



When Are LIGO/Virgo's Big Black Hole Mergers?

Maya Fishbach¹ , Zoheyr Doctor² , Thomas Callister³ , Bruce Edelman² , Jiani Ye⁴, Reed Essick⁵ , Will M. Farr^{3,4} ,
Ben Farr² , and Daniel E. Holz^{5,6,7,8}

¹ Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 1800 Sherman Ave., Evanston, IL 60201, USA

² Institute for Fundamental Science, Department of Physics, University of Oregon, Eugene, OR 97403, USA

³ Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA

⁴ Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA

⁵ Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, IL 60637, USA

⁶ Enrico Fermi Institute, The University of Chicago, IL 60637, USA

⁷ Department of Physics, The University of Chicago, IL 60637, USA

⁸ Department of Astronomy & Astrophysics, The University of Chicago, IL 60637, USA

Received 2021 January 25; accepted 2021 March 10; published 2021 May 10

Abstract

We study the evolution of the binary black hole (BBH) mass distribution across cosmic time. The second gravitational-wave transient catalog (GWTC-2) from LIGO/Virgo contains BBH events out to redshifts $z \sim 1$, with component masses in the range $\sim 5\text{--}80 M_\odot$. In this catalog, the biggest BBHs, with $m_1 \gtrsim 45 M_\odot$, are only found at the highest redshifts, $z \gtrsim 0.4$. We ask whether the absence of high-mass observations at low redshift indicates that the mass distribution evolves: the biggest BBHs only merge at high redshift, and cease merging at low redshift. Modeling the BBH primary-mass spectrum as a power law with a sharp maximum mass cutoff (TRUNCATED model), we find that the cutoff increases with redshift ($> 99.9\%$ credibility). An abrupt cutoff in the mass spectrum is expected from (pulsational) pair-instability supernova simulations; however, GWTC-2 is only consistent with a TRUNCATED mass model if the location of the cutoff increases from $45^{+13}_{-5} M_\odot$ at $z < 0.4$ to $80^{+16}_{-13} M_\odot$ at $z > 0.4$. Alternatively, if the primary-mass spectrum has a break in the power law (BROKEN POWER LAW) at $38^{+15}_{-8} M_\odot$, rather than a sharp cutoff, the data are consistent with a nonevolving mass distribution. In this case, the overall rate of mergers, at all masses, increases with redshift. Future observations will distinguish between a sharp mass cutoff that evolves with redshift and a nonevolving mass distribution with a gradual taper, such as a BROKEN POWER LAW. After ~ 100 BBH merger observations, a continued absence of high-mass, low-redshift events would provide a clear signature that the mass distribution evolves with redshift.

Unified Astronomy Thesaurus concepts: [Astrophysical black holes \(98\)](#); [Gravitational wave astronomy \(675\)](#); [Astrostatistics \(1882\)](#); [Massive stars \(732\)](#); [Stellar mass black holes \(1611\)](#); [Stellar remnants \(1627\)](#)

1. Introduction

In their first three observing runs, the Advanced LIGO (LIGO Scientific Collaboration et al. 2015) and Virgo (Acernese et al. 2015) gravitational-wave (GW) detectors observed binary black hole (BBH) mergers out to redshifts $z \sim 1$ (Abbott et al. 2019a, 2020a). Observing BBH systems over a range of redshifts allows us to probe the properties of these mergers across cosmic time and unravel how these merging BBH systems came to be. Previous studies (Fishbach et al. 2018; Abbott et al. 2019b, 2020b; Callister et al. 2020; Roulet et al. 2020; Tiwari 2020) have measured the BBH merger rate as a function of redshift, assuming that other properties of the population, including the mass and spin distributions, are constant throughout cosmic time. However, there are reasons to expect that the overall BBH population properties may themselves evolve with redshift:

1. The initial conditions of zero-age main-sequence stars (e.g., metallicity) evolve over cosmic time, which could affect the resultant masses and spins of black holes (BHs) from stellar evolution (Belczynski et al. 2010; Kudritzki & Puls 2000; Brott et al. 2011; Fryer et al. 2012; Dominik et al. 2015; Neijssel et al. 2019; Safarzadeh & Farr 2019; Farrell et al. 2021; Kinugawa et al. 2021; Vink et al. 2021).
2. If BBH mergers occur in dynamical environments, the dynamical environments could evolve over cosmic time,

(dis)favoring mergers of different masses and spins (Rodriguez & Loeb 2018; El-Badry et al. 2019; Romero-Shaw et al. 2020; Santoliquido et al. 2020; Weatherford et al. 2021).

3. The delay time from the inception of a BBH to its merger may depend on the masses and spins of the component BHs and the orbital eccentricity (Kushnir et al. 2016; Samsing 2018; Mapelli et al. 2019).
4. Different formation channels (including a formation channel that permits hierarchical mergers) may contribute more or less to the overall merger rate at different times (Rodriguez & Loeb 2018; Rodriguez et al. 2019; Santoliquido et al. 2020; Yang et al. 2020; Zevin et al. 2021).

The combined effect of these phenomena would manifest in the GW data as mass and/or spin distributions of merging BHs that are different at different redshifts. For example, Figure 1 of Rodriguez et al. (2019) shows that the mass distribution for mergers in dense star clusters extends to higher masses when considering mergers at all redshifts compared to mergers with $z < 1$. Similarly, some population synthesis models of BBH mergers from isolated binary evolution exhibit more support for higher-mass mergers at higher redshifts, although this effect is expected to be mild at the redshifts accessible to Advanced LIGO and Advanced Virgo (e.g., Mapelli et al. 2019). With a

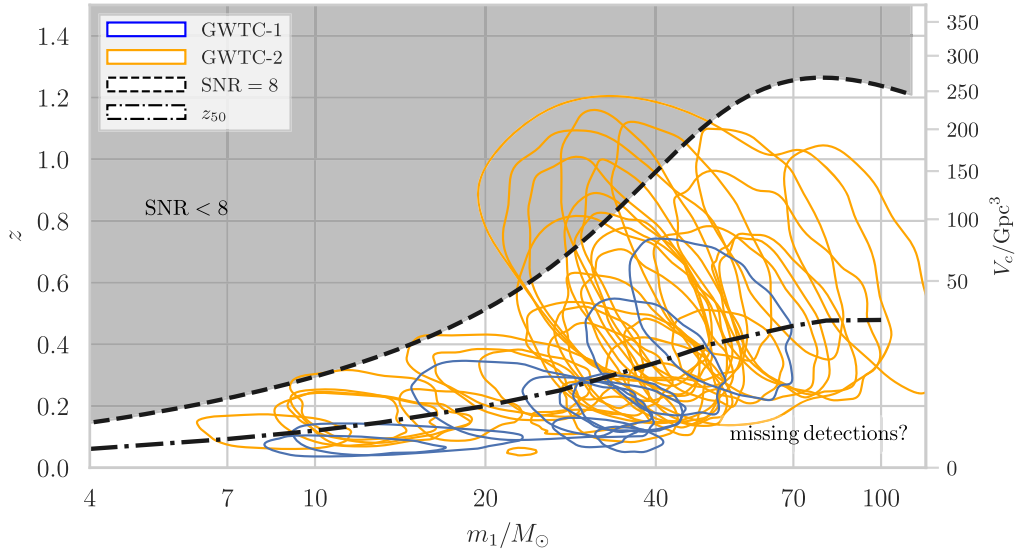


Figure 1. Posterior probability distribution of the primary mass m_1 and redshift z for the 34 confident BBH events announced in GWTC-2 (orange) and 10 confident BBHs in GWTC-1 (blue). The contours enclose 90% of the posterior probability, inferred under the default parameter estimation priors used in the GWTC-1 and GWTC-2 publications (Abbott et al. 2019a, 2020a). The shaded region denotes the redshifts for which the signal-to-noise ratio of an equal-mass BBH is less than 8 for an optimally oriented component BH mass under the Advanced LIGO “mid-high” noise curve (Abbott et al. 2018); we do not detect low-mass, high-redshift events because they are too quiet. As a point of reference, we overlay the curve representing z_{50} , the expected median distance of detected mergers that are distributed with constant rate density per comoving volume. The question of whether there are equal numbers of events above and below this z_{50} line motivates us to examine whether the intrinsic BBH mass distribution evolves with redshift.

growing catalog of BBH events, we can begin to empirically measure the existence (or lack thereof) of redshift evolution out to $z \sim 1$.

The second GW transient catalog (GWTC-2; Abbott et al. 2020a) contains 44 confident BBH mergers with $m_1 > m_2 > 3 M_\odot$ and a false-alarm rate (FAR) below 1 yr^{-1} . While the first LIGO/Virgo catalog release, GWTC-1, found evidence for a dearth of systems with $m_1 > 45 M_\odot$, the updated GWTC-2 catalog contains several systems with $m_1 > 45 M_\odot$. The presence of these high-mass BHs leads Abbott et al. (2020b) to infer that the primary-mass spectrum is more complicated than a power law with an abrupt mass cutoff. Although a mass cutoff at $\sim 40 M_\odot$ was consistent with GWTC-1, Abbott et al. (2020b) concluded that the GWTC-2 observations are more consistent with a *break* or a *peak* at $\sim 40 M_\odot$, with the mass distribution declining more steeply at higher masses.

Notably, all of the high-mass observations in GWTC-2 are also at relatively high redshifts, as seen in Figure 1, with a noticeable absence of events with $m_1 \gtrsim 45 M_\odot$ at low redshifts. For reference, we also overlay the expected median redshift z_{50} of sources distributed with a constant rate per comoving volume¹⁰ (Chen et al. 2021; Abbott et al. 2018). The apparent absence is not necessarily surprising, because the highest-mass systems are also detectable at the highest redshifts, and, if these systems are rare, we expect to detect them primarily at high redshifts where there is more cosmological volume. However, this also suggests an alternative explanation for the high-mass detections in GWTC-2, also proposed by Safarzadeh & Farr (2019): the underlying astrophysical mass distribution may

skew to higher masses at higher redshifts, implying that a higher fraction of high-mass mergers (per comoving time-volume) occur at higher redshifts than at low redshifts.

