Machine-learning interpolation of population-synthesis simulations to interpret gravitational-wave observations: A case study

Kaze W. K. Wong [1,*] and Davide Gerosa [2,†]

[1]*Department of Physics and Astronomy, Johns Hopkins University,*
3400 N. Charles Street, Baltimore, Maryland 21218, USA
[2]*School of Physics and Astronomy and Institute for Gravitational Wave Astronomy,*
University of Birmingham, Birmingham B15 2TT, United Kingdom

We report on advances to interpret current and future gravitational-wave events in light of astrophysical simulations. A machine-learning emulator is trained on numerical population-synthesis predictions and inserted into a Bayesian hierarchical framework. In this case study, a modest but state-of-the-art suite of simulations of isolated binary stars is interpolated across two event parameters and one population parameter. The validation process of our pipelines highlights how omitting some of the event parameters might cause errors in estimating selection effects, which propagates as systematics to the final population inference. Using LIGO/Virgo data from O1 and O2 we infer that black holes in binaries are most likely to receive natal kicks with one-dimensional velocity dispersion $\sigma = 105^{+44}_{-29}$ km/s. Our results showcase potential applications of machine-learning tools in conjunction with population-synthesis simulations and gravitational-wave data.

I. INTRODUCTION

The observed catalog of gravitational-wave (GW) detections is growing at a fast pace. Eleven [1] (or possibly more [2]) events have been announced so far and hundreds more are expected to be observed within a few years. This opens a unique possibility of inferring properties of the population of merging compact binaries in the Universe.

Current state-of-the-art analyses assume that the underlying population of GW sources is described by some phenomenological parametric expression (e.g., [3]). For instance, one can assume that the mass distribution of merging binary black holes (BHs) follows a power law, and use GW data to infer its spectral index. As the size of the GW catalog grows, increasingly complex parametrizations and nonparametric tests (e.g., [4]) will allow us to capture finer and finer details of the observed population.

Astrophysical predictions of GW populations are typically computed using population-synthesis codes [5–10]– collections of prescriptions that encode our understanding and ignorance of how compact binaries form and evolve from their stellar progenitors. The parameters needed to initialize a population-synthesis simulation are directly related to poorly understood astrophysical mechanisms ruling the lives of massive stars. These include, for instance, efficiency of the common-envelope phase

(if any), strength of the supernova kicks, fallback material, mass-loss rates, stellar winds, etc.

A possible strategy to exploit future large GW catalogs is to bypass phenomenological models and compare data directly against population-synthesis simulations. A first step in this direction consists of estimating mixing fractions between two or more precomputed models [11–14]. More ambitiously, one could use GW data to infer the set of code input flags that best matches the observations. This approach faces an immediate difficulty, namely that a new, computationally expensive population-synthesis simulation is required at each evaluation of the population likelihood.

Progress to overcome this limitation was recently presented by Taylor and Gerosa [15]. By combining Gaussian process regression (GPR), principal component analysis (PCA), space-filling algorithms, and a hierarchical Bayesian framework, they were able to efficiently interpolate a precomputed bank of population-synthesis simulations and use the resulting emulator to infer the posterior distributions of the population parameters. Their method, however, was only applied to a few simple scenarios [15].

In this paper, we present a more realistic application: we train a GPR interpolant on a small, but state-of-the-art, set of population-synthesis predictions of BH binaries formed in isolation [16]. The resulting emulator slots into a hierarchical Bayesian analysis and is fed with BH binary data from LIGO/Virgo first (O1) and second (O2) observing runs. Under these astrophysical assumptions, we measure the natal kicks that BHs receive at birth.

[*]kazewong@jhu.edu
[†]d.gerosa@bham.ac.uk

This case study sheds light on some of the challenges one has to overcome to fully compare GW data and astrophysical simulations. In particular, we show that omitting a subset of the single-event parameters, either because they are not modeled in the simulations or simply because they make the inference problem computationally prohibitive, will cause a systematic bias on the final population inference.

This paper is organized as follows. In Sec. II we briefly review the procedure of Ref. [15] and present the current application. In Sec. III we show inference results using both mock datasets and real observations. Finally, in Sec. IV we discuss some astrophysical implication of our findings and highlight future developments of this approach.

