Draft version October 10, 2023
Typeset using LaTeX default style in AASTeX631

Nonparametric Spectral Siren Cosmology

@afarah18

ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque, morbi tristique senectus et netus et. Mauris ut leo, cras viverra metus rhoncus sem, nulla et lectus vestibulum. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus, morbi dolor nulla, malesuada eu, pulvinar at, mollis ac. Curabitur auctor semper nulla donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1. INTRODUCTION

Like light, gravitational waves (GWs) are redshifted as they propagate across the universe. Unlike light, however, the selection effects of GWs are extremely well understood, allowing for a precise estimate of each catalogs' completeness and an unbiased measurement of the true GW population. This allows GW signals to be used as relatively clean probes of cosmological parameters (e.g. **?????**).

GW signals provide direct measurements of each source's luminosity distance and redshifted, or detector frame, masses, $m_{\rm det} = m_{\rm source}(1+z)$. Therefore, if the source frame mass is known, each event provides a direct mapping between luminosity distance and redshift, allowing for a measurement of the rate of expansion of the universe at the time the GW signal was emitted, $H(z)$. In practice, the source frame mass is unknown, but it is possible to know the location of features in the source frame mass distribution. The whole mass distribution therefore acts similarly to an electromagnetic spectrum, where the location of spectral features provides a redshift measurement. The method of using the mass distribution of GW sources to measure cosmological parameters has therefore been coined "spectral sirens" (Ezquiaga & Holz 2022). Spectral sirens were first demonstrated to be a feasible method to measure the Hubble constant by **?** using the BNS mass distribution, and extended to the BBH mass distribution by Farr et al. (2019).

Perhaps because of the analogy to electromagnetic spectra in which the location of the source frame spectral features is well-known, or because of the examples given in Farr et al. (2019) and **?**, it is commonly presumed that the exact shape of the mass distribution must be known *a priori* to do this measurement, from e.g. theoretical expectations of the location of a pulsational-pair instability pileup, upper mass gap, or maximum neutron star mass. However, the information in the spectral siren measurement comes solely from the assumption that, at a given luminosity distance, all CBC mergers follow the same mass distribution. Therefore the location, shape, and existence of all features in the mass distribution can be inferred simultaneously with $H(z)$.

In this work, we explicitly demonstrate that no *a priori* knowledge about the shape of the CBC mass spectrum is necessary to use the spectral siren methodology by inferring $H_0$with a non-parametric model for the mass distribution of CBCs. This flexible model makes minimal prior assumptions for the shape of the mass distribution and is therefore capable of accurately inferring a wide range of mass spectrum morphologies. Despite its flexibility, it is able to consistently obtaining an unbiased measurement for H_0, 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.

2. MOTIVATING EXAMPLE

Maybe this should just be part of results. Like just do this for O5 data and show how bad it can be. if its not that bad, we can do it for even more events to show when it will be a problem.

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. To demonstrate this, we generate a

catalog of 1,000 GW events from an underlying mass distribution described by the POWER LAW + PEAK model and its maximum *a posteriori* hyperparameters from **?**. We then use this catalog to infer H_0 using three different models for the source-frame mass distribution:

1. the POWER LAW + PEAK model, which includes the true mass distribution within its hyperprior
2. a single power law, which does not include the true mass distribution within its hyperprior, and
3. the flexible, Gaussian process-based model that we use for the remainder of this work, which can closely approximate the morphology of the true mass distribution, along with several other morphologies.

The results of each of these fits are shown in Figure **??**. The top/left panel shows the inferred source frame mass distribution for each of the considered models, and the bottom/right panel shows the corresponding posteriors on H_0.

