

# Ain't No Mountain High Enough: Semiparametric Modeling of LIGO–Virgo's Binary Black Hole Mass Distribution

Bruce Edelman<sup>1</sup> , Zohey Doctor<sup>1,2</sup> , Jaxen Godfrey<sup>1</sup> , and Ben Farr<sup>1</sup> 

<sup>1</sup> Institute for Fundamental Science, Department of Physics, University of Oregon, Eugene, OR 97403, USA; [bedelman@uoregon.edu](mailto:bedelman@uoregon.edu)

<sup>2</sup> Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Department of Physics and Astronomy, Northwestern University, Evanston, IL 60201, USA

Received 2021 September 13; revised 2021 November 2; accepted 2021 November 2; published 2022 January 17

## Abstract

We introduce a semiparametric model for the primary mass distribution of binary black holes (BBHs) observed with gravitational waves (GWs) that applies a cubic-spline perturbation to a power law. We apply this model to the 46 BBHs included in the second gravitational-wave transient catalog (GWTC-2). The spline perturbation model recovers a consistent primary mass distribution with previous results, corroborating the existence of a peak at  $35 M_{\odot}$  (>97% credibility) found with the POWERLAW+PEAK model. The peak could be the result of pulsational pair-instability supernovae. The spline perturbation model finds potential signs of additional features in the primary mass distribution at lower masses similar to those previously reported by Tiwari and Fairhurst. However, with fluctuations due to small-number statistics, the simpler POWERLAW+PEAK and BROKENPOWERLAW models are both still perfectly consistent with observations. Our semiparametric approach serves as a way to bridge the gap between parametric and nonparametric models to more accurately measure the BBH mass distribution. With larger catalogs we will be able to use this model to resolve possible additional features that could be used to perform cosmological measurements and will build on our understanding of BBH formation, stellar evolution, and nuclear astrophysics.

*Unified Astronomy Thesaurus concepts:* [Black holes \(162\)](#); [High energy astrophysics \(739\)](#); [Gravitational wave astronomy \(675\)](#); [Gravitational wave sources \(677\)](#); [Gravitational waves \(678\)](#); [LIGO \(920\)](#)

## 1. Introduction

The LIGO–Virgo Collaboration's second catalog of compact object mergers has shown that the universe is teeming with colliding compact objects with a variety of masses and spins (Abbott et al. 2016). In contrast to the 11 sources reported in the first LIGO–Virgo Collaboration (LVC) catalog (GWTC-1 Abbott et al. 2019a), the second catalog (GWTC-2 Abbott et al. 2021a) contains 50 sources, enabling a deeper look into the formation environments of compact object binaries. The sources in GWTC-2 include two binary neutron stars (BNSs; Abbott et al. 2017, 2020a), 46 binary black holes (BBHs), and 2 neutron star black hole (NSBHs) candidates (Abbott et al. 2020d). The 46 confirmed BBHs observed in GWTC-2 include the first clear evidence of an asymmetric mass binary, potentially the least massive black hole known, and the most massive stellar mass black hole to date (Abbott et al. 2020e, 2020d, 2020f, 2020g). With this large catalog of BBH mergers, one can now begin to robustly infer the properties of the astrophysical BBH distribution in addition to each individual event's properties (Abbott et al. 2019b, 2021b).

Prior to the release of GWTC-2, the inferred mass distribution for the more massive (primary) components in mergers was thought to be consistent with a declining power law that cuts off at  $\sim 45 M_{\odot}$  (Fishbach & Holz 2017; Abbott et al. 2019b). When analyzing the BBH primary mass distribution including events in GWTC-2, Abbott et al. (2021b) found that a truncated power law is no longer consistent with the additional observations. The primary mass

distribution was found to have some feature at  $\sim 35\text{--}40 M_{\odot}$ , which was best described by either a break to a steeper power law or a power law with the addition of a peak. The presence of a peak in the primary mass distribution in this mass range is not surprising; it would be expected if we are witnessing effects of pulsational pair-instability supernovae (PPISNe; Talbot & Thrane 2018). Massive stars that are too light to be fully disrupted by a pair-instability supernova (PISN) can shed large amounts of mass in a series of explosive pulsations before collapsing to a black hole (Woosley 2017, 2019; Farmer et al. 2019). This process leads to a wide range of initial stellar masses that map onto remnant black holes with masses  $30 M_{\odot} \leq m_{\text{BH}} \leq 45 M_{\odot}$  (Belczynski et al. 2016; Marchant et al. 2019; Stevenson et al. 2019). GWTC-2 also includes more massive binaries than previously observed, most notably GW190521 (Abbott et al. 2020f, 2020g). Both component black holes of GW190521 have masses that pose a challenge to the theoretical prediction that pair-instability (PI) would forbid isolated stellar evolution from creating remnant black holes with masses from  $\sim 50\text{--}125 M_{\odot}$  (Heger & Woosley 2002; Woosley et al. 2002; Heger et al. 2003; Spera & Mapelli 2017). There is some evidence that GW190521 could be a mass-gap-straddling binary or the result of other physical processes that get around the conflict with PISN theory (Farrell et al. 2021; Fishbach & Holz 2020; Gayathri et al. 2020; Sakstein et al. 2020; Secunda et al. 2020; Cruz-Osorio et al. 2021; Estellés et al. 2021; Nitz & Capano 2021; Edelman et al. 2021a). However, the presence of these high-mass-component black holes could also point toward there being a contribution to the observed population of BBHs detected by LIGO/Virgo that formed in a way that avoids PI. These formation possibilities include hierarchical mergers in dense stellar environments, relativistic accretion onto heavy BHs in active galactic nuclei



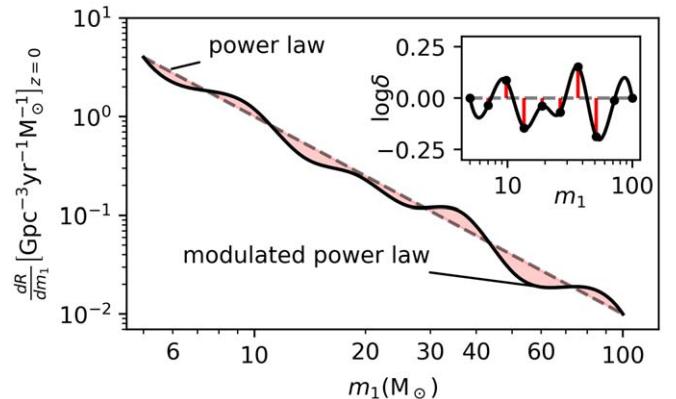
Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

disks, isolated binary evolution of low-metallicity Population II stars, or even the presence of new physics beyond the standard model (Rodriguez et al. 2019; Belczynski 2020; Croon et al. 2020; Doctor et al. 2020; Mapelli et al. 2020; McKernan et al. 2020; Kimball et al. 2020, 2021; Doctor et al. 2021).

Incorrectly inferring the BBH mass distribution has been shown to significantly bias both estimates of merger rates and the stochastic gravitational-wave background amplitude (Talbot & Thrane 2018). Additionally, the effects of PI can imprint features onto the mass distribution such as a high-mass cutoff in the mass distribution (PISN), or a possible a pileup of mergers at masses just below the cutoff (PPISN). Resolving either of these features can provide a mass scale, calibrated across cosmic time, that enables measurements of the redshift-luminosity-distance relation to infer cosmological parameters (Farr et al. 2019). As catalogs of GWs from BBHs grow in size (Abbott et al. 2020b), we will be able to infer the BBH mass distribution with greater fidelity to determine if there is the presence of additional structure beyond a power law. Such structure could yield insights about the nature of what environments BBHs form in and how they are connected to the rich fields of stellar evolution and nuclear astrophysics (Zevin et al. 2017; Farmer et al. 2019, 2020; Ng et al. 2021).

Bayesian nonparametric models provide a useful data-oriented approach to modeling when one has little information or prior knowledge about the structure of a set of data. These approaches provide little to no constraints on the functional form imposed by the model and instead use very flexible functions that have large prior support for a wide variety of unknown densities. Nonparametric modeling has been widely applied across different areas of GW astrophysics, including modeling deviations from GR waveform models, modeling the noise power spectrum of detectors, modeling the calibration of the detectors, and creating surrogate models for faster waveform execution (Littenberg & Cornish 2015; Doctor et al. 2017; Edelman et al. 2021b; Vitale et al. 2021b).

In this work, we approach the mass spectrum from a data-driven perspective, using a semiparametric method rather than the low-dimensional parametric models used in Abbott et al. (2019b, 2021b). Our semiparametric method is complementary to both parametric and fully nonparametric approaches (Mandel et al. 2016; Tiwari 2021) by incorporating a simple parametric description of the large-scale structure (i.e., a power law) with an additional nonparametrically modeled component on top. This approach can aid in searching for generic deviations to the underlying parametric descriptions that could be the result of astrophysical processes. Since nonparametric approaches make few assumptions on the form that the underlying distributions may take, our model minimizes biases to the structure such deviations could take. We expect a large fraction of stellar-mass BHs to form at the end of life of massive stars, which motivates our choice of a power-law form of the BBH primary mass distribution following a similar functional form to the stellar initial mass function (Kroupa 2001). We therefore reconstruct the primary mass distribution with the TRUNCATED power-law model (Fishbach & Holz 2017; Abbott et al. 2021b), in which we modulate with a nonparametric perturbation. This method takes advantage of using a simple parametric form to capture the majority of the structure in the primary mass distribution while the perturbation function can find data-informed deviations from the power law. In Section 2 we describe our semiparametric perturbation population model, and in Section 3 we present and



**Figure 1.** Sketch description of the spline perturbation primary mass model. The inset shows the interpolated cubic spline perturbation function for the plotted modulated power law.

discuss the inferred properties of the primary mass distribution when analyzing all 46 BBHs in GWTC-2. Finally, we explore possible interpretations of our results and conclude in Section 4.