The remainder of the paper is structured as follows. Section 2 introduces the phenomenological models that we use to describe the BBH population, and the statistical framework that we use to fit the models given the GWTC-2 data. Section 3 presents our main results regarding the evolution of the BBH mass distribution. We carry out posterior predictive checks in Section 4, and discuss future prospects before concluding in Section 5.

2. Methods

We first describe the parameterized population models we assume in Section 2.1 before laying out how we use these within the statistical framework of Section 2.2.

2.1. Population Models

We use simple phenomenological parametric models to describe the distribution of BBH masses m_1 and m_2 , effective spin χ_{eff} , and redshift z , based on the models used in Abbott et al. (2020b). We write the differential merger rate density (number of mergers per comoving volume per source-frame time) as:

$$\frac{d\mathcal{R}(m_1, m_2, \chi_{\text{eff}}, z)}{dm_1 dm_2 d\chi_{\text{eff}}} = \mathcal{R}_0 p(m_1, m_2|z) p(\chi_{\text{eff}}) f(z), \quad (1)$$

where \mathcal{R}_0 is the rate density at redshift $z = 0$, $p(m_1, m_2|z)$ is the two-dimensional source-frame mass distribution at a given redshift, $p(\chi_{\text{eff}})$ is the distribution of effective spins (assumed to be independent of z), and $f(z)$ describes the evolution of the overall merger rate with redshift. The probability distributions p are normalized, so that $p(m_1, m_2|z)$ integrates to unity over the allowed m_1, m_2 (here taken to be $2 M_\odot < m_2 < m_1 < 100 M_\odot$ to match the mass range of the simulated detections used to

⁹ We use the default parameter estimation priors of Abbott et al. (2020a) for each event in Figure 1 and therefore the quantitative details may change within a population analysis; see Figure 2. However, the default priors suffice for this simple visualization.

¹⁰ z_{50} is computed via <https://users.rcc.uchicago.edu/~dholz/gwc/>, an online calculator based on Chen et al. (2021), assuming equal-mass mergers and the “Advanced LIGO mid-high” noise curve, which can be found here: https://dcc.ligo.org/public/0094/P1200087/019/fig1_aligo_sensitivity.txt.

estimate searches' sensitivities), and $p(\chi_{\text{eff}})$ integrates to unity over $-1 < \chi_{\text{eff}} < 1$. The function $f(z)$ is chosen so that $f(z=0)=1$. Thus, integrating the differential merger rate $d\mathcal{R}/dm_1 dm_2 d\chi_{\text{eff}}$ over all masses and spins at a given z gives the overall merger rate density at that redshift. The rate density of Equation (1) can alternatively be written in terms of the number density, defined by the relation:

$$\frac{d\mathcal{R}}{dm_1 dm_2 d\chi_{\text{eff}}} = \frac{dN}{dm_1 dm_2 d\chi_{\text{eff}} dV_c dt_{\text{src}}}, \quad (2)$$

where V_c is the comoving volume element (Hogg 1999) and t_{src} is the time measured in the source frame, so that after integrating over the observing time in the detector frame:

$$\begin{aligned} \frac{dN(m_1, m_2, \chi_{\text{eff}}, z)}{dm_1 dm_2 d\chi_{\text{eff}} dz} &= \frac{dV_c}{dz} \left(\frac{T_{\text{obs}}}{1+z} \right) \\ &\times \frac{d\mathcal{R}}{dm_1 dm_2 d\chi_{\text{eff}}}, \end{aligned} \quad (3)$$

where T_{obs} is the total observing time and the factor of $(1+z)$ converts source-frame time to detector-frame time. Integrating the number density $dN/dm_1 dm_2 d\chi_{\text{eff}} dz$ over all masses, spins and redshifts up to some maximum z_{max} gives the total expected number of BBH mergers in the universe out to z_{max} .

With the factorization of Equation (1), our model for the joint mass–spin–redshift distribution of BBH systems consists of a redshift-dependent mass distribution, a spin distribution, and a rate evolution function. We introduce a parametric model and a set of hyperparameters to describe each of these components, listed in Tables 1 and 2. Exploring the redshift dependence of the mass distribution is the focus of this work, so we consider a few different parametric forms for the mass distribution $p(m_1, m_2|z)$, but fix the parametric form of the spin distribution and rate evolution. For the rate evolution, we assume $f(z) = (1+z)^\kappa$ following Fishbach et al. (2018) and Abbott et al. (2019b, 2020b). For the spin distribution, we assume a truncated Gaussian distribution for the effective spin during the inspiral χ_{eff} , described by a mean μ_{eff} and standard deviation σ_{eff} , and truncated to the physical range $[-1, 1]$ (Roulet & Zaldarriaga 2019; Abbott et al. 2020b; Miller et al. 2020). We ignore other spin degrees of freedom because only χ_{eff} correlates noticeably with our mass and redshift inference (Ng et al. 2018).

We use two different underlying parametric distributions for the component masses. BROKEN POWER LAW, similar to the model defined in Abbott et al. (2020b), models the primary mass distribution as a power law from m_{min} to m_{max} with an additional parameter m_{break} where the power law spectral index changes from α_1 to α_2 :

$$p(m_1) \propto \begin{cases} (m_1/m_{\text{break}})^{\alpha_1} & \text{if } m_{\text{min}} \leq m_1 < m_{\text{break}} \\ (m_1/m_{\text{break}})^{\alpha_2} & \text{if } m_{\text{break}} \leq m_1 < m_{\text{max}} \\ 0 & \text{else} \end{cases} \quad (4)$$

We also modify this model to approximate the TRUNCATED model also defined in Abbott et al. (2020b) by fixing the second power-law exponent $\alpha_2 = -20$, which mimics the hard cutoff of the TRUNCATED model. In our approximated TRUNCATED model, m_{break} becomes the high-mass cutoff, and we fix $m_{\text{max}} = 100 M_\odot$. Equation (4) describes both these models, and specific prior ranges

as well as assumed functional forms of the distributions' evolution with redshift are given in Table 1. For simplicity, we adopt a sharp lower bound on our mass distributions instead of the tapering function employed in Abbott et al. (2020b), as this has no effect on the high-mass inference on which we focus in this work.

In both of these models we describe the conditional mass-ratio distribution with a single power law with index β :

$$\begin{aligned} p(q|m_1) &= m_1 p(m_2|m_1) \\ &= \left(\frac{\beta+1}{1-(m_{\text{min}}/m_1)^{\beta+1}} \right) q^\beta. \end{aligned} \quad (5)$$

Assumed prior ranges for β are also shown in Table 1.

2.2. Statistical Framework

We use hierarchical Bayesian inference to fit these models, marginalizing over individual event properties and the expected number of detections during the observing period (Loredo 2004; Mandel 2010; Mandel et al. 2019). Given data $\{d_i\}$ from N_{det} GW events, we wish to infer the parameters describing our chosen population distributions Λ . Using Bayes' rule, we can write out the full hierarchical posterior distribution as:

$$\begin{aligned} p(\Lambda, \mathcal{R}_0|\{d_i\}) &\propto \mathcal{R}_0^{N_{\text{det}}} e^{-\mathcal{R}_0 \xi(\Lambda)} \\ &\left[\prod_{i=1}^{N_{\text{det}}} \int \mathcal{L}(d_i|m_1, m_2, \chi_{\text{eff}}, z) p(m_1, m_2, \chi_{\text{eff}}, z|\Lambda) dm_1 dm_2 d\chi_{\text{eff}} dz \right] p(\Lambda, \mathcal{R}_0). \end{aligned} \quad (6)$$

Here we define the i th individual event likelihood as $\mathcal{L}(d_i|m_1, m_2, \chi_{\text{eff}}, z)$ and $\xi(\Lambda)$ as the fraction of binary sources we would expect to successfully detect for a given population model defined by the hyperparameters Λ :

$$\xi(\Lambda) = \int p(m_1, m_2, \chi_{\text{eff}}, z|\Lambda) P(\text{det}|m_1, m_2, \chi_{\text{eff}}, z) dm_1 dm_2 d\chi_{\text{eff}} dz, \quad (7)$$

where $P(\text{det}|m_1, m_2, \chi_{\text{eff}}, z)$ is the probability of detecting a single system with parameters $m_1, m_2, \chi_{\text{eff}}$, and z . Assuming a log-uniform prior on \mathcal{R}_0 , we marginalize and write the posterior as:

$$\begin{aligned} p(\Lambda|\{d_i\}) &\propto \frac{p(\Lambda)}{\xi(\Lambda)^{N_{\text{det}}}} \\ &\times \prod_{i=1}^{N_{\text{det}}} \int \mathcal{L}(d_i|m_1, m_2, \chi_{\text{eff}}, z) p(m_1, m_2, \chi_{\text{eff}}, z|\Lambda) dm_1 dm_2 d\chi_{\text{eff}} dz \end{aligned} \quad (8)$$

$$\approx \frac{p(\Lambda)}{\xi(\Lambda)^{N_{\text{det}}}} \prod_{i=1}^{N_{\text{det}}} \left(\frac{1}{K_i} \sum_j \frac{p(m_1^{ij}, m_2^{ij}, \chi_{\text{eff}}^{ij}, z^{ij}|\Lambda)}{\pi(m_1^{ij}, m_2^{ij}, \chi_{\text{eff}}^{ij}, z^{ij})} \right), \quad (9)$$

where we approximate the integrals over single-event parameters via importance sampling with K_i single-event posterior samples generated with the prior $\pi(m_1, m_2, \chi_{\text{eff}}, z)$: $\{m_1^{ij}, m_2^{ij}, \chi_{\text{eff}}^{ij}, z^{ij}\}$ denoting the j th sample for the i th event. The single-event posterior samples for the 44 events used in this analysis are taken from Abbott et al. (2021, 2020c). To compute the sum in Equation (9), we use the same set of samples, derived under the same priors and waveform models, as in Abbott et al. (2020b).