II. METHODS

A. Hierarchical Bayesian inference

For each GW event, data d are routinely analyzed using Bayesian inference [1]. The chosen priors encode one's physical intuition of the underlying population and play an important role when interpreting the results [17]. A hierarchical analysis aims at parametrizing the choice of prior and using data to infer the resulting "hyper-parameters." For clarity, in the following we will denote parameters describing single events (e.g., masses, redshifts) as "event parameters," and the hyperparameters describing the entire sample as "population parameters" (e.g., the strength of supernova kicks). Let us assume that a set of population parameters λ predicts a distribution of event parameters θ

$$\frac{\mathrm{d}}{\mathrm{d}\theta} r(\lambda) = r(\lambda) p_{\mathrm{pop}}(\theta|\lambda), \tag{1}$$

where $\int p_{\mathrm{pop}}(\theta|\lambda)\mathrm{d}\theta = 1$ and the total rate $r(\lambda)$ is typically measured in yr^{-1}. This function encodes our astrophysical assumptions on the underlying populations. It can be estimated using a parameterized model, a population-synthesis simulation, or, in our case, by evaluating a machine-learning emulator. The predicted number of events is $N(\lambda) = r(\lambda) \times T_{\mathrm{obs}}$ where T_{obs} is the duration of the observing run(s).

We wish to analyze a GW catalog containing N_{obs} entries. For simplicity, we assume that all the events present in the catalog are of astrophysical origin. A more complete analysis including triggers with larger false-alarm probabilities is left to future work (e.g., [18]). Single-event posterior $p(\theta|d)$ are computed using some default prior $\pi(\theta)$ which is chosen by issuers of the catalog. In practice, both prior and posterior are usually provided under the form of Monte Carlo samples [1].

Detector selection effect are encoded in a function $0 \leq p_{\mathrm{det}}(\theta) \leq 1$, indicating the likelihood that an event with parameters θ appears in the catalog. This is used to define the observable distribution

$$\frac{\mathrm{d}}{\mathrm{d}\theta} r_{\mathrm{det}}(\lambda) = r(\lambda) p_{\mathrm{pop}}(\theta|\lambda) p_{\mathrm{det}}(\theta) \tag{2}$$

and the expected number of observations $N_{\mathrm{det}}(\lambda) = r_{\mathrm{det}}(\lambda) \times T_{\mathrm{obs}}$. Here we follow a common approach and approximate $p_{\mathrm{det}}(\theta)$ using the single-detector semi-analytic approximation of Refs. [19,20]) as implemented in Ref. [21] with a signal-to-noise ratio threshold equal to 8 and the waveform model of Ref. [22]. This was shown to be in good agreement with large-scale injection campaigns [3,23].

All these ingredients enter the population likelihood, which has the standard expression of an inhomogeneous Poisson process (cf. Refs. [15,24–26] for detailed derivations). In particular, the population posterior reads

$$p(\lambda|d) \propto \pi(\lambda) e^{-N_{\mathrm{det}}(\lambda)} N(\lambda)^{N_{\mathrm{obs}}} \prod_{i=1}^{N_{\mathrm{obs}}} \int \frac{p_i(\theta|d)}{\pi_i(\theta)} p_{\mathrm{pop}}(\theta|\lambda)\mathrm{d}\theta,$$

$$\tag{3}$$

where $\pi(\lambda)$ is some assumed population prior. If one wishes exclude rate information from the inference, a marginalization over $N(\lambda)$ with prior $\propto 1/N(\lambda)$ yields [27]

$$p(\lambda|d) \propto \pi(\lambda) \prod_{i=1}^{N_{\mathrm{obs}}} \int \frac{p_i(\theta|d)}{\pi_i(\theta)} \frac{p_{\mathrm{pop}}(\theta|\lambda)}{\int p_{\mathrm{pop}}(\theta|\lambda) p_{\mathrm{det}}(\theta)} \mathrm{d}\theta. \tag{4}$$

B. Training simulations

In this paper we consider GW sources formed in isolation via a common-envelope phase–a leading formation channel for binary BHs (e.g., [28]). We use the set of predictions presented in Refs. [16,29] (see references therein). The simulations are performed with the STARTRACK [8] and PRECESSION [30] codes (see also [31]). In particular, these runs employ the same set of assumptions used in model M10 of Ref. [32], except that BH natal kicks are not suppressed compared to neutron stars. Kicks are drawn from a Maxwellian distribution with one-dimensional velocity dispersion σ. We consider seven simulations with $\sigma = 0$, 25, 50, 70, 130, 200, and 265 km/s, the latter being the value inferred from pulsar proper motions [33].

