Gravitational-wave selection effects using neural-network classifiers

Davide Gerosa, Geraint Pratten, and Alberto Vecchio

School of Physics and Astronomy & Institute for Gravitational Wave Astronomy, University of Birmingham, Birmingham, B15 2TT, United Kingdom

(Dated: July 15, 2020)

We present a novel machine-learning approach to estimate selection effects in gravitational-wave observations. Using techniques similar to those commonly employed in image classification and pattern recognition, we train a series of neural-network classifiers to predict the LIGO/Virgo detectability of gravitational-wave signals from compact-binary mergers. We include the effect of spin precession, higher-order modes, and multiple detectors and show that their omission, as it is common in large population studies, tends to overestimate the inferred merger rate in selected regions of the parameter space. Although here we train our classifiers using a simple signal-to-noise ratio threshold, our approach is ready to be used in conjunction with full pipeline injections, thus paving the way toward including actual distributions of astrophysical and noise triggers into gravitational-wave population analyses.

I. INTRODUCTION

Much like any other observatory, gravitational-wave (GW) interferometers suffer from selection biases. Ground-based detectors such as LIGO and Virgo are sensitive only to GW frequencies between ~10 and ~1000 Hz. While this is the regime of stellar-mass compact objects, the non-flat frequency response of the detectors implies that black-holes (BHs) with, say, 50\(M_\odot\) are easier to detect than those with, say, 5\(M_\odot\). BH binaries with spins that are co- (counter-) aligned with the orbital angular momentum present longer (shorter) signals and are thus easier (harder) to observe. Furthermore, GW detectors are more sensitive to signals coming from specific parts of the sky: a single interferometer is more likely to observe sources if they are located overhead/underneath the detector and presents blind spots in the plane of the arms. This feature is attenuated by a multi-detector network, which in general provides a more uniform sky coverage (although the the two LIGO instruments have essentially the same arms’ orientation and Virgo is currently still a factor ~2 less sensitive than LIGO). Finally, because GW emission is beamed along the direction of the binary’s orbital angular momentum, observations of face-on sources are far more likely compared to their edge-on counterparts.

As the size of the catalogue of observed GW events grows, an accurate characterization of selection effects is a key ingredient to obtain unbiased estimates of the population properties of merging compact objects. The most accurate way to estimate if a GW signal is observable is to inject a fake copy into the data, which are then processed with the same analysis pipeline employed in the actual search. Large campaigns with hundreds of thousands of injections are used to characterize the detection efficiency of analysis pipelines, which is a crucial input to infer the astrophysical merger rates of compact objects [1, 2]. Although accurate, this strategy is computationally very expensive and can be pursued only for a limited number of assumed source populations. However, the population properties and the merger rate must be estimated at the same time. Selection effects thus need to be computed at each evaluation of the population likelihood [3], which makes large injection campaigns prohibitive.

In practice, state-of-the-art GW population studies all include selection biases in some approximate form [4–6]. The most commonly used approach dates back to Refs. [7, 8] and relies on factoring out the dependencies of the extrinsic angles on the event signal-to-noise ratio (SNR). Events can then be thresholded based on the SNR of an “optimal” source with the same intrinsic parameters. A recent comparison between outputs of injection campaigns and SNR thresholding has been presented in Ref. [9]. Possible paths of improvement include considering additional calibration factors [10, 11] and time-varying sensitivity noise curves [5].

The semi-analytical treatment of Refs. [7, 8] is, strictly speaking, only valid when a single detector —instead of the full LIGO-Virgo, and soon KAGRA, network— is considered. Furthermore, the notion of an “optimal” source is not straightforward to define for precessing sources (because the orbital plane orientation changes as the binary is observed) and GW emission modes beyond the dominant quadrupole.

In this paper, we present a novel approach to estimate selection effects in GW observations. We tackle this problem with machine learning, borrowing techniques that are commonly applied to image-classification and pattern-recognition problems (see e.g. [12]). More specifically, we train a series of neural-network classifiers on the parameters of compact binaries, with the goal of predicting if their GW signals are detectable. Can a computer “learn” if LIGO and Virgo will observe GWs?
II. GRAVITATIONAL-WAVE DETECTABILITY

A compact binary on a quasi-circular orbit is characterized by source-frame masses \(m_1\) and \(m_2\) (combined into total mass \(M = m_1 + m_2\) and mass ratio \(q = m_2/m_1 \leq 1\), dimensionless spins \(\chi_1\) and \(\chi_2\), and redshift \(z\). Position and orientation with respect to the detectors are defined in terms of right ascension \(\alpha\), declination \(\delta\), orbital-plane inclination \(i\), and polarization angle \(\psi\). To simplify the notation, in the following we combine intrinsic and extrinsic parameters using the symbols \(\theta = \{M, q, \chi_1, \chi_2\}\) and \(\lambda = \{\alpha, \delta, i, \psi\}\).