## 2. Spline Perturbation Model

We use a hierarchical Bayesian inference framework to infer the properties of the astrophysical distribution of BBHs that incorporates software injections to estimate selection effects (Farr 2019; Mandel et al. 2019; Vitale et al. 2021a). This procedure is described in detail in Appendix A. In order to capture both the overall trends and any sharper features that may be in the primary mass distribution, we modulate a base parametric hyperprior on primary mass,  $p(m_1|\Lambda)$ , by a highly flexible perturbation function—in this case, a cubic spline. We choose the simplest of previously used parametric models as our underlying mass distribution,  $p(m_1|\Lambda)$ , which is described by a power law in both primary mass and mass ratio with a sharp low- and high-mass cutoff (Fishbach & Holz 2017; Abbott et al. 2019b, 2021b). This model was referred to as the TRUNCATED model in Abbott et al. (2021b). While the TRUNCATED model alone does not describe GWTC-2 well (Abbott et al. 2021b), it captures the majority of the large-scale structure found in the primary mass distribution. For our underlying description, we extend the TRUNCATED model to allow for a tapering of the distribution at low masses following the same form used for the POWERLAW+PEAK model described in Talbot & Thrane (2018) and Abbott et al. (2021b). Figure 1 shows an illustration of our spline perturbation model on top of a power law without any mass cutoffs or tapering. We multiplicatively apply perturbations to the underlying distribution as:

$$p_{\text{spline}}(m_1|\Lambda, \{m_i, f_i\}) = k * p(m_1|\Lambda) \exp(f(m_1; \{m_i, f_i\})) \quad (1)$$

In the above equation,  $k$  is a normalization factor found by numerically integrating  $p_{\text{spline}}(m_1|\Lambda, \{m_i, f_i\})$  over the entire range of primary masses, and  $f(m_1; \{m_i, f_i\})$  is the perturbation function modeled as a cubic spline that is interpolated between  $n$  knots placed in  $m_1$  space. These knots are denoted by their locations in mass space,  $\{m_i\}_{i=1}^n$ , and their heights at each knot,  $\{f_i\}_{i=1}^n$ . For readability, we hereafter drop explicit dependence of  $f$  on  $\{m_i, f_i\}$  unless needed.

We fix the locations of each knot to be linear in  $\log m_1$  space from  $5-100 M_\odot$  and restrict the perturbations to zero at the minimum and maximum knots. With these restrictions our spline

model adds  $n - 2$  extra hyperparameters to the underlying primary mass model we are perturbing, one for each of the inner knots' heights. We log-space the knots and perturb our underlying model with the multiplicative factor,  $\exp(f(m_1))$ , to reflect the wide range in orders of magnitude of the underlying power law. We then impose Gaussian priors on the knot heights  $\{f_i\}$  centered at 0 and with standard deviations,  $\sigma_{\text{knot}}$ . Our model then has two specifications which control the resolution ( $n$ ) and the magnitude ( $\sigma_{\text{knot}}$ ) of perturbations to which the model is sensitive. We discuss the effect of changing these model settings on our prior assumptions and motivate the particular choices we made for this work in Appendix B.

In addition to the primary mass distribution, we simultaneously fit for the mass ratio and redshift distributions, without any spline perturbations applied. We apply a power-law distribution for the mass ratio as  $p(q|m_1, m_{\min}, \beta_q) \propto q^{\beta_q} \Theta(qm_1 - m_{\min}) \Theta(m_1 - qm_1)$ , with  $\Theta$  denoting the Heaviside step function that ensures  $m_2$  is within the range  $[m_{\min}, m_1]$ . We then fit for the evolution of the merger rate with redshift also with a power law such that  $p(z|\lambda) \propto \frac{dV_c}{dz} \frac{1}{1+z} (1+z)^\lambda$ , where  $dV_c$  is the co-moving volume element (Hogg 1999; Fishbach et al. 2018; Abbott et al. 2019b, 2021b). We do not fit for a population prior on the BBH spins, and assume the spin prior used for individual event parameter estimation in Abbott et al. (2021a), which is uniform in component spin magnitudes and isotropic in component spin orientations. We enumerate each of the model's hyperparameters and corresponding hyperprior distributions used in this work in Table 1.

### 3. Results

#### 3.1. Astrophysical BBH Primary Mass Distribution

We use a catalog containing each of the 46 BBH mergers reported in Abbott et al. (2021a) in which we perform a hierarchical Bayesian analysis to infer the astrophysical mass spectrum and merger rate evolution with redshift, as described in Appendix A. We perform multiple iterations of our semiparametric model with different numbers of knots and both a “conservative” ( $\sigma_{\text{knot}} = 1$ ) and “wide” ( $\sigma_{\text{knot}} = 2$ ) prior width on the knots. For both cases of prior width we additionally do a post hoc “marginalization” over the number of knots by combining posterior draws weighted according to the ratios in marginal likelihoods. Explicitly, we take  $\left\lfloor N_{\min} \frac{\mathcal{Z}_n}{\mathcal{Z}_{\max}} \right\rfloor$  samples from each inference where  $N_{\min}$  is the minimum number of samples of samples in each of the posteriors,  $\mathcal{Z}_n$  the marginal likelihood of inference with  $n$  knots, and  $\mathcal{Z}_{\max}$ , the maximum marginal likelihood of the combined posteriors. In Figure 2 we plot the posterior merger rate density as a function of primary mass (top row) for our combined spline model (combining 10, 15, and 20 knot models), compared to the POWERLAW+PEAK model. The most prominent feature in the primary mass distribution is the apparent peak at  $\sim 35 M_{\odot}$ , similar to the peak found at the same mass by the POWERLAW+PEAK model (Talbot & Thrane 2018; Abbott et al. 2021b).

In addition to the peak at  $\sim 35 M_{\odot}$ , there are signs of additional features—albeit less significant—at lower masses. We find signs of an inflated rate of mergers with primary masses  $\sim 10 M_{\odot}$  and reduced rate around  $\sim 7.5 M_{\odot}$ , when compared with the power-law structure. The model is less certain about the low-mass features as there are only a few events with support for  $m_1 < 10 M_{\odot}$ . The dearth of observed low-mass BBHs, combined

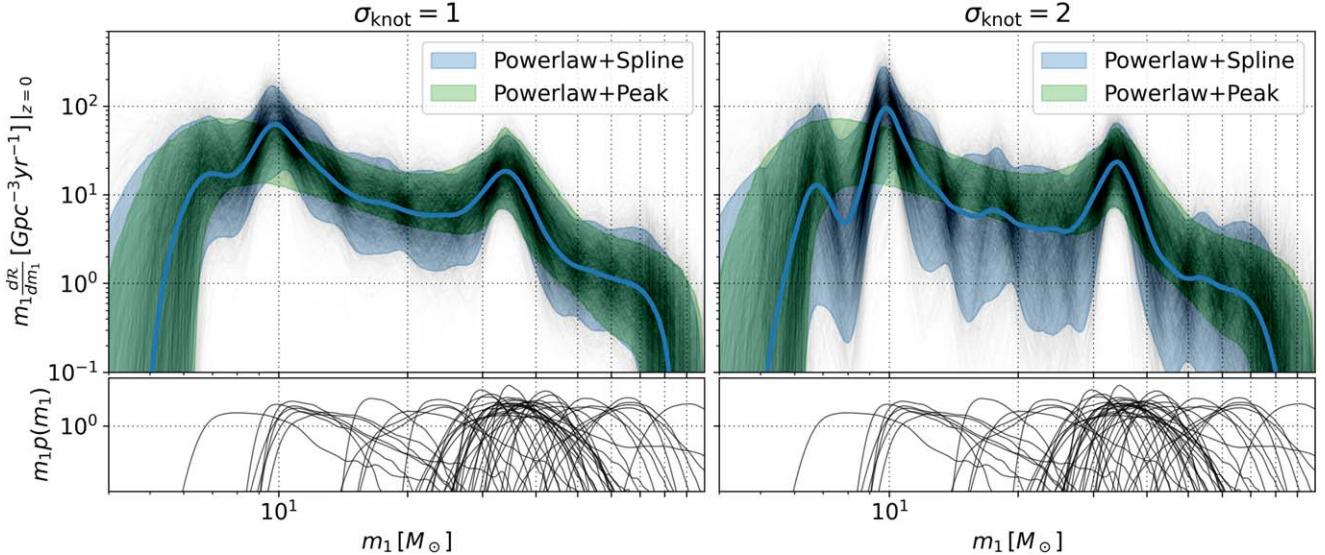
with their small sensitive volume, significantly inflates our uncertainty at the low end of the mass distribution. In Figure 3, we show the results inferred by each of the 10, 15, and 20 knot spline models individually, showing that the spline model consistently finds common features regardless of the number and location of knots. The combined spline model based on the marginal likelihoods is mostly composed of the 15 knot result because the ratios of marginal likelihoods favor 15 knots over 20 at 2:1 odds and 15 over 10 at 3:1. The spline models are best constrained in the regions of over/underdensities discussed above and much less certain (prior dominated) in regions between the features where the parametric component (i.e., power law) can fully explain the trend. The bottom row of Figure 3 shows the perturbation function,  $f(m_1)$ , inferred from the different spline models. While there are some differences between knot choices due to the different length scales, they are all in agreement when taking into account the uncertainties and each consistently recovers similar merger rates and perturbations at both the  $10 M_{\odot}$  and  $35 M_{\odot}$  peaks. In Figure 4 we plot the posterior distribution of the perturbation function sliced at the approximate masses of the three, ( $f(m_1 = 7.5 M_{\odot})$ ,  $f(m_1 = 10 M_{\odot})$  and  $f(m_1 = 35 M_{\odot})$ ). We find similar posteriors on the perturbation at these three mass regions, across the models varying the number of knots and the spline prior width. We calculate the percentile where  $f = 0$  falls in the posterior distribution for each of these three cuts, which would be near 50% in the presence of no deviations to the power law or equivalently for draws from the spline model prior. The percentiles of zero perturbation for the combined model with the conservative (wide) priors are 70.8% (72.6%), 19.4% (13.1%), and 0.8% (2.5%) at  $7.5 M_{\odot}$ ,  $10 M_{\odot}$ , and  $35 M_{\odot}$ , respectively.

