MNRAS 000, 1–10 (2023) Preprint 28 July 2023 Compiled using MNRAS LATEX style file v3.0

Is there an excess of black holes around $20M_\odot$? Optimising the complexity of population models with the use of reversible jump MCMC.

A. Toubiana,[1]★ Michael L. Katz,[1] Jonathan R. Gair[1]
[1]*Max Planck Institute for Gravitationsphysik (Albert Einstein Institute), Am Mühlenberg 1, 14476 Potsdam, Germany*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Some analyses of the third gravitational wave catalogue released by the LIGO-Virgo-KAGRA collaboration (LVK) suggest an excess of black holes around $15-20M_\odot$. In order to investigate this feature, we introduce two flexible population models, a semi-parametric one and a non-parametric one. Both make use of reversible jump Markov chain Monte-Carlo to optimise their complexity. We also illustrate how the latter can be used to efficiently perform model selection. Our parametric model broadly agrees with the fiducial analysis of the LVK, but finds a peak of events at slightly larger masses. Our non-parametric model shows this same displacement. Moreover, it also suggests the existence of an excess of black holes around $20M_\odot$. We assess the robustness of this prediction by performing mock injections and running simplified hierarchical analyses on those (i.e. without selection effects and observational uncertainties). We estimate that such a feature might be due to statistical fluctuations, given the small number of events observed so far, with a 5% probability. We estimate that with a few hundreds of observations, as expected for O4, our non-parametric model will be able to robustly determine the presence of this excess. It will then allow for an efficient agnostic inference of the properties of black holes.

Key words: black hole physics – gravitational waves – methods: statistical

1 INTRODUCTION

The total number of gravitational wave (GW) observations from compact binaries by the LIGO-Virgo-KAGRA (LVK) collaboration Aasi et al. (2015); Acernese et al. (2015); Akutsu et al. (2021) now adds up to 90 Abbott et al. (2021a). Together they form the third GW transients catalogue (GWTC-3) released by the LVK. As the number of observation increases, we become able to infer the astrophysical properties of GW sources not only individually but as a population. Different astrophysical models for the formation and evolution mechanisms of GW sources predict different distributions of parameters such as masses, spins or redshifts Benacquista & Downing (2013); Postnov & Yungelson (2014); de Mink & Mandel (2016); Samsing & Ramirez-Ruiz (2017); Gerosa et al. (2018); Tagawa et al. (2020); Sedda et al. (2023); Giacobbo & Mapelli (2018); Wiktorowicz et al. (2019); van Son et al. (2022); Romero-Shaw et al. (2021); Mapelli (2021); Bouffanais et al. (2021); Inayoshi et al. (2017). Thus, inferring the population of GW sources from data is a powerful tool to constrain astrophysical models.

Focusing on binary black holes (BBHs), one of the main results of population analyses is the growing evidence for an excess of black holes (BHs) around $35M_\odot$ Abbott et al. (2021b, 2023); Tiwari (2021, 2022, 2023); Edelman et al. (2022a); Farah et al. (2023); Sadiq et al. (2022); Ruhe et al. (2022); Callister & Farr (2023). This excess can be interpreted as a "pile-up" of BHs before the upper mass gap Woosley (2017); Woosley et al. (2002); Farmer et al. (2019); Talbot & Thrane (2018), although it happens at lower masses than predicted by current astrophysical models Belczynski et al. (2016); Marchant et al. (2018); Renzo et al. (2020); Farag et al. (2022). Moreover, first hints of correlations between parameters have started to appear Hoy et al. (2022); Callister et al. (2021); Fishbach et al. (2021); Adamcewicz & Thrane (2022); Bavera et al. (2022); Biscoveanu et al. (2022), e.g., between the effective spin and the mass ratio or between redshift and spins. Such correlations carry the signature of the astrophysical channels through which the binaries form. However, some of these were not anticipated by astrophysical models, illustrating once more how population analysis can shed light on astrophysics.

Another interesting feature is a possible excess of BHs around $15-20M_\odot$ Abbott et al. (2023); Edelman et al. (2022a); Tiwari (2022, 2023). It has been pointed out that it could be the signature of second generation mergers Tiwari & Fairhurst (2021); Tiwari (2022); Mahapatra et al. (2022). However, it is statistically less significant than the feature at around $35M_\odot$ Farah et al. (2023); Tiwari (2023) and has been found by only a few of the analyses performed on GWTC-3. Assessing the significance of this excess was the first motivation for our study.

From a Bayesian perspective, the goal of population inference Mandel et al. (2019); Vitale et al. (2020) is to obtain the posterior distribution on population hyperparameters (denoted Λ) assuming a population prior for the individual event parameters $p(\Theta|\Lambda)$. The event parameters of interest, Θ, are typically the masses, the spins and the redshift of the source. Broadly speaking, we can identify three main approaches, differing in how $p(\Theta|\Lambda)$ is modelled:

- astrophysical: the distribution of parameters is obtained from as-

★ E-mail: atoubiana@aei.mpg.de

trophysical simulations. These simulations typically output samples from the population prior (i.e., events).

- parametric: the population prior is written as a combination of simple functions, which depend on hyperparameters that are inferred from the data.
- non-parametric: a more complicated and flexible functional form for the population prior is assumed, with a variable number of degrees of freedom. The parameters of the model have (in general) no physical meaning.

In the first approach, the hyperparameters have a clear astrophysical meaning. They are related to the parameters of the astrophysical simulations, for example the efficiency of energy transfer from the binary to the gas during the common envelope stage, or the branching fraction between binaries formed in isolation and those formed dynamically, or parameters characterising properties of primordial BHs Zevin et al. (2021); Wong et al. (2021); Franciolini et al. (2022); Mould et al. (2022). Moreover, such an approach naturally incorporates correlations between parameters. However, it heavily relies on assumptions about astrophysical processes that are highly uncertain and has limited flexibility. Furthermore, in the standard way of performing population inference one needs to evaluate $p(\Theta|\Lambda)$ Mandel et al. (2019); Vitale et al. (2020), thus requiring an additional step where the probability density function (pdf) is estimated from samples of the population prior with the use of neural networks or kernel density estimators Zevin et al. (2021); Wong et al. (2021); Franciolini et al. (2022); Mould et al. (2022); Toubiana et al. (2021). The parametric approach Talbot & Thrane (2018); Abbott et al. (2021b, 2023) partially leverages these issues, using functional forms for the population prior that are generic enough and could describe a variety of astrophysical scenarios. This is the approach used in the fiducial POWER-LAW+PEAK (PP) analysis of the LVK, which describes the primary mass pdf as the weighted sum of a power-law and a Gaussian Abbott et al. (2023). Although more flexible than the astrophysical one, the parametric approach also makes assumptions about the general form of the population prior. For instance, the fiducial LVK analysis does not allow for excesses of BHs at both $20M_\odot$ and $35M_\odot$. Such assumptions are no longer required in the non-parametric approach. The price to pay is an increase in the number of parameters to be inferred and the loss of physical interpretation of those parameters.

However, it is precisely non-parametric models that have first indicated a potential excess of BHs around $15-20M_\odot$ in the LVK analysis. Additional analyses since then have also found evidence for such an excess Edelman et al. (2022a); Tiwari (2022, 2023), but not all Sadiq et al. (2022); Ruhe et al. (2022); Callister & Farr (2023). A generic problem of non-parametric models is that the number of parameters is allowed to be arbitrarily large. Adding parameters generally increases the complexity of the model but might lead to "overfitting" the data. Therefore, a compromise needs to be found, which requires iterating over the model dimensionality to find a suitable balance between the quality of the fit and the size of the model. In this work, we propose reversible jump Markov chain Monte-Carlo (RJMCMC) as a tool to optimally determine the complexity of the model. See Rinaldi & Del Pozzo (2021) for a different approach using Dirichlet processes. We consider two models for the distribution of primary masses that take advantage of RJMCMC. A semi-parametric one that is a more flexible version of the PP model of the LVK, and a non-parametric one, representing the pdf of primary masses as a piece-wise power-law function. Moreover, we illustrate how RJMCMC can be used to perform model selection without having to perform multiple runs.

