Searching for structure in the binary black hole spin distribution

Jacob Golomb[●]

LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, USA
and Department of Physics, California Institute of Technology, Pasadena, California 91125, USA

Colm Talbot[●]

LIGO Laboratory, Massachusetts Institute of Technology,
185 Albany Street, Cambridge, Massachusetts 02139, USA,
Department of Physics and Kavli Institute for Astrophysics and Space Research,
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA,
and Kavli Institute for Cosmological Physics, University of Chicago,
5640 South Ellis Avenue, Chicago, Illinois 60615, USA

The spins of black holes in merging binaries can reveal information related to the formation and evolution of these systems. Combining events to infer the astrophysical distribution of black hole spins allows us to determine the relative contribution from different formation scenarios to the population. Many previous works have modeled spin population distributions using low-dimensional models with statistical or astrophysical motivations. While these are valuable approaches when the observed population is small, they make strong assumptions about the shape of the underlying distribution and are highly susceptible to biases due to mismodeling. The results obtained with such parametric models are valid only if the allowed shape of the distribution is well motivated (i.e., for astrophysical reasons). Unless the allowed shape of the distribution is well motivated (i.e., for astrophysical reasons), results obtained with such models, thus, may exhibit systematic biases with respect to the true underlying astrophysical distribution, along with resulting uncertainties not being reflective of our true uncertainty in the astrophysical distribution. In this work, we relax these prior assumptions and model the spin distributions using a more data-driven approach, modeling these distributions with flexible cubic spline interpolants in order to allow for capturing structures that the parametric models cannot. We find that adding this flexibility to the model substantially increases the uncertainty in the inferred distributions but find a general trend for lower support at high spin magnitude and a spin tilt distribution consistent with isotropic orientations. We infer that 62%–87% of black holes have spin magnitudes less than $a = 0.5$ and 27%–50% (90% credible levels) of black holes exhibit negative χ_{eff}. Using the inferred χ_{eff} distribution, we place a conservative upper limit of 37% for the contribution of hierarchical mergers to the astrophysical binary black hole population. Additionally, we find that artifacts from unconverged Monte Carlo integrals in the likelihood can manifest as spurious peaks and structures in inferred distributions, mandating the use of a sufficient number of samples when using Monte Carlo integration for population inference.

I. INTRODUCTION

Gravitational waves offer a unique probe into the properties of merging black holes (BHs) and neutron stars. Since the first such detection in 2015, the LIGO-Virgo network [1–3] has reported the detection of ∼90 binary black hole (BBH) mergers, with each gravitational-wave (GW) signal encoding physical information about the BHs involved, such as their masses and angular momenta (spins) [4,5]. Extracting this information has has enabled the study of properties of BBH systems on both an individual and population-level basis. From an astrophysical perspective, combining GW detections to infer the mass, spin, and redshift distributions of BBH systems can help answer questions ranging from binary formation and stellar evolution [6,7] to the expansion rate of the Universe and possible deviations from general relativity [8,9].

The spin of the BHs in a BBH system offers insight into the history of the binary. For example, BH spins can help reveal whether the BHs in a BBH system formed directly from core collapse of a heavy star or from the previous merger of two lighter BHs [10–13]. Although the processes governing the angular momentum transport out of a stellar core during collapse are not well constrained, recent modeling work indicates that BHs resulting directly from core collapse supernovae should have negligible spin

magnitudes [14–16]. While processes such as tidal interactions and mass transfer can induce higher spins on BHs in binary systems, it is uncertain how appreciable the resulting spin-ups can be [17–20]. On the other hand, BHs formed from the merger of two nonspinning BHs are expected to form a final BH with a relatively high spin magnitude [13,21,22], motivating the possibility to use spin magnitude as a tracer of a BH's formation history.

The direction of the BH spin vectors also encode information related to the formation history of a BBH system. Models suggest that BBH systems formed from common evolution, in which the component BHs evolve together from a stellar binary in an isolated environment free from significant dynamical interactions, should have component spin vectors nearly aligned with the orbital angular momentum axis, with any tilt being efficiently brought into alignment by tidal interactions [23,24]. On the other hand, BBH systems formed from dynamical encounters are not expected to have any correlated spins, such that the BH spin vectors are isotropic with respect to the orbital angular momentum [11,25,26].

While only a couple of events in the third gravitational-wave transient catalog individually feature confidently high spin magnitudes or antialignment (i.e., a spin vector pointing opposite the angular momentum), hierarchically combining observations of GW events while folding in selection effects can reveal the degree to which these parts of spin parameter space contribute to the astrophysical distribution of BH spins. Previous work has used these inferred contributions to estimate the fraction of BBH systems in the local Universe which may have been formed hierarchically, dynamically, and by isolated evolution [6,10,11,27]. However, recent publications have disagreeing estimates for the contributions of antialigned and nonspinning BBHs to the astrophysical population.

In [6,7], the authors conclude that the BBH distribution must feature antialigned spins at $> 90\%$ credibility, in contrast to the conclusion drawn in [28] that such antialignment is not evident in the population. In addition, [29] finds evidence for a nonspinning subpopulation of BHs, a conclusion which was challenged by [30]. While technical differences exist between works, a major possible contribution to some of these differing conclusions is model misspecification (see, e.g., [31,32]), that is, different assumptions being imposed on the functional form of the spin distribution.

The Default model in [6,7] models the distribution of the magnitude of the BH spin vector and the tilt angle between the spin vector and the orbital angular momentum. They adopt a Beta distribution for the spin magnitude model [6,33]:

$$\pi(a_{1,2}|\alpha_\chi,\beta_\chi) = \text{Beta}(a_{1,2}|\alpha_\chi,\beta_\chi), \qquad (1)$$

where a_1 (a_2) is the magnitude of the spin vector of the primary (secondary) BH and α_χ and β_χ are population

hyperparameters determining the structure of the Beta distribution. The model for the distribution of tilt angles, θ, is motivated by two subpopulations: one preferentially aligned [$\cos(\theta) \approx 1$] and one isotropic [6,7,34]. The model is parametrized as

$$\pi(\cos\theta_{1,2}|\xi,\sigma_t) = \xi G_t(\cos\theta_1|\sigma_t)G_t(\cos\theta_2|\sigma_t) + \frac{1-\xi}{4},$$
(2)

where G_t is a truncated Gaussian centered at $\cos\theta = 1$ with standard deviation σ_t and bounded in $[-1, 1]$ and ξ is the relative mixing fraction between the subpopulations. The second term corresponds to the contribution from the uniform (isotropic) distribution.

This population model has been extended in other work to allow for other astrophysically motivated features to help draw conclusions related to the different formation scenarios present in the astrophysical distribution [28–30,35]. Adopting an astrophysically motivated, strongly parametric model necessarily limits the possible features resolvable in the inferred distribution to what the chosen function can model. Accordingly, in this work, we consider a strongly parametric model to be one that has a specific, prior-determined shape as provided by the parameterization (e.g., a normal distribution), which is then constrained by the data. When using such a distribution to draw astrophysical conclusions from the inferred population, this is a reasonable and intended consequence, as the model is chosen to encode prior beliefs on the parameters that should govern the astrophysical distribution; however, if additional features exist in the true astrophysical distribution and a strongly parametric model cannot account for them, such features can be missed and a biased result may be obtained.

Previous work has shown that substructures in the BH mass distribution can be captured by cubic splines acting as a perturbation on top of a simpler parametric model [6,36]. In Ref. [36], the authors consider an exponentiated spline perturbation modulating an underlying power law in the mass distribution. In this work, we model the spin magnitude and tilt distributions using exponentiated cubic splines modulating a flat distribution to obtain a more data-driven result for the inferred population of BH spins. In doing so, we limit the potential bias caused by mismodeling the spin distribution and allow for the possibility of capturing features not accessible with a strongly parametric model.

The remainder of the paper is organized as follows. In Sec. II, we detail the functional form and implementation of the cubic spline model. We provide the background of hierarchical Bayesian inference in Sec. III, as it applies to population inference with GW sources. In Sec. IV, we present the resulting spin distributions we obtain for various spline models adopted in this work. Finally, we use these results to draw conclusions related to the astrophysical distribution of BBH spins and provide a

relevant discussion in Sec. V. We additionally supply three appendixes; the first provides additional details about an efficient caching technique for the cubic spline model, the second explores the effect of uncertainty in our estimation of the selection function, and the third describes robustness of our results to different choices of prior distribution.

II. MODELS

Following the model for the black hole mass distribution considered in [36], we fit the distribution of spin magnitudes and cosine tilts using exponentiated cubic splines:

$$p(x) \propto e^{f(x)}. \quad (3)$$

A spline is a piecewise polynomial function defined by a set of node positions, the value of the function at those nodes, and boundary conditions at the end nodes. We use a cubic spline, as it is the lowest-order spline that enforces continuity of the function and its first derivative everywhere.

A. Node positions and amplitudes

In this work, we consider models with four, six, eight, and ten nodes spaced linearly in the domain of the parameter space. For the two distributions being modeled with splines, this gives 16 unique spin models (four amplitude node placement models × four tilt node placement models). Our choice for the prior on the amplitude of each node is a unit Gaussian distribution. Comparisons with other node amplitude prior choices are detailed in Appendix C.

In order to fully characterize a cubic spline, the first and second derivatives must be determined at each node. For all but the end points, these derivatives are specified by requiring continuity in the spline and its derivative. At the end points, there is no unique way to determine this, and a range of boundary conditions are commonly used. For our implementation, we want the prior distribution of the derivatives at the end points to match that of the internal nodes. This requires providing two additional free parameters at each end of the spline. In practice, we add two additional nodes outside each boundary, with amplitudes that are free to vary according to the prior. Throughout this work, the number of nodes in a model refers to the number of nodes within the domain (i.e., not including these outside nodes). The spacing between these nodes is the same as that between nodes within the domain.

