Gravitational Wave Cosmology: measuring the Hubble constant with dark standard sirens

Rachel Gray

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

September 2021

Abstract

The fact that the universe is expanding has long been agreed upon, but the rate at which it is doing so, the Hubble constant, is a cause of disagreement across the field of cosmology, with a tension between early- and late-time measurements that could either be due to systematic measurement uncertainty, or an as-yet unknown discrepancy between the cosmological model and our universe.

The use of gravitational wave standard sirens for cosmological measurements has long been considered. Gravitational wave signals from compact binary coalescences have their distance encoded within them, and require no other form of calibration, making them independent of the cosmic distance ladder. A measurement of the Hubble constant using standard sirens would, by agreeing with the early-time or late-time measurements, give an indication as to whether the solution to the Hubble constant tension is systematic error or exciting new physics.

The detection of the binary neutron star merger, GW170817, alongside its electromagnetic counterpart, gave rise to the first measurement of the Hubble constant using gravitational waves. Further gravitational wave detections are required in order to improve the precision enough to discriminate between the early- and late-time Hubble constant measurements, but as of 2021 no other gravitational wave detections have been linked unambiguously with an electromagnetic counterpart.

In 1986, Schutz proposed a method for measuring the Hubble constant using gravitational wave detections when observed without electromagnetic counterparts (hence the term *dark standard sirens*). Galaxy catalogues can be used instead to provide the redshifts of potential host galaxies, and the uncertainty of which galaxy is the real host can be marginalised over.

This thesis takes Schutz's original proposition and places it in a Bayesian framework which allows the incompleteness of the galaxy catalogue (and the possibility that the host galaxy of a gravitational wave event won't be contained within it) to be accounted for. Gravitational wave selection effects (due to the finite sensitivity of the current detectors) are also incorporated into the methodology. This methodology is implemented in a software package, `gwcosmo`, which is then tested using a series of mock data analyses of increasing complexity. Following the success of this, the gravitational wave detections from Advanced LIGO and Virgo's first and second observing runs are used, in combination with publicly available galaxy catalogues, to make the first measurement of the Hubble constant that combines data from multiple gravitational

wave detections. Finally, further improvements are made to `gwcosmo` that take into account variations in the completeness of galaxy catalogues across the sky. This work paves the way for the analysis of the gravitational wave data from Advanced LIGO and Virgo's third observing run, and beyond.

Contents

List of Tables

List of Figures

Acknowledgements

I began my PhD on the 2nd October 2017, less than two months after the detection of the binary neutron star merger, GW170817, and its electromagnetic counterpart. To say that it was an exciting start to my PhD is an understatement. The words "baptism of fire" (a phrase I believe was used during my first year progression review) spring to mind. I think I can safely say that this PhD has been a formative process for me. Now, at the end of my PhD, there are many people I would like to thank for helping me get to this point.

My first and biggest thank you goes to my supervisors, John Veitch and Chris Messenger, without whom this PhD would have been a very different experience. Their advice and guidance has been consistently excellent, and working with them has been a real highlight of the last four years. My thanks goes to Chris, in particular, for being the first person to make me seriously consider doing a PhD, and for introducing me to the field of gravitational wave cosmology. For John, thank you for encouraging me to think seriously about what comes next.

Over the past few years I have been fortunate to collaborate with many people within the LVK Cosmology working group. Much of the research I have done would not have been possible without their contributions. I look forward to our future collaborations! I also want to express my gratitude to the members of the Institute for Gravitational Research – I know I have benefited hugely from working in such a supportive group, where people are always happy to share their expertise.

Overall my experience at the University of Glasgow has been a very positive one. A big part of that has been the postdocs and other PhD students who have made working here such a joy. They have kept me company, listened to me vent over coffee and at the pub, and generally helped me to retain my sanity.

Finally, of course, there is my family, who have been incredibly supportive throughout this entire process. I've appreciated all the motivational chats, messages, and photos over the last few years (and the lack of questions about when I'm going to finish writing up)! Thank you for keeping me grounded while I've been off staring into space.

This PhD was funded by the Science and Technology Facilities Council, award number 1947165.

Declaration

All work in this thesis was carried out by the author unless otherwise explicitly stated. Chapter 1 is introductory material. Chapters 2, 3 and 4 contain material from published papers, and have individual introductory statements which summarise the author's contribution. Chapter 5 consists of the author's as-yet unpublished work.

Chapter 1

Introduction

The possibility that the universe might be expanding was first acknowledged in the 1920s, when Friedmann and Lemaître independently produced theoretical proofs that allowed it [1, 2]. Observationally, this was confirmed only a few years later by Hubble, who saw that the recession velocity of galaxies increased with their distance from the observer [3]. While the expansion of the universe has long been acknowledged as fact, the quest to measure the speed at which the universe is expanding, known as the Hubble constant (H_0) is one that has spanned nearly a century.