The SNR \(\rho\) \cite{13, 14} is the most commonly employed metric to estimate the detectability of GW signals. For a single interferometer, the SNR \(\rho_S\) is defined as the waveform inner product weighted by the detector response. In Appendix A. In brief, one can factor out the dependency of \(\rho\) where \(T\) is the time length of the observing run and \(\det\) indicates the Heaviside step function. This expression is analytically \cite{7, 8}; the explicit calculation is reported in Eq. (1) can be computed semi-analytically \cite{7, 8}; the explicit calculation is reported in Appendix A. In brief, one can factor out the dependency of \(\lambda\) on the SNR to obtain \(\rho_S(\theta, z, \lambda) = \omega(\lambda)\rho_{S, \text{opt}}(\theta, z)\) \cite{14}, where \(\rho_{S, \text{opt}}\) is the SNR of an “optimal” source located overhead the detector with face-on inclination. Because the projection function \(0 \leq \omega(\lambda) \leq 1\) is universal, one can easily convert any given probability distribution function \(p(\lambda)\) into \(p(\omega)\). Computing the marginalized distribution \(p_{\text{det}}(\theta, z)\) thus reduces to evaluating the cumulative distribution function \(P(\omega) = \int_{\omega'}^{1} p(\omega')d\omega'\) at \(\omega = \rho_{\text{thr}}/\rho_{S, \text{opt}}\) with \(\rho_{\text{thr}} = 8\). The blue circles mark the corresponding estimate obtained by numerically marginalizing over the prediction of our non-spinning neural network evaluated on the validation set. We use \(N_{\text{MC}} = 10^4\) Monte Carlo samples of \(p(\lambda)\). A long tail at large values of \(\omega\) is omitted for clarity.

III. MACHINE-LEARNING IMPLEMENTATION

This semi-analytical treatment breaks down in the more general case where a multi-detector network is considered (cf. e.g. Ref. \cite{20}). Furthermore, the concept of an optimally oriented source is not directly applicable if one considers precessing binaries and GW harmonics beyond the dominant mode.

We tackle this issue using machine learning. We implement a neural-network classifier using Google’s TensorFlow infrastructure \cite{21}. We use a sequential Keras model with three layers: an input layer with either 7 or 13 neurons (set by the dimensionality of input parameter.
space, see below), a single hidden layer with 32 neurons, and a final outer layer which classifies the source as either “detectable” or “not detectable”. The hidden (outer) layer makes use of a hyperbolic tangent (sigmoid) activation function. Inputs are preprocessed with an affine transformation and rescaled within $[-1, 1]$ in each dimension. Neuron weights are initialized using the Glorot algorithm [22]. Optimization is performed using the Adam optimizer [23] with an initial learning rate of $10^{-2}$ and the binary-cross entropy loss function [12]. The network is exposed to the training data in batches of size 32 for up to 150 epochs. After 10 training epochs, the learning rate is decreased exponentially. In general, neural-network performances increase with the size of the training sample, but so does the computational cost of the training. Results presented in this paper are based on neural networks that have been trained and tested on two independent samples of $N = 10^7$ sources each. We systematically explored a wide variety of architectures and the setup we just described has been found to maximize the validation accuracy at a reasonable computational cost. Training each of the neural networks presented in this paper took $\sim 30$ hours on a single off-the-shelf CPU.

We present results obtained with three classifiers. In all cases, training and validation sets are generated distributing detector-frame total masses $M_z = M(1+z)$ uniformly in $[2M_\odot, 1000M_\odot]$, mass ratios $q$ uniformly in $[0.1, 1]$, redshifts uniformly in $[10^{-4}, 4]$, and assuming isotropic orientations $i$, sky locations $\alpha, \delta$, and polarization angles $\psi$. The largest value of $z$ has been chosen to marginally exceed the horizon redshift of all the sources in the sample. We stress that this distribution does not need to represent a plausible astrophysical scenario but only allow for accurate training. Once trained, the network can then be evaluated on the chosen population.

