

Classifying binary black holes from Population III stars with the *Einstein* Telescope: A machine-learning approach

Filippo Santoliquido^{1,2,*}, Ulyana Dupletsa^{1,2}, Jacopo Tissino^{1,2}, Marica Branchesi^{1,2}, Francesco Iacovelli^{3,4},
Giuliano Iorio^{5,6}, Michela Mapelli^{7,5,6}, Davide Gerosa^{8,9,10}, Jan Harms^{1,2}, and Mario Pasquato^{5,6,11,12,13}

¹ Gran Sasso Science Institute (GSSI), 67100 L'Aquila, Italy

² INFN, Laboratori Nazionali del Gran Sasso, 67100 Assergi, Italy

³ Département de Physique Théorique, Université de Genève, 24 quai Ernest Ansermet, 1211 Genève, Switzerland

⁴ Gravitational Wave Science Center (GWSC), Université de Genève, 1211 Genève, Switzerland

⁵ Dipartimento di Fisica e Astronomia “G. Galilei”, Università degli studi di Padova, Vicolo dell’Osservatorio 3, 35122 Padova, Italy

⁶ INFN, Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy

⁷ Institut für Theoretische Astrophysik, ZAH, Universität Heidelberg, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany

⁸ Dipartimento di Fisica “G. Occhialini”, Università degli studi di Milano-Bicocca, piazza della Scienza 3, 20126 Milano, Italy

⁹ INFN, Sezione di Milano-Bicocca, piazza della Scienza 3, 20126 Milano, Italy

¹⁰ School of Physics and Astronomy & Institute for Gravitational Wave Astronomy, University of Birmingham, Birmingham B15 2TT, United Kingdom

¹¹ Département de Physique, Université de Montréal, 1375 Avenue Thérèse-Lavoie-Roux, Montréal, Canada

¹² Mila – Quebec Artificial Intelligence Institute, 6666 Rue Saint-Urbain, Montréal, Canada

¹³ Ciela – Montréal Institute for Astrophysical Data Analysis and Machine Learning, Montréal, Canada

Received 15 April 2024 / Accepted 21 August 2024

ABSTRACT

Third-generation (3G) gravitational-wave detectors such as the *Einstein* Telescope (ET) will observe binary black hole (BBH) mergers at redshifts up to $z \sim 100$. However, an unequivocal determination of the origin of high-redshift sources will remain uncertain because of the low signal-to-noise ratio (S/N) and poor estimate of their luminosity distance. This study proposes a machine-learning approach to infer the origins of high-redshift BBHs. We specifically differentiate those arising from Population III (Pop. III) stars, which probably are the first progenitors of star-born BBH mergers in the Universe, and those originated from Population I-II (Pop. I-II) stars. We considered a wide range of models that encompass the current uncertainties on Pop. III BBH mergers. We then estimated the parameter errors of the detected sources with ET using the Fisher information-matrix formalism, followed by a classification using XGBOOST, which is a machine-learning algorithm based on decision trees. For a set of mock observed BBHs, we provide the probability that they belong to the Pop. III class while considering the parameter errors of each source. In our fiducial model, we accurately identify $\geq 10\%$ of the detected BBHs that originate from Pop. III stars with a precision $>90\%$. Our study demonstrates that machine-learning enables us to achieve some pivotal aspects of the ET science case by exploring the origin of individual high-redshift GW observations. We set the basis for further studies, which will integrate additional simulated populations and account for further uncertainties in the population modeling.

Key words. black hole physics – gravitational waves – methods: numerical – methods: statistical – stars: Population III

1. Introduction

The *Einstein* Telescope (ET) will be a forefront instrument for observing gravitational waves (GWs) in Europe (Punturo et al. 2010; Maggiore et al. 2020). In particular, the ET will observe mergers of binary black holes (BBHs) at redshifts as high as $z \sim 100$ (Ng et al. 2021; Kalogera et al. 2021; Ng et al. 2022b; Yi et al. 2022; Branchesi et al. 2023). However, the detection of high-redshift events does not automatically imply an accurate inference of their parameters. Several works have demonstrated the challenge the ET will face in constraining the redshift distribution of the detected sources, which in turn is a crucial step to unveil their astrophysical origins (e.g. Ng et al. 2022a; Chen et al. 2022; Iacovelli et al. 2022a; Mancarella et al. 2023; Fairhurst et al. 2024; Marcoccia et al. 2024).

To this end, Ng et al. (2022a, 2023) investigated the accuracy of redshift measurements for BBH mergers. They specifically sought to determine whether their primordial origin could be established. Their findings indicate that for BBHs with total masses ranging between $20 M_{\odot}$ and $40 M_{\odot}$ that merge at $z > 40$, it is possible to infer that $z > 30$ with a credibility of up to 70% using a single ET observatory. Mancarella et al. (2023) studied a population of BBHs with parameter distributions extrapolated from those of current LIGO-Virgo-KAGRA events (Abbott et al. 2023b) and found that only $\sim 3\%$ of the sources with $z \geq 30$ can be reliably constrained to be beyond redshift 30 at a credibility of 99%. These findings limit the scientific potential of the ET and Cosmic Explorer (Reitze et al. 2019) because the detection of BBHs in the early Universe is a pivotal element of the science objectives for third-generation (3G) GW interferometers.

Population III (Pop. III) stars form from pristine metal-free gas follow cosmic nucleosynthesis at $z > 20$ (Haiman et al.

* Corresponding author; filippo.santoliquido@gssi.it

1996; Tegmark et al. 1997; Abel et al. 2002; Yoshida et al. 2003; Klessen & Glover 2023) and are thought to lead to the formation of the first star-origins black holes in the Universe (Kinugawa et al. 2014, 2016; Hartwig et al. 2016; Belczynski et al. 2017; Tanikawa et al. 2022a; Costa et al. 2023; Nandal et al. 2023; Iwaya et al. 2023; Tanikawa 2024; Liu et al. 2024a,b). There have been no direct observations of Pop. III stars so far (Rydberg et al. 2013; Schauer et al. 2022; Larkin et al. 2023; Meena et al. 2023; Trussler et al. 2023), and they can directly be detected with the *James Webb* Space Telescope (JWST) only with gravitational lensing (Zackrisson et al. 2024; Bovill et al. 2024; Wiggins et al. 2024). In recent years, BBHs originating from Pop. III stars have gained significant attention in the context of GW astronomy (Kinugawa et al. 2020; Liu & Bromm 2020a,b; Tanikawa et al. 2022a,b; Wang et al. 2022). In particular, Santoliquido et al. (2023) showed that between $\sim 20\%$ and $\sim 70\%$ of the detectable mergers from Pop. III BBHs occur at $z > 8$. These percentages depend on assumptions regarding the star formation rate density and the adopted initial conditions, resulting in distinct redshift distributions.

We present a novel method for distinguishing the origins of GW sources. Specifically, we differentiate between BBHs originating from Pop. III stars and those from Pop. I–II stars, which encompass sources that merge across a broad redshift range. It can be highly informative to determine the origin of single GW events, especially regarding the specific physical processes leading to their formation (Abbott et al. 2019b; Fishbach & Holz 2020; Kimball et al. 2020; Singh et al. 2022; Tong et al. 2022; Gamba et al. 2023; Ng et al. 2023). Two distinct methods have been proposed in the literature to achieve this. The first method involves a classification in which categorical variables are inferred through a hierarchical Bayesian analysis that considers hyperparameters describing each subpopulation (Farr et al. 2015; Mould et al. 2022, 2023; Godfrey et al. 2023). The second method is based on machine-learning (ML), where ML algorithms are trained on simulated sources and are subsequently applied to GW detections (Antonelli et al. 2023). These methods have both been directly applied to GW events published in LIGO–Virgo–KAGRA catalogs (Abbott et al. 2016, 2019a, 2021, 2023b).

In our analysis, we combine the outcomes of ML classifiers trained on Pop. III and Pop. I–II BBHs with the inferred posterior distribution of detected sources and take the parameter estimation capabilities of the ET into account. Consequently, we assign a probability for each source to be associated with either origin. We assess the performance of our classification and determine how the data inform the classification process. Our investigation covers systematic variations within various models describing Pop. III BBHs, including astrophysical merger rates.

With our proposed method, we aim to demonstrate that it is possible to infer the origin of individual events, considering the limited but still significant capability of the ET to perform the parameter estimation of high-redshift sources (Ng et al. 2022a, 2023; Mancarella et al. 2023). Our method combines information from astrophysical simulations with the results of the parameter estimation to achieve this goal. It is complementary to currently adopted methods that are based on Bayesian analysis (e.g., Thrane & Talbot 2019; Mould et al. 2023). The focus of our method on individual GW observations potentially enables us to identify rare events such as the most distant Pop. III BBHs, which are particularly interesting for constraining our understanding of the early Universe.

Machine-learning, deep learning, and artificial intelligence have become increasingly critical in astronomy and astrophysics (e.g., Ivezić et al. 2014; Baron 2019; Djorgovski et al. 2022; Moriwaki et al. 2023; Pasquato et al. 2024; Riggi et al. 2024), with a growing emphasis on applications within the GW field (Zevin et al. 2017; Cuoco et al. 2021a,b). In this context, a significant focus of ML techniques involves classification tasks (e.g., Baker et al. 2015; Chatterjee et al. 2020; Sasaoka et al. 2023; Alhassan et al. 2023; Antonelli et al. 2023; Berbel et al. 2024). Furthermore, the large number of detections expected with 3G GW observatories (Reitze et al. 2019; Maggiore et al. 2020; Branchesi et al. 2023) will make a full Bayesian parameter estimation computationally expensive (Couvares et al. 2021; Roulet & Venumadhav 2024). Hence, ML and deep learning hold particular promise for a rapid inference of GW parameters (Green et al. 2020; Dax et al. 2021, 2023; Williams et al. 2021; Gabbard et al. 2022; Wong et al. 2023; Wouters et al. 2024).

This paper is organized as follows: We delineate in Sect. 2.1 the simulation setup for generating BBHs from Pop. III and Pop. I–II stars. In Sect. 2.2 we evaluate the merger rate density as a function of redshift, while Sect. 2.3 provides a brief overview of the Fisher information-matrix (FIM) approximation we used to estimate the parameters. Section 2.4 details the method for assigning the probability of each source to belong to the Pop. III class, while Sect. 2.5 delineates the training process for the ML classifiers. Our results along with a discussion of our method caveats are presented in Sects. 3 and 4. Conclusion are drawn in Sect. 5.

2. Methods

2.1. Astrophysical populations

Santoliquido et al. (2023) presented a comprehensive investigation on the main sources of uncertainty affecting BBHs born from Pop. III stars. Here, we focus on three of their models: their fiducial model, along with the two models yielding the lowest and highest comoving merger rate density of those they calculated.

2.1.1. Population III stars

The catalogs of BBH mergers reported by Santoliquido et al. (2023) were generated using the binary population synthesis code SEVN (Spera et al. 2019; Mapelli et al. 2020) for both Pop. III (Costa et al. 2023) and Pop. I–II stars (Iorio et al. 2023). SEVN combines single and binary evolution by interpolating a set of precomputed single stellar-evolution tracks (Iorio et al. 2023). Pop. III stellar tracks were computed using the PARSEC code (Bressan et al. 2012; Costa et al. 2021; Nguyen et al. 2022) with a metallicity of $Z = 10^{-11}$ and covering a zero-age main-sequence (ZAMS) mass range of 2.2–600 M_{\odot} (for details, see Costa et al. 2023).

Santoliquido et al. (2023) explored a large set of initial conditions for Pop. III binary systems. The three models analyzed here correspond to the setups labeled LOG1, LAR1, and TOP5. The initial mass function (IMF) is defined as a flat-in-log probability distribution for model LOG1 (Stacy & Bromm 2013; Susa et al. 2014; Hirano et al. 2015; Wollenberg et al. 2020; Chon et al. 2021; Tanikawa et al. 2021; Jaura et al. 2022; Prole et al. 2022),

$$\xi(M_{\text{ZAMS},1}) \propto M_{\text{ZAMS},1}^{-1}, \quad (1)$$

Table 1. Astrophysical populations.