The first measurements of H_0 were highly inaccurate, dominated by measurement errors and misunderstandings. Recent measurements have reduced the uncertainty to a few percent [4, 5], but in doing so a new and exciting problem has appeared – different measurements produce results which seem to disagree, with early-universe measurements (also known as early-time measurements) favouring, in general, a lower value of H_0 than the local (late-time) measurements [6, 7]. The cause of this disagreement is still up for debate, with the proposed explanations falling into two groups: either some of the measurements must have underlying systematic errors which, when corrected for, will cause them to align with the rest; or the cosmological model of the universe, which describes the evolution of the universe over time, is incorrect, and this tension between measurements of H_0 is actually an indicator that new physics is required to describe our universe.

To make a local measurement of H_0 two things are required: a measurement of the distance to an object (a galaxy, for example), caught in the Hubble flow, such that its movement away from the observer is dominated by the expansion of the universe; and a measurement of how fast that object is moving away from us. At small distances (cosmologically speaking), the speed at which an object is travelling away from us is proportional to its redshift: the fractional lengthening of the wavelength of light emitted from it, due to the expansion of the universe as it travels through it. The redshifts of galaxies can be obtained through spectroscopy, for example (measuring the wavelengths of known emission lines, and seeing how they have shifted relative to their expected position). Measuring the distances to nearby galaxies poses a different problem.

There is no one way to measure distances in the universe which is applicable at all scales. Instead the *cosmic distance ladder* is used. This ladder is a grouping of different measurement techniques which can be used to measure distances at different scales across the universe. One of the main methods of measuring distances is through the use of *standard candles* – astronomical objects of known brightness. The relationship between the observed brightness of the object and its intrinsic brightness provides a measurement of its luminosity distance. This method is not perfect for a couple of reasons: firstly, objects can appear fainter due to other reasons than their distance (if interstellar gas gets in the way, for example); secondly, it is hard to know the intrinsic brightness of an astronomical object – and if the properties of these standard candles changes with redshift, for example, then there would be an inherent bias in the cosmic distance ladder.

This is where gravitational waves (GWs) enter the problem. GWs are ripples in space-time caused by extreme events in the universe. The GWs from compact binary coalescences (CBCs) (such as the merging of black holes or neutron stars) have an incredibly useful property, in that the amplitude of the signal they produce, as measured on Earth, is inversely proportional to their luminosity distance from us. This distance has some measurement uncertainty, but can be obtained from the gravitational wave signal alone, requiring no additional calibration [8]. As such, these GW sources are referred to as *standard sirens*, because they are analogous to standard candles as distance indicators.

The detection of the first GW signal by Advanced LIGO in 2015 [9] was heralded as marking a new era in astronomy. However it wasn't until two years later, with the detection of the first ever binary neutron star (BNS) merger by the Advanced LIGO and Virgo detectors [10], that the field of GW cosmology was revolutionised. Not only was this event unmistakeably loud, due to being nearby, it was also very well localised in the sky – so much so that the electromagnetic (EM) counterpart of the signal could be identified, and was seen by multiple telescopes, across multiple wavelengths, for many days [11, 12]. In doing so the host galaxy of the source was identified, allowing a redshift to be obtained for the merger. With the distance measure provided by the GW signal, this led to the first ever direct measurement of H_0 using GWs [13].

The uncertainty on the H_0 measurement from GW170817 and its counterpart alone was about 15%, nowhere near accurate enough to indicate which of the previous cosmological measurements it agreed with more. Several dozen more similar events would need to be observed in order to achieve that level of precision – and therein lies the issue: GW170817 was an exceptional event. Exceptionally nearby, exceptionally well-localised, and exceptionally fortunate that the counterpart was identified so early on, so that so many telescopes could turn and point towards it before it faded from view. Since then the third observing run of Advanced LIGO and Virgo has wrapped up and no other event has been observed with a confirmed EM counterpart [14, 15]. Few events even come close to being as nearby or as well localised on the

sky.

The number of GW events lacking viable EM counterparts is exactly where the work in this thesis comes in. In 1986 Schutz proposed a method of using GW standard sirens to measure H_0 in the case where they *don't* have an EM counterpart [8] (hence the term *dark standard siren*). Instead, all of the galaxies lying within the sky localisation of the GW event would be considered as possible hosts, and the uncertainty of which galaxy is the real host is averaged over.