B. Modeling spins with splines

In this work, we use the spline model detailed above to model the population of spin magnitudes a and tilt angles $\cos\theta$. Consistent with [6,7], we model these parameters as independent and identically distributed. The total spin population model is

$$\pi_{\text{spin}}(\eta|\Lambda_s) = p_a(a_1)p_a(a_2)p_t(\cos\theta_1)p_t(\cos\theta_2), \quad (4)$$

where Λ_s is the set of population hyperparameters controlling the spline node location and amplitudes and η is the set of single-event parameters. The functions p are determined from Eq. (3). The domain of the spin magnitude distribution extends over $a \in [0,1]$ and that of the spin tilt distribution covers $\cos\theta \in [-1,1]$.

III. METHODS

A. Hierarchical Bayesian inference

In order to constrain the spin magnitude and tilt distribution, we carry out hierarchical Bayesian inference in which we calculate the likelihood of the entire observed dataset given a set of population hyperparameters Λ while marginalizing over the uncertainty in the physical parameters of each event. After analytically marginalizing over the total merger rate R with a prior $\pi(R) \propto R^{-1}$, we express the likelihood of the hyperparameters Λ parametrizing the population is expressed as (e.g., [37])

$$\mathcal{L}(\{d\}|\Lambda) \propto p_{\text{det}}(\Lambda)^{-N} \prod_i^N \int \mathcal{L}(d_i|\eta_i)\pi(\eta_i|\Lambda)d\eta_i. \quad (5)$$

Here, $\mathcal{L}(d_i|\eta_i)$ is the likelihood of observing the data d from the ith event, given physical (i.e., single-event) parameters η_i. In this work, η_i consists of masses, spins, and redshift of the ith event. The quantity $p_{\text{det}}(\Lambda)$ encodes the sensitivity of the search algorithm that identified the signals and is described in more detail in Sec. III B.

Our population model $\pi(\eta|\Lambda)$ describes the astrophysical distribution of masses, redshifts, and spins. We model the primary mass distribution with the power law + peak model [38], the mass ratio ($q = \frac{m_2}{m_1}$) distribution with a power law, and the redshift distribution also with a power law, with source-frame comoving merger rate density $R(z) \propto (1+z)^3$ [6,7,39]. We choose to fix the redshift distribution, because we use our own injection set to estimate sensitivity, thresholding on signal-to-noise ratio (SNR) rather than false alarm rate (FAR) to determine "found" injections. Since this makes the threshold used to select real events (FAR < 1 yr) slightly different from that used to threshold sensitivity injections and the redshift distribution is particularly sensitive to the near-threshold events, we fix the redshift distribution in order to avoid biases (see Ref. [40] for an example of where a similar approximation was used). See Sec. III B for details on sensitivity injections. We list the hyperparameters Λ and their corresponding priors in Table I.

An initial choice that must be made when computing Eq. (5) is which events to include in the analysis. Typically, this is done by establishing some detection threshold on the SNR or FAR and including all events that pass this threshold. We choose to include the 59 events in the third

TABLE I. Priors for mass distribution used in hierarchical inference, consistent with those used in [6]. Priors on the spin distribution are described in Sec. II. Priors are uniform between the bounds listed in the third column.

Parameter	Description	Prior
α	Power law index for m_1	$(-4, 12)$
β	Power law index for q	$(-2, 7)$
$M_{\max}$	Maximum mass	$(60, 100)$
$M_{\min}$	Minimum mass	$(2, 7)$
λ	Fraction of sources in Gaussian peak	$(0, 1)$
M_{pp}	Location of Gaussian peak	$(20, 50)$
σ_{pp}	Standard deviation of Gaussian peak	$(1, 10)$
δ_m	Minimum mass turn-on length	$(0, 10)$

observing run (O3) of the LIGO-Virgo network which have a false alarm rate of less than 1 yr^{-1} and are included in the main BBH analysis of [6]. We limit ourselves to events in O3 for self-consistency, as the injections we perform to evaluate selection effects (see Sec. III B) use O3a and O3b detector sensitivities.

We compute Eq. (5) using the package GWPopulation [41], which constructs a Monte Carlo approximation of this integral by reweighting samples from the single-event posteriors into the population model. We use the nested sampling package DYNESTY [42] to obtain hyperparameter samples from the posterior distribution.

B. Selection effects

Since the sensitivity to an event is determined by the single-event parameters η, the observed population is biased toward events produced by the astrophysical population that are preferentially observable. To account for the bias arising from selection effects, we must compute the fraction of signals ($p_{\det}$) that will have a detection statistic ρ passing our detection threshold ρ_{th} by marginalizing over all possible signals and noise realizations n (e.g., [43,44]):

$$p_{\det}(\Lambda) = \int dn \int d\eta \, p(\eta|\Lambda) p(n) \Theta(\rho - \rho_{th}). \quad (6)$$

Here, Θ is the Heaviside step function. In practice, we use Monte Carlo importance sampling with respect to some simulated fiducial reference population Λ_0 to estimate Eq. (6). This method relies on injecting N_{inj} sources from this reference population into detector noise and determining which of these sources pass our detection threshold [45]. This is computed as

$$p_{\det}(\Lambda) = \frac{1}{N_{\text{inj}}} \sum_{\eta \sim \eta_{\text{found}}} \frac{\pi(\eta|\Lambda)}{\pi(\eta|\Lambda_0)}, \quad (7)$$

where η_{found} corresponds to the single-event parameters of the events from the injection set that pass detection threshold.

For our sensitivity injection set, we simulate $\mathcal{O}(5 \times 10^7)$ sources and inject them into Gaussian noise corresponding to O3 detector sensitivity specified by the representative power spectral densities in [4,46]. This results in $\sim 900\,000$ injections passing our detection threshold of network optimal SNR greater than 10, where the square of the network optimal SNR is defined as the quadrature sum of the SNRs in each detector. We choose this threshold to be a surrogate for the 1 yr^{-1} FAR threshold used for event selection. While this is not an exact mapping between the two detection statistics, the effects of spins on sensitivity is subdominant, so we expect that this approximation will not cause biases.

C. Uncertainties in the likelihood

Since we approximate Eqs. (5) and (7) using Monte Carlo summation, there exists a resulting statistical uncertainty associated with the use of finite samples to obtain estimates for the value of the log-likelihood [45,47–49]. For each sample of Λ, we compute this associated uncertainty in the log-likelihood. Considering the computed approximation of $\ln \mathcal{L}(\{d\}|\Lambda)$ to be a realization from a distribution that asymptotically tends to a Gaussian distribution, we approximate the covariance of this estimate using Eq. (15) in [50]. For uncertainty propagation, we compute the variance associated with each $\ln \mathcal{L}(d|\Lambda)$ from Eq. (10) in [50].

While we do not enforce any threshold directly on the variance $(\Delta \ln \mathcal{L})^2$, we retain this information for all points in the hyperposterior to investigate correlations between features in the population and uncertainty in the log-likelihood (see Appendix B).

D. Uncertainty in the evidence

The evidence, or marginal likelihood, associated with a particular model is expressed simply as the expectation value of the likelihood conditioned on the population prior:

$$\mathcal{Z} = \int d\Lambda \mathcal{L}(\{d\}|\Lambda)\pi(\Lambda). \quad (8)$$

Comparing this quantity for two different models allows one to compute a Bayes factor, which is commonly used as a discriminator between models based on their relative strength at describing the observed data.

Because Dynesty computes the evidence by iteratively summing over a finite number of weights, there exists a statistical uncertainty associated with the estimated evidence. Dynesty reports this uncertainty along with the computed evidence.

Since there is also an uncertainty in the quantity $\ln \mathcal{L}(\{d\}|\Lambda)$ used in computing the evidence, and the
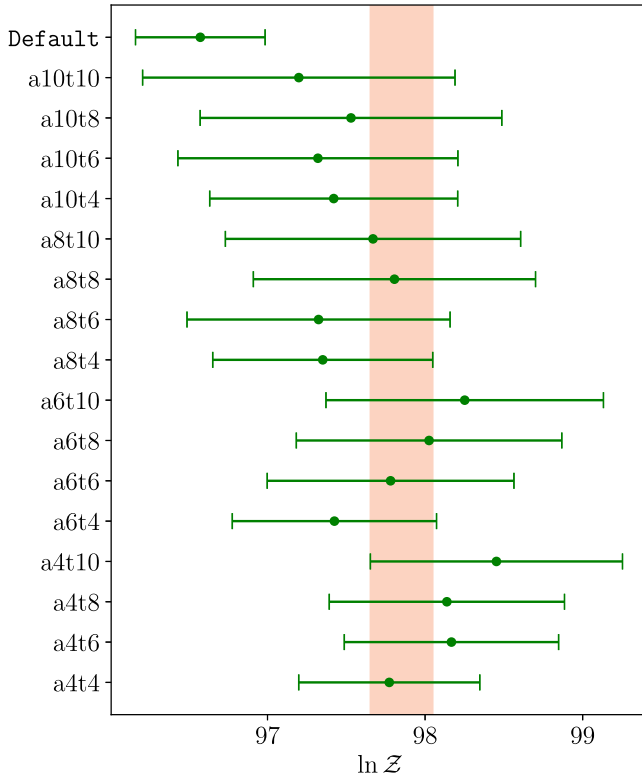
FIG. 1. Comparison of evidence obtained from the different spline node combinations considered, as well as from the Default model. Uncertainties are computed by adding the average covariance in the log-likelihood [see Eq. (15) in [50]] in quadrature with the uncertainty in the evidence as reported from Dynesty. The numbers after "a" and "t" are the number of nodes in the magnitude and tilt models, respectively. All evidences from the spline models are consistent with $\ln \mathcal{Z}$ in the red shaded region at 1σ.

evidence is the average of a set of $\ln \mathcal{L}$ values, we take the contribution of this uncertainty to the total evidence to be the average uncertainty in $\ln \mathcal{L}(\{d\}|\Lambda)$ over the draws from $\pi(\Lambda)$.

