A dark standard siren measurement of the Hubble constant following LIGO/Virgo/KAGRA O4a and previous runs

C. R. Bom [1,2‹] V. Alfradique [1] A. Palmese [3] G. Teixeira,[1] L. Santana-Silva [1] A. Santos[1] and P. Darc[1]

[1] *Centro Brasileiro de Pesquisas Físicas, Coordenação de Desenvolvimento Técnológico, Rua Dr Xavier Sigaud 150, 22290-180 Rio de Janeiro, RJ, Brazil*
[2] *Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rodovia Márcio Covas, lote J2, quadra J, 23812-101 Itaguaí, RJ, Brazil*
[3] *McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh PA 15213, USA*

ABSTRACT

We present a new constraint on the Hubble constant (H_0) from the standard dark siren method using a sample of five well-covered gravitational wave (GW) alerts reported during the first part of the fourth observing run of the Laser Interferometer Gravitational-Wave Observatory (LIGO), the Virgo and Kamioka Gravitational Wave Detector (KAGRA) collaborations (LVK) and with three updated standard dark sirens from third observation run in combination with the previous constraints from the first three runs. Our methodology relies on the galaxy catalogue method alone. We use a deep learning method to derive the full probability density estimation of photometric redshifts using the Legacy Survey catalogues. We add the constraints from well localized binary black hole mergers to the sample of standard dark sirens analysed in our previous work. We combine the H_0 posterior for 5 new standard sirens with other 10 previous events (using the most recent available data for the five novel events and updated three previous posteriors from O3), finding $H_0 = 70.4^{+13.6}_{-11.7}$ km s^{-1} Mpc^{-1} (68 per cent confidence interval) with the catalogue method only. This result represents an improvement of ∼ 23 per cent comparing the new 15 dark siren constraints with the previous 10 dark siren constraints and a reduction in uncertainty of ∼ 40 per cent from the combination of 15 dark and bright sirens compared with the GW170817 bright siren alone. The combination of dark and bright siren GW170817 with recent jet constraints yields H_0 of $68.0^{+4.4}_{-3.8}$ km s^{-1} Mpc^{-1}, a ∼ 6 per cent precision from standard sirens, reducing the previous constraint uncertainty by ∼ 10 per cent.

Key words: gravitational waves – catalogues – cosmological parameters – cosmology: observations.

1 INTRODUCTION

The current 4–6σ tension in the Hubble constant (Planck Collaboration VI 2018; Freedman et al. 2019; Riess et al. 2019, 2021) arises from the significant discrepancy between different cosmological probes, in particular from the cosmic microwave background (CMB; Planck Collaboration VI 2018), and those using supernovae (SNe) and cepheids for the local distance ladder (Riess et al. 2021). New independent measurements of the Hubble constant have the potential to shed light on this discrepancy (e.g. Verde, Treu & Riess 2019; Dainotti et al. 2021; Abdalla et al. 2022), and, depending on their precision, could arbitrate the tension. Among novel probes, the standard sirens (see Schutz 1986) methodology employs gravitational wave (GW) occurrences to obtain luminosity distances. This information is used to infer cosmological parameters, most notably the Hubble constant H_0, upon integration with redshift data derived from host galaxies. This emergent probe could play an important role as it is independent of the cosmic distance ladder (Chen, Fishbach & Holz

2018; Gray et al. 2020; Mukherjee et al. 2021a; Cigarrán Díaz & Mukherjee 2022; Alves et al. 2024; Bom & Palmese 2024).

Standard sirens are categorized as 'bright' in instances where an electromagnetic counterpart is definitively identified alongside a singular host galaxy, and as 'dark' or 'statistical' in the absence of such counterparts. Bright sirens, exemplified by GW170817 (Abbott et al. 2017a), yield measurements of high precision. However, the requirement of host identification poses a series of challenges due to the wide search volume and the cadence requirements (Andreoni et al. 2022; Bom et al. 2024). Furthermore, kilonovae have been the only widely confirmed electromagnetic sources detectable from GWs, although black hole mergers are proposed to produce electromagnetic counterparts in certain circumstances with a few identified candidates (Graham et al. 2023; Cabrera et al. 2024; Rodríguez-Ramírez, Nemmen & Bom 2024a; Rodríguez-Ramírez et al. 2024b). Therefore, the dark siren method can be applied to a larger number of events, including GW170817 (Fishbach et al. 2019), the binary black hole (BBH) mergers GW170814 (Soares-Santos et al. 2019) and GW190814 (Palmese et al. 2020), and several events from the first three LIGO/Virgo observing runs (Abbott et al. 2023c). Dark standard sirens assuming the catalogue method (Gair et al. 2022)

rely on the position and redshift of potential host galaxies, leading to less precise results than bright sirens on a single-event basis. However, combining dark and bright sirens can enhance constraints on cosmological parameters, leveraging the abundance of events without counterparts.

The fourth LIGO/Virgo/KAGRA (LVK) Observing run (O4) began on 2023 May 24, and is scheduled to span 20 months, including 2 months allocated for commissioning breaks dedicated to maintenance. The initial segment of O4 referred to as O4a, finished on 2024 January 16. During O4a there have been 82 GW alerts reported, of which 80 were classified as BBH candidates. It is important to note that the Virgo detector was not operational during the O4a run, therefore the typical sky localizations were worse than would have been expected if it had joined the run. Meanwhile, KAGRA has participated in the run for a limited duration, albeit at a notably reduced binary neutron star (BNS) inspiral range compared to the LIGO detectors. In this work, we use a set of well-localized and confident candidate events from O4a covered by public photometric catalogues and imaging data, mainly from the Legacy Survey (LS; Dey et al. 2019; Zhou et al. 2020) and DECam Local Volume Exploration Survey (DELVE; Drlica-Wagner et al. 2021, 2022). The photometric redshifts were computed using the same deep learning technique from Alfradique et al. (2024). We added five new sirens to the sample of dark sirens from Palmese et al. (2020, 2023) and Alfradique et al. (2024), increasing the total to 15 dark sirens. The dark siren catalogue method used in this work to constrain H_0 is described in the aforementioned papers, and we also refer to the Section 3 for a description of our methodology.

2 DATA

2.1 The LIGO/Virgo GW data

In this study, we expand the analysis upon the previous works with 8 events (Palmese et al. 2023, hereafter P23) and 10 events (Alfradique et al. 2024, hereafter A24) from LVK run O1–O3, using the catalogue method. We employ comparable selection criteria from prior studies (see P23 and A24) and, as such, investigate five new sirens identified in O4 with more than 70 per cent of their probability coverage falling within the Legacy or DELVE surveys, and with luminosity distances less than $d_L <$ 1500 Mpc. We restricted in distance based on the previous works, as the catalogues turn more incomplete and also the dependency on Ω_m becomes more relevant. The positional data for these events come from the maps publicly provided by the LVK collaboration. We present the 90 per cent credible interval (CI) sky region of all events used in this work in Fig. 1. These maps of right ascension (RA), declination (Dec.), and distance probability, are represented using HEALPIX pixelation. Within this framework, the probability distribution along each line of sight from the maps is assumed to follow a Gaussian distribution. We draw attention to the fact that the criteria adopted here are chosen with the intention of selecting those superevents with greater constraint capability, but there is no impediment for others to be added to the sample, as long as the selection function (see its definition in Section 3), defined in the H_0 posterior, correctly describes the cuts adopted.

The five novel siren events we selected are S231226av (Ligo Scientific Collaboration, VIRGO Collaboration & Kagra Collaboration 2023e), S231206cc (LIGO Scientific Collaboration & Virgo Collaboration 2023d), S230922g (LIGO Scientific Collaboration & Virgo Collaboration 2023c), S230919bj (LIGO Scientific Collaboration & Virgo Collaboration 2023b), and S230627c (LIGO Scientific

Collaboration & Virgo Collaboration 2023a). They are all classified as BBH mergers with probability $>$ 99 per cent, except for S230627c, which has a classification of 49 per cent for neutron star–black hole, 48 per cent for BBH, and 3 per cent of being noise. In particular, S231226av is among the lowest false-alarm-rate (FAR) events, while S230919bj and S230627c are in the top 20 per cent percentile of low FAR in O4. The novel events with 90 per cent volume are comparable to the previous sirens studied with 90 per cent volume $\sim 10^{-3}$ Gpc^3. We use the latest public skymaps (see Table 1 for details) produced by the PYTHON code BILBY (Ashton et al. 2019). We use the last skymaps from the GW Transient Catalogue (GWTC[1]; LIGO Scientific Collaboration, Virgo Collaboration & KAGRA Collaboration 2021) for three events from P23, which used the maps from the GW alerts S191204r, S200129m, and S200311bg. The updated events in this work are named GW191204_171526, GW200129_065448, and GW200311_115853, respectively. For the remaining events, we use the last publicly available skymaps on the GraceDB event page.[2] Furthermore, to perform an additional check on the quality of the selected GW alerts from O4 in this work, we reproduced the same cuts used on the O3 GW alerts. These cuts excluded retracted events after a few hours or days, using the same level of FAR considered, distance, and area.

2.2 Optical survey data

We made use of publicly available catalogues from the DESI Legacy Survey (Dey et al. 2019) and DELVE survey (Drlica-Wagner et al. 2021, 2022). The combined footprint is presented in Fig. 1. We use this data to obtain precise photometric redshifts using the same technique and deep learning method described in A24. We use a mixture density network (MDN; Bishop 1994) to derive the full probability density functions (PDFs) for each galaxy in the survey catalogue. A more comprehensive description of the model and photometric redshift quality assessment in both surveys and a comparison between the MDN method and the public data for the Legacy Survey is presented in the Section 2.3. Our final constraints use the Legacy Survey PDF instead of Gaussian approximations from the Legacy Survey photo-z catalogues (Zhou et al. 2020). The Legacy Survey results slightly outperformed the DELVE for the same model. We use DELVE-based photometric redshift as a validation and we found a small impact for the final constraints of < 0.5 km s^{-1} Mpc^{-1}.