This thesis takes Schutz's original proposition, and the subsequent Bayesian derivations thereof as a starting point. Chapter 2 presents a Bayesian methodology which incorporates both GW selection effects (from the finite sensitivity of GW detectors) and EM selection effects (from the finite sensitivity of EM telescopes, meaning that the host galaxy of the GW merger might not be contained within current galaxy catalogues). Chapter 3 introduces the software `gwcosmo`, a Python implementation of the methodology, and tests it rigorously with a series of increasingly complex mock data analyses (MDAs). Following the success of this, Chapter 4 analyses the GW detections from Advanced LIGO and Virgo's first and second observing runs, in combination with publicly available galaxy catalogues, to make the first measurement of the Hubble constant that combines data from multiple GW detections. Chapter 5 introduces a further improvement to `gwcosmo` which allows it to take into account variations in the completeness of galaxy catalogues across the sky, and reanalyses the first and second observing run detections. This work paves the way for the analysis of the GW data from Advanced LIGO and Virgo's third observing run, and beyond.

Before moving on, however, a more thorough understanding of the current fields of cosmology and GWs would be beneficial. Section 1.1 introduces the basics of modern cosmology and the current best model of how our universe has evolved over time. Section 1.2 describes the Hubble constant tension, the latest measurements driving it, and the possible underlying causes. Section 1.3 describes the physics which allow CBCs to be used as standard sirens. Section 1.4 introduces Schutz's original idea in greater detail, and discusses other cosmological measurements that have been made, or could be made in the future, using GWs.

1.1 The standard model of cosmology

The cosmological principle states that the universe is homogeneous and isotropic, when looked at on a large enough scale. Homogeneous in that every location within it is broadly the same (and our own location within it is in no way special), and isotropic in that it looks the same in every direction. The fact that galaxies in the universe appear to be moving away from the Earth is one of the most important pieces of observational evidence in cosmology, alongside the fact that things at higher distances are receding at faster rates. This is compatible with homogeneity and isotropy if the universe itself is expanding, because no matter where in the universe they are, an observer would see the same thing. Both Hubble and Lemaître recognised that galaxies

appeared to be receding from the Earth at velocities that were roughly proportional to their distances, and independently derived what is now known as the Hubble-Lemaître law [2,3]. The Hubble-Lemaître law states

$$v = H_0 d, \tag{1.1}$$

where v is an object's recession velocity (usually measured in km s^{-1}), d is the proper distance between the observer and the object, usually measured in Mpc, and H_0, the Hubble constant, is the present-day value of the Hubble parameter, with units of km s^{-1} Mpc^{-1}.

The fact that the universe is expanding implies that in the past it was much smaller than it is today – this is a foundational piece of thinking for the Big Bang theory, which states that the universe began in an energetic explosion from a singularity, and has been expanding ever since. The standard model of cosmology refers to the Λ-cold-dark-matter (ΛCDM) model of the universe, which begins with a Big Bang and evolves to be a universe filled with cold, dark matter (non-relativistic matter that does not emit, or interact strongly with, light) as well as ordinary matter. A cosmological constant, Λ, which indicates the presence of dark energy in the universe is required to fit observational evidence. This works against the pull of gravity in order to accelerate the expansion of the universe.

In the ΛCDM model, the Big Bang took place approximately 13.8 billion years ago [4]. In this scenario, the universe expands rapidly, and is filled with matter and photons in a hot dense plasma. Small perturbations in the plasma propagate like sound waves (known as Baryonic Acoustic Oscillations (BAO)), creating over- and under-densities. As the universe continues to expand, it cools, allowing for atoms to form, at which point light is no longer constantly absorbed and re-emitted, but can travel freely through open space. This time is known as recombination, when matter and light decouple, and the photons from that era are detected today as the cosmic microwave background (CMB). The perturbations that were in the plasma no longer oscillate, but remain imprinted in the distribution of matter, and can be seen as tiny variations in the temperature of the CMB. The expansion of the universe continues, with gravity causing over-densities of matter to contract and heat up, and eventually form stars and galaxies. Eventually these evolve to be a vast web of matter, with clusters of galaxies and filaments joining them. This large-scale structure retains the imprint of the BAO, which is observable today. The expansion of the universe does not separate objects which are gravitationally bound, but the voids of empty space continue to expand. As time progresses, Dark Energy starts to dominate, causing the expansion of the universe to accelerate.