We take these two sources of uncertainty in the evidence to be independent and compute the total uncertainty in $\mathcal{Z}$ by propagation of errors. As a result, we obtain both the evidence and its uncertainty for each model of a fixed set of spline nodes. We note that all of the evidences for the spline models are consistent within their 1σ uncertainties, and the Default model has a natural log Bayes factor of ≈ -1.5 with respect to the overlapping region of uncertainties in the evidence for the spline models.

IV. RESULTS

In this section, we present the results from analyzing the population of BBH spin magnitude and tilts using spline models. We use the standard Gaussian prior on node amplitudes as described in Sec. II A, with nodes placed linearly within the domain of parameter space. In Fig. 1, we

show the evidence and their uncertainties for the 16 node combinations we consider in this work. All models give similar evidence, with no significant preferences considering their associated uncertainties. The red shaded region shows where all of the evidence estimates overlap within 1σ. This indicates that adding more nodes does not tend to provide a better fit to the distribution and also does not overfit it. We therefore cite the numbers in this section using the most flexible model, with ten nodes for both the magnitude and tilt distributions. Unless otherwise noted, the plots of the spin magnitude (tilt) distributions assume ten nodes in the tilt (magnitude) distribution.

As a general trend, we notice that the inferred 90% region of parameter space exhibits oscillating peaks at the location of the spline nodes. As shown in Fig. 13, these oscillations appear for uninformative data. With the observations of BH spins being weakly informative, we see this effect from the strong influence of the prior on the posterior distribution of the spline nodes.

A. The distribution of spin magnitudes

In Fig. 2, we show the inferred distribution of spin magnitudes for our four choices of node numbers in spin magnitude, assuming ten nodes in spin tilt. Although each model involves different positions and numbers of spline nodes, we note that the uncertainties (solid lines) and the average line (dot-dashed lines) in the distribution are comparable between models.

The 90% credible interval of the distribution is relatively broad, making it difficult to discern obvious trends in spin magnitude. However, we note the general pattern of a preference for smaller spin magnitudes in the population and less support for higher spin magnitudes. Considering the model with ten magnitude and ten tilt nodes, we infer that $77.1\%^{+10.4\%}_{-14.8\%}$ of spin magnitudes are below $a = 0.5$ and 50% of spin magnitudes are below $a = 0.25^{+0.16}_{-0.10}$ (all
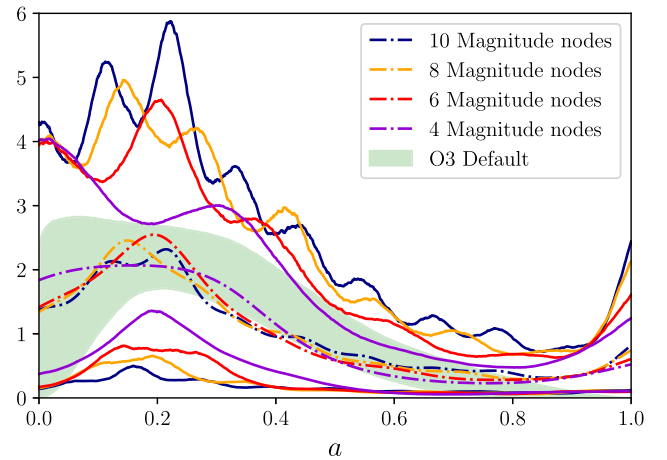

FIG. 2. Distribution of spin magnitudes, with different numbers of nodes corresponding to different colors. All use ten nodes in the tilt distribution.
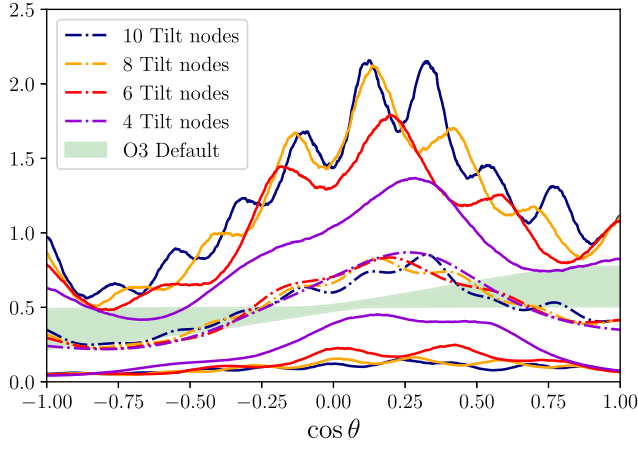
FIG. 3. Distribution of spin tilts, with different numbers of nodes corresponding to different colors. All use ten nodes in the spin magnitude distribution.

uncertainties in this work are reported at the 90% symmetric credible levels unless otherwise stated).

While the models using fewer spline nodes tend to place increased support around $a = 0.2$, the significance is substantially reduced as we add more spline nodes. When more spline nodes are added, the model becomes more flexible and more data are necessary to constrain the distribution. While this feature may be real, it is not confident enough to remain present as the flexibility of the model increases.

Comparing to the Default model (the green shaded region) used in [6], we observe substantially more uncertainty in the inferred spin magnitude distributions using our spline models. As a point of comparison, in Figs. 11 and 12 of Appendix C we show the distribution of spin magnitudes for different numbers of nodes, this time assuming four nodes in the tilt distribution. We find no significant differences from the distributions assuming ten nodes in the tilt distribution, which indicates that the number of nodes in the tilt distribution has a negligible effect on the inferred spin magnitude distribution.

B. The distribution of spin tilts

In Fig. 3, we show the inferred distribution of spin tilts. Similar to the case with spin magnitudes, the uncertainties in the distribution are wide, but the average distributions for the different node combinations agree. In general, the distribution is consistent with being flat and featureless, but there is a slight trend for an increase in support for $-0.25 < \cos\theta < 0.75$.

As demonstrated by comparing Fig. 12 with Fig. 3, the inferred distribution of spin tilts is very similar when we model the spin magnitude distribution with four nodes. We confirm this invariance for all sets of magnitude nodes tested, suggesting that the number of spin magnitude nodes does not meaningfully affect the recovered spin tilt distribution.
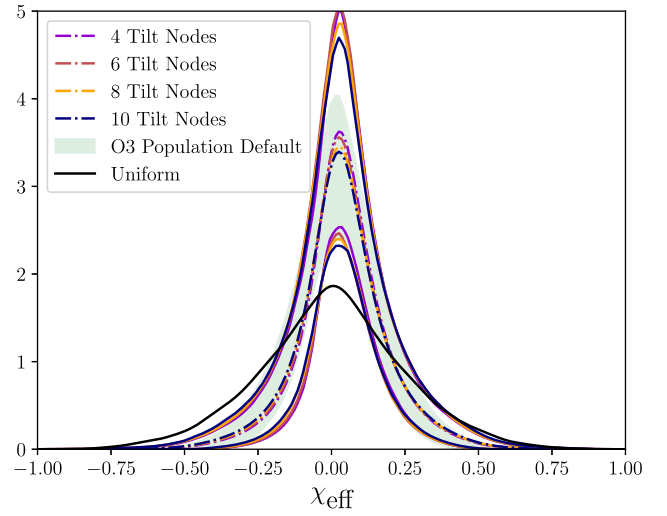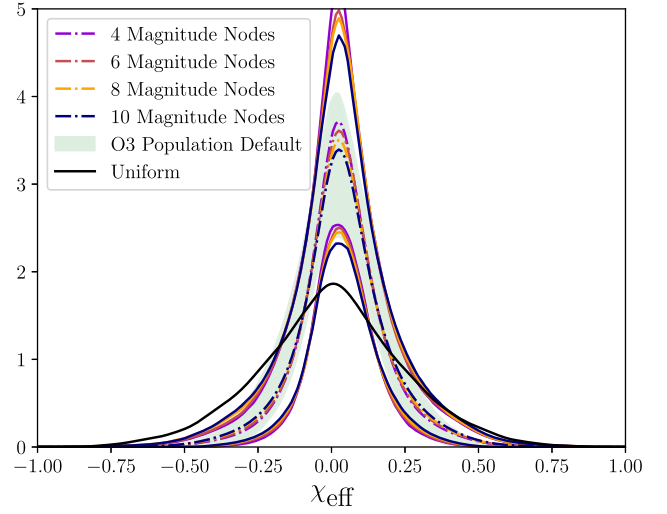

FIG. 4. Distribution of effective inspiral spin parameter as recovered from the distributions in Figs. 2 and 3.

We infer that $38.6\%^{+17.3\%}_{-15.6\%}$ of spin tilts are below $\cos\theta = 0$ and 50% of spin tilts are below $\cos\theta = 0.15^{+0.22}_{-0.22}$. Notably there is no trend for an increase in support for $\cos\theta = 1$ as would be predicted by a preferentially aligned-spin population (see Sec. V).

C. The distribution of χ_{eff}

As an alternative to modeling the component spins, it is common to consider instead the total spin contribution aligned with the orbital angular momentum, the so-called "effective" aligned spin parameter. This term is parametrized as

$$\chi_{\mathrm{eff}} = \frac{a_1 \cos\theta_1 + q a_2 \cos\theta_2}{1 + q}. \qquad (9)$$

While we do not directly model the distribution of χ_{eff} in this work, we can use the inferred distributions of $a, \cos(\theta)$, and q to reconstruct a distribution for χ_{eff} (cf. [7,29]).