In the GW sample processed in this work, only the S230919bj event has significant 90 per cent region uncovered by the photo-z catalogues from both Legacy galaxy survey and DELVE. To compensate for the insufficient coverage, we adopted the same procedure from Palmese et al. (2023). Our strategy involves distributing simulated galaxies in regions lacking data. To ensure that marginalization includes all possible host galaxies and leave our Hubble constant measurement free of underestimated uncertainty, the injected galaxies follow our prior distribution as given by the training sample. The photo-z distribution of these fake galaxies was sampled using the Monte Carlo technique. We assume a uniform spatial distribution, and the number density is given by the value of the Legacy Survey galaxy catalogue. For the photo-z precision, we first found the relation between the photo-z error and photo-z by computing the mean and standard deviation value of the photo-z error in photo-z bins of size 0.05 from our training sample, after that, we sampled from a Gaussian function for each photo-z bins. This same

---

[1]https://gwosc.org/eventapi/html/GWTC/
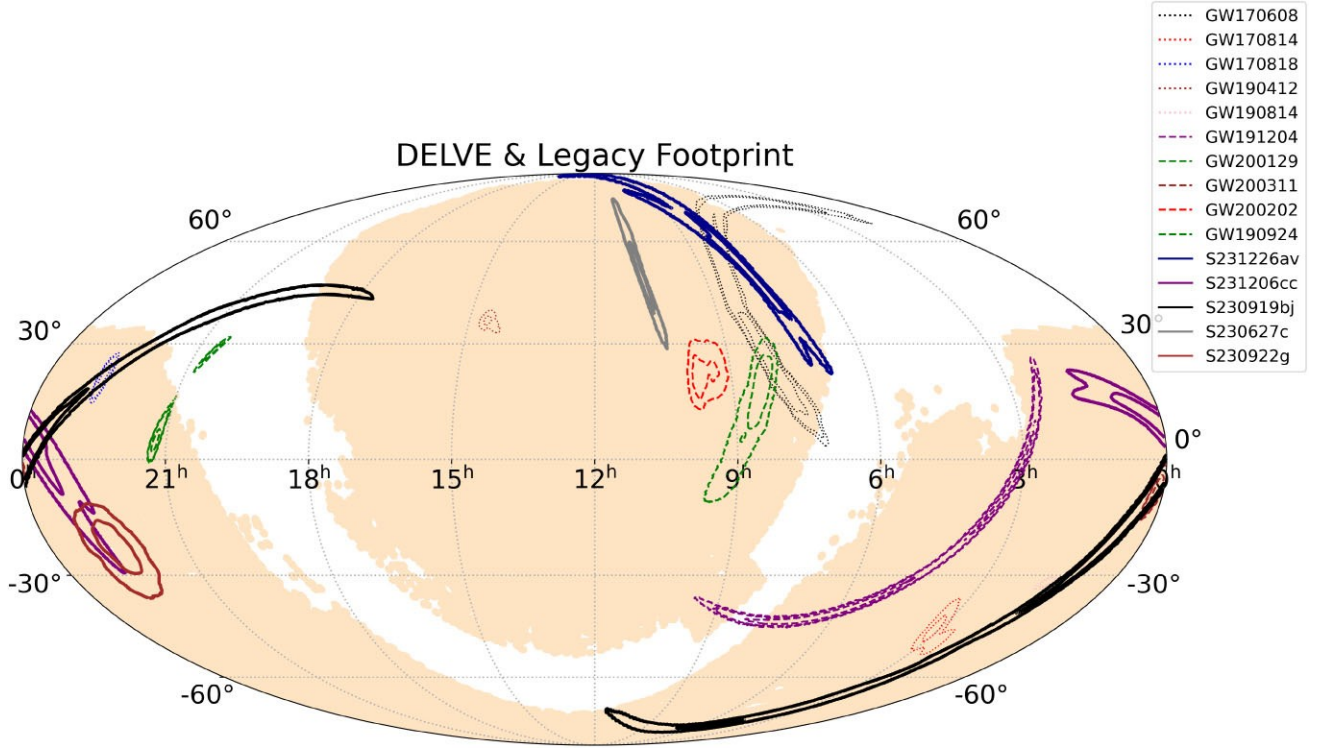[2]https://gracedb.ligo.org/

Figure 1. The LIGO/Virgo/KAGRA dark sirens analysed in this paper. The contours depict the 90 per cent CI localization from the sky maps. Dotted contour lines indicate events scrutinized by Palmese et al. (2023). Dashed lines represent events examined in Alfradique et al. (2024), which underwent reprocessing. Solid lines denote newly incorporated events in the analysis. The shaded areas represent regions covered by the DELVE and DESI Legacy Survey catalogues.

Table 1. Luminosity distance, 90 per cent CI area and volume, and FAR of gravitational wave events and candidates used in this analysis. We also report the reference paper or Genral Coordinates Network (GCN) circulars that reports to the sky map used for each event. Where a range of FAR is provided, this is because multiple FAR estimates are available from multiple search algorithms. The FAR reported for the candidates is different from the confirmed events as it is estimated from the online analysis. These candidates have all recently been confirmed as gravitational wave events in LIGO Scientific Collaboration (2021). References for each event: (1) Abbott et al. (2019), (2) Abbott et al. (2017b), (3) Abbott et al. (2021), (4) Abbott et al. (2020), (5) Abbott et al. (2023b), (6) Ligo Scientific Collaboration (2023e), (7) Ligo Scientific Collaboration (2023d), (8) Ligo Scientific Collaboration (2023c), (9) Ligo Scientific Collaboration (2023b), and (10) Ligo Scientific Collaboration (2023a).

Event	d_L (Mpc)	A (deg^2)	V (Gpc^3)	FAR
GW170608[1]	320^{+120}_{-110}	392	3×10^{-3}	< 1 per 10^5 yr
GW170814[2]	540^{+130}_{-210}	62	2×10^{-3}	< 1 per $10^4 - 10^7$ yr
GW170818[1]	1060^{+420}_{-380}	39	7×10^{-3}	< 1 per 10^5 yr
GW190412[3]	740^{+120}_{-130}	12	4×10^{-4}	< 1 per $10^3 - 10^5$ yr
GW190814[4]	241^{+26}_{-26}	19	3×10^{-5}	< 1 per $10^4 - 10^7$ yr
GW190924_021846[3]	564^{+145}_{-145}	348	1×10^{-2}	< 1 per 10^5 yr
GW191204_171526[5]	624^{+123}_{-123}	256	8×10^{-3}	< 1 per 10^5 yr
GW200129_065448[5]	929^{+179}_{-179}	31	3×10^{-3}	< 1 per 10^5 yr
GW200311_115853[5]	1154^{+206}_{-206}	35	6×10^{-3}	< 1 per 10^5 yr
GW200202_154313[8]	409^{+95}_{-95}	167	2×10^{-3}	< 1 per 10^5 yr
S231226av[6]	1218^{+171}_{-171}	199	3×10^{-2}	< 1 per 10^{42} yr
S231206cc[7]	1467^{+264}_{-264}	342	9×10^{-2}	< 1 per 10^{27} yr
S230922g[8]	1491^{+443}_{-443}	324	1×10^{-1}	< 1 per 10^{16} yr
S230919bj[9]	1491^{+402}_{-402}	708	2×10^{-1}	< 1 per 10^2 yr
S230627c[10]	291^{+64}_{-64}	82	4×10^{-4}	< 1 per 10^2 yr

procedure was adopted for the apparent magnitude distribution of these fake galaxies, since it is intrinsically associated with the redshift value. Here, we also assume that all the fake photometric redshift PDFs follow a Gaussian function; we make a Gaussian sampling around the assigned true value with the standard deviation given by their respective photo-z error. In Section 4, we present the impact of these fakes galaxies and found it to be minor.

2.3 Galaxy catalogues and photometric redshifts

In this work, we use catalogue data from both DELVE survey (Drlica-Wagner et al. 2021, 2022) and the Legacy Survey (Dey et al. 2019) as the main data used to select the galaxies within the LIGO/Virgo localization for each event. It is worth mentioning that the Legacy Survey catalogues incorporated publicly available DECam data, including DELVE footprint for the LS DR10. The DELVE survey observed a large fraction of the Southern sky covering an area of 21 000 deg^2, of which 17 000 deg^2 were homogeneously observed in the four broad-bands (g, r, i, and z) and photometric depth up to 24.3, 23.9, 23.5, and 22.8 for g, r, i, and z, respectively for 5σ point source detections. The DELVE catalogues comprise 2.5 billion sources, in which 618 million have data in all the four bands. The Legacy Survey explored a significant portion of the sky ($\sim 33\,500$ deg^2) in grz bands, reaching depths of 24.0, 23.4, and 22.5 in grz for 5σ detection. The sky coverage of DELVE and Legacy Survey catalogues are shown in Fig. 1, together with the 50 per cent and 90 per cent CI of the GW events studied in previous works and the five new events explored in this study.

Photometric redshifts used in this work were determined using the same deep learning method presented in Alfradique et al. (2024). The fundamental component of the deep learning model comprises a neural network that analyses tabular data, using a Legendre Memory Unit (Voelker, Kajić & Eliasmith 2019) with an MDN. Unlike traditional neural networks that provide single value estimations, MDNs output conditional probability densities through a linear combination of individual probability distributions (components), chosen to be Gaussian distributions in our case. This approach enables a more comprehensive characterization of predictions and errors assessment.

The neural network output is a linear combination of C Gaussian kernels [$N(\mu_i, \sigma_i)$, where $\{\mu_i\}$ is the mean and $\{\sigma_i\}$ is the standard deviation] weighted by mixture coefficients $\{\alpha_i\}$. We impose, for the mixture coefficients, that $\sum_{i=1}^{C} \alpha_i = 1$ and $0 < \alpha_i < 1$. Therefore, the PDF can be written as

$$\text{PDF}(z) = \sum_{i=1}^{C} \alpha_i N(\mu_i, \sigma_i). \tag{1}$$

One PDF is assigned to each galaxy, i.e. the MDN model predicts different sets of $\{\mu_i\}$, $\{\sigma_i\}$, and $\{\alpha_i\}$ for different photometric inputs. This method was implemented to estimate photo-z's PDFs for the DELVE DR2 and LS DR10 catalogues. The architecture used for the DELVE photo-z is detailed in Teixeira et al. (2024). The LS DR10[3] photo-z were estimated using a similar architecture, with minor adjustments after a fine tuning considering the point statistics metrics such as reducing the width of the dense layers and reducing the number of components C in the mixture, from $C = 20$ for DELVE to $C = 6$ for Legacy. Input features for this process include *griz* magnitudes and colour indices (g–r, g–i, g–z, r–i, r–z, and i–z). The spectroscopic data used as training set for DELVE catalogues is

the same as the one presented in Teixeira et al. (2024), which was created from the cross-match between DELVE catalogue and several spectroscopic data available in different large sky surveys. For the Legacy Survey, we use the same training sample from the public photometric redshift catalogue (Zhou et al. 2020).

The Legacy Survey (LS) public data releases do not directly provide the apparent magnitude in photometric band. Therefore, we initially used the linear fluxes [columns `FLUX_{G,R,I,Z}`] to compute the magnitudes m by employing the conversion $m = 22\cdot5 - 2\cdot5\log_{10}(f)$, and derived the magnitude errors also from the inverse variances of the fluxes.[4] To mitigate star contamination, we adopted the same approach outlined by Palmese et al. (2023), applying colour cuts based on *Gaia* data, such as removing all known stars. Additionally, all the magnitudes were corrected for Milky Way extinction. Finally, we restrict our GW analysis to r band magnitudes to be lower than 21.

We conducted a series of selection cuts based on the photometry quality and properties to ensure the utilization of galaxies with the best possible detections. We excluded all objects with unphysical colours, retaining only those that satisfy the conditions

$$-1 > g-r, \quad r-i, \quad i-z < 4.$$

For the spectroscopic sample (refer to Teixeira et al. 2024 for a detailed list of the spectroscopic catalogues also used in this work), we restricted our objects to $0\cdot01 < z_{\text{spec}} < 1\cdot5$. After applying these cuts, our spectroscopic sample contains $\sim 2\cdot2$ M galaxies. We selected the training sample in order to have a uniform z_{spec} distribution of 0.01 to 1.0, and comprise all objects available with $z_{\text{spec}} > 1$ (which are few in number), resulting in 550 k (580 k for DR9) galaxies for training the model. The same approach of uniform training was also used in Legacy Survey publicly available photo-z (Zou et al. 2019) to avoid possible systematic bias towards oversampled regions.

We generate PDFs for galaxies in the test sample to validate the model by checking probabilistic and marginal calibration. Finally, we use the trained model to generate PDFs for galaxies in our target data sets. We explore the performance of the full PDF estimations by examining both the point estimates photometric redshifts and the calibration of their PDFs. For a given galaxy, the photo-z value is defined as its respective PDF's most probable value (peak). Fig. 2 presents the final photo-z distribution, found with the full photo-z PDF described in Appendix A, in the 90 per cent CI area of each GW event studied here. In order to highlight the overdensity regions, the uniform distribution in comoving volume $(\text{d}N/\text{d}z)_{\text{com}}$ was subtracted from the photo-z distribution $\text{d}N/\text{d}z$.