In order to talk about the expansion of the universe in a more rigorous manner it is first useful to introduce the idea of a *metric*: a metric details the separation of events in space-time. In particular, for a universe that obeys the cosmological principle, the metric must be a Friedmann-Lemaître-Robertson-Walker (FLRW) metric (developed, in some cases independently, by the four men it is named after in the 1920s and 1930s [1, 2, 16–21]). The FLRW metric is an exact solution to Einstein's field equations of general relativity, which describe how space-time is

curved by the matter within it. The key thing to note for the FLRW metric is that it includes a time-dependent *scale factor*, which describes how the spatial distance between two events evolves with time. In general, the metric takes the form

$$ds^2 = -c^2dt^2 + a^2(t)ds_3^2, \tag{1.2}$$

where ds^2 is the space-time interval between two events, $-c^2dt^2$ is the time part of the separation, and $a^2(t)ds_3^2$ is the spatial part of the separation, with $a(t)$ as the scale factor. The exact form of ds_3^2 depends on the geometry of the universe, and whether it has spherical, flat, or hyperbolic geometry.

In an expanding universe, every point in space is moving away from every other one (ignoring objects which are gravitationally bound for the time being). If the universe is described using comoving coordinates, which expand with the universe in order to give fixed coordinates to objects caught in the Hubble flow, the scale factor, $a(t)$, is how one translates between comoving coordinates and proper distance – the separation between objects at a fixed time, which changes over time. The present-day value of the scale factor, $a(t_0)$ is defined to be 1. In general,

$$d(t) = a(t)d_0, \tag{1.3}$$

where $d(t)$ is the (proper) distance between two objects at a time t, and d_0 was the distance between them at t_0.

Interestingly the Hubble-Lemaître law can be derived quite simply from Eq. 1.3. Taking the time derivative of each side leads to

$$\dot{d}(t) = \dot{a}(t)d_0. \tag{1.4}$$

Recognising that Eq. 1.3 can also be rearranged to give $d_0 = d(t)/a(t)$, and substituting this into Eq. 1.4 gives

$$\dot{d}(t) = \frac{\dot{a}(t)}{a(t)}d(t), \tag{1.5}$$

where $\dot{d}(t)$ is just the recession velocity of an object as seen by the observer, $\dot{a}(t)/a(t)$ is the Hubble parameter and $d(t)$ is the proper distance to the object.

The FLRW metric on its own does not provide a full picture of the universe, as it does not contain information about how the universe is affected by the matter and energy within it. However, solving Einstein's equations assuming a FLRW metric gives rise to the Friedmann equations [1, 16]. These equations dictate how $a(t)$ evolves, and how it is related to the pressure, density, and geometry of the universe. These take the form

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \tag{1.6}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{1.7}$$

where a is the scale factor (using a shorthand notation where the (t) has been dropped), $\dot{a}$ is the time derivative of a, ρ is the density of the universe and p is its pressure, k is the curvature of the universe, and Λ is the cosmological constant. In both equations, a, p and ρ evolve with time, while π, c, G and Λ are constants, and k can have one of three values:

$$k = \begin{cases} 1 & \text{for a universe with spherical geometry (known as a closed universe)} \\ 0 & \text{for a universe with flat geometry} \\ -1 & \text{for a universe with hyperbolic geometry (known as an open universe).} \end{cases} \tag{1.8}$$

The curvature of the universe is actually a continuum, but the scale factor can be rescaled to cover this variation, such that k takes on an integer value and its sign (whether it is positive, negative, or zero) determines the type of geometry the universe follows.

By definition, the Hubble parameter H (or $H(t)$) is

$$H \equiv \left(\frac{\dot{a}}{a}\right), \tag{1.9}$$

and its present day value is what is termed the Hubble constant, H_0. For cosmological inference it is necessary to obtain an expression of the relationship between H_0, redshift (z), and luminosity distance (d_L) (see, *e.g.* , [22]). To do so it is helpful to define the density parameter, Ω_m, which dictates the density of the universe relative to the critical density which the universe would have to have in order to be flat if there were no cosmological constant Λ, *i.e.* solving Eq. 1.6 when k and $\Lambda = 0$:

$$\Omega_m = \frac{\rho}{\rho_c}, \tag{1.10}$$

where ρ_c is, by definition,

$$\rho_c \equiv \frac{3H^2}{8\pi G}. \tag{1.11}$$

Returning to Friedmann's equation, Eq. 1.6, and rearranging it gives

$$\frac{8\pi G}{3H^2}\rho - \frac{kc^2}{a^2H^2} + \frac{\Lambda c^2}{3H^2} = 1. \tag{1.12}$$

Substituting in Ω_m, and defining the curvature density parameter and dark energy density parameter as

$$\Omega_k = -\frac{kc^2}{a^2H^2} \quad \text{and} \quad \Omega_\Lambda = \frac{\Lambda c^2}{3H^2} \tag{1.13}$$

respectively gives

$$\Omega_m + \Omega_k + \Omega_\Lambda = 1. \tag{1.14}$$

These density parameters contain the Hubble parameter, and hence will change over time. It is also worth defining the present-day values of these parameters, that is

$$\Omega_{m,0} = \frac{\rho_0}{\rho_{c,0}} = \frac{8\pi G}{3H_0^2}\rho_0, \quad \Omega_{k,0} = -\frac{kc^2}{a_0^2 H_0^2} = -\frac{kc^2}{H_0^2} \quad \text{and} \quad \Omega_{\Lambda,0} = \frac{\Lambda c^2}{3H_0^2}. \tag{1.15}$$