Applying both models to GWTC-3, our semi-parametric model favours having a second Gaussian at $\sim 10M\odot$ relative to the PP model, leading to a displacement of the peak of low-mass events. Our non-parametric model shows more differences, in broad agreement with other non-parametric analyses of the LVK. In particular, it agrees on the displacement of the peak at low masses and does show some evidence for an excess of BHs around $20M_\odot$. We investigate the statistical significance of this excess under simplifying assumptions by performing mock injections, and estimate that there is a 5% chance that this excess is due to statistical fluctuations, when assuming a population compatible with the PP model. We show that with $\sim$ 500 events our non-parametric model will be able to reliably and efficiently identify such an excess.

This paper is organised as follows. In Sec. 2, we describe the details of our analysis. We present the results of our population analysis on GWTC-3 in Sec. 3. Then, in Sec. 4 we describe our mock injections and comment on the significance of the excess. Finally, in Sec. 5 we present our general conclusions.

2 SETUP

In this section, we will start by reviewing the basics of hierarchical Bayesian analyses, then we comment on the advantages of RJMCMC in population analyses and finally describe the population models we use.

2.1 Hierarchical Bayesian framework

Assuming a population prior $p(\Theta|\Lambda)$, we can write the number density of events as:

$$\frac{\mathrm{d}N}{\mathrm{d}\Theta}(\Lambda) = N(\Lambda)p(\Theta|\Lambda), \tag{1}$$

such that $N(\Lambda)$ is the total number of events during the observation period $T_{\rm obs}$ predicted by the population model. Inference on Λ is performed within a hierarchical Bayesian framework. Given a set of N_{obs} observed data, $\{d_i\}$, the posterior on the hyperparameters governing the population is Mandel et al. (2019); Vitale et al. (2020):

$$p(\Lambda|\{d_i\}) \propto \pi(\Lambda)e^{-\xi(\Lambda)N(\Lambda)}\prod_{i=1}^{N_{obs}}\int\frac{\mathrm{d}N}{\mathrm{d}\Theta}(\Lambda)\frac{p(\Theta|d_i)}{\pi_{PE}(\Theta)}\mathrm{d}\Theta, \tag{2}$$

where $p(\Theta|d_i)$ is the single event posterior, $\pi_{PE}(\Theta)$ is the prior used for parameter estimation, $\pi(\Lambda)$ is the prior on the hyperparameters and $\xi(\Lambda)$ is the selection function. The latter measures the fraction of events from the population that we expect to observe for a given value of Λ:

$$\xi(\Lambda) = \int\int_{d\ \text{detectable}} p(d|\Theta)p(\Theta|\Lambda)\mathrm{d}d\mathrm{d}\Theta. \tag{3}$$

The integral over d is performed only over the detectable datasets. It defines the probability of detecting such an event from the population characterised by the hyperparameters Λ. A handy way to compute the selection function for a variety of population models is to first perform an injection campaign, generating mock events with realistic noise and running the detection pipelines used during operational runs to determine if they are detectable. The selection function is then computed via Monte-Carlo integration:

$$\xi(\Lambda) = \frac{1}{N_{all}}\sum_{\Theta_i\ \text{detected}}\frac{p(\Theta_i|\Lambda)}{\pi_\emptyset(\Theta_i)}, \tag{4}$$

where $\pi_\emptyset(\Theta)$ is the prior used to perform the injections, N_{all} is the *total* number of injection performed and the sum runs only over the parameters that lead to detectable events. We marginalise over the statistical uncertainty in $\xi(\Lambda)$ coming from the Monte-Carlo estimation following the method in Farr (2019). We compute the selection function using the injection campaign performed by the LVK LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration (2021a,b) and applying the same criteria described in appendix A of Abbott et al. (2023) for the detectability of an event. Finally, we evaluate the posterior on Λ via Monte-Carlo integration:

$$p(\Lambda|\{d_i\}) \propto \frac{\pi(\Lambda)}{\xi(\Lambda)^{N_{obs}}} \prod_i^{N_{obs}} \sum_{\Theta_j \sim p(\Theta|d_i)} \frac{1}{\pi_{PE}(\Theta_j)} \frac{\mathrm{d}N}{\mathrm{d}\Theta}(\Lambda)\Big|_j , \quad (5)$$

where it is assumed that we use the same number of samples for each event. We retain the same 69 BBH events with false alarm rate below $0.25 yr^{-1}$ and use the same parameter estimation samples for those as Abbott et al. (2023). These are provided in the public data release of the Gravitational Wave Open Science Center LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration (2021a,b). Finally, we apply the criteria on the number of effective samples described in appendix B of Abbott et al. (2023).

In Eq. 5, we use the proportionality symbol instead of the equality one because we have omitted numerical factors that depend on the observed data $\{d_i\}$, but not on Λ, i.e., the individual event evidences and the overall model evidence. These factors are unimportant when the goal is to obtain the posterior distribution on Λ, but the overall model evidence is usually required to perform model selection. Assuming equal a priori probability for the models we want to compare, the ratio between evidences gives the Bayes' factor between models. As we will describe next, RJMCMC allows us to bypass evidence computation and returns the Bayes' factor without any extra cost.

2.2 Reversible jump MCMC

RJMCMC Green (1995) is a powerful method that explores the parameter space while allowing its dimensionality to vary. This is achieved by proposing not only changes in the parameters of the current model, but also the addition or removal of model parameters. As an example, imagine we have a datastream containing an unknown number of signals of known shape. RJMCMC allows the number of sources in the datastream to be estimated while also estimating their parameters. It returns a posterior distribution on the number of sources present in the datastream. This is needed when the data stream from a detector may contain many signals simultaneously, such as LISA and the "global fit" problem Littenberg & Cornish (2023); Littenberg et al. (2020); Katz, Karnesis, Korsakova & Gair (Katz et al.). RJMCMC can also be used to perform model selection: in this example, the ratio of the number of samples containing n_1 sources to the number of samples containing n_2 sources gives the ratio between the evidences of the two hypotheses. In this work we use the Eryn implementation of RJMCMC Karnesis et al. (2023a,b). For more details on the mathematics and methods used in RJMCMC within Eryn, we refer the reader to Karnesis et al. (2023a). We have also implemented our code for GPUs to speed up the computation. The results presented in this work take in average 4 days to run on a GPU.

In the context of population analysis, RJMCMC is useful for many purposes. First, it allows us to explore more flexible combinations to describe the population prior. For instance, we can consider an extended PP model where the number of Gaussians is free to vary. We can also let the presence of a power-law component be decided by the data, allowing for 0 or 1 (or even more) power-laws while sampling. Finally, we can perform model selection between having a broken power-law or a simple power-law, letting the number of each component be 0 or 1 and jumping between them. For non-parametric models, the number of parameters required is related to the complexity of the pdf being fitted. Increasing the number of parameters is tempting but might lead to "overfitting" of the data and finding spurious features and of course leads to a more complicated posterior to be sampled. Moreover, one usually has to try different configurations, performing runs for different numbers of parameters until the best compromise is found, based on the evidence. RJMCMC alleviates this burden by letting the number of parameters be a free parameter and providing a posterior over it. In this sense, the complexity of the model is chosen by the data.