To evaluate the quality of the point estimates, we employ the following metrics:

(i) **Median bias**: the bias, defined as $z = z_{\text{phot}} - z_{\text{spec}}$, directly measures the deviation of our estimations from the target values (z_{spec}).

(ii) **Scatter**: the normalized median absolute deviation σ_{NMAD}, defined as

$$\sigma_{\text{NMAD}} = 1\cdot48 \times \text{median}\left(\frac{z - \text{median}(z)}{1 + z_{\text{spec}}} \right), \tag{2}$$

is a standard measurement of the bias scattering (Brammer, van Dokkum & Coppi 2008; Li et al. 2022; Lima et al. 2022). We aim

---

[3]We used the latest available version of DR10, namely 10.1.

[4]See the description of the `FLUX_IVAR_{G,R,I,Z}` flag in https://www.legacysurvey.org/dr9/description/photometry.
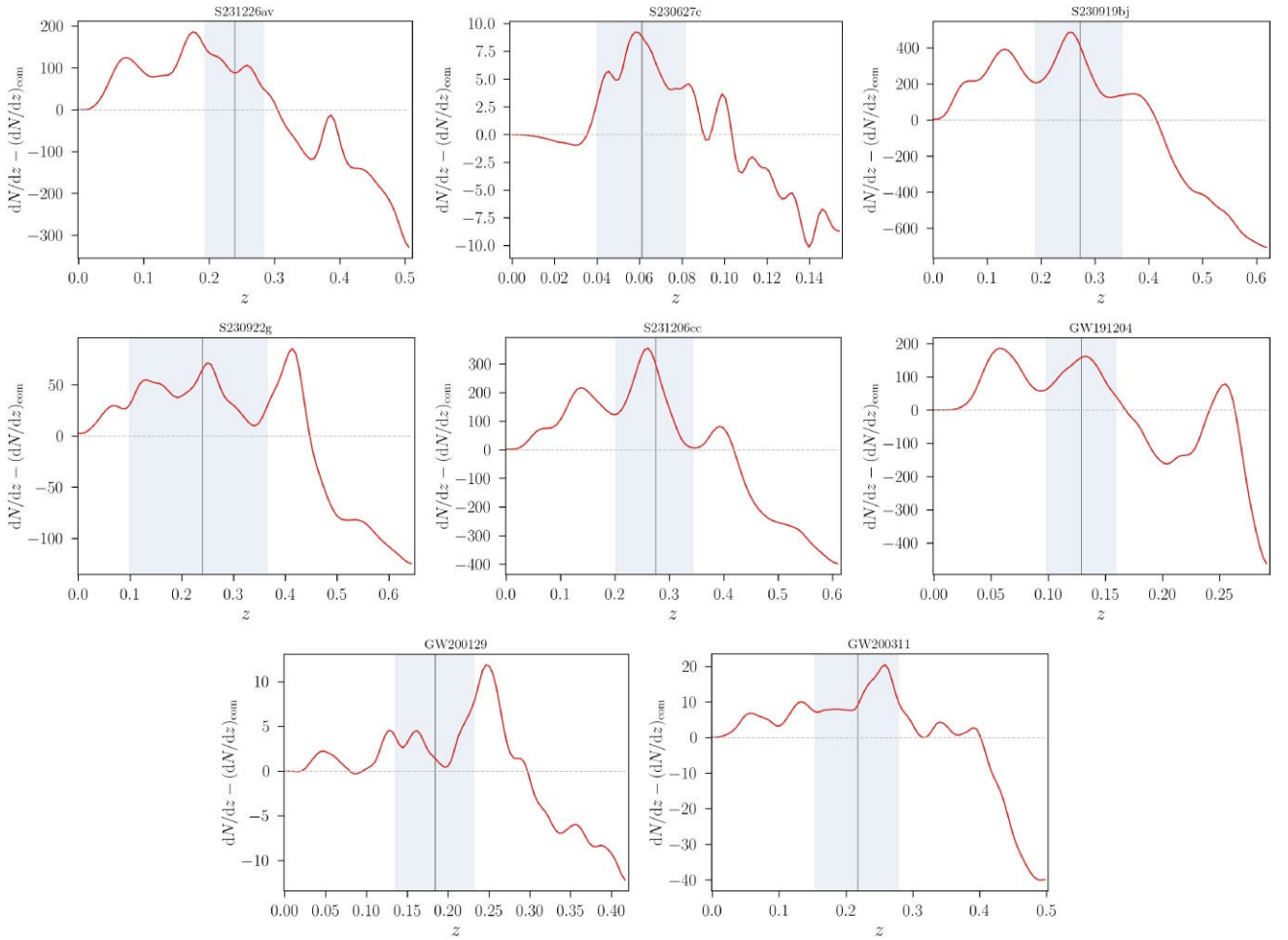
Figure 2. Redshift distribution of galaxies in the 90 per cent CI area of the five new dark siren events from O4a (S231226av, S230627c, S230919bj, S230922g, and S231206cc), and the three superevents discussed in Palmese et al. (2023) that have now been confirmed as GWs (GW191204, GW200129, and GW200311). To highlight the presence of overdensities and underdensities along the line of sight, the redshift distribution was subtracted with a uniform number density. The grey vertical lines represent the luminosity distance of each GW event marginalized over the entire sky, assuming an H_0 of 70 km s^{-1} Mpc^{-1}, and the shaded regions are the 1σ uncertainties considering the same H_0. These regions are only showed for reference.

for σ_{NMAD} to be as low as possible. The choice of σ_{NMAD} instead of σ_{68} is less sensitive to outliers.

(iii) **Outlier fraction**: outliers are defined as objects which

$$\frac{z}{1 + z_{\mathrm{spec}}} > 0.15. \tag{3}$$

We define η as being the fraction of outliers in any subsample of photo-z estimations. This definition of an outlier follows the same approach as adopted in Ilbert et al. (2006) and Lima et al. (2022).

All the aforementioned metrics were computed using the objects with spectroscopic correspondence inside the 90 per cent confidence area of the events S231206cc, S230919bj, S230922g, GW191204_171526, GW190924_021846, GW200129_065458, GW200202_154313, and GW200311_115853. Additionally, with the exception of the odds constraints, the analysis refers only to the objects that satisfy the restrictions imposed to the events' photo-z catalogues. Fig. 3 shows the median bias, scatter, and outlier fraction as a function of photo-z bins with a width of 0.025. We generated a curve representing the mean value of these metrics within each photo-z bin for each of the aforementioned GW events, using objects with available spec-z in each region. Subsequently, we

averaged these curves across all events (solid lines) and computed the standard deviation between the values in each photo-z bin (shaded regions).

Considering the measurements for both DELVE and LS surveys, we obtain σ_{NMAD} and an outlier fraction equal to 0.032 (0.016), 6.2 (2.4) per cent for DELVE (Legacy), respectively. These results show that our measurements using Legacy data outperform those obtained using DELVE data. We use the Legacy public photometric redshifts in Fig. 3 catalogues for performance comparison in the point-like metrics. The LS public photometric redshifts were computed using the Random forest method by Zhou et al. (2020). In this method, they used r-band magnitude, g–r, r–z, z–$W1$, and $W1$–$W2$ colours as input features. We compare our results with the public photo-z catalogue from LS-DR10 as well as the public redshifts from LS-DR9 (Zou et al. 2019) utilized in Palmese et al. (2023). On the left panel of Fig. 3 the MDN method exhibits an improvement of the averaged median bias in the $0.01 < z < 0.1$ interval, compared both to the DR10 public available data and the data used in Palmese et al. (2023), which is relevant since a considerable amount of the objects in the event skymaps lies in that interval. Conversely, for higher redshifts, both curves are compatible within the scatter. Throughout the entire range of photo-z, the bias for our method is very low and
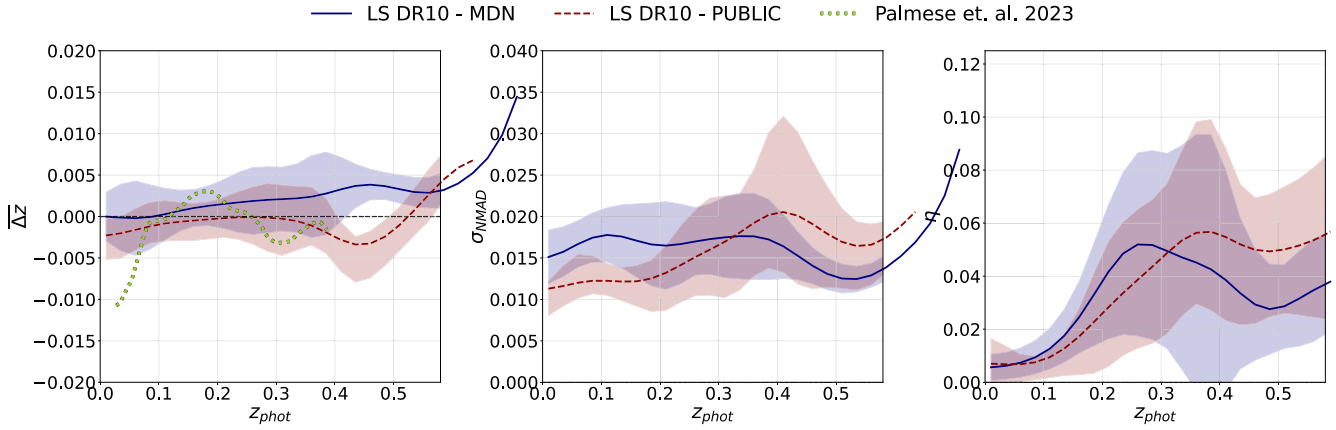
Figure 3. Mean values of the median bias ($\overline{\Delta z}$), σ_{NMAD}, and η as a function of 0.025 width photo-z bins. The solid lines depict the computed averages for all objects within each bin for each metric. The shadows represent the respective standard deviations.
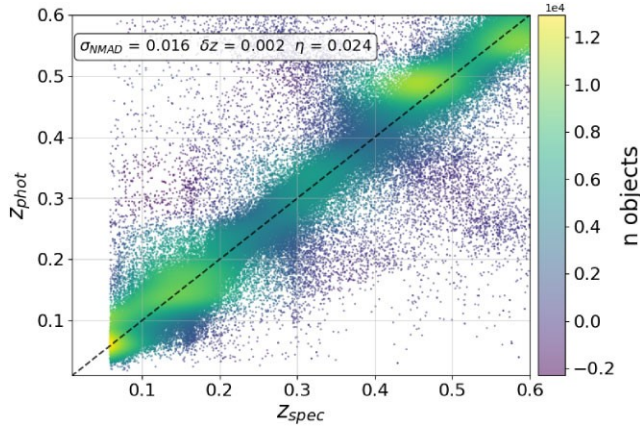

Figure 4. Photo-z's estimations versus the z_{spec} values. The intensity map represents the density of objetcs, and on the top of the figure, we displayed the point estimation metrics for the entire spec-z sample within the analysed GW and superevent regions.

does not exceed 0.007. It is also consistent with bias 0 within the scatter until $z \sim 0.4$. However, it tends to be positive everywhere, indicating a weak cumulative overestimation effect. Additionally, in Fig. 3 the scatter and outlier fraction align with publicly available redshifts values, thereby affirming the reliability of our method in generating accurate photo-z estimations. Fig. 4 presents the spec-z sample versus photo-z obtained.