To compare our findings against analytic estimates of $p_{\text{det}}$, we first develop a simpler network assuming non-spinning sources ($\chi_1 = \chi_2 = 0$), considering only the dominant $(\ell, |m|) = (2, 2)$ mode, and using the single-detector condition $\rho_S > 8$. Stepping up in complexity, we then include higher order modes $(\ell, |m|) = (2, 2), (2, 1), (3, 3), (3, 2), (4, 4)$ and precessing sources with spins distributed uniformly in magnitudes in $[0, 1]$ and isotropically in directions. In this case, we train networks using both conditions $\rho_S > 8$ and $\rho_N > 12$.

SNRs are computed with PyCBC [24] assuming the IMRPhenomXPHM [25] waveform model and the Planck 15 cosmology [26]. We consider a three-detector network made of LIGO Hanford, LIGO Livingston, and Virgo at their nominal design sensitivity [27]. When referring to single-detector SNRs $\rho_S$, we assume a single LIGO instrument. Other neural networks trained using sensitivity curves of LIGO/Virgo during their observing runs O1, O2, O3, and O4 are provided at Ref. [28].

Figure 2 shows the evolution of the neural-network accuracies as a function of the training epoch. In this context, accuracy is simply defined as the fraction of the inputs which are correctly identified as either “detectable” or “not detectable”. Among all iterations, we select the ones that maximize the validation accuracy. The final values of accuracies and losses are reported in Table I.

As expected, the network trained on non-spinning sources behaves better because it has to interpolate across a smaller 7-dimensional parameter space. When considering the full 13-dimensional parameter space of precessing binaries, we find that the condition $\rho_N > 12$ ($\rho_S > 8$) is easier (harder) to classify. This is because the beam pattern of a combination of instruments is smoother compared to that of a single interferometer [29]. Although more instruments contribute to the SNR, it is worth noting that $\rho_N > 12$ is a potentially stricter criterion than $\rho_S > 8$ with the expected distribution scaling as $\rho^{-4}$ [17, 29]. All three neural networks show very similar performances on validation and training sets, indicating that we are not overfitting the data.

We stress that the loss and accuracy values depend on the population one is trying to predict. The distributions used in the training/testing process were deliberately chosen to be “challenging” to classify: we overpopulate regions of the parameter space at high SNR (low $z$ and high $M$) to encourage the network to better learn the various correlations between the input parameters for loud

![FIG. 2. Neural-network performances during the training process. Solid (dashed) lines indicate the accuracy evaluated over the validation (training) dataset. Colors indicate three different neural networks trained considering: non-spinning binaries, dominant emission mode, and single detector (blue); precessing binaries, higher-order modes, and single detector (orange); precessing binaries, higher-order modes, and three detectors (green).](image-url)
TABLE I. Accuracies and losses of the final trained models. We report performances evaluated on the training and validation sets, which are generated with binaries distributed as specified in Sec. III. We also evaluate the classifiers on two astrophysically motivated populations (“powerlaw” and “log-uniform”) from Ref. [1, 2], see Sec. IV for details. All samples in this table have size $N = 10^7$.

| Source Description | Training | Validation | “powerlaw” | “log-uniform” |
|--------------------|----------|------------|------------|--------------|
| | Accuracy | Loss | Accuracy | Loss | Accuracy | Loss | Accuracy | Loss |
| Non-spinning, quadrupole only, $\rho_S > 8$ | 0.9835 | 0.0381 | 0.9836 | 0.0381 | – | – | 0.9917 | 0.0194 |
| Precessing, higher harmonics, $\rho_S > 8$ | 0.9782 | 0.0509 | 0.9781 | 0.0511 | 0.9978 | 0.0053 | – | – |
| Precessing, higher harmonics, $\rho_N > 12$ | 0.9817 | 0.0434 | 0.9817 | 0.0917 | 0.9979 | 0.0049 | 0.9920 | 0.0190 |

Not surprisingly, the events that the network cannot identify correctly all have $\rho \sim \rho_{\text{hr}}$. There is roughly an equal number of detectable sources which are classified as “not detectable” and vice versa, which suggests that the impact of this mismodeling will be further mitigated when integrating to compute $p_{\text{det}}$ and $VT$.