With 1,000 events, we find that using the incorrect parametric form of the mass distribution (a power law) causes a biased inferred value of H_0 at the X-σ level, whereas using the correct functional form (POWER LAW + PEAK) provides an unbiased inference of H_0. However, this is not due to the need to know the morphology of the mass distribution *a priori*, as the Gaussian process-based model recovers the correct value of H_0 despite making minimal assumptions about the mass distribution. The Gaussian process-based model does, however, obtain larger uncertainties on the inferred value of H_0, as it has many more free parameters. We have repeated this analysis with 50 separate simulated catalogs, and found that the power law model causes us to infer a higher/lower value of H_0 at greater than X-σ, Y% of the time, whereas the Gaussian process-based model reaches that level of bias only Z% of the time, and the POWER LAW + PEAK model reaches the same level of bias A% of the time. This indicates that prior knowledge of the shape of the mass distribution is not required to perform a spectral siren measurement, so long as strong assumptions about the shape of the mass distribution are not made.

3. METHODS

3.1. *Data simulation*

We generate a mock catalog of GW events characteristic of that expected from the LVK's fifth observing run (O5). These events are drawn from an underlying population described by

$$\frac{\mathrm{d}N}{\mathrm{d}m_1\mathrm{d}m_2\mathrm{d}z} \propto p(m_1, m_2|\bar{\Lambda}_m)p(z|H_0, \Omega_M), \tag{1}$$

where

$$p(m_1, m_2|\bar{\Lambda}_m) \propto \tag{2}$$

$$p(z|H_0, \Omega_M) \propto \frac{\mathrm{d}V_C}{\mathrm{d}z}\frac{1}{1+z}, \tag{3}$$

, $V_C(H_0, \Omega_M)$ is the comoving volume for given cosmological parameters H_0 and Ω_M, and $\bar{\Lambda_m} = \alpha$... are the parameters describing the power law in primary mass, ... respectively. When generating the simulated events, we have fixed α =... and used cosmological parameters H_0= 68.7, Ω_M= 0.3, $\Omega_\Lambda = 1 - \Omega_M$. These choices correspond to the maximum *a posteriori* values obtained by an analysis of GWTC-3 data using the _ model **?**, which is morphologically similar to that described in Equations 1–3. We neglect spins entirely, as they do not carry cosmological information. TODO: fill in details of mass distribuiton. Should we use PDB and include BNSs, which will give us a nice sharp feature to work with, or should we use PLP which has a bump? Could also do PLP and then put a uniform dist or power law for NS masses at the LMG, and leave an empty mass gap. Would need to get relative heights correct though.

After passing the simulated events through projected detector selection effects, the resulting catalog has X events, including X binary neutron star systems, Y NSBH systems, and Z BBH systems, consistent with the numbers projected for O5 by **?**. We use the software package `GWMockCat` (**?**) 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 (**?**) to calculate observed signal-to-noise ratios ρ_{obs}, and we consider events and injections with $\rho_{\mathrm{obs}} > 8$ to be detectable. The full procedure for this mock data generation process is described in **???**, and the exact settings used for `GWMockCat` are made available in the accompanying data release. TODO:put a showyourwork link here.

3.2. *Spectral sirens and statistical framework*

1. Briefly explain HBA with selection effects.

$$p(\Lambda|data) = p(\Lambda)\exp -Nexp\Pi L_i \tag{4}$$

We infer this posterior using the no-u-turn sampler for Hamiltonian Monte Carlo within `numpyro` (Hoffman & Gelman 2011; **?**).

The spectral siren method of measuring cosmological parameters relies on the assumption that all compact binaries follow the same mass distribution, regardless of redshift, or if the mass distribution evolves with redshift in any way besides a monotonic rightwards shift. The likelihood in Equation 4 is maximized when the relationship between redshift and luminosity distance makes all source frame masses to follow the same distribution[1]. This is the basis of the spectral siren method of measuring cosmological parameters; we can simultaneously infer the relationship between redshift and luminosity distance with the parameters of the source frame mass distribution. The full set of hyper-parameters therefore includes the cosmological parameters that dictate the D_L-z relation: H_0, Ω_M, Ω_Λ, Ω_r, and 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. Algorithmically, this is equivalent to drawing a value of H_0, using it to define a relationship between luminosity distance and redshift, then transforming detector frame masses to source frame masses according to $m_{\rm det}(1 + z(D_L)) = m_{\rm source}$, evaluating the source-frame mass distribution at these transformed values, and then evaluating the likelihood in Equation 4. The full process is outlined in Section 3.3.

