgo the need for a parametric representation of $H(z)$ altogether, for example by using bins in luminosity distance. A similar technique using redshift bins is commonly used in cosmological measurements using galaxy surveys and Type 1a supernovae (e.g., R. Beig et al. 2003; L. Anderson et al. 2014; A. Aghamousa et al. 2016), but we have not pursued that possibility in this work. Such a method would likely increase uncertainties on $H(z)$, especially for $z \gtrsim 1.5$. We find that for the fifth LIGO–Virgo–KAGRA observing run, spectral sirens will be most constraining at $z = 0.9$. This redshift is larger than the expected redshifts of detectable electromagnetic counterparts of binary neutron star mergers (R. W. Kiendrebeogo et al. 2023), implying that upgrades to current GW detectors will allow the spectral siren method to probe $H(z)$ at otherwise unexplored distances by GWs.

Proposed next-generation GW detectors such as Cosmic Explorer and Einstein Telescope will be sensitive to CBCs across cosmic time (out to $z \sim 100$; ET steering committee et al. 2020; M. Evans et al. 2021). Future cosmological surveys, such as those enabled by the Nancy Grace Roman Space Telescope and the Vera Rubin Observatory, are expected to be able to precisely measure $H(z)$ to $z \sim 3$ (D. Spergel et al. 2015), making the spectral siren method uniquely positioned to measure the expansion of the Universe at high redshift. When combined with bright and dark siren methods, the low-redshift expansion history will also be well constrained (H.-Y. Chen et al. 2024). Sensitivity to high redshifts is a feature of the spectral siren method in general, and nonparametric methods such as the one presented here will be imperative to avoid systematic biases in spectral siren cosmology.

Acknowledgments

The authors thank Reed Essick, Ben Farr, Maya Fishbach, Utkarsh Mali, and Colm Talbot for helpful conversations. A.M.F. is supported by the National Science Foundation Graduate Research Fellowship Program under grant No. DGE-1746045. J.M.E. is supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 847523 INTERACTIONS, and by VILLUM FONDEN (grant Nos. 53101 and 37766). M.Z. gratefully acknowledges funding from the Brinson Foundation in support of astrophysics research at the Adler Planetarium. D.E.H. is supported by NSF grants AST-2006645 and PHY-2110507, as well as by the Kavli Institute for Cosmological Physics through an endowment from the Kavli Foundation and its founder Fred Kavli. The Tycho supercomputer hosted at the SCIENCE HPC center at the University of Copenhagen was used for supporting this work. This material is based upon work supported by NSF's LIGO Laboratory, which is a major facility fully funded by the National Science Foundation.

Software: `numpyro` (E. Bingham et al. 2019; D. Phan et al. 2019), `tinygp` (D. Foreman-Mackey et al. 2021), `arviz` (R. Kumar et al. 2019), `jax`, show your work! (R. Luger et al. 2021).

Appendix A
Details of Data Simulation

The exact form of the injected population is

$$\frac{dN}{dm_1 dz} \propto p(m_1|\bar{\Lambda}_m)p(z|\bar{\Lambda}_z, H_0, \Omega_M), \tag{A1}$$

where

$$p(m_1, m_2|\bar{\Lambda}_m) \propto \mathcal{S}(m_{\rm min}, m_{\rm max}) \times \left(f_{\rm peak}\, e^{-\frac{1}{2}\left(\frac{m_1-\mu}{\sigma}\right)^2}\mathcal{N}_{\rm g} + (1-f_{\rm peak})m_1^{\alpha}\mathcal{N}_{\rm pl}\right), \tag{A2}$$

$$p(z|H_0, \Omega_M) \propto \frac{dV_C}{dz}\frac{1}{1+z}\frac{(1+z)_z^{\alpha}}{1+\left(\frac{1+z}{1+z_p}\right)^{\alpha_z+\beta_z}}. \tag{A3}$$