Each event in our synthetic catalogs is described by masses, spins, redshift, as well as its contribution to the total merger rate $\mathrm{d}r/\mathrm{d}\theta$ [34]. In this paper, we restrict the event parameters used in our statistical inference to masses, redshifts, and rates (see below). Even though BH spins are not directly considered, assumptions on their distribution enter the waveform, hence the detection rates. Using the various spin models developed by Ref. [16], we verified that this indirect spin effect has a negligible impact on our final results. For concreteness, in the following we use the "time-uniform" model [16]. All binaries are assumed to reach the LIGO/Virgo band in quasicircular orbits.

Selection effects $p_{\text{det}}(\theta)$ are computed using sensitivity curves for both LIGO in its design configuration and LIGO during O1/O2. In particular, we use the "Design Sensitivity" and the "Early High Sensitivity" from Ref. [35], respectively (cf. [3]).

C. Gaussian processing

Astrophysical simulations are used to train a Gaussian process interpolator to quickly evaluate the rates $dr(\lambda)/d\theta$. In particular, our problem has a single population parameter $\lambda = \{\sigma\}$. Our implementation closely follows that of Ref. [15]. We first assume a common binning scheme across all simulations and convert the distribution using PCA.[1] The resulting features are then interpolated across the hyperparameter space with GPR.

The choice of event parameters used in the inference needs to be addressed with care. A trade-off is present between the size of the vector θ and the resulting GPR accuracy. A larger number of event parameters would increase the amount of astrophysical information captured by the analysis. However, this requires larger training banks to keep the interpolation error under control. This means, not surprisingly, that if we want to increase the amount of information used in the hyperparameters inference, we also need to feed in more training data to ensure the same accuracy.

We found that our set of 7 simulations allows us to accurately interpolate across source-frame chirp mass and redshift, i.e., $\theta = \{M_c, z\}$. For this (admittedly modest) training bank, inserting additional parameters, like mass ratio or effective spin, significantly degrades the performance of the interpolator. In particular, we use 40 equispaced bins in $M_c \in [5, 45]$ and $z \in [0, 1]$. GPR is implemented using SCIKIT-LEARN [36] with a squared exponential kernel as in Ref. [15]. The population posterior of Eq. (3) is sampled using EMCEE [37].

We validate our pipeline using a standard out-of-sample test. We train our regression machine using 6 simulations and validate results against the one that was left out. Figure 1 shows the predicted distributions of chirp mass M_c, redshift z, and rates r. Other than some small-scale differences, the interpolator accurately captures all the main features of the training set. For instance, we found a fractional difference in the intrinsic rate as small as $\Delta r / r \sim 6\%$

Reducing the number of event parameters implies that the quantities used in the inference differ from those needed to compute selection effects via $p_{\text{det}}(\theta)$. Assumptions on other parameters beyond chirp mass and redshift such as mass ratio and spins will inevitably be