Name	IC	$M_{\text{ZAMS},1}$	q	P	e	SFRD	N
Pop. III fiducial	LOG1	Eq. (1)	Eq. (4)	Eq. (5)	Eq. (6)	H22	6916
Pop. III optimistic	LAR1	Eq. (2)	Eq. (4)	Eq. (5)	Eq. (6)	J19	119399
Pop. III pessimistic	TOP5	Eq. (3)	Eq. (7)	Eq. (8)	Eq. (9)	SW20	202
Pop. I-II	Fiducial	Eq. (10)	Eq. (4)	Eq. (5)	Eq. (6)	MF17	887364

Notes. For each model, we list the name of the initial condition configuration (IC), the distribution of the ZAMS masses of the primary star $M_{\text{ZAMS},1}$, the distribution of the mass ratios q , the distribution of the orbital periods P , the distribution of the orbital eccentricities e , and the adopted SFRD model. The rightmost column reports the total number of mergers from Eq. (16) assuming $T_{\text{obs}} = 10$ yr.

a Larson (1998) distribution for model LAR1,

$$\xi(M_{\text{ZAMS},1}) \propto M_{\text{ZAMS},1}^{-2.35} e^{M_{\text{cut}}/M_{\text{ZAMS},1}} \quad (2)$$

with $M_{\text{cut}} = 20 M_{\odot}$ (Valiante et al. 2016), and a top-heavy distribution for model TOP5 (Stacy & Bromm 2013; Jaacks et al. 2019; Liu & Bromm 2020a),

$$\xi(M_{\text{ZAMS},1}) \propto M_{\text{ZAMS},1}^{-0.17} e^{(M_{\text{cut}}/M_{\text{ZAMS},1})^2}, \quad (3)$$

with $M_{\text{cut}} = 20 M_{\odot}$.

For the LOG1 and LAR1 models, the mass ratio ($q = M_{\text{ZAMS},2}/M_{\text{ZAMS},1}$), orbital period (P), and eccentricity (e) were drawn from the distributions by Sana et al. (2012). These distributions are fits to O- and B-type binary stars in the local Universe,

$$\xi(q) \propto q^{-0.1} \text{ with } q \in [0.1, 1] \text{ and } M_{\text{ZAMS},2} > 2.2 M_{\odot}, \quad (4)$$

$$\xi(\log P) \propto (\log P)^{-0.55} \text{ with } \log P \in [0.15, 5.5], \quad (5)$$

$$\xi(e) \propto e^{-0.142} \text{ with } e \in (0, 1). \quad (6)$$

Model TOP5 instead adopts the mass ratio and the eccentricity distributions derived from cosmological simulations by Stacy & Bromm (2013),

$$\xi(q) \propto q^{-0.55} \text{ with } q \in [0.1, 1] \text{ and } M_{\text{ZAMS},2} > 2.2 M_{\odot}, \quad (7)$$

$$\xi(\log P) \propto \exp[-(\log P - \mu)^2 / (2\sigma^2)], \quad (8)$$

with $\mu = 5.5$ and $\sigma = 0.85$, and a thermal distribution for the eccentricity (Kinugawa et al. 2014; Hartwig et al. 2016; Tanikawa et al. 2021),

$$\xi(e) \propto 2e \text{ with } e \in [0, 1). \quad (9)$$

We considered three independent estimates of the Pop. III star formation rate density (SFRD). These were based on the models by Hartwig et al. (2022, hereafter H22), Jaacks et al. (2019, hereafter J19), and Skinner & Wise (2020, hereafter SW20) and were adopted in model LOG1, LAR1 and TOP5, respectively.

We call the LOG1, LAR1, and TOP5 models fiducial, optimistic, and pessimistic, respectively. A summary of our models is reported in Table 1.

2.1.2. Population I–II stars

We considered the fiducial model from Iorio et al. (2023) as representative of BBHs that formed from Pop. I–II stars. The initial ZAMS mass of primary stars follows a Kroupa (2001) IMF,

$$\xi(M_{\text{ZAMS},1}) \propto M_{\text{ZAMS},1}^{-2.3} \text{ with } M_{\text{ZAMS},1} \in [5, 150] M_{\odot}. \quad (10)$$

The secondary masses, initial orbital periods, and eccentricities were distributed using the prescriptions by Sana et al. (2012) (Eqs. (4), (5), and (6)).

2.1.3. Population synthesis with SEVN

We adopted the rapid model for core-collapse supernovae (Fryer et al. 2012) to convert the final properties of stars into BH masses. Additionally, we incorporated the outcomes of electron-capture supernovae, as described by Giacobbo & Mapelli (2019). For pulsational pair-instability and pair-instability supernovae, we adopted the model by Mapelli et al. (2020). The black-hole natal kicks were drawn from the formalism presented by Giacobbo & Mapelli (2020).

SEVN integrates wind mass transfer (Bondi & Hoyle 1944), stable Roche-lobe overflow (Lubow & Shu 1975; Ulrich & Burger 1976), common-envelope evolution (Webbink 1984), and GW emission leading to orbital decay and circularization (Peters 1964). We used the same setup as in the fiducial model by Iorio et al. (2023), using their default values for all relevant parameters. We adopted a common-envelope efficiency parameter of $\alpha = 1$, which corresponds to assuming that all the orbital energy that is lost from the system contributes to unbinding the common envelope.

2.2. COSMORATE

We determined the evolution of BBH merger rate density using COSMORATE (Santoliquido et al. 2020, 2021), which interfaces catalogs of simulated BBHs with a metallicity-dependent SFRD model. The merger rate density in the source frame is given by

$$\mathcal{R}(z) = \int_{z_{\text{max}}}^{\infty} \left[\int_{Z_{\text{min}}}^{Z_{\text{max}}} \psi(z') p(Z|z') \mathcal{F}(z', z, Z) dZ \right] \frac{dt(z')}{dz'} dz', \quad (11)$$

where $\psi(z')$ represents the chosen SFRD evolution, selected from those presented in Table 1, and $p(Z|z')$ is the distribution of metallicity Z at a fixed formation redshift z' . Considering that we modeled Pop. III stars with a single metallicity value, we simply set $p(Z|z')$ equal to a Dirac delta function centered on the chosen value $Z = 10^{-11}$. For the case of Pop. I–II stars, we used (Madau & Fragos 2017)

$$\psi(z) = a \frac{(1+z)^b}{1 + [(1+z)/c]^d}, \quad (12)$$

with $a = 0.01 M_{\odot} \text{Mpc}^{-3} \text{yr}^{-1}$, $b = 2.6$, $c = 3.2$, and $d = 6.2$, while for the metallicity distribution, we adopted (Madau & Fragos 2017)

$$p(Z|z') = \frac{1}{\sqrt{2\pi}\sigma_Z} \exp\left\{-\frac{[\log(Z(z')/Z_{\odot}) - \langle \log Z(z')/Z_{\odot} \rangle]^2}{2\sigma_Z^2}\right\}, \quad (13)$$

where $\langle \log Z(z')/Z_{\odot} \rangle = \log \langle Z(z')/Z_{\odot} \rangle - \ln(10)\sigma_Z^2/2$ and $\sigma_Z = 0.2$ (Bouffanais et al. 2021a).

Table 2. Signal parameters and their adopted priors.

Parameter	Units	Prior
$m_{1,d}$	[M_{\odot}]	$\mathcal{U}(0, \infty)$
$m_{2,d}$	[M_{\odot}]	$\mathcal{U}(0, m_{1,d})$
d_L	[Mpc]	Eq. (18)
ra	[rad]	$\mathcal{U}(0, 2\pi)$
sin dec	[rad]	$\mathcal{U}(-1, 1)$
cos ι	[rad]	$\mathcal{U}(-1, 1)$
ϕ_c	[rad]	$\mathcal{U}(0, 2\pi)$
ψ	[rad]	$\mathcal{U}(0, \pi)$
t_c	[s]	$\mathcal{U}(0, \infty)$

Notes. $\mathcal{U}(a, b)$ denotes a uniform distribution between a and b . See Sect. 2.3 for details.

In Eq. (11) we introduced the expression $dt(z')/dz' = (1 + z')^{-1}H(z')^{-1}$, where the Hubble parameter $H(z)$ in a flat Λ CDM Universe is given by

$$H(z) = H_0 \sqrt{(1+z)^3 \Omega_M + (1 - \Omega_M)}, \quad (14)$$

with H_0 denoting the Hubble constant, and Ω_M representing the a -dimensional matter density parameter. We used the cosmological parameters reported by Aghanim et al. (2020). The expression $\mathcal{F}(z', z, Z)$ from Eq. (11) is given by

$$\mathcal{F}(z', z, Z) = \frac{1}{\mathcal{M}_{\text{TOT}}(Z)} \frac{d\mathcal{N}(z', z, Z)}{dt(z)}, \quad (15)$$

where $\mathcal{M}_{\text{TOT}}(Z) = \mathcal{M}_{\text{sim}}(Z)/(f_{\text{bin}}f_{\text{IMF}})$. Here, $\mathcal{M}_{\text{sim}}(Z)$ is the total simulated initial stellar mass in SEVN, $f_{\text{bin}} = 0.4$ accounts for the assumed binary fraction (Sana et al. 2012), and $f_{\text{IMF}} = 0.255$ accounts for the incomplete sampling of the IMF (lorio et al. 2023). Additionally, $d\mathcal{N}(z', z, Z)/dt(z)$ denotes the rate of BBH mergers originating from progenitor stars with a metallicity Z at redshift z' that merge at redshift z .

The total number of BBH mergers for each model, regardless of whether they are detectable, is reported in Table 1. This is given by

$$\mathcal{N} = T_{\text{obs}} \int \mathcal{R}(z)p(m_1, m_2|z) \frac{1}{1+z} \frac{dV_c}{dz} dm_1 dm_2 dz, \quad (16)$$

where $\mathcal{R}(z)$ is the merger rate density (Eq. (11)), and $p(m_1, m_2|z)$ is the two-dimensional source-frame mass distribution at a given redshift extracted with COSMORATE for each astrophysical model (Table 1). The factor $1/(1+z)$ converts source-frame time into detector-frame time, dV_c/dz is the differential comoving volume element, and T_{obs} is the observing time.

For each of our models, we drew a realization of \mathcal{N} BBHs assuming $T_{\text{obs}} = 10$ yr, obtaining a number of Pop. III mergers that is orders of magnitude smaller for the fiducial ($\sim 7 \times 10^4$), optimistic ($\sim 1 \times 10^5$), and pessimistic ($\sim 2 \times 10^2$) model compared to Pop. I–II BBHs ($\sim 9 \times 10^5$).

2.3. Parameter estimation

We simulated the parameter-estimation performance of the ET using the FIM formalism (Cutler & Flanagan 1994; Vallisneri 2008; Chan et al. 2018; Grimm & Harms 2020; Borhanian 2021; Iacovelli et al. 2022a) as implemented in GWFISH (Dupletsa et al. 2023). In this approach, the likelihood

of the data realization d is approximated as a multivariate Gaussian distribution. The posterior is therefore given by

$$p(\theta|d) \propto \pi(\theta) \exp\left(-\frac{1}{2} (\theta - \bar{\theta})^T \mathcal{F} (\theta - \bar{\theta})\right), \quad (17)$$

where $\bar{\theta}$ are the injected parameters. We assumed the prior distributions $\pi(\theta)$ reported in Table 2 (Dupletsa et al. 2024). In particular, for the luminosity distance, we chose a prior uniform in comoving volume and source-frame time,

$$\pi(d_L) \propto \frac{dV_c}{dd_L} \frac{1}{1+z} = \frac{dV_c}{dz} \frac{dz}{dd_L} \frac{1}{1+z}, \quad (18)$$

where $d_L \in [0, d_L(z = 1000)]$ Mpc. The FIM is given by

$$\mathcal{F}_{ij} = \left(\frac{\partial h}{\partial \theta_i} \middle| \frac{\partial h}{\partial \theta_j} \right) \bigg|_{\theta = \bar{\theta}}, \quad (19)$$

where GWFISH adopts $\theta = \{m_{1,d}, m_{2,d}, d_L, \text{ra}, \text{dec}, \iota, \phi_c, \psi, t_c\}$, where $m_{1,d}$, $m_{2,d}$, and d_L denote the detector-frame masses and luminosity distance, respectively; ra and dec are sky position coordinates, ι is the inclination angle of the binary with respect to the line of sight, ϕ_c is the phase at coalescence, ψ is the polarization angle, and t_c is the time of coalescence. The waveforms also depend on the spin parameters, which we set to zero. We explain our motivation for this choice and discuss it in Sect. 4.4.

We used the waveform approximant IMRPHENOMHM (Kalaghatgi et al. 2020). The inner product in Eq. (19) is defined as

$$(a|b) = 4\text{Re} \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{a(\theta, f)b^*(\theta, f)}{S_n(f)} df, \quad (20)$$

with $f_{\text{low}} = 2$ Hz and $f_{\text{high}} = 2048$ Hz, and $a(\theta, f)$ and $b(\theta, f)$ are the Fourier transforms of the time-domain signals $a(\theta, t)$ and $b(\theta, t)$. The noise power spectral density $S_n(f)$ was taken to be that of a single triangular ET detector with an arm length of 10 km located in Sardinia (latitude $40^\circ 31'$, longitude $9^\circ 25'$), which is a possible candidate site (Branchesi et al. 2023). We used the HFLF-cryogenic sensitivity curve¹, which includes both a high-frequency instrument and a cryogenic low-frequency instrument.