Table 1
Prior Ranges for Mass Models Conditioned on Redshift: $p(m_1, m_2|z) = p(m_2|m_1, z)p(m_1|z)$ Described by Equations (4) and (5)

Name	Nonevolving Parameters	Evolving Parameters	
TRUNCATED	$m_{\min} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$	None	
	$m_{\max} = 100 M_{\odot}$		
	$m_{\text{break}} \sim \mathcal{U}(30 M_{\odot}, 100 M_{\odot})$		
	$\alpha_1 \sim \mathcal{U}(-5, 2)$ $\alpha_2 = -20$ $\beta \sim \mathcal{U}(-4, 12)$		
BINNED EVOLUTION TRUNCATED	$m_{\max} = 100 M_{\odot}$ $\alpha_2 = -20$ $z_0 = 0.4$	$m_{\min}(z) = m_{\min}^{\text{low}} + (m_{\min}^{\text{high}} - m_{\min}^{\text{low}})\Theta(z > z_0)$	$m_{\min}^{\text{low}} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$ $m_{\min}^{\text{high}} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$
		$m_{\text{break}}(z) = m_{\text{break}}^{\text{low}} + (m_{\text{break}}^{\text{high}} - m_{\text{break}}^{\text{low}})\Theta(z > z_0)$	$m_{\text{break}}^{\text{low}} \sim \mathcal{U}(30 M_{\odot}, 100 M_{\odot})$ $m_{\text{break}}^{\text{high}} \sim \mathcal{U}(30 M_{\odot}, 100 M_{\odot})$
		$\alpha_1(z) = \alpha_1^{\text{low}} + (\alpha_1^{\text{high}} - \alpha_1^{\text{low}})\Theta(z > z_0)$	$\alpha_1^{\text{low}} \sim \mathcal{U}(-5, 2)$ $\alpha_1^{\text{high}} \sim \mathcal{U}(-5, 2)$
		$\beta_z(z) = \beta^{\text{low}} + (\beta^{\text{high}} - \beta^{\text{low}})\Theta(z > z_0)$	$\beta^{\text{low}} \sim \mathcal{U}(-4, 12)$ $\beta^{\text{high}} \sim \mathcal{U}(-4, 12)$
EVOLVING TRUNCATED	$m_{\min} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$ $m_{\max} = 100 M_{\odot}$ $\alpha_1 \sim \mathcal{U}(-5, 2)$ $\alpha_2 = -20$ $\beta \sim \mathcal{U}(-4, 12)$	$m_{\text{break}}(z) = m_{\min} + (m_{\max} - m_{\min})b(z)$ $b(z) = (1 + (b_0^{-1} - 1)\exp(-b_1 z))^{-1}$	$b_0 \sim \mathcal{U}(0, 1)$ $b_1 \sim \mathcal{U}(-4, 4)$
BROKEN POWER LAW	$m_{\min} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$ $m_{\max} \sim \mathcal{U}(65 M_{\odot}, 100 M_{\odot})$ $m_{\text{break}} \sim \mathcal{U}(20 M_{\odot}, 65 M_{\odot})$ $\alpha_1 \sim \mathcal{U}(-5, 2)$ $\alpha_2 \sim \mathcal{U}(-12, 2)$ $\beta \sim \mathcal{U}(-4, 12)$	None	
EVOLVING BROKEN POWER LAW	$m_{\min} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$ $m_{\max} \sim \mathcal{U}(65 M_{\odot}, 100 M_{\odot})$ $\alpha_1 \sim \mathcal{U}(-5, 2)$ $\alpha_2 \sim \mathcal{U}(-10, 2)$ $\beta \sim \mathcal{U}(-4, 12)$	$m_{\text{break}}(z) = m_{\min} + (m_{\max} - m_{\min})b(z)$ $b(z) = (1 + (b_0^{-1} - 1)\exp(-b_1 z))^{-1}$	$b_0 \sim \mathcal{U}(0, 1)$ $b_1 \sim \mathcal{U}(-4, 4)$
ALTERNATE EVOLVING BROKEN POWER LAW	$m_{\min} \sim \mathcal{U}(2 M_{\odot}, 10 M_{\odot})$ $m_{\max} \sim \mathcal{U}(65 M_{\odot}, 100 M_{\odot})$ $m_{\text{break}} \sim \mathcal{U}(20 M_{\odot}, 65 M_{\odot})$ $\alpha_1 \sim \mathcal{U}(-5, 2)$ $\beta \sim \mathcal{U}(-4, 12)$	$\alpha_2(z) = \alpha_2^0 + \alpha_2' z$	$\alpha_2^0 \sim \mathcal{U}(-10, 2)$ $\alpha_2' \sim \mathcal{U}(-12, -12)$

Note. For each model considered, we separate the parameters that do not evolve with redshift from those that do, showing the assumed functional forms of the evolution. $\Theta(\cdot)$ represents the Heaviside function. All hyperparameters are drawn from uniform priors between \mathcal{X} and \mathcal{Y} : $\mathcal{U}(\mathcal{X}, \mathcal{Y})$.

We additionally approximate $\xi(\Lambda)$ via importance sampling over sets of detected simulated events,¹¹ marginalizing over our uncertainty from the finite number of simulated events available (Farr 2019).

Given this population likelihood, we sample from the posterior on the population hyperparameters Λ using the Monte Carlo samplers PYMC3 (Salvatier et al. 2016), EMCEE (Foreman-Mackey et al. 2013), and STAN (Carpenter et al. 2017).

Once we have posterior samples Λ , we perform a series of posterior predictive checks. In essence, the posterior predictive

¹¹ The simulated detection sets covering the O3a observing run can be found at <https://dcc.ligo.org/LIGO-P2000217/public>. For the first two observing runs, we use the mock injection sets used in Abbott et al. (2020b), which can be found at <https://dcc.ligo.org/LIGO-P2000434/public>.

Table 2
Prior Ranges Assumed for $f(z)$ and $p(\chi_{\text{eff}})$

$f(z) = (1 + z)^{\kappa}$	$\kappa \sim \mathcal{U}(-6, 6)$
$p(\chi_{\text{eff}}) \propto \exp[-(\chi_{\text{eff}} - \mu^2/2\sigma^2)]$ $\times \Theta(-1 \leq \chi_{\text{eff}} \leq 1)$	$\mu \sim \mathcal{U}(-0.5, 0.5)$ $\sigma \sim \mathcal{U}(0.02, 1)$

Note. Together with one of the models specified in Table 1, these distributions form our population model for the instantaneous rate density in the source frame $d\mathcal{R}/dm_1 dm_2 d\chi_{\text{eff}}$ (Equation (1)). $\Theta(\cdot)$ represents the Heaviside function. All hyperparameters are drawn from uniform priors between \mathcal{X} and \mathcal{Y} : $\mathcal{U}(\mathcal{X}, \mathcal{Y})$.

checks are goodness-of-fit checks that compare the observed data and synthetic events drawn from the fits to the population models. If the models correctly account for all variation within

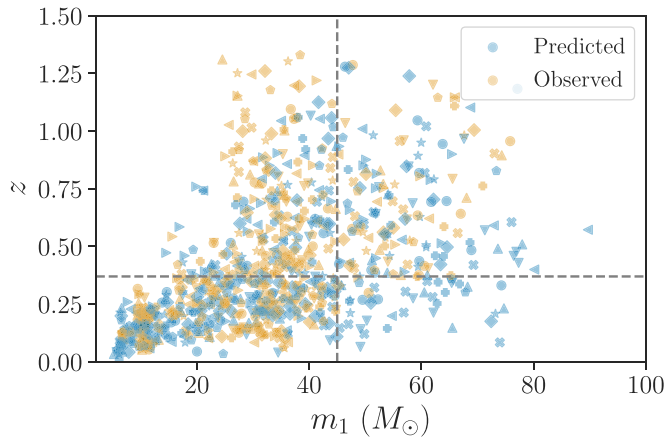


Figure 2. Primary masses and redshifts of the 44 confident BBH observations (orange) compared to 44 draws from the predicted observable distribution (blue), inferred under the nonevolving TRUNCATED model. Each marker shape corresponds to a different set of 44 draws, where each set is inferred under a different population model drawn from the hyperposterior. We plot 10 total sets. The dashed lines denote $z = 0.37$ (the median observed redshift in the sample) and $m_1 = 45 M_\odot$. The top-left corner (low masses, high redshifts) contains no predicted or observed events because it is beyond the detection horizon. On the other hand, the bottom right corner (high masses, low redshifts) contains predicted events, but no observed events, showing that the model generally overpredicts the number of high-mass ($m_1 > 45 M_\odot$) events at low redshifts ($z < 0.37$) compared to our observations.

the observations, synthetic “predicted events” drawn from the population hyperposterior should resemble the observed behavior of real events. The first example of such a check (Figure 2) is introduced in the following Section 3.1.

3. Constraints on the Evolution of the Mass Distribution

We describe the consistency between the data and a series of increasingly complex population models in what follows, beginning with a simple nonevolving TRUNCATED power law in Section 3.1 and then proceeding to more complicated models that do evolve with redshift in Sections 3.2 and 3.3. We also confirm the robustness of our conclusions to a few possible systematics in Sections 3.4 and 3.5.

3.1. Tension with the Nonevolving TRUNCATED Model

To see if the dearth of high-mass events at low redshifts shown in Figure 1 is statistically significant, we first present a posterior predictive check using the nonevolving TRUNCATED model. We use the statistical methods described above to fit the TRUNCATED model (see Table 1) to the GWTC-2 events. Figure 2 shows the observed primary masses and redshifts of BBH events (orange), compared to the prediction from the nonevolving TRUNCATED model (blue). For every hyperposterior sample in the TRUNCATED model, we draw 44 synthetic events from the predicted distribution and record their primary masses and redshifts. We likewise draw one primary mass and redshift sample for each of the 44 events in our catalog, inferred under the same draw from the population hyperposterior. Figure 2 shows this comparison for 10 sets of fair draws from the population hyperposterior, each set marked with a distinct symbol. We note that the GWTC-2 events with $m_1 \gtrsim 45 M_\odot$ occur at redshifts $z \gtrsim 0.37$. The model tends to overpredict the maximum observed mass at low redshift in order to match the maximum observed mass at high redshift. Splitting each set of predicted and observed events at their

median redshifts to define “low” and “high” redshift events, the TRUNCATED model overpredicts the largest mass seen at low redshifts 91% of the time, typically overestimating the observed maximum mass at low redshifts by $18^{+18}_{-22} M_\odot$. On the other hand, at high redshifts, the predicted maximum mass typically matches the observed maximum mass for both models, with an average difference of only $\sim 2 M_\odot$ between the predicted and observed maximum mass for the TRUNCATED model.