Returning to Eq. 1.6, and dividing both sides by H_0^2, then substituting in the present day density parameters produces the following:

$$\frac{H^2}{H_0^2} = \frac{\rho}{\rho_0}\Omega_{m,0} + \frac{1}{a^2}\Omega_{k,0} + \Omega_{\Lambda,0}, \tag{1.16}$$

It is now necessary to start thinking in terms of redshift, and how *cosmological redshift* (the redshift due purely to the expansion of space as light travels through it) relates to the expansion of the universe. In fact, the relationship depends only on the scale factor, and is as follows:

$$1 + z = \frac{a(t_o)}{a(t_e)}, \tag{1.17}$$

where $a(t_o)$ is the size of the universe when light from the object was observed and $a(t_e)$ is the size of the universe at the time the light was emitted. If the observation is taking place in the present day (as most are wont to do) then it can be simply written as $a(t) = 1/(1 + z)$.

Recognising that, for a matter-dominated universe, the present day density of the universe is related to its density at time t by $\rho/\rho_0 = 1/a(t)^3$, and then substituting that and $a(t) = 1/(1 + z)$ into Eq. 1.16, the dimensionless Hubble parameter, $E(z)$ can be defined:

$$E(z) \equiv \frac{H(z)}{H_0} = \sqrt{\Omega_{m,0}(1 + z)^3 + \Omega_{k,0}(1 + z)^2 + \Omega_{\Lambda,0}}. \tag{1.18}$$

The latest measurements of the density parameters give values of $\Omega_{m,0} \sim 0.3$, $\Omega_{k,0} \sim 0$ and $\Omega_{\Lambda,0} \sim 0.7$ [4]. In general it is agreed that the universe is either flat, or very close to being so. Assuming so allows many simplifications for the following equations, and so it will in general be assumed that k has been taken to equal 0 unless specified otherwise.

The luminosity distance, d_L, of an object is the distance it would have if the inverse square-law relating flux and distance held across the universe (which it does not). For a flat universe the luminosity distance of an object is equivalent to the redshifted comoving distance, $(1 + z)D_C$. This leads to the following relationship between d_L, z and H_0:

$$d_L = \frac{c\,(1 + z)}{H_0} \int_0^z \frac{dz'}{E(z')}. \tag{1.19}$$

Another useful quantity to define before moving forward is that of comoving volume V_C. Isotropy and homogeneity imply that matter is distributed (more or less) uniformly in the universe, and

that this has remained true as the universe evolves. That is, matter is believed to have a uniform in comoving volume distribution. In reality, the clustering of galaxies means that this is not entirely the case, but when the universe is examined on a large enough scale (> 100 Mpc) the principles of isotropy and homogeneity hold. A uniform in comoving volume distribution takes the form

$$\frac{dV_C}{dz} = \frac{c^3}{H_0^3}\frac{1}{E(z)}\left(\int_0^z \frac{dz'}{E(z')}\right)^2, \tag{1.20}$$

which, when normalised, drops its dependence on H_0 and can be used as prior on galaxy distribution in redshift.

Returning now to the early measurements that were made of galaxy redshifts, and the observation that they increased with the galaxy's distance from Earth. The observed redshift of an object with recession velocity v is given by

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{emit}}}{\lambda_{\text{emit}}}, \tag{1.21}$$

where λ_{emit} is the redshift of the light when it is emitted from the object, and λ_{obs} it its redshift when observed. For $v \ll c$, it can be approximated as the following:

$$z \approx \frac{v}{c}. \tag{1.22}$$

That is, at small redshifts, redshift and recession velocity are proportional. At small redshifts one can also say that luminosity distance and proper distance are roughly equivalent, which leads to the linear Hubble relation

$$cz \approx H_0 d_L. \tag{1.23}$$

This assumes that the redshift in question is a cosmological redshift, and is not affected by peculiar motion (motion other than recession due to the Hubble flow), which impacts the measured redshift, and can easily dominate for galaxies at low redshifts. In general, however, z for a galaxy is measurable (using spectroscopy or photometry, and correcting for peculiar velocities) and d_L is measurable (using standard candles), and so local measurements of H_0 can be made this way, without any further assumptions about other cosmological parameters – as was done for the first ever measurements of H_0. Of course, the first measurements made by Lemaître and Hubble had large measurement uncertainties, but were also far from today's currently accepted value of $H_0 \sim 70$ km s^{-1} Mpc^{-1}due to miscalibrated distances. This brings us nicely to consider current day measurements of H_0, to see how they disagree and what could be driving this disagreement.

