

# Flexible and Accurate Evaluation of Gravitational-wave Malmquist Bias with Machine Learning

Colm Talbot<sup>1,2,3</sup>  and Eric Thrane<sup>4,5</sup> 

<sup>1</sup> LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; [colm.talbot@ligo.org](mailto:colm.talbot@ligo.org)

<sup>2</sup> Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>3</sup> LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

<sup>4</sup> School of Physics and Astronomy, Monash University, VIC 3800, Australia

<sup>5</sup> OzGrav: The ARC Centre of Excellence for Gravitational-Wave Discovery, Clayton, VIC 3800, Australia

Received 2021 June 29; revised 2022 January 2; accepted 2022 January 14; published 2022 March 7

## Abstract

Many astronomical surveys are limited by the brightness of the sources, and gravitational-wave searches are no exception. The detectability of gravitational waves from merging binaries is affected by the mass and spin of the constituent compact objects. To perform unbiased inference on the distribution of compact binaries, it is necessary to account for this selection effect, which is known as Malmquist bias. Since systematic error from selection effects grows with the number of events, it will be increasingly important over the coming years to accurately estimate the observational selection function for gravitational-wave astronomy. We employ density estimation methods to accurately and efficiently compute the compact binary coalescence selection function. We introduce a simple pre-processing method, which significantly reduces the complexity of the required machine-learning models. We demonstrate that our method has smaller statistical errors at comparable computational cost than the method currently most widely used allowing us to probe narrower distributions of spin magnitudes. The currently used method leaves 10%–50% of the interesting black hole spin models inaccessible; our new method can probe >99% of the models and has a lower uncertainty for >80% of the models.

*Unified Astronomy Thesaurus concepts:* [Gravitational waves \(678\)](#); [Gravitational wave astronomy \(675\)](#); [Gravitational wave sources \(677\)](#)

## 1. Motivation

The phenomenon of Malmquist selection biases in brightness-limited astronomical surveys has been known for nearly a century (Malmquist 1922, 1925), and methods for mitigating this bias have remained an active area of research in astronomy (e.g., Freudling et al. 1995; Loredo 2004; Foreman-Mackey et al. 2014; Farr et al. 2015; March et al. 2018; Connor 2019; Mandel et al. 2019). Gravitational-wave searches are affected by Malmquist bias as the parameters of the merging binaries determine the luminosity of the source. More massive binaries produce larger gravitational-wave strain, all else equal, and so they can be observed at greater distances, at least until the mass becomes so large that the signal begins to shift out of the observing band. Additionally, binaries with black holes that are spinning along axes aligned with the orbital angular momentum are visible to farther distances (Campanelli et al. 2006; Ng et al. 2018).

Over the past several years, there has been increasing interest in population studies, which seek to measure the distribution of astrophysical parameters such as the mass, spin, and distance of merging compact objects using events observed with Advanced LIGO/Virgo (Aasi et al. 2015; Acernese et al. 2015; see, e.g., Abbott et al. 2021a; Roulet et al. 2021; The LIGO Scientific Collaboration et al. 2021b and references therein). To perform unbiased inference on the distribution of astrophysical parameters, it is necessary to account for selection biases when performing population inference; see, e.g., Loredo (2004), Farr et al. (2015), Mandel et al. (2019), Thrane & Talbot (2019),

and Vitale et al. (2020). The standard method employed in gravitational-wave searches requires computing the total sensitivity of search pipelines for a given population model. Evaluating the sensitivity for different population parameters involves integrating the time-dependent sensitivity to all binary parameters over the whole observing time.

Astrophysical population inference is typically performed using Markov Chain Monte Carlo (MCMC; Metropolis et al. 1953; Hastings 1970) or nested sampling algorithms (Skilling 2004), which can require  $O(10^3\text{--}10^7)$  likelihood evaluations for a well converged run, and the selection function must be evaluated on the fly at every iteration. As the binary black hole catalog grows, this integral must be evaluated with increasing precision (Farr 2019), and correspondingly increased computational cost. Additionally, as the catalog grows, so does our resolving power, meaning that subdominant effects, e.g., the effect of black hole spin on the sensitivity, must be considered. Guaranteeing sufficient precision is especially challenging for narrow population models as the Monte Carlo integrals currently performed are poorly suited to probing these distributions. In this work, we demonstrate that by performing a density estimation step on the set of found injections, we can dramatically increase the efficiency of these calculations, enabling us to probe narrow population models.

The rest of this paper is structured as follows. In the next section, we define some relevant quantities and outline methods for accounting for Malmquist bias in gravitational-wave searches. We then briefly summarize a few preliminaries for gravitational-wave population inference in Section 3. Following this, in Section 4, we describe the problem of density estimation and discuss various commonly used methods. In Section 5 we estimate the gravitational-wave transient selection



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

function as a function of binary parameters using a Gaussian mixture model. After this, we apply our methods to the binary black hole systems identified in The LIGO Scientific Collaboration et al. (2021a). Some closing thoughts are then provided. The notebook that performed the analysis presented here can be found at Talbot (2022).

## 2. Sensitivity Estimation

Most gravitational-wave population analyses impose a detection threshold on the analyzed triggers to avoid contamination from terrestrial noise sources, for example, demanding that the false-alarm rate of a trigger is less than once per year. Applying this threshold leads to a selection bias in the observed sample. We quantify this by considering the probability that a signal with parameters  $\theta$  (e.g., binary mass and black hole spin) would surpass our threshold  $\rho_{\text{th}}$ <sup>6</sup>

$$p_{\text{det}}(\theta) = \int_{\rho > \rho_{\text{th}}} d\mathbf{d} p(\mathbf{d}|\theta). \quad (1)$$

The integral is over all observed data, and  $p_{\text{det}}$  is the fraction of the data that surpasses the threshold under the assumption that a signal with parameters  $\theta$  is present. For population analyses, we require the fraction of all sources that are detectable, for a given population model, characterized by parameters,  $\Lambda$ ,

$$P_{\text{det}}(\Lambda) = \int d\theta p(\theta|\Lambda) p_{\text{det}}(\theta), \quad (2)$$

where  $p(\theta|\Lambda)$  is a conditional prior for  $\theta$  given population (hyper-)parameters  $\Lambda$ , e.g., the shape of the black hole mass distribution. For a detailed derivation of these quantities see, e.g., Finn & Chernoff (1993), Messenger & Veitch (2013), Farr et al. (2015), Tiwari (2018), Thrane & Talbot (2019), and Mandel et al. (2019).

We emphasize that all population analyses that apply a threshold necessarily have a corresponding selection bias that must be accounted for, including analyses that explicitly model contamination of the sample from terrestrial sources (Gaebel et al. 2019; Galauadage et al. 2020; Roulet et al. 2020). However, see Smith et al. (2020) for a method that avoids thresholds entirely.

The integral over  $\mathbf{d}$  in Equation (1) requires that we understand the sensitivity of gravitational-wave searches throughout the observing history. In practice, there are currently two widely used methods to compute this integral: inject simulated signals into the data and see how many of them are recovered by the search pipelines; or use a semi-analytic approximation based on the power spectral density of the interferometers, e.g., Finn & Chernoff (1993).

The former method gives the most faithful representation of the search sensitivity. However, the latter has several computational advantages. Because of the large parameter space that must be covered, the injection and recovery procedure gives us only the parameter values of the found/missed signals, whereas the semi-analytic approach can efficiently generate a numerical value for  $p_{\text{det}}$  marginalized over specific nuisance parameters. Thus, the semi-analytic approach can also be performed much more computationally cheaply due to the cost of performing and recovering injections. Previous methods to improve the reliability of semi-analytic estimates include calibration of semi-analytic

estimates with the output of injection campaigns (Wysocki & O’Shaughnessy 2018) and phenomenological fits, e.g., Fishbach et al. (2018) and Veske et al. (2021).