This baseline analysis shows that the mismatch between predicted and observed masses and redshifts suggested by Figure 1 also appears when we consider an overly simple population distribution. A similar conclusion regarding the failure of the TRUNCATED model to fit the GWTC-2 data was found by Abbott et al. (2020b), who pointed out the tension between the observed primary mass distribution and the TRUNCATED model prediction. Figure 2 recasts this tension in terms of the joint mass and redshift distribution, and corroborates our expectation from Figure 1 that a redshift-dependent mass distribution may be needed to accurately describe the observed population.

3.2. Two Redshift Bins

To further explore whether the data support a different mass distribution at high redshifts compared to low redshifts, we first perform a change-point analysis. We fit the mass distribution in two redshift bins ($z < 0.4$ and $z > 0.4$), which splits the events roughly evenly between the two bins. We assume that the TRUNCATED model describes the mass distribution in both bins, but we allow the parameters describing the mass distribution to jump discontinuously between the bins. We refer to this model as the BINNED EVOLUTION TRUNCATED model (see Table 1 for prior ranges). Unsurprisingly, we find a strong preference that the maximum mass in the high-redshift bin is larger than the maximum mass in the low-redshift bin (99.4% credibility). This preference remains (95.2% credibility) even when we exclude the most massive event in the high-redshift bin: GW190521 (see Section 3.4). Including GW190521, we infer that the maximum mass in the high-redshift bin is larger by $35^{+17}_{-17} M_\odot$ than the maximum mass in the low-redshift bin. Without GW190521, the high-redshift maximum mass is larger than the low-redshift maximum mass by $16^{+12}_{-15} M_\odot$.

Meanwhile, the other parameters describing the mass distribution are consistent between the two bins, albeit with large uncertainties. For example, the power-law slope of the mass distribution in the high-redshift bin is poorly constrained, because it is degenerate with the redshift evolution of the merger rate. Steep (negative) m_1 power-law slopes correspond to steep (positive) redshift evolution slopes, because steeper mass distributions with fewer high-mass events must have a larger overall merger rate at high redshift to support the number of high-mass events observed at large redshifts (Fishbach et al. 2018). In the following, we focus on the high-mass end of the mass distribution and its possible evolution with redshift.

3.3. Continuous Evolution with Redshift

Motivated by the BINNED EVOLUTION TRUNCATED analysis, we next allow the high-mass end of the mass distribution to evolve continuously and monotonically with redshift. We parameterize the location of the break in the power law as a function of redshift,

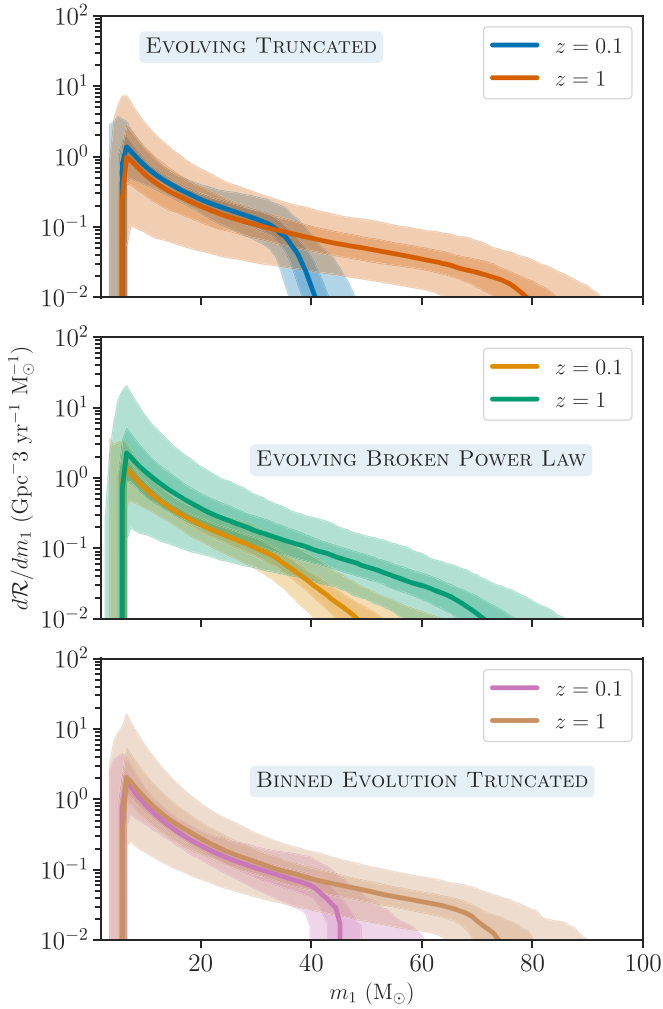


Figure 3. Rate density as a function of m_1 inferred under models that allow the mass distribution and the overall rate to evolve with redshift. Solid lines denote the median $d\mathcal{R}/dm_1$ at each m_1 and shaded regions correspond to the 50% and 90% symmetric credible regions. (Top) A sharp high-mass cutoff at $m_{\text{break}}(z)$, fixing the high-mass ($m_1 > m_{\text{break}}$) power-law slope to $\alpha_2 = -20$. (Middle) Models with variable α_2 and m_{break} . The top and middle panels both assume that only m_{break} evolves with redshift (see Table 1), while the other mass distribution parameters are constant in redshift. (Bottom) The BINNED EVOLUTION TRUNCATED model, in which a separate TRUNCATED mass distribution is fit to systems with $z < 0.4$ and $z > 0.4$, and all mass distribution parameters are allowed to vary between the redshift bins.

allowing it to vary between $2 M_\odot < m_{\text{min}} < m_{\text{break}} < m_{\text{max}} < 100 M_\odot$.

We consider two scenarios for the evolving mass distribution: an EVOLVING TRUNCATED model in which the high-mass slope is fixed to be steep ($\alpha_2 = -20$) to approximate a sharp cutoff that evolves with redshift, and an EVOLVING BROKEN POWER LAW model in which we fit for the high-mass slope, considering values $\alpha_2 > -10$. Table 1 shows the functional form of the models and their full prior ranges. For the EVOLVING TRUNCATED model, $m_{\text{break}}(z)$ parameterizes the location of the cutoff.

Figure 3 shows the rate density as a function of m_1 , $d\mathcal{R}/dm_1$, at two redshifts, $z = 0.1$ and $z = 1$, inferred under the EVOLVING TRUNCATED and EVOLVING BROKEN POWER LAW models described above and the BINNED EVOLUTION TRUNCATED model of Section 3.2. Note that our models allow for an overall evolution of the merger rate through $f(z)$ in Equation (1), and we generally

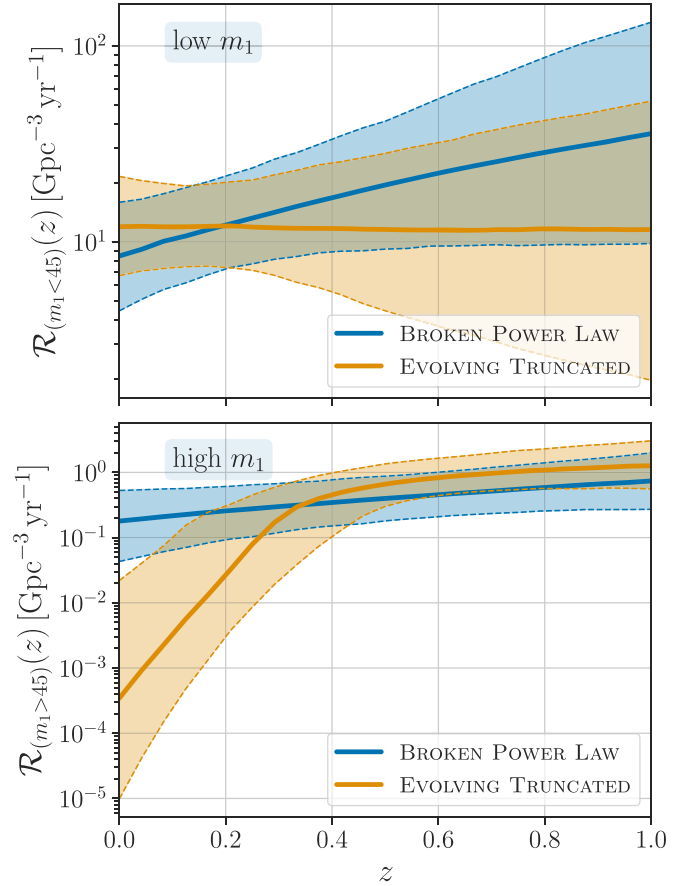


Figure 4. Rate evolution as a function of redshift for BBH systems with (top) $m_1 < 45 M_\odot$ and (bottom) $m_1 > 45 M_\odot$, in the nonevolving BROKEN POWER LAW model (blue) and the EVOLVING TRUNCATED model (orange). If we assume that the mass distribution does not evolve with redshift, the evolution of the merger rate in any given mass range follows $f(z) \propto (1+z)^\kappa$. When we allow the mass distribution to evolve with redshift, the merger rate for BBH systems with $m_1 > 45 M_\odot$ increases with redshift more rapidly, with a small rate $\mathcal{R}_{(m_1 > 45)} < 10^{-2} \text{ Gpc}^{-3} \text{ yr}^{-1}$ at $z = 0$ but a similar rate to the nonevolving mass distribution model at $z \gtrsim 0.3$. Meanwhile, for low-mass events, the evolving mass distribution predicts a slightly smaller merger rate at high z compared to the nonevolving mass distribution.

infer different values for the total merger rate at $z = 1$ and $z = 0.1$. The EVOLVING TRUNCATED and BINNED EVOLUTION TRUNCATED models (first and third panels, respectively) both assume that the primary mass distribution has a sharp maximum mass cutoff; under these models, we infer that the mass distribution extends to higher masses at $z = 1$ than at $z = 0.1$. Meanwhile, the EVOLVING BROKEN POWER LAW model allows for a consistent shape to the mass distribution at $z = 1$ and $z = 0.1$, although evolution toward higher masses at high redshifts is also possible.