IV. GRAVITATIONAL-WAVE POPULATIONS

We now compare our results against the analytic estimate of $p_{\text{det}}$ described in Sec. II and Appendix A. Once a classifier has been trained, the detectability $p_{\text{det}}(\theta, z)$ can be estimated by repeatedly sampling $p(\lambda)$ and counting the number of draws which are identified as “detectable”. Figure 1 shows such comparison for the non-spinning, $(2, 2)$-mode neural network trained on the condition $\rho_S > 8$ and applied to (a fraction of) the test sample. In this case, all the assumptions made in Sec. II to analytically average $p_{\text{det}}$ hold and the two approaches should give the same result. The blue histogram in Fig. 3 shows the deviation $\Delta p_{\text{det}}$ between numerical and semi-analytical evaluations. Our neural network reproduces the LIGO detectability both qualitatively and quantitatively. We report minor systematic deviations at large SNRs (small value of $\omega$), where the neural network tends to slightly overestimate the effective spacetime volume (top-left region in Fig. 1). More suitable network infrastructures and/or training strategies might suppress this spurious feature.

Now that we have established that our approach is accurate, we explore whether determining $p_{\text{det}}$ using a single-detector SNR threshold for non-precessing binaries and including only the leading order $(2, 2)$ mode —strategy adopted in the overwhelming majority of population studies— describes the detection probability correctly across the source parameter space in observations performed with multiple instruments. We therefore apply our model to cases where the assumptions behind the semi-analytical estimate of $p_{\text{det}}$ are not valid. We evaluate the difference $\Delta p_{\text{det}}$ between the network prediction of $p_{\text{det}}$ and the corresponding estimate if one were to naively apply the semi-analytic approach. More precisely, we compute $\omega = 8/\rho_{S, \text{opt}}$ assuming overhead sources with the orbital angular momentum aligned with the line of sight at the GW emission frequency $f_{\text{ref}} = 20$ Hz. The left (right) region of Fig. 3 with $\Delta p_{\text{det}} < 0$ ($\Delta p_{\text{det}} > 0$) correspond to cases where the simplistic approach overestimates (underestimates) the detector performance compared to the prediction of our neural networks.

The orange histogram in Fig. 3 is computed assuming...
the same distribution used in the training and validation process, but considering precessing sources, additional harmonics beyond the dominant emission mode, and the three-detector LIGO/Virgo network. Although the vast majority of the sources are correctly identified ($\Delta p_{\text{det}} \simeq 0$), the shoulder extending towards positive values of $\Delta p_{\text{det}}$ indicates that the analytical single-detector approximation tends to, on average, underestimates the LIGO/Virgo detectability. Our models are publicly available at 
\begin{center}
\url{github.com/dgerosa/pdetclassifier}
\end{center}
where we provide trained networks and training/validation samples for the three cases described in this paper, as well as additional outputs calibrated on the LIGO/Virgo detectability. Our models are publicly available at 
\begin{center}
\url{github.com/dgerosa/pdetclassifier}
\end{center}
including spins, higher-order modes, and multi-detector SNR calculations results in higher detection probability, suggesting that current detection-rate estimates that do not rely on actual systematic injection studies might be slightly overestimated. Such mismodeling appears to be more relevant in the corner of the parameter space characterized by high and unequal masses.

Compared to raw evaluations of the SNR, our approach allows for a computational speed-up of more than a factor 1000 (evaluated on a standard off-the-shelf laptop). If a similar number of Monte Carlo samples is used to marginalize over $\lambda$, this implies that our tool (in its present form) might allow the full inclusions of spins, higher-order modes, and network SNRs in VT estimates at the same computational cost of the currently employed semi-analytic average.

We hope our work might be useful to both GW data analysts and astrophysicists. Researches developing codes of binary-star and cluster evolution to predict GW events will be able to use our neural network to efficiently filter their synthetic catalogs based on the LIGO/Virgo detectability. Our models are publicly available at 
\begin{center}
\url{github.com/dgerosa/pdetclassifier}
\end{center}
where we provide trained networks and training/validation samples for the three cases described in this paper, as well as additional outputs calibrated on the LIGO/Virgo performance during their observing runs O1, O2, O3, and forecasted O4. We hope this will facilitate the immediate adoption of our approach.

Our neural-network classifiers are trained to learn a single yes/no variable (“is this input binary detectable?”). An alternative strategy to estimate GW selection biases...
using machine learning would be to implement a regression neural network to interpolate the SNR [32]. In this case, the detectability condition \( \rho > \rho_{\text{thr}} \) can be evaluated at runtime using the predicted values of \( \rho \). A classifier like ours has the obvious drawback that a new model needs to be trained for each choice of the SNR threshold \( \rho_{\text{thr}} \) (but note that the vast majority of the computational cost lies in building the training sample and does not need to be repeated). The advantage is that our infrastructure can be trained on detectability conditions that are more accurate than a simple SNR threshold.