2.3 Population models

As in Abbott et al. (2023), the event parameters used to perform the population inference are the source-frame mass of the primary m_1, the mass ratio $q \leq 1$, the spin magnitudes χ_1 and χ_2, the angles between the BH spins and the orbital angular momentum of the binary (tilt angles), θ_1 and θ_2, and the redshift of the source, z. We assume the number density to be separable:

$$\begin{aligned}\frac{\mathrm{d}N}{\mathrm{d}\Theta}(\Lambda) =& \frac{\mathrm{d}N}{\mathrm{d}m_1}(\Lambda_{m_1}) p(q|m_1, \Lambda_q, \Lambda_{m_1}) p(\chi_1|\Lambda_\chi) p(\chi_2|\Lambda_\chi) \\ & p(\theta_1, \theta_2|\Lambda_\theta) p(z|\Lambda_z). \end{aligned} \quad (6)$$

and focus on modelling the primary mass number density, $\frac{\mathrm{d}N}{\mathrm{d}m_1}(\Lambda_{m_1})$. We describe next the different possibilities that we explore for it. As for the remaining parameters (mass ratio, spins and redshift) we use the same model as the fiducial analysis in Abbott et al. (2023), which we describe in App. A.

2.3.1 *Semi-parametric model*

We consider an extended version of the PP model of the LVK, with two main differences: we let the number of Gaussians vary, and, instead of requiring the power-law component to be present, we allow for either one power-law, or one broken power-law, or none. We label this model FLEXIBLE POWER-LAW+GAUSSIANS (FPG). The m_1 number density can be written:

$$\begin{aligned}\frac{\mathrm{d}N}{\mathrm{d}m_1}(\Lambda_{m_1}) =& S(m_1, m_{min}, \delta_m) \times \\ & \Bigg[\sum_{i=0}^{N_G} \lambda_i^G G(m_1, \mu_i, \sigma_i) + \lambda^{PL} \sum_{i=0}^{1} PL(m_1, m_{min}, m_{max}, \alpha) \\ & + \lambda^{BPL} \sum_{i=0}^{1} BPL(m_1, m_{min}, m_{max}, m_{break}, \alpha_1, \alpha_2) \Bigg] \end{aligned} \quad (7)$$

where N_G is the maximum number of Gaussians allowed a priori, $G(m, \mu, \sigma)$ is a Gaussian distribution centred around μ of width σ, PL is a power-law and BPL is a broken power-law, defined in Eqs. 8 and 9 below. $S(m, m_{\rm min}, \delta_m)$ is the smoothing function introduced in Talbot & Thrane (2018), which we give explicitly in App. A, and δ_m defines a scale over which the m_1 pdf goes smoothly to zero. The

Parameter	λ	μ	σ	m_{min}	m_{max}	m_{break}	α	δ_m
Range	$0-10^5$	2-100 $M_\odot$	1-10 $M_\odot$	2-10 $M_\odot$	30-100 $M_\odot$	2-100 $M_\odot$	1.1-10	0.5-10

Table 1. Range of the priors on the hyperparameters of the FLEXIBLE POWER-LAW+GAUSSIANS model. We assume uniform priors in the provided ranges.

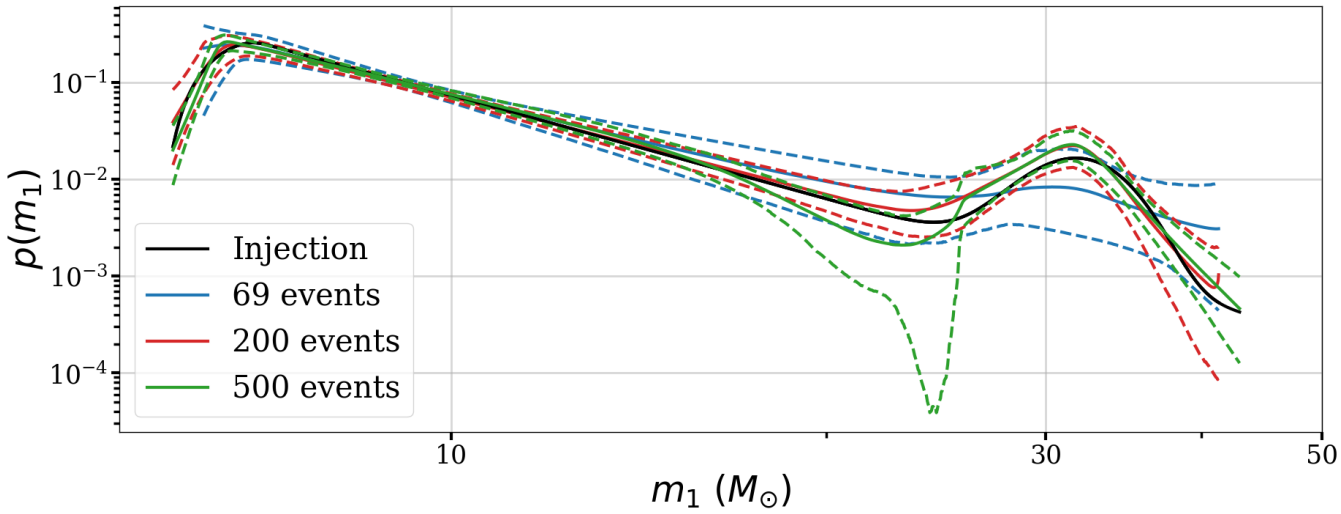

Figure 1. Comparison between the m_1 pdf from which we draw events (black line) and the inferred distribution with our non-parametric model. The solid coloured lines show the mean and the dashed coloured lines delimit the 90% confidence interval for different numbers of injections. The injected pdf is typically within the 90% confidence interval. The injection pdf is given by the PP model with $\lambda^G = 5.4$, $\lambda^{PL} = 63.6$, $\mu = 31.8M_\odot$, $\sigma = 2.8M_\odot$, $m_{min} = 3.6M_\odot$, $m_{max} = 99.4M_\odot$, $\alpha = 3.5$, $\delta_m = 1.7M_\odot$.

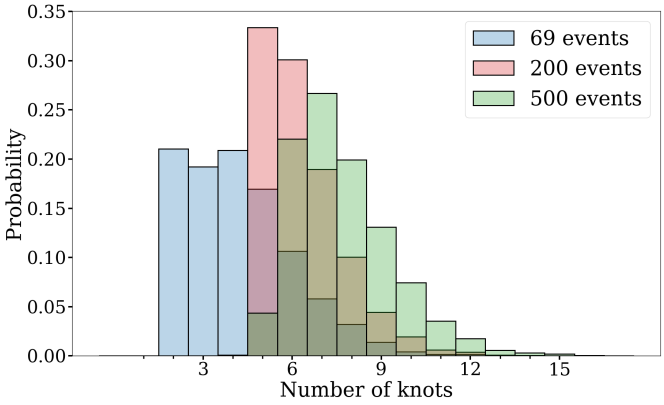

Figure 2. Posterior distribution on the number of knots for each of the curves shown in Fig. 1. Increasing the number of events barely increases the number of knots required to represent the pdf.

power-law and broken power-law are defined as:

$$PL(m_1, m_{min}, m_{max}, \alpha) = \begin{cases} \mathcal{N} m_1^{-\alpha}, \text{ if } m_{min} \leq m_1 \leq m_{max}; \\ 0 \text{ otherwise,} \end{cases} \tag{8}$$

$$BPL(m_1, m_{min}, m_{max}, m_{break}, \alpha_1, \alpha_2) = \begin{cases} \mathcal{N} m_1^{-\alpha_1} \left(1 + \frac{m_1}{m_{break}}\right)^{\alpha_1 - \alpha_2} \\ \text{if } m_{min} \leq m_1 \leq m_{max}; \\ 0 \text{ otherwise,} \end{cases} \tag{9}$$

where in both cases $\mathcal{N}$ is the appropriate normalisation factor. Note that in the case $\alpha_1 = \alpha_2$ we recover the simple power-law. In Eq. 7, it is understood that the $i = 0$ case corresponds to the absence of the component. For instance, we can have two Gaussians with or without a power-law, or just a broken power-law and so on and so forth. The only restrictions are that we must have at least one component and that we cannot have the power-law and the broken power-law simultaneously. In any of these cases we set the likelihood to zero.