Another way of understanding the evolution of the mass distribution is seen in Figure 4, which shows the merger rate as a function of redshift for systems with $m_1 < 45 M_\odot$ (top panel) compared to $m_1 > 45 M_\odot$ (bottom panel). The blue bands show the BROKEN POWER LAW mass distribution, in which the merger rate can evolve with redshift, but the evolution is independent of the masses. With this model, we find that the overall merger rate likely evolves, with rate evolution parameter $\kappa = 2.1^{+2.2}_{-1.9}$ ($\kappa = 0$ corresponds to a nonevolving rate). The orange bands show the EVOLVING TRUNCATED model, in which the mass distribution, as well as the merger rate, can evolve with redshift. This model finds

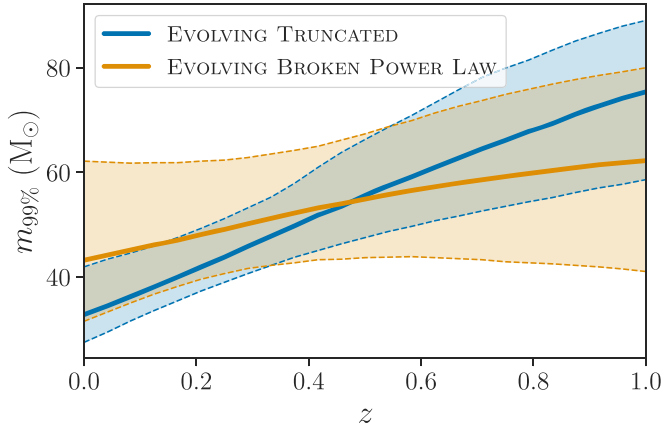


Figure 5. The 99th percentile of the primary mass distribution, $m_{99\%}$, as a function of redshift. Solid lines show the median and shaded bands show symmetric 90% credible intervals. Assuming that the BH mass distribution has a sharp cutoff (EVOLVING TRUNCATED model), it likely increases with redshift ($> 99.9\%$ credibility). Assuming an EVOLVING BROKEN POWER LAW model, there is a weaker preference for the location of the break to evolve (83% credibility).

that the merger rate for systems with $m_1 > 45 M_\odot$ evolves significantly (orange band, bottom panel). At low redshifts, there are few such systems, with a merger rate $\lesssim 0.01 \text{ Gpc}^{-3} \text{ yr}^{-1}$ at $z=0$, but at $z=1$, the merger rate of $m_1 > 45 M_\odot$ systems reaches $\sim 10 \text{ Gpc}^{-3} \text{ yr}^{-1}$. Meanwhile, as seen by the slope of the orange band in the top panel, the merger rate for systems with $m_1 < 45 M_\odot$ might not evolve at all, or may even decrease with increasing redshift. Therefore, the EVOLVING TRUNCATED model infers less evolution in the total merger rate, finding smaller values of $\kappa = 0.0^{+2.4}_{-2.6}$, compared to the BROKEN POWER LAW model. The correlation between the evolution of the total merger rate and the evolution of the mass distribution is to be expected; see the discussion in Section 3.2. In order to explain the number of observations at high redshifts, we infer either that the total merger rate must increase with redshift, or that the fraction of high-mass events must increase while the total merger rate stays approximately constant.

For the EVOLVING TRUNCATED and EVOLVING BROKEN POWER LAW models, we show the inferred 99th percentile of the m_1 distribution, $m_{99\%}$, as a function of redshift in Figure 5. As we saw in Figures 3 and 4, the EVOLVING TRUNCATED model finds a strong preference for the high-mass cutoff to increase with increasing redshift ($> 99.9\%$). On the other hand, if the primary mass distribution follows an EVOLVING BROKEN POWER LAW model, the preference for mass evolution is much weaker. Marginalizing over the high-mass power-law slope α_2 , we find that m_{break} increases with increasing redshift at 83% credibility. We stress that even this mild preference for evolution in the EVOLVING BROKEN POWER LAW model depends on the choice of α_2 prior, because of the correlation between the steepness of α_2 and the evolution of m_{break} . This degeneracy between the abruptness of the high-mass cutoff and the preference for evolution can be seen in Figure 6, which shows the correlation between the evolution of m_{break} and α_2 , the power-law slope above m_{break} . In the limit of large negative $\alpha_2 \lesssim -8$, the EVOLVING BROKEN POWER LAW model approaches an EVOLVING TRUNCATED model, and we find a strong preference for m_{break} to evolve. Meanwhile, for shallower values of $-6 \lesssim \alpha_2 \lesssim -4$, the data is consistent with $m_{\text{break}}(z=1) = m_{\text{break}}(z=0)$. Our final posterior, then, depends on how much prior volume we include below $\alpha_2 \lesssim 8$.

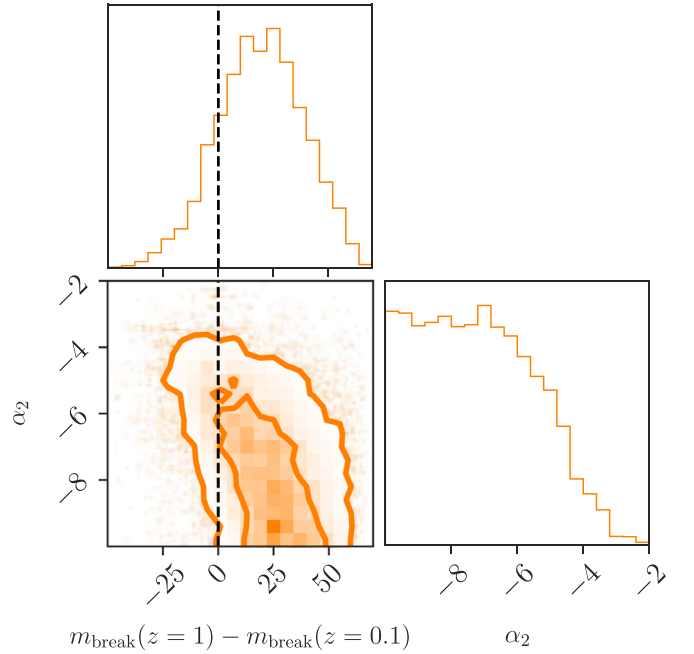


Figure 6. Corner plot showing the correlation between the evolution of m_{break} in the EVOLVING BROKEN POWER LAW model and the high-mass power-law slope α_2 . The vertical dashed line corresponds to no evolution of the mass distribution. For a sharp cutoff $\alpha_2 \ll 0$, we find a strong preference for evolution, but for a shallower cutoff ($\alpha_2 \gtrsim -6$), the observed events are consistent with no evolution.

While we have chosen to parameterize evolution in the EVOLVING BROKEN POWER LAW model with a redshift-dependent m_{break} parameter, we find similar results when we instead consider a redshift-dependent $\alpha_2(z)$. This is the ALTERNATE EVOLVING BROKEN POWER LAW described in Table 1.

This alternate parameterization may approximate the scenario in which the break mass m_{break} represents the lower edge of the pair-instability gap (fixed across cosmic time), so that systems with $m_1 > m_{\text{break}}$ belong to a subpopulation that contaminates the gap. In this scenario, α_2 sets the rate of this subpopulation, which can become more (less negative α_2) or less (more negative α_2) dominant at different redshifts. Figure 7 shows the posterior on $\alpha_2(z=0)$ and $\alpha_2(z=1)$ under this model, marginalizing over the other model parameters. As in Figure 6, we infer that if the $z=0$ mass distribution has a sharp cutoff ($\alpha_2^0 \lesssim -8$), α_2 probably becomes less negative with increasing redshift, corresponding to a higher merger rate for systems with $m_1 > m_{\text{break}}$. However, the data remain consistent with a mass distribution that is independent of redshift as long as $\alpha_2 \gtrsim -6$. Additionally, we recover similar results for the 99th percentile of the mass distribution as a function of redshift.

3.4. Sensitivity to GW190521

Since GW190521 is the most massive event in our catalog, one might suspect that our conclusions regarding the high-mass end of the BBH population are driven by this event. To test this, we repeat our analyses while excluding GW190521 from the sample, and we verify that our results are robust in these leave-one-out analyses. If we model the mass distribution with a sharp maximum mass cutoff and exclude GW190521 from the analysis, we recover that the location of the cutoff must

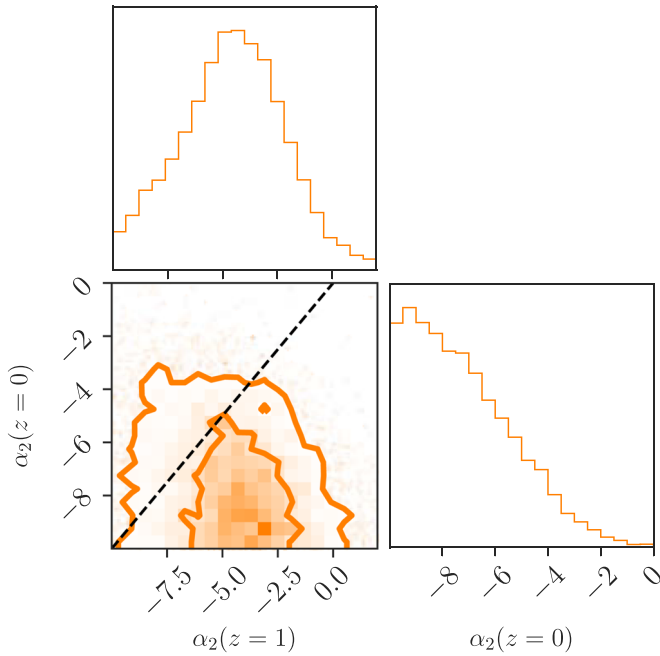


Figure 7. Corner plot showing the correlation between $\alpha_2(z=1)$ and $\alpha_2(z=0)$ inferred under an ALTERNATE EVOLVING BROKEN POWER LAW model in which α_2 evolves with redshift. The dashed black line shows $\alpha_2(z=1) = \alpha_2(z=0)$, corresponding to no redshift evolution. Similarly to the model with an evolving m_{break} , when $\alpha_2(z=0)$ is steep, we recover a preference for the mass distribution to evolve.

evolve with redshift (96.4% credibility in the EVOLVING TRUNCATED model; 95.2% credibility in the BINNED EVOLUTION TRUNCATED model). For the EVOLVING BROKEN POWER LAW model, any preference for evolution of the mass distribution is slightly weakened with the exclusion of GW190521, and our conclusion remains that the data is consistent with a nonevolving mass distribution.