---

[1]PCA naturally allows reducing the size of the computational problem by filtering out unnecessary features [15]. In this case, we are only using 7 training simulations and are able to process the entire distributions without any compression.
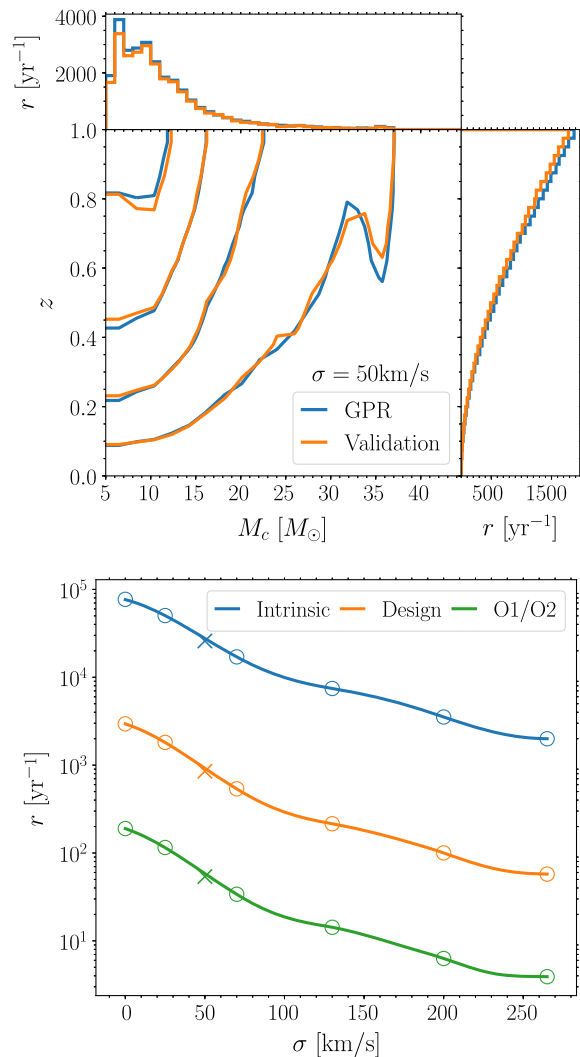

FIG. 1. Out-of-sample test of our machine-learning interpolator. The simulation with $\sigma = 50$ km/s is excluded from the training dataset and used to validate results. The top panel shows the intrinsic distributions of chirp mass and redshift $dr/d\theta$. Blue curves show the interpolated result, while orange curves show the control set. Contours mark 30%, 50%, 70% and 90% confidence intervals; side histograms show the marginalized distributions. The bottom panel shows detection rates across the hyperparameter space. In particular, the blue line shows intrinsic rates r, while orange and green lines show observable rates r_{det} for LIGO at design sensitivity and during O1/O2, respectively. Circles mark the simulations used to train the interpolator; crosses mark the validating dataset.

necessary to compute waveforms. For this reason, one cannot simply interpolate $dr(\lambda)/d\theta$ across $\theta = \{M_c, z\}$ and compute $N_{\text{det}}(\lambda)$ at each likelihood evaluation. We bypass the issue by computing the detectable rates from the training simulations (where mass ratios and spins are provided) and running a second GPR/PCA interpolation on $dr_{\text{det}}(\lambda)/d\theta$ (cf. Fig. 1). This approach is tested in the next section.
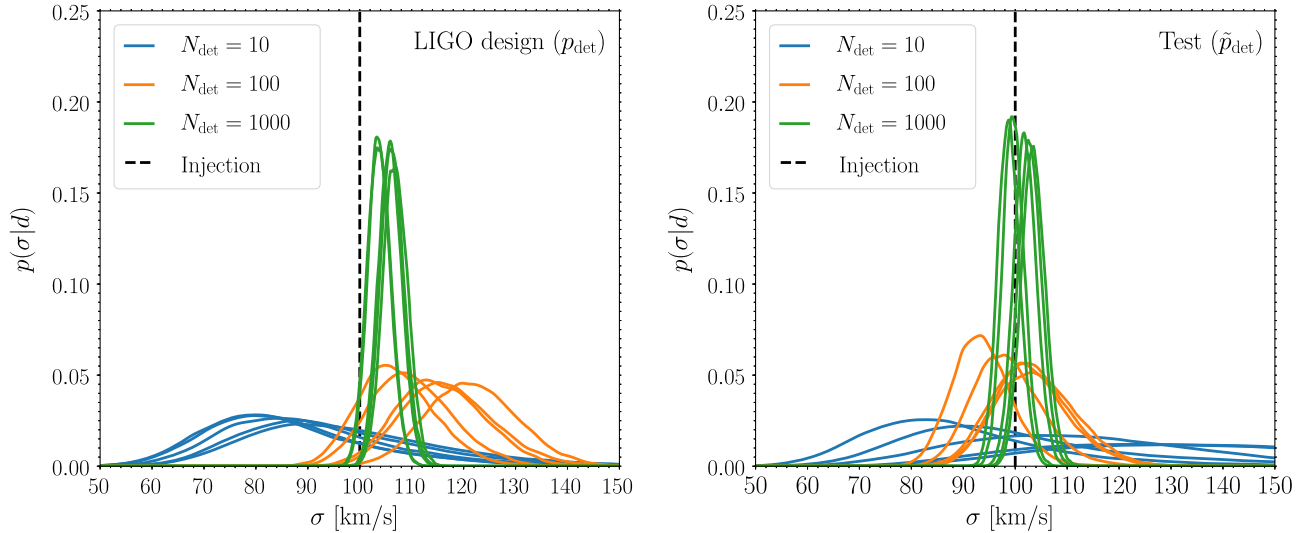
FIG. 2. Injection-recovery test. The left panel considers LIGO at design sensitivity, where selection effects are included through p_{det}. In this case, both $dr/d\theta$ and $dr_{det}/d\theta$ are interpolated from population-synthesis simulations. The right panel shows a test run where we assume a fictitious LIGO detector described by $\tilde{p}_{det}$ (see the description in the text), which allows us to only interpolate $dr/d\theta$. The injected value $\bar{\sigma} = 100$ km/s is marked by vertical dashed lines. Solid curves show posterior distributions of the strength of supernova kicks σ assuming the predicted number of observable sources is $N_{det}(\bar{\sigma}) = 10$ (blue), 100 (orange), and 1000 (green). A few realizations are reported for each of these cases.