There have been several recent methods to leverage supervised machine-learning methods to estimate  $p_{\text{det}}$  (Gerosa et al. 2020; Wong et al. 2020b). Gerosa et al. (2020) use a neural-network classifier to give a binary outcome of “detectable” or “not detectable” for a given set of binary parameters; this method requires retraining when the threshold is changed. In Wong et al. (2020b), the authors train a neural-network regressor to estimate the signal-to-noise ratio (a commonly used detection threshold), allowing for the threshold to be changed trivially. However, both of these methods require specifying all of the binary parameters in order to evaluate  $p_{\text{det}}$ . In this work, we use density estimation on the set of found injections to provide a continuous, generative model for  $p_{\text{det}}$  in arbitrary subsets of the binary parameters.

The integral over  $\theta$  in Equation (2) marginalizes over all of the parameters describing the source—15 parameters to completely characterize a quasi-circular binary black hole merger—in addition to any parameters describing the state of the instruments. In practice, many of the parameters are not modeled in current population analyses; the most complex models considered currently fit for the distribution of seven of these parameters, the two component masses, spin magnitudes, spin-tilt angles, and redshift, requiring the evaluation of a seven-dimensional integral within each likelihood evaluation. The other parameters are assumed to be well described by the prior distributions used during sampling. These are mostly geometric parameters describing the position and orientation of the binary, although it is possible that some of these parameters may deviate from isotropy. For example, we could search for deviations from isotropy over the sky position, e.g., Payne et al. (2020) and Stiskalek et al. (2021), or features in the distribution of the azimuthal spin parameters due be influenced by spin-orbit resonances (Schnittman 2004; Gerosa et al. 2018; Varma et al. 2022). This integral is, therefore, recast as a Monte Carlo integral over the set of found injections (Tiwari 2018; Farr 2019)

$$P_{\text{det}}(\Lambda) = \frac{1}{N_{\text{inj}}} \sum_{i=1}^{N_{\text{found}}} \frac{p(\theta_i|\Lambda)}{p(\theta_i|\Lambda_0)}. \quad (3)$$

Here,  $p(\theta|\Lambda_0)$  is the distribution of the injected signals that will depend on the specific analysis,  $N_{\text{inj}}$  is the total number of injected signals, and  $N_{\text{found}}$  is the number of injections surpassing the threshold. The sum in Equation (3) is over samples drawn from the distribution of found injections

$$\theta_i \sim p_{\text{det}}(\theta_i) \pi(\theta_i|\Lambda_0). \quad (4)$$

To ensure sufficient convergence of the Monte Carlo integral, we must have an effective sample size of at least four times the number of observed events (Farr 2019). This means that to fit tightly peaked distributions, we need a large number of samples for the distribution of found injections or a continuous representation of  $p_{\text{det}}$ . Performing more injections to increase the number of recovered injections quickly becomes computationally prohibitive. In this work, we resolve this issue by performing density estimation using the set of found injections. Using these density estimates, we can directly evaluate  $p_{\text{det}}$  and/or generate additional samples from the distribution of found injections.

<sup>6</sup> The specific choice of threshold is irrelevant so long as it is robustly defined.

### 3. Population Inference

#### 3.1. Models

For demonstrative purposes, we consider two simple population models from within the gravitational-wave literature. Following Abbott et al. (2021a) and Talbot & Thrane (2018), we model the binary black hole mass distribution as a power law in the larger mass,  $m_1$ , between the minimum and maximum mass along with a normally distributed component and a power law in the mass ratio,  $q = m_2/m_1$ ,

$$p(m_1|\alpha, m_{\min}, m_{\max}) = (1 - \lambda) \frac{(1 - \alpha)m_1^{-\alpha}}{m_{\max}^{1-\alpha} - m_{\min}^{1-\alpha}} + \frac{\lambda}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{(m_1 - \mu_m)^2}{2\sigma_m^2}\right) \quad (5)$$

$$p(q|m_1, \beta, m_{\min}) = \frac{(1 + \beta)q^{\beta}}{1 - \left(\frac{m_{\min}}{m_1}\right)^{1+\beta}}. \quad (6)$$

This is the POWER-LAW + PEAK model in Abbott et al. (2021a) without the low-mass smoothing. We assume that both component spins are drawn from the same distribution. We model the spin magnitudes as following a Beta distribution (Wysocki et al. 2019)

$$p(a_i|\alpha_{\chi}, \beta_{\chi}) = \frac{a_i^{\alpha_{\chi}-1}(1 - a_i)^{\beta_{\chi}-1}}{B(\alpha_{\chi}, \beta_{\chi})}. \quad (7)$$

We model the distribution of spin orientations as a combination of a truncated half-normal and a uniform distribution (Talbot & Thrane 2017)

$$p(\cos \theta_i|\sigma_s) = \frac{(1 - \xi)}{2} + \xi N(\sigma_s) \exp\left(-\frac{(\cos \theta_i - 1)^2}{2\sigma_s^2}\right). \quad (8)$$

The factor  $N$  ensures that the distribution is properly normalized. This is the DEFAULT model in Abbott et al. (2021a).

The reference distribution using the LIGO/Virgo/Kagra collaborations' most recent injection campaign (LIGO Scientific Collaboration et al. 2021a) is the product of these distributions with population hyper-parameters  $\alpha = 2.35$ ,  $m_{\min} = 2$ ,  $m_{\max} = 100$ ,  $\lambda = 0$ ,  $\beta = 1$ ,  $\alpha_{\chi} = 1$ ,  $\beta_{\chi} = 1$ , and  $\xi = 0$ . These define  $p(\theta|\Lambda_0)$  for our application.

#### 3.2. Likelihood

The standard likelihood used in population inference for gravitational-wave sources in the presence of selection biases is (e.g., Mandel et al. 2019; Thrane & Talbot 2019; Vitale et al. 2020),

$$\mathcal{L}(\{d_i\}|\Lambda) = \frac{1}{P_{\det}(\Lambda)^N} \prod_i^N \int d\theta_i \mathcal{L}(d_i|\theta_i) p(\theta_i|\Lambda). \quad (9)$$

Where the product over  $i$  runs over the  $N$  observed events with data  $d_i$ . The integral over  $\theta_i$  is typically performed by importance sampling from the single-event posterior distribution for  $p(\theta_i|d_i)$  as is done to calculate  $P_{\det}$ . We take the publicly available samples from the single-event posterior distributions from Abbott et al. (2019a, 2021b) and The LIGO Scientific Collaboration et al. (2021a). Since the likelihood explicitly depends on  $P_{\det}$ , calculating this quantity is the main target of

this work. This likelihood is then used to explore the posterior distribution for the population parameters given all of the observed data, e.g., in Section 6.3.

### 4. Density Estimation

Reconstructing a function or probability density from a finite set of samples from the distribution is a widespread problem in data analysis. For example, injection campaigns to determine the sensitivity of gravitational-wave detectors do not give us a continuous description of the sensitivity, but rather a discrete set of samples from the distribution of found injections.

Density estimation methods can be loosely divided into parametric and nonparametric methods.<sup>7</sup> Parametric density estimation involves fitting a parameterized phenomenological model to the data. An example of this is the method used to reconstruct the population distribution of binary black holes in this work. To estimate the gravitational-wave selection function, we will rely on nonparametric density estimation.

Many methods for nonparametric density estimation are commonly used; however, most traditional methods such as binning or kernel density estimation scale poorly as the dimensionality of the problem increases. More sophisticated density estimation techniques involving the optimization of many parameters, such as Gaussian mixture models or flow-based inference, have proved successful at approximating complex functions in large dimensional spaces; see, e.g., Powell et al. (2019), Gabbard et al. (2022), Green et al. (2020), Green & Gair (2021), and Wong et al. (2020a, 2020b, 2021) for applications in gravitational-wave inference. These models also provide natural ways to generate additional samples from the underlying densities at minimal cost and are therefore sometimes referred to as generative models.