As the signal observed at the detector depends on the redshifted masses and the luminosity distance to the source, we converted the source-frame masses and redshift obtained through COSMORATE in Sect. 2.2 into detector-frame masses and luminosity distances,

$$m_{i,d} = m_i(1+z), \quad (21)$$

$$d_L = c(1+z) \int_0^z \frac{dz'}{H(z')}, \quad (22)$$

where $i = 1, 2$ denotes the primary and secondary mass, respectively; and the Hubble parameter $H(z)$ is given in Eq. (14).

The parameters $m_{1,d}$, $m_{2,d}$, and d_L were distributed accordingly to their astrophysical populations (Table 1). The other parameters were extracted from their prior distributions (Table 2). Furthermore, we assumed a 100% duty cycle such that T_{obs} corresponds to the full data-taking time.

In Fig. 1, we show the parameter estimation for BBHs originating in Pop. III and Pop. I–II stars observed with the ET. Notably, a few Pop. III sources in the optimistic model are measured with extremely high precision. They have $\Delta d_L/d_L \leq 0.01$

¹ The sensitivity curve can be downloaded at <https://apps.et-gw.eu/tds/?content=3&r=14065>.

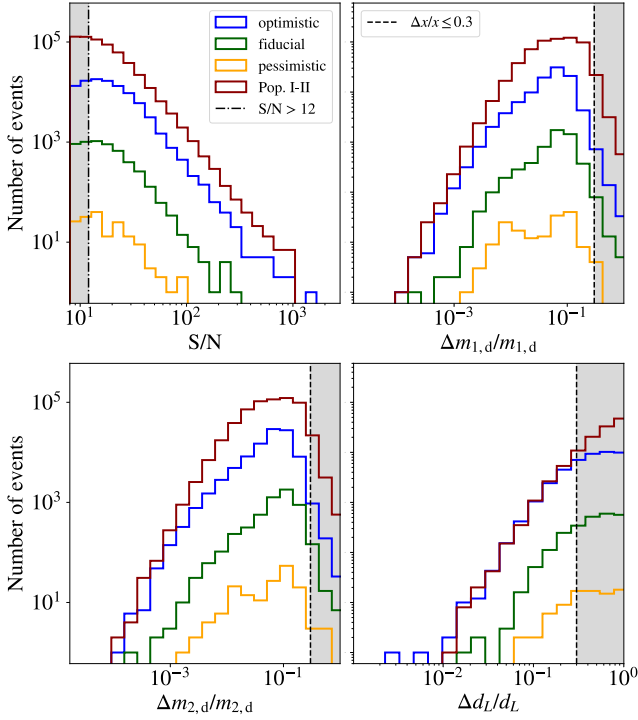


Fig. 1. Distribution of S/Ns (top left panel) and relative errors for detector-frame primary masses (top right panel), detector-frame secondary masses (bottom left panel), and luminosity distances (bottom right panel). The blue, green, and orange histograms correspond to optimistic, fiducial, and pessimistic models for Pop. III BBHs, respectively, and the red histograms refer to BBHs from Pop. I-II stars (see Sects. 2.1 and 2.3 for details). In the top left panel, the dash-dotted black line marks the adopted S/N limit, and the dashed black lines in the other panels represent the adopted threshold for relative errors, where Δx are 1σ errors and $x = \{m_{1,d}, m_{2,d}, d_L\}$.

because they are at low redshift ($z \lesssim 1$) and have high source-frame masses ($m_1 \gtrsim 30 M_\odot$).

For the remainder of the paper, we select sources with $S/N \geq 12$ and relative errors smaller than 0.3 on $m_{1,d}$, $m_{2,d}$ and d_L . We further discuss this choice in Sect. 4.5. As a result, the number of detected and selected Pop. I-II BBHs is $N_{\text{Pop. I-II}} = 11397$ and the number of Pop. III BBHs is $N_{\text{Pop. III}} = 464, 10185, 22$ for the fiducial, optimistic, and pessimistic case, respectively. Despite these thresholds, our analysis considered sources at low and high redshifts, with maximum redshift values of $z_{\text{max}} \approx 23, 19,$ and 21 for the fiducial, pessimistic, and optimistic models, respectively. We also created balanced datasets, that is, we increased T_{obs} in Eq. (16) to obtain exactly 11397 events for the Pop. III models as well.

Figs. 2, 3, and 4 illustrate the parameter distributions of Pop. III BBHs in the fiducial, optimistic, and pessimistic balanced models compared to those of Pop. I-II. The one-dimensional histograms along the diagonal axes of the pair plots show distinct peaks in the distributions of all the parameters. For instance, in the detector frame, the primary mass peaks at $20\text{--}30 M_\odot$ for Pop. I-II BBHs, whereas for Pop. III objects, the maximum is at $200\text{--}300 M_\odot$. On average, Pop. III BBHs exhibit detector-frame masses that are approximately 100 times higher than those of Pop. I-II. The peak of the source-frame primary mass distribution for BBHs generated from Pop. III stars falls within the range of $m_1 \sim 30\text{--}40 M_\odot$, and the peak of the merger rate density occurs at $z \sim 8\text{--}16$ (Santoliquido et al.

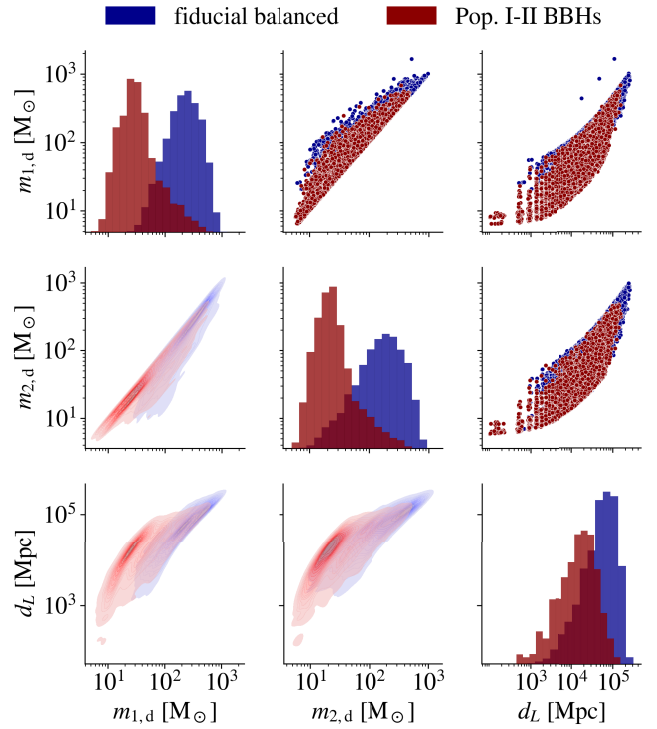


Fig. 2. Pair plot illustrating the distribution for the fiducial balanced case. The red and blue distributions indicate mergers of BBHs from Pop. I-II and Pop. III stars, respectively. The lower plots show kernel density estimations, the upper plots show the individual samples, and the diagonal plots show the marginalized distributions.

2023). In the detector frame, the primary masses of Pop. III BBHs are therefore $m_{1,d} \sim 240\text{--}640 M_\odot$. In contrast, Pop. I-II BBHs exhibit a peak in the source-frame primary mass distribution at $10\text{--}15 M_\odot$ (Broekgaarden et al. 2022; van Son et al. 2023; Santoliquido et al. 2023), and the merger rate density peaks at $z \sim 2\text{--}3$ (Santoliquido et al. 2021). The secondary mass distribution of Pop. I-II objects peaks at $\sim 20\text{--}30 M_\odot$, in contrast with Pop. III, for which it peaks at $\sim 100\text{--}200 M_\odot$. The luminosity distance peaks at $\sim 2 \times 10^4$ Mpc for Pop. I-II BBHs, and for Pop. III sources, it peaks at $\sim 10^5$ Mpc.

Nevertheless, the two populations overlap: The tails of the distributions of Pop. I-II $m_{1,d}$, $m_{2,d}$, and d_L intersect with the maxima of Pop. III. This suggests that it is not trivial to efficiently distinguish between these populations.

2.4. Classification of a single detection

The probability that the detected event j belongs to the Pop. III class (hereafter, k for brevity) is given by

$$p(j \in k | d_j, \{\beta\}) = \int d\mathbf{x} p(j \in k | \mathbf{x}, d_j, \{\beta\}) p(\mathbf{x} | d_j, \{\beta\}), \quad (23)$$

where d_j is the data stream of the event j , $\{\beta\} = \{\beta_{\text{I-II}}, \beta_{\text{III}}\}$ are the fractions of Pop. I-II and Pop. III BBH mergers, respectively, and \mathbf{x} is a subset of the waveform parameters (see Sect. 2.5). The term $p(i \in k | \mathbf{x}, d_i, \{\beta\})$ is the probability that the event j belongs to the Pop. III class given the waveform parameters \mathbf{x} , the data stream d_j , and the mixing fractions. We evaluated it by using a ML algorithm trained on detectable waveform parameters (see Sect. 2.5).

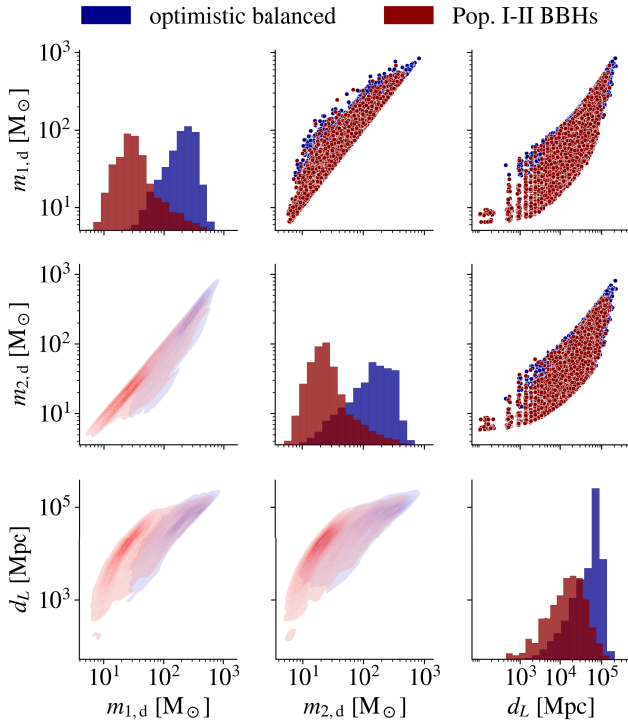


Fig. 3. Same as Fig. 2, but for the optimistic balanced model.

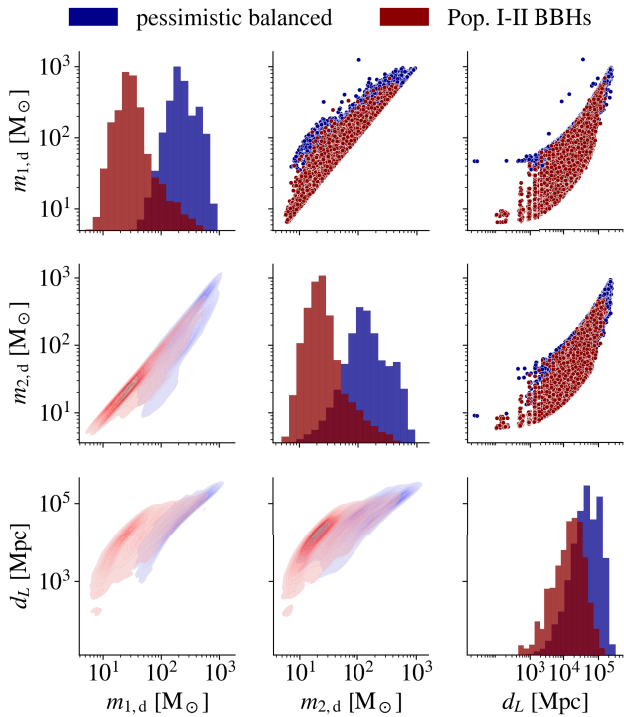


Fig. 4. Same as Fig. 2, but for the pessimistic balanced model.

We approximated the integral in Eq. (23) with a Monte Carlo summation. We made the simplifying assumption that the posterior samples of the single event j do not depend on the mixing fractions β , that is, they were obtained using uninformative priors. We further discuss this choice in Sect. 4.6. Therefore, to solve Eq. (23), we drew N_s posterior samples from $p(\mathbf{x}|d_j)$ that

Table 3. Hyperparameters adopted in XGBOOST training after a randomized search.