The presence of the  $\sim 35 M_{\odot}$  feature was previously found and reported in Abbott et al. (2021b) as being either a peak (likely due to the PPISN pileup Talbot & Thrane 2018) or a break to a steeper power law. The POWERLAW+PEAK model returned the highest marginal likelihood of parametric mass models considered in Abbott et al. (2021b), but was only favored with roughly 3:1 odds. Due to the inherent nature of the spline perturbation model, we would be more likely to find features that look like peaks rather than a power-law break in the distribution. We additionally fit for spline perturbations on top of the BROKENPOWERLAW model, which found little to no support for two different power-law slopes and recovered a nearly identical primary mass distribution to what was found when modulating a single power law. The low mass feature recovered by our spline model was not identified in Abbott et al. (2021b) because the models considered there did not have the flexibility to fit such features at low mass: they only include a smooth rise to the low-mass end of the power law. This could be explained by additional structure that cannot be described by a smooth rise to a power law, coming from the upper edge of the proposed neutron star black hole mass gap. The flexibility of our semiparametric approach enables us to find additional structure in the astrophysical mass distribution beyond what can be found with simpler toy models. While the spline perturbation model clearly finds structure beyond the power law around  $\sim 35 M_{\odot}$  in the analyzed catalog of 46 BBH mergers, we cannot say for certain that the low mass feature is inherent to the astrophysical mass distribution. There is still the possibility that our model could be latching onto fluctuations in our data due to small-number statistics. This possibility is reflected in the percentiles in Figure 4. The perturbation function

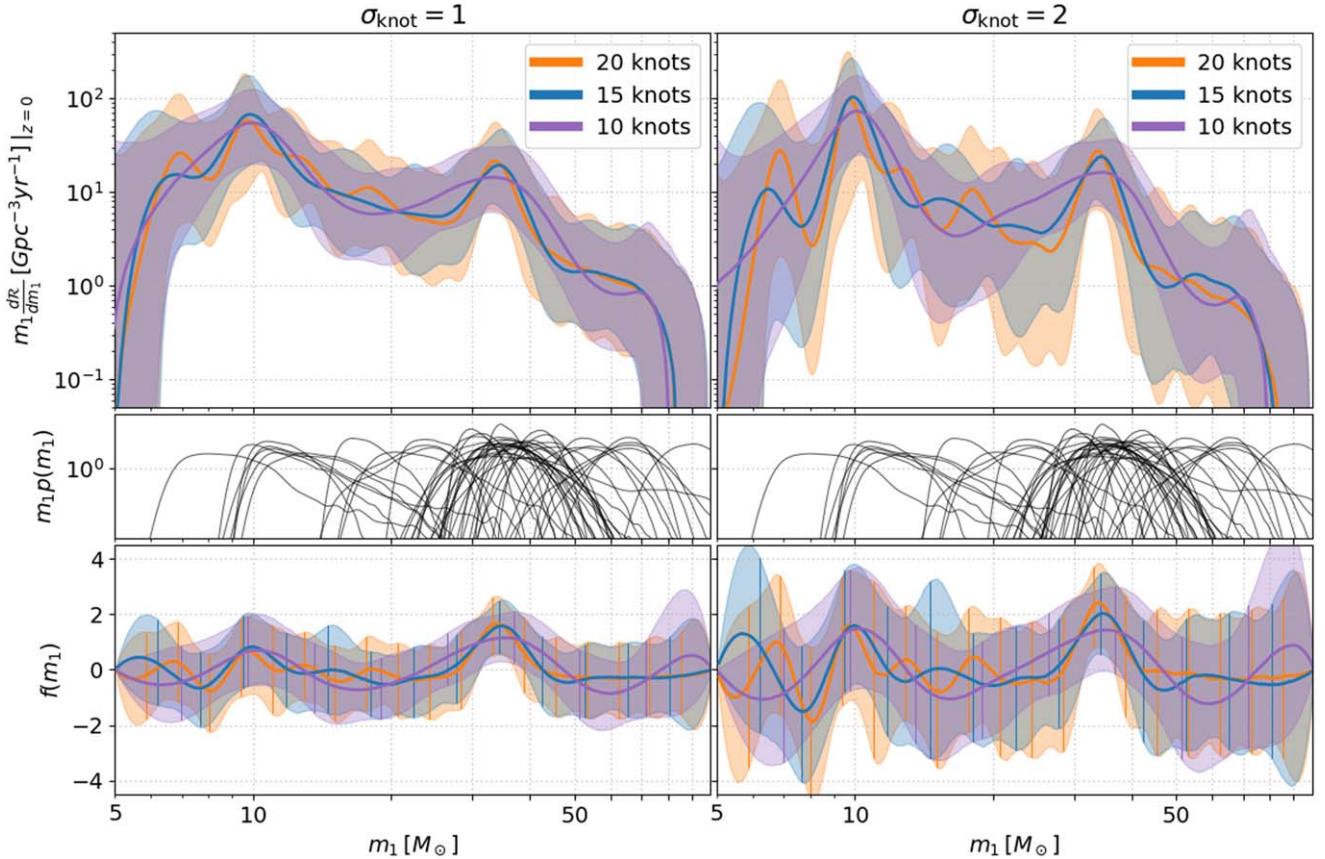
**Table 1**  
Hyperparameters, Their Descriptions, and Chosen Priors for This Work for Each Respective Population Model

Model	Parameter	Description	Prior
Primary Mass Model Parameters			
TRUNCATED	$\alpha$	slope of the power law	U(-4, 12)
	$m_{\min}$	minimum mass cutoff	U( $2 M_{\odot}$ , $10 M_{\odot}$ )
	$m_{\max}$	maximum mass cutoff	U( $35 M_{\odot}$ , $100 M_{\odot}$ )
	$\delta_m$	low-mass smoothing from $[m_{\min}, m_{\min} + \delta_m]$	U( $0 M_{\odot}$ , $10 M_{\odot}$ )
BROKENPOWERLAW	$\alpha_1$	slope of first power law	U(-4, 12)
	$\alpha_2$	slope of second power law	U(-4, 12)
	$b$	fraction between $m_{\min}$ and $m_{\max}$ where the power law break lies	U(0, 1)
	$m_{\min}$	minimum mass cutoff	U( $2 M_{\odot}$ , $10 M_{\odot}$ )
	$m_{\max}$	maximum mass cutoff	U( $50 M_{\odot}$ , $100 M_{\odot}$ )
	$\delta_m$	low-mass smoothing from $[m_{\min}, m_{\min} + \delta_m]$	U( $0 M_{\odot}$ , $10 M_{\odot}$ )
POWERLAW+PEAK	$\alpha$	slope of the power law	U(-4, 12)
	$m_{\min}$	minimum mass cutoff	U( $2 M_{\odot}$ , $10 M_{\odot}$ )
	$m_{\max}$	maximum mass cutoff	U( $30 M_{\odot}$ , $100 M_{\odot}$ )
	$\mu_p$	location in mass space where the Gaussian peak lies	U( $20 M_{\odot}$ , $70 M_{\odot}$ )
	$\sigma_p$	width of the Gaussian peak	U( $0.4 M_{\odot}$ , $10 M_{\odot}$ )
	$\lambda_p$	fraction of BBH in the Gaussian component	U(0, 1)
	$\delta_m$	low-mass smoothing from $[m_{\min}, m_{\min} + \delta_m]$	U( $0 M_{\odot}$ , $10 M_{\odot}$ )
POWERLAW+MULTIPEAK	$\alpha$	slope of the power law	U(-4, 12)
	$m_{\min}$	minimum mass cutoff	U( $2 M_{\odot}$ , $10 M_{\odot}$ )
	$m_{\max}$	maximum mass cutoff	U( $30 M_{\odot}$ , $100 M_{\odot}$ )
	$\mu_{p,1}$	location in mass space where the first Gaussian peak lies	U( $5 M_{\odot}$ , $40 M_{\odot}$ )
	$\sigma_{p,1}$	width of the first Gaussian peak	U( $0.4 M_{\odot}$ , $10 M_{\odot}$ )
	$\lambda_{p,1}$	fraction of BBH in the first Gaussian component	U(0, 1)
	$\mu_{p,2}$	location in mass space where the second Gaussian peak lies	U( $20 M_{\odot}$ , $70 M_{\odot}$ )
	$\sigma_{p,2}$	width of the second Gaussian peak	U( $0.4 M_{\odot}$ , $10 M_{\odot}$ )
	$\lambda_{p,2}$	fraction of BBH in the second Gaussian component	U(0, 1)
	$\delta_m$	low-mass smoothing from $[m_{\min}, m_{\min} + \delta_m]$	U( $0 M_{\odot}$ , $10 M_{\odot}$ )
Mass Ratio Model Parameters			
POWERLAWMASSRATIO	$\beta_q$	slope of the mass ratio power law	U(-4, 12)
Redshift Evolution Model Parameters			
POWERLAWREDSHIFT	$\lambda$	slope of redshift evolution power law $(1 + z)^{\lambda}$	U(-2, 6)
Spline Perturbation Model Parameters			
CUBIC SPLINE	$\{m_n\}$	location in primary mass of the n spline interpolant knots	FIXED
	$\{f_n\}$	y-value of the spline interpolant knots	$N(\mu = 0, \sigma = 1)$

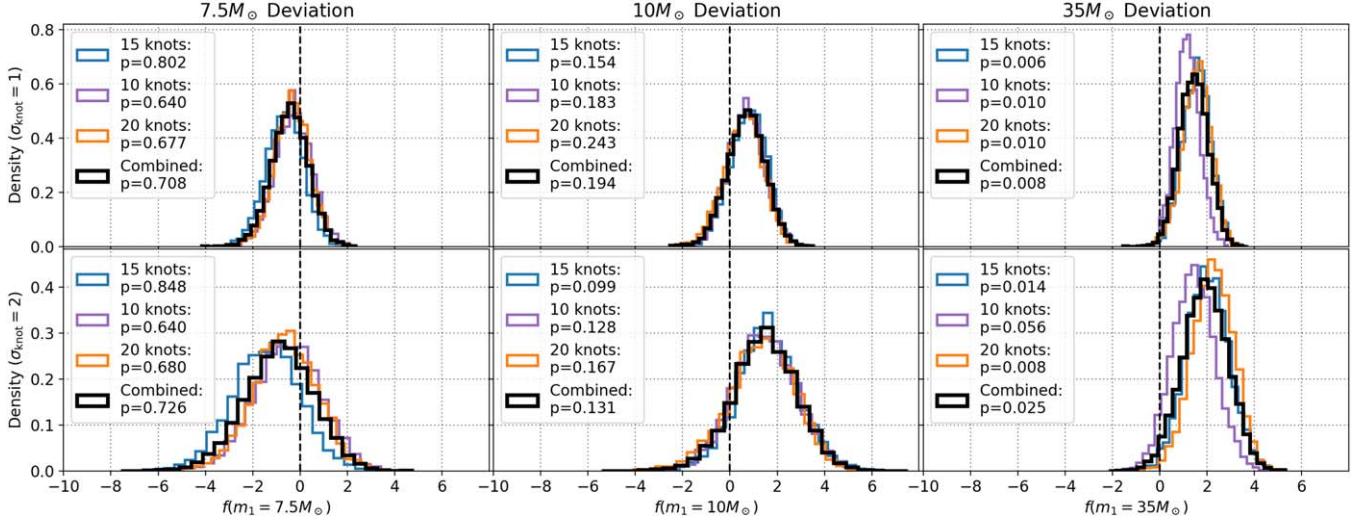
**Note.** The TRUNCATED model is extended from the version used in Abbott et al. (2021b) to have the option of a low-mass taper of the same form as the POWERLAW+PEAK model. Note that we do not describe a spin population model in this table since in this work we are not inferring a hyperprior on the spins and instead assume they are described by the default (uniform in component magnitudes, isotropic in orientations) parameter estimation prior used to producing (Abbott et al. 2021a).