Here, $V_C(H_0, \Omega_M)$ is the comoving volume for given cosmological parameters H_0 and Ω_M, and $\bar{\Lambda}_m = \{\alpha, m_{\rm min}, m_{\rm max}, \mu, \sigma, f_{\rm peak}\}$ are the (hyper-)parameters describing the power law in primary mass, minimum and maximum black hole mass, Gaussian peak location and width, and fraction of events in the Gaussian peak, respectively. $\mathcal{S}$ is a smoothing function at low and high masses, and $\mathcal{N}_{\rm pl}$ and $\mathcal{N}_{\rm g}$ are the normalizations between $m_{\rm min}$ and $m_{\rm max}$ for the power-law component and truncated Gaussian component, respectively. The smoothing creates support for masses below $m_{\rm min}$ and above $m_{\rm max}$. $\bar{\Lambda}_z = \{z_p, \alpha_z, \beta_z\}$ are the parameters governing the peak of the redshift distribution, low-z power-law slope, and high-z power-law slope. When generating the simulated events, we have fixed $\alpha = -2.7$, $m_{\rm max} = 78\,M_\odot$, $m_{\rm min} = 10\,M_\odot$, $\mu = 30\,M_\odot$, $\sigma = 7.0\,M_\odot$, $f_{\rm peak} = 0.05$, $z_p = 2.4$, $\alpha_z = 1.0$, and $\beta_z = 3.4$. These choices correspond to the maximum a posteriori values obtained by an analysis of GWTC-3 data using the POWER LAW + PEAK model (R. Abbott et al. 2023a). For simplicity, we assume a uniform mass ratio distribution. Thus, $\theta = \{m_1, z\}$. We consistently apply these assumptions to the data generation process and the population inference. We do not fit for or simulate spins, as they do not redshift and hence do not carry additional cosmological information, and we fix the distribution of mass ratios.

We use cosmological parameters $H_0 = 67.66\,{\rm km}^{-1}\,{\rm s}^{-1}\,{\rm Mpc}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 1 - \Omega_M$, consistent with those found by Planck Collaboration et al. (2016). We emphasize, however, that the choice of cosmological parameters for data generation is arbitrary and does not impact the results, since we are concerned only with the ability of our method to recover the injected values. Throughout the data generation and inference, we use the approximations presented in M. Adachi & M. Kasai (2012) to efficiently convert between D_L and z for a given set of cosmological parameters Ω_M and H_0.

We do not include neutron stars in our simulation set, as their contribution to the spectral siren measurement is expected to be subdominant in O5. However, if a lower mass gap between the heaviest neutron stars and lightest black holes exists, it will provide an additional feature with which to inform the measurement, and will be the most informative feature for spectral siren measurements with next-generation detectors (J. M. Ezquiaga & D. E. Holz 2022).

After passing the simulated events through projected detector selection effects, the resulting catalog has 591 events, consistent with the numbers projected for O5 by R. W. Kiendrebeogo et al. (2023). We use the software package `GWMockCat` (A. M. Farah et al. 2023) to simulate posterior samples for these events with measurement uncertainties typical of those expected from O5 detectors. `GWMockCat` also simulates a set of software injections, which we use to estimate selection effects in the inference. To determine the detectability of both injections and simulated events in O5, we use the projected O5 LIGO power spectral density (B. P. Abbott et al. 2020a, 2020b) for a single detector to calculate observed signal-to-noise ratios (SNRs) $\rho_{\rm obs}$, and we consider events and injections with a single-detector SNR $\rho_{\rm obs} > 8$ to be detectable. The full procedure for this mock data generation process is described in M. Fishbach & D. E. Holz (2017), A. M. Farah et al. (2023), and R. Essick & M. Fishbach (2024).

Appendix B
Effects of Fitting for Secondary Mass

We examine the effect of only fitting the distribution of primary masses on our results. To do so, we perform two spectral siren analyses: one that includes a fit to mass ratio, and one that does not. Both analyses use the parametric POWER LAW + PEAK model for the distribution of primary masses. We model the distribution of mass ratios with a power law, similar to the majority of analyses presented in R. Abbott et al. (2021b, 2023a).

We use the same set of simulated events for both analyses. These are generated in the same way as described in Section 2.1. The inference results are shown in Figure 4. We find the posteriors on H_0 to be similar between the two cases. Additionally, the recovered mass spectra are nearly identical. We therefore conclude that fitting for mass ratio does not significantly impact our main conclusions.