In this work, we approximate  $p_{\det}(\theta)$  using a Gaussian mixture model. A Gaussian mixture model is an unsupervised density estimator that approximates the distribution as a set of multivariate Gaussian distributions each with a unique mean and covariance. The model assumes that the target distribution can be well modeled by a finite sum of multivariate Gaussian distributions

$$\mathcal{D}(\theta') = \sum_{k=1}^K \frac{1}{w_k} \mathcal{N}(\theta'; \mu_k, \Sigma_k). \quad (10)$$

Here  $K$  is the number of components in the mixture and can be manually tuned,  $w_k$  is the weight associated with the  $k$ th component, and  $\mu_k$  and  $\Sigma_k$  are the mean vector and covariance matrix for that component, respectively. The values of  $w_k$ ,  $\mu_k$ , and  $\Sigma_k$  are optimized using the expectation-maximization algorithm (Webb 1999) to maximize the value of

$$\langle \ln \mathcal{D} \rangle = \frac{1}{N} \sum_{i=1}^N \ln \mathcal{D}(\theta'_i) \quad (11)$$

over the training data.

The nonparametric methods discussed above are all examples of unsupervised learning techniques as they do not require estimates of the target density as inputs. There are also

<sup>7</sup> We note that the word “nonparametric” is something of a misnomer, as often these models involve large numbers of unphysical parameters to perform the fit. An alternative delineation is between models where the parameters are either physically motivated (parametric) or not physically motivated (nonparametric).

supervised nonparametric density estimation methods that require the target function to be evaluated at the training points. For example, Gaussian process regression and neural-network regression have also been applied widely in gravitational-wave data analysis; see, e.g., Graff et al. (2012), Veitch et al. (2015), Moore et al. (2016), Doctor et al. (2017), Lange et al. (2018), Landry & Essick (2019), Williams et al. (2020), Gerosa et al. (2020), and D’Emilio et al. (2021).

## 5. Sensitivity to Binaries

In this section, we develop methods to evaluate Equations (2) and (1) and generate new samples from the distribution of found injections. We begin by training a function to estimate  $p_{\text{det}}(\theta)p(\theta|\Lambda_0)$  using a set of  $\sim 8 \times 10^4$  found injections. We use the same sensitivity data products used in The LIGO Scientific Collaboration et al. (2021b). Specifically, we take the found injections from Advanced LIGO/Virgo’s third observing run with a threshold of false-alarm rate  $< 1 \text{ yr}^{-1}$  in any of the search pipelines employed by the LIGO/Virgo collaboration (LIGO Scientific Collaboration et al. 2021a; The LIGO Scientific Collaboration et al. 2021a, 2021b). See the relevant publications and data releases for more details.

### 5.1. Pre-processing

Gravitational-wave parameters are typically only defined over finite domains and many have significant support at the edges, e.g., spin magnitudes are contained in the unit interval and the majority of observed black holes are consistent with being non-spinning. However, the algorithms we use for density estimation work best over an infinite domain without sharp boundaries. Our aim is to transform the found injections such that the transformed samples are drawn from a unit multivariate normal distribution. Therefore, we begin by performing the following mapping to the found injections:

1. Transform the injections from the original distribution to the unit interval. We denote generic transformations as  $U$  and discuss specific suitable transformations below.
2. Map the samples from the unit interval to a unit normal distribution using the probit function  $\Phi^{-1}$  (Bliss 1934).

Mathematically the full transformation is

$$\theta' = \Phi^{-1}(U(\theta)) \quad (12)$$

and the Jacobian is

$$\mathcal{J}(\theta) = \frac{d\theta'}{d\theta} = \frac{dU/d\theta}{\mathcal{N}(\theta'; \mu = 0, \sigma = 1)} \quad (13)$$

where  $\mathcal{N}$  is the normal distribution. We consider the four following scaling methods.

*Naive.* The simplest mapping onto the unit interval is a simple shift and scale from the original domain to the unit interval

$$U(\theta) = \frac{\theta - \theta_{\min}}{\theta_{\max} - \theta_{\min}}. \quad (14)$$

The mapping is attractive as it can be trivially applied to any parameter and has been used in other applications, e.g., D’Emilio et al. (2021). The Jacobian of this transformation is a constant

$$\frac{dU}{d\theta} = \frac{1}{\theta_{\max} - \theta_{\min}}. \quad (15)$$

*CDF.* For some parameters, the original distribution may be more complex; however, we may know an analytic form for that distribution. In this case, we perform the mapping using the cumulative distribution function (CDF) of the injected population

$$U(\theta) = \int_{\theta_{\min}}^{\theta} d\bar{\theta} p(\bar{\theta}|\Lambda_0). \quad (16)$$

In this case the Jacobian is

$$\frac{dU}{d\theta} = p(\theta|\Lambda_0). \quad (17)$$

This transformation is appealing as the form of the Jacobian means that evaluating  $p_{\text{det}}$  is trivial; however, as we will see for parameters that strongly influence the detectability, this mapping does a poor job approximating a unit uniform distribution.

*Approximate.* In cases where  $p_{\text{det}}$  is a strong function of  $\theta$  with a known (or approximately known) functional form, we may choose to map onto the unit interval using an approximate expression for the observed cumulative distribution

$$U(\theta) = \frac{\int_{\theta_{\min}}^{\theta} d\theta' p(\theta'|\Lambda_0) \tilde{p}_{\text{det}}(\theta')}{\int_{\theta_{\min}}^{\theta_{\max}} d\theta' p(\theta'|\Lambda_0) \tilde{p}_{\text{det}}(\theta')}. \quad (18)$$

Here  $\tilde{p}_{\text{det}}(\theta)$  is our analytic approximation to the selection function and the Jacobian

$$\frac{dU}{d\theta} = \frac{p(\theta'|\Lambda_0) \tilde{p}_{\text{det}}(\theta')}{\int_{\theta_{\min}}^{\theta_{\max}} d\theta' p(\theta'|\Lambda_0) \tilde{p}_{\text{det}}(\theta')}. \quad (19)$$

In our case, we use  $\tilde{p}_{\text{det}}(\theta) \propto m^{2.35} q^2$  so we have  $p(\theta|\Lambda_0) \tilde{p}_{\text{det}}(\theta) \approx \text{const}$ . We note that the naive and approximate methods give the same transformation for primary mass and aligned spin components. There is a difference in the transformation of the mass ratio. While we only account for the dependence of  $p_{\text{det}}$  on the primary mass, we could employ a more complex expression, e.g., as discussed in Veske et al. (2021).

*Empirical.* Finally, we use an approximation to the one-dimensional target distribution. We construct a Gaussian kernel density estimate

$$\tilde{p}(\theta) \approx p(\theta|\Lambda_0) p_{\text{det}}(\theta) \quad (20)$$