**Figure 2.** We plot the differential merger rate,  $\frac{d\mathcal{R}}{d \log m_1} = m_1 \frac{d\mathcal{R}}{d m_1}$ , as a function of primary mass (top row) for the combined spline model marginalized over the 10, 15, and 20 knot models in blue. The solid line shows the median while the shaded region shows the 90% credible interval and the 90% credible interval found from the POWERLAW+PEAK model in green. The black traces show 1000 draws from the combined spline model posterior and we plot kernel density estimates (KDEs) of the posterior samples of primary source frame mass for each of the 46 BBHs in GWTC-2. We plot  $m_1 p(m_1)$  on a log-scaled y-axis with a Gaussian KDE approximating  $p(m_1)$  for each event. These posterior samples are not re-weighted to a population and come directly from the accompanying data release to Abbott et al. (2021a).



**Figure 3.** The median (solid) and 90% credible interval (shaded) of the inferred differential merger rate density as a function of primary mass (top row) with spline models with 10 (purple), 15 (blue), and 20 (orange) knots. We show the conservative prior case ( $\sigma_{\text{knot}} = 1$ ) in the left column and the wide prior case ( $\sigma_{\text{knot}} = 2$ ) on the right. The middle row shows kernel density estimates (KDEs) of the posterior samples of primary source frame mass for each of the 46 BBHs in GWTC-2. We plot  $m_1 p(m_1)$  on a log-scaled y-axis with a Gaussian KDE approximating  $p(m_1)$  for each event. These posterior samples are not re-weighted to a population and come directly from the accompanying data release to Abbott et al. (2021a). The bottom row shows the median (solid) and 90% credible intervals (shaded) of the inferred perturbation function,  $f(m_1)$ , for each choice of  $n$ , with the vertical lines showing the locations of the spline knots.



**Figure 4.** The posterior distribution of  $f(m_1)$  at sliced at the three most apparent inferred perturbations in the posterior which roughly lie at  $\sim 7.5 M_\odot$  (left column),  $\sim 10 M_\odot$  (middle column), and  $\sim 35 M_\odot$  (right column). We show the posteriors for 10 (purple), 15 (blue), and 20 (orange) knots and for both cases of prior width:  $\sigma_{\text{knot}} = 1$  (top row) and  $\sigma_{\text{knot}} = 2$  (bottom row). We additionally show the result when combining the models (weighted by their marginal likelihoods) across the three choices number of nodes in black. We report the quantile in which  $f = 0$  falls for each of the models and perturbation regions knots in each legend.

at  $7.5 M_\odot$  and  $10 M_\odot$  does not rule out  $f = 0$  at high credibility regardless of prior choices in the spline model, while in contrast, the perturbation at  $35 M_\odot$  rules out  $f(35 M_\odot) = 0$  at 97%–99.4% credibility across each variation of spline model used. We investigate the possibility that these subsequent deviations from a power law could appear due to our model’s systematics in Appendix C. We report no signs of correlations between the successive perturbations, which would be expected if the spline function was imposing biases onto our inferred perturbations. With future larger catalogs of gravitational-wave sources, we will be able to further resolve these low-mass features to determine if they are indeed present in the astrophysical mass distribution or a reflection of the current small catalog size.

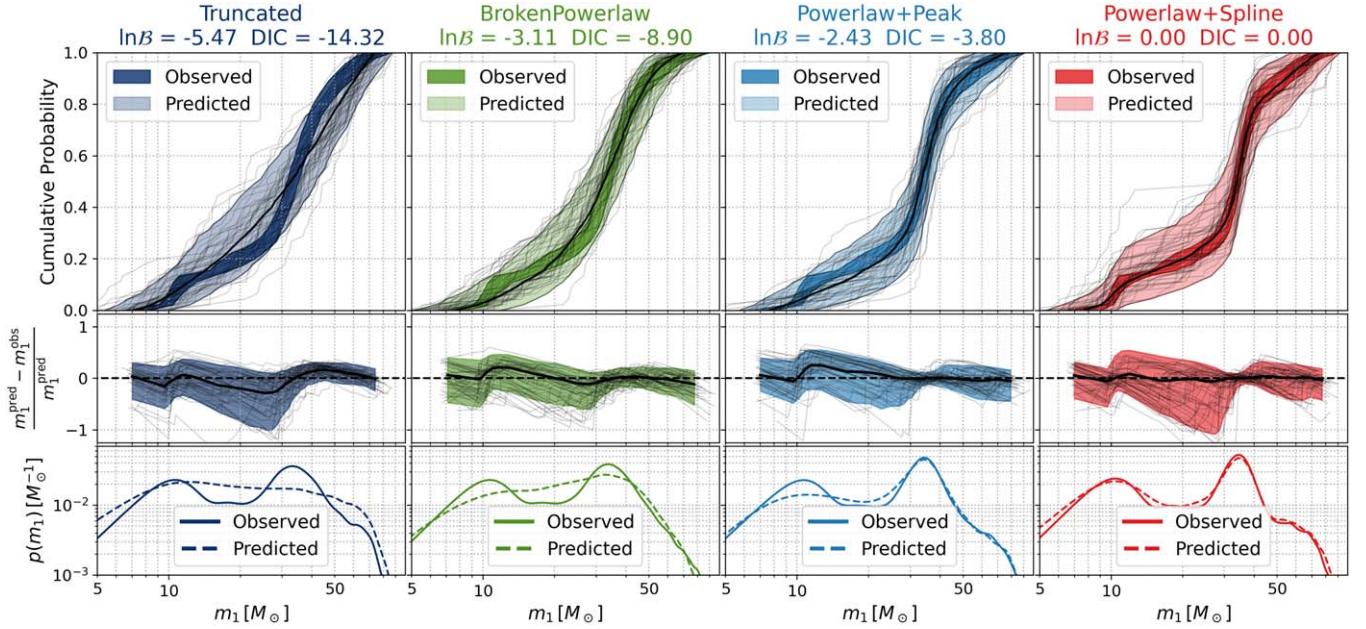
### 3.2. Posterior Predictive Checks

With the large flexibility that comes from taking nonparametric approaches to modeling, one must be careful in validating inferences, especially in cases with small numbers of observations. One way we evaluate our semiparametric model is through posterior predictive checks. We employ the same injection set used to estimate our search’s selection effects mentioned in Appendix A to create mock detected populations under a given population described by a posterior on hyperparameters. We then compare these mock populations to the observed population. To do this, we first re-weight the injections to our inferred population for  $N_{\text{draw}}$  draws of hyperparameters from our population posterior, then take  $N_{\text{obs}}$  (46 for this BBH-only analysis) draws for each of the re-weightings. This generates  $N_{\text{draw}}$  sets of  $N_{\text{obs}}$  “Predicted” observations for a given population inference. Next, we re-weight the individual event posterior samples to the same inferred population for  $N_{\text{draw}}$  draws of hyperparameters from the posterior. For each draw of hyperparameters, we take a fair draw from each re-weighted event posterior to generate our corresponding  $N_{\text{draw}}$  sets of  $N_{\text{obs}}$  “Observed” observations for a given population inference. From this procedure we generated 500 sets of 46 “Observed” and “Predicted” catalogs, which we compare to each other in Figure 5 to confirm that our inferred population model predictions are consistent with the observations. We show the

cumulative probability as a function of observed primary mass in the top row, the relative error in predicted primary masses to observed in the second row, and the last row shows the observed and predicted primary mass distributions averaged over all of the hyperparameters inferred in our posterior. The colored bands in the top row of Figure 5 show that a model is inconsistent with observations when the dark “Observed” band extends outside of the lighter “Predicted” band. We see this behavior at  $\sim 40$ – $50 M_\odot$  for the TRUNCATED model which illustrates a conclusion from Abbott et al. (2021b): the TRUNCATED model is inconsistent with the mergers in GTWC-2. The spline perturbation model is the only primary mass model that recovers both the low and high mass features seen in the observed distribution, but it does exhibit more uncertainty than other models in the regions between the  $\sim 10 M_\odot$  and  $\sim 35 M_\odot$  peak. When considering possible fluctuations due to small-number statistics, the observations at low mass are still consistent with both the POWERLAW+PEAK and BROKENPOWERLAW models.

### 3.3. Astrophysical Implications

The BBH mass distribution is particularly well suited to answering a wide range of astrophysical questions. In particular, the masses of detected events are relatively well measured, and different channels of BBH formation result in different mass distributions (e.g., Zevin et al. 2017), implying that the formation history is encoded within these distributions. With the tens of detected events available now, disentangling the overlapping subpopulations in the full population (if they exist) is a challenge. Our perturbation model can be used to see if or where a distribution describing a single (dominant) subpopulation or formation channel may fail to fully fit the data, which would provide evidence that there may be non-negligible contributions due to additional formation channels. The hints of structure we see at the low end of the mass distribution could point to such a superposition of multiple formation channels. Another factor that can affect the mass distribution is the physics of PISN or PPISN. Stellar evolution models describing mass loss and PPISN have uncertainties that can drastically change predictions on the masses at which the PPISN and PISN play a role. For example, choices in



**Figure 5.** Posterior predictive checks for three of the parametric models used in Abbott et al. (2021b) and the spline perturbation model of this work. We show the spline model result with the highest marginal likelihood which was the 15 knot and  $\sigma_{\text{knot}} = 1$  model. The observed and predicted values for primary masses are generated by re-weighting either the injection set or the set of posterior samples for each BBH analyzed, for 500 draws from each models inferred posterior on the hyperparameters, and then drawing 46 values from the re-weighted injections and a single fair draw from each of the 46 event re-weighted posterior samples. The top panel shows the CDF generated from these sets observed and predicted events for each of the 4 models, with the 90% credible levels enclosed by the bands, the median shown with thick black lines, and the thin black lines showing 50 of the 500 sets of 46 predicted events. The middle row uses the same set of predicted and observed events and the y-axis shows the relative error in predicted to observed mass ( $(m_1^{\text{pred}} - m_1^{\text{obs}})/m_1^{\text{pred}}$ ) as a function of  $m_1^{\text{obs}}$ . The last row of plots shows the PDF of the top row averaged over the 500 draws from the posterior on the hyperparameters for both sets of events.

nuclear reaction rates within stellar cores can affect the BBH mass distribution (Farmer et al. 2019, 2020). Our spline model enables us to measure these imprints of BBH formation in the observed distribution without enforcing the specific distribution shapes that are inherent to parametric models. Our findings corroborate the existence of a feature consistent with a PPISN “pileup” in the  $30\text{--}40 M_\odot$  range, and we infer its shape without assuming a simple functional form. With more GW detections, further resolution of this peak with the spline model could offer insights into supernova physics.