III. RESULTS

A. Mock data

We first apply our statistical pipeline to mock data. We assume that a population of BH binaries with true value $\bar{\sigma} = 100$ km/s is observed by LIGO at design sensitivity. This is a location in parameter space where we have not performed a population-synthesis simulation. The observing time T_{obs} is chosen such that the predicted number of observation is $N_{det}(\bar{\sigma}) = \{10, 100, 1000\}$. We assume the number of entries N_{obs} in our mock catalogs is Poisson-distributed with mean $N_{det}(\bar{\sigma})$. Injections events $\theta = \{M_c, z\}$ are generated from the $dr_{det}/d\theta$ emulator. For simplicity, we assume posterior $p_i(\theta|d)$ are bivariate Gaussians centered on those extracted values with standard deviations equal to 10%. We sample the posterior of Eq. (3) assuming a flat priors on M_c, z and σ.

Our injection-recovery results are shown in left panel of Fig. 2 for several catalog realizations. The injected value $\bar{\sigma} = 100$ km/s is well within the predicted posterior only for the case with $N_{det} = 10$. As the number of detections grows, a systematic bias becomes more and more evident. For $N_{det} = 1000$, the posteriors peak at ∼105 km/s and the true values lies well outside the 90% confidence interval.

This bias is somewhat expected because we are neglecting some of the event parameters and only considering $\theta = \{M_c, z\}$. This forces us to interpolate $dr(\lambda)/d\theta$ and $dr_{det}(\lambda)/d\theta$ separately. In normal circumstance these two distributions are related by a single detectability function, i.e., $dr_{det}(\lambda)/d\theta = p_{det}(\theta)dr(\lambda)/d\theta$. Our pipeline, however, violates this condition because the two interpolants have different interpolation errors occurring in each bin.

We test this interpretation by considering a fictitious detector where selection effects $\tilde{p}_{det}$ depends only on M_c and z. This is constructed by assuming the same LIGO sensitivity curve which, however, responds to all binary BHs as if it they were equal mass and nonspinning, i.e., $\tilde{p}_{det}(M_c, z) = p_{det}(M_c, z, q = 1, \chi_1 = 0, \chi_2 = 0)$. In this case, we can interpolate only $dr/d\theta$ using population synthesis data. Injections are constructed extracting couples $\{M_c, z\}$ from $dr/d\theta$ and accepting/rejecting each draw according to $\tilde{p}_{det}(\theta)$. The factor $e^{-N_{det}(\lambda)}$ in Eq. (3) is estimated by integrating $\tilde{p}_{det}(\theta) \times dr/d\theta$ at each likelihood evaluation. Results are shown in the right panel of Fig. 2. We recover a largely unbiased estimates of the population parameter.