3.5. Robustness to Sensitivity Estimate

As described in Section 2, when fitting the population models to the data, we must account for GW selection effects. For a given population model described by Λ , we estimate the detectable fraction $\xi(\Lambda)$ of Equation (7) by performing a Monte Carlo integral over a set of injections detected by the GWTC-2 search pipelines, reweighted by the population model at Λ (Farr 2019). To be detected, an injection must have a FAR of less than one per year, matching the FAR threshold we use when selecting events for our analysis. Crucially, if there were a systematic error in the injections' estimated FARs that was correlated in source-frame mass and redshift, the mass distribution inference could artificially prefer redshift evolution. In particular, an apparent dearth of sources with high mass and low redshift could be explained as an *overestimation* of the sensitivity to high-mass, low-redshift systems.

As a crude test of the effect of misestimated sensitivity, we perform multiple population inferences on O3a events with a modified TRUNCATED primary-mass model that allows linear evolution of the maximum mass and power-law slope with redshift. In each run, we throw away different fractions of found injections at high total mass ($M > 100 M_\odot$) and low redshift ($z < 0.4$), going all the way up to throwing away 90% of those injections. This simulates selection functions with ever smaller low-redshift, high-mass sensitivities. We find that

tossing out injections does not eliminate the preference for mass distribution evolution with this TRUNCATED model, suggesting that sensitivity misestimation is likely not driving our conclusions that the data prefer a mass distribution that either evolves with redshift or has additional features beyond a simple truncated power law.

Other systematic uncertainties that may lead to spurious conclusions about the evolution of the mass distribution include the possibility of strongly lensed GW events in the sample (Dai et al. 2017) or deviations from the assumed cosmology (Farr et al. 2019). We do not account for these possibilities here. The probability that our sample contains one or more strongly lensed GW events is very small, as only $\sim 1/1000$ events are expected to be lensed (Li et al. 2018; Oguri 2018). For our analyses, we assume the cosmological parameters from Planck Collaboration et al. (2016) for consistency with Abbott et al. (2020a). In principle the cosmological parameters can be simultaneously inferred with the mass distribution (Farr et al. 2019).

4. Posterior Predictive Checks

We find that the data is consistent with two interpretations: a nonevolving BROKEN POWER LAW with a relatively shallow high-mass slope, or an EVOLVING TRUNCATED model. In this section, we carry out posterior predictive checks to examine the features of the data that are most consistent with each interpretation, and discuss how future data will distinguish between the two scenarios.

As a first check, we revisit the feature highlighted in Figures 1 and 2: the missing high-mass, low-redshift observations in GWTC-2. Fundamentally, consistency between observations and our models boils down to the relative fractions of events detected with high masses ($m_1 \gtrsim 45 M_\odot$) at low and high redshift. The left panel of Figure 8 shows the uncertainty in this ratio under the nonevolving BROKEN POWER LAW and EVOLVING TRUNCATED models. The BROKEN POWER LAW model generally predicts a few low-redshift events with $m_1 > 45 M_\odot$ while the EVOLVING TRUNCATED model predicts fewer, or even zero, such events. In principle, then, we should be able to distinguish between these two models given enough observations. That is to say, we will prefer the nonevolving BROKEN POWER LAW if we see more than one high-mass, low-redshift event for every ~ 5 high-mass, high-redshift events. If we see fewer high-mass, low-redshift events, the EVOLVING TRUNCATED model will be preferred.

However, Poisson uncertainty with the current limited set of events is large enough that we do not strongly favor either interpretation. The right panel of Figure 8 shows the distribution of the number of detected events with $m_1 > 45 M_\odot$ at low redshift out of 44 events under each model. We find that even without redshift evolution, the BROKEN POWER LAW model predicts zero low-redshift, high-mass events out of 44 detections 15% of the time, and ≤ 1 such events 38% of the time. Both models, then, are consistent with the current absence of detections.

While both evolving and nonevolving models are consistent with our current (lack of) observations of high-mass, low-redshift events, we additionally test the general goodness-of-fit of the entire mass distributions. Figures 9 and 10 demonstrate this second check.

Specifically, Figure 9 demonstrates the consistency of the inferred population model with the observations by drawing 44 predicted events from the population hyperposterior, sorting them, and then comparing the predicted ($m_1^{(\text{pred})}$) and observed

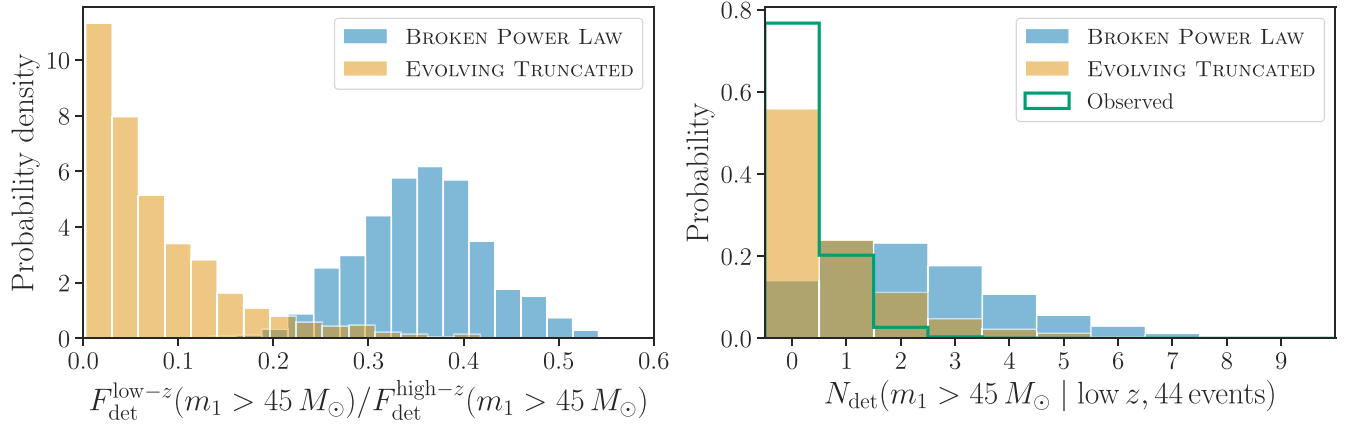


Figure 8. (Left) Ratio of the detection rate of high-mass events ($m_1 > 45 M_\odot$) between high-redshift and low-redshift events (split at the median redshift). Model predictions refer to the true detection rate and do not account for Poisson uncertainty that arises for a finite number of observations. (Right) The expected number of low-redshift observations with $m_1 > 45 M_\odot$ given a total of 44 events. The unfilled green histogram corresponds to the GWTC-2 events; the uncertainty reflects measurement uncertainty in their source parameters inferred under the BROKEN POWER LAW model. Both models are consistent with GWTC-2’s observation of nearly zero low-redshift, $m_1 > 45 M_\odot$ events.

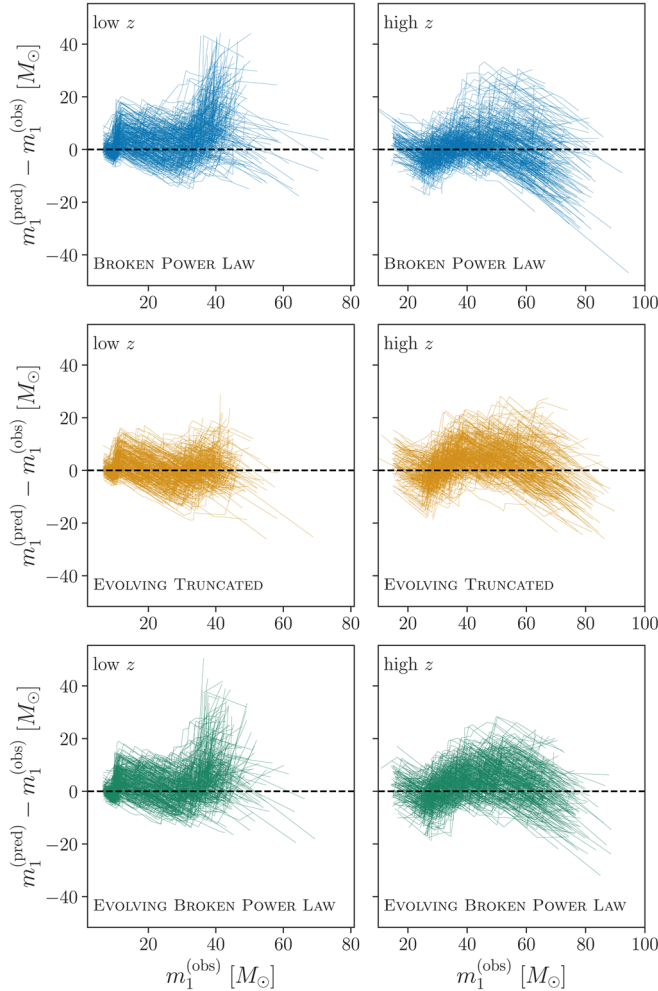


Figure 9. Difference between sorted sets of predicted ($m_1^{(\text{pred})}$) and observed ($m_1^{(\text{obs})}$) primary masses as a function of the observed primary mass for (top) BROKEN POWER LAW, (middle) EVOLVING TRUNCATED, and (bottom) EVOLVING BROKEN POWER LAW models. Events are separated into (left) low-redshift and (right) high-redshift subsets. Each line represents a different realization of predicted and observed masses drawn from the corresponding hyperposterior.