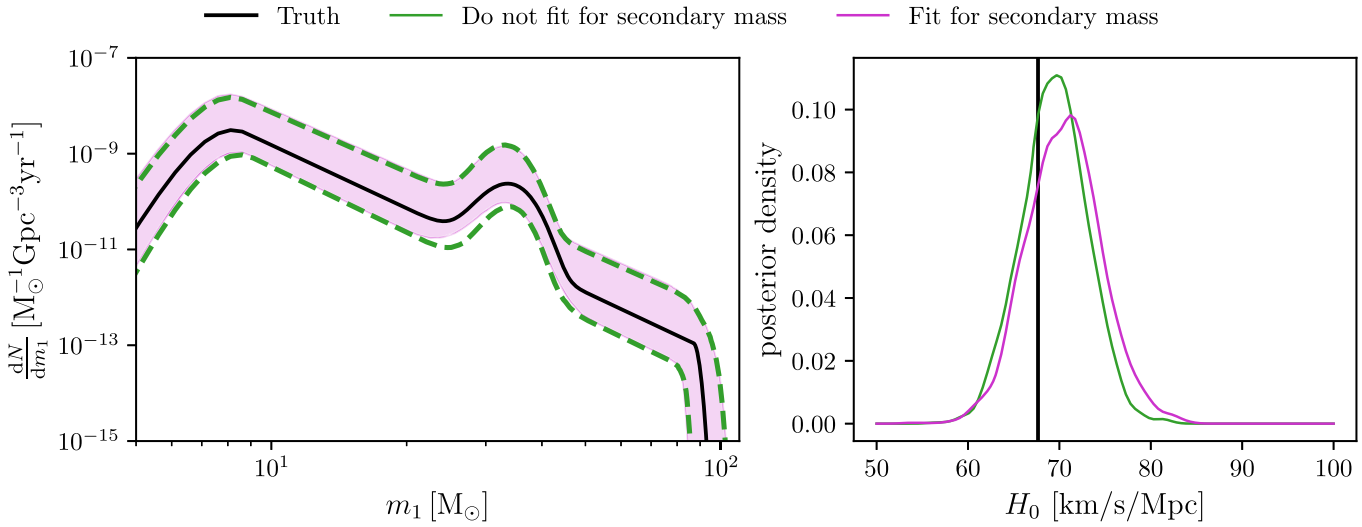

Figure 4. Comparison of a parametric spectral siren analysis performed while fitting for the distributions of primary and secondary masses (magenta solid) and primary masses only (green dashed). The posterior on H_0 is relatively unchanged between the two, and only minor differences are observed in the mass spectra.

Appendix C
Gaussian Process–based Mass Distribution

In this Section, we discuss the properties of the GP-based mass distribution and describe our modeling choices in more detail.

Practically, the difference in the inference of the population when using a GP versus other modeling choices is that the population model ($dN/dt_{\rm det}d\theta_i(\theta_i;\Lambda)$ in Equation (1)) is determined directly by a realization of the GP, rather than by a handful of hyperparameters Λ and evaluated on an analytical function. In other words, when using parametric models, $dN/dt_{\rm det}d\theta_i(\theta_i;\Lambda)$ is calculated by evaluating a specific functional form described by a small set of hyperparameters. With the GP approach, the hyperparameters describing the mass distribution are the rate at each event-level posterior sample's source-frame mass, and the rate at each found injection's source-frame mass. The GP *is* $p(\Lambda)$, the prior on population parameters (except in the case of cosmological parameters, which all have uniform priors).

Because the GP is defined only at specific data points, we have $N_{\rm ev}M + N_{\rm inj}$ mass hyperparameters, where M is the number of posterior samples per event and $N_{\rm inj}$ is the number of injections used to calculate the selection function (see, e.g., S. Vitale et al. 2020; R. Essick & M. Fishbach 2021). In this way, our GP-based mass distribution is similar to the autoregressive population models used in T. A. Callister & W. M. Farr (2024). Indeed, an autoregressive process is a GP with a specific choice of kernel.