3.3. *Gaussian process-based mass distribution*

Gaussian processes (GPs) are random processes for which any linear combination of outcomes are Gaussian distributed (Rasmussen & Williams 2006). Their smoothness properties make them widely useful in GW data analysis and beyond for regression problems, such as modeling time-domain waveforms (**??**), density estimation problems, such as estimating posterior densities of single-event parameters from parameter estimation samples (**?**), and as a prior on histogram bin heights (**?**Li et al. 2021). Our use case is slightly different. We will utilize a GP as a prior on the functions that describe the primary mass distribution of CBCs. This choice adds very little prior information about the shape of the mass distribution, besides enforcing that it must be smooth. Practically, the only difference in the inference of the population when using a GP versus other modeling choices is that $p(m_1|\Lambda)$ in Equation 4 is determined directly by a realization of the GP, rather than by a handful of hyper-parameters Λ and evaluated on an analytical function.

In other words, when using parametric models, $p(\theta|\Lambda)$ in Equation **??** is calculated by evaluating a specific functional form described by a small set of hyper-parameters, Λ. With the GP approach, the hyper-parameters 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. We therefore have $N_{\rm evs}M + N_{\rm inj}$ mass hyper-parameters, where $N_{\rm ev}$ is the number of events, M is the number of PE samples per event, and $N_{\rm inj}$ is the number of injections used to calculate the selection function ξ. We label these $\{\mathcal{R}_i j, \mathcal{R}_k\}$, and our full set of hyper-parameters is $\Lambda = \{H_0, \Omega_M, \mathcal{R}_i j, \mathcal{R}_k\}$. In this way, our GP-based mass distribution is similar to the autoregressive population models used in **?**. Indeed, an autoregressive model is a Gaussian process with a particular choice of kernel.

A 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 (Rasmussen & Williams 2006). We use a Matérn kernel (**??**) 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. Matérn kernels have two parameters besides the mean variance that determine their properties: a length scale l and variance s. In our use case, these are one level further removed from hyper-parameters, so we adopt the terminology used in **?** and call them "hyper-hyper-parameters." We fit these hyper-hyper-parameters along with the hyper-parameters Λ to minimize prior assumptions about the form of the mass distribution. We use "penalized-complexity" priors on the hyper-hyper-parameters to enforce that the model does not create small-scale structure uninformed by data, thereby

[1] Ezquiaga & Holz (2022) show that the source frame masses can follow different distributions at different redshifts, so long as this redshift evolution of the mass distribution does not exactly mimic cosmology (i.e. all features cannot exhibit a monotonic rightward shift with redshift).

Model	σ_{H_0} [km/s/Mpc]	bias [km/s/Mpc]

Table 1. Performance of various mass distribution models used for measuring the Hubble constant, H_0. We consider the correct parametric model, an incorrect parametric model, and the non-parametric, Gaussian process-based model presented in this work. Bias is defined as the number of standard deviations between the true injected value of $H_0$and the recovered posterior mean. The parametric models yield more precise constraints on H_0, as shown by smaller standard deviations (σ_{H_0}), but the incorrect parametric model is less accurate than the non-parametric model. In reality, it is impossible to know the true functional form of the mass distribution, so the additional statistical uncertainty introduced by non-parametric approaches may be desirable over the systematic errors associated with choosing a parametric model.

avoiding over-fitting (Simpson et al. 2017; Simpson 2022). Explicitly, the priors on l and s have less than 5% support for correlation lengths smaller than the minimum spacing between event-level posterior samples.