In Eq. 7, the amplitudes λs are related to the number of events in each component. Strictly speaking there is not equality because of the smoothing function $S(m, m_{\rm min}, \delta_m)$, but this is a small correction. To avoid cases where a component is added with very small amplitude, we use a flat prior on the λ's. We also use flat priors for the remaining hyperparameters characterising the m_1 distribution, with ranges given in Table 1. The prior on the number of Gaussians, power-law and broken-power-law is also taken to be flat. For all other population hyperparameters we use the same priors as in Abbott et al. (2023).

2.3.2 *Non-parametric models*

In our non-parametric model, we describe the m_1 number density as a PIECE-WISE POWER-LAW (PWP) function. We write $\Lambda_{m_1} = \{x_i, v_i\}_n$, where v_i is the value of the pdf at a knot x_i, and n is the total number of knots. The number density at any point is obtained by interpolation:

$$\frac{\mathrm{d}N}{\mathrm{d}\Theta}(\Lambda) = \begin{cases} v_i \left(\frac{m_1}{x_i}\right)^{\frac{\log(v_{i+1}/v_i)}{\log(x_{i+1}/x_i)}}, \\ \text{if } x_1 < \ldots < x_i \leq m_1 < x_{i+1} < \ldots < x_n; \\ 0 \text{ if } m_1 < x_1 \text{ or } m_1 > x_n. \end{cases} \tag{10}$$

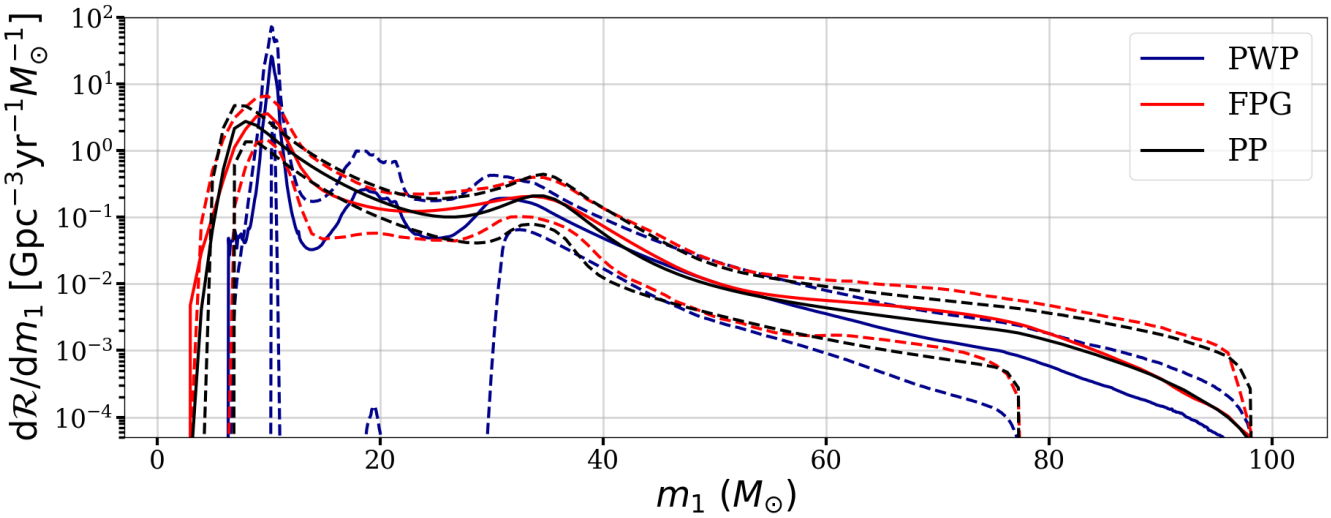

Figure 3. Comparison of the volumetric rate as a function of m_1 predicted by the PP (in black), the FPG (in orange) and the PWP (in blue) models. Full lines indicate the mean and dashed lines the 90% confidence interval of each model. The FPG model is in good agreement with the PP one, except for a displacement of the peak towards slightly larger masses. This comes from favouring a second Gaussian at $\sim 10M_\odot$ in addition to the one at $\sim 35M_\odot$, as shown in Fig. 1. As for the PWP model, it shows more deviations from the PP model. In particular, it agrees with the FPG model concerning the shift of the peak, favouring an even more pronounced peak, and suggests an excess of BHs around $20M_\odot$, as also suggested by some previous analyses Abbott et al. (2023); Edelman et al. (2022a); Tiwari (2022, 2023).

This is equivalent to assuming that $\log\left(\frac{dN}{d\Theta}(\Lambda)\right)$ is a piece-wise linear function of $\log(m_1)$.

We assume a log-flat prior on the $\{v_i\}_n$ and on the $\{x_i\}_n$ and a flat prior on the number of knots. In principle, the range of the prior on the position of the knots is determined by the minimum and maximum m_1 sample over all events. However, because there are very few samples above $100M_\odot$ (less than 0.5% of the total samples) and those are sparsely distributed, we find that letting the knots take values above $100M_\odot$ leads to spurious features. In fact, above $100M_\odot$ the determination of the population posterior is almost completely driven by the selection function. Therefore, we take $100M_\odot$ as the upper limit for the position of the knots.

We illustrate our method by generating mock injections compatible with the PP population inferred by the LVK, and recovering the pdf with our non-parametric model. We consider an increasing number of events: 69, as in the current dataset, 200 and 500. The latter two define a realistic range for the number of BBHs we expect to have observed after the fourth operational run (O4). For simplicity, in this illustrative case we do not include either selection effects or measurement errors. In Fig. 1, we compare the recovered pdfs to the pdf from which the mock injections are drawn. The injected pdf typically lies within the 90% confidence interval, showing that our method is able to properly infer it. Moreover, increasing the number of events does not lead to a dramatic increase in the number of knots used for interpolation, as can be seen in Fig. 2. Thus, the number of free parameters of our model remains reasonable as the size of the dataset increases. We have performed such injections for 200 sets of hyperparameters drawn from the LVK posterior and produced pp-plots by computing to which quantile of the recovered distribution do the quantiles of the observed set of events correspond. We obtain diagonal pp-plots, reinforcing our confidence that our non-parametric model can be used to infer LVK-like populations.

One of the non-parametric models used by the LVK is the POWER-LAW+SPLINE (PS) Edelman et al. (2022b), which models the number density as:

$$\frac{dN}{d\Theta}(\Lambda) \propto S(m_1, m_{min}, \delta_m) PL(m_1, m_{min}, m_{max}, \alpha) e^{f(m_1|\{f_i\})}, \tag{11}$$

where $f(m_1|\{f_i\})$ is a cubic spline function and the $\{f_i\}$ are the values of this function at fixed knots, spaced log-uniformly. The number of knots was fixed to 20. Our PWP model differs from the PS by: (i) not assuming an underlying power-law shape, (ii) letting the number and the position of knots vary. Notice that both models are able to represent a simple power-law function.

3 INFERENCE ON GWTC3

We apply both our PWP and FPG models to the LVK data. As described in Sec. 2.1, for ease of comparison, we consider exactly the same events and the same samples as in the LVK analyses Abbott et al. (2023). The inferred volumetric rates as a function of m_1 are shown in Fig. 3, where they are compared to the fiducial LVK result. We recap how the volumetric rate is derived from the number density in App. B.