Our approach will show its full potential when used in conjunction with large software injections campaigns, like those presented in Ref. [1, 2] to estimate detection rates. Although undoubtedly more accurate at modeling selection biases, such estimates are not currently used in most GW population studies because of their high computational cost. We believe a machine-learning treatment like the one explored in this paper is a promising way forward. We recommend that analyses reporting GW events should make public the estimate of the detection probability across the parameter space covered by the search, to be used down-stream in population studies through, e.g., an approach like ours.

ACKNOWLEDGMENTS

We thank Riccardo Buscicchio, Kaze Wong, Patricia Schmidt, and Emanuele Berti for discussions. D.G. is supported by Leverhulme Trust Grant No. RPG-2019-350. A.V. acknowledges the support of the Royal Society and Wolfson Foundation, and the UK Science and Technology Facilities Council through Grant No. ST/N021702/1. Computational work was performed on the University of Birmingham BlueBEAR cluster, the Athena cluster at HPC Midlands+ funded by EPSRC Grant No. EP/P020232/1 and the Maryland Advanced Research Computing Center (MARCC).

Appendix A: Semi-analytic estimate of \( p_{\text{det}} \)

The integral in Eq. (1) can be carried out semi-analytically for the case a single detector, non-precessing sources, and considering only the dominant quadrupole emission mode.

In this simpler case, let us define the sky location of the source with a polar angle \( \vartheta \) and an azimuthal angle \( \phi \). This is equivalent to \( \delta = \pi/2 - \vartheta \) and \( \phi = \alpha \) for a detector located at the north pole with one arm directed towards the vernal equinox. Given an incoming GW with polarizations \( h_+ \) and \( h_\times \), an interferometer with 90°-arms is sensitive to the combination

\[
h(t) = F_+ h_+(t) + F_\times h_\times(t), \tag{A1}
\]

where the beam patterns are given by [14]

\[
F_+ = \frac{1}{2} (1 + \cos^2 \vartheta) \cos 2\phi \cos 2\psi - \cos \vartheta \sin 2\phi \sin 2\psi,
\]

\[
F_\times = \frac{1}{2} (1 + \cos^2 \vartheta) \cos 2\phi \sin 2\psi + \cos \vartheta \sin 2\phi \cos 2\psi.
\]

If spin precession and higher harmonics are neglected, the GW emission of binary coalescence is given by

\[
h_+(t) = A(t) \frac{1 + \cos^2 \vartheta}{2} \cos \Phi(t), \tag{A4}
\]

\[
h_\times(t) = A(t) \cos \vartheta \sin \Phi(t), \tag{A5}
\]

where amplitude \( A(t) \) and phase \( \Phi(t) \) are predicted by the chosen waveform approximant. One can rewrite Eq. (A1) as

\[
h(t) = \omega A(t) \cos[\Phi(t) - \Phi_0]. \tag{A6}
\]

where \([7, 8]\)

\[
\omega = \sqrt{\left(\frac{F_+}{2} + \frac{1 + \cos^2 \vartheta}{2}\right)^2 + (F_\times \cos \vartheta)^2}, \tag{A7}
\]

\[
\tan \Phi_0 = \frac{2F_\times \cos \vartheta}{F_+ (1 + \cos^2 \vartheta)}. \tag{A8}
\]

It is straightforward to show that \( \max_{\vartheta, \phi, \psi} \omega = 1 \). This maximum is obtained when \( \vartheta = 0 \) (i.e. the source is face-on) and \( \vartheta = 0 \) (i.e. the source is overhead the detector). Let us call such a source “optimal”. The factorization of Eq. (A6) implies that the SNR of generic binary \( \rho \) can be written as \( \rho = \rho_{\text{opt}} \). One has \( p(\lambda) d\lambda = p(\omega) d\omega \) and thus, from Eq. (1),

\[
p_{\text{det}} = \int p(\omega) \Theta[\omega \rho_{\text{opt}} - \rho_{\text{thr}}] d\omega
\]

\[
= \int_{\omega \rho_{\text{opt}} > \rho_{\text{thr}}} p(\omega) d\omega.
\]

Let us note that Refs. [7, 8] make use of the equivalent notation \( \Theta \equiv 4\omega \).