Figure 4 shows this inferred distribution of χ_{eff} as we vary the number of tilt and magnitude nodes. As a point of comparison, we show the corresponding reconstruction of χ_{eff} when the distributions of component spin magnitudes and tilts are inferred using the Default model with the same catalog of events. In addition, we also plot the χ_{eff} distribution recovered from uniformly sampling in a and $\cos(\theta)$ (solid black curve), assuming mass ratios drawn from a power law with an index of 2 (q consistent with the results in [6]). The χ_{eff} distribution inferred from the spline model agrees well with the Default reconstruction but is a noticeably narrower distribution than a uniform spin magnitude and tilt distribution would result in. Using the model with ten magnitude and tilt nodes each, we infer that $38.7\%^{+12.8\%}_{-11.5\%}$ of BBH systems have $\chi_{\text{eff}} < 0$.

D. The distribution of χ_{p}

Another "effective" spin parameter commonly modeled in the gravitational wave literature is the effective precessing spin parameter χ_{p}, which quantifies the amount of in-plane spin present in a BBH merger [51]. Here,

$$\chi_{\text{p}} = \max\left[a_1\sin\theta_1, \left(\frac{3+4q}{4+3q}\right)qa_2\sin\theta_2\right]. \quad (10)$$

Similar to the previous subsection, we can reconstruct the distribution of χ_{p} using the spline models of a, $\cos(\theta)$, as well as our inference on the population of q.

Figure 5 shows the inferred distribution of χ_{p} as a function of different magnitude and tilt nodes, respectively. We also show the inferred distribution recovered from the Default model [6,7,33,34] analysis using the same event list, as well as the distribution corresponding to uniform distributions in a and $\cos\theta$. The inferred distribution of χ_{p} is consistent with an isotropic distribution and shows agreement with the χ_{p} distribution reconstructed from the Default model. An exception to this agreement is the slightly increased support at high χ_{p} that is not present in the Default reconstruction. The higher support for large a and $\cos(\theta) \approx 0$ in the spline model relative to the Default model explains this increased support at high χ_{p}. Similarly, our result allows for more support at high χ_{p} relative to what is presented in [6], in which it is assumed that the χ_{eff} and χ_{p} distributions follow a multivariate normal distribution [52]. In particular, Fig. 16 in [6] shows vanishing support for $\chi_{\text{p}} > 0.4$ in the population, whereas we find some support in this region is included at the 90% credibility. We note that varying the number of nodes in spin magnitude has the largest impact on the averaged recovered χ_{p} distribution (dash-dotted curves), indicating that this measurement depends on our choice of prior for the spin magnitude distribution.
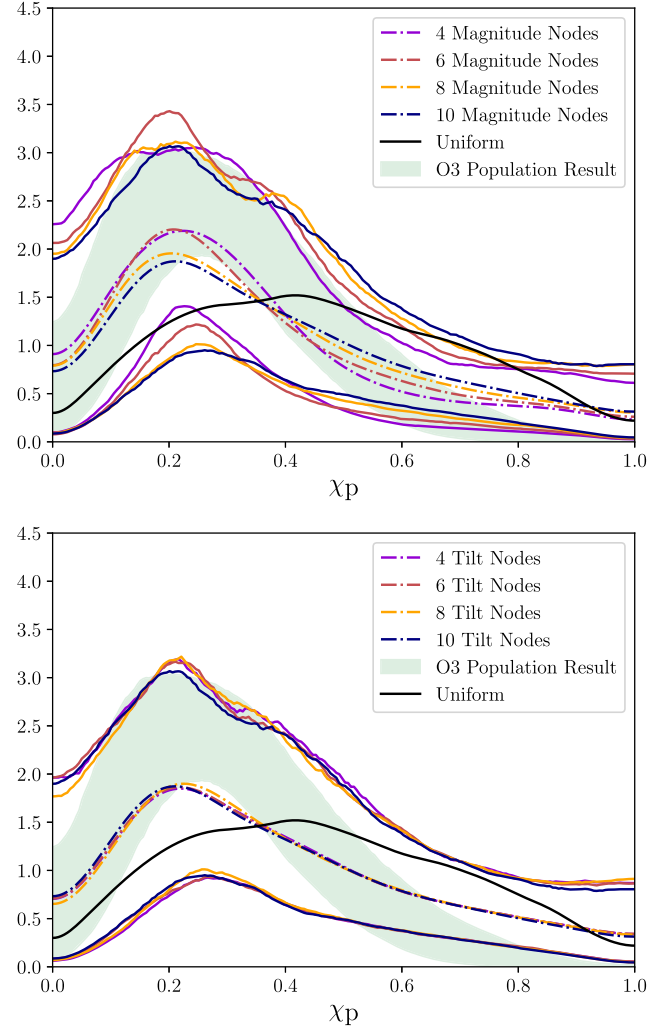

FIG. 5. Distribution of precession spin parameter as recovered from the distributions in Fig. 3.

V. DISCUSSION

Most previous analyses of the astrophysical distribution of merging binary black hole systems have focused on fitting parametric phenomenological models strongly constrained by the functional form of the model to the observed data (e.g., [6] and references therein) or directly compared with detailed simulations (e.g., [53,54]). However, more data-driven methods have been employed to infer the binary black hole mass [36,55–57] and spin distributions [56].

In this work, we use cubic splines (a model previously used to fit the black hole mass distribution [36]) to fit the astrophysical spin magnitude and spin tilt distributions of black holes as inferred from LIGO-Virgo observations. In doing so, we limit the influence of prior modeling assumptions on the inferred distribution and present a more data-driven result. While the uncertainties in the inferred distributions are large, we are able to interpret

trends as they relate to astrophysical mechanisms of BBH formation.

Models of stellar physics suggest that angular momentum transport out of the core of a collapsing star is highly efficient, indicating that first-generation stellar BHs should primarily have negligible spin [14]. Based on this, some models tend to favor nonspinning BHs when born in isolated environments and not susceptible to tidal spin-up.

Motivated by the work in [14,58], the authors of [6,29,30,59] search for contributions from a nonspinning subpopulation of BHs in the distributions of χ_{eff} and spin magnitude. Some of this previous work has found support for $a = 0$ when using low-dimensional parametric spin distribution models that allow for support at that point [10,29,60–62]. The results of such inference are strongly model dependent, with the preference for the presence of a nonspinning component depending on the morphology of that component. We do not confidently recover such a feature as demonstrated in Fig. 2. Given the width of the 90% credible interval at low spin magnitudes, we are unable to rule out the presence of this feature. This is consistent with what was found in [30,59,63], in which the authors find that there is insufficient data to resolve such a nonspinning subpopulation when employing a strongly parametric spin model with a spike at $a = 0$.

On the other hand, the merger of equal mass, nonspinning BBH systems are expected to result in a remnant BH with $a \sim 0.7$. As a result, population simulations predict that hierarchical mergers resulting from products of nonspinning first-generation mergers will leave a signature of a subpopulation of BH spins peaked around $a \sim 0.7$ with tails extending from $a \sim 0.5$ to ~ 0.9 [13]. Referring back to Fig. 2, we do not see evidence of an obvious subpopulation in this high-spin region of interest but rather some preference for low-spin magnitudes, possibly indicating that hierarchical mergers are not providing the dominant formation mechanism for the observed BBHs. The lack of support for a relatively high spin subpopulation is consistent with the conclusions drawn in [6,30]. Assuming that BBHs from hierarchical mergers all have $a > 0.5$, we infer that no more than $23\%^{+14\%}_{-11\%}$ of the astrophysical population of merging BBHs form through a hierarchical merger channel.

A notable feature in our analysis is the increased uncertainty in the spin magnitude distribution as compared to that from the Beta distribution in spin magnitude. The motivation for using a Beta distribution in [6,7,33] is not physical but is statistical: the Beta distribution has support only in the interval [0, 1] and offers a flexible, parametric fit for the mean and variance of a distribution and has an analytic form. The spline model introduced in this work offers more flexibility than the Beta distribution, so lacking a physical motivation for the Beta distribution, we expect that the uncertainties in the spin magnitude distribution obtained in this work are more appropriate than those

obtained from the Default model. Furthermore, the Beta distribution used in [6,7] cannot model structures such as increased support for nonspinning BHs or a secondary peak at high spin, making it a suboptimal model for the astrophysical spin magnitude distribution in the presence of a nonspinning subpopulation. Comparing to our data-driven approach, we therefore conclude that the resulting spin distributions presented in both [6] as well as this work are partially model driven.

When dynamical encounters take place within dense environments such as globular clusters, it is likely that some of the remnant BHs are retained in the cluster and merge in a subsequent dynamical encounter, contributing to the hierarchical merger population. The authors of [11] find that, for a broad range of populations considered, 16% of mergers in the hierarchical merger population have $\chi_{\text{eff}} < -0.3$. Using our inferred χ_{eff} distribution, we infer $2.1\%^{+3.9\%}_{-1.5\%}$ of BBH mergers have $\chi_{\text{eff}} < -0.3$. Using this interpretation from the χ_{eff} distribution, we place a conservative upper limit on the contribution of hierarchical mergers to the BBH merger population of $13\%^{+24\%}_{-9\%}$, which agrees with the one obtained when using just the spin magnitude information. This limit broadly agrees with the upper limit of 26% for the fraction of hierarchical mergers presented in [11], in which the authors use low-dimensional parametric models to infer the χ_{eff} distribution. This is also consistent with the results of [10], who found that depending on the escape velocity of the hierarchical merger environment up to $\approx 10\%$ of merging black holes may come from hierarchical mergers.