To validate the PDFs, we investigated the following metrics:

(i) **PIT**: the probability integral transform (PIT; Dawid 1984) represents the cumulative density function of the PDF up to the z_{spec} value for each galaxy. We can assess the quality of the PDFs by examining the PIT distribution for a representative sample of galaxies. The distribution of PIT values is expected to be uniform between 0 and 1 (Mucesh et al. 2021), indicating that the z_{spec} values can be considered as random events generated from independent PDFs. A slope in the PIT distribution indicates a bias in the estimations, while the concavity of the distribution reveals whether the PDFs are over or underdispersed (Polsterer, D'Isanto & Gieseke 2016).

(ii) **Odds**: the odds (Benitez et al. 2014) measure the degree of confidence in the photo-z estimate derived from a given PDF. It is defined as the probability of the redshift lying within an interval around the photo-z value. For a given PDF, we can compute:

$$\text{Odds} = \int_{z_{phot}-0.06}^{z_{phot}+0.06} \text{PDF}(z)dz. \tag{4}$$

The interval of 0.12 around the photo-z value was chosen based on the approach used by Coe et al. (2006) in their analysis of photometric redshift estimates for galaxies in the Hubble Ultra Deep Field (Beckwith et al. 2006). The ideal distribution of odds for a galaxy sample well-represented by the training set should exhibit a pronounced peak near 1. However, it could also represent a distribution of underdispersed PDFs. For this reason, we have to analyse the odds and PIT distribution simultaneously, in order to infer the quality of the PDFs.

(iii) **Coverage diagnostic**: the coverage diagnostic [or high posterior density (HPD) diagnostic] stands as a well-established metric commonly employed to assess the quality of credible regions generated by simulation-based inference algorithms. Additionally, it provides a means to evaluate the accuracy of the estimated redshift distribution, as outlined in Dalmasso et al. (2020). The fundamental concept involves evaluating the probability that a specified credible region within the inferred distribution contains the true value. This assessment provides insights into whether the estimated distribution is overconfident, calibrated, or underconfident (Hermans et al. 2022). The coverage diagnostic was executed by selecting the pair (spec-z, PDF) and sampling 1000 values from the photo-z's PDF, thereby generating a frequency distribution with 1000 bins spanning the range from 0 to 1.

First, a credible region is defined for the estimated distribution using the highest density regions. This region represents the smallest area that contains at least $100(1-\alpha)$ per cent of the mass of the inferred photo-z distribution, establishing an interval for a given credibility level $(1-\alpha)$. The expected coverage is the frequency with which the true parameter (spec-z) value falls within this highest density region, essentially indicating how often it falls inside the calculated interval. If our model produces well calibrated distributions, we expect the spec-z value should be contained inside the calculated interval from the $(1-\alpha)$ HPD region of the estimated distribution exactly $(1-\alpha)$ per cent of the time. If the coverage probability is less

than the $(1 - \alpha)$ credibility level is the sign of underestimation of the PDF's variance, and it could lead to unreliable approximations since it excludes physical values of photo-z. Conversely, if the coverage probability is larger than the $(1 - \alpha)$ credibility level, then this indicates that the estimated PDFs are overestimating their variance, in average.

The middle panel of Fig. 5 shows the odds distribution for the same set of galaxies present on the point-estimate analysis, while the left panel illustrates the PIT distribution for a subset of these objects satisfying the constraint odds > 0.7. The latter exhibits a marginal concavity at lower values of PIT and a negative slope throughout higher values. Although not pronounced, these features collectively suggest a tendency towards positive bias, implying a slight overestimation of the most probable photo-z's values. As discussed above, there is trending in the median bias curve (Fig. 3) of overestimating photo-z. This overestimation effect is further illustrated in Fig. 4, particularly evident around photo-z equal to 0.48. Despite this anomaly at the higher redshifts, the scattering of photo-z remains well-behaved, as expected. Consequently, the inclination of the PIT distribution might be attributed to a localized systematic bias rather than indicating any inherent global bias within the model. Although this could impact the analysis, it is worth stressing that our Bayesian analysis performs a bias correction based on this curve in the same manner as previous work by Alfradique et al. (2024). A similar shape of PIT distribution was also encountered by D'Isanto & Polsterer (2018) using MDN to estimate redshifts from images in the Sloan Digital Sky Survey Data Release 9.

The odds distribution shows a gradual increase in frequency with the odds values. This result suggests the absence of a systematically overconfident or overdispersed PDFs, indicating a high level of confidence in the estimated PDFs. From Fig. 6, a perfectly calibrated estimated distribution aligns with a diagonal line (dashed black line), indicating an expected coverage probability matching the credibility level. We notice that the expected coverage curve matches the diagonal, indicating that our model produces well-calibrated photo-z PDFs. This linear relation confirms the reliability of our methodology.

In summary, our methodology estimates photo-z with an accuracy comparable to those publicly available in LS-DR10, while also generating well-calibrated PDFs. It is noteworthy that our approach can predict redshifts in objects with detections in all *g, r, i, z* bands, and in any combination of three bands containing detections in *g* band.

3 METHOD

Here, we outline the statistical methodology known as the dark siren approach, first idealized by Schutz (1986) and revisited in several works (Holz & Hughes 2005; MacLeod & Hogan 2008; Del Pozzo 2012). This methodology was initially detailed in Chen et al. (2018) and subsequently modified by Soares-Santos et al. (2019), Palmese et al. (2020), (2023), and Alfradique et al. (2024). The method consists of using the Bayesian formalism to infer the Hubble constant parameter through the GW detection data d_{GW} and the d_{EM} set of photo-z measurements of the possible host galaxies made using the deep learning algorithm presented in 2.3. As the GW and Electromagnetic (EM) measurements are done independently, the joint GW and EM likelihood can be defined as the product of the two individual likelihood, $p(d_{GW}, d_{EM}|H_0) = p(d_{GW}|H_0) p(d_{EM}|H_0)$.

From the Bayesian framework, the H_0 posterior of one GW event can be written in the final form as

$$p(H_0|d_{GW}, d_{EM}) \propto \frac{p(H_0)}{\beta(H_0)} \sum_i \frac{1}{Z_i} \int dz_i\, dz$$
$$\times\, p(d_{GW}|d_L(z_i, H_0), \hat{\Omega}_i) p_i(d_{EM,i}|z_i, z)$$
$$\times\, p(z) \frac{r^2(z_i)\psi(z_i)}{H(z_i)}. \tag{5}$$

Where $p(H_0)$ is prior on H_0 which we assume to be flat over the range $[20, 140]$ km s^{-1} Mpc^{-1}, $\beta(H_0)$ is the normalization factor that describes the selection effects in the measurement process, $r(z)$ is the comoving distance, $H(z) = H_0 [\Omega_m(1+z)^3 + 1 - \Omega_m]^{1/2}$ is the Hubble parameter in a flat Lambda cold dark matter (ΛCDM) model, $Z_i = \int p(d_{EM}|z_i) r^2(z_i)/H(z_i) dz_i$ are evidence terms that normalize the posterior where $p(d_{EM}|z_i)$ is the EM likelihood marginalized over the photo-z bias z, and $p(z)$ is the prior on the photo-z bias that we measure from our photo-z validation sample (see Fig. 3 in Section 2.3). The last term on the equation above comes from the marginalization over the galaxies' redshifts and sky positions, assuming that the galaxies are uniformly distributed in comoving volume with a merger rate evolution, $\psi(z)$, following the Madau–Dickinson cosmic star formation rate (Madau & Dickinson 2014). In the following paragraphs, we will detail how the function $\beta(H_0)$ was computed.

The GW and EM likelihoods are written as, respectively:

$$p(d_{GW}|d_L(z, H_0), \hat{\Omega}_i) \propto \frac{p(\hat{\Omega}_i) N(\hat{\Omega}_i)}{2\pi\sigma(\hat{\Omega}_i)} \exp\left[-\frac{(d_L - \mu(\hat{\Omega}_i))^2}{2\sigma^2(\hat{\Omega}_i)}\right],$$
$$p(d_{EM}|z_i, z) = \sum_i p(z_i|z_i) p(z_i|z\Omega_i), \tag{6}$$

where, we explicitly consider the dependence of cosmological parameters on the luminosity distance, as measured by GW, and the solid angle $\hat{\Omega}_i$ corresponding to each observed galaxy *i*. Following Singer et al. (2016b), the GW likelihood is approximated by a Gaussian function. The EM likelihood is the product of the probability distribution function of the photometric redshift z_k for each *k* galaxy, where we consider the correction of the shifted in redshift for the photo-z biases, z, in the data. This phenomenon arises from the absence of a uniform distribution in redshift or colour beyond a certain magnitude or colour selection threshold, which causes the deep learning algorithm to oversample the redshift around the distribution peaks, thereby introducing systematic biases. The individual H_0 posterior distributions found in this work are presented (in colours) in Fig. 7.

The selection effects were computed based on the criteria outlined in Chen et al. (2018) and Gray et al. (2020). This involves the joint GW–electromagnetic likelihood, which is marginalized over all conceivable GW and EM data. Assuming that the events are isotropically distributed on a large scale, this term can be written in its compact form as:

$$\beta(H_0) = \int_0^{z_{max}} p_{sel}^{GW}(d_L(z, H_0)) p(z) dz, \tag{7}$$

where p_{sel}^{GW} is the probability of a GW event at a given luminosity distance d_L to be detected, z_{max} is the maximum true redshift at which we can detect the host galaxies, and $p(z)$ is the galaxy catalogue distribution that for simplicity, we assume to follow a uniformly distributed in comoving volume including the merger rate evolution described by the Madau–Dickinson star formation rate (Madau & Dickinson 2014). The function p_{sel}^{GW} is equal to one if all events in a given redshift z satisfy the detection condition, i.e. the detector network signal-noise-ratio (SNR) is > 12 and the localization volume satisfies our selection criteria ($A_{50\,per\,cent} < 1000$ deg^2 and
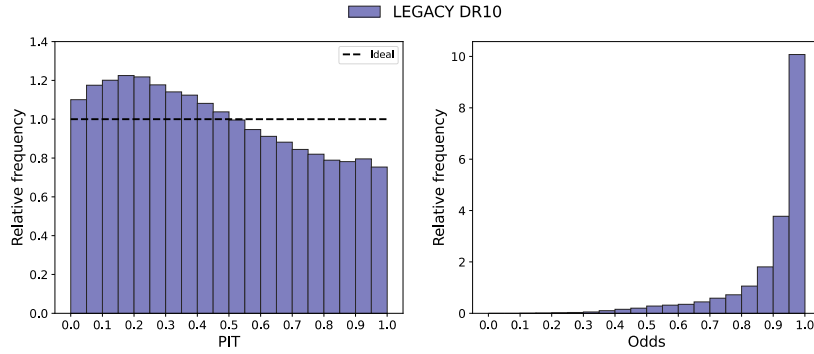
Figure 5. PIT and odds distributions for the objects with spectroscopic redshifts within the event regions. The odds distributions (right panel) remains for all objects (around 245 k galaxies), meanwhile the PIT distribution (left panel) remains for the objects which obey odds > 0.7 beyond the errors constrains employed to the H_0 posterior estimations (around 130 k galaxies).
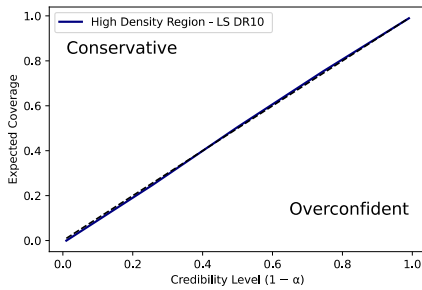

Figure 6. Coverage test results, where the solid line represents the resulting curve for our estimated PDFs and the dashed line represents the ideal case of perfectly calibrated PDFs in terms of credibility.