For the case of isotropic sources, an analytic fit to the integral in Eq. (A9) is provided in Ref. [20]; a public Monte Carlo implementation is available at Ref. [33].
[1] B. P. Abbott et al. (LIGO and Virgo Collaboration), PRX 6, 041015 (2016), arXiv:1606.04856 [gr-qc].
[2] B. P. Abbott et al. (LIGO and Virgo Collaboration), PRX 9, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].
[3] I. Mandel, W. M. Farr, and J. R. Gair, MNRAS 486, 1086 (2019), arXiv:1809.02063 [physics.data-an].
[4] S. R. Taylor and D. Gerosa, PRD 98, 083017 (2018), arXiv:1806.08365 [astro-ph.HE].
[5] B. P. Abbott et al. (LIGO and Virgo Collaboration), ApJ 882, L24 (2019), arXiv:1811.12940 [astro-ph.HE].
[6] J. Roulet and M. Zaldarriaga, MNRAS 484, 4216 (2019), arXiv:1806.10610 [astro-ph.HE].
[7] L. S. Finn and D. F. Chernoff, PRD 47, 2198 (1993), arXiv:gr-qc/9301003 [gr-qc].
[8] L. S. Finn, PRD 53, 2878 (1996), arXiv:gr-qc/9601048 [gr-qc].
[9] B. P. Abbott et al. (LIGO and Virgo Collaboration), ApJS 227, 14 (2016), arXiv:1606.03939 [astro-ph.HE].
[10] I. Mandel, W. M. Farr, and J. R. Gair, MNRAS 486, 1086 (2019), arXiv:1809.02063 [physics.data-an].
[11] D. Wysocki and R. O'Shaughnessy, LIGO Document T1800427 (2018).
[12] I. J. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (MIT Press, 2016).
[13] K. S. Thorne, “Gravitational radiation,” in Three Hundred Years of Gravitation, edited by S. W. Hawking and W. Israel (1989) p. 330.
[14] B. S. Sathyaprakash and B. F. Schutz, LRR 12, 2 (2009), arXiv:0903.0338 [gr-qc].
[15] A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst, and D. A. Brown, ApJ 849, 118 (2017), arXiv:1705.01513 [gr-qc].
[16] C. Messick et al., PRD 95, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].
[17] H.-Y. Chen and D. E. Holz, (2014), arXiv:1409.0522 [gr-qc].
[18] H.-Y. Chen, D. E. Holz, J. Miller, M. Evans, S. Vitale, and J. Creighton, (2017), arXiv:1709.08079 [astro-ph.CO].
[19] R. O’Shaughnessy, V. Kalogera, and K. Belczynski, ApJ 716, 615 (2010), arXiv:0908.3635 [astro-ph.CO].
[20] M. Dominik, E. Berti, R. O’Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. E. Holz, T. Bulik, and F. Pannarale, ApJ 806, 263 (2015), arXiv:1405.7016 [astro-ph.HE].
[21] M. Abadi et al., in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (2016) pp. 265–283.
[22] X. Glorot and Y. Bengio, in Proceedings of the thirteenth international conference on artificial intelligence and statistics (2010) pp. 249–256.
[23] D. P. Kingma and J. Ba, in 3rd International Conference for Learning Representations (2014) arXiv:1412.6980 [cs.LG].
[24] S. A. Usman et al., CQG 33, 215004 (2016), arXiv:1508.02357 [gr-qc].
[25] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellès, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, (2020), arXiv:2004.06503 [gr-qc].
[26] P. A. R. Ade et al. (Planck Collaboration), A&A 594, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
[27] B. P. Abbott et al. (LIGO and Virgo Collaboration), LRR 21, 3 (2018), arXiv:1304.0670 [gr-qc].
[28] D. Gerosa, github.com/dgerosa/pdetclassifier, doi.org/10.5281/zenodo.3941495 (2020).
[29] B. F. Schutz, CQG 28, 125023 (2011), arXiv:1102.5421 [astro-ph.IM].
[30] K. K. Y. Ng, S. Vitale, A. Zimmerman, K. Chatziioannou, D. Gerosa, and C.-J. Haster, PRD 98, 083007 (2018), arXiv:1805.03046 [gr-qc].
[31] D. Gerosa, E. Berti, R. O’Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, PRD 98, 084036 (2018), arXiv:1808.02491 [astro-ph.HE].
[32] K. W. K. Wong, K. K. Y. Ng, and E. Berti, in preparation.
[33] D. Gerosa, github.com/dgerosa/gwdet, doi.org/10.5281/zenodo.889965 (2018).