The FPG model is in reasonable agreement with the LVK result, with the most noticeable difference at low masses. Regarding the power-law versus broken power-law comparison, the former is mildly favoured, with a Bayes' factor of 1.671 ± 0.003. However, in the broken power-law case, the posterior distribution is nearly flat in m_{break} and is concentrated around $\alpha_1 = \alpha_2$, reducing to a simple power-law. Thus, the favoured configuration is a power-law with 2 Gaussians, with a (43.49 ± 0.15)% probability relative to all possible combinations. The second best is a power-law with 3 Gaussians, at (23.91 ± 0.11)% probability. Finally, we find only a (8.5 ± 0.07)% probability of having only Gaussians. In the preferred configuration, we find a Gaussian at $\sim 35M_\odot$ and another at $\sim 10M_\odot$. As a consequence, the FPG model predicts a peak of events slightly displaced to larger masses compared to the fiducial LVK analysis. As commented above and illustrated in Fig. 5, there is also some support for

[t]

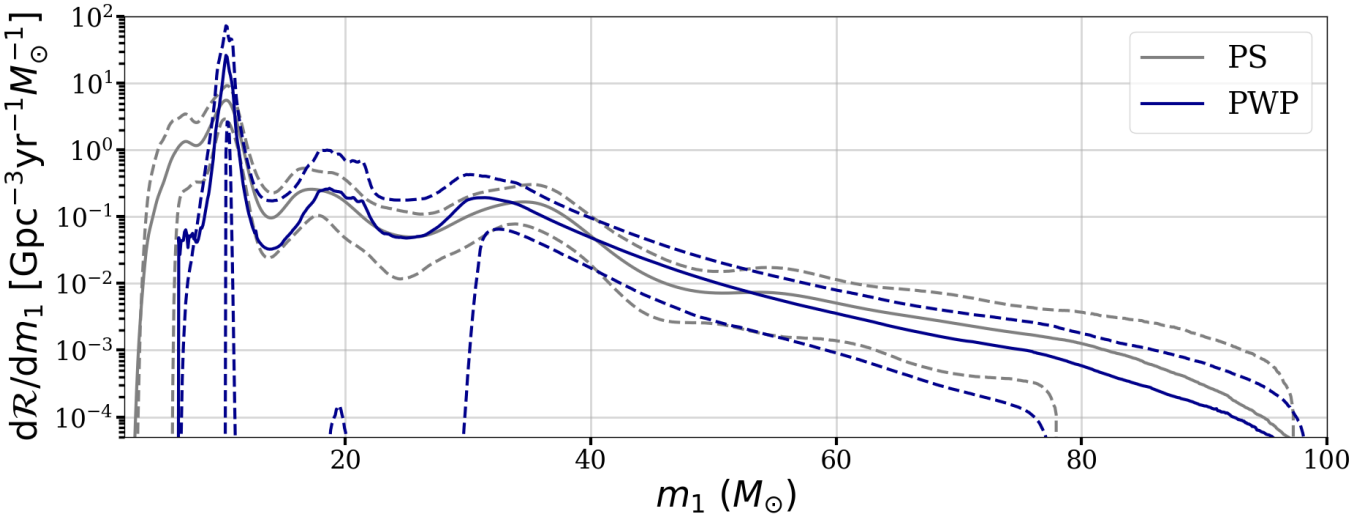

Figure 4. Same as Fig. 3 for the comparison between the PS (grey) and PWP (blue) models. Both models agree on the location of the peak at low masses, with our model predicting a more peaked shape. They also agree on the location of the secondary peaks. The uncertainty in our model is larger, in particular around $20M_\odot$ due to less a priori asusmptions, e.g., we don't assume an underlying power-law shape.

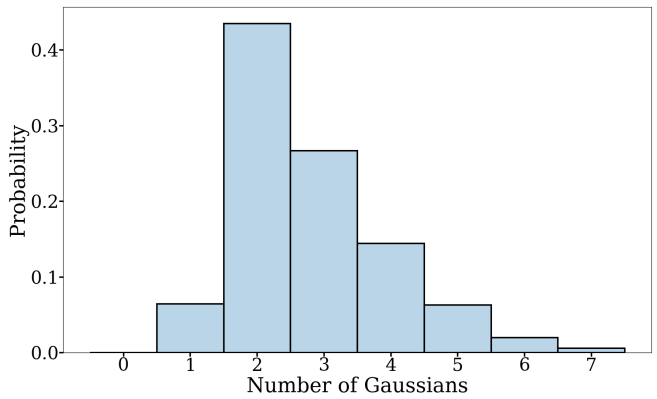

Figure 5. Posterior distribution on the number of Gaussians for our FPG model. The model favours having 2 Gaussians: one at $\sim 35M_\odot$ and another at $\sim 10M_\odot$. It also has some support for a third Gaussian around $65M_\odot$.

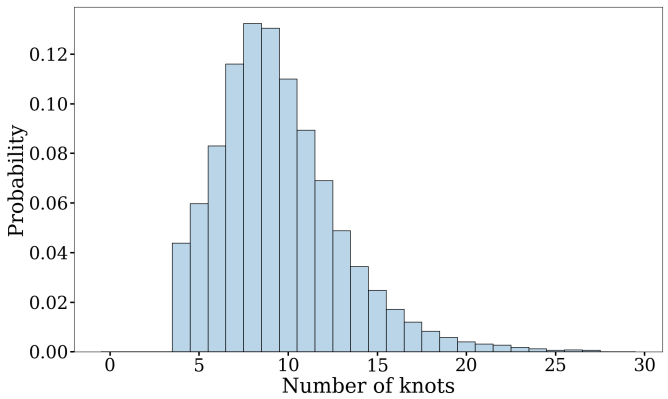

Figure 6. Posterior distribution on the number of knots for our PWP model.

a third Gaussian around $65M_\odot$, leading to the observed excess at high masses. However, we note that the y-axis is in log-scale, which exaggerates the size of this effect. Finally, we do not find evidence for an excess around $20M_\odot$. The result shown in Fig. 5 is summed over all possible model component combinations, with relative probabilities obtained from sampling.

Our non-parametric model shows more differences with respect to the PP analysis. Its prediction resembles more the ones returned by the FLEXIBLE MIXTURES and POWER-LAW+SPLINE models of Abbott et al. (2023). We compare the results of the latter to ours in Fig. 4 and in Fig. 6, we show the posterior distribution on the number of knots for the PWP model. The PWP model agrees with the FPG and the PS models on the position of the main peak, but predicts a more pronounced shape. Such a peak is typically expected for binaries that form in isolation Giacobbo & Mapelli (2018); Wiktorowicz et al. (2019); van Son et al. (2022). Together, these results suggest that the PP model is not flexible enough to capture these fine features. The PWP also recovers a secondary maximum around $35M_\odot$. It is slightly shifted to lower masses compared to the FPG, PP and PS models, but the 90% confidence intervals still have a broad overlap in this region. Regarding the original goal of our analysis, the PWP model does suggest an excess of BHs around $20M_\odot$, as indicated by the peak in the mean and the upper boundary of the 90% confidence interval, in broad agreement with the PS prediction. We stress however that the uncertainty of the PWP model in this region is very large, and the inferred distribution is also compatible with not having a peak around $20M_\odot$. We attribute the larger uncertainty of our model with respect to the PS as due to the fact it makes fewer a priori assumptions, e.g., we do not assume that the underlying function is a power-law. Similarly, the difference in the volumetric rate before the peak is likely due to the underlying assumption on the shape of the m_1 pdf in the PS model. Given the small number of observations, this excess at $20M_\odot$ suggested by our model could be due to statistical fluctuations. In order to investigate this possibility, we perform a series of mock injections.