$d_L < 1500\,\mathrm{Mpc}$), and zero if none of the events located in a redshift z satisfies those conditions. Therefore, the $p_{\mathrm{sel}}^{\mathrm{GW}}$ function is, in sum, an efficiency curve, i.e. a smooth function that falls from 1 to 0 over the redshift range for each H_0 value. It is important to mention that in the definition of $\beta(H_0)$ the selection effects of EM data are incorporated assuming that all possible host galaxies are detected up to a maximum redshift, where for the Legacy catalogue the value of $z_{\mathrm{max}} = 0.6$ was adopted.

The selection function was modelled in a similar way in the other previous dark siren works (Palmese & Kim 2020; Palmese et al. 2023; Alfradique et al. 2024) by simulating 30 000 BBH mergers using BAYESTAR software (Singer & Price 2016; Singer et al. 2016a, 2016b) with the frequency domain approximant IMRPhenomD. We simulate the BBH mergers assuming that they follow a power law plus peak mass distribution with the same parameters as described in Bom & Palmese (2024). The spins distribution follow a uniform distribution between $(-1, 1)$, and a uniform distribution in comoving volume assuming a dependency with the merger rate evolution described by the Madau–Dickinson star formation rate and fixing the Planck 2018 cosmology for 20 different H_0 values within our prior range. All the 30 000 injected events passed by the matched filter analysis, where we computed the SNR assuming the O4 sensibility curve as published by LIGO in Document P1200087[5] for the Hanford and Livingston LIGO network detectors. The network SNR defines the detection condition above 12 and at least two detectors have a single detector SNR above 4. A Gaussian noise was added in all the measurements. Lastly, the BAYESTAR skymaps were reconstructed

for each detection event, where we assume a luminosity distance that follows $\propto d_L^2$. The last selection cut that we should consider is that events serving as dark sirens have at least 70 per cent of their 90 per cent CI comoving volume covered by the Legacy Survey. As noted by Palmese et al. (2023), this selection cut can be ignored since the GW antenna pattern is not correlated with the survey sky footprint, which corresponds to selecting events isotropically. This is in agreement with the way that the selection function is computed, so we do not expect a strong H_0 dependence. Here, we also ignore this selection effect. However, it is worth noting that the value of H_0 changes the redshift range associated with a given luminosity distance horizon, consequently changing the SNR distribution of the source. Therefore, this may affect the size of the credible localization regions causing an impact on the validation of the survey's coverage condition. This effect will be investigated in more detail in a future work.

We analysed the effect that our choice of the power law plus peak mass distribution has on the selection function when compared to the result assuming a power-law distribution. We found an average relative error of 7 per cent, which reflects a slight variation in the individual H_0 posteriors, where we observed a deviation of 1.2 km s^{-1} Mpc^{-1} at the peak of the H_0 posterior only for the event S230627c (the other events remained unchanged). We also examined the impact of our choice on the rate of BBH mergers by comparing our selection function with that found under the assumption that the merger rate density follows $(1 + z)^\kappa$, with κ fixed at the value inferred in Abbott et al. (2023a). The results indicate a negligible relative error of 0.7 per cent, implying a maximum variation of 0.9 km s^{-1} Mpc^{-1} on the standard deviation of the individual posteriors. Therefore, the different choices of BBH population into the selection functions have a negligible effect in the combined H_0 constraint by standard dark sirens, lower than an order of magnitude of the present errors both in individual posteriors and the combined constraint.

We also accounted for a volume limited sample, we adopted a method similar to Palmese et al. (2020), (2023) and Alfradique et al. (2024). Initially, we set a maximum redshift of interest for each GW event based on its 90 per cent confidence interval in luminosity distance, averaged over the entire sky. This upper bound is translated into a maximum redshift using the highest H_0 value considered in our prior. For each event, we calculate the absolute magnitude that represents the galaxy sample's limiting apparent magnitude at that redshift. We then remove galaxies with absolute magnitudes lower than this threshold. Notably, while we employ a fiducial CDM cosmology to determine these magnitudes, the H_0 dependence does

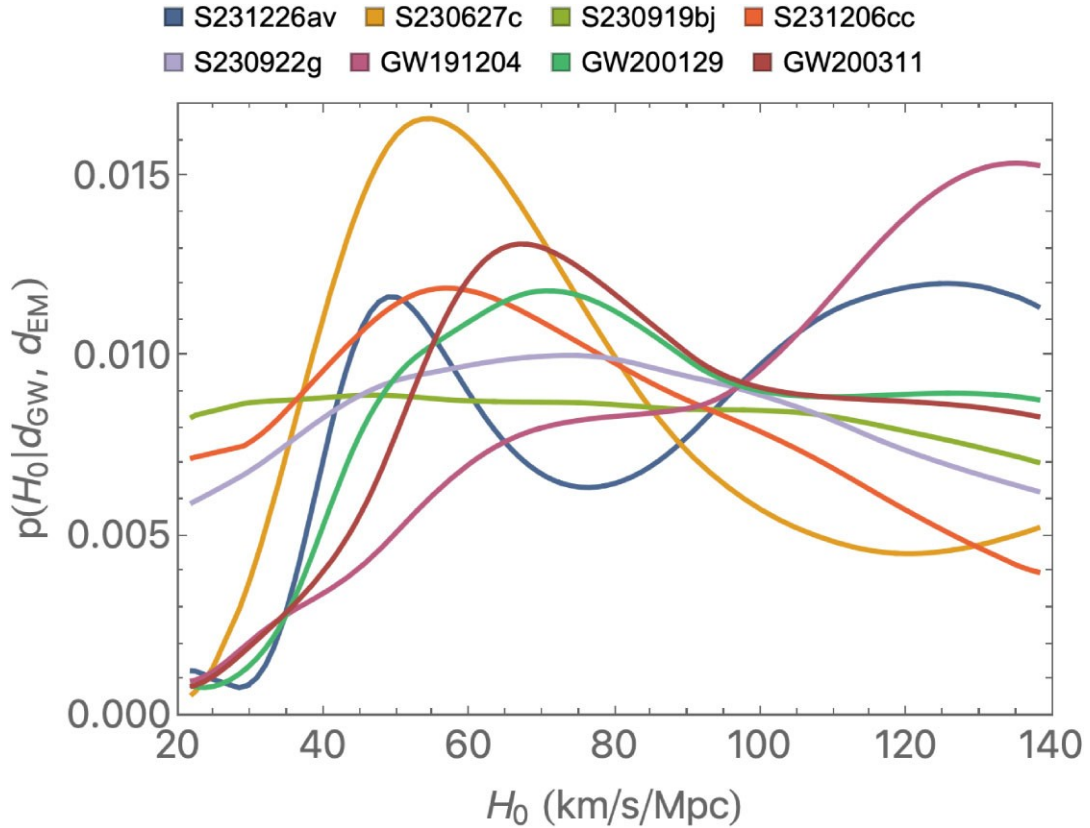[5]The data are available at https://dcc.ligo.org/LIGO-T2000012/public.

Figure 7. The Hubble constant posterior distributions for each dark siren considered in this work, which were found using the galaxies from Legacy Survey.

not affect the results because the threshold value and the galaxies' absolute magnitudes scale with H_0 in the same way (Palmese et al. 2020).

4 RESULTS AND DISCUSSION

In this section, we present the H_0 posterior produced using the dark siren methodology described in the previous section with the Legacy Survey photo-z's PDFs constructed using the deep learning methodology from A24 and also described in Section 2.3. We computed the H_0 posterior for five GW candidate BBH events detected in the LIGO O4a run and reanalysed three events from P23 with the updated skymaps, as they were not yet confirmed as GW events (see Section 2). The 68 per cent CI of the five O4a superevents represent about 69 per cent to 91 per cent, depending on the event analysed, of the 68 per cent CI of the prior width. The most constraining event from O4a is S230627c, because of its better localization, galaxy catalogue coverage, and quality of the photo-z at the relevant redshift ranges. For the revisited GW events, despite the increase in 90 per cent CI volume, all three events previously studied with the Bayestar maps in P23 present an improvement of ~ 18 per cent, 5 per cent, and 9 per cent in 68 per cent CI, respectively. The main cause of these improvements is the quality of our photo-z measurement (see Fig. 3) as discussed in detail in Section 2.3.

The majority of the H_0 posteriors show a clear peak, except for S231226av, which exhibit multiple peaks in different ranges of H_0 (raging from 49 to 136 km s^{-1} Mpc^{-1}). As already mentioned in previous works (Soares-Santos et al. 2019; Palmese et al. 2020, 2023; Alfradique et al. 2024), these peaks are associated with the

presence of overdensities regions along the line of sight (see Fig. 2 in Section 2.3). In the dark siren methodology, this is indicative of the host galaxy being more likely to be in these regions. Additionally, it is evident that certain events display a more prominent peak, which may stem from the presence of a prominent overdensity within an improved localization. In other words, the need for marginalization over fewer galaxies and/or the fact that the galaxies live at similar redshift, lead to a more informative H_0 posterior.

The H_0 posterior for GW191204 presents a considerable probability at $H_0 \approx 130$ km s^{-1} Mpc^{-1}, which is associated with an overdensity of galaxies at $z \approx 0.25$ (see Fig. 2 in Section 2.3). We can also see that this H_0 posterior begins to decrease at the high-H_0 end, due to the absence of galaxies at high redshift (the galaxy density in Fig. 2 is negative at the high-z end, indicating the presence of an underdense region).