Mergers of BBH systems which have spins that are isotropic in orientation, as is expected from dynamical formation scenarios, implies a distribution of χ_{eff} symmetric about 0 (see the black line in Fig. 4). This prediction comes from the idea that, during a dynamical capture, there is no reason to expect that the two BHs should have correlated spin directions when they randomly encounter each other. In contrast, spins of BBH systems forming from common evolution are expected to remain primarily aligned with the orbital angular momentum, resulting in exclusively positive values for χ_{eff} from this population. While the distribution of χ_{eff} we recover appears symmetric, it is centered at $\chi_{\text{eff}} = 0.033^{+0.034}_{-0.038}$, favoring a positive central location but consistent with being centered at $\chi_{\text{eff}} = 0$ at the 90% level. This constraint is similar to that obtained by [64], using basis splines to model the component spin distributions. This result, coupled with the result that $38.7\%^{+12.8\%}_{-11.5\%}$ of events have $\chi_{\text{eff}} < 0$, presents the possibility that dynamical encounters are a significant contribution to the formation mechanisms of BBH merger systems. This is in contrast to the results reported in [6] that the χ_{eff} distribution is centered at 0.06 and rules out being centered at 0 at the 90% level; this result is obtained by modeling the χ_{eff} distribution as a Gaussian with the mean and standard deviation as free parameters [6,52]. While we use a largely

identical event list in our analyses, the modeling assumptions for the χ_{eff} distribution made in [6] probably explain some of the differences in our results. While [28,30] also find a χ_{eff} distribution consistent with being centered at zero, [28] does not find any support for $\chi_{\text{eff}} < 0$; such a difference may also be due to different modeling choices for the population of χ_{eff}.

We see increased uncertainty in $\cos\theta$ with respect to those obtained from the Default model from [6]. While the tilt distribution of the Default model is astrophysically motivated, it is incapable of capturing any possible substructure that may be present at locations other than $\cos\theta = 1$, as it is modeled by a monotonic function. Given the additional flexibility of the spline model, we notice a trend in the average line of the $\cos(\theta)$ distribution toward increased support for $-0.25 < \cos\theta < 0.5$. This trend is of low significance given the uncertainties surrounding it in the inferred $\cos\theta$ parameter space but may indicate a nontrivial contribution from BBHs with in-plane spins to the astrophysical population. This trend is consistent with what is found in [35,64], in which the authors use more flexible models to infer the $\cos\theta$ distribution. Our result that $38.6\%^{+17.3\%}_{-15.6\%}$ of BHs exhibit negative spin tilts is broadly consistent with previous studies that indicate the need for negative alignment in the astrophysical population. The presence of support for $\cos\theta < 0$ in the population was reported in [7] and confirmed in subsequent studies (e.g., [6,30]).

It is generally considered unlikely for BBH systems formed under common or isolated evolution scenarios to exhibit spin-orbit misalignment, as any such misalignment in these systems is expected to be corrected by mass transfer and tidal effects [23,24]. Lack of confidently increased support for $\cos\theta = 1$ indicates that aligned-spin BBH systems do not contribute a statistically resolvable subpopulation of mergers. A possible explanation for this is a comparable or more significant contribution of BBH mergers from dynamical encounters in dense environments to the inferred astrophysical population of BBH mergers, as this would manifest as a more isotropic distribution in tilts.

Unlike other work which adopts data-driven models but enforce regularization or smoothing conditions across parameter space (e.g., [30,64]), we set the correlation length explicitly by setting the number of spline nodes and their locations *a priori*. While smoothing over the scale of perturbations in parameter space may avoid the biases noted in Appendix B, it also limits the scale of the features that can be resolved by the model. By controlling for this scale limit on a model-by-model basis, we can evaluate if we need to resolve finer-scaled features in order to get a better fit to the data, for example, by comparing the model evidence in Fig. 1. We believe that the averaging or smoothing adopted in [30,64] is what prevents the spurious features noted in Appendix B from being recovered in those works, at the cost of not being able to resolve fine features

in the population distribution. Nonetheless, as we see from the similar evidence values in Fig. 1, it is not important that a model need to be able to resolve such narrow features in the spin distribution, as they do not seem to inform the posterior to a large extent. We anticipate that this will not be the case in analyses with future catalogs, as additional data will become more informative to the structures in the spin distribution, limiting the outsized influence of the model prior on the posterior. In this case, data-driven models may become necessary to best describe the distribution and this will be reflected by a higher evidence for these models relative to that of the Default model.

The morphological differences in the recovered spin distributions between models are a natural outcome of using different models to infer a distribution given uninformative data, as the prior provides much of the information to the posterior. We note there are visible differences in the recovered distribution between spline models and their credible regions in Figs. 2 and 3, such as the locations of large oscillations which appear depending on node placement. Because of the comparable evidence values in Fig. 1, we conclude that none of the spline models do a significantly better job describing the data than any of the others. We anticipate that, in future studies, additional observations will contribute to resolvable structure (if present) in the spin distributions; in this case, the spline models which best describe this structure will have the highest evidence. Similarly, the Bayes factor between the Default model and the spline model of the lowest evidence is only $\ln \text{BF} \sim 1.5$ in favor of the spline model, indicating that the Default model does not provide a significantly worse fit to the data than the spline models. In other words, the additional flexibility of the spline models presented in this work may not be necessary to correctly describe the spin distribution, as the lower-dimensional model provides a fit with similar evidence. We caution against trusting the recovered features in the resulting distributions as different models with similar evidences can recover fairly different features in the spin distribution.

The work presented in this paper motivates the need for more data-driven models for inferring the BBH spin distribution from GW sources, as there may be features of astrophysical importance that cannot be captured by currently used parametric models. While we cannot confidently discern many trends in the spin magnitude and tilt distributions, we can place constraints on the support in different parts of spin parameter space by substantially relaxing modeling assumptions. Using our more flexible model, we find that substantially increased uncertainties are a necessary cost to being able to model arbitrary features in the spin distribution, given current GW data. Data collected from events in future observing runs may help resolve such features which may exist in the spin distribution, as well as motivate a better choice of priors to use on these data-driven models.

APPENDIX A: EFFICIENT EVALUATION OF THE SPLINE MODEL

Our model requires evaluating a different spline model at many values during each likelihood evaluation. This process can be divided into three stages: constructing the spline model, identifying where each of evaluation points lies in relation to the nodes, and evaluating the appropriate piece in the spline. The first stage must be performed at every iteration but does not depend on the number of points that the spline will be evaluated at. The second stage is independent of the value of the spline nodes but must be performed for each of the evaluation points. For a uniform spacing of spline nodes, this can be efficiently evaluated; however, for a generic spline, this can be computationally intensive. At the third stage, we simply combine the results of the two previous stages.
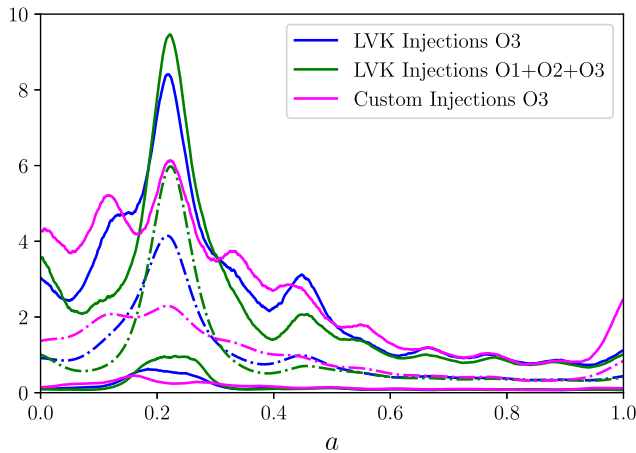
This can be trivially parallelized using a graphics processing unit (GPU).

For our use case, the locations at which the splines are evaluated and the node points are the same at every iteration. We can, therefore, cache the result of the second stage. We find that for our application the caching method accelerates the evaluation of the model by a factor of $\gtrsim 100$. Our implementation cached_interpolate is available via PYPI and CONDA-FORGE.

APPENDIX B: COMPARISON BETWEEN INJECTION SETS

Sensitivity estimates for Advanced LIGO and Virgo were released along with [4,6] for the first three observing runs [4,65,66]. These sensitivity estimates consist of injections of simulated sources into detector noise, along with the SNRs and FARs of these injections as reported by the detection pipelines used in the LIGO-Virgo observing runs. In this section, we compare the use these injections to compute the p_{det} term as written in Eq. (7) to the use of our own injections to compute the same term.