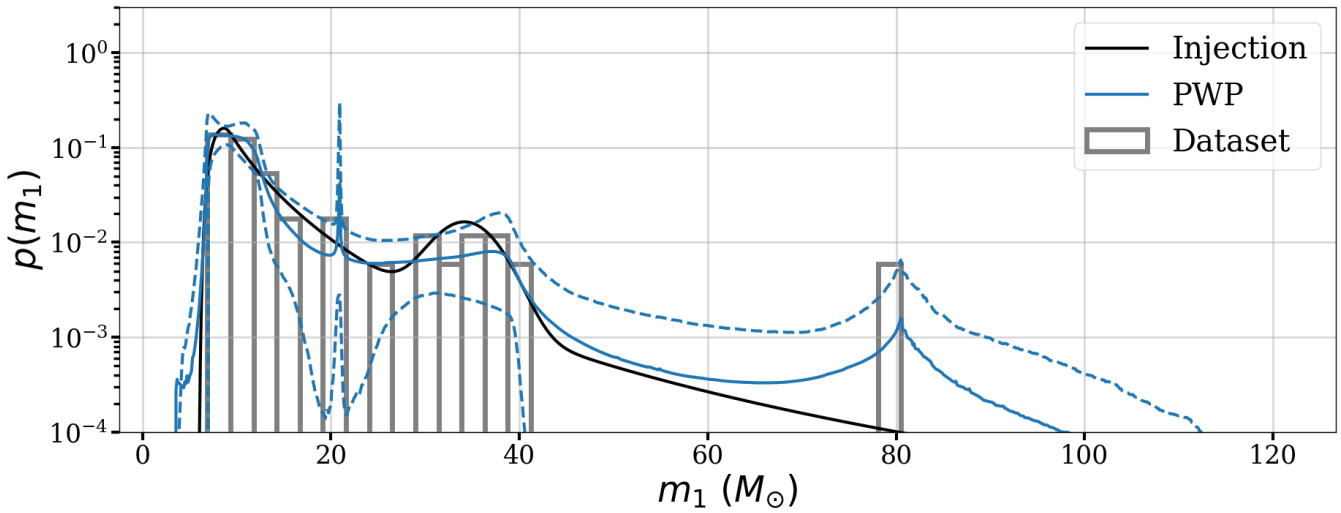

Figure 7. Examples of a mock injection catalogue with 69 events where statistical fluctuations lead to erroneously thinking that there in an excess of BHs around $20M_\odot$. The same happens around $80M_\odot$.

4 INFERENCE ON MOCK DATA

Firstly, we want to assess how likely it is that we find a spurious peak-like feature between $13M_\odot$ and $25M_\odot$ in the distribution inferred with our PWP model when the underlying population does not have such a feature. This range is motivated by the results of Mahapatra et al. (2022) that show that BBHs with primary mass in this range likely contain at least one second generation BH. We decide on the presence of a peak-like feature by looking for a local maximum in the mean of the inferred pdf. We refer to Farah et al. (2023) for an analysis of the statistical significance of other noticeable features in the m_1 distribution. We draw 200 sets of hyperparameters from the LVK PP posterior and for each of those we simulate datasets of 69, 200 and 500 events. The latter two define a realistic range for the number of BBH observations we expect by O4. Then, we analyse the generated events with our non-parametric model and count the fraction of realisations in which we find a peak between $13M_\odot$ and $25M_\odot$ in the inferred distribution. Since we only want to get a rough idea of the significance of the peak observed in the LVK data, to speed up computations we do not account for selection effects nor measurement errors. We find that the probability to find a peak between $13M_\odot$ and $25M_\odot$ is:

- 0.04 ± 0.01 with 69 events
- 0.02 ± 0.01 with 200 events
- < 0.005 with 500 events

We show in Fig. 7 an example where applying our PWP model to a dataset of 69 events leads to erroneously thinking there is an excess of BHs around $20M_\odot$. We show in gray the histogram of events. Although the underlying population has no local maximum around $20M_\odot$, statistical fluctuations lead to an excess of events in this region, driving the erroneous inference. In this example, the same happens around $80M_\odot$. Thus, the false-alarm probability for detecting an excess between $13M_\odot$ and $25M_\odot$ with our non-parametric model after observing 69 events is roughly 5%, but virtually null after 500 events. When analysing these same datasets with the FPG model, none of the cases shows a peak between $13M_\odot$ and $25M_\odot$. At high masses, where the events are more sparsely distributed, it is common for our non-parametric model to show spurious peaks, due to fitting for isolated events, as illustrated in Fig. 7. This problem does not concern the range we are interested in, between $13M_\odot$ and $25M_\odot$, which is in the bulk of distribution and where the events are more continuously distributed.

Next, we want to estimate under which circumstances we can confidently detect the presence of the peak around $20M_\odot$. For the same 200 samples from the LVK population, we define 50 new populations by adding a Gaussian of mean $19M_\odot$ and width $2M_\odot$ with increasing weight. For this purpose we define the normalised fraction $f_i = \lambda_i / \sum_i \lambda_i$, where the sum runs over the amplitudes of all components. In practice we grid the fraction f_{19}^G logarithmically from 10^{-3} to 1 and redistribute the weight that was initially in the power-law between the Gaussian at $19M_\odot$ and the power-law. The weight of the Gaussian at $\sim 35M_\odot$ remains the same. This simplistic procedure is driven by the idea that if there is an additional peak around $20M_\odot$, it has been "swallowed" by the power-law component. One might also expect that the inferred power-law index is actually smaller than it should, i.e., the power-law is less steep in order to accommodate the excess. However, this effect should be small and we are looking for a rough estimate of the detectability of an excess around $20M_\odot$, so we do not account for it. Once again we draw sets of 69, 200 and 500 events from each of these populations, and analyse them both with the PWP and with the FPG model. Fig. 8 shows the probability of finding a peak between $13M_\odot$ and $25M_\odot$ as a function of the weight of the Gaussian around $19M_\odot$ with each model. At small weights, we recover the false-alarm probability discussed above. As expected, the chances of detecting an excess increase as the weight increases. With 500 events, our non-parametric model will reliably indicate if the excess is truly physical, with higher probability than the FPG model. From this plot, we also see that it is not surprising that the latter does not find the putative excess of BHs in the current LVK data, since it requires large weights to identify it with only 69 events. On this same figure, the solid blue curve shows the probability of detecting a peak with the non-parametric model and not with the semi-parametric one when observing 69 events, as it the case for the GWTC-3 dataset. Under the hypothesis that the BBH population is compatible with the ones used to perform the mock injections, this can be interpreted as a likelihood for the observed datum that one fit finds a peak and the other does not. Constructing a posterior from this likelihood and a flat prior on f_{19}^G, we estimate that $f_{19}^G \geq 0.08$

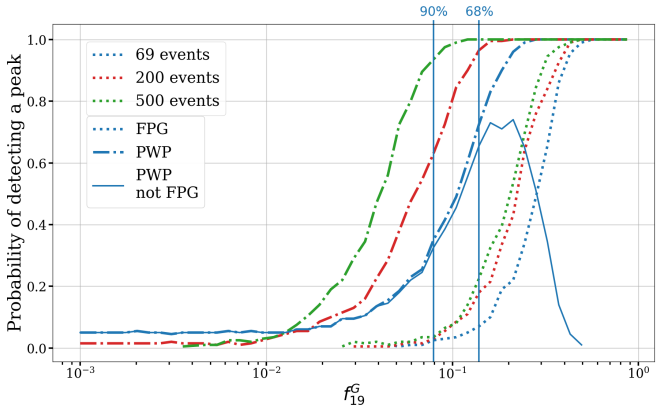

Figure 8. Probability of detecting a peak between $13M_\odot$ and $25M_\odot$ as a function of the weight of the Gaussian at $19M_\odot$, for both the FPG and the PWP model, considering an increasing number of events. At small weights, the probability matches the false-alarm probability. The bumpy aspect at small weights is due to low statistics in this region, but the general trend is clear. Curves interrupted somewhere over the x-range go to 0. The solid blue curve shows the probability of detecting a peak with the non-parametric model and not with the semi-parametric one for 69 events, as it is the case for the GWTC-3 dataset. The vertical lines show the lower bounds on f_{19}^G at 68% and 90% credibility, obtained by treating this curve as a likelihood, as described in the text.

at 90% credibility and $f_{19}^G \geq 0.14$ at 68% credibility. These are the values indicated with vertical lines. From the value at which the full line curve goes to 0, we deduce in addition that $f_{19}^G \lesssim 0.5$.