#### 4. Conclusions

Accurate estimation of the BBH mass distribution is paramount to getting accurate estimates of merger rates, the GW stochastic background, and false-alarm rates for potential new triggers. Low-dimensional parametric models have the advantage of being easily interpreted but are limited in their flexibility and subject to a priori expectations. We presented a semiparametric approach to modeling the primary mass distribution of BBH mergers, using cubic splines that modulate a power law. We show that our flexible semiparametric approach, when applied to the BBHs in GWTC-2, consistently recovers the previously reported excess of observed mergers near  $\sim 35 M_\odot$ , and shows potential signs of additional features in the low-mass end of the BBH distribution. These low-mass features beyond the power-law structure, correspond with similar features found in the chirp mass distribution using a separate nonparametric approach based on a flexible Gaussian mixture model (Tiwari & Fairhurst 2021). We show through posterior predictive checks that the spline model is at least as good as the POWERLAW+PEAK model at fitting the high-mass structure in our catalog while having the flexibility beyond a smooth rise into a power law to capture the apparent

excess at  $10 M_\odot$ . Structure in the mass distribution could arise from many different astrophysical phenomena but if we are able to confidently identify either a high-mass cutoff or pileup in the mass distribution, it is likely to be related to effects of PISN or PPISN. These two features can both be used as calibrated mass scales to measure a redshift–luminosity–distance relation with which it is possible to infer cosmological parameters with Farr et al. (2019).

Our semiparametric approach has advantages compared to other fully nonparametric approaches modeling the BBH mass distribution (Mandel et al. 2016; Tiwari 2021; Tiwari & Fairhurst 2021). The semiparametric approach leverages the information learned from the parametric models to explain the majority of the structure, while reserving the flexibility to see where observations may start to diverge from previous inferences. This same method of applying cubic spline perturbations to simpler population models can be used on any of the other commonly modeled population distributions such as the mass ratio, spins, or redshift evolution. With the relatively small catalog sizes currently available, structure, if present, would likely only appear in the best measured parameters. In future work, we plan to extend this method to incorporate multi-dimensional splines, that could uncover correlations between different parameters such as peaks in the mass distribution associated with a high spin magnitude. Such correlations would be a tell-tale sign of hierarchical mergers, for example (Fishbach & Holz 2017; Gerosa & Berti 2017; Doctor et al. 2020, 2021; Kimball et al. 2020). Future work could also extend this method to more than just the BBH mass distribution and allow for adaptive resolution splines that allow the knot locations to vary (Edelman et al. 2021b). With additional observations of GWs associated with BNSs and NSBHs (Abbott et al. 2021c) our spline perturbation model is well suited to modeling the joint mass distribution of all GWs.

observed compact objects. This would complement the parametric model of Fishbach et al. (2020), giving insights on the structure (or lack thereof) of the “lower mass gap” that may exist between the heaviest neutron stars and lightest black holes.

We thank Will Farr for countless useful discussions related to this work and hierarchical modeling. This research has made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Center (<https://www.gw-openscience.org/>), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. This work benefited from access to the University of Oregon high performance computer, Talapas. This material is based upon work supported in part by the National Science Foundation under Grant PHY-1807046 and work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation.

*Software:* ASTROPY (Astropy Collaboration et al. 2018), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), BILBY (Ashton et al. 2019), GWPOLULATION (Talbot et al. 2019), DYNESTY (Speagle 2020).

## Appendix A Hierarchical Inference

We use hierarchical Bayesian inference to simultaneously infer the population distributions of the primary masses ( $m_1$ ), mass ratios ( $q$ ) and the redshifts ( $z$ ) of observed BBHs. For a set of hyperparameters,  $\Lambda$ , and local ( $z=0$ ) merger rate density (units of mergers per co-moving volume per time),  $\mathcal{R}_0$ , we write the overall number density of BBH mergers in the universe as:

$$\begin{aligned} & \frac{dN(m_1, q, z|\mathcal{R}_0, \Lambda)}{dm_1 dq dz} \\ &= \frac{dV_c}{dz} \left( \frac{T_{\text{obs}}}{1+z} \right) \frac{d\mathcal{R}(m_1, q, z|\mathcal{R}_0, \Lambda)}{dm_1 dq} \\ &= \mathcal{R}_0 p(m_1|\Lambda) p(q|m_1, \Lambda) p(z|\Lambda) \end{aligned} \quad (\text{A1})$$

Above,  $dV_c$  is the co-moving volume element (Hogg 1999) and  $T_{\text{obs}}$ , the observing time period that produced the catalog with the related factor of  $1+z$  converting this detector-frame time to source frame. We assume a Lambda CDM cosmology using the cosmological parameters from Ade et al. (2016). The redshift evolution of the merger rate follows  $p(z|\lambda) \propto \frac{dV_c}{dz} \frac{1}{1+z} (1+z)^\lambda$ . Integrating Equation (A1) across all masses, and up to some redshift,  $z_m$ , returns the total number of BBH mergers in the universe out to that redshift. Let  $\{d_i\}$  represent a set of data from  $N_{\text{obs}}$  observed gravitational waves associated with BBH mergers. We model the merger rate as an inhomogenous point process and when imposing a log-uniform prior on the merger rate, we can marginalize over the merger rate to get the posterior distribution of our hyperparameters,  $\Lambda$  (Mandel et al. 2019; Vitale et al. 2021b).

$$\begin{aligned} p(\Lambda|\{d_i\}) &\propto \frac{p(\Lambda)}{\xi(\Lambda)^{N_{\text{obs}}}} \prod_{i=1}^{N_{\text{obs}}} \left[ \int \mathcal{L}(d_i|m_1^i, q^i, z^i) \right. \\ &\quad \times p(m_1|\Lambda) p(q|m_1, \Lambda) p(z|\Lambda) dm_1 dq dz \right], \end{aligned} \quad (\text{A2})$$

where,  $\mathcal{L}(d_i|m_1, q, z)$ , is the single-event likelihood function from each events original analysis, and  $\xi(\Lambda)$  is the detection efficiency given a population distribution described by  $\Lambda$ . The procedure for calculating  $\xi(\Lambda)$  is described in more detail below. The LVC reports out posterior samples for each

observed event, with which we can use importance sampling to estimate the integrals above in Equation (A2). We replace the integrals with ensemble averages over  $K_i$  posterior samples associated with each event in the catalog:

$$\begin{aligned} p(\Lambda|\{d_i\}) &\propto p(\Lambda) \prod_{i=1}^{N_{\text{obs}}} \\ &\times \left[ \frac{1}{K_i} \sum_{j=1}^{K_i} \frac{p(m_1^{i,j}|\Lambda) p(q^{i,j}|\Lambda) p(z^{i,j}|\Lambda)}{\pi(m_1^{i,j}, q^{i,j}, z^{i,j})} \right] \end{aligned} \quad (\text{A3})$$

Here  $j$  indexes the  $K_i$  posterior samples from each event and  $\pi(m_1, q, z)$  is the default prior used by parameter estimations that produced the posterior samples for each event. In the analyses of GWTC-2 the default prior used is uniform in detector-frame masses and Euclidean volume. The corresponding prior evaluated in source frame masses and redshift is  $\pi(m_1, q, z) \propto m_1(1+z)^2 D_L^2(z) \frac{dD_L}{dz}$ , where  $D_L$  is the luminosity distance.

To carefully incorporate selection effects to our model we need to quantify the detection efficiency,  $\xi(\Lambda)$ , of the search pipelines that were used to create GWTC-2, at a given population distribution described by  $\Lambda$ .

$$\begin{aligned} \xi(\Lambda) &= \int dm_1 dq dz P_{\text{det}}(m_1, q, z) \\ &\times p(m_1|\Lambda) p(q|m_1, \Lambda) p(z|\Lambda) \end{aligned} \quad (\text{A4})$$

The above integral is not tractable since there is no analytic prescription for  $P_{\text{det}}(m_1, q, z)$ , the detection probability of an individual event. To estimate this integral we use a software injection campaign where GWs from a fixed, broad population of sources are simulated, put into real detector data, and then run through the same search pipelines that were used to produce the catalog we are analyzing.<sup>3</sup> With these search results in hand, we use importance sampling to evaluate the integral in Equation (A4):

$$\xi(\Lambda) = \frac{1}{N_{\text{inj}}} \sum_{i=1}^{N_{\text{found}}} \frac{p(m_1^i|\Lambda) p(q^i|m_1, \Lambda) p(z^i|\Lambda)}{p_{\text{inj}}(m_1^i, q^i, z^i)} \quad (\text{A5})$$

Where the sum indexes only over the  $N_{\text{found}}$  injections that were successfully detected out of  $N_{\text{inj}}$  total injections, and  $p_{\text{inj}}(m_1, q, z)$  is the reference distribution from which the injections were drawn. Additionally, we follow the procedure outlined in Farr (2019) to marginalize the uncertainty in our estimate of  $\xi(\Lambda)$  due to a finite number of simulated events. We make the assumption that repeated sampling of  $\xi(\Lambda)$  will follow a normal distribution with  $\xi(\Lambda) \sim \mathcal{N}(\mu(\Lambda), \sigma(\Lambda))$ , where the mean,  $\mu$ , is the estimate from Equation (A5), while the variance,  $\sigma^2$ , is defined as:

$$\begin{aligned} \sigma^2(\Lambda) &\equiv \frac{\mu^2(\Lambda)}{N_{\text{eff}}} \simeq \frac{1}{N_{\text{inj}}^2} \sum_{i=1}^{N_{\text{found}}} \\ &\times \left[ \frac{p(m_1^i|\Lambda) p(q^i|m_1, \Lambda) p(z^i|\Lambda)}{p_{\text{inj}}(m_1^i, q^i, z^i)} \right]^2 - \frac{\mu^2(\Lambda)}{N_{\text{inj}}} \end{aligned} \quad (\text{A6})$$

<sup>3</sup> For O3a we used the injection sets used by Abbott et al. (2021b), which can be found at <https://dcc.ligo.org/LIGO-P2000217/public>. For O1/O2 we used the mock injection sets used by Abbott et al. (2019b) which can be found at <https://dcc.ligo.org/LIGO-P2000434/public>.

