# Expect the unexpected: augmented mixture models for black hole population studies

Stefano Rinaldi • \*

Institut für Theoretische Astrophysik, ZAH, Universität Heidelberg, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany Received September 23, 2025; accepted XXX

#### **ABSTRACT**

Context. Black hole population studies are currently performed either using astrophysically motivated models (informed but rigid in their functional forms) or via non-parametric methods (flexible but not directly interpretable).

Aims. In this paper, we present a statistical framework to complement the predictive power of astrophysically motivated models with the flexibility of non-parametric methods.

Methods. Our method makes use of the Dirichlet distribution to robustly infer the relative weights of different models as well as of the Gibbs sampling approach to efficiently explore the parameter space.

Results. After having validated our approach using simulated data, we apply this method to the BBH mergers observed during the first three Observing Runs of the LIGO-Virgo-KAGRA collaboration using both phenomenological and astrophysical models as parametric models, finding results in agreement with the currently available literature.

**Key words.** Methods: statistical – gravitational waves – stars: black holes

Context. Black but rigid in the Aims. In this prodels with the Methods. Our well as of the Results. After during the fire astrophysical Key words. Matter Matter (BBH) mergers de Wirgo (Acernese et 2020) collaboration run (O3) (Abbott by GWTC-4.0 (The 2025b), BBH pop astrophysicists who massive binary every physical BH distrifuted for the details of matter due to stellar wind 2024; Vink et al. 2025), its metallical Hirschi et al. 2025 supernova process. At the same time, mation channels for can be broadly distributed for the details of the details of matter and the same time, mation channels for can be broadly distributed for the details of the details of the details of materials and the same time, mation channels for can be broadly distributed for the details of the details o With 69 gravitational waves (GWs) from binary black hole (BBH) mergers detected by the LIGO (Aasi et al. 2015), Virgo (Acernese et al. 2015) and KAGRA (Akutsu et al. 2020) collaboration (LVK) at the end of the third observing run (O3) (Abbott et al. 2023a) and 84 events newly added by GWTC-4.0 (The LIGO Scientific Collaboration et al. 2025b), BBH population studies are now a prime tool for astrophysicists when it comes to investigate the physics of massive binary evolution. The characterisation of the astrophysical BH distribution requires a profound understanding of the details of massive star evolution such as the mass loss due to stellar winds (Kruckow et al. 2024; Romagnolo et al. 2024; Vink et al. 2024; Merritt et al. 2025; van Son et al. 2025), its metallicity dependence (Belczynski et al. 2016; Hirschi et al. 2025) and the effect of the pair-instability supernova process (Mapelli et al. 2020; Vink et al. 2021). At the same time, it is also crucial to model the possible formation channels for compact binaries: the proposed models can be broadly divided into two classes, isolated evolution and dynamical formation. The isolated evolution scenario considers the possibility for the progenitors of the two compact objects to be already part of a binary system during the stellar stage: in this case, for example, the mass transfer (either stable or unstable) can tamper with the mass ratio of the binary (Röpke & De Marco 2023), and thus leave an imprint in the resulting BH distribution (Marchant et al. 2021; Gallegos-Garcia et al. 2021; Willcox et al. 2023). Conversely, dynamically assembled systems are those binaries where the components are brought together already at the stage of compact objects as a result of dynamical interactions happening in dense environments, as it is the case of

three-body encounters, dynamical captures and star clusters (Ziosi et al. 2014; Rodriguez et al. 2016; Kremer et al. 2020; Banerjee 2022). For a comprehensive review of the available models, we refer to Mapelli (2020) and Mandel & Farmer (2022). Due to the huge complexity of the astrophysical processes, the distribution induced by the aforementioned models cannot be expressed in terms of simple functions: the modelling community rely on numerical methods capable of producing synthetic catalogues of merging BBHs. For the same reason, developing an all-encompassing model accounting for all possible physical processes and formation channels is considered a titanic task.

Despite the effort put in accelerating and optimising population synthesis codes, these algorithms are not fast enough yet to be embedded in the MCMC methods commonly used to analyse the GW data – with the notable exception of machine-learning-enhanced approaches (e.g., Colloms et al. 2025). The available literature makes often use of phenomenological parametrised models inspired by the expected features of the BH distribution: among others we mention the POWERLAW+PEAK model (Abbott et al. 2023b), describing the primary mass distribution as a weighted superposition of a tapered power-law and a Gaussian distribution, as well as its updated version Broken POWER LAW + 2 PEAKS (The LIGO Scientific Collaboration et al. 2025a), currently favoured by the data. These models, albeit simplified with respect to the full-fledged astrophysical models, are still capable to provide insights on the processes at play behind the observed BHs. This approach is the one mainly employed by the LVK collaboration (Abbott et al. 2019, 2021, 2023b; The LIGO Scientific Collaboration et al. 2025a) as well as many authors (e.g., Fishbach & Holz 2017; Talbot & Thrane 2018; Farah et al. 2023; Gennari et al. 2025; Rinaldi et al. 2025a). For a com-

<sup>\*</sup> E-mail: stefano.rinaldi@uni-heidelberg.de

prehensive review, please see Callister (2024). Due to the fact that the functional forms are merely inspired by the astrophysical models and do not have any direct connection with the physical processes at play, the possibility of assuming a functional form that do not encompass the true underlying distribution and thus biasing the analysis cannot be neglected entirely. Moreover, new, previously unforeseen features in the spectrum must be added by hand when needed

To ensure the robustness of the population analysis in this respect, a complementary approach has been developed based on the concept of non-parametric methods. These, despite the potentially misleading name, are models that employ a functional form with a countably infinite number of parameters that can approximate arbitrary probability densities: in other words, a base for the space of normalised functions. These models have the advantage of being solely data-driven and not encoding any previous belief concerning the expected shape of the distribution, making them useful tools to discover new features in the BH spectrum. Among the ones developed within the GW community, we mention autoregressive processes (Callister & Farr 2024), Dirichlet process Gaussian mixture models (Rinaldi & Del Pozzo 2022), Gaussian processes (Li et al. 2021), reversible jump Markov chain Monte Carlo (Toubiana et al. 2023), and binned approaches (Ray et al. 2023; Heinzel et al. 2025). The flexibility of such models comes with the cost that the resulting reconstructed probability density is merely a description of the underlying distribution, not immediately interpretable in terms of astrophysical processes – although in Rinaldi et al. (2025b) we propose a way to circumvent this limitation.

In summary, we can broadly categorise the models employed as either 'informed but rigid' (astrophysical and parametrised) or 'flexible but not interpretable' (nonparametric), with somewhat complementary pros and cons. Several works employs a semi-parametric approach, in which the model is defined as the product of a parametric and a non-parametric distribution, each applied on a disjoint partition of the binary parameter space (e.g. a non-parametric model for the masses and a parametric model to describe the redshift and spin parameters). What is currently missing, to the best of our knowledge, is a way of bringing the two categories together to describe the same binary parameters in a 'flexible and informed' way: in this work, we will present the augmented mixture model (AMM), a weighted superposition of parametric (or astrophysical) and non-parametric models designed to infer the BH distribution accounting for the possible presence of unforeseen features in the spectrum while retaining the interpretability offered by the informed models.

The paper is organised as follows: in Section 2 we summarise the statistical framework used to analyse GW data; Section 3 introduces the AMM along with an outline of the algorithmic implementation. Section 4 demonstrates the applicability of our method using synthetic data, whereas Section 5 applies the AMM to the GW events detected during the third LVK observing run. Lastly, Section 6 summarises our findings.