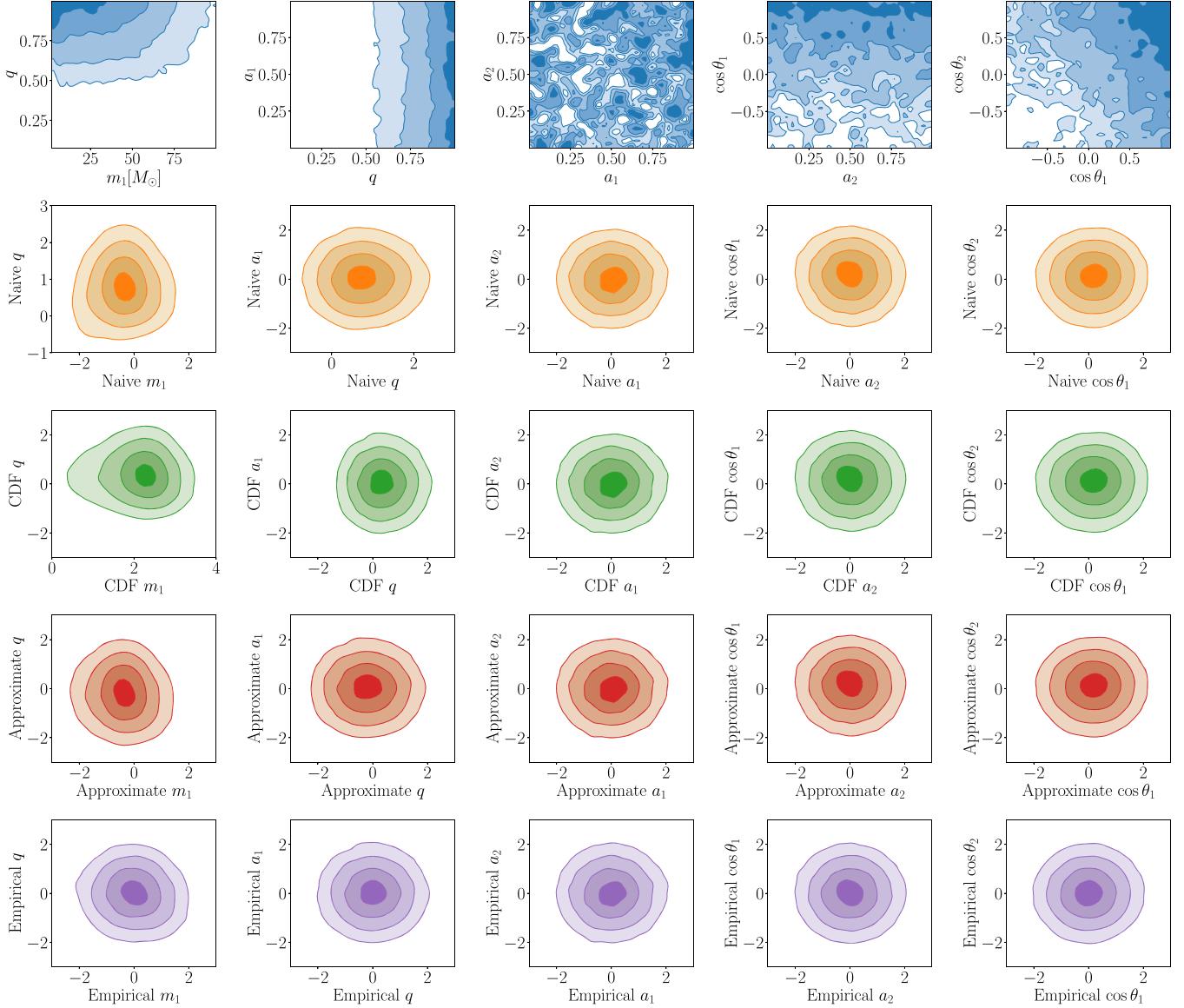
$$\equiv \frac{1}{N_{\text{found}}} \sum_{\theta_i}^{N_{\text{found}}} \mathcal{N}(\theta; \mu = \theta_i, \sigma = \sigma) \quad (21)$$

from the found injections. The standard deviation is chosen using Scott’s rule (Scott 1992). In order to account for parameters that have significant support at the edges, we apply a reflecting boundary condition to the estimate

$$\tilde{p}(\theta) = \tilde{p}(\theta) + \tilde{p}(2\theta_{\min} - \theta) + \tilde{p}(2\theta_{\max} - \theta). \quad (22)$$

Using this estimate, we then compute an empirical CDF

$$U(\theta) = \int_{\theta_{\min}}^{\theta_{\max}} d\theta' \tilde{p}(\theta'). \quad (23)$$



**Figure 1.** Two-dimensional density plots of the distribution of found injections before (top) and after (lower) applying the transformations described in Section 5.1. Our aim is to transform the data to approximate a unit multivariate normal distribution. In descending order, these transformations are naive, CDF, approximate, and empirical (see the main text for definitions). We note that the empirical scaling maps the data most closely onto a unit multivariate normal distribution.

The Jacobian for this transformation is

$$\frac{dU}{d\theta} = \tilde{p}(\theta) \quad (24)$$

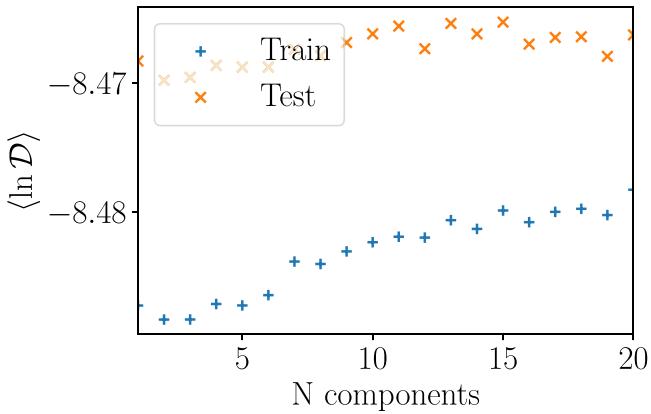
and can be trivially evaluated from our kernel density estimate.

In Figure 1 we show the set of found injections in the original ( $\theta$ ) space and each of the transformed spaces. In descending order, the rows are the original data, naive scaling, CDF scaling, approximate scaling, and empirical scaling, respectively. Each of the transformations has removed the railing against the boundaries in all of the parameters. However, there are visible features remaining, especially in the mass parameters. We note that the empirical scaling most closely transforms the data to an uncorrelated multivariate unit normal. We will use the empirical scaling going forward unless otherwise specified.

## 5.2. Density Estimation

Our aim is to take the regularized samples  $\{\theta'_i\}$  and estimate the density  $\mathcal{D}$  using a Gaussian mixture model. Training is performed by maximizing the mean natural log-density of the test samples as implemented in SCIKIT-LEARN (Pedregosa et al. 2011). Adding more components will improve the quality of the fit. However, by using too many components, we risk overfitting statistical fluctuations in the training set. To avoid this, we split the samples into a training (80%) and a test (20%) set. The fluctuations in the test set should be independent of those in the training set and therefore we will choose the number of components when the quality of the fit in the test set stops improving when adding more components.

In Figure 2, we show the average log-density over the training and test sample sets for the trained Gaussian mixture models with varying numbers of components. The offset between the two sets of points is simply due to random



**Figure 2.** The average log-likelihood over the test and training data of the trained Gaussian mixture model as a function of the number of Gaussian components. We note that the performance on the test data set flattens out after  $\sim 10$  components, while the performance on the training data continues to improve.

fluctuations in the train/test split. Other realizations can lead to a smaller mean log-density for the test data set. We note that the performance on the test data set flattens out after  $\sim 10$  components, while the performance on the training data continues to improve. We, therefore, use 10 components in the remainder of this work to avoid over-fitting. In the subsequent Sections, we use a 10-component Gaussian mixture model density estimate  $\mathcal{D}$  trained using all of the found injections  $\{\theta'_i\}$ .

### 5.3. Evaluation

After training the Gaussian mixture model, we can trivially generate new samples from the target distribution by drawing samples  $\theta'$  from the Gaussian mixture model and applying the inverse of Equation (12). Alternatively, since these are density estimates, we can also directly evaluate the estimated density

$$\mathcal{D}(\theta') = \frac{p_{\text{det}}(\theta)p(\theta|\Lambda_0)}{\mathcal{J}(\theta)}. \quad (25)$$

In practice, we want to evaluate the selection function

$$p_{\text{det}}(\theta) = \frac{\mathcal{J}(\theta)\mathcal{D}(\theta')}{p(\theta|\Lambda_0)}. \quad (26)$$

Here  $\mathcal{J}$  is the Jacobian from Equation (13) and  $p(\theta|\Lambda_0)$  is the original distribution of injections. We use Equation (26) as an alternate means of computing Equation (2) with an equivalent Monte Carlo integral over samples from the population distribution

$$P_{\text{det}}(\Lambda) = \langle p_{\text{det}}(\theta_i) \rangle_{\theta_i \sim p(\theta|\Lambda)}. \quad (27)$$

With the empirical mapping described in Section 5.1 we have

$$p_{\text{det}}(\theta) = \frac{\mathcal{D}(\theta')}{\mathcal{N}(\theta'; \mu = 0, \sigma = 1)} \frac{\tilde{p}(\theta)}{p(\theta|\Lambda_0)}. \quad (28)$$

This can be very efficiently evaluated as required.