Hyperparameter	fid. bal.	opt. bal.	pes. bal.	range
max_depth	13	13	12	[2, 15]
learning_rate	0.034	0.034	0.016	[0, 0.1]
gamma	0.36	0.36	0.25	[0, 3]

Notes. The hyperparameters are distributed log-uniformly in the provided ranges. The column names stand for fiducial (fid.), optimistic (opt.), pessimistic (pes.), and balanced (bal.)

are given as in Eq. (17) for each detection j ,

$$p(j \in k|d_j, \{\beta\}) \approx \left\langle p(j \in k|\mathbf{x}_i, d_j, \{\beta\}) \right\rangle_{\mathbf{x}_i \sim p(\mathbf{x}|d_j)}, \quad (24)$$

where i spans the N_s posterior samples. The presence of noise typically leads to a deviation of the likelihood maximum (Eq. (17)) from the injected values in GWFISH (Rodríguez et al. 2013; Gupta et al. 2022). To simulate this, the N_s posterior samples were centered around a new maximum likelihood that was randomly sampled from Eq. (17) for each detected source (for further details, see Section 2.1 in Iacovelli et al. 2022a).

2.5. Classification probability of a single event

The determination of the probability that the event j belongs to the Pop. III class given its fixed waveform and population parameters ($p(j \in k|\mathbf{x}, d_j, \beta)$ in Eq. (23)) can be seen as a classification problem that is easily addressed with a ML algorithm. The classification we performed, despite it can be done on every waveform parameter in principle (see Sect. 2.3), is based only on the detector-frame primary and secondary mass and on the luminosity distance, that is, $\mathbf{x} = \{m_{1,d}, m_{2,d}, d_L\}$. These parameters are those linked to astrophysical processes, and the remaining parameters were assigned randomly in our simulations.

We trained three different classifiers for which we kept the BBHs from Pop. I–II stars fixed and varied those from Pop. III stars (fiducial, optimistic, and pessimistic). We trained the classifiers on balanced datasets because classification performances are maximized when the number of elements in the two classes does not vary by some orders of magnitude (Haixiang et al. 2017). Following customary practice, we reserved 70% of the sources of the balanced datasets for training and validation and used the remaining 30% to test the performance of the balanced classifiers.

To perform the classification, we adopted the ML algorithm XGBOOST (eXtreme Gradient Boosting; Chen & Guestrin 2016). The key idea behind XGBOOST is to sequentially add shallow decision trees to an ensemble of learners, each of which corrects its predecessor. At every iteration, this algorithm fits the new decision tree to the residual errors from the previous iterations.

The performance of XGBOOST relies on a set of user-adjustable hyperparameters aimed at enhancing the classifier performance on a specific dataset. To determine the optimal hyperparameters, we employed the RANDOMIZEDSEARCHCV method from SKLEARN (Pedregosa et al. 2011), which randomly selects a fixed number of hyperparameter combinations from a predefined search space. This random sampling makes the process computationally efficient and well suited for extensive search spaces (Bergstra & Bengio 2012; Lones 2021). We looped over three hyperparameters. Table 3 outlines the

ranges we tested and the identified optimal values. In particular, `max_depth` is the maximum depth of a tree in XGBOOST. Increasing `max_depth` allows the model to learn more complex relations in the data, but can also lead to overfitting, where the classifier learns the training data excessively, but struggles to generalize to new so-far unseen data. `learning_rate` determines the size of the steps taken during training. A lower `learning_rate` makes the model more robust, but requires more iterations to converge. `gamma` also controls overfitting by penalizing the complexity of the trees: higher values of `gamma` lead to fewer splits.

To prevent overfitting, we also employed a k -fold cross-validation. We divided the training dataset into $k = 5$ subsets. The model undergoes training k times, with each iteration using $k - 1$ subsets for training and the remaining set for validation. Implemented for additional robustness, we employed the so-called nested cross-validation, which introduces an inner level of cross-validation with $k = 2$ during the hyperparameter tuning process.

When trained on balanced datasets, XGBOOST provides $p(j \in k|\mathbf{x}, d_j)$, a probability independent of the mixing fraction. To evaluate the performances of the balanced classifiers, we considered a fixed threshold, that is, whenever $p(j \in k|\mathbf{x}, d_j) > 0.5$, the j event is predicted to be a Pop. III BBH.

The results are presented using confusion matrices that consist of four entries: true positive (hereafter TP, correctly predicted Pop. III BBHs), true negative (hereafter TN, correctly predicted Pop. I-II BBHs), false positive (hereafter FP, incorrectly predicted Pop. III), and false negative (hereafter FN, incorrectly predicted Pop. I-II). We evaluated the performance of our classifiers using precision, recall, and F1 score. Precision is defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (25)$$

where a high value indicates that a positive prediction from the classifier is likely to be correct. When $\text{TP} = 0$, we set $\text{Precision} = 0$. Recall (or sensitivity) is defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (26)$$

and measures the ability of the classifier to capture all the positive instances (i.e., Pop. III), with a high value indicating that the model is effective at identifying most of the positive instances. The F1 score

$$\text{F1} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (27)$$

is the harmonic mean of the precision and recall, providing a balance between the two.

Training a ML classifier on artificially balanced dataset introduces a bias favoring the minority class (Weiss 1984). Among the suggested mitigation strategies (Tian et al. 2020), we opt for the following Bayesian approach, which relies on simulation-informed priors (Chan et al. 2019; Berbel et al. 2024). The probability we wished to evaluate is $p(j \in k|\mathbf{x}, d_j, \beta)$, which enters Eq. (23). We started from the Bayes theorem written in terms of odds,

$$\frac{p(i \in k|\mathbf{x}, d_i, \beta)}{p(i \notin k|\mathbf{x}, d_i, \beta)} = \frac{\pi(i \in k|\beta)}{\pi(i \notin k|\beta)} \frac{p(\mathbf{x}, d_i|i \in k, \beta)}{p(\mathbf{x}, d_i|i \notin k, \beta)}. \quad (28)$$

The likelihood of the single event j , denoted as $p(\mathbf{x}, d_j|i \in k, \beta)$, does not depend on the fractions β , which can only vary its

normalization. We can then write $p(\mathbf{x}, d_i|i \in k, \beta) = p(\mathbf{x}, d_i|i \in k)$ and apply the Bayes theorem once more to obtain

$$p(i \in k|\mathbf{x}, d_i) = \frac{p(\mathbf{x}, d_i|i \in k)\pi(i \in k)}{p(\mathbf{x}, d_i)} = \frac{p(\mathbf{x}, d_i|i \in k)}{\sum_k p(\mathbf{x}, d_i|i \in k)}, \quad (29)$$

where the second equality follows from the fact that $\pi(i \in k) = 1/2$. We substitute Eq. (29) twice in Eq. (28), making use of the fact that the normalization $\sum_k p(\mathbf{x}, d_i|i \in k)$ is independent of k and therefore cancels out. The relevant prior probabilities are $\pi(i \notin k|\beta) = \beta_{\text{I-II}}$ and $\pi(i \in k|\beta) = \beta_{\text{III}}$. Eq. (28) then becomes

$$p(i \in k|\mathbf{x}, d_i, \beta) = \frac{p(i \in k|\mathbf{x}, d_i)\beta_{\text{III}}}{p(i \in k|\mathbf{x}, d_i)(\beta_{\text{III}} - \beta_{\text{I-II}}) + \beta_{\text{I-II}}}, \quad (30)$$

where $\beta_{\text{III}} = N_{\text{Pop. III}}/N_{\text{tot}}$ and $\beta_{\text{I-II}} = N_{\text{Pop. I-II}}/N_{\text{tot}}$ with $N_{\text{tot}} = N_{\text{Pop. III}} + N_{\text{Pop. I-II}}$ (see Sect. 2.3).

3. Results

3.1. Classification performances on balanced datasets

Fig. 5 shows that the precision-recall curves evaluated on the test sets and with the best hyperparameters listed in Table 3 remain consistent across the five folds. Moreover, the classifiers exhibit almost the same performances for different data splits, indicating strong generalization capabilities to unseen data.

The confusion matrices evaluated on the test sets are presented in Fig. 6 and indicate good classification performances across all three models. The majority of instances are correctly classified, as shown with high fractions (≥ 0.45) on the diagonal of each matrix. These robust performances are further corroborated by the scoring parameters evaluated on the test sets presented in Table 4. In particular, the achieved precision evaluated on the test sets consistently exceeds 0.90 for each classifier.

The classifiers exhibit a higher rate of confusing Pop. I-II BBHs as Pop. III than of the reverse (top right corners of the confusion matrices). This is because the tails of the parameter distribution ($m_{1,d}$, $m_{2,d}$, and d_1) of Pop. I-II BBHs extend in the ranges occupied by Pop. III BBHs (see Figs. 2, 3, and 4). This fact might imply worse classifier performances in the presence of BBHs that were generated through alternative formation channels, as discussed in more detail in Sect. 4.6.

3.2. Classification of single detections

When we take the expected merger rate ratio between detected Pop. III and Pop. I-II BBHs into account, we count that $\sim 4\%$, $\sim 47\%$, and $\sim 0.2\%$ of the total number of BBH mergers are generated from the fiducial, optimistic, and pessimistic model, respectively (Sect. 2.3). We used the probability defined in Eq. (23), which provides a confidence in determining whether a source belongs to the Pop. I-II or Pop. III class. We can thus introduce a threshold on $p(j \in k|\mathbf{x}, d_j, \beta)$: A higher threshold results in fewer Pop. III BBHs being identified, but with higher certainty. We evaluated Eq. (23) only for those sources that were in the test sets.

Table 5 presents the scores as function of this threshold. Our results reveal that we can effectively distinguish BBHs from Pop. III stars with high precision. For instance, Table 5 illustrates that with a threshold equal to 0.7, we can identify $\geq 10\%$ of Pop. III BBHs in the fiducial model with a precision $> 90\%$.

As shown in Fig. 7, sources that are confidently identified as Pop. III BBHs, that is, sources with high $p(j \in k|\mathbf{x}, d_j, \beta)$,

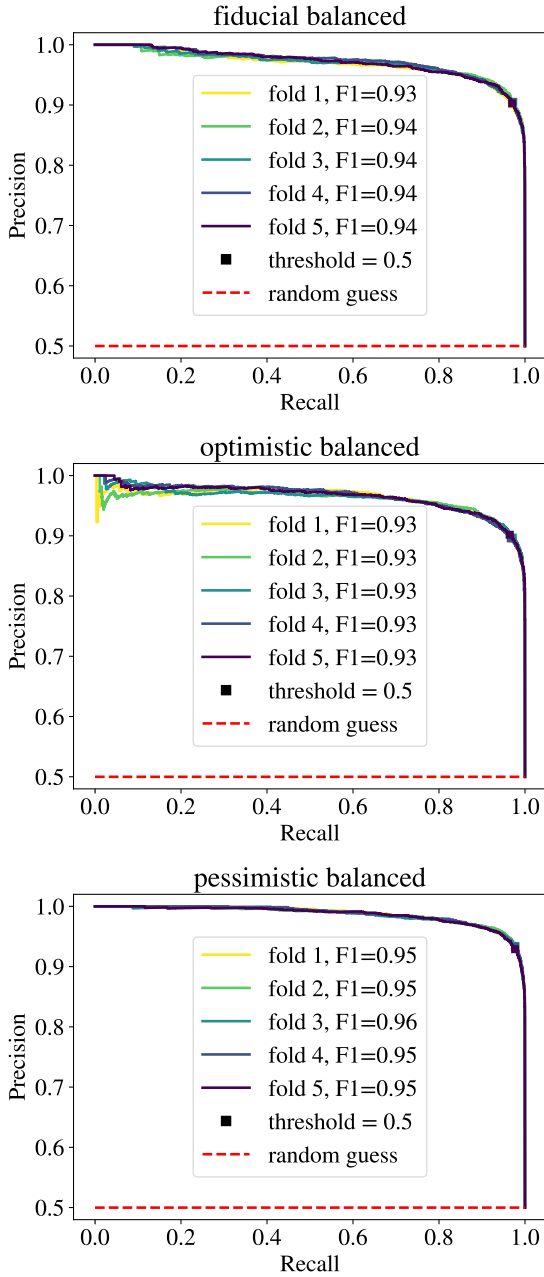


Fig. 5. Precision-recall curves evaluated on the test sets for the fiducial balanced model (top panel), optimistic balanced (middle panel), and pessimistic balanced (bottom panel). The colors indicate the results we obtained for different folds. F1 is reported as a scoring parameter for each fold and has been evaluated by applying a threshold equal to 0.5 on $p(j \in k|\mathbf{x}, d_j)$ (see Sect. 2.5 for details). The dashed horizontal red line represents the precision-recall curve of a classifier making random guesses.

typically exhibit a high median detector-frame primary mass ($m_{1,d} \gtrsim 500 M_{\odot}$) and secondary mass ($m_{2,d} \gtrsim 400 M_{\odot}$) as well as high median luminosity distance ($d_L \gtrsim 1 \times 10^5$ Mpc).