1.2 The Hubble constant tension

While the ΛCDM model of the universe is very well matched to current observations, a tension has arisen in measured values of H_0 coming from different types of observation. So far, the majority of measurements of H_0 show a divide along the lines of early- vs late-time measurements [6, 7], where early-time measurements are those made at high redshift looking far back into the history of the universe, and propagating forwards to find the current-day value of the Hubble parameter, while late-time measurements are made locally at lower redshifts and hence measure the current day value of the Hubble parameter. This tension between recent measurements is demonstrated in Fig. 1.1.

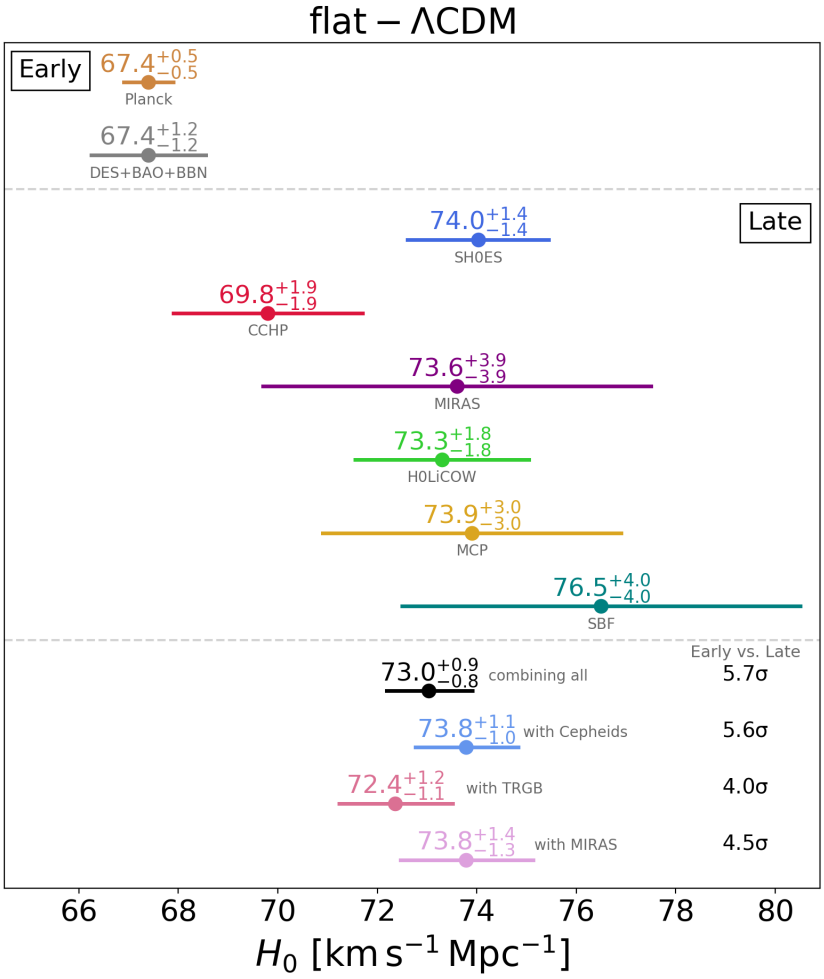

Figure 1.1: Demonstration of the early- vs late-time Hubble constant tension between recent measurements. Figure credit: Vivien Bonvin and Martin Millon [23].

First considering early-time measurements of H_0. In a universe that is completely homogeneous and isotropic, the CMB would be completely uniform. However, minor fluctuations in the density of plasma of the early universe occurred, caused by BAO. When the epoch of recombination occurred and matter and light decoupled, allowing photons to travel uninhibited through

the universe, these minute density fluctuations were imprinted on the matter of the universe, and seeded the formation of the large-scale structure of galaxy clusters that we see today.

The presence of BAO in the early universe means that the CMB measured today is not completely smooth, but shows small fluctuations in its measured temperature across the sky. The scale of these variations has an angular dependence, which can be used to measure H_0. The Planck collaboration uses measurements of anisotropies in the CMB to measure H_0 and in 2018 produced a value of $H_0 = 67.36 \pm 0.54$ km s^{-1} Mpc^{-1} [4]. Other measurements of CMB anisotropies produce similarly consistent results, such as the Atacama Cosmology Telescope (ACT) + Wilkinson Microwave Anisotropy Probe (WMAP) result which gives $H_0 = 67.6 \pm 1.1$ km s^{-1} Mpc^{-1}, [24].