We note that this method requires an efficient method of generating samples from the population distribution. This can be trivially performed using inverse-transform sampling if the population model has an analytically invertible CDF, or is a sum of such distributions. For other population models (for example, those using the low-mass smoothing introduced in

Talbot & Thrane 2018 or the redshift distribution introduced in Fishbach et al. 2018 and used in many population analyses, e.g., Abbott et al. 2019b, 2021a), another method is required. The simplest alternative approach is numerically estimating the inverse CDF; this can be implemented for all models at an increased, and perhaps prohibitive, computational cost. In Wong et al. (2020a), the authors train a deep flow-based generative network capable of very efficiently generating samples from the mass model in Talbot & Thrane (2018). Additionally, one could relegate the generation of samples from the population model to an offline pre-processing step by training a deep neural network to estimate  $P_{\text{det}}$  directly.

In practice, we find that this method requires far fewer samples in the Monte Carlo integral than when resampling the found injections, 5000–10,000 samples from the population model versus  $\sim 80,000$  found injections with the same number of effective samples for each method. The number of effective samples is defined slightly differently for the two Monte Carlo methods considered here. For Equation (27), we adopt the usual definition (Elvira et al. 2018)

$$N_{\text{eff}} = \frac{\left( \sum_{i=1}^N p_{\text{det}}(\theta_i) \right)^2}{\sum_{i=1}^N p_{\text{det}}(\theta_i)^2}. \quad (29)$$

However, for Equation (3), a correction is required to account for the initial injections with  $p_{\text{det}} = 0$  (Farr 2019). Following Farr (2019), for both of these methods, we only allow for samples with  $N_{\text{eff}} > 4N_{\text{events}}$  and marginalize over the statistical uncertainty in  $P_{\text{det}}$  in the likelihood.

## 6. Results

To demonstrate the efficacy of our new method, we consider the accuracy and precision of three different methods to compute the population selection function. We evaluate  $P_{\text{det}}$  three times for each set of samples:

1. Using Equation (3) with the original found injection set;
2. Using Equation (3) with samples generated using our Gaussian mixture model; and
3. Using Equation (27) with 10,000 samples from the population model.

We compare our estimators using 5000 samples drawn from the prior distribution for our population parameters specified in Table 1. These are the parameters describing the models presented in Section 3.1. We note that the prior is specified on the mean and variance of the Beta distribution

$$\mu_\chi = \frac{\alpha_\chi}{\alpha_\chi + \beta_\chi} \quad (30)$$

$$\sigma_\chi^2 = \frac{\alpha_\chi \beta_\chi}{(\alpha_\chi + \beta_\chi)^2 (\alpha_\chi + \beta_\chi + 1)} \quad (31)$$

$$\alpha_\chi = \frac{\mu_\chi^2 (1 - \mu_\chi) - \mu_\chi \sigma_\chi^2}{\sigma_\chi^2} \quad (32)$$

$$\beta_\chi = \frac{\mu_\chi (1 - \mu_\chi)^2 - (1 - \mu_\chi) \sigma_\chi^2}{\sigma_\chi^2}. \quad (33)$$

We highlight one specific limiting case that will most clearly demonstrate the differences between the methods. We note that

**Table 1**  
Prior Distribution for Population Parameters Used in the Analysis  
Presented Here

Parameter	Distribution
$\alpha$	$\mathcal{U}(-2, 8)$
$\beta$	$\mathcal{U}(-2, 8)$
$m_{\min}$	$\mathcal{U}(2, 10)$
$m_{\max}$	$\mathcal{U}(50, 100)$
$\lambda_m$	$\mathcal{U}(0, 1)$
$\mu_m$	$\mathcal{U}(20, 50)$
$\sigma_m$	$\mathcal{U}(0, 10)$
$\mu_{\chi}$	$\mathcal{U}(0, 1)$
$\sigma_{\chi}^2$	$\mathcal{U}(0, 0.25)$
$\xi$	$\mathcal{U}(0, 1)$
$\sigma_s$	$\mathcal{U}(0, 4)$

**Note.**  $\mathcal{U}(a, b)$  indicates a uniform distribution in  $[a, b]$ . We note that the Beta distribution is parameterized in terms of the mean ( $\mu_{\chi}$ ) and variance ( $\sigma_{\chi}^2$ ). Additional cuts are imposed such that  $\alpha_{\chi}, \beta_{\chi} > 0$ .

$\alpha_{\chi} \rightarrow 0$  indicates a spin distribution that peaks very sharply at zero black hole spin. If  $\alpha_{\chi} < 1$ , the distribution is singular at  $a = 0$ ; due to Monte Carlo convergence issues, these singular configurations have not been used in many previous analyses (e.g., Abbott et al. 2019b, 2021a; The LIGO Scientific Collaboration et al. 2021b).

### 6.1. Computing $P_{\text{det}}$

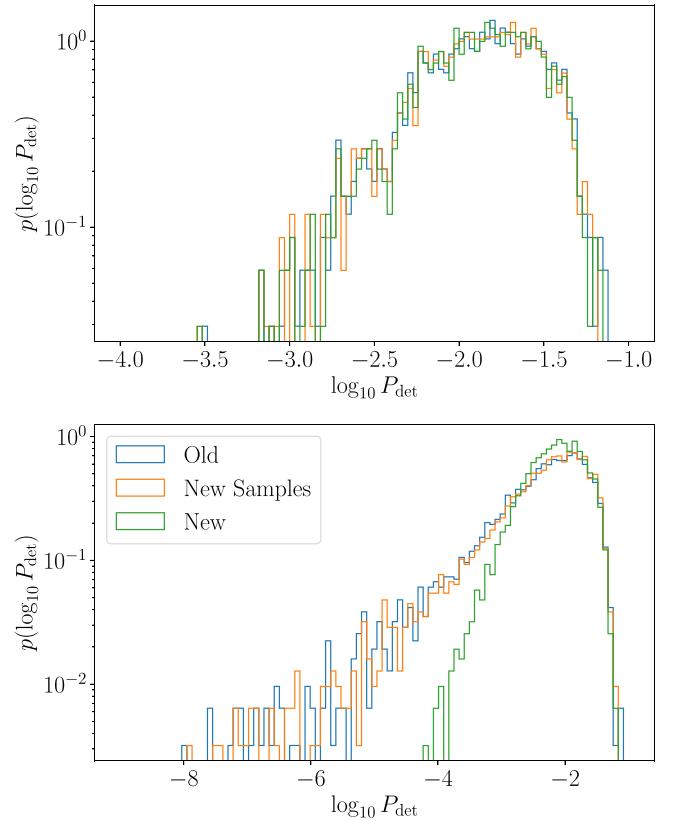
While the mean log-density is a suitable metric for training our estimators, we can perform a stronger test by considering how well the population-averaged  $P_{\text{det}}(\Lambda)$  compares across a range of  $\Lambda$  values using our Gaussian mixture model.

In Figure 3, we show the distribution  $P_{\text{det}}$  for these three methods (in order blue, orange, and green). In the top and bottom panels, we show samples for non-singular ( $\alpha_{\chi}, \beta_{\chi} > 1$ ) and singular spin magnitude distributions, respectively. We note that all of the methods agree well for the non-singular distributions. For singular spin distributions, both methods that resample a fiducial set of sample (blue and orange) methods differ significantly from our new method. These distributions are very sharply peaked and hence are the natural failure mode of importance sampling. The fact that the blue and orange results agree closely demonstrates that this effect is due to the difference in the construction of the Monte Carlo integral and not due to inaccuracy in the Gaussian mixture model estimate.