Compared to the full analysis of Sec. 3, our mock studies do not take into account the presence of selection effects and measurement uncertainties in individual events. We expect the latter to smooth out the mass distribution, making fine features harder to detect. Thus, we might be slightly overestimating the detectability of peaks by our methods. Measurement uncertainties would also smooth out sharp features like the one seen in Fig. 7. Selection effects would make the distribution of observed events different to the astrophysical one. For instance, events in the peak around $10M_\odot$ would be harder to detect compared to events with mass $20M_\odot$ or $35M_\odot$. On average, we could expect this to decouple from the statistical fluctutations we investigated in the first part of this section, and not have too much impact on our conclusions regarding the detectability of spurious peaks. Quantities like the position of the peaks and the "excess probability" in these peaks would in turn be affected, but these are not the focus of our investigation. On the other hand, selection effects would make it easier to identify real excesses around $20M_\odot$ for a fixed number of observed events.

5 CONCLUSION

Capturing distinctive features in the population of BBHs is crucial to discriminate between astrophysical models. Astrophysically motivated priors are useful to directly constrain physical parameters, but lack flexibility when the population includes unanticipated structure. In this paper we have proposed two flexible models for the distribution of primary masses and applied them to GWTC-3. A crucial ingredient of our method is RJMCMC, which allows the complexity of the model to be chosen by data as well as to perform model selection. With these models in hand, we have assessed the statistical significance of the excess of events around $15 - 20M_\odot$ found by some of the population analyses performed on GWTC-3 Abbott et al. (2023); Edelman et al. (2022a); Tiwari & Fairhurst (2021), but not all Sadiq et al. (2022); Ruhe et al. (2022); Callister & Farr (2023).

Our first model is an extended and more flexible version of the POWER-LAW+PEAK model of the LVK, where the number of Gaussians is free to vary and we can choose between having a power-law, a broken power-law or none. It illustrates how RJMCMC can be used to perform model selection, allowing to choose between a variety of models at once, instead of running them individually and comparing the evidences afterwards. We have found that the current data favours having a power-law component with two Gaussians: one around $35M_\odot$ and another around $10M_\odot$. As a consequence, it predicts a peak of events at slighlty higher masses than in the fiducial LVK analysis. There is also mild support for a third Gaussian at $\sim 65M_\odot$, but it has small significance in the current data. Furthermore, it is disfavoured to have only Gaussians. Finally, we find no sign for a break in the slope of the power-law, but this might change as the number of observations increases. Moreover, more elaborated combinations of parametric functions could be considered in order to fully take advantage of the flexibility offered by RJMCMC.

Next, we have proposed a non-parametric model representing the m_1 pdf as a piece-wise power-law function. We infer the position of the knots and the value of the pdf at those knots, but also the number of knots, thanks to the RJMCMC. The complexity of our model is not pre-determined, but it is decided by data. This model shows a few differences with respect to the POWER-LAW+PEAK model, in particular it agrees with our semi-parametric model regarding the displacement of the low-mass peak and also suggests an excess of BHs around $20M_\odot$. It is in better agreement with the POWER-LAW+SPLINE model of the LVK. However, by performing mock-injections under simplifying assumptions (i.e. negelcting selection effects and measurement uncertainties), we have found that there is roughly a 5% chance that the peak at $20M_\odot$ is due to statistical fluctuations when assuming a population compatible with the LVK POWER-LAW+PEAK analysis. With 500 events the false-alarm probability is nearly zero. Moreover, when analysing mock-populations that do have an excess at $20M_\odot$, our model can more easily find it than our FLEXIBLE POWER-LAW+GAUSSIANS model. As the number of events increases, additional features in the m_1 distribution might appear. Hierarchical mergers could lead to a series of regularly spaced peaks with decreasing amplitude Tiwari & Fairhurst (2021); Tiwari (2022). Detecting more massive events such as GW190521 Abbott et al. (2020) will inform us on the mass-gap, and whether there is a dearth of events in this region, as predicted by models. Our non-parametric model would be able to capture these features without any a priori modelling of the signature of these effects on the m_1 distribution, proving a powerful tool to better understand astrophysics.

In this work we have assumed the population prior to be separable in the event parameters, primary mass, mass ratio, spins and redshift. However, finding correlations between parameters would increase our ability to discriminate between astrophysical scenarios. Different approaches have already been explored Hoy et al. (2022); Callister et al. (2021); Fishbach et al. (2021); Adamcewicz & Thrane (2022); Bavera et al. (2022); Biscoveanu et al. (2022); Ray et al. (2023), with some of them yielding first hints of correlations. With the increase in the number of events, such fine features will become more prominent. It is therefore the natural next step for us to extend out method to multi-dimensional distributions.

ACKNOWLEDGMENTS

This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

DATA AVAILABILITY

This work makes use of the publicly available data released by the LVK LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration (2021a,b) and the publicly available code Eryn Karnesis et al. (2023b). In addition to our own, we show results from Abbott et al. (2023), available on LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration (2021b).