The kernel is a function that defines the covariance between input points in the GP (in our case, two source-frame mass values). It defines the notion of similarity between adjacent points and thereby encodes our assumptions about the smoothness of the source-frame mass distribution (C. E. Rasmussen & C. K. I. Williams 2006). We use a Matérn kernel (M. S. Handcock & M. L. Stein 1993; M. L. Stein 1999) with $\nu = 5/2$, but have repeated the analysis with $\nu = 3/2$ and ∞, finding little impact on the results, except that the $\nu = \infty$ case (also called the squared exponential kernel) produces a slightly more jagged mass distribution. In addition to the mean, Matérn kernels have two parameters that determine their properties: a length scale l and a variance s. In our use case, these are one level further removed from hyperparameters, so we adopt the terminology used in T. A. Callister & W. M. Farr (2024) and call them "hyper-hyperparameters." We fit these hyper-hyperparameters along with the hyperparameters Λ to minimize prior assumptions about the form of the mass distribution. We use penalized-complexity priors on the hyper-hyperparameters to enforce that the model does not create small-scale structure uninformed by data, thereby avoiding overfitting (D. Simpson et al. 2017; D. Simpson 2022). Explicitly, the priors on l and s are Fréchet and Gamma distributions, respectively, and are defined to have less than 5% support for correlation lengths smaller than the average spacing between event-level posterior means.

The time to evaluate a GP is notorious for scaling as the cube of the number of data points, making GPs unwieldy with large data sets, such as the $\mathcal{O}(10^9)$ posterior samples and software injections expected for O5. We therefore make two approximations to a full GP to increase computational efficiency. First, for each likelihood evaluation, we evaluate a full GP on a regular grid between 0.1 $M_\odot$ and 250 $M_\odot$ and then interpolate it at each data point. Second, we use the quasi-separability of Matérn kernels to analytically perform the transformation between covariance matrix and GP draw. This second step is done using the `QuasisepSolver` module (D. Foreman-Mackey et al. 2017) in the `tinygp` code base (D. Foreman-Mackey et al. 2021), and requires data to be sortable (i.e., one-dimensional).

Algorithmically, each posterior evaluation contains the following steps:

1. Draw cosmological parameters H_0 and Ω_M from uniform prior distributions.
2. Convert the luminosity distances and detector-frame masses of each event posterior sample to redshifts and source-frame masses according to the cosmology specified by step 1.
3. Draw hyper-hyperparameters l and s from the penalized-complexity priors described above.
4. Draw a single GP realization with a kernel defined by l and s. This is defined on a regular grid of source-frame masses and evaluated using the `QuasisepSolver` in `tinygp`.
5. Interpolate the GP at each event posterior sample and injection source-frame mass (from step 3).
6. Calculate the population likelihood according to Equation (1).

We perform these steps within `numpyro` (E. Bingham et al. 2019; D. Phan et al. 2019), sampling the posterior using the no-u-turn sampler for Hamiltonian Monte Carlo (M. D. Hoffman & A. Gelman 2011). This can be seen explicitly in the source code accompanying this paper, in the `scripts/priors.py` script. At each likelihood evaluation, we calculate the effective number of parameter estimation samples and found injections. After sampling, we determine if the effective number of samples exceed the efficiency criteria detailed in R. Essick & W. Farr (2022). All results presented in this work exceed these efficiency criteria without the need to excise regions of parameter space with low numbers of effective samples.

Appendix D
Biases Induced by the BROKEN POWER LAW Model

Figure 5 shows recovered posteriors on H_0 as inferred from 50 mock catalogs using both the BROKEN POWER LAW and POWER LAW + PEAK parametric models. We find that the POWER LAW + PEAK model recovers H_0 posteriors that are largely symmetric about the true value, whereas the BROKEN POWER LAW model typically finds more support for larger H_0 values. We note that G. Pierra et al. (2024) report an underestimate of H_0 when fitting their Broken Power Law model to a single data set generated with POWER LAW + PEAK, but the prior bounds on the maximum mass parameter of their model do not extend below the injected value of that parameter, which in turn does not allow for the possibility of an overestimated H_0. It is therefore not possible to directly compare the direction of our observed H_0 offset with theirs, despite the similarities between our chosen population models.