As discussed above, the number of parameters we can confidently interpolate is limited by size of the training dataset. A larger set of simulations would allow us to model more event parameters, consequently reducing systematic uncertainties on the resulting inference. Figure 2 shows, however, that the present simulation set is appropriate for $\mathcal{O}(10)$ events, as in this case statistical uncertainties largely dominates over systematics. We thus proceed by analyzing the 10 BH binary events detected during O1 and O2.

B. Events from LIGO/Virgo O1 and O2

After removing data segments contaminated by significant noise sources, LIGO/Virgo O1 and O2 resulted in $T_{obs} = 48.6$ and 118 days of coincident data, respectively [1,38]. We make use of posterior and prior samples of 10 binary BH coalescences publicly released by the LIGO and Virgo collaborations [39]. In particular, they provide luminosity
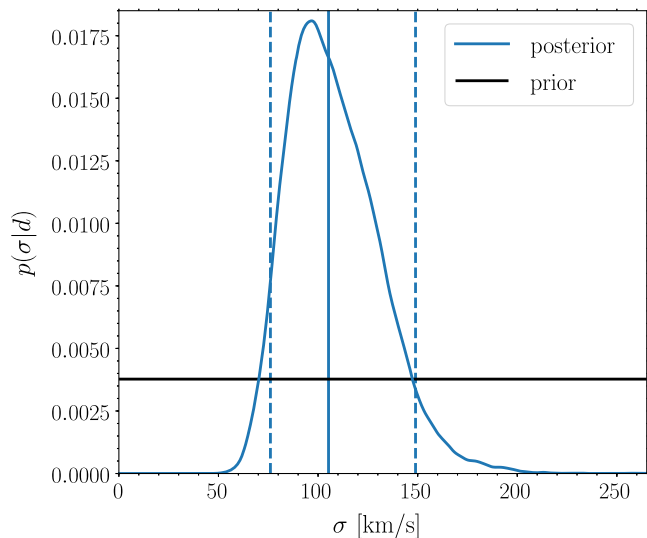
FIG. 3.   Constraints on the strength of BH natal kicks σ using 10 BH binary data from LIGO/Virgo O1 and O2. The blue curve shows our posterior distribution. Vertical blue lines show the corresponding median (solid) and 90% confidence interval (dashed). We assume a flat prior (horizontal black line).

distance and detector-frame masses, which we convert to redshift and source-frame masses. A Gaussian kernel-density estimation is then employed to obtain $\pi(\theta)$ at the locations of the posterior samples, which allows approximating the integrals in Eqs. (3)–(4) as Monte Carlo sums.

The resulting inference is illustrated in Fig. 3, where we show the posterior distribution of the population parameter σ. This is our GW measurement of the kicks imparted to BHs at formation. Quoting median and 90% confidence interval, we find $\sigma = 105^{+44}_{-29}$ km/s. The information gain between prior and posterior, as quantified by the Kullback-Leibler divergence [40], is $D_{KL} = 1.54$.

The posterior is skewed toward high values of σ. This is because the event rate r changes more (less) rapidly at low (high) values of σ (cf. Fig. 1). Consequently, data can more (less) easily accommodate kicks that are larger (smaller) than the inferred posterior maximum.

Our inference is largely driven by the integrated event rate $r(\lambda)$. For this set of simulations, the current GW catalog does not contain enough discriminating power to perform an informative analysis restricted to p_{pop}. Repeating our study using the marginalized likelihood of Eq. (4) returns a much lower information gain $D_{KL} = 0.24$.

IV. DISCUSSION

By applying the statistical framework of Ref. [15], we analyzed current GW data using population-synthesis simulations of binary BHs formed in isolation. Simulations enter the training process of a machine-learning algorithm, in this case GPR. The resulting emulator is then used to evaluate the likelihood in the context of a standard hierarchical Bayesian analysis. The present case study puts this idea into practice using a modest set of 7 training simulations, allowing us to showcase both prospects and pitfalls of this approach. Overall, the pipeline returns posterior distributions of the input flags one needs to initialize and run population-synthesis simulations. These are related to poorly understood mechanisms in the lives of massive stars, like BH natal kicks, which are here measured directly from GW data.