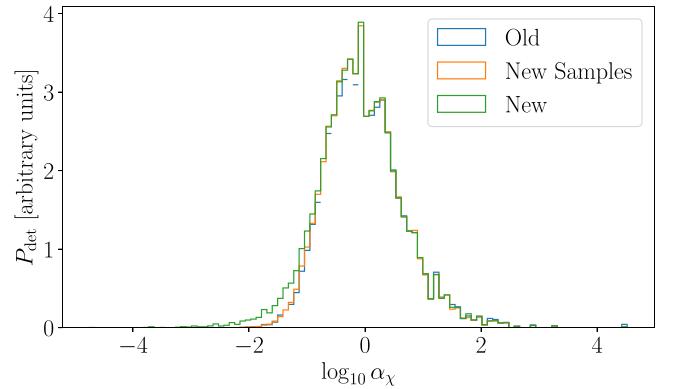
In Figure 4, we show the population-averaged sensitivity marginalized over all population parameters except for the logarithm of the Beta distribution  $\alpha$  parameter. The color scheme is the same as in Figure 3. The fiducial sample re-weighting methods underestimate the sensitivity to singular spin distributions compared to methods that directly evaluate  $P_{\text{det}}$ . We find that all of the methods agree well for the parameters describing the black hole mass distribution.

### 6.2. Monte Carlo Convergence

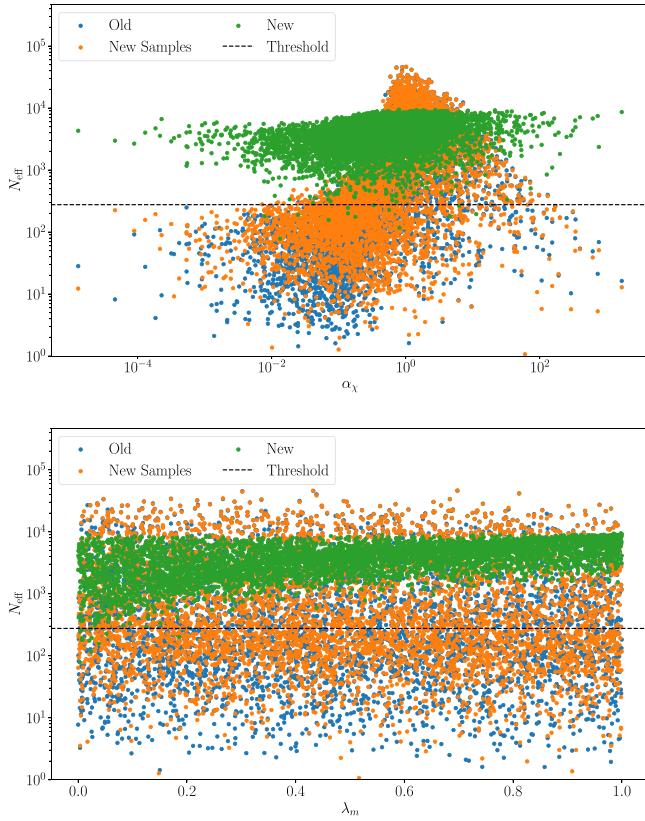
As described in Farr (2019), in order to have a reliable estimate of  $P_{\text{det}}$ , we need a sufficient number of effective samples in our Monte Carlo integral. When re-weighting a fixed set of recovered injections, this amounts to certain parts of the parameter space being inaccessible to our analyses. By



**Figure 3.** The distribution of the logarithm of the population-averaged sensitivity  $P_{\text{det}}$  over the prior used for our training data (see Table 1). In the top panel, we show samples for non-singular spin magnitude distributions ( $\alpha_{\chi}, \beta_{\chi} > 1$ ). In the bottom panel, we use samples for singular spin distributions. In blue and orange, we calculate  $P_{\text{det}}$  using Equation (3) as in previous analyses using the recovered simulated injections from the LIGO/Virgo collaboration (blue) and samples drawn from our Gaussian mixture model fit to  $P_{\text{det}}$  (orange). In green, we calculate  $P_{\text{det}}$  using Equation (27) using 10,000 samples from the population model. We note that the methods agree well for the non-singular distributions; however, the old method breaks down for singular spin distributions leading to the difference in calculated  $P_{\text{det}}$ .



**Figure 4.** The population-averaged sensitivity marginalized over all population parameters except the logarithm of the Beta distribution  $\alpha$  parameter for spin magnitudes. In blue and orange, we calculate  $P_{\text{det}}$  using Equation (3) as in previous analyses using the recovered simulated injections from the LIGO/Virgo collaboration (blue) and samples drawn from our Gaussian mixture model fit to  $P_{\text{det}}$  (orange). In green, we calculate  $P_{\text{det}}$  using Equation (27) using 10,000 samples from the population model. The computed values of  $P_{\text{det}}$  agree for non-singular spin distributions  $\log_{10} \alpha_{\chi} > 0$  but the sample re-weighting method underestimates the sensitivity for singular spin configurations.



**Figure 5.** Number of effective samples in the calculation of  $P_{\text{det}}$  as a function of two population parameters describing the distribution of merging black holes. (Top) One of the parameters describing the black hole spin distribution  $\alpha_{\chi}$ . (Bottom) The fraction of primary black holes whose mass falls in the Gaussian component of a two-component power-law and Gaussian mixture model. In blue, we show  $N_{\text{eff}}$  calculated by re-weighting injections directly. In orange, we draw a new fixed set of samples using our Gaussian mixture model density estimate and re-weight them. In green, we draw samples from the population model and directly evaluate  $p_{\text{det}}(\theta)$  using our Gaussian mixture model density estimate. The dashed black line shows the threshold for sufficient convergence of the Monte Carlo integral. We note that the old method is systematically biased away from small  $\alpha_{\chi}$  (distributions that assign small spins to most black holes) and is unable to probe the region with  $\alpha_{\chi} \lesssim 0.5$ . In total,  $\approx 40\%$  of the prior volume is inaccessible with the old method, compared to  $< 1\%$  with our new method.

contrast, we can draw arbitrary numbers of samples from our density estimates of  $p_{\text{det}}$  in order to achieve sufficient convergence.

To demonstrate this, we compute the number of effective samples for each of the population samples used in the rest of this section using Equations (2) and (27). For the former, we use the  $\approx 80,000$  found injections during O3 (LIGO Scientific Collaboration et al. 2021a). Throughout, we use  $N_{\text{events}} = 69$ . (binary black hole events confidently identified in O3.) For the latter, we draw  $10^4$  samples from the population model. In Figure 5 we show the number of effective samples for the old fiducial sample re-weighting method (blue with real injections and orange with samples drawn from our density estimate) and our new method (green) for two population parameters. In the top and bottom panels, we show the parameters that demonstrate the most obvious trend for the old and new methods, respectively ( $\alpha_{\chi}$  and  $\lambda_m$  — the fraction of primary masses in the Gaussian component). Once again, we note that the blue and orange results agree closely, indicating the robustness of the Gaussian mixture model fit.

The spread in  $N_{\text{eff}}$  for the old method is over four orders of magnitude, while for the new method, it is only two orders of magnitude. This means that if using a fixed number of samples at each iteration, that fixed number can be smaller for the new method leading to increased computational efficiency. As expected, we see that the efficiency of the fiducial sample re-weighting drops significantly for  $\alpha_{\chi} < 1$ . In contrast, our new method only weakly depends on the parameters of the spin magnitude distribution. We find that when re-weighting the found injections, we reject  $\approx 10\%$  of the non-singular samples and  $\approx 50\%$  of the singular samples compared to  $< 1\%$  of samples using our new method. The new method has more effective samples (smaller uncertainty in  $> 80\%$  of the space).