Above we define  $N_{\text{eff}}$  as the effective number of independent draws contributing to the importance sampled estimate, in which we verify to be sufficiently high after re-weighting the injections to a given population (i.e.,  $N_{\text{eff}} > 4N_{\text{obs}}$ ). We write the hyperposterior marginalized over the merger rate and uncertainty in estimation of  $\xi(\Lambda)$ , neglecting terms of  $\mathcal{O}(N_{\text{eff}}^{-2})$  or greater (Farr 2019), as:

$$\begin{aligned} \log p(\Lambda | \{d_i\}) &\propto \sum_{i=1}^{N_{\text{obs}}} \log \\ &\times \left[ \frac{1}{K_i} \sum_{j=1}^{K_i} \frac{p(m_1^{i,j} | \Lambda) p(q^{i,j} | \Lambda) p(z^{i,j} | \Lambda)}{\pi(m_1^{i,j}, q^{i,j}, z^{i,j})} \right] \\ &- N_{\text{obs}} \log \mu + \frac{3N_{\text{obs}} + N_{\text{obs}}^2}{2N_{\text{eff}}} + \mathcal{O}(N_{\text{eff}}^{-2}). \end{aligned} \quad (\text{A7})$$

We explicitly enumerate each of the models used in this work for  $p(m_1 | \Lambda)$ ,  $p(q | m_1, \Lambda)$ , and  $p(z | \Lambda)$  along with their respective hyperparameters and prior distributions in Table 1. To calculate marginal likelihoods and draw samples of the hyperparameters from the hierarchical posterior distribution shown in Equation (A7), we use the BILBY (Ashton et al. 2019; Romero-Shaw et al. 2020) and GWPOPULATION (Talbot et al. 2019) Bayesian inference software libraries with the DYNESTY dynamic nested sampling algorithm (Speagle 2020).

## Appendix B Model Comparisons and Prior Specifications

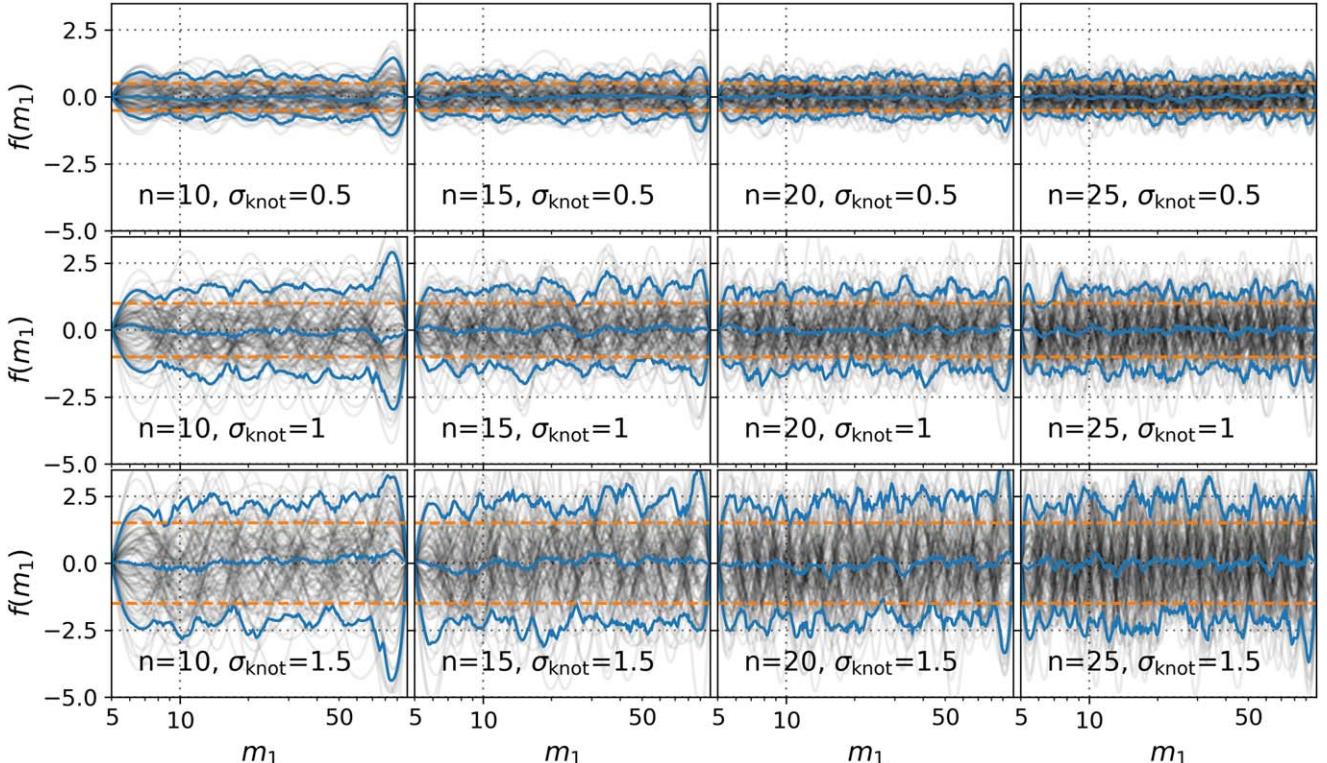
To compare competing population models in the aforementioned Bayesian framework we calculate two different measures of model goodness of fit, namely the marginal likelihoods ( $\mathcal{Z}$ ) and deviance information criterion (DIC) (Spiegelhalter et al. 2002).

The marginal likelihood for a given model is a constant that enforces that the posterior distribution is normalized (i.e.,  $\mathcal{Z} = \int d\Lambda p(\Lambda | \{d_i\})$ ), which has the property that it is higher for models that fit the data better or find higher likelihoods, while penalizing more complicated models by their prior volumes. As our semiparametric approach has arbitrary prior choices one needs to make, this can significantly affect the marginal likelihoods inferred. We also calculate the DIC, a metric developed specifically for Bayesian hierarchical models (Spiegelhalter et al. 2002), which is less sensitive to the arbitrary prior choices for our semiparametric model. While there are some limitations to the DIC (Spiegelhalter et al. 2014), it provides a secondary metric to validate our model choices. The DIC is defined as:

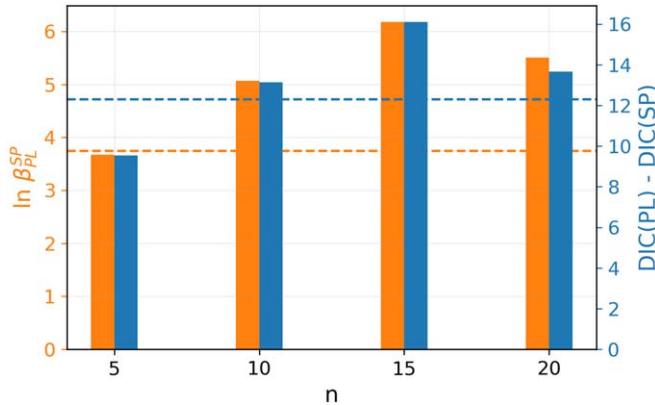
$$\begin{aligned} \text{DIC} &= -2\overline{\log(\mathcal{L})} + p_D \\ &= -2(\overline{\log(\mathcal{L})} - \text{var}(\log(\mathcal{L}))) \end{aligned} \quad (\text{B1})$$

With  $\overline{\log \mathcal{L}}$  the mean log-likelihood, and  $p_D$  the effective number of dimensions, defined as  $p_D = \frac{1}{2}\text{var}(-2 \log \mathcal{L})$  with  $\text{var}(\dots)$  denoting the variance. Lower DICs indicate better models which, similarly to the marginal likelihood, favors models that find higher likelihoods while penalizing the more complicated models through the effective dimension term. We compare two models by calculating the ratio of their marginal likelihoods (i.e., Bayes Factors<sup>4</sup>), defined as the ratio of each models marginal likelihoods. To compare DICs, we take the difference of two models values (DIC dif =  $\text{DIC}_A - \text{DIC}_B$ ) where positive differences indicate preference for model B, and negative differences indicate preference for model A.

We use these model comparisons to motivate a sensible choice for our spline models prior flexibility, namely the number of knots ( $n$ ). Figure 6 shows changing prior widths on



**Figure 6.** 100 draws from the prior predictive distribution of the cubic spline function,  $f(m_1)$ , for different choices for the number of knots,  $n$ , and the width of the Gaussian priors on the knots,  $\sigma_{\text{knot}}$ . The orange dashed line shows  $\pm \sigma_{\text{knot}}$  while the blue solid lines show the median and  $1\sigma$  credible bounds of the draws from the prior.



**Figure 7.** We show the two model comparison methods, Bayes factors (orange, left y-axis) and the DIC difference (blue, right y-axis), each comparing our spline perturbation model (denoted as SP) to the TRUNCATED model (denoted as PL). The comparisons are calculated such that positive values of either metric denote the spline perturbation model being favored over the TRUNCATED. Both values are shown for varying specifications for the spline prior. Along the x-axis we show different discrete choices (5, 10, 15, and 20) for the number of nodes, n. Each of these spline model analyses shown was performed with  $\sigma_{\text{knot}} = 1$ . The horizontal dashed lines show the Bayes factor (orange) and DIC difference (blue) found when comparing Abbott et al. (2021b)'s favorite mass model, POWERLAW+PEAK, to the TRUNCATED model.

our knots only only effects magnitude of perturbations the spline is sensitive to. Additionally, we see that as we add more knots, the model is free to fit sharper fluctuations. This flexibility comes with a penalty in our comparison metrics due to increased model complexity. Therefore, we would expect to see the model comparisons increasingly favor our spline perturbation model as we increase our models complexity/flexibility up to a point where the penalty will start to overpower the higher likelihoods found with more flexibility. Figure 7 shows how the DIC differences and log Bayes factors ( $\ln \mathcal{B}$ ) change when comparing the spline perturbation model to the TRUNCATED model with different choices for  $n$ . We see that the comparisons favoring our spline model peaks around 15 knots, indicating that 15 knots is a good trade-off between our models flexibility and goodness of fit. We also report the marginal likelihoods and model comparisons (relative to the most preferred model) for each of the parametric primary mass models from Abbott et al. (2021b) and various specifications of the spline model in Table 2. From this table we see that the spline model is consistently favored despite our arbitrary model specifications, giving credence to the hypothesis that there are features in the data our semiparametric method is capable of finding that previously used parametric mass models are not sensitive to. We do not use the comparisons in Table 2 to determine the validity of the POWERLAW+SPLINE models over others, and further studies on simulated populations and the effect of small-number statistics are needed to fully assess the significance and robustness of these features. However, as catalogs of BBH

<sup>4</sup> The true “Bayesian” way to compare models is using odds, which are Bayes factors multiplied by the ratio of prior odds of each model. Because we do not a priori have expectations of which population model would be more likely, we use Bayes factors which are odds ratios with equal prior weights for each model.