4. Discussion

4.1. Interpretability and manual classifier

One of the well-known limitations of ML algorithms is their lack of interpretability; that is, it is not always clear why or on

which basis an algorithm predicts a specific outcome. Various methods exist to address this issue with various sophistication (Angelino et al. 2018; Molnar et al. 2021; Angelov et al. 2021; Li et al. 2022). However, given the simplicity of the problem we aim to solve here, that is, to classify BBHs based on only three features, we can opt for a straightforward but highly interpretable approach: a manual classifier.

We devised a simple algorithm that systematically explores a fine grid of values within the feature space. The objective of this manual classifier is to identify thresholds for the primary mass, secondary mass, and luminosity distance that yield the highest precision (Eq. (25)). We used the same balanced training and testing dataset as described in Sect. 2.5.

The results of this approach were straightforward: We found optimal scores for the manual classifier that were exclusively determined by the threshold on the detector-frame primary mass, and the secondary mass and luminosity distance were just required to be above zero.

For the fiducial, optimistic, and pessimistic models, the thresholds we found are $m_{1,d}^{\text{thr}} \simeq 62 M_{\odot}$, $m_{1,d}^{\text{thr}} \simeq 57 M_{\odot}$, and $m_{1,d}^{\text{thr}} \simeq 72 M_{\odot}$, respectively. This cut-based manual classifier assigns any source with $m_{1,d} > m_{1,d}^{\text{thr}}$ to the class of Pop. III BBHs with $p(j \in k|\mathbf{x}, d_j) = 1$.

The performances of this manual classifier are comparable to those of XGBOOST, with the corresponding confusion matrices reported in Fig. 8. The F1 scores (Eq. (27)) evaluated on the balanced test sets are 0.87, 0.86, and 0.89 for the fiducial, optimistic, and pessimistic models, respectively.

The manual classifier highlights the primary mass as the most influential factor in distinguishing Pop. III BBHs. However, it cannot replace XGBOOST because it yields inadequate performance with imbalanced classes, as shown in Fig. 9. The precision with $p(j \in k|d_j, \{\beta\}) > 0.9$ evaluated with the manual classifier on the fiducial model is equal to 0.16. The performance of the manual classifier in unbalanced cases might be improved by introducing additional complexity, such as incorporating a sigmoid function and fitting its free parameters. However, in our view, this approach is more intricate than simply employing an easy-to-train ML algorithm.

4.2. Bayes factors

We compared our results with the method described by Mould et al. (2023), who computed the Bayes factors of single detected events generated from two different simulated populations. Specifically, we evaluated

$$\mathcal{D}_{\text{III/I-II}} = \frac{P(\det|\text{I-II})}{P(\det|\text{III})} \mathcal{B}_{\text{III/I-II}} \frac{\pi(\text{III})}{\pi(\text{I-II})}, \quad (31)$$

where the ratio $\pi(\text{III})/\pi(\text{I-II}) = \beta_{\text{III}}/\beta_{\text{I-II}}$ is the prior odds ratio as chosen in Eq. 30. We included the detection probabilities $P(\det|\text{III}) \approx 0.55$ and $P(\det|\text{I-II}) \approx 0.43$, which represent the fractions of detectable sources with the ET of Pop. III and Pop. I-II BBHs, respectively (see e.g., Thrane & Talbot 2019; Mandel et al. 2019; Vitale et al. 2022). We approximated the Bayes factor between Pop. III and Pop. I-II BBHs $\mathcal{B}_{\text{III/I-II}}$ using a Monte Carlo summation,

$$\mathcal{B}_{\text{III/I-II}} = \frac{\sum_{i=1}^{N_s} \pi(\mathbf{x}_i|\text{III})/\pi(\mathbf{x}_i|U)}{\sum_{i=1}^{N_s} \pi(\mathbf{x}_i|\text{I-II})/\pi(\mathbf{x}_i|U)}, \quad (32)$$

where i spans N_s posterior samples of a detected event j [$\mathbf{x}_i \sim p(\mathbf{x}|d_j)$], and $\pi(\mathbf{x}_i|U)$ is the probability density of the uninformative priors used during parameter estimation (see Sect. 2.3

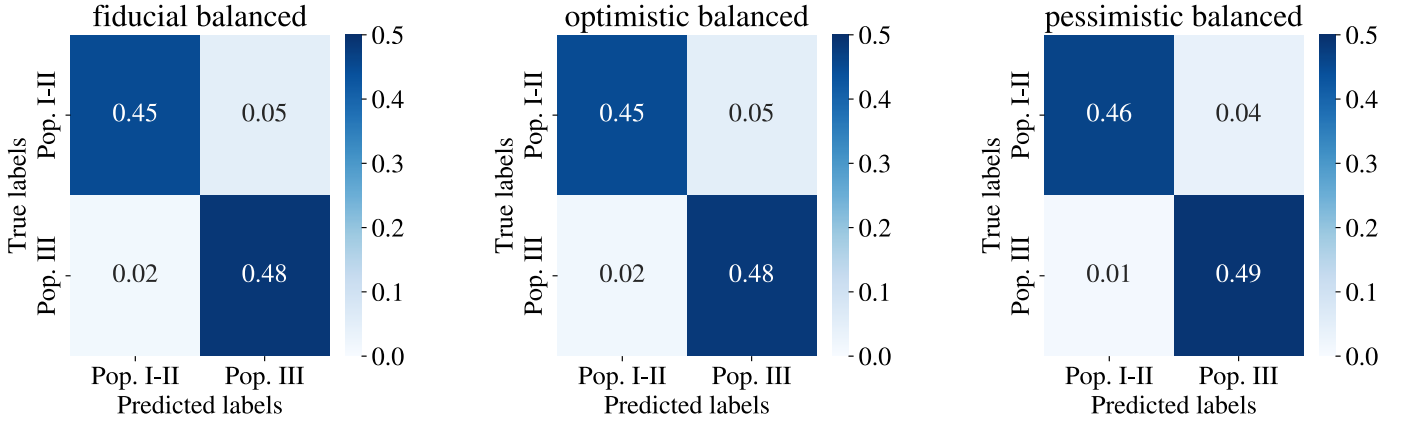


Fig. 6. Confusion matrix evaluated on the test sets for the fiducial balanced model (left panel), optimistic balanced model (middle panel), and pessimistic balanced model (right panel). The entries are color-coded and normalized to one.

Table 4. Scoring parameters evaluated on the balanced test sets for the three classifiers of Pop. III BBHs.

Model	Precision	Recall	F1 score
Fiducial balanced	0.90	0.97	0.94
Optimistic balanced	0.90	0.96	0.93
Pessimistic balanced	0.92	0.98	0.95

Notes. See Sect. 2.5 for details.

and Table 2); $\pi(\mathbf{x}|\text{III})$ and $\pi(\mathbf{x}|\text{I-II})$ are the astrophysical distributions of Pop. III and Pop. I-II BBH parameters, respectively (Fig. 2). To evaluate these probability densities, we used the Gaussian kernel density estimation (KDE) from SKLEARN (Pedregosa et al. 2011), and the bandwidth was determined using the method by Scott (2015).

We compared the Bayes factors in Eq. 31 with those obtained from Eq. 23,

$$\mathcal{X}_{\text{III/I-II}} = \frac{p(j \in k|d_j, \{\beta\})}{p(j \notin k|d_j, \{\beta\})} = \frac{p(j \in k|d_j, \{\beta\})}{1 - p(j \in k|d_j, \{\beta\})}, \quad (33)$$

where the second equality follows from the fact that $p(j \notin k|\mathbf{x}, d_j, \{\beta\}) = 1 - p(j \in k|\mathbf{x}, d_j, \{\beta\})$ (see Eq. (30)). Selection effects, represented by the factor $P(\det|\text{I-II})/P(\det|\text{III})$ in Eq. (31) are inherently accounted for in Eq. (33) because the ML classifiers are trained exclusively on detectable sources. To compare the Bayes factors, we evaluated the precision metric from Eq. (25). We placed different thresholds on the logarithms of Bayes factors to classify the same events of the fiducial model as in Fig. 7. Table 6 shows that the two methodologies perform similarly overall, as we would expect because they ultimately approximate the same quantity with different numerical approaches. However, the classification using decision trees inherently scales better with higher-dimension data, automatically handles nonstandardized data, and is generally a factor of ~ 10 faster to optimize and evaluate than KDE.

4.3. Impact of extrinsic parameters

In Sect. 2.3 we randomly chose the extrinsic parameters for each injected source based on their prior distributions (see Table 2). These parameters, such as sky position and inclination angle, can affect our results. The sensitivity of the detec-

Table 5. Scoring of the fiducial (top), optimistic, (middle) and pessimistic (bottom) models for different thresholds on $p(j \in k|d_j, \{\beta\})$, as reported in the first column.

Fiducial						
Thr.	%TP	%TN	%FP	%FN	Precision	Recall
0.1	98	83	17	2	0.19	0.98
0.2	92	89	12	8	0.25	0.92
0.5	38	98	3	62	0.43	0.38
0.7	14	100	1	87	0.90	0.14
0.9	7	100	0	93	1.00	0.07
Optimistic						
Thr.	%TP	%TN	%FP	%FN	Precision	Recall
0.1	100	74	26	1	0.78	1.00
0.2	100	77	23	1	0.79	1.00
0.3	99	79	21	1	0.81	0.99
0.5	97	83	17	3	0.84	0.97
0.7	91	88	12	9	0.87	0.91
0.9	51	95	5	49	0.91	0.51
Pessimistic						
Thr.	%TP	%TN	%FP	%FN	Precision	Recall
0.1	83	100	0	17	0.62	0.83
0.2	50	100	0	50	1.00	0.50
0.3	17	100	0	83	1.00	0.17
0.5	17	100	0	83	1.00	0.17
0.7	0	100	0	100	0	0
0.9	0	100	0	100	0	0

Notes. The scoring metrics are reported in Sect. 2.5.

tor varies across the sky, and the luminosity distance is degenerate with the inclination angle. Consequently, altering these extrinsic parameters can either reduce or increase the total number of detected sources and the accuracy of their parameter estimations.

We generated ten catalogs drawing different independent values for the extrinsic parameters. This procedure was performed exclusively for the pessimistic model because it is more likely to be sensitive to the choice of extrinsic parameters: It has fewer expected mergers than the other models. Subsequently, we injected these new populations in GWFISH.

When we varied the extrinsic parameters ten times while keeping the intrinsic parameters fixed (i.e., $m_{1,d}$, $m_{2,d}$, and d_L),