# 2. Summary of statistical framework

In this work, we will make use of the statistical framework presented in Mandel et al. (2019); Vitale et al. (2022), briefly

recapped in this Section. For what concerns the notation, we will denote with  $\mathbf{d} = \{d_1, \dots d_N\}$  the data associated with the N observed GW events included in the catalogue. Each GW signal is described by a set of parameters  $\theta$  (e.g., masses and spins of the binary components, distance, sky position etc.). The fact that a specific GW event is detectable is denoted with  $\mathbb{D}$ . The astrophysical probability distribution will be denoted with  $p(\theta|\Lambda)$ , where  $\Lambda$  is the set of parameters used to describe the astrophysical distribution. Albeit not used in this section, in the specific case of non-parametric methods we will denote the (infinitely many) parameters with  $\Theta$ .

Following Mandel et al. (2019), the likelihood reads

$$p(\mathbf{d}|\Lambda, \mathbb{D}) = \prod_{i}^{N} p(d_{i}|\Lambda, \mathbb{D}) = \prod_{i}^{N} \int p(d_{i}|\theta_{i}, \mathbb{D}) p(\theta_{i}|\Lambda, \mathbb{D}) d\theta_{i},$$
(1)

having assumed that the events are independent and identically distributed. Making use of the Bayes' theorem the integrand can be refactored as

$$\frac{p(\mathbb{D}|d_i, \theta_i)p(d_i|\theta_i)}{p(\mathbb{D}|\theta_i)} \frac{p(\mathbb{D}|\theta_i)p(\theta_i|\Lambda)}{p(\mathbb{D}|\Lambda)} = \frac{p(d_i|\theta_i)p(\theta_i|\Lambda)}{p(\mathbb{D}|\Lambda)}.$$
 (2)

The assumption here is that the detectability of an observed event is, by definition, equal to 1. The denominator is the detectability fraction,

$$p(\mathbb{D}|\Lambda) = \int p(\mathbb{D}|\theta)p(\theta|\Lambda)d\theta \equiv \xi(\Lambda), \qquad (3)$$

and it is usually estimated via Monte Carlo integration using a set of simulated signals to marginalise over noise realisations. Making use of the Bayes' theorem on  $p(d_i|\theta_i)$ , eq. (1) becomes

$$p(\mathbf{d}|\Lambda, \mathbb{D}) = \prod p(d_i|\Lambda, \mathbb{D}) = \prod_i^N \frac{p(d_i)}{\xi(\Lambda)} \int \frac{p(\theta_i|d_i)p(\theta_i|\Lambda)}{p(\theta_i)} d\theta_i.$$
(4)

In the specific case in which the astrophysical model is a weighted superposition of M models, whose parameters we denote with  $\Lambda = \{\Lambda_1, \ldots, \Lambda_M\}$ ,

$$p(\theta_i|\mathbf{\Lambda}) = \sum_{j=0}^{M} w_j p_j(\theta_i|\Lambda_j) \quad \text{with} \quad \sum_{j=0}^{M} w_j = 1,$$
 (5)

the likelihood takes the simple form of a superposition of likelihoods where the relative weights account for the different detectability fractions (Rinaldi et al. 2025a):

$$p(\mathbf{d}|\mathbf{\Lambda}, \mathbb{D}) = \prod_{i}^{N} \sum_{j}^{M} \frac{w_{j}\xi_{j}(\Lambda_{j})}{\sum_{n} w_{n}\xi_{n}(\Lambda_{n})} p_{j}(d_{i}|\Lambda_{j}, \mathbb{D})$$

$$\equiv \prod_{i}^{N} \sum_{j}^{M} \phi_{j} p_{j}(d_{i}|\Lambda_{j}, \mathbb{D}). \quad (6)$$

Here  $p_j(d_i|\Lambda_j,\mathbb{D})$  refers to the likelihood defined in Eq (4) evaluated using the j-th model  $p_j(\theta|\Lambda_j)$ ,  $\phi \equiv \{\phi_1,\ldots\phi_M\}$  are the 'observed' mixture fractions – namely, the fraction of events that are generated from the corresponding mixture component after applying selection effects – and  $\mathbf{w} \equiv \{w_1,\ldots,w_M\}$  denotes the 'intrinsic' mixture fractions (as above but before the application of selection effects).

# 3. The augmented mixture model

In the previous Section we did not make any assumption regarding the specific models  $p_j(\theta|\Lambda_j)$ . In what follows, we will consider a mixture of parametric models plus one non-parametric model, whose parameters will be denoted by  $\Theta$ . In this case, differently from Eq. (5), we require that  $\Sigma_j w_j \leq 1$  to account for the presence of the additional non-parametric channel.

$$p(\theta|\mathbf{\Lambda},\Theta) = \sum_{j}^{M} w_{j} p_{j}(\theta|\Lambda_{j}) + (1 - \Sigma_{j} w_{j}) \text{NP}(\theta|\Theta).$$
 (7)

Here, we denoted with  $NP(\theta|\Theta)$  the non-parametric model, as opposed to the parametric models  $p_j(\theta|\Lambda_j)$ . We will refer to this superposition as 'augmented mixture model' (AMM).

This specific choice, namely including a non-parametric component in the mixture, addresses the possibility that the parametric models might not capture all the features that are present in the underlying distribution encoded in the data: observations that are unlikely to be explained by the available parametric models – either because they come from a region with little support or because there is an unforeseen overabundance of detections – can be captured by the non-parametric channel, acting in this case as a sort of 'additional storage' where the data that do not fit the analytical predictions can be collected. This ensures that only observations that are consistent with the functional form of the specific parametric model j are considered while estimating its parameters  $\Lambda_i$ , thus preventing mismodelling biases in the inferred posterior distribution  $p(\Lambda_i|\mathbf{d},\mathbb{D})$ . In the same fashion, the non-parametric reconstruction will be obtained making use only of the data that are unlikely to be explained by the available physically informed models and thus describing only the features that are yet to be accounted for. The property of the non-parametric model of being able to – at least in principle – approximate arbitrary probability densities translates to the AMM, making this model potentially overcomplete: this means that there might be more than one arbitrarily precise representation of the underlying data. This might be the case, for example, when a parametric model already including the features required to describe the data is augmented with a non-parametric model: both the case in which all the data are explained by the parametric model and the one where all the observations are captured by the non-parametric channel are precise descriptions of the underlying distribution. For the same reason, it is in principle possible for the non-parametric channel to 'take over' the reconstruction and explain the entirety of the data alone even if the parametric model could, in principle, account for a subset of the observations. This however is not expected to happen due to the parametric model carrying more information about the expected shape of the distribution and thus being a priori favoured in certain areas of the parameter space with respect to the completely agnostic non-parametric method.

In the remainder of this Section, we will illustrate our algorithmic implementation of the AMM. In general, the joint parameter space  $(\Lambda, \Theta)$  can be explored using a variety of techniques, mainly depending on the specific non-parametric model used: here, we will make use of the collapsed Gibbs sampling approach and (H)DPGMM (Rinaldi & Del Pozzo 2022) as non-parametric method.

# 3.1. Summary of (H)DPGMM and FIGARO

We will now recap the key aspects of both (H)DPGMM and its associated sampling scheme. For the full derivation we refer to Rinaldi & Del Pozzo (2022).