The distance sound waves could have travelled in the time before recombination is called the sound horizon, r_s, and its value at the time of recombination is both well predicted by theory, and well measured by CMB anisotropies. The length scale set by the sound horizon has changed over time as the universe has expanded, and measuring it at different redshifts provides important insight into the rate of that expansion. Low-redshift BAO measurements can be made by looking at the clustering of galaxies from large surveys (see, *e.g.* [25–27]). These low-redshift measurements of BAO can make a joint constraint on H_0 and r_s, but additional measurements are needed to break the degeneracy. Usually this comes from CMB measurements, but recently a method of using deuterium abundance measurements and the Big Bang Nucleosynthesis (BBN) theory to calibrate r_s have been used [28, 29]. Crucially, the BAO+BBN measurement is completely independent of CMB measurements of H_0, but agrees well with the values quoted by Planck and others.

In terms of local measurements of H_0, the most well-known method is through observations of type 1a supernovae. Their intrinsic luminosities, and therefore their distances, can be determined from measurements of their light curves, making them useable as standard candles. Type 1a luminosities are typically calibrated using the known distances to Cepheid variable stars (requiring parallax measurements to obtain the distance to nearby cepheids, and then the identification of host galaxies containing both a cepheid and type 1a supernova, which then means the calibration can be propagated to type 1a supernovae further into the Hubble flow). The 2019 Supernovae, H_0, for the Equation of State of Dark energy (SH0ES) collaboration constraint produced $H_0 = 74.03 \pm 1.42$ km s^{-1} Mpc^{-1} [30], from observations of 70 long-period cepheids in the host galaxies of Type 1a supernovae. A recalibration of the distances to the cepheids in [5] improves the measurement to $H_0 = 73.2 \pm 1.3$ km s^{-1} Mpc^{-1}, which is the most up to date measurement at the time of writing, and is in tension to a level of 4.2σ with the Planck 2018 result (assuming ΛCDM).

Other late-time measurements of H_0 include (but are not limited to): H0LiCOW, which presented a measurement of H_0 from 6 gravitationally lensed quasars with measured time-delays, giving $H_0 = 73.3^{+1.7}_{-1.8}$ km s^{-1} Mpc^{-1} [31] which is independent of the distance ladder but in agree-

ment with the local SH0ES measurement. The Maser Cosmology Project produced a local H_0 measurement based on the distances to local megamaser-hosting galaxies (again independent of the cosmic distance ladder) of $H_0 = 73.9 \pm 3.0$ km s^{-1} Mpc^{-1} [32]. A method of using the tip of the red giant branch to calibrate distances to type 1a supernovae (instead of the usual distance ladder) produced a result in the middle of the H_0 tension region [33], although the analysis of a different research group using the same method produced a higher value aligning more closely with other late-time measurements [34].

While the Hubble constant tension is clearly present in the results summarised above, there is no single satisfactory explanation for its occurrence. Broadly speaking there are two possible causes: systematics in the measurements, or a deviation from the ΛCDM model. If large systematics exist, they are more likely to be present in the late-time measurements, where the astrophysics is not precisely known and often multiple measurements need to be combined to produce a value for H_0. However, there is currently no single systematic in the distance ladder that could account for all of the tension as it currently stands – not to mention the existence of local measurements which are independent of the distance ladder but still agree with their results.

On the other hand, the CMB measurements by Planck and others are highly precise, and have fewer sources for potential systematic uncertainty associated with them. However, the H_0 values derived from those measurements are model-specific. If they *are* wrong it is perhaps less likely to be due to measurement error, and more likely to be a breakdown in the current cosmological model. In that case, new physics will need to be explored in order to explain the H_0 tension. However thus far there is no easy modification to ΛCDM which explains the H_0 tension without introducing further tensions in other areas of cosmology. Possible modifications to the ΛCDM model include dynamical dark energy explanations, modified gravity, higher numbers of species of relativistic neutrinos, or non-zero curvature (see *e.g.* [35–40]). Currently these models reduce the H_0 tension, but this tends to be due to a broadening of the posterior on H_0 rather than due to a shifting of the central value, and in some cases these models are disfavoured by other datasets.

Anecdotally, when asked whether or not they believed that the ΛCDM model of cosmology was correct, many cosmologists agreed that, much like Newton's theory of gravity, ΛCDM is clearly a theory that describes our universe incredibly well, but is unlikely to represent the fullest picture, when examined in the most extreme cases. The question does not seem to be *if* the ΛCDM model will break down, but *when*. Is the Hubble tension a sign that we have reached the limitations of ΛCDM? That remains to be see. Perhaps gravitational wave standard sirens will shed some light on the issue.