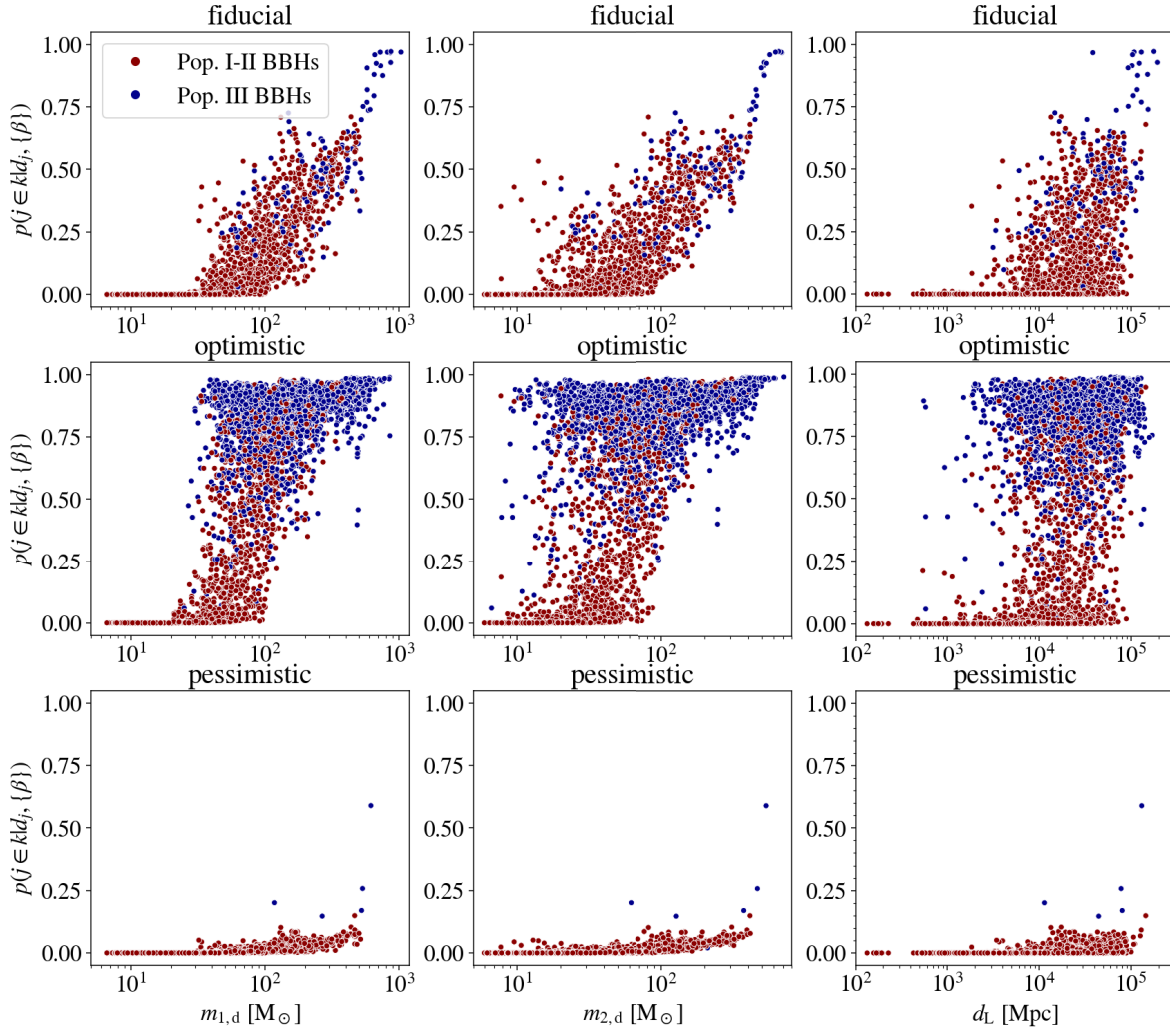


Fig. 7. Probability ($p(j \in k|d_j, \{\beta\})$ in Eq. (23)) of each mock observation to belong to the Pop. III class in the left to right columns as a function of primary mass ($m_{1,d}$), secondary mass ($m_{2,d}$), and luminosity distance (d_L). Each dot corresponds to the median of the posterior samples. The top, middle, and bottom rows correspond to the fiducial, optimistic, and pessimistic scenarios, respectively. The red (blue) dots indicate observations known to belong to Pop. I–II (Pop. III). We refer to Sect. 2.4 for details.

the total number of detected sources with a relative error of ≤ 0.3 fluctuates between $N_{\text{Pop. III}} = 13$ and $N_{\text{Pop. III}} = 25$. The percentage of BBHs correctly identified with $p(j \in k|d_j, \{\beta\}) \geq 0.1$ as being generated from Pop. III ranges between 33% and 75% with a corresponding precision between 0.4 and 1.

4.4. Spin parameters

In our analysis, we set spin parameters equal to zero. This is because SEVN only provides as output the tilt angle in the setup adopted by Costa et al. (2023). This is the angle between the rotation axis of the single BH with the direction of the angular momentum of the orbital system (Iorio et al. 2023). There is no significant difference in the distribution of cosines of tilt angles between Pop. I–II and Pop. III BBHs because $>99\%$ of them are equal to one in both populations and across redshift. In our case, the spin magnitudes must be added in post-processing (e.g., Wysocki et al. 2018; Gerosa et al. 2018; Belczynski et al. 2020; Bavera et al. 2020; Périgois et al. 2023), which means that the differences, if any, between Pop. III and Pop. I–II BBHs are artificial.

4.5. Fisher information-matrix approximation

The FIM formalism is widely used in forecasting studies, where it shapes the design and scientific objectives of advanced detectors (Borhanian & Sathyaprakash 2022; Pieroni et al. 2022; Iacovelli et al. 2022a; Branchesi et al. 2023). However, this approximation is only valid under certain conditions, most notably, the high-S/N limit (Vallisneri 2008). Even for loud sources, disparities may emerge between the parameter uncertainties estimated using the FIM formalism and those derived through full Bayesian inference (van der Sluys et al. 2008; Rodriguez et al. 2013; Mandel et al. 2014; Veitch et al. 2015; Romero-Shaw et al. 2020). Nonetheless, the chosen priors in Table 2 and the cuts on relative errors and S/N mitigate this effect (Dupletsa et al. 2024).

4.6. Population modeling uncertainty

The general performances of the classifiers may be influenced by the interplay of three factors. First, the simulated populations of Pop. I–II and Pop. III BBHs might differ from those that will be observed with the ET, potentially leading to biased

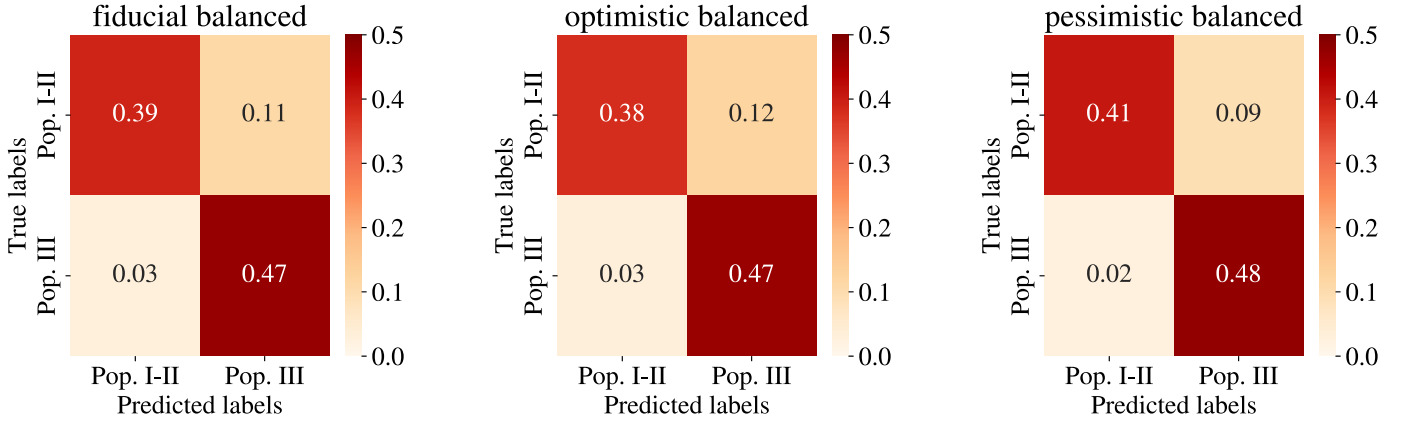


Fig. 8. Confusion matrix for the balanced fiducial model (left panel), balanced optimistic (middle panel), and balanced pessimistic (right panel) adopting manual classification (see Sect. 4.1 for details). The entries are color-coded and normalized to one.

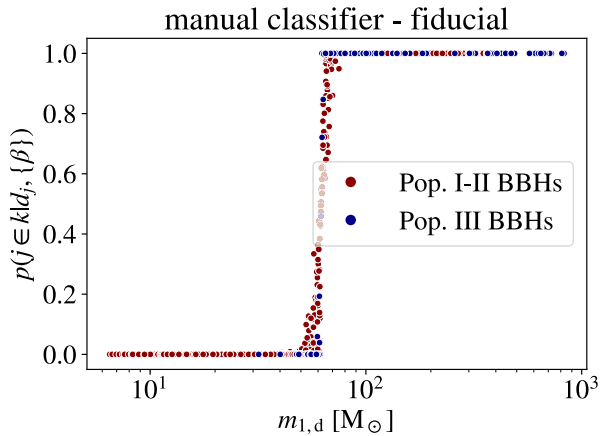


Fig. 9. Probability ($p(j \in kd_j, \{\beta\})$ in Eq. (23)) of each mock observation of the fiducial model to belong to the Pop. III class as a function of primary mass ($m_{1,d}$), evaluated with the manual classifier (see Sect. 4.1). Each dot corresponds to the median of the posterior samples. The red (blue) dots indicate observations known to belong to Pop. I-II (Pop. III).

predictions. Most existing machine-learning models operate under the closed-world assumption, assuming that the test data are drawn from the same distribution as the training data. However, in open-world scenarios, events can be unknown and deviate significantly from the training data. Therefore, it is essential to ensure that the simulated data accurately reflect reality. Several out-of-distribution detection techniques have been proposed, ranging from classification-based to density-based to distance-based methods (see Yang et al. 2021, for a review). One such technique was presented in Pasquato et al. (2024), who estimated the probability density of the training set and evaluated it on both the test set and real dataset. They used KDE scores to identify and reject out-of-distribution data points. Similarly, we will apply our method only once it has been determined that the simulated data matches reality. Additionally, retraining on alternative data is easily achievable because XGBOOST requires only few minutes on ten CPUs for training, considering both optimization and cross-validation.

The classification of detected events can also face challenges by sources that originate from different astrophysical phenomena. For instance, the dynamical formation channels predict

Table 6. Precision (Prec.) of the fiducial model for different thresholds (Thr.) on the logarithms of the Bayes factors obtained with KDE ($\ln \mathcal{D}_{\text{III/I-II}}$, Eq. (31)) and with XGBOOST ($\ln \mathcal{X}_{\text{III/I-II}}$, Eq. (33)), respectively.

Thr.	Prec. with $\ln \mathcal{D}_{\text{III/I-II}}$	Prec. with $\ln \mathcal{X}_{\text{III/I-II}}$
-3	0.18	0.19
-2	0.24	0.26
-1	0.32	0.35
0	0.46	0.56
1	0.56	1.00
2	0.90	1.00
3	1.00	1.00

a subpopulation of BBHs with primary masses ranging from 40 to 100 M_{\odot} (e.g., Ziosi et al. 2014; Rodriguez et al. 2015, 2019; Antonini & Rasio 2016; Mapelli 2016; Askar et al. 2017; Banerjee 2017, 2021; Antonini et al. 2018; Di Carlo et al. 2019, 2020; Arca Sedda & Benacquista 2019; Rastello et al. 2020; Kumamoto et al. 2020; Arca Sedda et al. 2023; Arca sedda et al. 2024; Kritos et al. 2023). When we consider the peak of their merger rate density at redshift $z \sim 2$ (Santoliquido et al. 2020), dynamically formed BHs would exhibit a detector-frame primary mass in the range of ~ 80 to 200 M_{\odot} . Consequently, the tails of the distribution for Pop. I-II BBHs would be thicker in the overlapping regions with Pop. III BBHs, potentially leading to less accurate classification performances (see Figs. 2, 3 and 4). Moreover, Pop. III BBHs could be subject to contamination from primordial black holes, which are expected to form at high redshift with a wide range of masses (e.g., De Luca et al. 2020; De Luca 2022; Franciolini et al. 2022; Ng et al. 2022b). When we consider the possibility of sources with various astrophysical origins, which we leave for future studies, it might become crucial to account for additional features to effectively disentangle them, for instance, spin properties (Gerosa et al. 2013; Qin et al. 2018, 2019; Fuller & Ma 2019; Bavera et al. 2020; Belczynski et al. 2020; Olejak & Belczynski 2021; Stevenson 2022; Périgois et al. 2023) and eccentricity (Samsing et al. 2014; Samsing & D’Orazio 2018; Zevin et al. 2019; Arca Sedda et al. 2021; Romero-Shaw et al. 2023; Codazzo et al. 2023; Dall’Amico et al. 2024). Additionally, a network of third-generation detectors might be

considered to enhance the precision in the parameter estimation (Iacovelli et al. 2022b; Borhanian & Sathyaprakash 2022).

In our analysis, while evaluating $p(j \in k|d_j, \{\beta\})$ from Eq. (23), we assumed that the hyperparameters of the population governing the distribution of single-event parameters are perfectly known. This is not generally true even with the ET, where uncertainties on hyperparameters are expected. Posterior distributions of population hyperparameters can be inferred by considering a set of gravitational-wave observations via a hierarchical Bayesian analysis (e.g. Loredò 2004; Mandel et al. 2019; Bouffanais et al. 2021b; Zevin et al. 2021; Vitale et al. 2022). We leave for future studies the integration of the hyperparameter uncertainty on $p(j \in k|d_j, \{\beta\})$ and the inclusion of population-informed priors (Miller et al. 2020; Moore & Gerosa 2021; Abbott et al. 2023a) to analyze the single-event posterior ($p(\mathbf{x}|d_j, \{\beta\})$ in Eq. (23)). In this way, we will quantify the variance that affects our results due to modeling systematics.

5. Conclusions

The ability to trace the origins of individual GW events upon their detection is pivotal for understanding the formation and evolution of compact objects. To address this problem, and with the goal of enhancing the scientific output of future GW astronomy, we coupled the heightened sensitivity of the ET with innovative ML techniques. We proposed a robust and straightforward ML approach that leverages the power of XGBOOST (Chen & Guestrin 2016), a fast-to-train algorithm based on decision trees that leads to very good performances.