REFERENCES

Aasi J., et al., 2015, Class. Quant. Grav., 32, 074001
Abbott R., et al., 2020, Phys. Rev. Lett., 125, 101102
Abbott R., et al., 2021a
Abbott R., et al., 2021b, Astrophys. J. Lett., 913, L7
Abbott R., et al., 2023, Phys. Rev. X, 13, 011048
Acernese F., et al., 2015, Class. Quant. Grav., 32, 024001
Adamcewicz C., Thrane E., 2022, Mon. Not. Roy. Astron. Soc., 517, 3928
Akutsu T., et al., 2021, PTEP, 2021, 05A101
Bavera S. S., Fishbach M., Zevin M., Zapartas E., Fragos T., 2022, Astron. Astrophys., 665, A59
Belczynski K., et al., 2016, Astron. Astrophys., 594, A97
Benacquista M. J., Downing J. M. B., 2013, Living Rev. Rel., 16, 4
Biscoveanu S., Callister T. A., Haster C.-J., Ng K. K. Y., Vitale S., Farr W. M., 2022, Astrophys. J. Lett., 932, L19
Bouffanais Y., Mapelli M., Santoliquido F., Giacobbo N., Di Carlo U. N., Rastello S., Artale M. C., Iorio G., 2021, Mon. Not. Roy. Astron. Soc., 507, 5224
Callister T. A., Farr W. M., 2023
Callister T. A., Haster C.-J., Ng K. K. Y., Vitale S., Farr W. M., 2021, Astrophys. J. Lett., 922, L5
Edelman B., Farr B., Doctor Z., 2022a
Edelman B., Doctor Z., Godfrey J., Farr B., 2022b, Astrophys. J., 924, 101
Farag E., Renzo M., Farmer R., Chidester M. T., Timmes F. X., 2022, Astrophys. J., 937, 112
Farah A. M., Edelman B., Zevin M., Fishbach M., Ezquiaga J. M., Farr B., Holz D. E., 2023
Farmer R., Renzo M., de Mink S. E., Marchant P., Justham S., 2019,] 10.3847/1538-4357/ab518b
Farr W. M., 2019, Research Notes of the AAS, 3, 66
Fishbach M., et al., 2021, Astrophys. J., 912, 98
Franciolini G., et al., 2022, Phys. Rev. D, 105, 083526
Gerosa D., Berti E., O'Shaughnessy R., Belczynski K., Kesden M., Wysocki D., Gladysz W., 2018, Phys. Rev. D, 98, 084036
Giacobbo N., Mapelli M., 2018, Mon. Not. Roy. Astron. Soc., 480, 2011
Green P. J., 1995, Biometrika, 82, 711
Hoy C., Fairhurst S., Hannam M., Tiwari V., 2022, Astrophys. J., 928, 75
Inayoshi K., Hirai R., Kinugawa T., Hotokezaka K., 2017, Mon. Not. Roy. Astron. Soc., 468, 5020
Karnesis N., Katz M. L., Korsakova N., Gair J. R., Stergioulas N., 2023b, Eryn, https://github.com/mikekatz04/Eryn,
Karnesis N., Katz M. L., Korsakova N., Gair J. R., Stergioulas N., 2023a
Katz M., Karnesis N., Korsakova N., Gair J., , to appear
LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration 2021a, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O1+O2+O3 Search Sensitivity Estimates, 10.5281/zenodo.5636816
LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration 2021b, The population of merging compact binaries inferred using gravitational waves through GWTC-3 - Data release, 10.5281/zenodo.5655785
Littenberg T. B., Cornish N. J., 2023
Littenberg T., Cornish N., Lackeos K., Robson T., 2020, Phys. Rev. D, 101, 123021
Mahapatra P., Gupta A., Favata M., Arun K. G., Sathyaprakash B. S., 2022
Mandel I., Farr W. M., Gair J. R., 2019, Mon. Not. Roy. Astron. Soc., 486, 1086
Mapelli M., 2021, Formation Channels of Single and Binary Stellar-Mass Black Holes. (arXiv:2106.00699), doi:10.1007/978-981-15-4702-7_16-1
Marchant P., Renzo M., Farmer R., Pappas K. M. W., Taam R. E., de Mink S., Kalogera V., 2018,] 10.3847/1538-4357/ab3426
Mould M., Gerosa D., Taylor S. R., 2022, Phys. Rev. D, 106, 103013
Postnov K. A., Yungelson L. R., 2014, Living Rev. Rel., 17, 3
Ray A., Magaña Hernandez I., Mohite S., Creighton J., Kapadia S., 2023
Renzo M., Farmer R., Justham S., Götberg Y., de Mink S. E., Zapartas E., Marchant P., Smith N., 2020, Astron. Astrophys., 640, A56
Rinaldi S., Del Pozzo W., 2021, Mon. Not. Roy. Astron. Soc., 509, 5454
Romero-Shaw I. M., Kremer K., Lasky P. D., Thrane E., Samsing J., 2021, Mon. Not. Roy. Astron. Soc., 506, 2362
Ruhe D., Wong K., Cranmer M., Forré P., 2022
Sadiq J., Dent T., Wysocki D., 2022, Phys. Rev. D, 105, 123014
Samsing J., Ramirez-Ruiz E., 2017, Astrophys. J. Lett., 840, L14
Sedda M. A., Naoz S., Kocsis B., 2023
Tagawa H., Haiman Z., Kocsis B., 2020, Astrophys. J., 898, 25
Talbot C., Thrane E., 2018, Astrophys. J., 856, 173
Tiwari V., 2021, Class. Quant. Grav., 38, 155007
Tiwari V., 2022, Astrophys. J., 928, 155
Tiwari V., 2023
Tiwari V., Fairhurst S., 2021, Astrophys. J. Lett., 913, L19
Toubiana A., Wong K. W. K., Babak S., Barausse E., Berti E., Gair J. R., Marsat S., Taylor S. R., 2021, Phys. Rev. D, 104, 083027
Vitale S., Gerosa D., Farr W. M., Taylor S. R., 2020
Wiktorowicz G., Wyrzykowski L., Chruslinska M., Klencki J., Rybicki K. A., Belczynski K., 2019,] 10.3847/1538-4357/ab45e6
Wong K. W. K., Breivik K., Kremer K., Callister T., 2021, Phys. Rev. D, 103, 083021
Woosley S. E., 2017, Astrophys. J., 836, 244
Woosley S. E., Heger A., Weaver T. A., 2002, Rev. Mod. Phys., 74, 1015
Zevin M., et al., 2021, Astrophys. J., 910, 152
de Mink S. E., Mandel I., 2016, Mon. Not. Roy. Astron. Soc., 460, 3545
van Son L. A. C., et al., 2022, Astrophys. J., 940, 184

APPENDIX A: POPULATION PRIORS

We describe the models used for the single event parameters other than the primary mass, these are the same as in the fiducial analysis of Abbott et al. (2023). For the mass ratio, we assume a power-law distribution:

$$p(q|m_1, \Lambda_q) \propto q^{\beta} S(qm_1, m_{min}, \delta_m) \tag{A1}$$

with

$$S(m_1, m_{min}, \delta_m) = \begin{cases} 0; \text{ if } m_1 < m_{min} \\ [f(m_1 - m_{min}, \delta_m) + 1]^{-1} \\ \text{ if } m_{min} \leq m_1 \leq m_{min} + \delta_m; \\ 1; \text{ if } m_1 > m_{min} + \delta_m, \end{cases} \tag{A2}$$

and

$$f(m, \delta_m) = e^{\frac{\delta_m}{m} + \frac{\delta_m}{m - \delta_m}}. \tag{A3}$$

In the PWP model, we do not use the smoothing function, and impose a sharp cut-off:

$$p(q|m_1, \Lambda_q) \begin{cases} 0; \text{ if } qm_1 < 2 \\ \propto q^{\beta}; \text{ otherwise.} \end{cases} \tag{A4}$$

We assume a common Beta-dsitribution for the spins magnitude:

$$p(\chi|\Lambda_\chi) = \text{Beta}(\alpha_\chi, \beta_\chi). \tag{A5}$$

As for their orientation, we model the joined distribution of the cosines of the tilt angles as a mixture between a two-dimensional Gaussian of width σ_t, centred at 0 and truncated at -1 and 1, and a two-dimensional flat distribution between -1 and 1.:

$$\begin{aligned} p(\cos(\theta_1), \cos(\theta_2)|\Lambda_\theta) =& \zeta G_t(\cos(\theta_1), \cos(\theta_2), 0, \sigma_t) \\ &+ (1 - \zeta) U(-1, 1) \end{aligned} \tag{A6}$$

Finally, we assume that the rate of BBHs evolves with redshift as $\mathcal{R}(z) = (1 + z)^{\kappa}$, which leads to the pdf on z:

$$p(z|\Lambda_z) \propto (1 + z)^{\kappa - 1} \frac{\mathrm{d}V_c}{\mathrm{d}z}, \tag{A7}$$

where V_c is the comoving volume.

The priors for β, m_{min}, δ_m, ζ, σ_t and κ are taken to be flat. For α_χ and β_χ, we assume the prior on the mean μ_χ and the variance σ^2_χ of the Beta-distribution to be flat, subject to the condition $\alpha_\chi > 1$, $\beta_\chi > 1$.

APPENDIX B: VOLUMETRIC RATE

The number density can be transformed into a volumetric rate by taking its ratio with the observed space-time volume VT_{obs}, defined as:

$$VT_{\text{obs}} = T_{\text{obs}} \int_0^{z_{\text{max}}} \frac{1}{1 + z} \frac{\mathrm{d}V_c}{d_z} \mathcal{R}(z)\, \mathrm{d}z \tag{B1}$$

We take $z_{\text{max}} = 2.3$ and use the value of T_{obs} provided in the public data released along Abbott et al. (2023). Finally,

$$\frac{\mathrm{d}\mathcal{R}}{\mathrm{d}m_1}(\Lambda) = \frac{1}{VT_{\text{obs}}} \frac{\mathrm{d}N}{\mathrm{d}m_1}(\Lambda). \tag{B2}$$

This paper has been typeset from a TEX/LATEX file prepared by the author.