($m_1^{(\text{obs})}$) masses as a function of $m_1^{(\text{obs})}$. As in the earlier posterior predictive checks, for every population hyperposterior sample, we draw one $m_1^{(\text{obs})}$ sample per event from its population-reweighted posterior. If the models capture the behavior of the data well, then the difference $m_1^{(\text{pred})} - m_1^{(\text{obs})}$ should remain near zero for all masses, regardless of the events’ redshifts. We demonstrate this by dividing the sample into low- and high-redshift bins. The largest deviations occur for the nonevolving BROKEN POWER LAW model at low redshift (top-left panel), where the model tends to overpredict the largest observed m_1 . The EVOLVING TRUNCATED model appears to be a better fit at low z (middle-left panel), but there may be hints of a deviation at high z (middle-right panel) where the model tends to systematically overpredict the masses of the $m_1 \sim 40 M_\odot$ events.

Figure 10 provides a complementary view, showing the uncertainty in the predicted cumulative distribution functions (CDFs) over m_1 , calculated from drawing sets of 44 predicted events from the population hyperposterior, along with the empirical cumulative distribution of observed m_1 , again binned into low- and high-redshift sets. Consistency in these plots corresponds to predicted CDF bands that encompass the black uncertainty bars from the individual events. At low z , all of the models shown are able to fit the events well, although, again, the EVOLVING TRUNCATED and EVOLVING BROKEN POWER LAW models better limit the maximum predicted m_1 at low redshifts to the observed value, while the BROKEN POWER LAW model often overpredicts the most massive observation at low z . At high z , we see that the EVOLVING TRUNCATED model tends to predict a longer tail to high masses, as the predicted CDFs are slightly shifted to the right compared to the observed high- z events. Although the BROKEN POWER LAW, EVOLVING TRUNCATED, and EVOLVING BROKEN POWER LAW models differ in their predictions, they all provide adequate fits to the data within the current uncertainties.

5. Conclusion and Future Prospects

We have fit the GWTC-2 events to redshift-dependent mass models, investigating the evolution of the BBH mass distribution across cosmic time. We explored the apparent dearth of

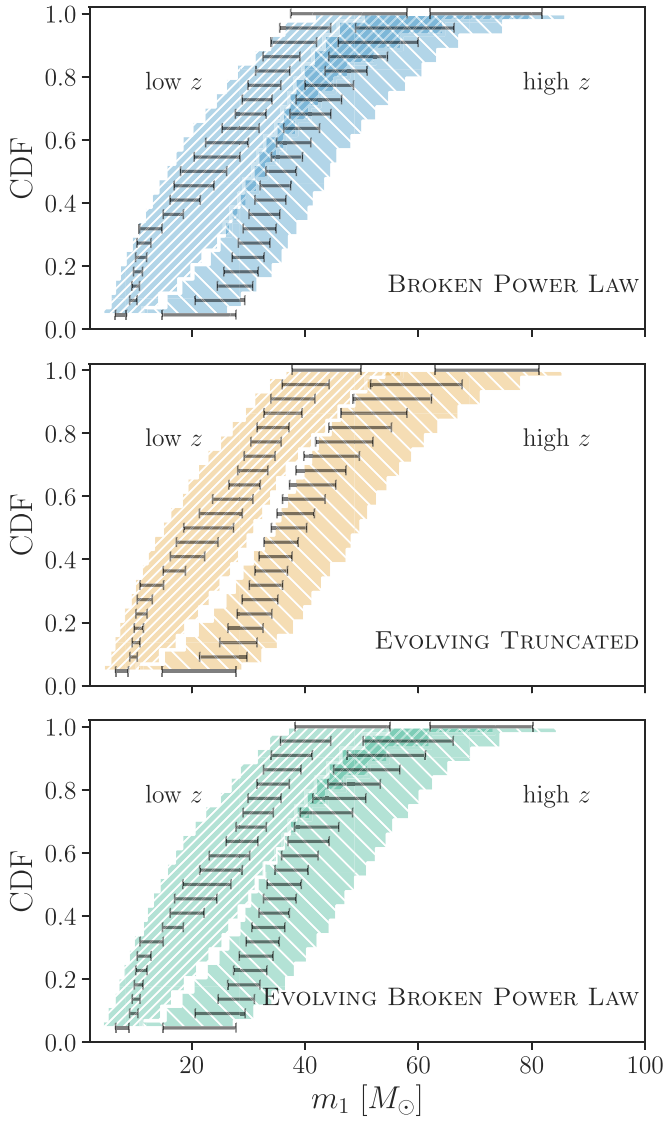


Figure 10. Comparison between 90% symmetric credible regions for the predicted cumulative distribution function (shaded bands) and the empirical cumulative distribution (black bars) of primary masses for (top) BROKEN POWER LAW, (middle) EVOLVING TRUNCATED, and (bottom) EVOLVING BROKEN POWER LAW models. Both predicted and observed events are separated into low-redshift and high-redshift bins, split at the median redshift.

high-mass BH mergers at low redshift, showing that it can be explained by either a BBH mass distribution that evolves with redshift or by a mass distribution that contains features beyond a truncated power law. In a nonevolving mass distribution, these beyond-power-law features must suppress the rate of high-mass systems ($m_1 > 45 M_\odot$) by including, for example, a break in the power law, in agreement with the conclusions of Abbott et al. (2020b).

We additionally confirmed that our conclusions are not driven by any particular event, showing that the results are qualitatively unchanged when we remove the heaviest system detected to date: GW190521. At the same time, we confirmed that possible gross misestimation of the sensitivity of our searches, which could systematically bias our belief about our ability to detect high-mass, low-redshift events, could not account for the preference for primary mass distributions with a sharp maximum cutoff to evolve with redshift.

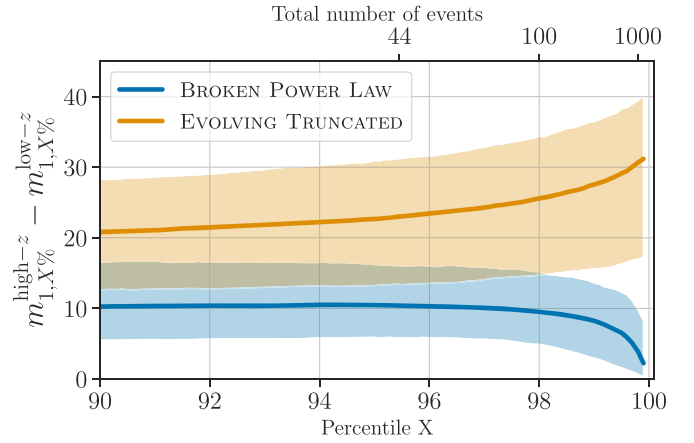


Figure 11. Differences in the primary-mass percentiles between detected events at high and low redshift as a function of the percentile. Shaded regions correspond to 90% uncertainty from the hyperposterior for the BROKEN POWER LAW (blue) and EVOLVING TRUNCATED (orange) models. The approximate total number of events needed to resolve each quantile is shown by the top x-axis. The current set of events only probes up to percentiles $\approx 95\%$, and more events will be needed in order to distinguish between the models.

If a TRUNCATED power law with a sharp maximum mass cutoff is assumed for the primary masses of BBH systems, we find that the mass distribution must evolve between $z = 0$ and $z = 1$ with $> 99\%$ credibility. In this case, the total merger rate is consistent with binaries being uniformly distributed in a comoving volume, but the types of binaries must change as a function of redshift. If, on the other hand, the data is described by a power law with a *break* rather than a sharp maximum mass cutoff, the data are consistent with a nonevolving mass distribution. This model prefers an overall rate density that increases with increasing redshift, rather than being uniformly distributed in a comoving volume. Both models fit the current data equally well, but with additional events, we will be able to distinguish them.

Figure 11 explores when we expect to be able to distinguish between an evolving mass distribution with a sharp maximum mass cutoff and a nonevolving mass distribution with a break in the power law, rather than a cutoff. We plot the difference in the mass scales corresponding to the percentiles of detected events in high-redshift and low-redshift bins. As we observe more events, we begin to resolve higher percentiles in the observed mass distribution; out of $N/2$ events in each redshift bin, we expect the most massive observed event in each bin to be at the $\sim(1 - \frac{2}{N})$ quantile. Since we can see more massive events out to higher redshifts, even if the underlying mass distribution is the same at all redshifts, the $X\%$ primary-mass percentile in the high-redshift bin, $m_{1,X\%}^{\text{high-}z}$, will correspond to a larger mass than the same percentile in the low-redshift bin, $m_{1,X\%}^{\text{low-}z}$. Therefore, for a nonevolving mass distribution, or one that favors larger masses at higher redshifts, the difference $m_{1,X\%}^{\text{high-}z} - m_{1,X\%}^{\text{low-}z}$ is positive. Shaded regions correspond to the uncertainty from different fair draws from the hyperposterior for each model. As we start probing higher percentiles of the observed mass distribution, if the mass distribution does not evolve with redshift (BROKEN POWER LAW model, in blue), the heaviest events observed in the low-redshift bin approach the masses of the heaviest events in the high-redshift bin, and so the difference decreases. However, if the mass distribution evolves with redshift (EVOLVING TRUNCATED model, in

orange), the underlying *astrophysical* mass distribution skews to higher masses at higher redshifts, and so the difference in masses among the *observed* high- and low-redshift events is larger than in the nonevolving case, and only increases as we consider higher percentiles of the observed distribution. Within current statistical uncertainties, the BROKEN POWER LAW and EVOLVING TRUNCATED models are consistent for small percentiles (< 95 th percentile) but diverge at higher percentiles. The current data set only probes approximately the $1 - \frac{2}{44} \approx 0.95$ quantile, as we have $N=44$ total events, or two redshift bins that each contain 22 events. Therefore, we cannot resolve the discrepancy between the models that only appears at higher percentiles with the current sample size. However, we will probe higher percentiles as we detect more events, and may be able to confidently distinguish between these models when we obtain a factor of $\gtrsim 2$ more events.