Using a set of injections, we include those which pass a detection threshold in the summation over the found injections. For the injections provided in [4], we use a threshold of FAR < 1 yr^{-1}, consistent with the choice made in [6]. We do not run the detection pipelines to assign a FAR to each of the injections from our custom injection set in this paper, so we threshold these on an optimal SNR > 10. We visually check that the distribution of found injections is not biased by our approximation of the detection threshold in the selection function, such that the distribution of detection probabilities across parameter space are qualitatively similar for the different injection sets. Since importance sampling Monte Carlo integration relies on drawing enough samples from the fiducial

FIG. 6. Distribution of spin magnitude using ten nodes. Different colors correspond to different sensitivity injection sets. Injection sets use events from their corresponding observing runs. Note that the peak at around $a = 0.2$ is most pronounced with using the O1 + O2 + O3 set and least with our custom O3 set.
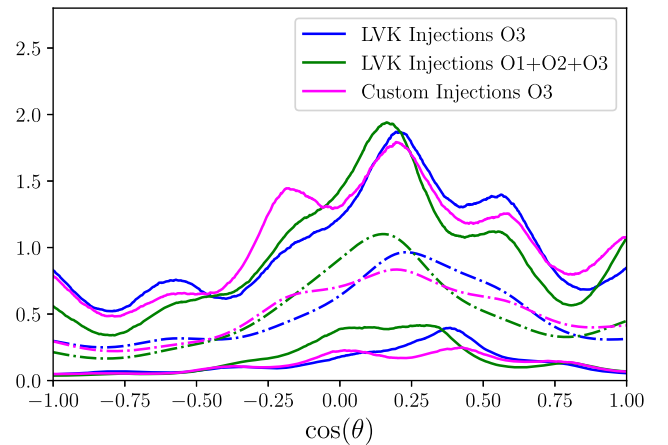

FIG. 7. Distribution of spin tilt using six nodes. Different colors correspond to different sensitivity injection sets. Injection sets use events from their corresponding observing runs.
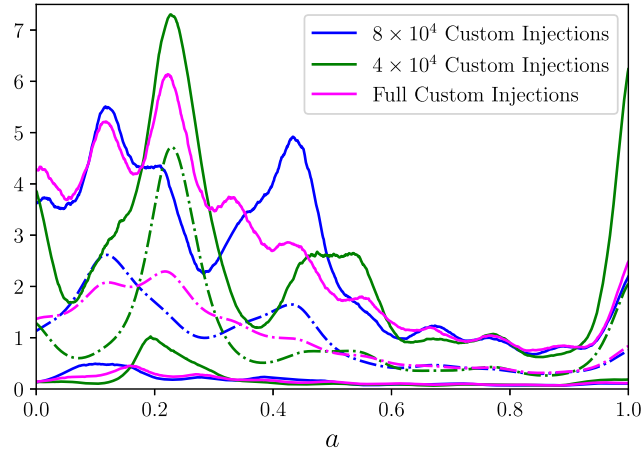
FIG. 8. Distribution of spin tilt using six nodes. Different colors correspond to a different number of found injections: the 900 000 from the custom injection set, ∼80 000, and ∼40 000, where the latter two are close to the number of found injections in the LVK O3-only and O1 + O2 + O3 injection sets, respectively.
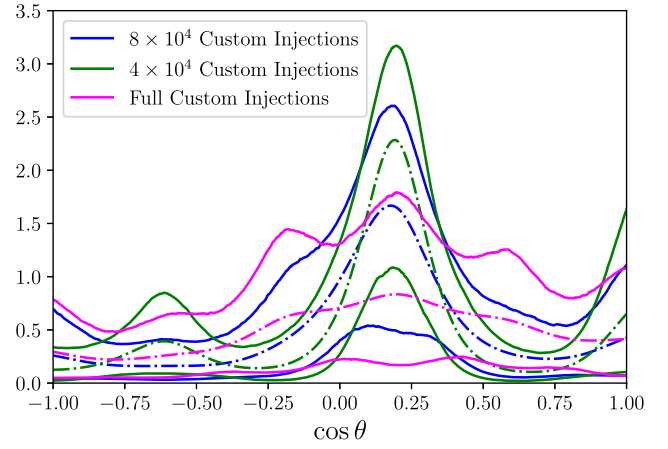

FIG. 9. Distribution of spin tilt using six nodes. Different colors correspond to a different number of found injections: the 900 000 from the custom injection set, ∼80 000, and ∼40 000, where the latter two are close to the number of found injections in the LVK O3-only and O1 + O2 + O3 injection sets, respectively.

distribution that cover the support of the target distribution, using Eq. (7) as a reliable estimator for Eq. (6) requires a suitable number of "found" injections to get a well-converged estimate (i.e., see Ref. [45]). With too few samples being used to compute the Monte Carlo approximation, the variance of our estimator is large and the resulting estimate may be a poor approximation of the true log-likelihood. The statistical uncertainty in the log-likelihood estimates at each point in parameter space can cause a systematic bias to appear in the resulting posterior distribution for the population.

The injection sets provided in [4,6] for the combined O1 + O2 + O3 sensitivity and O3 sensitivity, respectively, contain 41 972 and 81 117 simulated events which pass our detection threshold. As expected, the use of more samples reduces the uncertainty in the computed p_{det} estimate and is reflected by the distribution of variances in $\ln \mathcal{L}$, as shown in Fig. 10. We drastically reduce the variances in the log-likelihood estimates by using our own injection set which contains 911 386 injections passing our detection threshold. See Figs. 6 and 7 for comparisons between inferred spin distributions when using these different injection sets.

In order to validate that the number of samples used to compute p_{det} is a cause of systematic bias in the inferred population (as opposed to the difference in detection statistic used for the threshold), we repeat the above spin distribution inference but using injection sets that have been downsampled to have ∼40 000 and ∼80 000 found injections. In Figs. 8 and 9, we note recovery of strongly peaked features in the spin magnitude and tilt distributions, respectively. The significance of these features becomes drastically reduced as the number of found injections increases, indicating that a lower number of effective samples used to compute Eq. (7) can lead to biases that propagate into spurious features in the spin distribution. We

therefore infer that a sufficient number of injections is necessary to recover an unbiased spin distribution using our spline model implementation, motivating our use of the custom injection set with substantially more found injections than what was released in [66].

We quantify the uncertainty in the log-likelihood across the parameter space by calculating the uncertainty in the difference in log-likelihoods for all of Λ. In Fig. 10, we quantify this uncertainty as the covariance in the log-likelihood between each point Λ and Λ′, the value of Λ that has the lowest variance.

We see in Fig. 10 that the amplitude of the node at $a = 0.22$ is correlated with higher statistical variance; as the amplitude of this node increases, the uncertainty in the log-likelihood increases as well, making the log-likelihoods computed in this part of parameter space less trustworthy. As we decrease the variance by using injection sets with higher N_{inj}, the uncertainty in the log-likelihood estimates decreases. With better estimates of the log-likelihood, the support for the high amplitude of the node at $a = 0.22$ decreases, indicating that this feature in the spin magnitude distribution may be an artifact of poorly converged Monte Carlo integrals. If this peak were a true feature in the astrophysical population, we would expect the inferred distribution computed with the custom injections would maintain support for high amplitude at this node, along with reduced uncertainty. We confirm that this uncertainty is associated with the selection function rather than associated with reweighting posterior samples in the population model by comparing the contributions of the uncertainties in both these Monte Carlo summations to the total propagated uncertainty in the log-likelihood; for the models tested in this paper, we consistently find that the uncertainty associated with the contribution from Eq. (7) dominates.
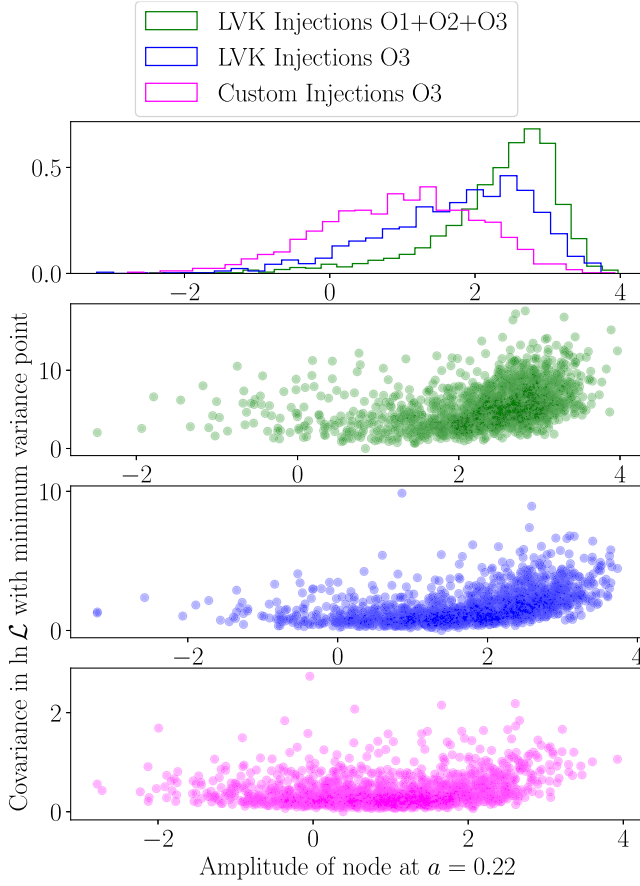
FIG. 10. Posterior samples for the amplitude of the spline node at $a = 0.22$ and their associated variances in the log-likelihood. Note the positive correlation between spline amplitude and variance. The results obtained using the injection set from all three observing runs, which has the least number of found injections, exhibits the highest variance in the log-likelihood.

We note that we have noticed several other examples of similar behavior in our analyses, notably manifesting as spurious peaks in the spin distribution. This demonstrates the need for sufficient coverage of injections when using importance sampling to compute sensitivity estimates especially when evaluating a population distribution that can model narrow peaks.

APPENDIX C: EFFECT OF PRIORS ON INFERRED DISTRIBUTION

In this work, we adopt a prior on the spline node amplitudes that is a standard normal distribution. In this appendix, we show the results obtained using different choices of prior on the node amplitudes and different positions.

In Fig. 13, we show the distribution of spin magnitudes and tilts from prior draws only. We note that the average of the distribution is flat, reflecting the lack of any further structure imposed by the prior on the mean of the distribution. On the other hand, the upper limit of the
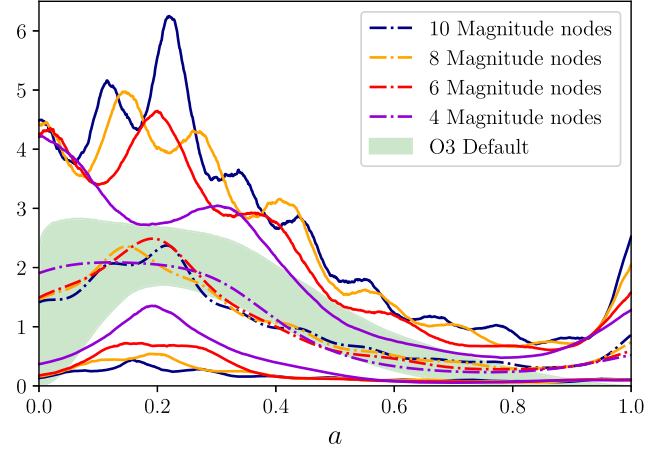

FIG. 11. Distribution of spin magnitudes, with different numbers of nodes corresponding to different colors. All use four nodes in the tilt distribution.