(H)DPGMM is a non-parametric model based on the Gaussian mixture model (GMM), a potentially infinite weighted sum of multivariate Gaussian distributions able to approximate arbitrary probability densities (Nguyen et al. 2020):

$$p(\theta) \simeq \text{NP}(\theta|\Theta) = \sum_{k=1}^{\infty} \lambda_k \mathcal{N}(\theta|\mu_k, \sigma_k).$$
 (8)

Here the parameters are the weights  $\lambda \equiv \{\lambda_1, \lambda_2, \ldots\}$  with  $\Sigma_k \lambda_k = 1$ , the mean vectors  $\boldsymbol{\mu} \equiv \{\mu_1, \mu_2, \ldots\}$  and the covariance matrices  $\sigma \equiv \{\sigma_1, \sigma_2, \ldots\}$ , collectively denoted with  $\Theta = \{\lambda, \mu, \sigma\}$ . In Rinaldi & Del Pozzo (2022), we introduced this non-parametric model as well as a scheme based on the collapsed Gibbs sampling approach to draw samples from the posterior distribution  $p(\Theta|\mathbf{d}, \mathbb{D})$ . This scheme is implemented in the FIGARO<sup>1</sup> code (Rinaldi & Del Pozzo 2024). The potentially infinite number of Gaussian components in the mixture is accounted for by making use of a Dirichlet process (Ferguson 1973) prior on the weights  $\lambda$ , controlled by its concentration parameter  $\alpha_{DP}$ : despite the number of components in a specific realisation of the GMM being always finite given a finite number of observations d, this choice allows, for every new data point added to the pool, to compute both the probability of the new  $d_{N+1}$ of having been drawn from one of the K already observed Gaussian components – meaning that at least one of the other data points has been drawn from each of these Kcomponents – as well as the probability for the new  $d_{N+1}$ of having been drawn from one of the infinitely many (and equally probable) unobserved components, thus effectively adding a new Gaussian component to the mixture when required by the available data.

In particular, if we introduce a set of indicator variables  $\mathbf{z} = \{z_1, \dots z_N\}$ , where each indicator variable  $z_i = k$  reads "the data  $d_i$  has been drawn from the k-th Gaussian component", it is possible to compute the probability for  $d_{N+1}$  to be drawn from the component k (Rinaldi & Del Pozzo 2022, Eq. 28),

$$p(z_{N+1} = k | \mathbf{z}, \mathbf{d}, \mathbb{D}, \alpha_{DP})$$

$$= \frac{1}{\mathcal{K}} \int p(\mathbf{d}_{z_i = k} | \mu_k, \sigma_k, \mathbb{D}) p(\mu_k, \sigma_k) d\mu_k d\sigma_k$$

$$\times p(z_{N+1} = k | \mathbf{z}, \alpha_{DP}), \quad (9)$$

where  $p(\mathbf{d}_{z_i=k}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbb{D})$  is the likelihood in Eq. (4) evaluated considering only the events assigned to the k-th component and the new  $d_{N+1}$  if k is one of the previously observed K Gaussian components or only on  $d_{N+1}$  if k=K+1, i.e., a new, previously unobserved component. The categorical probability distribution expressed by the second term on the right hand side reads (Rinaldi & Del Pozzo 2022, Eq. 25 and 26):

$$p(z_{N+1} = k | \mathbf{z}, \alpha_{\text{DP}}) = \begin{cases} \frac{\alpha_{\text{DP}}}{N + \alpha_{\text{DP}}} & \text{if } k = K+1\\ \frac{n_k}{N + \alpha_{\text{DP}}} & \text{otherwise} \end{cases}$$
 (10)

<sup>&</sup>lt;sup>1</sup> Publicly available at https://github.com/sterinaldi/FIGARO and via pip.

Here  $n_k$  denotes the number of events already associated with the component k. Summing over all the possible components plus the new, unobserved one gives the normalisation constant K:

$$\mathcal{K} = \sum_{k=1}^{K+1} p(z_{N+1} = k | \mathbf{z}, \mathbf{d}, \mathbb{D}, \alpha_{DP})$$
(11)

The introduction of indicatory variables is particularly useful when it comes to sampling the posterior distribution  $p(\Theta|\mathbf{d}, \mathbb{D})$ : since each Gaussian component of the mixture is independent from each other, conditioning on  $\mathbf{z}$  makes the parameter space partially separable:

$$p(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\sigma} | \mathbf{d}, \mathbb{D}, \mathbf{z}, \alpha_{\text{DP}}) = p(\boldsymbol{\lambda} | \mathbb{D}, \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \alpha_{\text{DP}}) \times \prod_{k=1}^{K} p(\mu_{k}, \sigma_{k} | \mathbf{d}_{z_{i}=k}, \mathbb{D}). \quad (12)$$

The posterior distribution on  $\lambda$ , under the assumption of a Dirichlet prior, reads

$$p(\boldsymbol{\lambda}|\mathbb{D}, \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \alpha_{\mathrm{DP}}) = \int p(\boldsymbol{\lambda}|\boldsymbol{\phi}, \mathbb{D}, \boldsymbol{\mu}, \boldsymbol{\sigma}) p(\boldsymbol{\phi}|\mathbf{z}, \alpha_{\mathrm{DP}}) d\boldsymbol{\phi}$$
$$= \int \prod_{k} \delta(\lambda_{k} - \phi_{k}/\xi(\mu_{k}, \sigma_{k})) p(\boldsymbol{\phi}|\mathbf{z}, \alpha_{\mathrm{DP}}), \quad (13)$$

where we made use of the definitions given in Eq. (3) for  $\xi(\mu, \sigma)$  and Eq. (6) for  $\phi$ , and

$$p(\boldsymbol{\phi}|\mathbf{z}, \alpha_{\mathrm{DP}}) = \Gamma(N + \alpha_{\mathrm{DP}}) \prod_{k=1}^{K} \frac{\phi_k^{(n_k + \alpha_{\mathrm{DP}}/K) - 1}}{\Gamma(n_k + \alpha_{\mathrm{DP}}/K)}.$$
(14)

This parameter space is explored by FIGARO using the collapsed Gibbs sampling approach. The Gibbs sampling scheme (Geman & Geman 1984; Gelfand & Smith 1990; Smith & Roberts 1993) is useful in all these situations where directly sampling from the joint parameter space—in our case,  $p(\Theta, \mathbf{z}|\mathbf{d}, \mathbb{D})$ —is expensive or impossible, but sampling from the conditioned distributions  $p(\mathbf{z}|\Theta, \mathbf{d}, \mathbb{D})$  and  $p(\Theta|\mathbf{z}, \mathbf{d}, \mathbb{D})$  is simple. If one of the conditional probabilities can be efficiently marginalised over the conditioned variable—thus making the exploration of the parameter space even simpler—the scheme is then referred to as collapsed Gibbs sampling (Liu 1994): this is the case, for example, of  $p(\mathbf{z}|\mathbf{d}, \mathbb{D})$  in Eq. (9).

Operatively, FIGARO draws a single sample from the posterior distribution  $p(\Theta|\mathbf{d}, \mathbb{D})$  in the following way:

- 1. Draw a sample for **z**:
  - (a) Consider an empty mixture model, with no observations. Out of the N available events, randomly pick one and assign it to the first component. At this stage,  $\mathbf{z} = \{z_1 = 1\}$ ;
  - (b) Randomly pick a new event from the N-1 left. Using Eq. (9) conditioned on the current value of  $\mathbf{z}$ , compute the probability of assigning this second event to the same component as the first one  $(z_2 = 1)$  or to a new one  $(z_2 = 2)$  and draw  $z_2$  accordingly;
  - (c) Repeat this procedure until all the observations have been added to the mixture, which will eventually have K active components. This will be the  $\mathbf{z}$  sample.
- 2. Draw a sample for  $\Theta$  conditioned on the **z** sample:

- (a) For each of the K active components of the mixture, draw a sample for  $(\mu_k, \sigma_k)$  using each of the terms in the product of Eq. (12). These will be the  $\mu$  and  $\sigma$  samples;
- (b) From Eq. (14), draw a sample for  $\phi$ ;
- (c) Using the  $\mu$  and  $\sigma$  samples to compute  $\xi(\mu_k, \sigma_k)$ , convert the  $\phi$  sample into a  $\lambda$  sample to obtain  $\Theta = {\lambda, \mu, \sigma}$ .