GP evaluations notoriously scale as the cube of the number of data points, making them unwieldy with large data sets such as the ones 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 $0\,M_\odot$ 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 (Foreman-Mackey et al. 2017) in the `tinygp` code base, and requires data to be sort-able (i.e. one-dimensional).

Algorithmically, each hyper-likelihood evaluation contains the following steps:

1. Draw cosmological parameters H_0 and Ω_M from the prior distributions described in Equation **??**.
2. Transform the luminosity distances and detector frame masses of each event posterior sample and injection into redshifts and source frame masses according to the cosmology specified by step 1.
3. Draw hyper-hyper-parameters 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`.
5. Interpolate the GP at each event posterior sample and injection source frame mass (from step 2)
6. Calculate the population likelihood according to Equation 4.

4. RESULTS

We fit the non-parametric population model described in Section 3.3 to a simulated GW catalog of 1,000 GW events from an underlying mass distribution described by the Power Law + Peak model and its maximum *a posteriori* hyperparameters from **?**.

The resulting inferred mass distribution is shown in Figure 1, along with the corresponding posterior on H_0. The inferred mass distribution (left panel) closely resembles the true, simulated distribution, and the injected value of $H_0$is consistent with its inferred posterior distribution (right panel). Additionally, the posterior on $H_0$is distinct from its prior distribution, indicating that the data is informative despite the flexibility of the population model. This measurement represents

We repeat this analysis with two parametric models: the Power Law + Peak model, which includes the true mass distribution within its hyperprior, as well as a single power law, which does not include the true mass distribution within its hyperprior. The results of each of these fits is shown in Figure **??**, and their performance is summarized in Table 1. When using a parametric model of the same form as the true simulated distribution (Power Law + Peak), we recover the injected value of H_0 within X-σ. However, when using a model that cannot accurately represent the true shape of the mass distribution, we recover a biased estimate of H_0: the true value is offset from the mean of the posterior by X-σ.

- Can get constraints on H_0 to X% by the end of O5
- If possible, could do constraints on H0 as a function of # of events.
- could also plot H(z) using posterior draws from H_0 and Ω_M- can see at which redshift you get the best constraints. Then plot posterior of $H(z = z_{\text{best measured}})$ rather than H_0