Costa et al. (2023) investigated a broad parameter space of initial conditions of Pop. III progenitor stars in binary systems using advanced population-synthesis simulations, based on stellar tracks (SEVN, Iorio et al. 2023). Santoliquido et al. (2023) combined these Pop. III BBHs with different star formation rate histories to estimate their merger density evolution with redshift (COSMORATE, Santoliquido et al. 2021). From the array of available models, we specifically chose three populations of Pop. III BBHs (fiducial, optimistic, and pessimistic) based on their respective detection rates. They were classified against a population of Pop. I-II stars taken from the fiducial model by Iorio et al. (2023). The parameter estimation of these populations was performed using the FIM formalism (GWFISH, Dupletsa et al. 2023).

We trained XGBOOST on balanced datasets, which yielded very good performances with a precision exceeding 90% for all the models (Sect. 2.5). We applied these classifiers in an event-by-event scenario, in which posterior samples of detected sources, the relative ratio of Pop. III and Pop. I-II BBH merger rate, and classification confidence via $p(j \in k|d_j, \{\beta\})$ were taken into account (Sect. 2.4). Our analysis revealed that the classifiers consistently achieve a precision of >90% in classifying a fraction of BBHs from Pop. III stars. In particular, we confidently identified $\geq 10\%$, $\geq 50\%$, and $\leq 50\%$ of BBHs from Pop. III stars in the fiducial, optimistic, and pessimistic models, respectively, with a precision of >90% (Table 5).

This study proposes a new method that integrates the simulation-based information of GW source populations with the parameter estimation inference. We demonstrated its validity in discerning the origins of individual GW detections.

Acknowledgements. We acknowledge the anonymous referee for their valuable comments. We thank Biswajit Banerjee, Manuel Arca Sedda, Samuele Ronchini, Stefano Torniamenti, Elena Cuoco, Guglielmo Costa, Davide Piras and Franco Raimondi for valuable discussions. This work was performed in part at the Aspen

Center for Physics, which is supported by National Science Foundation grant PHY-2210452 and by a grant from the Simons Foundation (1161654, Troyer). F.S. and M.B. acknowledge financial support from the AHEAD2020 project (grant agreement n. 871158). M.B. also acknowledges support from the PRIN grant METE under contract No. 2020KB33TP. M.M., F.S., and G.I. acknowledge financial support from the European Research Council for the ERC Consolidator grant DEMOBLACK, under contract No. 770017 (PI: Mapelli). M.M. also acknowledges support from the German Excellence Strategy via the Heidelberg Cluster of Excellence (EXC 2181 – 390900948) STRUCTURES. D.G. is supported by ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, MUR PRIN Grant No. 2022-Z9X4XS, MSCA Fellowships No. 101064542–StochRewind and No. 101149270–ProtoBH, and the ICSC National Research Centre funded by NextGenerationEU. F.I. is supported by Swiss National Science Foundation Grant No. 200020_191957 and the SwissMap National Center for Competence in Research. The research leading to these results has been conceived and developed within the *Einstein* Telescope Observational Science Board ET-0101A-24. U.D. acknowledges Stefano Bagnasco, Federica Legger, Sara Vallero, and the INFN Computing Center of Turin for providing support and computational resources. G.I. acknowledges financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.4, – Call for tender No. 3138 of 18/12/2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU

Data availability

SEVN is publicly available at <https://gitlab.com/sevn/codes/sevn>: the version used in this work is the commit 0f9ae3bf in the branch COSTA23POPIII (<https://gitlab.com/sevncodes/sevn/-/tree/Costa23popIII>). COSMORATE is publicly available at gitlab.com/Filippo.santoliquido/cosmo_rate_public. GWFISH is publicly available at github.com/janosch314/GWFish.

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *Phys. Rev. Lett.*, **116**, 061102
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019a, *Phys. Rev. X*, **9**, 031040
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019b, *ApJ*, **882**, L24
- Abbott, R., Abbott, T. D., Zhu, X. J., et al. 2021, *Phys. Rev. X*, **11**, 021053
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023a, *Phys. Rev. X*, **13**, 011048
- Abbott, R., Abbott, T. D., Acernese, F., et al. 2023b, *Phys. Rev. X*, **13**, 041039
- Abel, T., Bryan, G. L., & Norman, M. L. 2002, *Science*, **295**, 93
- Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, *A&A*, **641**, A6
- Alhassan, W., Bulik, T., & Suchenek, M. 2023, ArXiv e-prints [arXiv:2310.10409]
- Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., & Rudin, C. 2018, arXiv e-prints [arXiv:1704.01701]
- Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. 2021, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **11**
- Antonelli, A., Kritos, K., Ng, K. K. Y., Cotesta, R., & Berti, E. 2023, *Phys. Rev. D*, **108**, 084044
- Antonini, F., & Rasio, F. A. 2016, *ApJ*, **831**, 187
- Antonini, F., Rodriguez, C. L., Petrovich, C., & Fischer, C. L. 2018, *MNRAS*, **480**, L58
- Arca Sedda, M., & Benacquista, M. 2019, *MNRAS*, **482**, 2991
- Arca Sedda, M., Li, G., & Kocsis, B. 2021, *A&A*, **650**, A189
- Arca Sedda, M., Mapelli, M., Benacquista, M., & Spera, M. 2023, *MNRAS*, **520**, 5259
- Arca sedda, M., Kamlah, A. W. H., Spurzem, R., et al. 2024, *MNRAS*, **528**, 5140
- Askar, A., Szkudlarek, M., Gondek-Rosińska, D., Giersz, M., & Bulik, T. 2017, *MNRAS*, **464**, L36
- Baker, P. T., Caudill, S., Hodge, K. A., et al. 2015, *Phys. Rev. D*, **91**, 062004
- Banerjee, S. 2017, *MNRAS*, **467**, 524
- Banerjee, S. 2021, *MNRAS*, **500**, 3002
- Baron, D. 2019, ArXiv e-prints [arXiv:1904.07248]
- Bavera, S. S., Fragos, T., Qin, Y., et al. 2020, *A&A*, **635**, A97
- Belczynski, K., Ryu, T., Perna, R., et al. 2017, *MNRAS*, **471**, 4702
- Belczynski, K., Klencki, J., Fields, C. E., et al. 2020, *A&A*, **636**, A104
- Berbel, M., Miravet-Tenés, M., Chaudhary, S. S., et al. 2024, *Class. Quant. Grav.*, **41**, 085012
- Bergstra, J., & Bengio, Y. 2012, *J. Mach. Learn. Res.*, **13**, 281
- Bondi, H., & Hoyle, F. 1944, *MNRAS*, **104**, 273
- Borhanian, S. 2021, *CQG*, **38**, 175014