These steps can be repeated to produce as many samples for  $\Theta$  as needed. The **z** samples, at this stage, are just a byproduct and can be discarded.

### 3.2. Inferring the parameters of an AMM

We now turn our attention to the problem of exploring the parameter space of the AMM, leveraging on the same sampling scheme used for the non-parametric method. From a mathematical point of view, most of the derivation is identical to the one we just summarised for the non-parametric methods. The main difference is that now the number of components – parametric models plus a non-parametric one – is finite and equal to M+1. Therefore, the prior on the weights  $\mathbf{w}$  is not a Dirichlet process but its finite equivalent the Dirichlet distribution. This distribution takes as input a vector of M+1 positive numbers  $\boldsymbol{\gamma}=\{\gamma_1,\ldots,\gamma_{M+1}\}$  acting as a priori pseudo-counts for each channel. The choice in which  $\gamma_i=1$   $\forall i$  corresponds to the uniform distribution on the M-dimensional simplex.

In particular, denoting with  $\zeta = \{\zeta_1, \dots, \zeta_N\}$  the indicator variables for the augmented mixture model, Eq. (9) becomes

$$p(\zeta_{N+1} = j | \boldsymbol{\zeta}, \mathbf{z}, \mathbf{d}, \mathbb{D}, \boldsymbol{\gamma}, \alpha_{DP})$$

$$= \frac{1}{C} \int p_j(\mathbf{d}_{\zeta_i = j} | \Lambda_j, \mathbb{D}) p(\Lambda_j) d\Lambda_j$$

$$\times p(\zeta_{N+1} = j | \boldsymbol{\zeta}, \boldsymbol{\gamma}), \quad (15)$$

if j < M + 1, thus corresponding to a parametric model, and

$$p(\zeta_{N+1} = M + 1 | \boldsymbol{\zeta}, \mathbf{z}, \mathbf{d}, \mathbb{D}, \boldsymbol{\gamma}, \alpha_{\mathrm{DP}})$$

$$= \frac{1}{\mathcal{C}} \int p(\mathbf{d}_{\zeta_i = M+1} | \boldsymbol{\Theta}, \mathbb{D}, \mathbf{z}, \alpha_{\mathrm{DP}}) p(\boldsymbol{\Theta}) d\boldsymbol{\Theta}$$

$$\times p(\zeta_{N+1} = M + 1 | \boldsymbol{\zeta}, \boldsymbol{\gamma}), \quad (16)$$

when considering the non-parametric model. The integrand can be further broken down using z:

$$p(\mathbf{d}_{\zeta_{i}=M+1}|\Theta, \mathbb{D}, \mathbf{z}, \alpha_{\mathrm{DP}})$$

$$= \sum_{k=1}^{K+1} \int p(\mathbf{d}_{\zeta_{i}=M+1, z_{i}=k} | \mu_{k}, \sigma_{k}, \mathbb{D}) p(\mu_{k}, \sigma_{k}) d\mu_{k} d\sigma_{k}$$

$$\times p(z_{N+1} = k | \mathbf{z}, \alpha_{\mathrm{DP}}). \quad (17)$$

These are the same terms used in the non-parametric inference presented in the previous Subsection just with the additional condition of considering only the events actually associated with the non-parametric channel. The categorical probability distribution corresponding to the one in Eq. (10) reads, due to the finite number of mixture components,

$$p(\zeta_{N+1} = j | \boldsymbol{\zeta}, \boldsymbol{\gamma}) = \frac{n_j + \gamma_j}{N + \Sigma_n \gamma_n}.$$
 (18)

As in Eq. (11), the normalisation constant C in Eqs. (15) and (16) is simply the sum over all the possible models:

$$C = \sum_{j=1}^{M+1} p(\zeta_{N+1} = j | \boldsymbol{\zeta}, \mathbf{z}, \mathbf{d}, \mathbb{D}, \boldsymbol{\gamma}, \alpha_{DP}).$$
 (19)

Finally, the posterior distribution on  $(\mathbf{w}, \mathbf{\Lambda}, \Theta)$  can be expressed as

$$p(\mathbf{w}, \boldsymbol{\Lambda}, \Theta | \mathbf{d}, \mathbb{D}, \boldsymbol{\zeta}, \mathbf{z}, \boldsymbol{\gamma}, \alpha_{DP}) = p(\mathbf{w} | \boldsymbol{\Lambda}, \Theta, \mathbb{D}, \boldsymbol{\zeta}, \boldsymbol{\gamma})$$
$$\times p(\Theta | \mathbf{d}_{\zeta_i = M+1}, \mathbb{D}, \mathbf{z}, \alpha_{DP}) \prod_{j=1}^{M} p(\Lambda_i | \mathbf{d}_{\zeta_i = j}, \mathbb{D}), \quad (20)$$

where  $p(\Theta|\mathbf{d}_{\zeta_i=M+1}, \mathbb{D}, \mathbf{z}, \alpha_{\mathrm{DP}})$  is given by Eq. (12) and

$$p(\mathbf{w}|\mathbf{\Lambda}, \Theta, \mathbb{D}, \boldsymbol{\zeta}, \boldsymbol{\gamma}) = \int p(\mathbf{w}|\boldsymbol{\phi}, \mathbb{D}, \mathbf{\Lambda}, \Theta) p(\boldsymbol{\phi}|\boldsymbol{\zeta}, \boldsymbol{\gamma}) d\boldsymbol{\phi}$$
$$= \int \delta(w_{M+1} - \phi_{M+1}/\xi(\Theta)) \prod_{j=1}^{M} \delta(w_j - \phi_j/\xi_j(\Lambda_j))$$
$$\times p(\boldsymbol{\phi}|\boldsymbol{\zeta}, \boldsymbol{\gamma}) d\boldsymbol{\phi}. \quad (2$$

The posterior distribution for  $\phi$  is the same as Eq. (14):

$$p(\boldsymbol{\phi}|\boldsymbol{\zeta},\boldsymbol{\gamma}) = \Gamma(N + \Sigma_n \gamma_n) \prod_{j=1}^{M+1} \frac{w_j^{n_j + \gamma_j - 1}}{\Gamma(n_j + \gamma_j)}.$$
 (22)

In this derivation we tacitly assumed that the parametric models have completely disjoint sets of parameters. If this is not the case, the only care to be taken is that the models sharing at least one parameter have to be inferred jointly.

With all these ingredients, we can now draw samples from the posterior distribution  $p(\mathbf{\Lambda}, \Theta | \mathbf{d}, \mathbb{D})$ :

- 1. Draw a sample for  $\zeta$  and z:
  - (a) Starting from a empty mixture, randomly select one of the available events and compute the probability of assigning it to one of the parametric models (Eq. (15)) or to the non-parametric component (Eq. (16)): then, draw  $\zeta_1$  accordingly. If  $\zeta_1 = M + 1$ , also assign the event to one of the Gaussian components of the non-parametric model;
  - (b) Repeat the previous step adding all the N-1 remaining events in random order, each time conditioning on the current values of  $\boldsymbol{\zeta}$  and  $\mathbf{z}$ . Events for which  $\zeta_i = M+1$  should also be used to update  $\mathbf{z}$  following the procedure described in Section 3.1. This will produce a sample for  $\boldsymbol{\zeta}$  and one for  $\mathbf{z}$ .
- 2. Draw a sample for  $\Lambda$ ,  $\Theta$  and  $\mathbf{w}$  conditioned on  $\boldsymbol{\zeta}$  and  $\mathbf{z}$ :
  - (a) For each parametric model, draw a sample for  $\Lambda_j$  using the corresponding term of the product in Eq. (20), producing a sample for  $\Lambda$ ;
  - (b) Draw a sample for  $\Theta$  as in Section 3.1;
  - (c) From Eq. (22), draw a sample for  $\phi$ ;
  - (d) Using the  $\Lambda$  and  $\Theta$  sample to compute  $\xi_j(\Lambda_j)$  and  $\xi(\Theta)$ , convert the  $\phi$  sample into a **w** sample.