The combination of the 15 dark sirens from O1 to O4a is depicted in Fig. 8 by the green line and detailed in Table 2. The maximum a posteriori and the 68 per cent CI is $H_0 = 70.4^{+13.6}_{-11.7}$ km s^{-1} Mpc^{-1}. As it can be seen, our results are consistent within 1σ with Planck (Planck Collaboration VI 2018) and the local Cepheid–supernova distance ladder (Riess et al. 2022). This result agrees with the latest dark sirens study using the BBH population and catalogue method (Abbott et al. 2023c), that found $H_0 = 67^{+13}_{-12}$ km s^{-1} Mpc^{-1} for 47 dark sirens from the third LVK GW transient catalogue (Abbott et al. 2023b) and the GLADE+ galaxy catalogue. A possible justification for the similar uncertainties, despite the difference in number of the events used, is the completeness and inhomogeneity of the GLADE+ catalogue at the redshift of interest of the studied dark sirens. This contributes to the dependence between the final H_0 constrain and the BBH population assumptions observed in Abbott
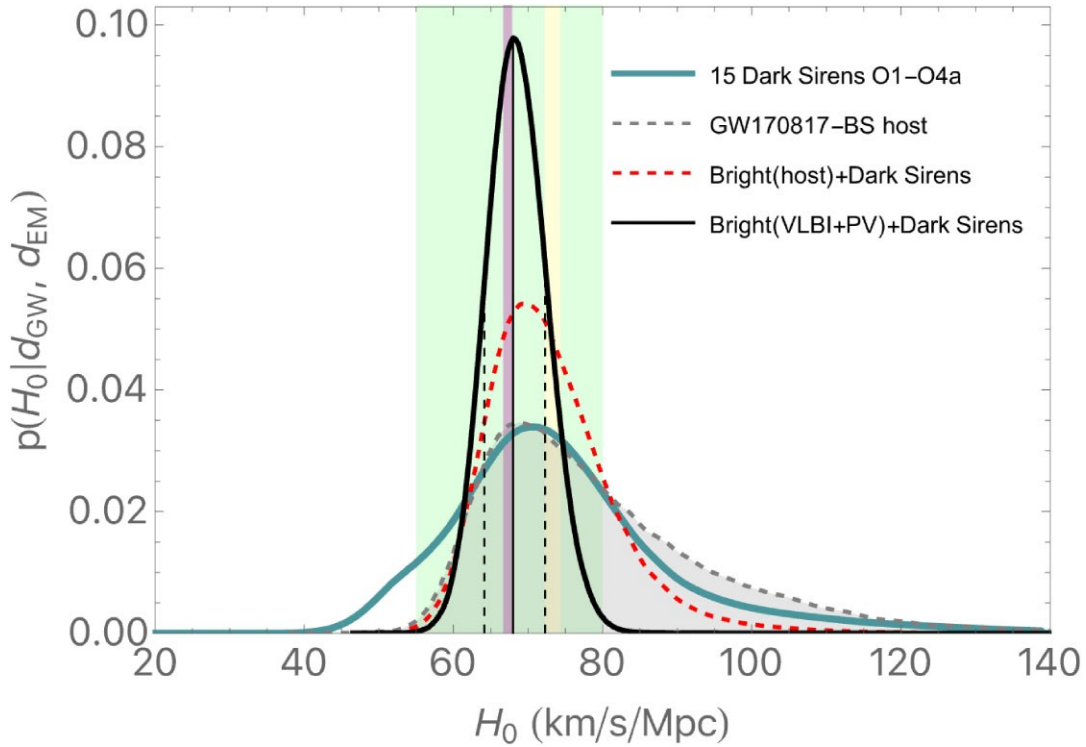
Figure 8. Hubble constant posterior distributions of the 15 dark sirens of O1–O4 observations. The dark green line shows the result from the combination of all the 15 dark sirens. The shaded grey posterior represents the GW170817 standard siren result, where only GW data are used, adapted from Nicolaou et al. (2020), which corrects the peculiar velocity to the constraint. The joint constraint from both the bright (i.e. GW170817, with the results from Nicolaou et al. 2020) and the dark standard sirens is shown by the dotted red line. The black line is the 15 dark siren H_0 posterior (green line) combined with the results for the GW170817 bright siren found in Mukherjee et al. (2021b) using the GW + VLBI data with the peculiar velocity corrections, and the vertical dashed lines show the 68 per cent region for this posterior. For reference, we show the 1σ Planck Collaboration VI (2018) (shaded pink), Riess et al. (2022) (R22; shaded light yellow), and the GWCT-3 (47 dark sirens + BBH population; Abbott et al. 2023c) constraints on H_0.

et al. (2023b), and minimizes the galaxy method contribution to the final constraint. Furthermore, we also combine our dark siren results with the bright siren GW170817 from Nicolaou et al. (2020), which extended the analysis of the peculiar velocity effect in the H_0 measure presented in Abbott et al. (2017c). From this analysis, we find $H_0 = 69.20^{+8.98}_{-5.85}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The dark siren information provides a notable reduction of 41 per cent in the 68 per cent CI and a 7 per cent improvement in the relative precision. We further combine our final dark siren H_0 posterior with the bright siren analysis for GW170817 presented in Mukherjee et al. (2021b), which corrects the peculiar velocity contribution in the measure of H_0 presented in Hotokezaka et al. (2019). In that work, the superluminal motion measured by the very large baseline interferometry (VLBI; Mooley et al. 2018) and the afterglow data were used to measure the inclination angle of the GW170817 event. The addition of an independent EM observation helps break the degeneracy between the distance and the inclination angle, which is one of the main contributors to the uncertainty of the GW distance measurement. We find that the dark sirens information causes a reduction in the relative uncertainty from 18 per cent to 11 per cent, leading to our final constraint: $H_0 = 68.00^{+4.28}_{-3.85}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Furthermore, we combine our 15 dark sirens result with a more recent H_0 measurement (Palmese et al. 2024), which combines the GW170817 measurements with the electromagnetic counterpart associated with afterglow rather than the superluminal jet motion as done in Mukherjee et al. (2021b). This result presents an agreement of 1.4σ (σ is uncertainty of H_0 from Planck data) with the H_0 measurements of Planck (Planck

Collaboration VI 2018), indicating a reduction of 0.1σ of the original result.

To guarantee the use of the most appropriate galaxy catalogue in terms of coverage of the localization volumes of the gravitational GW events treated here and the quality of the estimated photo-z's, we also analysed the measurements of H_0 for our dark sirens sample only using the DELVE catalogue. This allows us to evaluate if the H_0 measurement and photo-z quality outperform our results obtained using the Legacy Survey. The photometric redshifts for the DELVE galaxies were computed with the deep learning methods presented in A24, adapted for the redshift range of interest here. The results of the individual H_0 posterior distribution found with the Legacy and DELVE galaxy catalogues are in agreement; however, the combination of the 15 dark sirens found with DELVE leads to a higher H_0 peak value (approximately $2.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) and a lower uncertainty of H_0 ($\sigma_{H_0} \sim 0.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) compared to those presented in Table 2. This dark siren result combined with the bright siren analysis from Mukherjee et al. (2021b), yields an H_0 measurement of $68.27^{+4.20}_{-3.91}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, which is also in agreement with the Legacy Survey result. Although the DELVE results present a slightly more precise measurement than those found with the Legacy Survey, the quality of the metrics of the photo-z measurements did not show the same robustness. The difference in reliability of photometric redshift measurements between DELVE and Legacy Survey may be attributed to the absence of coadded data in DELVE, leading to lower SNR compared to the Legacy Survey catalogue. Furthermore, Legacy offers uniform sky coverage, enabling analysis

Table 2. Hubble constant measurements using GW standard sirens from this work and previous works. H_0 values and uncertainties are given in km s^{-1} Mpc^{-1}, and H_0 priors are flat, unless otherwise stated. The uncertainty from the flat prior only is derived by assuming the same H_0 maximum found in the analysis. Quoted uncertainties represent 68 per cent HDI around the maximum of the posterior. The '$\sigma_{H_0}/\sigma_{\rm prior}$' column shows the 68 per cent CI from the posterior divided by 68 per cent CI of the prior width. (1) Finke et al. (2021), (2) Palmese et al. (2023), (3) Abbott et al. (2023c), (4) Alfradique et al. (2024), (5) adapted from Nicolaou et al. (2020), (6) Palmese et al. (2024), (7) Mukherjee et al. (2021b), and (8) this work.

Event(s)	Method(s)	H_0 km s^{-1} Mpc^{-1}	σ_{H_0} km s^{-1} Mpc^{-1}
O1–O3a[1]	Catalogue	$67.3^{+27.6}_{-17.9}$	22.5 (34 per cent)
O1–O3 – 8 dark sirens[2]	Catalogue	$79.8^{+19.1}_{-12.8}$	15.8 (20 per cent)
O1–O3 – 47 dark sirens[3]	Catalogue + BBH population	67^{+13}_{-12}	12.5 (18 per cent)
O1–O3 – 47 dark sirens[3]	BBH population	67^{+14}_{-13}	13.5 (20 per cent)
O1–O3 – 10 dark sirens[4]	Catalogue	$76.00^{+17.6}_{-13.4}$	15.6 (20 per cent)
GW170817[5]	Bright (v_p corrected)	$68.80^{+17.3}_{-7.6}$	12.5 (18 per cent)
GW170817[6]	Bright (*Chandra* + *HST* + VLA)	$75.46^{+5.34}_{-5.39}$	5.36 (7 per cent)
GW170817[7]	Bright (v_p + VLBI)	$68.3^{+4.6}_{-4.5}$	4.6 (7 per cent)
O1–O4a – 15 dark sirens[8]	Catalogue	$70.4^{+13.6}_{-11.7}$	12.6 (18 per cent)
O1–O4a – 15 dark + 1 bright sirens[8]	Bright (host) + Catalogue	$69.2^{+9.3}_{-5.6}$	7.5 (11 per cent)
O1–O4a – 15 dark + 1 BS (EM)[8]	Bright (*Chandra* + *HST* + VLA) + Catalogue	$74.3^{+5.1}_{-4.9}$	5.0 (7 per cent)
O1–O4a – 15 dark + 1 BS (EM)[8]	Bright (v_p + VLBI) + Catalogue	$68.0^{+4.4}_{-3.8}$	4.1 (6 per cent)

of events across both the Southern and Northern skies. Given that the precision of photometric redshifts primarily influences H_0 measurements, we selected the Legacy results as our primary outcome.

We also investigated the effect of adding galaxy fakes to the galaxy catalogue for the S230919bj event. The presence of the galaxy fakes causes the peak of the posterior distribution to be slight shifted (a per cent-level change) and the uncertainty is reduced by about 0.3 per cent. Concerning the combined result of the 15 dark sirens, the presence of the galaxy fakes leads to a 1 per cent increase in precision, and the result combined with GW170817 remains unchanged, which demonstrates the small impact that galaxy fakes have on the final conclusions of this work.

As other dark siren measurements, our H_0 measurement are valid under the flat CDM model and also presents a dependency on the background cosmology, since the events go beyond $z \sim 0.1$ where changes to Ω_m and other cosmological parameters have a more significant impact on H_0 estimates. Therefore, we check if our results change with the choice of the Ω_m value within the 5σ interval found by the CMB measurement (Planck Collaboration VI 2018). There is a minor shift, with a relative difference of less than 4 per cent in the peak of the H_0 posterior distributions, and a relative difference less than 1 per cent in the 68 per cent CI. Therefore, the H_0 constraints presented here do not depend on the value of Ω_m, as long as it agrees with the Planck constraints.

The H_0 posteriors presented in this work were computed considering the full redshift PDF, derived through the deep learning algorithm described in Section 2.3, for each of the possible host galaxies. The application of a photo-z PDF ensures the use of a more reliable galaxy distribution, first proposed in Palmese et al. (2020). This treatment is different from other previous dark sirens works (LIGO Scientific Collaboration & Virgo Collaboration 2019; Soares-Santos et al. 2019; Abbott et al. 2023c) that assumed a Gaussian approach for these distributions. In Appendix A, we discuss the comparison between the results of these different methodologies; the H_0 posteriors' behaviour is slightly different, being more significant for closer events where the effect of marginalization over thousands of galaxies is minimized. As already noted by Palmese et al. (2020) and also seen here, the use of the Gaussian approximation causes

the distribution of galaxies dN/dz to be smoothed out, resulting in a flatter H_0 posterior. Despite these differences, the H_0 measurements are consistent with each other.

5 CONCLUSIONS

In this contribution, we use the data from the current best-localized and covered GW events from LVK O4a observing run to derive a dark siren measurement of the Hubble constant using the galaxy catalogue method and precise photometric redshifts. We obtained $H_0 = 70.4^{+13.6}_{-11.7}$ km s^{-1} Mpc^{-1}, i.e. a $\sim$ 18 per cent uncertainty on H_0 from dark sirens based on the catalogue method alone and a total of 15 sirens. This is, at the best of our knowledge, an unprecedented precision for the catalogue method and for the dark siren approach.

We combine our results from 15 dark sirens with recent constraints over the one bright standard siren available GW170817, considering the constraints to the viewing angle from VLBI and the host galaxy's peculiar velocity (Mukherjee et al. 2021b). We obtained $H_0 = 68.00^{+4.28}_{-3.85}$ km s^{-1} Mpc^{-1}, representing a 6 per cent measurement of H_0, which reduces the previous constraint uncertainty by approximately 10 per cent. We note that the precision and uncertainty of the 15 dark sirens are similar to the GW170817 constraint without the viewing angle constraints from electromagnetic observations. It is worth noting that the current results are derived under the assumption of a flat CDM scenario.