- Borhanian, S., & Sathyaprakash, B. S. 2022, ArXiv e-prints [arXiv:2202.11048]
- Bouffanais, Y., Mapelli, M., Santoliquido, F., et al. 2021a, *MNRAS*, 507, 5224
- Bouffanais, Y., Mapelli, M., Santoliquido, F., et al. 2021b, *MNRAS*, 505, 3873
- Bovill, M. S., Stiavelli, M., Wiggins, A. I., Ricotti, M., & Trenti, M. 2024, *ApJ*, 962, 49
- Branchesi, M., Maggiore, M., Alonso, D., et al. 2023, *JCAP*, 2023, 068
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Broekgaarden, F. S., Berger, E., Stevenson, S., et al. 2022, *MNRAS*, 516, 5737
- Chan, M. L., Messenger, C., Heng, I. S., & Hendry, M. 2018, *Phys. Rev. D*, 97, 123014
- Chan, R., Rottmann, M., Hüger, F., Schlicht, P., & Gottschalk, H. 2019, ArXiv e-prints [arXiv:1901.08394]
- Chatterjee, D., Ghosh, S., Brady, P. R., et al. 2020, *ApJ*, 896, 54
- Chen, T., & Guestrin, C. 2016, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16* (New York, NY, USA: ACM), 785
- Chen, H.-Y., Ricarte, A., & Pacucci, F. 2022, ArXiv e-prints [arXiv:2202.04764]
- Chon, S., Omukai, K., & Schneider, R. 2021, *MNRAS*, 508, 4175
- Codazzo, E., Di Giovanni, M., Harms, J., Dall'Amico, M., & Mapelli, M. 2023, *Phys. Rev. D*, 107, 023023
- Costa, G., Bressan, A., Mapelli, M., et al. 2021, *MNRAS*, 501, 4514
- Costa, G., Mapelli, M., Iorio, G., et al. 2023, *MNRAS*, 525, 2891
- Couvares, P., Bird, I., Porter, E., et al. 2021, ArXiv e-prints [arXiv:2111.06987]
- Cuoco, E., Iess, A., Morawski, F., & Razzano, M. 2021a, *Handbook of Gravitational Wave Astronomy* (Springer), 46
- Cuoco, E., Powell, J., Cavaglià, M., et al. 2021b, *Mach. Learn.: Sci. Technol.*, 2, 011002
- Cutler, C., & Flanagan, E. E. 1994, *Phys. Rev. D*, 49, 2658
- Dall'Amico, M., Mapelli, M., Torniamenti, S., & Arca Sedda, M. 2024, *A&A*, 683, A186
- Dax, M., Green, S. R., Gair, J., et al. 2021, *Phys. Rev. Lett.*, 127, 241103
- Dax, M., Green, S. R., Gair, J., et al. 2023, *Phys. Rev. Lett.*, 130, 171403
- De Luca, V. 2022, ArXiv e-prints [arXiv:2207.08638]
- De Luca, V., Franciolini, G., Pani, P., & Riotto, A. 2020, *JCAP*, 2020, 044
- Di Carlo, U. N., Giacobbo, N., Mapelli, M., et al. 2019, *MNRAS*, 487, 2947
- Di Carlo, U. N., Mapelli, M., Giacobbo, N., et al. 2020, *MNRAS*, 498, 495
- Djorgovski, S. G., Mahabal, A. A., Graham, M. J., Polsterer, K., & Krone-Martins, A. 2022, ArXiv e-prints [arXiv:2212.01493]
- Dupletsa, U., Harms, J., Banerjee, B., et al. 2023, *Astron. Comput.*, 42, 100671
- Dupletsa, U., Harms, J., Ng, K. K. Y., et al. 2024, ArXiv e-prints [arXiv:2404.16103]
- Fairhurst, S., Mills, C., Colpi, M., et al. 2024, *MNRAS*, 529, 2116
- Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. 2015, *Phys. Rev. D*, 91, 023005
- Fishbach, M., & Holz, D. E. 2020, *ApJ*, 904, L26
- Franciolini, G., Musco, I., Pani, P., & Urbano, A. 2022, *Phys. Rev. D*, 106, 123526
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., et al. 2012, *ApJ*, 749, 91
- Fuller, J., & Ma, L. 2019, *ApJ*, 881, L1
- Gabbard, H., Messenger, C., Heng, I. S., Tonolini, F., & Murray-Smith, R. 2022, *Nat. Phys.*, 18, 112
- Gamba, R., Breschi, M., Carullo, G., et al. 2023, *Nat. Astron.*, 7, 11
- Gerosa, D., Kesden, M., Berti, E., O'Shaughnessy, R., & Sperhake, U. 2013, *Phys. Rev. D*, 87, 104028
- Gerosa, D., Berti, E., O'Shaughnessy, R., et al. 2018, *Phys. Rev. D*, 98, 084036
- Giacobbo, N., & Mapelli, M. 2019, *MNRAS*, 482, 2234
- Giacobbo, N., & Mapelli, M. 2020, *ApJ*, 891, 141
- Godfrey, J., Edelman, B., & Farr, B. 2023, ArXiv e-prints [arXiv:2304.01288]
- Green, S. R., Simpson, C., & Gair, J. 2020, *Phys. Rev. D*, 102, 104057
- Grimm, S., & Harms, J. 2020, *Phys. Rev. D*, 102, 022007
- Gupta, P. K., Puecher, A., Pang, P. T. H., et al. 2022, ArXiv e-prints [arXiv:2205.01182]
- Haiman, Z., Thoul, A. A., & Loeb, A. 1996, *ApJ*, 464, 523
- Haixiang, G., Yijing, L., Shang, J., et al. 2017, *Exp. Syst. Appl.*, 73, 220
- Hartwig, T., Volonteri, M., Bromm, V., et al. 2016, *MNRAS*, 460, L74
- Hartwig, T., Magg, M., Chen, L.-H., et al. 2022, *ApJ*, 936, 45
- Hirano, S., Hosokawa, T., Yoshida, N., Omukai, K., & Yorke, H. W. 2015, *MNRAS*, 448, 568
- Iacovelli, F., Mancarella, M., Foffa, S., & Maggiore, M. 2022a, *ApJ*, 941, 208
- Iacovelli, F., Mancarella, M., Foffa, S., & Maggiore, M. 2022b, *ApJS*, 263, 2
- Iorio, G., Mapelli, M., Costa, G., et al. 2023, *MNRAS*, 524, 426
- Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2014, *Statistics Data Mining and Machine Learning in Astronomy* (Princeton University Press)
- Iwaya, M., Kinugawa, T., & Tagoshi, H. 2023, ArXiv e-prints [arXiv:2312.17491]
- Jaacks, J., Finkelstein, S. L., & Bromm, V. 2019, *MNRAS*, 488, 2202
- Jaura, O., Glover, S. C. O., Wollenberg, K. M. J., et al. 2022, *MNRAS*, 512, 116
- Kalaghatgi, C., Hannam, M., & Raymond, V. 2020, *Phys. Rev. D*, 101, 103004
- Kalogera, V., Sathyaprakash, B. S., Bailes, M., et al. 2021, ArXiv e-prints [arXiv:2111.06990]
- Kimball, C., Talbot, C., Berry, C. P. L., et al. 2020, *ApJ*, 900, 177
- Kinugawa, T., Inayoshi, K., Hotokezaka, K., Nakauchi, D., & Nakamura, T. 2014, *MNRAS*, 442, 2963
- Kinugawa, T., Miyamoto, A., Kanda, N., & Nakamura, T. 2016, *MNRAS*, 456, 1093
- Kinugawa, T., Nakamura, T., & Nakano, H. 2020, *MNRAS*, 498, 3946
- Klessen, R. S., & Glover, S. C. O. 2023, *ARA&A*, 61, 65
- Kritos, K., Berti, E., & Silk, J. 2023, *Phys. Rev. D*, 108, 083012
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kumamoto, J., Fujii, M. S., & Tanikawa, A. 2020, *MNRAS*, 495, 4268
- Larkin, M. M., Gerasimov, R., & Burgasser, A. J. 2023, *AJ*, 165, 2
- Larson, R. B. 1998, *MNRAS*, 301, 569
- Li, X.-H., Cao, C. C., Shi, Y., et al. 2022, *IEEE Trans. Knowledge Data Eng.*, 34, 29
- Liu, B., & Bromm, V. 2020a, *MNRAS*, 495, 2475
- Liu, B., & Bromm, V. 2020b, *MNRAS*, 497, 2839
- Liu, B., Hartwig, T., Sartorio, N. S., et al. 2024a, *MNRAS*, 534, 1634
- Liu, S., Wang, L., Hu, Y.-M., Tanikawa, A., & Trani, A. A. 2024b, *MNRAS*, 533, 2262
- Lones, M. A. 2021, ArXiv e-prints [arXiv:2108.02497]
- Loredo, T. J. 2004, *Am. Inst. Phys. Conf. Ser.*, 735, 195
- Lubow, S. H., & Shu, F. H. 1975, *ApJ*, 198, 383
- Madau, P., & Fragos, T. 2017, *ApJ*, 840, 39
- Maggiore, M., Van Den Broeck, C., Bartolo, N., et al. 2020, *JCAP*, 2020, 050
- Mancarella, M., Iacovelli, F., & Gerosa, D. 2023, *Phys. Rev. D*, 107, L101302
- Mandel, I., Berry, C. P. L., Ohme, F., Fairhurst, S., & Farr, W. M. 2014, *CQG*, 31, 155005
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, 486, 1086
- Mapelli, M. 2016, *MNRAS*, 459, 3432
- Mapelli, M., Spera, M., Montanari, E., et al. 2020, *ApJ*, 888, 76
- Marcoccia, P., Nardini, G., & Pieroni, M. 2024, *MNRAS*, 531, 4444
- Meena, A. K., Zitrin, A., Jiménez-Teja, Y., et al. 2023, *ApJ*, 944, L6
- Miller, S., Callister, T. A., & Farr, W. M. 2020, *ApJ*, 895, 128
- Molnar, C., König, G., Herbinger, J., et al. 2021, ArXiv e-prints [arXiv:2007.04131]
- Moore, C. J., & Gerosa, D. 2021, *Phys. Rev. D*, 104, 083008
- Moriwaki, K., Nishimichi, T., & Yoshida, N. 2023, *Rep. Prog. Phys.*, 86, 076901
- Mould, M., Gerosa, D., & Taylor, S. R. 2022, *Phys. Rev. D*, 106, 103013
- Mould, M., Gerosa, D., Dall'Amico, M., & Mapelli, M. 2023, *MNRAS*, 525, 3986
- Nandal, D., Regan, J. A., Woods, T. E., et al. 2023, *A&A*, 677, A155
- Ng, K. K. Y., Vitale, S., Farr, W. M., & Rodriguez, C. L. 2021, *ApJ*, 913, L5
- Ng, K. K. Y., Chen, S., Goncharov, B., et al. 2022a, *ApJ*, 931, L12
- Ng, K. K. Y., Franciolini, G., Berti, E., et al. 2022b, *ApJ*, 933, L41
- Ng, K. K. Y., Goncharov, B., Chen, S., et al. 2023, *Phys. Rev. D*, 107, 024041
- Nguyen, C. T., Costa, G., Girardi, L., et al. 2022, *A&A*, 665, A126
- Olejak, A., & Belczynski, K. 2021, *ApJ*, 921, L2
- Pasquato, M., Trevisan, P., Askar, A., et al. 2024, *ApJ*, 965, 89
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825
- Pérgois, C., Mapelli, M., Santoliquido, F., Bouffanais, Y., & Rufolo, R. 2023, *Universe*, 9, 507
- Peters, P. C. 1964, *Phys. Rev.*, 136, 1224
- Pieroni, M., Ricciardone, A., & Barausse, E. 2022, *Sci. Rep.*, 12, 17940
- Prole, L. R., Clark, P. C., Klessen, R. S., & Glover, S. C. O. 2022, *MNRAS*, 510, 4019
- Punturo, M., Abernathy, M., Acernese, F., et al. 2010, *CQG*, 27, 194002
- Qin, Y., Fragos, T., Meynet, G., et al. 2018, *A&A*, 616, A28
- Qin, Y., Marchant, P., Fragos, T., Meynet, G., & Kalogera, V. 2019, *ApJ*, 870, L18
- Rastello, S., Mapelli, M., Di Carlo, U. N., et al. 2020, *MNRAS*, 497, 1563
- Reitze, D., Adhikari, R. X., Ballmer, S., et al. 2019, *Bull. Am. Astron. Soc.*, 51, 35
- Riggi, S., Umama, G., Trigilio, C., et al. 2024, *PASA*, 41, e029
- Rodriguez, C. L., Farr, B., Farr, W. M., & Mandel, I. 2013, *Phys. Rev. D*, 88, 084013
- Rodriguez, C. L., Morscher, M., Pattabiraman, B., et al. 2015, *Phys. Rev. Lett.*, 115, 051101
- Rodriguez, C. L., Zevin, M., Amaro-Seoane, P., et al. 2019, *Phys. Rev. D*, 100, 043027
- Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, *MNRAS*, 499, 3295
- Romero-Shaw, I. M., Gerosa, D., & Loutrel, N. 2023, *MNRAS*, 519, 5352
- Roulet, J., & Venumadhav, T. 2024, ArXiv e-prints [arXiv:2402.11439]

- Rydberg, C.-E., Zackrisson, E., Lundqvist, P., & Scott, P. 2013, *MNRAS*, **429**, 3658
- Samsing, J., & D’Orazio, D. J. 2018, *MNRAS*, **481**, 5445
- Samsing, J., MacLeod, M., & Ramirez-Ruiz, E. 2014, *ApJ*, **784**, 71
- Sana, H., de Mink, S. E., de Koter, A., et al. 2012, *Science*, **337**, 444
- Santoliquido, F., Mapelli, M., Bouffanais, Y., et al. 2020, *ApJ*, **898**, 152
- Santoliquido, F., Mapelli, M., Giacobbo, N., Bouffanais, Y., & Artale, M. C. 2021, *MNRAS*, **502**, 4877
- Santoliquido, F., Mapelli, M., Iorio, G., et al. 2023, *MNRAS*, **524**, 307
- Sasaoka, S., Koyama, N., Dominguez, D., et al. 2023, *Phys. Rev. D*, **108**, 123033
- Schauer, A. T. P., Bromm, V., Drory, N., & Boylan-Kolchin, M. 2022, *ApJ*, **934**, L6
- Scott, D. W. 2015, *Multivariate Density Estimation: Theory, Practice, and Visualization* (Hoboken, NJ: John Wiley and Sons)
- Singh, N., Bulik, T., Belczynski, K., & Askar, A. 2022, *A&A*, **667**, A2
- Skinner, D., & Wise, J. H. 2020, *MNRAS*, **492**, 4386
- Spera, M., Mapelli, M., Giacobbo, N., et al. 2019, *MNRAS*, **485**, 889
- Stacy, A., & Bromm, V. 2013, *MNRAS*, **433**, 1094
- Stevenson, S. 2022, *ApJ*, **926**, L32
- Susa, H., Hasegawa, K., & Tominaga, N. 2014, *ApJ*, **792**, 32
- Tanikawa, A. 2024, *Rev. Mod. Plasma Phys.*, **8**, 13
- Tanikawa, A., Susa, H., Yoshida, T., Trani, A. A., & Kinugawa, T. 2021, *ApJ*, **910**, 30
- Tanikawa, A., Chiaki, G., Kinugawa, T., Suwa, Y., & Tominaga, N. 2022a, *PASJ*, **74**, 521
- Tanikawa, A., Yoshida, T., Kinugawa, T., et al. 2022b, *ApJ*, **926**, 83
- Tegmark, M., Silk, J., Rees, M. J., et al. 1997, *ApJ*, **474**, 1
- Thrane, E., & Talbot, C. 2019, *PASA*, **36**, e010
- Tian, J., Liu, Y.-C., Glaser, N., Hsu, Y.-C., & Kira, Z. 2020, arXiv e-prints [arXiv:2010.11820]
- Tong, H., Galadage, S., & Thrane, E. 2022, *Phys. Rev. D*, **106**, 103019
- Trussler, J. A. A., Conselice, C. J., Adams, N. J., et al. 2023, *MNRAS*, **525**, 5328
- Ulrich, R. K., & Burger, H. L. 1976, *ApJ*, **206**, 509
- Valiante, R., Schneider, R., Volonteri, M., & Omukai, K. 2016, *MNRAS*, **457**, 3356
- Vallisneri, M. 2008, *Phys. Rev. D*, **77**, 042001
- van der Sluys, M., Raymond, V., Mandel, I., et al. 2008, *CQG*, **25**, 184011
- van Son, L. A. C., de Mink, S. E., Chruślińska, M., et al. 2023, *ApJ*, **948**, 105
- Veitch, J., Raymond, V., Farr, B., et al. 2015, *Phys. Rev. D*, **91**, 042003
- Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2022, *Handbook of Gravitational Wave Astronomy*, 45
- Wang, L., Tanikawa, A., & Fujii, M. 2022, *MNRAS*, **515**, 5106
- Webbink, R. F. 1984, *ApJ*, **277**, 355
- Weiss, G. M. *Foundations of Imbalanced Learning* (John Wiley & Sons, Ltd), 13
- Wiggins, A. I., Bovill, M. S., Strolger, L.-G., Stiavelli, M., & Bowling, C. 2024, *A&A*
- Williams, M. J., Veitch, J., & Messenger, C. 2021, *Phys. Rev. D*, **103**, 103006
- Wollenberg, K. M. J., Glover, S. C. O., Clark, P. C., & Klessen, R. S. 2020, *MNRAS*, **494**, 1871
- Wong, K. W. K., Isi, M., & Edwards, T. D. P. 2023, *ApJ*, **958**, 129
- Wouters, T., Pang, P. T. H., Dietrich, T., & Van Den Broeck, C. 2024, ArXiv e-prints [arXiv:2404.11397]
- Wysocki, D., Gerosa, D., O’Shaughnessy, R., et al. 2018, *Phys. Rev. D*, **97**, 043014
- Yang, J., Zhou, K., Li, Y., & Liu, Z. 2021, ArXiv e-prints [arXiv:2110.11334]
- Yi, S.-X., Stoppa, F., Nelemans, G., & Cator, E. 2022, *A&A*, **663**, A156
- Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, *ApJ*, **592**, 645
- Zackrisson, E., Hultquist, A., Kordt, A., et al. 2024, *MNRAS*, submitted [arXiv:2312.09289]
- Zevin, M., Coughlin, S., Bahaadini, S., et al. 2017, *CQG*, **34**, 064003
- Zevin, M., Samsing, J., Rodriguez, C., Haster, C.-J., & Ramirez-Ruiz, E. 2019, *ApJ*, **871**, 91
- Zevin, M., Bavera, S. S., Berry, C. P. L., et al. 2021, *ApJ*, **910**, 152
- Ziosi, B. M., Mapelli, M., Branchesi, M., & Tormen, G. 2014, *MNRAS*, **441**, 3703