Again, this procedure can be iterated to produce multiple  $(\mathbf{w}, \mathbf{\Lambda}, \Theta)$  samples. In this case, the auxiliary variable  $\boldsymbol{\zeta}$  can be useful in determining which channel is more likely to have produced each of the available events.

To accompany this derivation, we developed ANUBIS, a Python code implementing the algorithm presented in this Section. ANUBIS relies on FIGARO for the non-parametric inference and on ERYN (Foreman-Mackey et al. 2013; Karnesis et al. 2023; Katz et al. 2023) as MCMC sampler. The code is publicly available at https://github.com/sterinaldi/ANUBIS and via pip, and will be used in the analyses of the following Sections.

# 4. Simulated data

To demonstrate the capability of AMMs to capture features in distributions, we will apply the framework described in the previous Section to two different simulated scenarios inspired by the BH distribution presented in Abbott et al. (2023b). The inferred values for the parameters will be reported quoting the median and 68% credible interval.

#### 4.1. One-dimensional distribution

The first, simplified example presented here is a one-dimensional distribution describing the BBH primary mass. The underlying distribution is assumed to be a truncated PowerLaw (PL) distribution with spectral index  $\alpha$ , bounded between between  $M_{\rm min}=3~{\rm M}_{\odot}$  and  $M_{\rm max}=80~{\rm M}_{\odot}$  (assumed known), superimposed to a Gaussian distribution:

$$p(M|\alpha, \mu, \sigma, w) = w_{\text{PL}} PL(M|\alpha) + w_{\text{Peak}} \mathcal{N}(M|\mu, \sigma).$$
 (23)

The PL distribution is defined as

$$PL(M|\alpha) = \begin{cases} \frac{(1-\alpha)M^{-\alpha}}{M_{\text{max}}^{1-\alpha} - M_{\text{min}}^{1-\alpha}} & \text{if} \quad M_{\text{min}} < M < M_{\text{max}} \\ 0 & \text{otherwise} \end{cases}$$
(24)

To generate the mock data presented in this Section we use  $\alpha=3.5,\,\mu=35~{\rm M}_{\odot},\,\sigma=4~{\rm M}_{\odot}$  and  $w_{\rm Peak}=0.05$ . The selection function through which we filtered the data, reported in Figure 1a as a gray dashed line, is modelled after the one presented in Veske et al. (2021). From this distribution we draw 100 observed values  $\boldsymbol{M}=\{M_1,\ldots,M_{100}\}$ , and for each of these we simulated a posterior distribution  $p(\theta_i|d_i)$  as a log-normal distribution as

$$p(\theta_i|d_i) = \mathcal{N}(\log M|\log m_i, \sigma_r) \tag{25}$$

where  $\sigma_r = 0.15$  for all events and

$$m_i \sim \mathcal{N}(\log m | \log M_i, \sigma_r)$$
 (26)

In this simplified example, we do not account for the correlation between detection efficiency and measurement uncertainty: we will include this feature in the more realistic simulation presented in the next Section. In what follows, we will analyse these simulated data with different AMMs.

# 4.1.1. PL + NP

In this first test, we assume a PL model augmented with the non-parametric channel (PL+NP), simulating the scenario where we have a good theoretical understanding of only one BBH formation channel. The free parameters are the PL index  $\alpha$  and the relative weights  $\mathbf{w} = \{w_{\mathrm{NP}}, w_{\mathrm{PL}}\}$ . The inferred distribution is reported in Figure 1a. The inferred value for the PL index is  $\alpha = 3.5 \pm 0.4$ , consistent with the

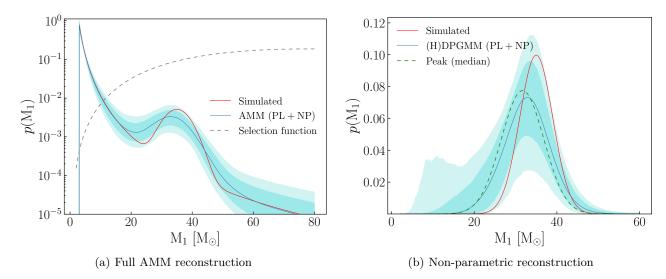
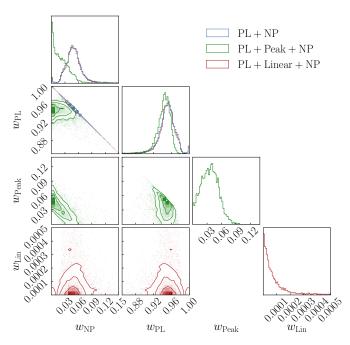


Fig. 1. Inferred distribution for the one-dimensional POWERLAW+PEAK example presented in Section 4.1 in the PL+NP case. The blue solid line marks the median reconstruction, the shaded areas correspond to the 68% and 90% credible regions and the solid red line shows the true underlying distribution. The gray dashed line in the left panel corresponds to the selection function used.



**Fig. 2.** Inferred weights for the different mixture models considered in Section 4.1.

simulated value. For comparison, if we do not include the non-parametric channel, we get  $\alpha=2.4\pm0.1$  – value biased by the presence in the data of a feature not accounted for in the model.

The non-parametric reconstruction, reported in Figure 1b, highlights the presence of a feature at around  $30-35~{\rm M}_{\odot}$  consistent with the simulated Gaussian distribution. The weights  $(w_{\rm NP},~w_{\rm PL})$  are reported in Figure 2 (blue histograms). The inferred non-parametric weight is  $w_{\rm NP}=0.05^{+0.02}_{-0.01}$ , in agreement with the simulated value  $w_{\rm Peak}=0.05$  that the non-parametric channel is expected to capture.

#### 4.1.2. PL + Linear + NP

Secondly, we show that the AMM is robust with respect to the inclusion of a "useless" component – a channel not part of the underlying distribution – in the mixture. In particular, we add a linear distribution bounded between the same  $M_{\rm min}$  and  $M_{\rm max}$  as the PL distribution:

$$\operatorname{Lin}(M) = \begin{cases} \frac{2M}{M_{\text{max}}^2 - M_{\text{min}}^2} & \text{if} \quad M_{\text{min}} < M < M_{\text{max}} \\ 0 & \text{otherwise} \end{cases} . (27)$$

The red histogram in Figure 2 shows how the inference is unaffected by the presence of an additional, unused channel: the weight associated with the linear distribution is found to be negligible and thus such channel do not contribute to the overall inference, leading to a reconstruction almost identical to the PL+NP case.

# 4.1.3. PL + Peak + NP

The last case considered here is the one in which we have all the necessary components in the parametric mixture to describe the underlying distribution, thus making the non-parametric channel redundant: in particular, we include in the analysis a PL distribution, a Gaussian peak and a non-parametric model. We find, for the parameters of this mixture,  $\alpha=3.7\pm0.4,\,\mu=32^{+3}_{-4}~{\rm M}_{\odot}$  and  $\sigma=5^{+2}_{-3}~{\rm M}_{\odot}$  – all in agreement with the simulated values. Figure 1b reports, in dashed green, the Gaussian distribution corresponding to the median inferred  $\mu$  and  $\sigma$ : this is in good agreement with the non-parametric feature reconstructed in the PL+NP case.