Future GW events will allow us to not only better resolve the BBH mass distribution, they will also probe the BBH population over a much higher redshift range. In addition to the expected increases in sensitivity for current detectors (Abbott et al. 2018), the next generation of GW detectors will be able to probe the redshift evolution of the BBH population out to $z \sim 30$ (Hall & Evans 2019), offering us a deeper look into the formation environments of BBH systems (Vitale et al. 2019; Ng et al. 2020; Romero-Shaw et al. 2020).

Characterizing features in the BBH mass distribution, such as a sharp cutoff, a break or a peak, and tracing their evolution with redshift provides insights into the physics of BBH formation and merger. Future data will allow us to distinguish between several possible scenarios. One possibility is that the maximum core mass for pair-instability supernovae (PISN) is highly sensitive to metallicity, creating a redshift-dependent cutoff in the BH mass distribution. It may also be that multiple binary formation channels contribute differently over the age of the universe such that higher-mass mergers are favored at earlier times. For example, some combination of stellar evolution, PISN physics, and/or hierarchical or stellar mergers may produce a small tail of BBHs with primary masses $\gtrsim 45 M_{\odot}$, and these processes may be more common at higher redshift. Alternatively, there may be no redshift evolution of the mass distribution at $0 < z \lesssim 1$, and instead the overall rate of mergers increases with redshift independently of the component masses. In this case, it must be that the PISN feature in the BBH mass distribution is not a sharp cutoff, or that the processes that contaminate the PISN mass gap and give rise to systems above $m_1 \gtrsim 45 M_{\odot}$ operate at similar relative rates throughout the observable redshift range.



The first three observing runs of the Advanced LIGO and Virgo interferometers have already illuminated many new aspects of compact objects within our universe. The first detections demonstrated *what* the most prevalent detectable GW sources are: they are BBH mergers. Subsequent studies have then asked *where* the biggest BHs are, motivated by the lack of observations of massive BBH during the first two LIGO/Virgo observing runs (Fishbach & Holz 2017). However, we can now phrase the question more precisely as *when* did the largest BBH merge in the history of the universe, which will ultimately help us determine *how* BHs form and *why* they merge.

We thank Daniel Wysocki for providing helpful comments on the manuscript. M.F. is supported by NASA through NASA Hubble Fellowship grant HST-HF2-51455.001-A awarded by

the Space Telescope Science Institute. R.E. and D.E.H. are supported at the University of Chicago by the Kavli Institute for Cosmological Physics through an endowment from the Kavli Foundation and its founder Fred Kavli. D.E.H. is also supported by NSF grants PHY-1708081 and PHY-2011997, and gratefully acknowledges the Marion and Stuart Rice Award. Z.D. was in part supported by NSF Graduate Research Fellowship grant DGE-1144082. B.E. and B.F. are supported by NSF grant PHY-1807046. The Flatiron Institute is supported by the Simons Foundation. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation grants PHY-0757058 and PHY-0823459. This research was conducted in part at the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara and was supported in part by the National Science Foundation under grant No. NSF PHY-1748958. 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.

Software: ASTROPY (Astropy Collaboration et al. 2018), CORNER (Foreman-Mackey 2016), EMCEE (Foreman-Mackey et al. 2013), NUMPY (Harris et al. 2020), MATPLOTLIB (Hunter 2007), PESUMMARY (Hoy & Raymond 2020), PYMC3 (Salvatier et al. 2016), PYSTAN (Carpenter et al. 2017), SCIPY (Virtanen et al. 2020), SEABORN (Waskom 2020), THEANO (Theano Development Team 2016).

ORCID iDs

Maya Fishbach  <https://orcid.org/0000-0002-1980-5293>
 Zoheyr Doctor  <https://orcid.org/0000-0002-2077-4914>
 Thomas Callister  <https://orcid.org/0000-0001-9892-177X>
 Bruce Edelman  <https://orcid.org/0000-0001-7648-1689>
 Reed Essick  <https://orcid.org/0000-0001-8196-9267>
 Will M. Farr  <https://orcid.org/0000-0003-1540-8562>
 Ben Farr  <https://orcid.org/0000-0002-2916-9200>
 Daniel E. Holz  <https://orcid.org/0000-0002-0175-5064>

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, *LRR*, **21**, 3
- Abbott, B. P., Abbott, R., Abbott, T. D., 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. 2018, *LRR*, **21**, 3
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020a, arXiv:2010.14527
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020b, arXiv:2010.14533
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2020c, GWTC-2 Data Release: Parameter Estimation Samples and Skymaps, <https://dcc.ligo.org/LIGO-P2000223/public>
- Abbott, R., Abbott, T. D., Abraham, S., et al. 2021, *SoftX*, **13**, 100658
- Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *CQGra*, **32**, 024001
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Belczynski, K., Dominik, M., Bulik, T., et al. 2010, *ApJL*, **715**, L138
- Brott, I., de Mink, S. E., Cantiello, M., et al. 2011, *A&A*, **530**, A115
- Callister, T., Fishbach, M., Holz, D. E., & Farr, W. M. 2020, *ApJL*, **896**, L32
- Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, *J. Stat. Softw.*, **76**, 1
- Chen, H.-Y., Holz, D. E., Miller, J., et al. 2021, *CQG*, **38**, 055010
- Dai, L., Venumadhav, T., & Sigurdson, K. 2017, *PhRvD*, **95**, 044011
- Dominik, M., Berti, E., O’Shaughnessy, R., et al. 2015, *ApJ*, **806**, 263
- El-Badry, K., Quataert, E., Weisz, D. R., Choksi, N., & Boylan-Kolchin, M. 2019, *MNRAS*, **482**, 4528
- Farr, W. M. 2019, *RNAAS*, **3**, 66
- Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, *ApJL*, **883**, L42
- Farrell, E. J., Groh, J. H., Hirschi, R., et al. 2021, *MNRAS*, **502**, L40
- Fishbach, M., & Holz, D. E. 2017, *ApJL*, **851**, L25

- Fishbach, M., Holz, D. E., & Farr, W. M. 2018, *ApJL*, **863**, L41
- Foreman-Mackey, D. 2016, *JOSS*, **1**, 24
- Foreman-Mackey, D., Conley, A., Meierjürgen Farr, W., et al. 2013, emcee: The MCMC Hammer, Astrophysics Source Code Library, ascl:1303.002
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, *ApJ*, **749**, 91
- Hall, E. D., & Evans, M. 2019, *CQGra*, **36**, 225002
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hogg, D. W. 1999, arXiv:astro-ph/9905116
- Hoy, C., & Raymond, V. 2020, arXiv:2006.06639
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Kinugawa, T., Nakamura, T., & Nakano, H. 2021, *MNRAS*, **501**, L49
- Kudritzki, R.-P., & Puls, J. 2000, *ARA&A*, **38**, 613
- Kushnir, D., Zaldarriaga, M., Kollmeier, J. A., & Waldman, R. 2016, *MNRAS*, **462**, 844
- Li, S.-S., Mao, S., Zhao, Y., & Lu, Y. 2018, *MNRAS*, **476**, 2220
- LIGO Scientific Collaboration, Aasi, J., Abbott, B. P., et al. 2015, *CQGra*, **32**, 074001
- Loredo, T. J. 2004, in AIP Conf. Ser. 735, 24th Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (Melville, NY: AIP), 195
- Mandel, I. 2010, *PhRvD*, **81**, 084029
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, **486**, 1086
- Mapelli, M., Giacobbo, N., Santoliquido, F., & Artale, M. C. 2019, *MNRAS*, **487**, 2
- Miller, S., Callister, T. A., & Farr, W. M. 2020, *ApJ*, **895**, 128
- Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., et al. 2019, *MNRAS*, **490**, 3740
- Ng, K. K., Vitale, S., Farr, W. M., & Rodriguez, C. L. 2020, arXiv:2012.09876
- Ng, K. K. Y., Vitale, S., Zimmerman, A., et al. 2018, *PhRvD*, **98**, 083007
- Oguri, M. 2018, *MNRAS*, **480**, 3842
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Rodriguez, C. L., & Loeb, A. 2018, *ApJL*, **866**, L5
- Rodriguez, C. L., Zevin, M., Amaro-Seoane, P., et al. 2019, *PhRvD*, **100**, 043027
- Romero-Shaw, I. M., Kremer, K., Lasky, P. D., Thrane, E., & Samsing, J. 2020, arXiv:2011.14541
- Roulet, J., Venumadhav, T., Zackay, B., Dai, L., & Zaldarriaga, M. 2020, *PhRvD*, **102**, 123022
- Roulet, J., & Zaldarriaga, M. 2019, *MNRAS*, **484**, 4216
- Safarzadeh, M., & Farr, W. M. 2019, *ApJL*, **883**, L24
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PyMC3: Python probabilistic programming framework, Astrophysics Source Code Library, ascl:1610.016
- Samsing, J. 2018, *PhRvD*, **97**, 103014
- Santoliquido, F., Mapelli, M., Bouffanais, Y., et al. 2020, *ApJ*, **898**, 152
- Theano Development Team 2016, arXiv:1605.02688
- Tiwari, V. 2020, arXiv:2012.08839
- Vink, J. S., Higgins, E. R., Sander, A. A. C., & Sabhahit, G. N. 2021, *MNRAS*, **504**, 146
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Vitale, S., Farr, W. M., Ng, K. K. Y., & Rodriguez, C. L. 2019, *ApJL*, **886**, L1
- Waskom, M. 2020, seaborn: statistical data visualization, v0.11.1, Zenodo, doi:10.5281/zenodo.592845
- Weatherford, N. C., Fragione, G., Kremer, K., et al. 2021, *ApJL*, **907**, L25
- Yang, Y., Bartos, I., Haiman, Z., et al. 2020, *ApJ*, **896**, 138
- Zevin, M., Bavera, S. S., Berry, C. P. L., et al. 2021, *ApJ*, **910**, 152