Regardless, this systematic offset demonstrates that incorrect parametric models unfortunately induce a systematic bias in cosmological inference, not an increase in statistical uncertainty. Other parametric models may show different trends depending on which parts of the mass distribution they incorrectly model and how. Therefore, while we are able to

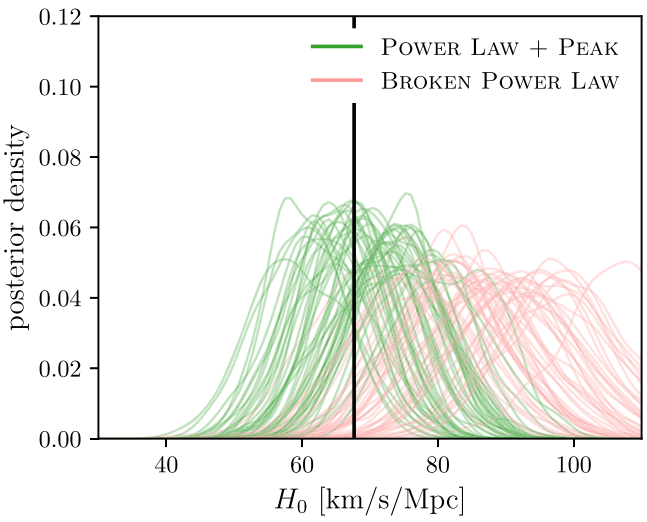

Figure 5. Recovered posteriors on H_0 from spectral siren analyses on 50 mock catalogs with the POWER LAW + PEAK (green) and BROKEN POWER LAW (pink) parametric models. The true, injected value is shown by a vertical black line.

identify the shortcomings of BROKEN POWER LAW when POWER LAW + PEAK describes the true mass distribution, it is not possible in practice to know which direction an incorrect parametric model will bias H_0 toward. Mitigating the bias by using more flexible models may thus be the simplest solution to this problem.

ORCID iDs

Amanda M. Farah https://orcid.org/0000-0002-6121-0285
Thomas A. Callister https://orcid.org/0000-0001-9892-177X
Jose María Ezquiaga https://orcid.org/0000-0002-7213-3211
Michael Zevin https://orcid.org/0000-0002-0147-0835
Daniel E. Holz https://orcid.org/0000-0002-0175-5064