The green histograms in Figure 2 shows the posterior distribution for the three weights  $(w_{\rm NP}, w_{\rm PL}, w_{\rm Peak})$ : in this case, the presence of additional, non-parametric features is disfavoured by the presence of the Gaussian channel  $(w_{\rm NP}=0.01^{+0.02}_{-0.01})$  and  $w_{\rm Peak}=0.04\pm0.02$ ). The suppression of the unused channel is however not as confident as it is in the previous PL+Linear+NP case due to the flexibility of the non-parametric model: being the mixture model overcomplete, it is always possible to associate some of the

available events with the NP channel, thus resulting in a non-zero associated weight. Nonetheless, the shape of the posterior distribution for  $w_{\rm NP}$  suggests that the parametric models are sufficient, on their own, to explain the data.

#### 4.2. Three-dimensional distribution

We now consider a more realistic example modelled after the real LVK GW detections, including in the analysis the primary mass  $M_1$ , the mass ratio q and the redshift z. The primary mass follows a POWERLAW+PEAK distribution (described in Appendix B1.b of Abbott et al. 2023b) with parameters  $\alpha=3.5,\,\delta=5~M_{\odot},\,M_{\rm min}=3~M_{\odot},\,\mu=35~M_{\odot},\,\sigma=4~M_{\odot}$  and  $w_{\rm Peak}=0.02.$  The maximum mass  $M_{\rm max}=80~M_{\odot}$  is assumed known. The mass ratio is distributed according to a power-law

$$p(q|\beta) \propto q^{\beta}$$
 (28)

with  $\beta = 1.1$  and the redshift is

$$p(z|\kappa) \propto \frac{\mathrm{d}V}{\mathrm{d}z} (1+z)^{\kappa-1}$$
 (29)

with  $\kappa=2$  between z=0 and z=2. For all the other parameters (spins and extrinsic parameters) we used isotropic/uniform distributions, assumed known. From these distribution, we draw  $10^5$  binaries and inject the corresponding signals in simulated noise representative of O3 sensitivity. For each signal we compute the network signal-to-noise ratio  $\rho$ , randomly select 59 events with  $\rho>10$  and estimate their parameters using BILBY (Ashton et al. 2019) to produce a mock catalogue with properties similar to the BBHs observed during O3. The selection effects can therefore accounted for using the search sensitivity estimates for O3 (LIGO Scientific et al. 2023b, v2) released along with the third Gravitational Wave Transient Catalog (GWTC-3).

We analyse this mock catalogue using the tapered PL (TPL) of the POWERLAW+PEAK model augmented with (H)DPGMM. The posterior distribution for the parameters  $\Lambda_{\rm TPL} = \{\alpha, \delta, M_{\rm min}, \kappa, \beta\}$  is reported in Figure 3a, whereas Figure 3b shows the non-parametric reconstruction of the marginal  $M_1-z$  distribution. The inferred distributions are consistent with the expectations, being both in agreement with the simulated values and highlighting the presence of the Gaussian feature at around 35  $\rm M_{\odot}$ .

The interpretation of the relative weights  $(w_{\rm NP}, w_{\rm TPL})$  requires some care, however, due to the presence of a censored area in the binary parameter space – namely the low-mass, high-redshift region. Since the GW detectors are not able to detect gravitational signals with such properties, the inferred population distribution will have to rely on extrapolation to cover that specific region of the binary parameter space. The non-parametric methods, however, are by construction informed only by the available data, thus unable to extrapolate beyond where data are observable, the detector horizon: reconstructing the intrinsic distribution in censored areas would yield diverging uncertainties. To prevent this issue,

FIGARO is coded such that the inferred distribution vanishes in the censored areas, effectively preventing extrapolation towards large redshifts. This prescription, limiting the number of high-redshift binaries, affects the calculation of the detectability fraction  $\xi(\Theta)$ , skewing the inference of  $w_{\rm NP}$  towards smaller values. In such cases it can become difficult to distinguish whether the presence of the non-parametric channel is required by the data or not. The posterior distribution for the observed mixture fraction  $\phi$ , reported in Figure 4, solves the issue showing that the non-parametric channel is actually required to account for  $\sim 35\%$  of the observed GW events.

### 5. BBHs from GWTC-3

Having demonstrated the capability of our approach in simulated scenarios, we will now analyse the primary mass, mass ratio and redshift distribution of BBHs using the publicly available GW events detected by the LVK collaboration. In this section, we will make use of the 69 GW events released in GWTC-3 (Abbott et al. 2023a) with false alarm rate  $< 1~\rm{yr}^{-1}$ . The selection effects are accounted for using the sensitivity estimates for the first three observing runs (O1+O2+O3 – LIGO Scientific et al. 2023a, v2) released along with GWTC-3. This choice is driven by the fact that the focus of this work is on the new methodology that we are introducing rather than on the astrophysical interpretation of the findings. An application of this method to the newly released GWTC-4.0 and a discussion of the results yielded will be the subject of a future paper.

#### 5.1. Parametrised model

In this first analysis, we apply the same AMM used in Section 4.2, TPL+NP, with  $\Lambda_{TPL}=\{\alpha,\delta,M_{\min},\kappa,\beta\}.$   $M_{\max}$  is fixed at  $100~M_{\odot}$  to ensure that the parametric distribution has non-zero support on the whole mass range. This analysis is similar in concept to the one presented in Rinaldi et al. (2025a), where we analysed the same dataset using a weighted superposition of a TPL and a Gaussian peak with independent redshift evolutions: following this paper, we include in the analysis only the primary mass, mass ratio, and redshift and we assume a fixed isotropic spin distribution (see Section 3 of Rinaldi et al. 2025a). Moreover, we will compare our findings with the ones presented in the aforementioned paper.

The posterior distribution for  $\Lambda_{\rm TPL}$  is reported in Figure 5, which is found to be in good agreement with Rinaldi et al. (2025a). The differences between the two distributions are to be ascribed to the fact that the NP channel used in this analysis has more flexibility than the Gaussian peak, thus affecting differently the inference of the TPL parameters. The reconstructed marginal non-parametric distribution for  $M_1$  and z is reported in Figure 6a, highlighting the presence of the  $\sim 35~{\rm M}_{\odot}$  pileup. This non-parametric reconstruction can be converted into a posterior on the parameters of a Gaussian distribution using the remapping procedure presented in Rinaldi et al. (2025b): the result of this remapping is shown in Figure 6b. We find  $\mu=40^{+7}_{-8}~{\rm M}_{\odot}$  and  $\sigma=6^{+4}_{-2}~{\rm M}_{\odot}$ , in agreement with the expectation of  $\mu\sim35~{\rm M}_{\odot}$  and  $\sigma\sim4~{\rm M}_{\odot}$  (Abbott et al. 2023b). The inferred relative weight of the NP channel is  $w_{\rm NP}=0.003^{+0.2}_{-0.003}$ , corresponding to an observed

The noise curves used are publicly available here: https://dcc.ligo.org/LIGO-T2000012-v1/public

<sup>&</sup>lt;sup>3</sup> A simple way of picturing this is saying that the non-parametric method is "unable to decide" whether the lack of observations in an empty area of the binary parameter space is either due to the intrinsic distribution not having support in such area or the selection function completely depleting it.