Our current results emphasize that a combination of well-localized dark sirens and high-quality photometric redshifts can achieve a competitive H_0 constraint from GWs. In particular, considering the absence of high confidence BNS merger detections during O4a, the BNS merger rate can be in the lower-end of previous estimations. Therefore, the number of well-localized BBHs detections could be one order of magnitude higher than that of BNS. Furthermore, neutron star–black hole (NSBH) systems do not present substantial promise as multimessenger sources (Biscoveanu, Landry & Vitale 2023). None the less, it is prudent to acknowledge that these detection rates may change significantly in the near future, and the emergence of a singular, observable BNS event with an electromagnetic counterpart could potentially offer more compelling constraints than a dozen dark standard sirens.

The current constraints from dark sirens by the catalogue method only, achieve a precision of ∼ 18 per cent, and in combination with bright sirens and additional constraints on the viewing angle, can achieve ∼ 6 per cent. As the number of dark siren events increases and we get closer to the level of statistical precision required to arbitrate the Hubble tension of ∼ 2 per cent, detailed studies to address potential systematics not included in this work should be carried out, especially considering different formation channels for BBH populations and catalogue depth (Gray et al. 2020; Mastrogiovanni et al. 2023).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abbott B. P. et al., 2017a, Phys. Rev. Lett., 119, 161101
Abbott B. P. et al., 2017b, Phys. Rev. Lett., 119, 141101
Abbott B. P. et al., 2017c, Nature, 551, 85
Abbott B. et al., 2019, Phys. Rev. X, 9, 031040
Abbott R., 2020, ApJ, 896, L44
Abbott R. et al., 2021, Phys. Rev. X, 11, 021053
Abbott R. et al., 2023a, Phys. Rev. X, 13, 011048
Abbott R. et al., 2023b, Phys. Rev. X, 13, 041039
Abbott R., Abe H., Acernese F. et al., 2023c, ApJ, 949, 37
Abdalla E. et al., 2022, J. High Energy Astrophys., 34, 49
Alfradique V. et al., 2024, MNRAS, 528, 3249 (A24)
Alves L. M. B., Sullivan A. G., Yang Y., Gayathri V., Márka Z., Márka S., Bartos I., 2024, MNRAS, 531, 3679
Andreoni I. et al., 2022, ApJS, 260, 18
Ashton G., Hübner M., Lasky P. D., Talbot C. et al., 2019, ApJS, 241, 27
Beckwith S. V. et al., 2006, AJ, 132, 1729
Benitez N. et al., 2014, preprint (arXiv:1403.5237)
Biscoveanu S., Landry P., Vitale S., 2023, MNRAS, 518, 5298
Bishop C., 1994, Aston Univ., Birmingham
Bom C. R., Palmese A., 2024, Phys. Rev. D, 110, 083005

Bom C. R. et al., 2024, ApJ, 960, 122
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Cabrera T. et al., 2024, preprint (arXiv:2407.10698)
Chen H.-Y., Fishbach M., Holz D. E., 2018, Nature, 562, 545
Cigarrán Díaz C., Mukherjee S., 2022, MNRAS, 511, 2782
Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, AJ, 132, 926
D'Isanto A., Polsterer K. L., 2018, A&A, 609, A111
Dainotti M. G., De Simone B., Schiavone T., Montani G., Rinaldi E., Lambiase G., 2021, ApJ, 912, 150
Dalmasso N., Pospisil T., Lee A., Izbicki R., Freeman P., Malz A., 2020, Astron. Comput., 30, 100362
Dawid A. P., 1984, J. R. Stat. Soc. Ser. A Stat., 147, 278
Del Pozzo W., 2012, Phys. Rev. D, 86, 043011
Dey A. et al., 2019, AJ, 157, 168
Drlica-Wagner A. et al., 2021, ApJS, 256, 2
Drlica-Wagner A. et al., 2022, ApJS, 261, 38
Finke A., Foffa S., Iacovelli F., Maggiore M., Mancarella M., 2021, J. Cosmol. Astropart. Phys., 2021, 026
Fishbach M. et al., 2019, ApJ, 871, L13
Freedman W. L. et al., 2019, ApJ, 882, 34
Gair J. R. et al., 2022, AJ, 166, 22
Graham M. J. et al., 2023, ApJ, 942, 99
Gray R. et al., 2020, Phys. Rev. D, 101, 122001
Hermans J., Delaunoy A., Rozet F., Wehenkel A., Begy V., Louppe G., 2022, preprint (arXiv:2110.06581)
Holz D. E., Hughes S. A., 2005, ApJ, 629, 15
Hotokezaka K., Nakar E., Gottlieb O., Nissanke S., Masuda K., Hallinan G., Mooley K. P., Deller A. T., 2019, Nature, 3, 940
Ilbert O. et al., 2006, A&A, 457, 841
Li C. et al., 2022, MNRAS, 518, 513
LIGO Scientific Collaboration, Virgo Collaboration, 2019, ApJ, 909, 218
LIGO Scientific Collaboration, Virgo Collaboration, 2023a, GRB Coordinates Network, 34086, 1
LIGO Scientific Collaboration, Virgo Collaboration, 2023b, GRB Coordinates Network, 34739, 1
LIGO Scientific Collaboration, Virgo Collaboration, 2023c, GRB Coordinates Network, 34757, 1
LIGO Scientific Collaboration, Virgo Collaboration, 2023d, GRB Coordinates Network, 35298, 1
LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration, 2021, Phys. Rev. X, 13, 041039
Ligo Scientific Collaboration, VIRGO Collaboration, Kagra Collaboration, 2023a, GRB Coordinates Network, 34087, 1
Ligo Scientific Collaboration, VIRGO Collaboration, Kagra Collaboration, 2023b, GRB Coordinates Network, 34739, 1
Ligo Scientific Collaboration, VIRGO Collaboration, Kagra Collaboration, 2023c, GRB Coordinates Network, 34757, 1
Ligo Scientific Collaboration, VIRGO Collaboration, Kagra Collaboration, 2023d, GRB Coordinates Network, 35298, 1
Ligo Scientific Collaboration, VIRGO Collaboration, Kagra Collaboration, 2023e, GRB Coordinates Network, 35428, 1
Lima E. et al., 2022, Astron. Comput., 38, 100510
MacLeod C. L., Hogan C. J., 2008, Phys. Rev. D, 77, 043512
Madau P., Dickinson M., 2014, Annu. Rev. Astron. Astrophys., 52, 71
Mastrogiovanni S. et al., 2023, Phys. Rev. D, 108, 042002
Mooley K. P. et al., 2018, Nature, 561, 355
Mucesh S. et al., 2021, MNRAS, 502, 2770
Mukherjee S., Wandelt B. D., Nissanke S. M., Silvestri A., 2021a, Phys. Rev. D, 103, 043520
Mukherjee S., Lavaux G., Bouchet F. R., Jasche J., Wandelt B. D., Nissanke S., Leclercq F., Hotokezaka K., 2021b, A&A, 646, 11
Nicolaou C., Lahav O., Lemos P., Hartley W., Braden J., 2020, MNRAS, 495, 90
Palmese A., Kim A. G., 2021, Phys. Rev. D, 103, 103507
Palmese A. et al., 2020, ApJ, 900, L33
Palmese A., Bom C. R., Mucesh S., Hartley W. G., 2023, ApJ, 943, 56 (P23)
Palmese A., Kaur R., Hajela A., Margutti R., McDowell A., MacFadyen A., 2024, Phys. Rev. D, 109, 063508
Planck Collaboration VI, 2018, A&A, 641, A6
Polsterer K. L., D'Isanto A., Gieseke F., 2016, preprint (arXiv:1608.08016)
Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, ApJ, 876, 85
Riess A. G., Casertano S., Yuan W., Bowers J. B., Macri L., Zinn J. C., Scolnic D., 2021, ApJ, 908, L6
Riess A. G. et al., 2022, ApJL, 934, L7
Rodríguez-Ramírez J. C., Nemmen R., Bom C. R., 2024a, preprint (arXiv:2407.09945)
Rodríguez-Ramírez J. C., Bom C. R., Fraga B., Nemmen R., 2024b, MNRAS, 527, 6076
Schutz B. F., 1986, Nature, 323, 310
Singer L. P., Price L. R., 2016, Phys. Rev. D, 93, 024013
Singer L. P. et al., 2016a, ApJS, 226, 10
Singer L. P. et al., 2016b, ApJ, 829, L15
Soares-Santos M. et al., 2019, ApJ, 876, L7
Teixeira G. et al., 2024, Astron.Comput., 49, 100886
Verde L., Treu T., Riess A. G., 2019, Nat. Astron., 3, 891
Voelker A. R., Kajić I., Eliasmith C., 2019, in Wallach H., Larochelle H., Beygelzimer A., d Alché-Buc F., Fox E., Garnett R., eds, Advances in Neural Information Processing Systems. Curran Associates, Inc., New York, USA, p. 15544
Zhou R. et al., 2020, MNRAS, 501, 3309
Zou H., Gao J., Zhou X., Kong X., 2019, ApJS, 242, 8

APPENDIX A: COMPARISON WITH H_0 CONSTRAINTS FROM POINT ESTIMATES OF PHOTOMETRIC REDSHIFTS

During this work, the H_0 posterior calculation was performed using the full photo-z PDFs estimated through the deep learning method described in Section 2.3. In this appendix, we will analyse the effect on the H_0 posterior originated by the choice of the galaxies photo-z PDFs. We will consider the photo-z PDF estimated by the MDN technique described in Section 2.3, which we will call the full photo-z PDF, or a Gaussian approximation, where we approximate each full photo-z PDFs to a Gaussian function whose mean coincides with the peak value and the standard deviation is equal to that found with the full photo-z PDF. In this context, the Gaussian approximations are constructed with the same peak and standard deviation as those found with the PDFs estimated with the MDN. We include the results of the public Legacy[6] in the figures. The top panel of Fig. A1 shows the posterior of H_0 found considering the full photo-z PDF (blue solid curve) and the Gaussian approximation (black dashed curve for the full PDF, this work, and the Legacy DR10.1 public results in grey dashed curve), along with the residuals curves (see the bottom panels).