References

Aasi, J., Abadie, J., Abbott, B. P., et al. 2015, CQGra, 32, 074001
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, ApJL, 848, L12
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, Natur, 551, 85
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, ApJL, 882, L24
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020a, LIGO-T2000012-v2: Noise Curves used for Simulations in the Update of the Observing Scenarios Paper, DCC, https://dcc.ligo.org/LIGO-T2000012-v2/public
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020b, LRR, 23, 3
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2021a, ApJ, 909, 218
Abbott, R., Abbott, T. D., Abraham, S., et al. 2021b, ApJL, 913, L7
Abbott, R., Abbott, T. D., Acernese, F., et al. 2023a, PhRvX, 13, 011048
Abbott, R., Abbott, T. D., Acernese, F., et al. 2023b, PhRvX, 13, 041039
Abbott, R., Abe, H., Acernese, F., et al. 2023c, ApJ, 949, 76
Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, CQGra, 32, 024001
Adachi, M., & Kasai, M. 2012, PThPh, 127, 145
Aghamousa, A., Aguilar, J., Ahlen, S., et al. 2016, arXiv:1611.00036
Akutsu, T., Ando, M., Arai, K., et al. 2021, PTEP, 2021, 05A101
Alsing, J., Silva, H. O., & Berti, E. 2018, MNRAS, 478, 1377
Anderson, L., Aubourg, E., Bailey, S., et al. 2014, MNRAS, 441, 24
Beig, R., Englert, B.-G., Frisch, U., et al. 2003, LNP, 598, 195
Bingham, E., Chen, J. P., Jankowiak, M., et al. 2019, Journal of Machine Learning Research, 20, 1
Biscoveanu, S., Talbot, C., & Vitale, S. 2022, MNRAS, 511, 4350
Callister, T., Fishbach, M., Holz, D. E., & Farr, W. M. 2020, ApJL, 896, L32
Callister, T. A., & Farr, W. M. 2024, PhRvX, 14, 021005
Chatterjee, D., Hegade, K. R. A., & Holder, G. 2021, PhRvD, 104, 083528
Chen, H.-Y., Ezquiaga, J. M., & Gupta, I. 2024, CQGra, 41, 125004
Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, Natur, 562, 545
Chen, H.-Y., Holz, D. E., Miller, J., et al. 2021, CQGra, 38, 055010
Chen, X., Li, S., & Cao, Z. 2019, MNRAS, 485, L141
Chernoff, D. F., & Finn, L. S. 1993, ApJL, 411, L5
Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, Sci, 358, 1556
Del Pozzo, W. 2012, PhRvD, 86, 043011
D'Emilio, V., Green, R., & Raymond, V. 2021, MNRAS, 508, 2090
Doctor, Z., Farr, B., Holz, D. E., & Pürrer, M. 2017, PhRvD, 96, 123011
Edelman, B., Doctor, Z., Godfrey, J., & Farr, B. 2022, ApJ, 924, 101
Edelman, B., Farr, B., & Doctor, z. 2023, ApJ, 946, 16
Essick, R. 2023, PhRvD, 108, 043011
Essick, R., & Farr, W. 2022, arXiv:2204.00461
Essick, R., & Fishbach, M. 2021, On Estimating Rates from Monte-Carlo Integrals over Injection Sets T2000100, LIGO, https://dcc.ligo.org/T2000100
Essick, R., & Fishbach, M. 2024, ApJ, 962, 169
ET steering committee 2020, Einstein Telescope: Science Case, Design Study and Feasibility Report, ET-0028A-20, Einstein Telescope, https://apps.et-gw.eu/tds/?content=3&r=17196
Evans, M., Adhikari, R. X., Afle, C., et al. 2021, arXiv:2109.09882
Ezquiaga, J. M., & Holz, D. E. 2021, ApJL, 909, L23
Ezquiaga, J. M., & Holz, D. E. 2022, PhRvL, 129, 061102
Farah, A., Fishbach, M., Essick, R., Holz, D. E., & Galaudage, S. 2022, ApJ, 931, 108
Farah, A. M., Edelman, B., Zevin, M., et al. 2023, ApJ, 955, 107
Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, ApJ, 887, 53
Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, ApJL, 883, L42
Farr, W. M., Sravan, N., Cantrell, A., et al. 2011, ApJ, 741, 103
Fishbach, M., Doctor, Z., Callister, T., et al. 2021, ApJ, 912, 98
Fishbach, M., Essick, R., & Holz, D. E. 2020, ApJL, 899, L8
Fishbach, M., Gray, R., Hernandez, I. M., et al. 2019, ApJL, 871, L13
Fishbach, M., & Holz, D. E. 2017, ApJL, 851, L25
Foreman-Mackey, D. 2021, tinygp, readthedocs, https://tinygp.readthedocs.io/en/stable/