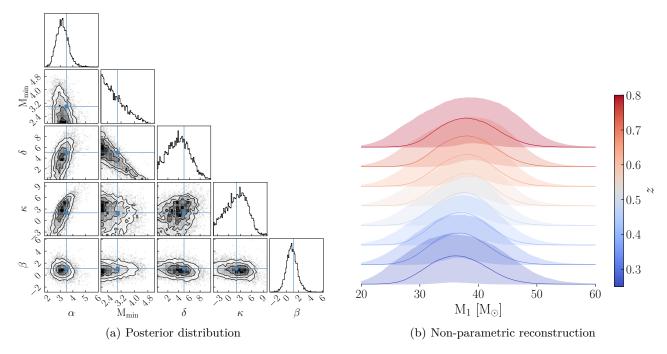


Fig. 3. Inferred distributions for the three-dimensional mock catalogue presented in Section 4.2. Left: posterior distribution for  $\Lambda_{\text{TPL}}$ . The blue cross-hairs mark the true values. Right: non-parametric marginal  $M_1 - z$  distribution. For each redshift value, the mass distribution has been normalised, and the shaded areas mark the 68% credible regions.

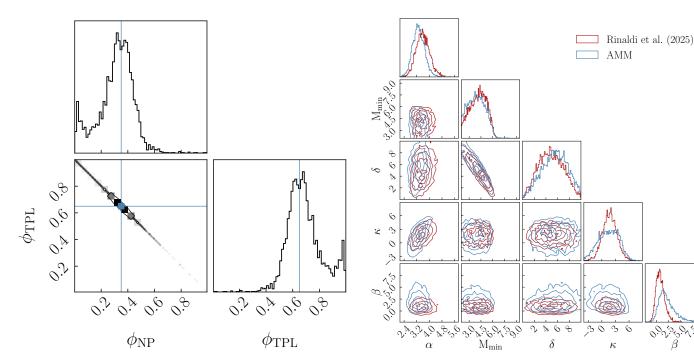


Fig. 4. Posterior distribution for  $\phi$  using the three-dimensional mock catalogue of Section 4.2. The blue cross-hairs mark the true values.

relative weight  $\phi_{\rm NP}=0.20^{+0.08}_{-0.11}$  – 13 GW events associated with the non-parametric channel.

# 5.2. Population synthesis

In this last example, we will show that astrophysical models can also be augmented with non-parametric methods. We consider an isolated evolution channel using the rapid binary

Fig. 5. Posterior distribution for  $\Lambda_{TPL}$  using GWTC-3.

population synthesis code SEVN<sup>4</sup> (Spera et al. 2019; Mapelli et al. 2020; Iorio et al. 2023b), and in particular the publicly available catalogues<sup>5</sup> (Iorio et al. 2023a, v2) released alongside Iorio et al. (2023b). For simplicity, we make use of the Fiducial model assuming only one metallicity ( $Z=10^{-3}$ ) and one value for the common envelope efficiency ( $\alpha_{\rm CE}=3$ ). This model is described in detail in Section 3.2 of Iorio

<sup>&</sup>lt;sup>4</sup> Publicly available at https://gitlab.com/sevncodes/sevn

https://zenodo.org/records/7794546

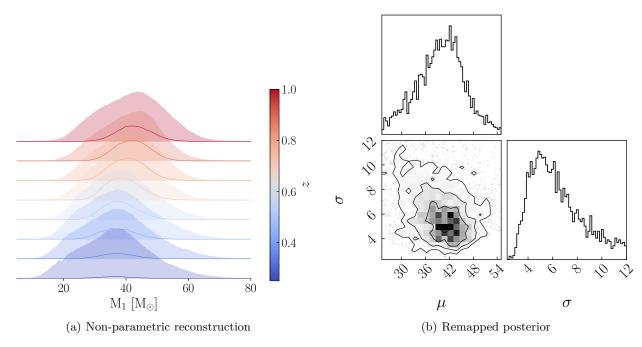


Fig. 6. Left:  $M_1 - z$  non-parametric reconstruction using GWTC-3 events. For each redshift value, the mass distribution has been normalised, and the shaded areas mark the 68% credible regions. Right: posterior distribution for  $\mu$  and  $\sigma$  of a Gaussian distribution obtained from the non-parametric reconstruction using the remapping procedure presented in Rinaldi et al. (2025b).

et al. (2023b). To get a probability density function, we fit a GMM approximant to the the  $M_1$  and q SEVN samples. Other choices for the approximant are also possible, such as the normalising flow emulator presented in Colloms et al. (2025) that allows for the exploration of different values for  $\alpha_{CE}$  and Z. The redshift model is the one given in Eq. (29): therefore, in this case,  $\Lambda_{SEVN} = \{\kappa\}$ .

We find that, for the SEVN+NP model,  $\kappa = 4.2^{+2.5}_{-3.9}$ . Concerning the relative importance of the two channels, we get  $w_{\rm NP} = 0.01^{+0.1}_{-0.01}$ , corresponding to an observed relative weight  $\phi_{\rm NP} = 0.35 \pm 0.14$ . This suggests that around 44 out of 69 GW events can be explained using the isolated evolution model, whereas the remaining ones have to be accounted for using the non-parametric channel.

When comparing the primary mass non-parametric reconstructions of the SEVN+NP and the TPL+NP models, reported in Figure 7a, we see that the former has more support towards more massive BHs. This is a direct consequence of the difference in support between the PL model – deliberately extended all the way to  $M_{\rm max} = 100~M_{\odot}$  – and the SEVN catalogue, which is cut at around 40  $M_{\odot}$ due to the pair instability model used (both reported in Figure 7b). This points towards the fact that not only the isolated evolution model used in this Section cannot describe the  $\sim 35~{\rm M}_{\odot}$  feature, but also that some additional channels (i.e., dynamical models) are needed to explain the high-mass end of the spectrum. We remind, however, that the example presented here is purely for demonstration purposes: an in-depth study investigating the features that population synthesis codes are able or unable to predict is outside of the scope of this work and will be the subject of a future paper.

# 6. Summary

In this paper, we introduced the concept of augmented mixture model (AMM), a model designed to equip parametric and/or astrophysical models with the capability of accommodating unforeseen features in the analysed data: we achieved this by building a weighted superposition of parametric models plus an additional non-parametric model. We validated our formalism applying it to the reconstruction of simulated datasets mimicking the currently available GW observations. We also analysed the real GW events from GWTC-3, finding good agreement with the available literature about the most pronounced features of the primary mass distribution.

The possibility of equipping astrophysical models with non-parametric methods will be extremely valuable to both data analysts and theoreticians. Having a channel able to collect events that are unlikely to be explained by the physically informed models included in the analysis will prevent biases due to mismodelling as well as pointing to the direction in which astrophysical models need to be further developed: with the grand total of observed GWs after the first third of O4 just above 150 detection, the newly released GWTC-4.0 will be the perfect playground for AMMs to possibly reveal new features in the astrophysical distribution of BBHs while strengthening our understanding of the formation paths of the compact objects that populate our Universe.

Acknowledgements. The author is grateful to Walter Del Pozzo, Gabriele Demasi, and Michela Mapelli for useful discussion, to Vasco Gennari, Giuliano Iorio, and Amedeo Romagnolo for comments on the initial draft of this paper, and to Gabriele also for providing the script used to generate the mock data used in Section 4.2. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 546677095. The author acknowledges financial support from the German Excellence Strategy via the Heidelberg Cluster of Excellence (EXC 2181 - 390900948) STRUC-

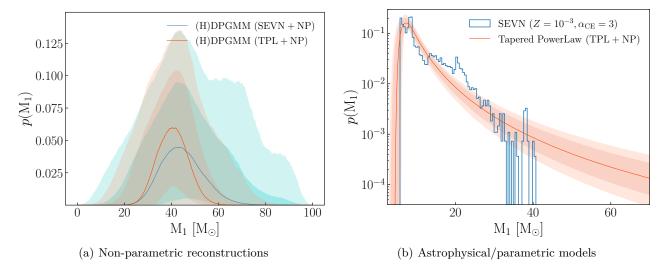


Fig. 7. Left: marginal non-parametric reconstruction for  $M_1$  in the SEVN+NP case (blue) and TPL+NP case (red). Right: comparison between the SEVN catalogue used in Section 5.2 (blue histogram) and the TPL inferred in Section 5.1 (solid red). In both panels, the shaded areas represent the 68% and 90% credible regions.