**Table 2**

Model Comparison Results, Listing Each Model Tested (Semiparametric Spline Model or Parametric Mass Model from Abbott et al. (2021b)) and Their Respective Marginal Likelihoods ( $\mathcal{Z}$ ) along with  $\ln \mathcal{B}$  and DIC dif

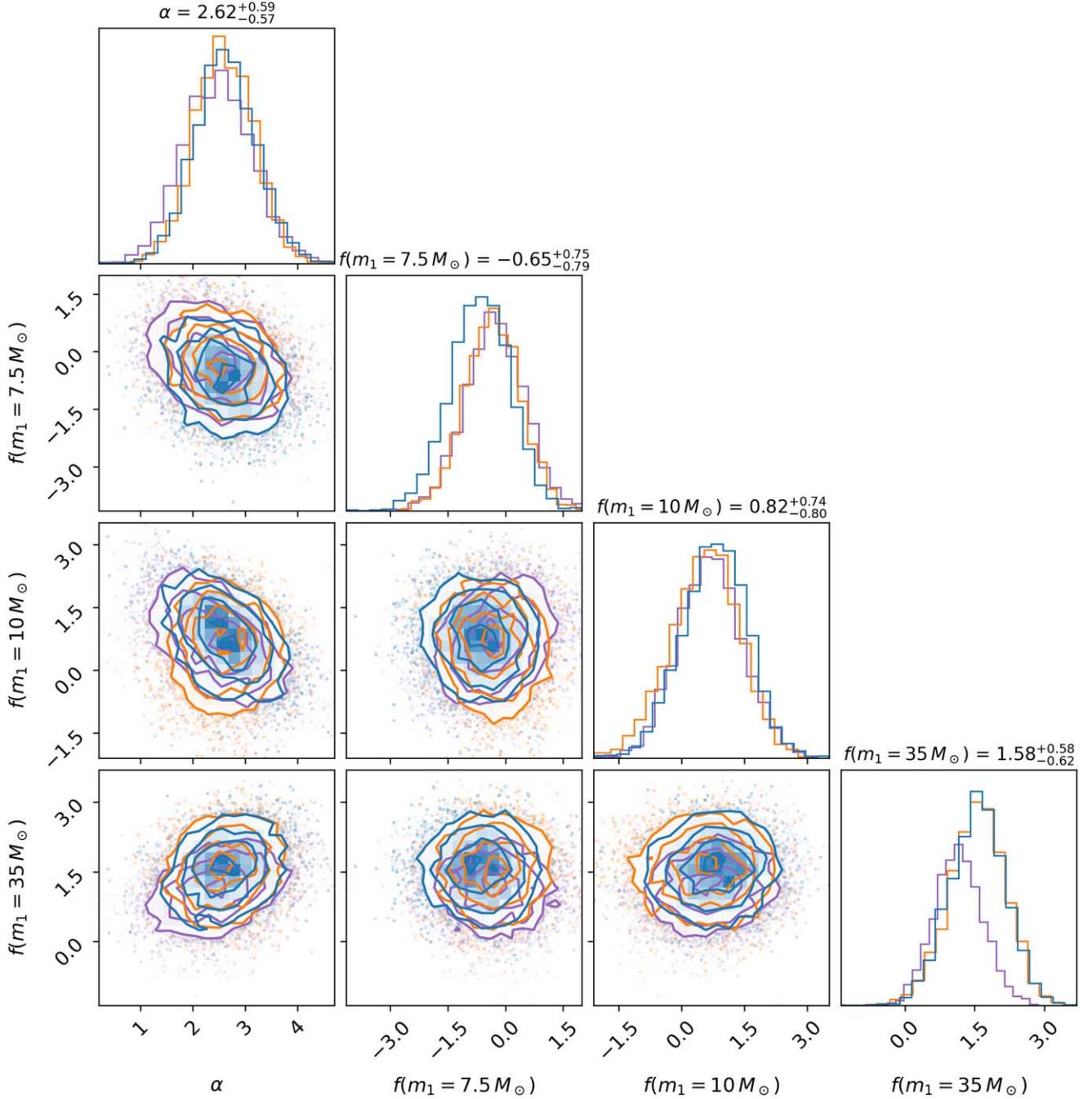
Model Name	$\ln \mathcal{Z}$	$\ln \mathcal{B}$	DIC dif
Powerlaw+Spline ( $n = 15$ , $\sigma_{\text{knot}} = 1$ )	-347.10	0.00	0.00
Powerlaw+Spline ( $n = 20$ , $\sigma_{\text{knot}} = 1$ )	-347.77	-0.67	-2.43
Powerlaw+Spline ( $n = 15$ , $\sigma_{\text{knot}} = 1$ , $\delta_m = 0$ )	-347.83	-0.74	-3.27
Powerlaw+Spline ( $n = 10$ , $\sigma_{\text{knot}} = 1$ )	-348.21	-1.11	-2.96
Powerlaw+Spline ( $n = 10$ , $\sigma_{\text{knot}} = 1$ , $\delta_m = 0$ )	-348.24	-1.14	-3.25
Powerlaw+Spline ( $n = 20$ , $\sigma_{\text{knot}} = 1$ , $\delta_m = 0$ )	-348.26	-1.16	-3.57
Powerlaw+Spline ( $n = 15$ , $\sigma_{\text{knot}} = 2$ )	-348.40	-1.30	-1.00
Powerlaw+Spline ( $n = 15$ , $\sigma_{\text{knot}} = 2$ , $\delta_m = 0$ )	-348.53	-1.43	-2.85
Powerlaw+Spline ( $n = 20$ , $\sigma_{\text{knot}} = 2$ )	-348.62	-1.52	-3.00
Powerlaw+Spline ( $n = 20$ , $\sigma_{\text{knot}} = 2$ , $\delta_m = 0$ )	-349.01	-1.91	-3.39
Powerlaw+MultiPeak	-349.27	-2.18	-6.87
Powerlaw+Spline ( $n = 10$ , $\sigma_{\text{knot}} = 2$ , $\delta_m = 0$ )	-349.43	-2.34	-3.12
Powerlaw+MultiPeak ( $\delta_m = 0$ )	-349.44	-2.35	-6.27
Powerlaw+Peak	-349.53	-2.43	-3.80
Powerlaw+Spline ( $n = 10$ , $\sigma_{\text{knot}} = 2$ )	-349.66	-2.57	-3.22
Powerlaw+Peak ( $\delta_m = 0$ )	-349.70	-2.60	-4.96
Broken Powerlaw ( $\delta_m = 0$ )	-349.92	-2.83	-9.32
Broken Powerlaw	-350.21	-3.11	-8.90
Truncated	-352.57	-5.47	-14.32
Truncated ( $\delta_m = 0$ )	-353.27	-6.18	-16.10

**Note.** Both comparison metrics for each of the listed models are relative to the “best performing” model or the one with the highest (lowest) marginal likelihood (DIC), which, in both cases, was the Powerlaw+Spline ( $n = 15$ ,  $\sigma_n = 1$ ) model. We note that the POWERLAW+MULTIPEAK finds higher marginal likelihoods than the POWERLAW+PEAK model which was not the case in Abbott et al. (2021b). This is because we used different priors for the POWERLAW+MULTIPEAK model that allowed for a peak at lower masses than the  $\sim 35 M_\odot$  peak instead of higher.

mergers increase in size, the impact of small-number statistics will diminish.

## Appendix C Correlations of Peaks

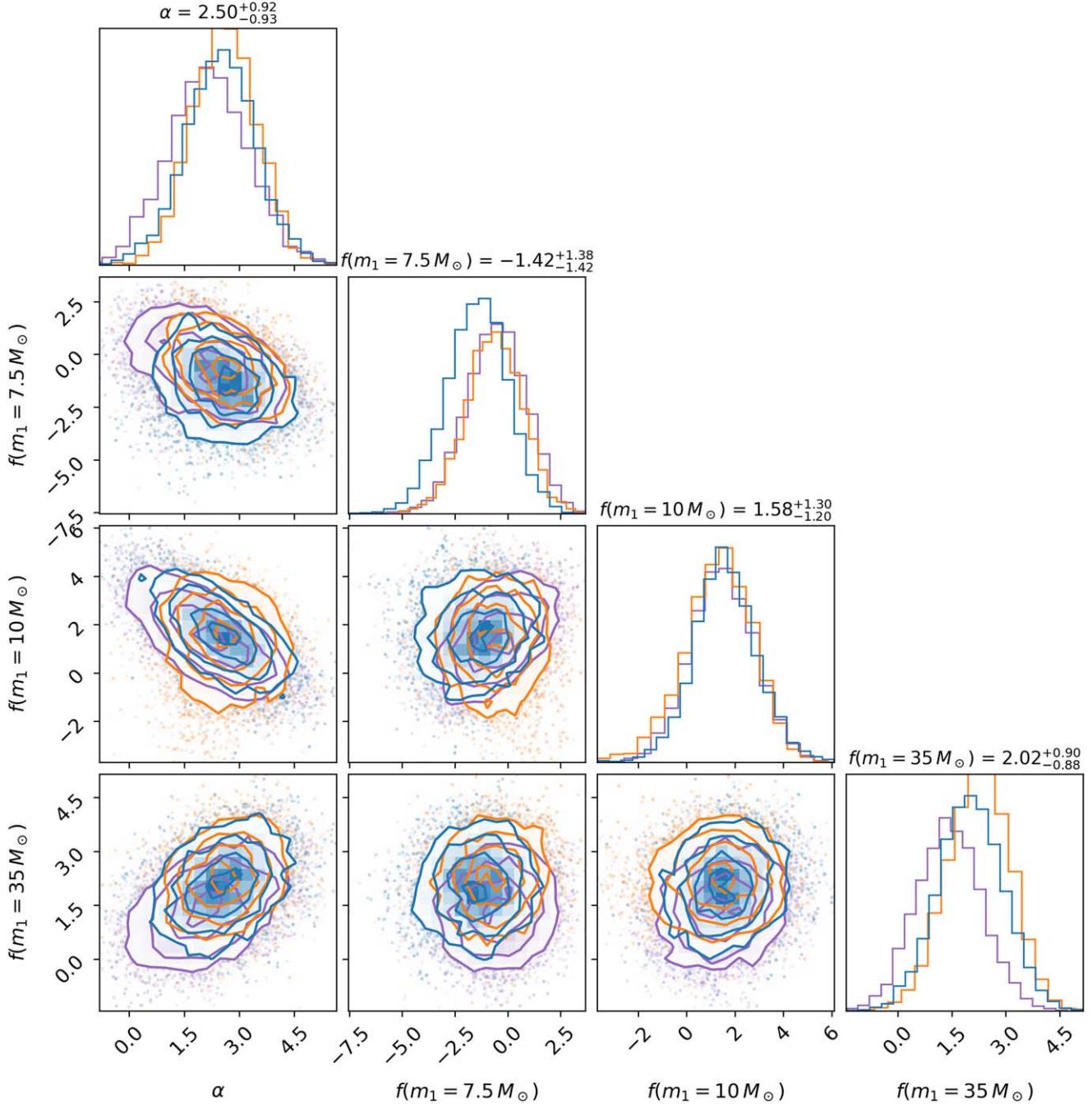
We look for the effect of our spline function biasing the inferred perturbation function by plotting a corner plot of the value of  $f(m_1)$  sliced through the masses that show the largest deviations. This is shown in Figure 8 for the conservative knot prior and Figure 9 for the wide knot prior. If the dip followed by the peak feature at 7.5 and 10 solar masses was found due to the nature of cubic splines we would expect to see correlations