### 6.3. Population Inference

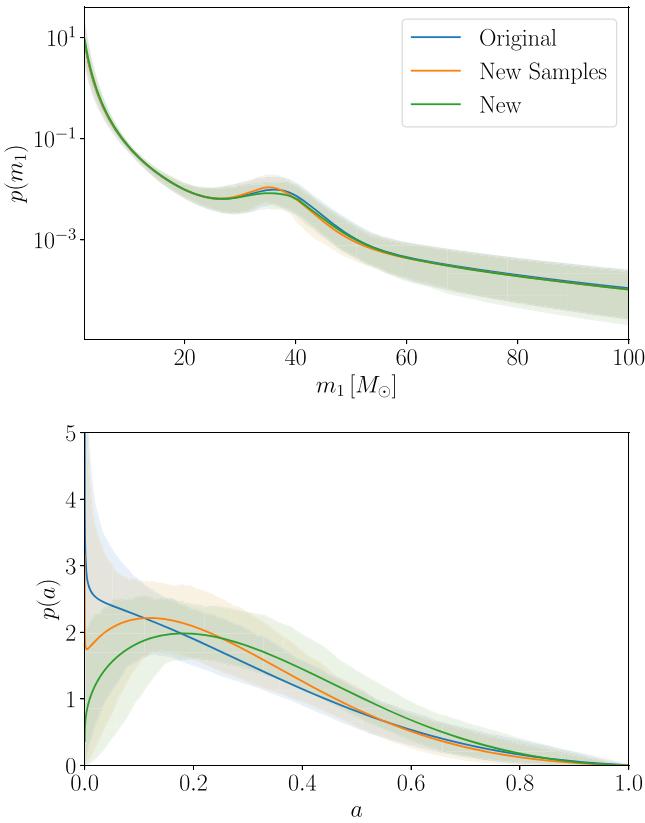
We now consider the impact of our different  $P_{\text{det}}$  evaluation methods on population inference. We analyze the 69 binary black hole mergers with false-alarm rate  $< 1 \text{ yr}^{-1}$  identified in GWTC-3 (The LIGO Scientific Collaboration et al. 2021a). We perform population inference three times, once with each of our  $P_{\text{det}}$  estimation methods using the BILBY-MCMC sampler (Ashton & Talbot 2021).

In Figure 6, we show the inferred astrophysical population of binary black holes when using our two methods to evaluate the selection function  $P_{\text{det}}$ . In blue, we evaluate Equation (3) using the original simulated injections. In orange, we evaluate Equation (3) using samples from our Gaussian mixture model. In green, we evaluate Equation (27) using our Gaussian mixture model. The solid curves show the mean inferred distribution, while the shaded regions indicate the 90% credible intervals. We find no significant difference between the inferred population using these methods, although we note a slight shift away from singular spin configurations when using our new method. This is exactly the region in which the old method underestimates the sensitivity, leading to an overestimated likelihood.

## 7. Discussion

Malmquist biases are ubiquitous in astronomical surveys, and methods of understanding and mitigating these biases are vital to performing astrophysical inference. Calculating and mitigating this bias is typically done by performing large simulations where synthetic signals are injected into the data and counting the number of recovered signals. While this method gives an optimal estimate of the performance of searches for signals in real data, they can be difficult to work with and extend to generic population models. In this work, we use these recovered signals to train a density estimate that can be reused to more efficiently compute the sensitivity to arbitrary populations.

To improve the accuracy of this density estimate, we introduce a pre-processing step that improves the convergence. Using this density estimate, we tested three methods to compute the population-level selection function. We found that our new density estimation method matches the previous injection resampling method for population models where the Monte Carlo integrals are well converged. We further demonstrated that our method is able to probe sharply peaked black hole spin distributions far more precisely than the existing method. This method can be more computationally expensive, especially for complex population models; however, deep-learning surrogate models present a solution to this problem (Wong et al. 2020a).



**Figure 6.** The inferred distribution of primary mass (top) and cosine spin tilt (bottom) when analyzing the binary black hole mergers identified in GWTC-3 using three different methods for estimating the population-averaged sensitivity,  $P_{\text{det}}$ : (blue) using the original found injections, (orange) new samples generated using a Gaussian mixture model fit to the found injections, and (green) evaluating  $P_{\text{det}}$  using samples from the population models. The results in blue and green do not include the dependence of  $P_{\text{det}}$  on the spin distribution. We note that this leads to a slight change in the inferred distribution of spin orientations: when including spin dependence in the selection function, there is slightly more preference for isotropic spins, although this is well below statistical uncertainties. The solid curves show the posterior predictive distribution, and the shaded regions show the symmetric 90% credible region.

Using our method, it is trivial to compute the fraction of sources that are observed  $P_{\text{det}}$  by marginalizing over parameters other than those parameterized in the population model, e.g., evaluate  $P_{\text{det}}$  using the parameters that most directly affect the sensitivity (chirp mass, mass ratio, and effective aligned/precessing spin) and model the population in terms of parameters with the most intuitive physical meaning (component masses, spin magnitudes, and orientations). We leave a detailed analysis of the best combination of parameters to use for the density estimation to future work. Our results are consistent with the results presented in The LIGO Scientific Collaboration et al. (2021b); however, the uncertainty on the measured selection function is less in  $>80\%$  of the space when using our new method. As the catalog of observed compact binary coalescences grows, it will be vital to understand the systematic error in our estimation of the selection function.

Machine-learning methods for density estimation are rapidly gaining popularity in the gravitational-wave data analysis community, e.g., Powell et al. (2019), Gabbard et al. (2022), Green et al. (2020), Green & Gair (2021), Wong et al. (2020a, 2020b, 2021), and Cuoco et al. (2020). Most of these methods require the use of complex neural-network-based density estimators, which require tuning many more free parameters

and thus extremely large training data sets. The pre-processing method introduced here removes sharp spectral features, e.g., at prior boundaries, and thus enables high-precision estimation of the target distribution using Gaussian mixture models, rather than having to employ deep-learning density estimators. Combining this pre-processing in other density estimation problems may have a similarly simplifying effect.

One limitation of the current method is that the Gaussian mixture model employed in this work provides only a best-fit model and does not provide an indication of uncertainty in the fit over the parameter space. We leave the exploration of density estimation techniques that model this uncertainty, e.g., Bayesian Gaussian mixture models, to a future study.

We thank Maya Fishbach for producing mock injections used in an early version of this work. We thank Sylvia Biscoveanu, Tom Dent, Reed Essick, Jacob Golomb, Cody Messick, Richard O’Shaughnessy, Alan Weinstein, Daniel Wysocki, and Salvatore Vitale for useful comments and discussions. This work is supported through the Australian Research Council (ARC) Centre of Excellence CE170100004 and ARC Future Fellowship FT150100281. This is document LIGO-P2000505. This research has made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Center (<https://www.gw-openscience.org>), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. Computing was performed computing clusters at the California Institute of Technology (LIGO Laboratory) supported by National Science Foundation Grants PHY-0757058 and PHY-0823459 and Swinburne University of Technology (OzSTAR). This work used publicly available samples from LIGO Scientific Collaboration & Virgo Scientific Collaboration (2018, 2020a, 2020b) and LIGO Scientific Collaboration et al. (2021a, 2021b). A Jupyter notebook to fully reproduce the results presented here along with a number of additional diagnostic figures can be found on GitHub.<sup>8</sup> This work made use of Google Colaboratory. Software used in this work includes: NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), SCIKIT-LEARN (Pedregosa et al. 2011), MATPLOTLIB (Hunter 2007), PANDAS (pandas development team, T., 2020), CUPY (Okuta et al. 2017), BILBY (Ashton et al. 2019; Ashton & Talbot 2021), and GWPOPULATION (Talbot et al. 2019).