We first tested our approach on mock catalogs and show the presence of a systematic bias that exceeds statistical uncertainties when the number of observations is $\gtrsim 100$. The omission of some event parameters causes errors when modeling detector selection effects which in turn, propagate to the population inference. The test reported in Sec. III A suggests that this issue can be alleviated with a larger set of training simulations, which will allow using a larger set of event parameters while keeping the interpolation error under control. The case where some parameters need to be omitted reflects a common situation. In this study, information on mass ratio and spins are available but limited number of simulations prevented us from carrying out a more complete analysis. A more damaging scenario occurs when some variables affecting the GW signal are not modeled at all in the astrophysical simulations. Our results show the importance of developing astrophysical models where all the observables (BH spins, eccentricity, etc) are taken into account.

We also applied our procedure to BH binary data from LIGO/Virgo O1 and O2. Our results suggest that binary BHs were imparted moderate kicks at formation ($\sigma \gtrsim 70$ km/s). This is in tentative agreement with position and proper-motion measurements in X-ray binaries [41–43], as well as GW measurements of BH-binary spin misalignment [44,45]. Our findings are, however, in contrast with current supernova models which predict that BH natal kicks should be highly suppressed due to fallback material in the late stage of the explosion (e.g., [46]). The simulations used in this paper do not have a dedicated flag to tune the amount of fallback, which is instead controlled directly by the value of σ.

We stress that our result is model dependent. This is intentional: *we are interpreting GW events in light of a specific set of astrophysical assumptions.* Consequently, only those assumptions are put to test. Among the set of predictions explored here, current GW data prefer models where moderately large natal kicks are imparted onto BHs. It is natural to expect that a more complete set of training simulations might change this result qualitatively. Even assuming the same population-synthesis setup, a larger set of training simulations will allow (i) capturing degeneracies between different population parameters and (ii) efficiently interpolating across additional event parameters.

Prior assumptions are inevitably part of any statistical analysis. Instead of relying on parametrized distributions,

our approach makes use of state-of-the-art simulations in a data-driven fashion. We believe our approach presents promising avenues to infer astrophysical formation and evolutionary processes of GW sources, thus making a step forward toward the goal of GW astronomy.

ACKNOWLEDGMENTS

[1] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), Phys. Rev. X **9**, 031040 (2019).

[2] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, arXiv:1904.07214 [Phys. Rev. D (to be published)].

[3] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), Astrophys. J. **882**, L24 (2019).

[4] I. Mandel, W. M. Farr, A. Colonna, S. Stevenson, P. Tiňo, and J. Veitch, Mon. Not. R. Astron. Soc. **465**, 3254 (2017).

[5] J. R. Hurley, C. A. Tout, and O. R. Pols, Mon. Not. R. Astron. Soc. **329**, 897 (2002).

[6] S. Stevenson, A. Vigna-Gómez, I. Mandel, J. W. Barrett, C. J. Neijssel, D. Perkins, and S. E. de Mink, Nat. Commun. **8**, 14906 (2017).

[7] N. Giacobbo and M. Mapelli, Mon. Not. R. Astron. Soc. **480**, 2011 (2018).

[8] K. Belczynski, V. Kalogera, F. A. Rasio, R. E. Taam, A. Zezas, T. Bulik, T. J. Maccarone, and N. Ivanova, Astrophys. J. Suppl. Ser. **174**, 223 (2008).

[9] B. Pattabiraman, S. Umbreit, W.-k. Liao, A. Choudhary, V. Kalogera, G. Memik, and F. A. Rasio, Astrophys. J. Suppl. Ser. **204**, 15 (2013).

[10] M. Giersz, D. C. Heggie, J. R. Hurley, and A. Hypki, Mon. Not. R. Astron. Soc. **431**, 2184 (2013).

[11] M. Zevin, C. Pankow, C. L. Rodriguez, L. Sampson, E. Chase, V. Kalogera, and F. A. Rasio, Astrophys. J. **846**, 82 (2017).

[12] S. Stevenson, C. P. L. Berry, and I. Mandel, Mon. Not. R. Astron. Soc. **471**, 2801 (2017).

[13] D. Gerosa and E. Berti, Phys. Rev. D **95**, 124046 (2017).

[14] Y. Bouffanais, M. Mapelli, D. Gerosa, U. N. Di Carlo, N. Giacobbo, E. Berti, and V. Baibhav, arXiv:1905.11054.

[15] S. R. Taylor and D. Gerosa, Phys. Rev. D **98**, 083017 (2018).