TURES, from the state of Baden-Württemberg through bwHPC, from the German Research Foundation (DFG) through grants INST 35/1597-1 FUGG and INST 35/1503-1 FUGG and from the European Research  $Council \ for \ the \ ERC \ Consolidator \ grant \ DEMOBLACK, \ under \ contract$ no. 770017. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, 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.

# References

Aasi, J. et al. 2015, Classical and Quantum Gravity, 32, 074001

Abbott, B. P. et al. 2019, ApJ, 882, L24

Abbott, R., Abbott, T. D., Acernese, F., et al. 2023a, Physical Review X, 13, 041039

Abbott, R. et al. 2021, ApJ, 913, L7  $\,$ 

Abbott, R. et al. 2023b, Physical Review X, 13, 011048

Acernese, F. et al. 2015, Classical and Quantum Gravity, 32, 024001 Akutsu, T. et al. 2020, Prog. Theor. Exp. Phys., 2021, 05A101

Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27

Banerjee, S. 2022, A&A, 665, A20

Belczynski, K. et al. 2016, A&A, 594, A97

Callister, T. A. 2024, arXiv e-prints, arXiv:2410.19145

Callister, T. A. & Farr, W. M. 2024, Physical Review X, 14, 021005
 Colloms, S., Berry, C. P. L., Veitch, J., & Zevin, M. 2025, ApJ, 988, 189

Farah, A. M., Edelman, B., Zevin, M., et al. 2023, The Astrophysical Journal, 955, 107

Ferguson, T. S. 1973, Ann. Stat., 1, 209

Fishbach, M. & Holz, D. E. 2017, The Astrophysical Journal Letters, 851, L25 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013,
Publications of the Astronomical Society of the Pacific, 125, 306
Gallegos-Garcia, M., Berry, C. P. L., Marchant, P., & Kalogera, V. 2021, ApJ, 922, 110

 Gelfand, A. E. & Smith, A. F. M. 1990, J. Am. Stat. Assoc., 85, 398
 Geman, S. & Geman, D. 1984, IEEE Trans. Patt. An. Mach. Int., PAMI-6, 721

Gennari, V., Mastrogiovanni, S., Tamanini, N., Marsat, S., & Pierra, G. 2025, Phys. Rev. D, 111, 123046

Heinzel, J., Mould, M., Álvarez-López, S., & Vitale, S. 2025, Phys. Rev. D, 111, 063043

Hirschi, R., Goodman, K., Meynet, G., et al. 2025, MN-RAS[arXiv:2508.21233]

Iorio, G., Costa, G., Mapelli, M., et al. 2023a, Dataset from the paper "Compact object mergers: exploring uncertainties from stellar and binary evolution with SEVN", https://zenodo.org/records/ 7794546

Iorio, G., Mapelli, M., Costa, G., et al. 2023b, MNRAS, 524, 426
 Karnesis, N., Katz, M. L., Korsakova, N., Gair, J. R., & Stergioulas,
 N. 2023, MNRAS, 526, 4814

Katz, M., Karnesis, N., & Korsakova, N. 2023, mikekatz04/Eryn: first full release

Kremer, K., Spera, M., Becker, D., et al. 2020, ApJ, 903, 45
 Kruckow, M. U., Andrews, J. J., Fragos, T., et al. 2024, A&A, 692, A141

Li, Y.-J., Wang, Y.-Z., Han, M.-Z., et al. 2021, The Astrophysical Journal, 917, 33

LIGO Scientific, C., Virgo, C., & KAGRA, C. 2023a, 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, https://zenodo.org/records/7890398

LIGO Scientific, C., Virgo, C., & KAGRA, C. 2023b, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O3 Search Sensitivity Estimates, https://zenodo.org/records/7890437

Liu, J. S. 1994, J. Am. Stat. Assoc., 89, 958

Mandel, I. & Farmer, A. 2022, Physics Reports, 955, 1

Mandel, I., Farr, W. M., & Gair, J. R. 2019, MNRAS, 486, 1086

Mapelli, M. 2020, Frontiers in Astronomy and Space Sciences, 7

Mapelli, M., Spera, M., Montanari, E., et al. 2020, ApJ, 888, 76
Marchant, P., Pappas, K. M. W., Gallegos-Garcia, M., et al. 202

Marchant, P., Pappas, K. M. W., Gallegos-Garcia, M., et al. 2021, A&A, 650, A107

Merritt, J., Stevenson, S., Sander, A., et al. 2025, arXiv e-prints, arXiv:2507.17052

Nguyen, T. T., Nguyen, H. D., Chamroukhi, F., & McLachlan, G. J. 2020, Cog. Math. St., 7, 1750861

Ray, A., Magaña Hernandez, I., Mohite, S., Creighton, J., & Kapadia, S. 2023, ApJ, 957, 37

Rinaldi, S. & Del Pozzo, W. 2022, MNRAS, 509, 5454

- Rinaldi, S. & Del Pozzo, W. 2024, Journal of Open Source Software, 9, 6589
- Rinaldi, S., Liang, Y., Demasi, G., Mapelli, M., & Del Pozzo, W. 2025a, arXiv e-prints, arXiv:2506.05929
- Rinaldi, S., Toubiana, A., & Gair, J. R. 2025b, arXiv e-prints, arXiv:2506.05153
- Rodriguez, C. L. et al. 2016, Phys. Rev. D, 93, 084029
- Romagnolo, A., Gormaz-Matamala, A. C., & Belczynski, K. 2024, ApJ, 964, L23
- Röpke, F. K. & De Marco, O. 2023, Living Reviews in Computational Astrophysics, 9, 2
- Smith, A. F. M. & Roberts, G. O. 1993, J. Royal Stat. Soc., 55, 3
- Spera, M., Mapelli, M., Giacobbo, N., et al. 2019, MNRAS, 485, 889 Talbot, C. & Thrane, E. 2018, The Astrophysical Journal, 856, 173
- The LIGO Scientific Collaboration, the Virgo Collaboration, & the KAGRA Collaboration. 2025a, arXiv e-prints, arXiv:2508.18083
- The LIGO Scientific Collaboration, The Virgo Collaboration, & the
- KAGRA Collaboration. 2025b, arXiv e-prints, arXiv:2508.18082 Toubiana, A., Katz, M. L., & Gair, J. R. 2023, MNRAS, 524, 5844
- van Son, L. A. C., Roy, S. K., Mandel, I., et al. 2025, ApJ, 979, 209
- Veske, D., Bartos, I., Márka, Z., & Márka, S. 2021, ApJ, 922, 258
- Vink, J. S., Higgins, E. R., Sander, A. A. C., & Sabhahit, G. N. 2021, MNRAS, 504, 146
- Vink, J. S., Sabhahit, G. N., & Higgins, E. R. 2024, A&A, 688, L10
  Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2022, in Handbook of Gravitational Wave Astronomy, ed. C. Bambi, S. Katsanevas, & K. D. Kokkotas, 45
- Willcox, R., MacLeod, M., Mandel, I., & Hirai, R. 2023, ApJ, 958, 138
  Ziosi, B. M., Mapelli, M., Branchesi, M., & Tormen, G. 2014, MNRAS, 441, 3703