## ORCID iDs

Colm Talbot  <https://orcid.org/0000-0003-2053-5582>  
Eric Thrane  <https://orcid.org/0000-0002-4418-3895>

## References

Aasi, J., Abadie, J., Abbott, B. P., et al. 2015, *CQGra*, 32, 74001  
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019a, *PhRvX*, 9, 031040  
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019b, *ApJL*, 882, L24  
 Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, *ApJL*, 913, L7  
 Abbott, R., Abraham, T. D., Abraham, S., et al. 2021b, *PhRvX*, 11, 021053  
 Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *CQGra*, 32, 024001  
 Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, *ApJS*, 241, 27  
 Ashton, G., & Talbot, C. 2021, *MNRAS*, 507, 2037  
 Bliss, C. I. 1934, *Sci*, 79, 38  
 Campanelli, M., Lousto, C. O., & Zlochower, Y. 2006, *PhRvD*, 74, 041501  
 Connor, L. 2019, *MNRAS*, 487, 5753  
 Cuoco, E., Powell, J., Cavaglià, M., et al. 2020, *MLS&T*, 2, 011002  
 D’Emilio, V., Green, R., & Raymond, V. 2021, *MNRAS*, 508, 2090  
 Doctor, Z., Farr, B., Holz, D. E., & Pürrer, M. 2017, *PhRvD*, 96, 123011

<sup>8</sup> [https://github.com/ColmTalbot/gmm\\_sensitivity\\_estimation/blob/main/gmm\\_sensitivity\\_estimation.ipynb](https://github.com/ColmTalbot/gmm_sensitivity_estimation/blob/main/gmm_sensitivity_estimation.ipynb)

Elvira, V., Martino, L., & Robert, C. P. 2018, arXiv:1809.04129

Farr, W. M. 2019, *RNAAS*, **3**, 66

Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. 2015, *PhRvD*, **91**, 023005

Finn, L. S., & Chernoff, D. F. 1993, *PhRvD*, **47**, 2198

Fishbach, M., Holz, D. E., & Farr, W. M. 2018, *ApJL*, **863**, L41

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, *ApJ*, **795**, 64

Freudling, W., da Costa, L. N., Wegner, G., et al. 1995, *AJ*, **110**, 920

Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F., & Murray-Smith, R. 2022, *NatPh*, **18**, 112

Gaebel, S. M., Veitch, J., Dent, T., & Farr, W. M. 2019, *MNRAS*, **484**, 4008

Galaudage, S., Talbot, C., & Thrane, E. 2020, *PhRvD*, **102**, 083026

Gerosa, D., Berti, E., O’Shaughnessy, R., et al. 2018, *PhRvD*, **98**, 084036

Gerosa, D., Pratten, G., & Vecchio, A. 2020, *PhRvD*, **102**, 103020

Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. 2012, *MNRAS*, **421**, 169

Green, S. R., & Gair, J. 2021, *MLS&T*, **2**, 03LT01

Green, S. R., Simpson, C., & Gair, J. 2020, *PhRvD*, **102**, 104057

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357

Hastings, W. K. 1970, *Biometrika*, **57**, 97

Hunter, J. D. 2007, *CSE*, **9**, 90

Landry, P., & Essick, R. 2019, *PhRvD*, **99**, 084049

Lange, J., O’Shaughnessy, R., & Rizzo, M. 2018, arXiv:1805.10457

LIGO Scientific Collaboration, & Virgo Scientific Collaboration 2018, Parameter Estimation Sample Release for GWTC-1, <https://dcc.ligo.org/LIGO-P1800370/public>

LIGO Scientific Collaboration, & Virgo Scientific Collaboration 2020a, GWTC-2 Data Release: Parameter Estimation Samples and Skymaps, <https://dcc.ligo.org/LIGO-P2000223/public>

LIGO Scientific Collaboration, & Virgo Scientific Collaboration 2020b, Data Release for “Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog”, <https://dcc.ligo.org/LIGO-P2000434/public>

LIGO Scientific Collaboration, Virgo Collaboration, & KAGRA Collaboration 2021a, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run—O3 Search Sensitivity Estimates, v1.0, Zenodo, doi:10.5281/zenodo.5546676

LIGO Scientific Collaboration, Virgo Collaboration, & KAGRA Collaboration 2021b, GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run—Parameter Estimation Data Release, v1.0, Zenodo, doi:10.5281/zenodo.5546663

Loredo, T. J. 2004, in AIP Conf. Proc. 735, 24th Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (College Park, MD: AIP), 195

Malmquist, K. G. 1922, *McLuF*, **100**, 1

Malmquist, K. G. 1925, *McLuF*, **106**, 1

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

March, M. C., Wolf, R. C., Sako, m., D’Andrea, C., & Brout, D. 2018, arXiv:1804.02474

Messenger, C., & Veitch, J. 2013, *NJPh*, **15**, 053027

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *JChPh*, **21**, 1087

Moore, C. J., Berry, C. P. L., Chua, A. J. K., & Gair, J. R. 2016, *PhRvD*, **93**, 064001

Ng, K. K. Y., Vitale, S., Zimmerman, A., et al. 2018, *PhRvD*, **98**, 083007

Okuta, R., Unno, Y., Nishino, D., Hido, S., & Loomis, C. 2017, in Proc. Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conf. on Neural Information Processing Systems (NIPS) (Red Hook, NY: Curran Associates)

pandas development team, T. 2020, *pandas-dev/pandas*: Pandas, v1.3.5, Zenodo, doi:10.5281/zenodo.3509134

Payne, E., Banagiri, S., Lasky, P. D., & Thrane, E. 2020, *PhRvD*, **102**, 102004

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, **12**, 2825

Powell, J., Stevenson, S., Mandel, I., & TiÁo, P. 2019, *MNRAS*, **488**, 3810

Roulet, J., Chia, H. S., Olsen, S., et al. 2021, *PhRvD*, **104**, 083010

Roulet, J., Venumadhav, T., Zackay, B., Dai, L., & Zaldarriaga, M. 2020, *PhRvD*, **102**, 123022

Schnittman, J. D. 2004, *PhRvD*, **70**, 124020

Scott, D. W. 1992, *Multivariate Density Estimation Theory, Practice, and Visualization* (New York: Wiley)

Skilling, J. 2004, *Bayesian Anal.*, **1**, 833

Smith, R. J. E., Talbot, C., Hernandez Vivanco, F., & Thrane, E. 2020, *MNRAS*, **496**, 3281

Stiskalek, R., Veitch, J., & Messenger, C. 2021, *MNRAS*, **501**, 970

Talbot, C. 2022, *gmm\_sensitivity\_estimation.ipynb*. [https://github.com/ColinTalbot/gmm\\_sensitivity\\_estimation/blob/main/gmm\\_sensitivity\\_estimation.ipynb](https://github.com/ColinTalbot/gmm_sensitivity_estimation/blob/main/gmm_sensitivity_estimation.ipynb)

Talbot, C., Smith, R., Thrane, E., & Poole, G. B. 2019, *PhRvD*, **100**, 043030

Talbot, C., & Thrane, E. 2017, *PhRvD*, **96**, 023012

Talbot, C., & Thrane, E. 2018, *ApJ*, **856**, 173

The LIGO Scientific Collaboration, The Virgo Collaboration, & The KAGRA Collaboration 2021a, arXiv:2111.03606

The LIGO Scientific Collaboration, The Virgo Collaboration, & The KAGRA Scientific Collaboration 2021b, arXiv:2111.03634

Thrane, E., & Talbot, C. 2019, *PASA*, **36**, e010

Tiwari, V. 2018, *CQGra*, **35**, 145009

Varma, V., Biscoveanu, S., Isi, M., Farr, W. M., & Vitale, S. 2022, *PhRvL*, **128**, 031101

Veitch, J., Raymond, V., Farr, B., et al. 2015, *PhRvD*, **91**, 042003

Veske, D., Bartos, I., Márka, Z., & Márka, S. 2021, *ApJ*, **922**, 258

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261

Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2020, arXiv:2007.05579

Webb, A. 1999, *Statistical Pattern Recognition*, A Hodder Arnold Publication (London: Newnes)

Williams, D., Heng, I. S., Gair, J., Clark, J. A., & Khamesra, B. 2020, *PhRvD*, **101**, 063011

Wong, K. W. K., Breivik, K., Kremer, K., & Callister, T. 2021, *PhRvD*, **103**, 083021

Wong, K. W. K., Contardo, G., & Ho, S. 2020a, *PhRvD*, **101**, 123005

Wong, K. W. K., Ng, K. K. Y., & Berti, E. 2020b, arXiv:2007.10350

Wysocki, D., Lange, J., & O’Shaughnessy, R. 2019, *PhRvD*, **100**, 043012

Wysocki, D., & O’Shaughnessy, R. 2018, Calibrating Semi-analytic VT’s against Reweighted Injection VT’s, <https://dcc.ligo.org/LIGO-T1800427/public>