[16] D. Gerosa, E. Berti, R. O'Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, Phys. Rev. D **98**, 084036 (2018).

[17] S. Vitale, D. Gerosa, C.-J. Haster, K. Chatziioannou, and A. Zimmerman, Phys. Rev. Lett. **119**, 251103 (2017).

[18] S. M. Gaebel, J. Veitch, T. Dent, and W. M. Farr, Mon. Not. R. Astron. Soc. **484**, 4008 (2019).

[19] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).

[20] L. S. Finn, Phys. Rev. D **53**, 2878 (1996).

[21] D. Gerosa, GWDET, https://github.com/dgerosa/gwdet, https://doi.org/10.5281/zenodo.889966 (2018).

[22] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).

[23] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), Astrophys. J. Lett. **833**, L1 (2016).

[24] T. J. Loredo, Am. Inst. Phys. Conf. Ser. **735**, 195 (2004).

[25] I. Mandel, W. M. Farr, and J. R. Gair, Mon. Not. R. Astron. Soc. **486**, 1086 (2019).

[26] E. Thrane and C. Talbot, Pub. Astron. Soc. Aust. **36**, e010 (2019).

[27] M. Fishbach, D. E. Holz, and W. M. Farr, Astrophys. J. Lett. **863**, L41 (2018).

[28] K. A. Postnov and L. R. Yungelson, Living Rev. Relativity **17**, 3 (2014).

[29] D. Gerosa, S. Ma, K. W. K. Wong, E. Berti, R. O'Shaughnessy, Y. Chen, and K. Belczynski, Phys. Rev. D **99**, 103004 (2019).

[30] D. Gerosa and M. Kesden, Phys. Rev. D **93**, 124066 (2016).

[31] D. Gerosa, SPOPS, https://github.com/dgerosa/spops, 10.5281/zenodo.2551791 (2019).

[32] K. Belczynski, A. Heger, W. Gladysz, A. J. Ruiter, S. Woosley, G. Wiktorowicz, H.-Y. Chen, T. Bulik, R. O'Shaughnessy, D. E. Holz, C. L. Fryer, and E. Berti, Astron. Astrophys. **594**, A97 (2016).

[33] G. Hobbs, D. R. Lorimer, A. G. Lyne, and M. Kramer, Mon. Not. R. Astron. Soc. **360**, 974 (2005).

[34] K. Belczynski, S. Repetto, D. E. Holz, R. O'Shaughnessy, T. Bulik, E. Berti, C. Fryer, and M. Dominik, Astrophys. J. **819**, 108 (2016).

[35] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), Living Rev. Relativity **21**, 3 (2018).

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

[37] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125**, 306 (2013).

[38] B. P. Abbott *et al.* (LIGO and Virgo Scientific Collaborations), Phys. Rev. X **6**, 041015 (2016); **8**, 039903 (2018).

[39] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, J. Phys. Conf. Ser. **610**, 012021 (2015).

[40] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).

[41] S. Repetto, A. P. Igoshev, and G. Nelemans, Mon. Not. R. Astron. Soc. **467**, 298 (2017).

[42] I. Mandel, Mon. Not. R. Astron. Soc. **456**, 578 (2016).

[43] P. Atri, J. C. A. Miller-Jones, A. Bahramian, R. M. Plotkin, P. G. Jonker, G. Nelemans, T. J. Maccarone, G. R. Sivakoff, A. T. Deller, S. Chaty, M. A. P. Torres, S. Horiuchi, J. McCallum, T. Natusch, C. J. Phillips, J. Stevens, and S. Weston, Mon. Not. R. Astron. Soc. **489**, 3116 (2019).

[44] R. O'Shaughnessy, D. Gerosa, and D. Wysocki, Phys. Rev. Lett. **119**, 011101 (2017).

[45] D. Wysocki, D. Gerosa, R. O'Shaughnessy, K. Belczynski, W. Gladysz, E. Berti, M. Kesden, and D. E. Holz, Phys. Rev. D **97**, 043014 (2018).

[46] C. L. Fryer, K. Belczynski, G. Wiktorowicz, M. Dominik, V. Kalogera, and D. E. Holz, Astrophys. J. **749**, 91 (2012).