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Freedman, W. L. 2021, ApJ, 919, 16
Fryer, C. L., & Kalogera, V. 2001, ApJ, 554, 548
Gair, J. R., Ghosh, A., Gray, R., et al. 2023, AJ, 166, 22
Giacobbo, N., Mapelli, M., & Spera, M. 2018, MNRAS, 474, 2959
Gray, R., Beirnaert, F., Karathanasis, C., et al. 2023, JCAP, 2023, 023
Gray, R., Hernandez, I. M., Qi, H., et al. 2020, PhRvD, 101, 122001
Gray, R., Messenger, C., & Veitch, J. 2022, MNRAS, 512, 1127
Handcock, M. S., & Stein, M. L. 1993, Technometrics, 35, 403
Hoffman, M. D., & Gelman, A. 2011, arXiv:1111.4246
Holz, D. E., & Hughes, S. A. 2005, ApJ, 629, 15
Huerta, E. A., Moore, C. J., Kumar, P., et al. 2018, PhRvD, 97, 024031
Kiendrebeogo, R. W., Farah, A. M., Foley, E. M., et al. 2023, ApJ, 958, 158
Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. 2019, JOSS, 4, 1143
Landry, P., & Essick, R. 2019, PhRvD, 99, 084049
Li, Y.-J., Wang, Y.-Z., Han, M.-Z., et al. 2021, ApJ, 917, 33
Loredo, T. 2009, BAAS, 213, 283
Luger, R., Bedell, M., Foreman-Mackey, D., et al. 2021, arXiv:2110.06271
Magaña Hernandez, I., & Ray, A. 2024, arXiv:2404.02522
Mandel, I., Farr, W. M., Colonna, A., et al. 2017, MNRAS, 465, 3254
Mandel, I., Farr, W. M., & Gair, J. R. 2019, MNRAS, 486, 1086
Mapelli, M. 2020, FrASS, 7, 38
Mapelli, M., & Giacobbo, N. 2018, MNRAS, 479, 4391
Marchant, P., & Bodensteiner, J. 2024, ARA&A, 62, 21
Mastrogiovanni, S., Laghi, D., Gray, R., et al. 2023a, PhRvD, 108, 042002
Mastrogiovanni, S., Leyde, K., Karathanasis, C., et al. 2021, PhRvD, 104, 062009
Mastrogiovanni, S., Pierra, G., Perriès, S., et al. 2023b, arXiv:2305.17973
Messenger, C., & Read, J. 2012, PhRvL, 108, 091101
Mukherjee, S. 2022, MNRAS, 515, 5495
Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., et al. 2019, MNRAS, 490, 3740
Özel, F., Psaltis, D., Narayan, R., & McClintock, J. E. 2010, ApJ, 725, 1918
Phan, D., Pradhan, N., & Jankowiak, M. 2019, arXiv:1912.11554
Pierra, G., Mastrogiovanni, S., Perriès, S., & Mapelli, M. 2024, PhRvD, 109, 083504
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning (Cambridge, MA: MIT Press)
Ray, A., Magaña Hernandez, I., Mohite, S., Creighton, J., & Kapadia, S. 2023, ApJ, 957, 37
Rinaldi, S., & Del Pozzo, W. 2022, MNRAS, 509, 5454
Sadiq, J., Dent, T., & Wysocki, D. 2022, PhRvD, 105, 123014
Schutz, B. F. 1986, Natur, 323, 310
Simpson, D. 2022, Priors Part 4: Specifying Priors That Appropriately Penalise Complexity, https://dansblog.netlify.app/2022-08-29-priors4/2022-08-29-priors4.html
Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. 2017, StaSc, 32, 1
Soares-Santos, M., Palmese, A., Hartley, W., et al. 2019, ApJL, 876, L7
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
Stein, M. L. 1999, Interpolation of Spatial Data (Berlin: Springer)
Talbot, C., & Thrane, E. 2018, ApJ, 856, 173
Tanvir, N. R., Levan, A. J., González-Fernández, C., et al. 2017, ApJL, 848, L27
Taylor, S. R., Gair, J. R., & Mandel, I. 2012, PhRvD, 85, 023535
Tiwari, V. 2021, CQGra, 38, 155007
Torniamenti, S., Mapelli, M., Périgois, C., et al. 2024, A&A, 688, A148
Valenti, S., Sand, D. J., Yang, S., et al. 2017, ApJL, 848, L24
van Son, L. A. C., de Mink, S. E., Callister, T., et al. 2022, ApJ, 931, 17
van Son, L. A. C., de Mink, S. E., Chruślińska, M., et al. 2023, ApJ, 948, 105
Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2020, in Handbook of Gravitational Wave Astronomy, ed. C. Bambi, S. Katsanevas, & K. D. Kokkotas (Berlin: Springer)
Ye, C. S., & Fishbach, M. 2024, ApJ, 967, 62
You, Z.-Q., Zhu, X.-J., Ashton, G., Thrane, E., & Zhu, Z.-H. 2021, ApJ, 908, 215
Zevin, M., Bavera, S. S., Berry, C. P. L., et al. 2021, ApJ, 910, 152
Zevin, M., Pankow, C., Rodriguez, C. L., et al. 2017, ApJ, 846, 82