90% credible regions shows considerable oscillations. These are coincident with the node locations and, thus, correspond to where the distribution is informed directly by the spline amplitude sample. These oscillations are, thus, an expected feature of the spline model. The regions in between these oscillations correspond to where the spline provides an interpolation between node locations.

Figure 14 shows the inferred distribution for the four tilt node and four magnitude node model for three different choices of prior on the node amplitude and placement: a broader Gaussian (yellow), a narrower Gaussian (magenta), a uniform distribution in $[-3, 3]$, and the unit Gaussian without the additional end nodes (see Sec. II A). They each result in comparable distributions within the statistical uncertainties; however, we note that the wider prior distributions lead to wider uncertainty bands, with the $\mathcal{N}(0, 0.5)$ prior giving the tightest constraints. We expect this to also hold for the other node configurations.
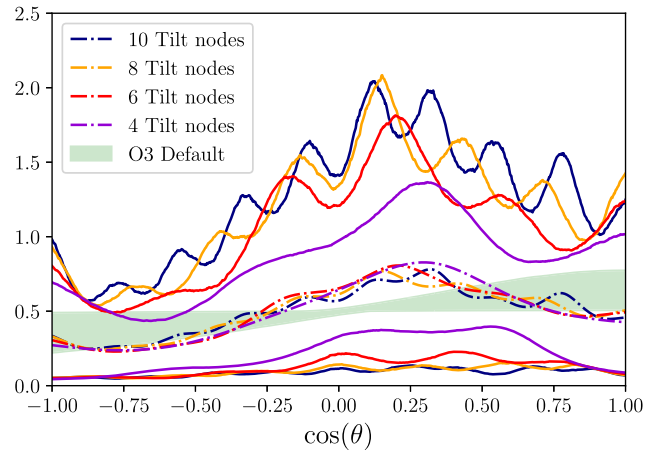

FIG. 12. Distribution of spin tilts, with different numbers of nodes corresponding to different colors. All use four nodes in the spin magnitude distribution.
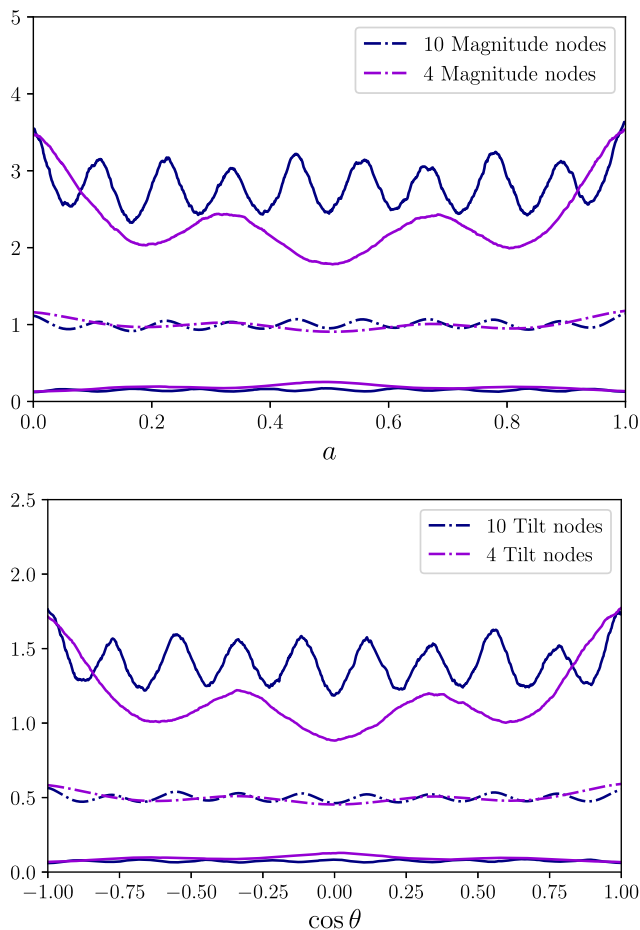
FIG. 13. Comparison of the priors on the spin magnitude (top) and spin tilt (bottom) distributions. The envelopes (increased width in the distribution) are coincident with the node placement.
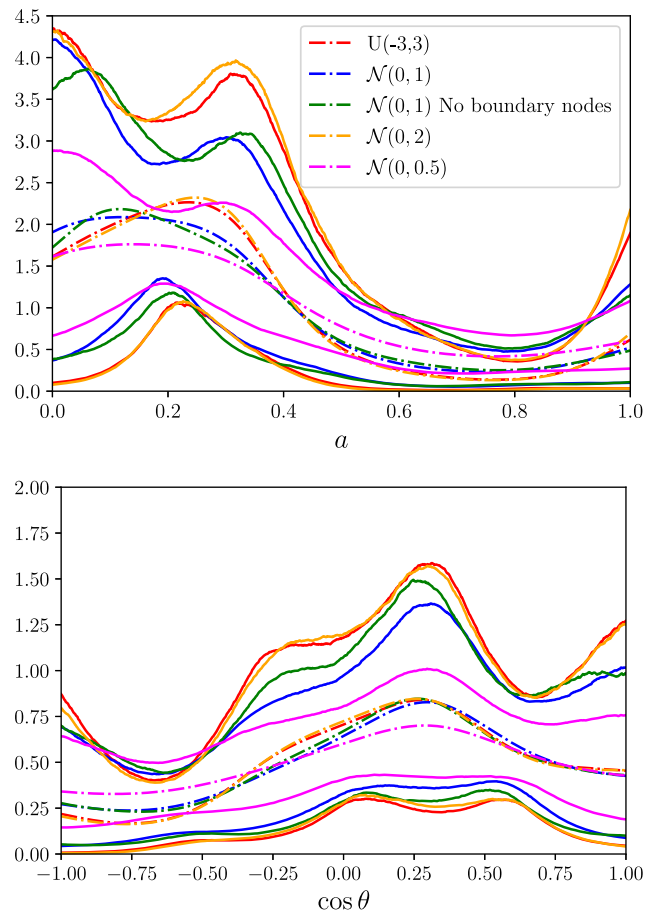

FIG. 14. Comparison of spin reconstructions for different choices of priors on the spline nodes. The blue distribution corresponds to choice used in this work. The distribution "no boundary nodes" also places the boundaries linearly in the domain but does not place any nodes on the domain boundary.

[1] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, Virgo, and VIRGO Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, Living Rev. Relativity **21**, 3 (2018).

[2] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).

[3] F. Acernese *et al.* (VIRGO Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[4] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration *et al.*, GWTC-3: Compact binary coalescences observed by LIGO and Virgo

during the second part of the third observing run, arXiv:2111.03606 [Phys. Rev. X (to be published)].

[5] LIGO Scientific Collaboration and Virgo Collaboration *et al.*, GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, Phys. Rev. X **9**, 031040 (2019).

[6] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration *et al.*, The population of merging compact binaries inferred using gravitational waves through GWTC-3, Phys. Rev. X **13**, 011048 (2023).

[7] LIGO Scientific Collaboration and Virgo Collaboration *et al.*, Population properties of compact objects from the

second LIGO-Virgo gravitational-wave transient catalog, Astrophys. J. Lett. **913,** L7 (2021).

[8] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration *et al.*, Constraints on the cosmic expansion history from GWTC-3, Astrophys. J. **949,** 76 (2023).

[9] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration *et al.*, Tests of general relativity with GWTC-3, arXiv:2112.06861.

[10] C. Kimball, C. Talbot, C. P. L. Berry, M. Zevin, E. Thrane, V. Kalogera, R. Buscicchio, M. Carney, T. Dent, H. Middleton, E. Payne, J. Veitch, and D. Williams, Evidence for hierarchical black hole mergers in the second LIGO-Virgo gravitational wave catalog, Astrophys. J. Lett. **915,** L35 (2021).

[11] M. Fishbach, C. Kimball, and V. Kalogera, Limits on hierarchical black hole mergers from the most negative $\chi_{\rm eff}$ systems, Astrophys. J. Lett. **935,** L26 (2022).

[12] D. Gerosa and M. Fishbach, Hierarchical mergers of stellar-mass black holes and their gravitational-wave signatures, Nat. Astron. **5,** 749 (2021).

[13] M. Fishbach, D. E. Holz, and B. Farr, Are LIGO's black holes made from smaller black holes?, Astrophys. J. Lett. **840,** L24 (2017).

[14] J. Fuller and L. Ma, Most black holes are born very slowly rotating, Astrophys. J. Lett. **881,** L1 (2019).

[15] Y. Qin, P. Marchant, T. Fragos, G. Meynet, and V. Kalogera, On the origin of black hole spin in high-mass x-ray binaries, Astrophys. J. Lett. **870,** L18 (2019).

[16] A. Heger, S. E. Woosley, and H. C. Spruit, Presupernova evolution of differentially rotating massive stars including magnetic fields, Astrophys. J. **626,** 350 (2005).

[17] M. Zevin and S. S. Bavera, Suspicious siblings: The distribution of mass and spin across component black holes in isolated binary evolution, Astrophys. J. **933,** 86 (2022).