**Figure 8.** Corner plot that shows the posterior distribution on the power-law slope,  $\alpha$ , and the height of the perturbation function,  $f(m_1)$ , sliced at the three masses of most significant deviation:  $7.5 M_\odot$ ,  $10 M_\odot$ , and  $35 M_\odot$ . We show the results for spline models with  $\sigma_{\text{knot}} = 1$  and 10 (purple), 15 (blue) and 20 (orange) nodes. The median and 90% credible intervals quoted are for the 15 knot model.

between the heights at these values which do not appear in either Figures 8 or 9. Since we are fitting for the underlying power law model simultaneously with the perturbations we also might expect to see some correlations with the power law slope and the peaks. There are slight signs of an expected anti-correlation of the  $10 M_\odot$  peak height and the power law slope, and corresponding correlation of the  $35 M_\odot$  peak height with

the slope. This happens due to the degeneracy between the parametric and nonparametric portions of our model. If the  $10 M_\odot$  peak is small, the power law slope becomes steeper so that the “turnover” power law at low mass can contribute to fitting this over-density. With a steeper power law slope, the power law portion of the model under fits the excess at  $35 M_\odot$ , leading to a larger peak found in the perturbation. We also note



**Figure 9.** Corner plot that shows the posterior distribution on the power law slope,  $\alpha$ , and the height of the perturbation function,  $f(m_1)$ , sliced at the three masses of most significant deviation:  $7.5 M_\odot$ ,  $10 M_\odot$ , and  $35 M_\odot$ . We show the results for spline models with  $\sigma_{\text{knot}} = 2$  and 10 (purple), 15 (blue) and 20 (orange) nodes. The median and 90% credible intervals quoted are for the 15 knot model.

these correlations with the power law slope are more apparent in the lower resolution spline models and under the wider prior on the knots.

### ORCID IDs

Bruce Edelman  <https://orcid.org/0000-0001-7648-1689>  
 Zoheyr Doctor  <https://orcid.org/0000-0002-2077-4914>  
 Jaxen Godfrey  <https://orcid.org/0000-0002-2202-8640>  
 Ben Farr  <https://orcid.org/0000-0002-2916-9200>

### References

- Abbott, B., Abbott, R., Abbott, T., et al. 2016, *PhRvL*, **116**, 131103
- Abbott, B., Abbott, R., Abbott, T., et al. 2017, *PhRvL*, **119**, 161101
- Abbott, B., Abbott, R., Abbott, T., et al. 2019a, *PhRvX*, **9**, 031040
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019b, *ApJL*, **882**, L24
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020a, *ApJL*, **892**, L3
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020b, *LRR*, **23**, 3
- Abbott, R., Abbott, T., Abraham, S., et al. 2020e, *PhRvD*, **102**, 043015
- Abbott, R., Abbott, T., Abraham, S., et al. 2020f, *PhRvL*, **125**, 101102
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020d, *ApJL*, **896**, L44
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020g, *ApJL*, **900**, L13
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, *PhRvX*, **11**, 021053

- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021b, *ApJL*, **913**, L7
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021c, *ApJL*, **915**, L5
- Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, *A&A*, **594**, A13
- Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, *ApJS*, **241**, 27
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Belczynski, K. 2020, *ApJL*, **905**, L15
- Belczynski, K., Heger, A., Gladysz, W., et al. 2016, *A&A*, **594**, A97
- Croon, D., McDermott, S. D., & Sakstein, J. 2020, *PhRvD*, **102**, 115024
- Cruz-Ororio, A., Lora-Clavijo, F. D., & Herdeiro, C. 2021, *JCAP*, **2021**, 032
- Doctor, Z., Farr, B., & Holz, D. E. 2021, *ApJ*, **914**, L18
- Doctor, Z., Farr, B., Holz, D. E., & Pürrer, M. 2017, *PhRvD*, **96**, 123011
- Doctor, Z., Wysocki, D., O’Shaughnessy, R., Holz, D. E., & Farr, B. 2020, *ApJ*, **893**, 35
- Edelman, B., Doctor, Z., & Farr, B. 2021a, *ApJL*, **913**, L23
- Edelman, B., Rivera-Paleo, F., Merritt, J., et al. 2021b, *PhRvD*, **103**, 042004
- Estellés, H., Husa, S., Colleoni, M., et al. 2021, arXiv:2105.06360
- Farmer, R., Renzo, M., de Mink, S. E., Fishbach, M., & Justham, S. 2020, *ApJL*, **902**, L36
- Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, *ApJ*, **887**, 53
- Farr, W. M. 2019, *RNAAS*, **3**, 66
- Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, *ApJL*, **883**, L42
- Farrell, E., Groh, J. H., Hirschi, R., et al. 2021, *MNRAS Lett.*, **502**, L40
- Fishbach, M., Essick, R., & Holz, D. E. 2020, *ApJL*, **899**, L8
- Fishbach, M., & Holz, D. E. 2017, *ApJL*, **851**, L25
- Fishbach, M., & Holz, D. E. 2020, *ApJL*, **904**, L26
- Fishbach, M., Holz, D. E., & Farr, W. M. 2018, *ApJL*, **863**, L41
- Gayathri, V., Healy, J., Lange, J., et al. 2020, arXiv:2009.05461
- Gerosa, D., & Berti, E. 2017, *PhRvD*, **95**, 124046
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, *ApJ*, **591**, 288
- Heger, A., & Woosley, S. E. 2002, *ApJ*, **567**, 532
- Hogg, D. W. 1999, arXiv:astro-ph/9905116
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Kimball, C., Talbot, C. L., Berry, C. P., et al. 2020, *ApJ*, **900**, 177
- Kimball, C., Talbot, C., Berry, C. P. L., et al. 2021, *ApJL*, **915**, L35
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Littenberg, T. B., & Cornish, N. J. 2015, *PhRvD*, **91**, 084034
- Mandel, I., Farr, W. M., Colonna, A., et al. 2016, *MNRAS*, **465**, 3254
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, **486**, 1086
- Mapelli, M., Santoliquido, F., Bouffanais, Y., et al. 2020, *Symm*, **13**, 1678
- Marchant, P., Renzo, M., Farmer, R., et al. 2019, *ApJ*, **882**, 36
- McKernan, B., Ford, K. E. S., O’Shaughnessy, R., & Wysocki, D. 2020, *MNRAS*, **494**, 1203
- Ng, K. K. Y., Vitale, S., Farr, W. M., & Rodriguez, C. L. 2021, *ApJL*, **913**, L5
- Nitz, A. H., & Capano, C. D. 2021, *ApJL*, **907**, L9
- Rodriguez, C. L., Zevin, M., Amaro-Seoane, P., et al. 2019, *PhRvD*, **100**, 043027
- Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, *MNRAS*, **499**, 3295
- Sakstein, J., Croon, D., McDermott, S. D., Straight, M. C., & Baxter, E. J. 2020, *PhRvL*, **125**, 261105
- Secunda, A., Bellovary, J., Mac Low, M.-M., et al. 2020, *ApJ*, **903**, 133
- Speagle, J. S. 2020, *MNRAS*, **493**, 3132
- Spera, M., & Mapelli, M. 2017, *MNRAS*, **470**, 4739
- Spiegelhalter, D., Best, N., Carlin, B., & Linde, A. 2014, *J. R. Stat. Soc. Series B Stat. Methodol.*, **76**, 485
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. 2002, *J. R. Stat. Soc. Series B Stat. Methodol.*, **64**, 583
- Stevenson, S., Sampson, M., Powell, J., et al. 2019, *ApJ*, **882**, 121
- Talbot, C., Smith, R., Thrane, E., & Poole, G. B. 2019, *PhRvD*, **100**, 043030
- Talbot, C., & Thrane, E. 2018, *ApJ*, **856**, 173
- Tiwari, V. 2021, *CQGra*, **38**, 155007
- Tiwari, V., & Fairhurst, S. 2021, *ApJL*, **913**, L19
- Virnstein, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2021a, arXiv:2007.05579
- Vitale, S., Haster, C.-J., Sun, L., et al. 2021b, *PhRvD*, **103**, 063016
- Woosley, S. E. 2017, *ApJ*, **836**, 244
- Woosley, S. E. 2019, *ApJ*, **878**, 49
- Woosley, S. E., Heger, A., & Weaver, T. A. 2002, *RvMP*, **74**, 1015
- Zevin, M., Pankow, C., Rodriguez, C. L., et al. 2017, *ApJ*, **846**, 82