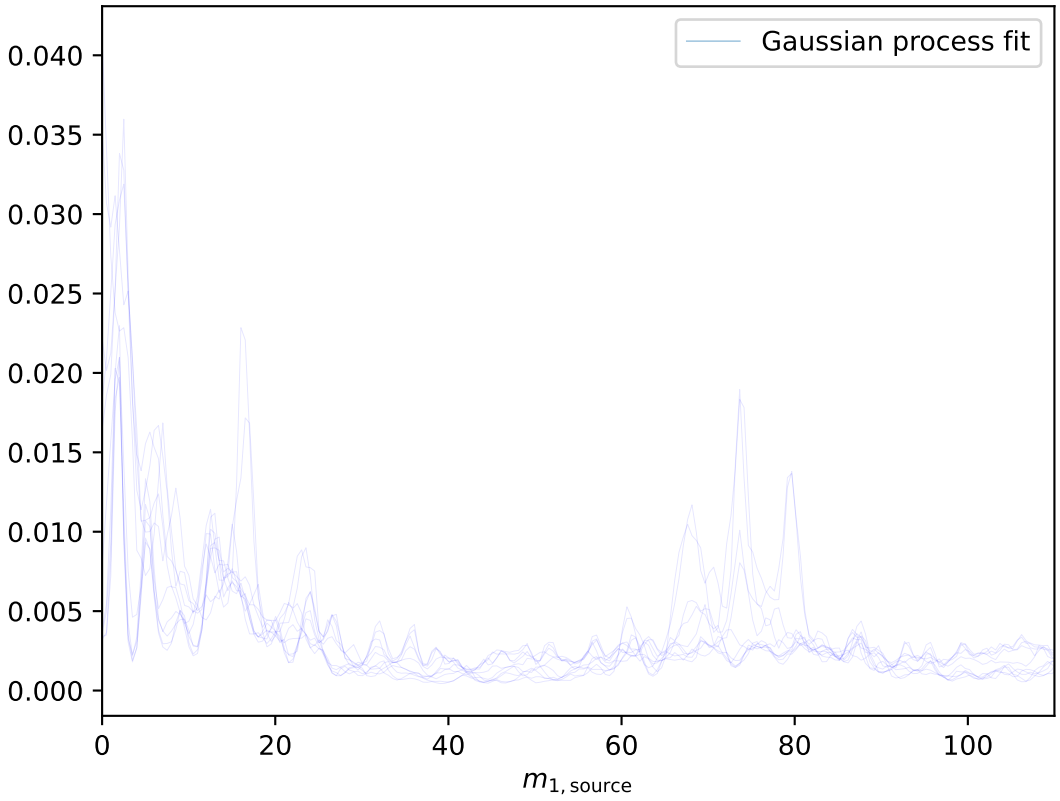

Figure 1. Two-panel figure of mass distribution and H_0 posterior. If we end up fitting Ω_M we can put a corner plot instead.

5. DISCUSSION

- We are able to get constraints on comsological parameters with a very flexible model for the mass distribution.
- The true population of compact objects in the Universe is unknown, so we will never have access to a measurement of H_0 made with a model of the same form as the true mass distribution when using real data. Systematic uncertainties arising from an incorrect choice for the functional form of the mass distribution are therefore inevitable. With current detectors, these systematic effects are dwarfed by statistical uncertainties. However, next-generation detectors will herald in enough GW observations to substantially decrease statistical uncertainty in these measurements. Non-parametric approaches such as the ones presented here will aid in mitigating systematic effects.
- hopefully these will be less biased than using a parametric model
- Also, this shows that you don't need to know the shape of the mass distribution a priori to do spectral sirens.
- This is because the information in spectral sirens comes from assuming events come from the same overall mass distribution across redshift, and that mass distribution doesn't evolve with redshift in the same exact way as would be mimicked by cosmological redshifting. This assumption is very basic, not an assumption that we know the exact astrophyisics of BBH formation (i.e. PPISN feature location)
- future/in prep. work will do the same thing but for a mass distribution that is allowed to evolve with redshift, similarly to the parametric analysis done in Ezquiaga & Holz (2022), but with the evolution with redshift allowed to be arbitrary so long as it does not mimic cosmology. - GPs are good bc they can do correlations/generalize

to multiple dimensions naturally. Our choice of kernel allows us to do this, but its not possible with the autoregressive kernel.

The authors thank Reed Essick, Utkarsh Mali, Ben Farr, Maya Fishbach, for helpful conversations.

REFERENCES

Ezquiaga, J. M., & Holz, D. E. 2022, Physical Review Letters, 129, 061102, doi: 10.1103/PhysRevLett.129.061102

Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, The Astrophysical Journal, 883, L42, doi: 10.3847/2041-8213/ab4284

Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, The Astronomical Journal, 154, 220, doi: 10.3847/1538-3881/aa9332

Hoffman, M. D., & Gelman, A. 2011, The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, arXiv. http://arxiv.org/abs/1111.4246

Li, Y.-J., Wang, Y.-Z., Han, M.-Z., et al. 2021, The Astrophysical Journal, 917, 33, doi: 10.3847/1538-4357/ac0971

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian processes for machine learning, Adaptive computation and machine learning (Cambridge, Mass: MIT Press)

Simpson, D. 2022, Un garçon pas comme les autres (Bayes) - Priors part 4: Specifying priors that appropriately penalise complexity. https://dansblog.netlify.app/posts/2022-08-29-priors4/priors4.html#the-dream-pc-priors-in-practice

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. 2017, Statistical Science, 32, 1, doi: 10.1214/16-STS576

APPENDIX

A. DETAILS OF DATA SIMULATION