[18] A. Olejak and K. Belczynski, The implications of high black hole spins for the origin of binary black hole mergers, Astrophys. J. Lett. **921,** L2 (2021).

[19] S. S. Bavera, T. Fragos, Y. Qin, E. Zapartas, C. J. Neijssel, I. Mandel, A. Batta, S. M. Gaebel, C. Kimball, and S. Stevenson, The origin of spin in binary black holes. Predicting the distributions of the main observables of Advanced LIGO, Astron. Astrophys. **635,** A97 (2020).

[20] S. S. Bavera, T. Fragos, M. Zevin, C. P. L. Berry, P. Marchant, J. J. Andrews, S. Coughlin, A. Dotter, K. Kovlakas, D. Misra, J. G. Serra-Perez, Y. Qin, K. A. Rocha, J. Román-Garza, N. H. Tran, and E. Zapartas, The impact of mass-transfer physics on the observable properties of field binary black hole populations, Astron. Astrophys. **647,** A153 (2021).

[21] I. Hinder, B. Vaishnav, F. Herrmann, D. M. Shoemaker, and P. Laguna, Circularization and final spin in eccentric binary-black-hole inspirals, Phys. Rev. D **77,** 081502 (2008).

[22] F. Hofmann, E. Barausse, and L. Rezzolla, The final spin from binary black holes in quasi-circular orbits, Astrophys. J. Lett. **825,** L19 (2016).

[23] W. M. Farr, S. Stevenson, M. C. Miller, I. Mandel, B. Farr, and A. Vecchio, Distinguishing spin-aligned and isotropic black hole populations with gravitational waves, Nature (London) **548,** 426 (2017).

[24] V. Kalogera, Spin-orbit misalignment in close binaries with two compact objects, Astrophys. J. **541,** 319 (2000).

[25] I. Mandel and R. O'Shaughnessy, Compact binary coalescences in the band of ground-based gravitational-wave detectors, Classical Quantum Gravity **27,** 114007 (2010).

[26] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, Illuminating black hole binary formation channels with spins in Advanced LIGO, Astrophys. J. Lett. **832,** L2 (2016).

[27] Z. Doctor, D. Wysocki, R. O'Shaughnessy, D. E. Holz, and B. Farr, Black hole coagulation: Modeling hierarchical mergers in black hole populations, Astrophys. J. **893,** 35 (2020).

[28] J. Roulet, H. S. Chia, S. Olsen, L. Dai, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Distribution of effective spins and masses of binary black holes from the LIGO and Virgo o1–o3a observing runs, Phys. Rev. D **104,** 083010 (2021).

[29] S. Galaudage, C. Talbot, T. Nagar, D. Jain, E. Thrane, and I. Mandel, Building better spin models for merging binary black holes: Evidence for nonspinning and rapidly spinning nearly aligned subpopulations, Astrophys. J. Lett. **921,** L15 (2021).

[30] T. A. Callister, S. J. Miller, K. Chatziioannou, and W. M. Farr, No evidence that the majority of black holes in binaries have zero spin, Astrophys. J. Lett. **937,** L13 (2022).

[31] E. Payne and E. Thrane, Model exploration in gravitational-wave astronomy with the maximum population likelihood, Phys. Rev. Res. **5,** 023013 (2023).

[32] I. M. Romero-Shaw, E. Thrane, and P. D. Lasky, When models fail: An introduction to posterior predictive checks and model misspecification in gravitational-wave astronomy, Pub. Astron. Soc. Aust. **39,** e025 (2022).

[33] D. Wysocki, J. Lange, and R. O'Shaughnessy, Reconstructing phenomenological distributions of compact binaries via gravitational wave observations, Phys. Rev. D **100,** 043012 (2019).

[34] C. Talbot and E. Thrane, Determining the population properties of spinning black holes, Phys. Rev. D **96,** 023012 (2017).

[35] S. Vitale, S. Biscoveanu, and C. Talbot, Spin it as you like: The (lack of a) measurement of the spin tilt distribution with LIGO-Virgo-KAGRA binary black holes, Astron. Astrophys. **668,** L2 (2022).

[36] B. Edelman, Z. Doctor, J. Godfrey, and B. Farr, Ain't no mountain high enough: Semiparametric modeling of LIGO-Virgo's binary black hole mass distribution, Astrophys. J. **924,** 101 (2022).

[37] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, Pub. Astron. Soc. Aust. **36,** e010 (2019).

[38] C. Talbot and E. Thrane, Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization, Astrophys. J. **856,** 173 (2018).

[39] M. Fishbach, D. E. Holz, and W. M. Farr, Does the black hole merger rate evolve with redshift?, Astrophys. J. Lett. **863,** L41 (2018).

[40] LIGO Scientific Collaboration and Virgo Collaboration, Binary black hole population properties inferred from the

first and second observing runs of Advanced LIGO and Advanced Virgo, Astrophys. J. Lett. **882,** L24 (2019).

[41] C. Talbot, R. Smith, E. Thrane, and G. B. Poole, Parallelized inference for gravitational-wave astronomy, Phys. Rev. D **100,** 043030 (2019).

[42] J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493,** 3132 (2020).

[43] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, Inferring the properties of a population of compact binaries in presence of selection effects, in *Handbook of Gravitational Wave Astronomy*, edited by C. Bambi (Springer, New York, 2022), p. 45.

[44] LIGO Scientific Collaboration and Virgo Collaboration *et al.*, Binary black hole mergers in the first Advanced LIGO observing run, Phys. Rev. X **6,** 041015 (2016).

[45] W. M. Farr, Accuracy requirements for empirically measured selection functions, Res. Not. Am. Astron. Soc. **3,** 66 (2019).

[46] LIGO Scientific Collaboration and Virgo Collaboration *et al.*, GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Phys. Rev. X **11,** 021053 (2021).

[47] J. Golomb and C. Talbot, Hierarchical inference of binary neutron star mass distribution and equation of state with gravitational waves, Astrophys. J. **926,** 79 (2022).

[48] L. Martino, V. Elvira, and F. Louzada, Effective sample size for importance sampling based on discrepancy measures, arXiv:1602.03572.

[49] R. Essick and W. Farr, Precision requirements for Monte Carlo sums within hierarchical Bayesian inference, arXiv:2204.00461.

[50] C. Talbot and J. Golomb, Growing pains: Understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy, Mon. Not. R. Astron. Soc., **526,** 3495 (2023).

[51] P. Schmidt, F. Ohme, and M. Hannam, Towards models of gravitational waveforms from generic binaries: II. Modelling precession effects with a single effective precession parameter, Phys. Rev. D **91,** 024043 (2015).

[52] S. Miller, T. A. Callister, and W. M. Farr, The low effective spin of binary black holes and implications for individual gravitational-wave events, Astrophys. J. **895,** 128 (2020).

[53] M. Zevin, S. S. Bavera, C. P. L. Berry, V. Kalogera, T. Fragos, P. Marchant, C. L. Rodriguez, F. Antonini, D. E. Holz, and C. Pankow, One channel to rule them all? Constraining the origins of binary black holes using multiple formation pathways, Astrophys. J. **910,** 152 (2021).

[54] K. W. K. Wong, K. Breivik, K. Kremer, and T. Callister, Joint constraints on the field-cluster mixing fraction, common envelope efficiency, and globular cluster radii from a population of binary hole mergers via deep learning, Phys. Rev. D **103,** 083021 (2021).

[55] I. Mandel, W. M. Farr, A. Colonna, S. Stevenson, P. Tiňo, and J. Veitch, Model-independent inference on compact-binary observations, Mon. Not. R. Astron. Soc. **465,** 3254 (2017).

[56] V. Tiwari, VAMANA: Modeling binary black hole population with minimal assumptions, Classical Quantum Gravity **38,** 155007 (2021).

[57] S. Rinaldi and W. Del Pozzo, (H)DPGMM: A hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function, Mon. Not. R. Astron. Soc. **509,** 5454 (2022).

[58] K. Belczynski *et al.*, Evolutionary roads leading to low effective spins, high black hole masses, and O1/O2 rates for LIGO/Virgo binary black holes, Astron. Astrophys. **636,** A104 (2020).

[59] H. Tong, S. Galaudage, and E. Thrane, The population properties of spinning black holes using gravitational-wave transient catalog 3, Phys. Rev. D **106,** 103019 (2022).

[60] C. Talbot and E. Thrane, Fast, flexible, and accurate evaluation of gravitational-wave Malmquist bias with machine learning, Astrophys. J. **927,** 76 (2022).

[61] Y.-Z. Wang, Y.-J. Li, J. S. Vink, Y.-Z. Fan, S.-P. Tang, Y. Qin, and D.-M. Wei, Tight constraint on the maximum mass of stellar-origin binary black holes and evidence for hierarchical mergers in gravitational wave observations, Astrophys. J. Lett. **941,** L39 (2022).

[62] C. Kimball, C. Talbot, C. P. L. Berry, M. Carney, M. Zevin, E. Thrane, and V. Kalogera, Black hole genealogy: Identifying hierarchical mergers with gravitational waves, Astrophys. J. **900,** 177 (2020).

[63] M. Mould, D. Gerosa, F. S. Broekgaarden, and N. Steinle, Which black hole formed first? Mass-ratio reversal in massive binary stars from gravitational-wave data, Mon. Not. R. Astron. Soc. **517,** 2738 (2022).

[64] B. Edelman, B. Farr, and Z. Doctor, Cover your basis: Comprehensive data-driven characterization of the binary black hole population, Astrophys. J. **946,** 16 (2023).

[65] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run—O1 + O2 + O3 search sensitivity estimates, 10.5281/zenodo.5636815.

[66] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run—O3 search sensitivity estimates, 10.5281/zenodo.5546675.