Comparing the results of the H_0 posterior generated using the Gaussian approach (black dashed curve) and the full photo-z PDF, we note that the results are in agreement at the percentage level for the more distant events, where the effect is suppressed by marginalization over a larger number of galaxies due to their larger localization volume. The impact of using the Gaussian approximation is more evident for the events S230627c, GW200129, and GW191204, which present a discrepancies of approximately 4 per cent, 16 per cent, and 40 per cent, respectively, in the peak distribution and can reach values

---

[6]The photo-z were computed using the Random Forest technique with data from the Legacy DR10.1, these results are available in https://www.legacysurvey.org/dr10.
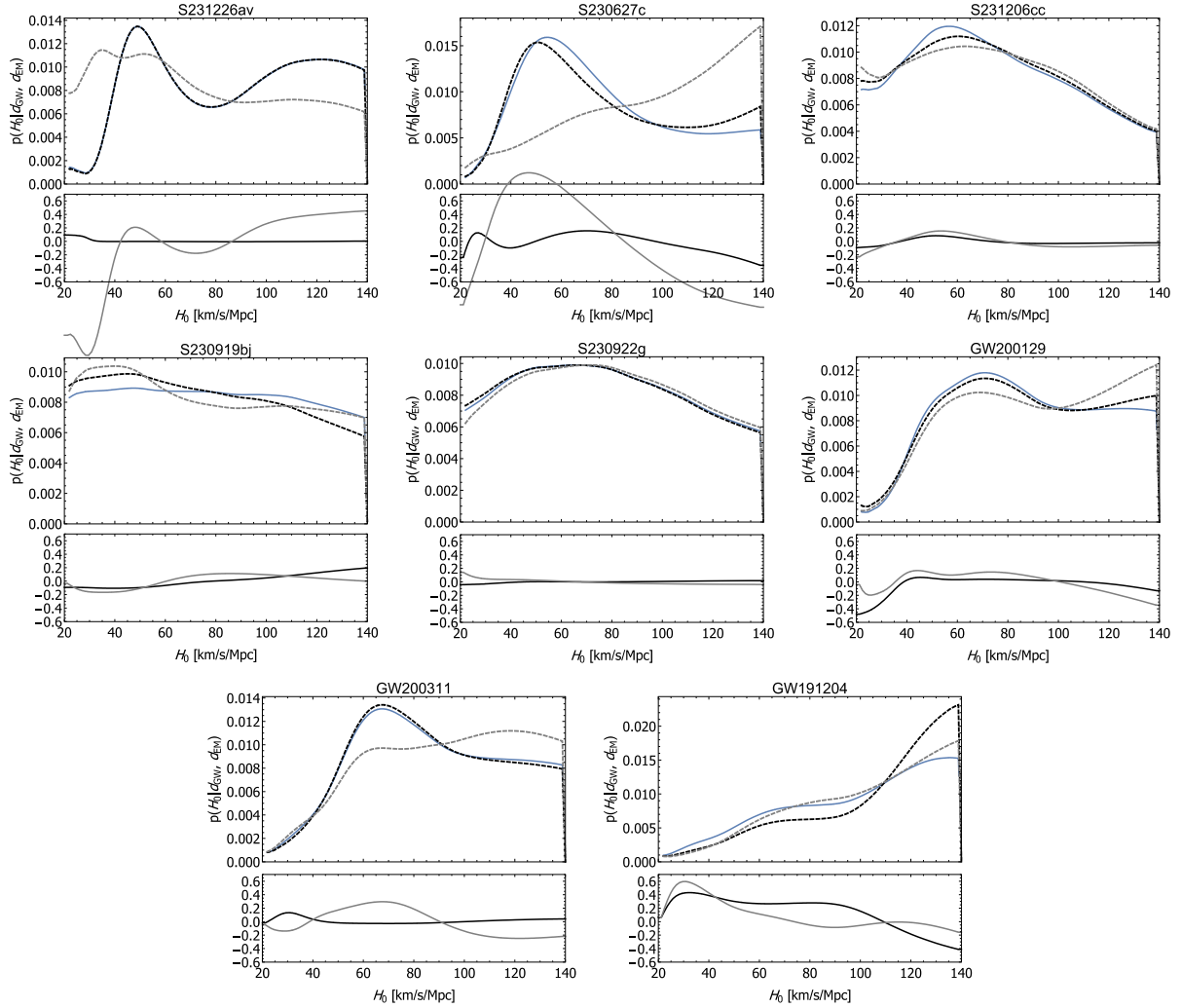
Figure A1. Comparison between the Hubble constant posterior distributions found using the full galaxies redshift PDFs (blue solid line) and a Gaussian approximation [dashed line: (1) black colour using the means and standard deviation of the full PDFs, and (2) grey colour using the photo-z measurements made by the public Legacy DR10.1] for all the dark sirens studied here. The bottom panels show the residuals value between the two curves shown on the top panel, calculated as 2 × (solid line − dashed line)/(solid line + dashed line).

> 13 per cent at the ends. As expected, the Gaussian approximation makes the peak of the H_0 posterior wider, implying a less precise H_0 measurement. This result is a consequence of the smoothing of overdensity regions in the photo-z distribution. Although the results do not indicate that the choice of photo-z PDF leads to disagreement in H_0 measurements, future dark siren measurements may reach a level of precision where the differences between these results could be statistically significant. We can also see that the Legacy public results leads to, in most events, a less restrictive H_0 posterior compared to those achieved through the full photo-z PDFs, resulting in an increase of up to 6 per cent in the H_0 uncertainty.

As shown by Soares-Santos et al. (2019), the choice for a wider redshift cut causes galaxies to be added at deeper redshift, whose photo-z PDF has significant Gaussian tails at high redshift, which implies an increase in the H_0 posterior at high H_0. However, the full photo-z PDF is able to reduce (the residual values, between the posterior considered a LIGO/Virgo luminosity distance posterior of 90 per cent and 99.7 per cent, at the high-H_0 end

reduced by $\approx$ 25 per cent compared to the result achieved with the Gaussian approximation) this dependence of the H_0 posterior behaviour with redshift cut, which evidences its advantage against the Gaussian approximation in performing a measurement of H_0 free of any systematic imposed by assumptions made in the methodology.

In Fig. A2, we present the photo-z PDFs estimated by the deep learning method (dark blue curve) and the Gaussian approximation (dashed curve in light blue) for galaxies at different redshifts. We notice that the full photo-z PDF has a tendency to be narrower than the Gaussian approximation, which favours a more precise measurement of H_0. Additionally, we see that the difference between the estimated PDF and the Gaussian approximation becomes more evident at higher redshifts. This highlights the importance of the MDN technique, especially for galaxies located at deep redshifts, which are already used for the LIGO/Virgo GW dark sirens and will be increasingly indispensable in the coming years where it is expected to observe increasingly distant GW events.
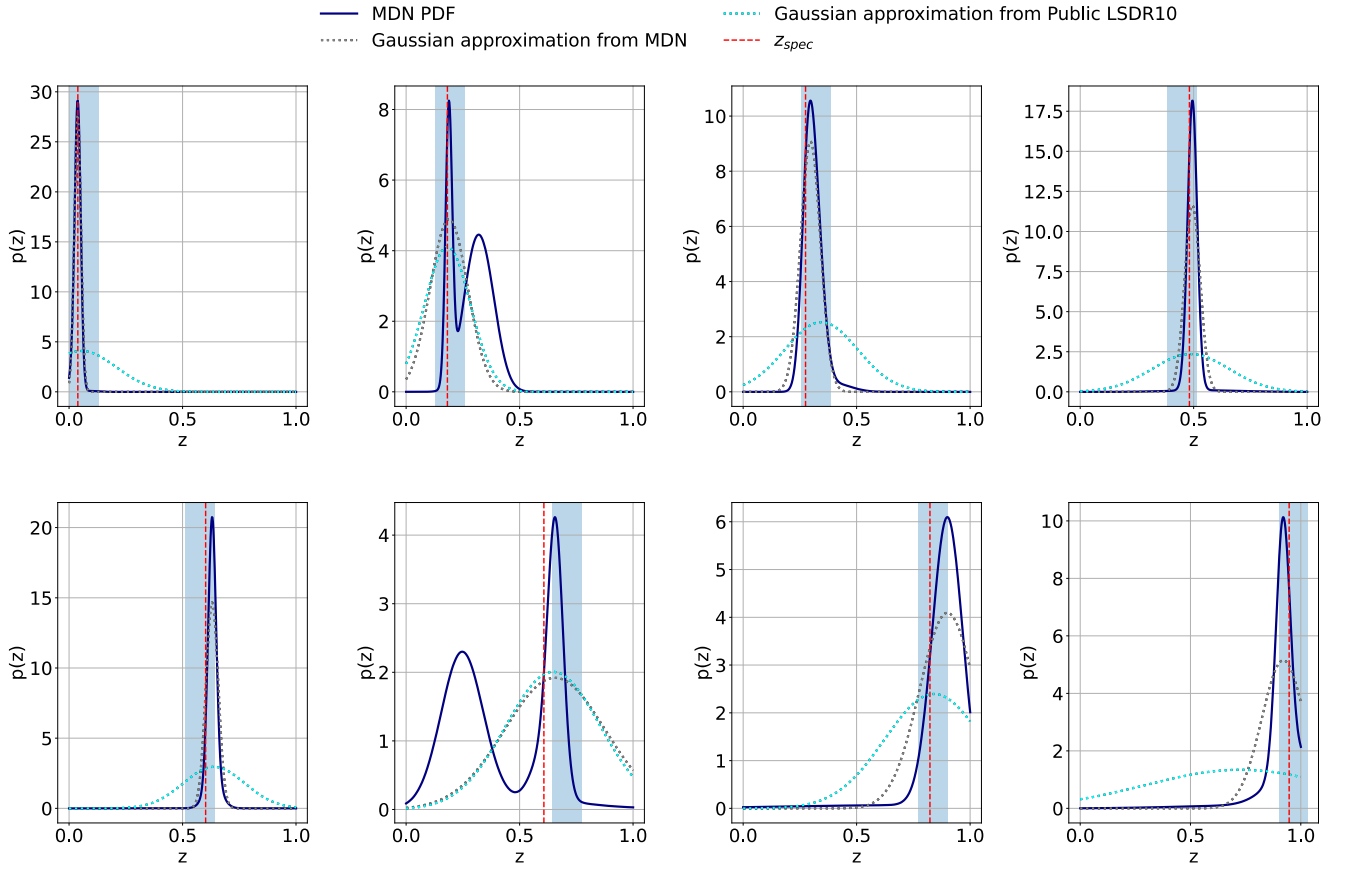
Figure A2. Photometric redshift PDFs estimated with the MDN technique (dark blue solid line), Gaussian approximation of MDN photo-z PDF (dashed curve in cyan), and Gaussian approximation using the photo-z estimated by the Public LSDR10 (dashed curve in grey) for different redshifts (increases from left to right). The vertical dashed red line represents the corresponding $z_{\rm spec}$ value.
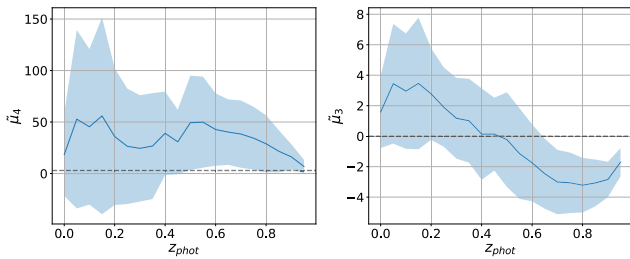

Figure A3. Kurtosis (left panel) and skewness (right panel) values of the photo-z PDFs, estimated with the MDN technique, as a function of $z_{\rm phot}$. The solid line represents the mean value of each quantity, and the shaded region represents the respective standard deviations computed in photo-z bins of size 0.05.

We can also observe that the full photo-z PDFs become increasingly non-Gaussian at deeper redshifts. To quantify this non-Gaussianity, in Fig. A3, we present the relation between the kurtosis

(fourth standardized moment $\bar{\mu}_4$, that is presented in the left panel) and the skewness (third standardized moment $\bar{\mu}_3$ shown in the right panel) of the full PDFs with $z_{\rm phot}$. We compute the mean (solid line) and the standard deviation (shaded region) of this quantities in photo-z bins of size 0.05. The values of kurtosis and skewness are distinct from those of a Gaussian distribution (kurtosis and skewness equals to 3 and 0, respectively) across almost the entire $z_{\rm phot}$ interval. We note a tendency of kurtosis values to decrease with increasing $z_{\rm phot}$, indicating the broadening of PDFs, which implies less precise $z_{\rm phot}$ measurements. This behaviour reflects the expected difficulty of obtaining precise measurements of distant galaxies. The skewness results indicate that the full PDFs are right-skewed at low $z_{\rm phot}$, decreasing with increasing $z_{\rm phot}$ until reaching a value of zero (indicating a symmetric PDF) at $z_{\rm phot} \approx 0.5$, after which they begin to be left-skewed.

This paper has been typeset from a TeX/LaTeX file prepared by the author.