1.3 Gravitational waves as standard sirens

The existence of GWs were predicted by Einstein in 1916 [41] as a consequence of his General Theory of Relativity which had been published the year before [42]. The behaviour of space-time in the presence of mass is described by Einstein's field equations, with the conclusion that accelerating masses will perturb space-time in such a way as to create ripples which propagate through it at the speed of light. Following this assertion, there was much debate about the validity of Einstein's conclusion, and whether GWs were real (including a time when Einstein himself was convinced that they were not). It was not until the 1950s, when Pirani described gravitational radiation in the framework of the Reimann tensor and demonstrated that it was not an artefact of the coordinate system [43] that it was generally accepted that GWs were real and, theoretically at least, measurable.

While there has been indirect evidence for the existence of GWs in the orbital decays of binary pulsars as far back as the 1980s, the first direct observation of GWs did not occur until 2015, nearly 100 years after Einstein's prediction. The observation was made by the Advanced LIGO detectors (where LIGO stands for Laser Interferometer Gravitational-wave Observatory) in Hanford and Livingston in the United States. The source of the signal was a pair of black holes approximately 440 Mpc away, rapidly inspiraling and then merging.

The intrinsic property of a GW signal from a CBC to encode within it the luminosity distance between the source and the observer makes GWs hugely interesting for astrophysical and cosmological reasons. Given the difficulty of measuring cosmological distances in astronomy, GWs provide a unique way in which to make a direct distance measurement which is completely independent of the cosmic distance ladder or any EM distance measurement used thus far in cosmology. From here on, it should be assumed that the GW signals under consideration are always from CBCs.

GWs distort the space-time they travel through, perpendicular to their direction of propagation, creating strain. Simply put, if two test masses were placed at a distance apart, the propagation of a gravitational wave past them would cause the distance between the two to change (this is the general principle of how GW laser interferometers detect GW signals). Similar to how light can be polarised, and the orientation of any linearly polarised light can be written as the sum of the horizontally and vertically polarised parts, GWs are also polarised, and the strain from any GW can be broken into a plus-polarised component, h_+, and a cross-polarised component, $h_\times$.

The strain of a GW signal depends on both the intrinsic properties of the source, and it's orientation relative to the location at which the strain is measured. The plus-polarised component

as a function of time, assuming the quadrupole approximation, is given by (following [44]),

$$\begin{aligned} h_+(t) &\equiv \frac{2\mathcal{M}_z}{d_L}\left(1+\cos^2(\iota)\right)(\pi\mathcal{M}_z f)^{2/3}\cos\left(\Phi+\Psi\right), \\ &= \frac{2\mathcal{M}_z}{d_L}\left(1+\cos^2(\iota)\right)\left(\frac{5}{256}\frac{\mathcal{M}_z}{(T-t)}\right)^{1/4}\cos\left(-2\left(\frac{(T-t)}{5\mathcal{M}_z}\right)^{5/8}+\Psi\right), \end{aligned} \tag{1.24}$$

where natural units $c = G = 1$ have been assumed. Here d_L is the luminosity distance between the source and observer, and $\mathcal{M}_z$ is the redshifted chirp mass of the source, defined as

$$\mathcal{M}_z = (1+z)\frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}, \tag{1.25}$$

where m_1 and m_2 are the primary and secondary masses of the binary system, and z is their redshift relative to the observer. The inclination of the binary system with respect to the observer is represented by ι, while f denotes the time-varying frequency of the signal, Ψ is the phase of the signal at time $t = T$, where T is the moment the binary coalesces, and Φ denotes how the phase of the signal evolves over time. The cross-polarised strain takes a similar form,

$$\begin{aligned} h_\times(t) &\equiv \frac{4\mathcal{M}_z}{d_L}\cos(\iota)\,(\pi\mathcal{M}_z f)^{2/3}\sin\left(\Phi+\Psi\right), \\ &= \frac{4\mathcal{M}_z}{d_L}\cos(\iota)\left(\frac{5}{256}\frac{\mathcal{M}_z}{(T-t)}\right)^{1/4}\sin\left(-2\left(\frac{(T-t)}{5\mathcal{M}_z}\right)^{5/8}+\Psi\right). \end{aligned} \tag{1.26}$$

The important takeaway here is that the amplitude of the strain produced by a GW signal is inversely proportional to the luminosity distance to the source.

In reality, the strain measured by a single GW interferometer is a linear combination of the plus and cross polarisation strains,

$$h(t) = F_+ h_+ + F_\times h_\times, \tag{1.27}$$

where F_+ and $F_\times$ are the antenna response functions of the detector, dependent on the source sky position (relative to the detector) and polarisation of the waves [45]. Additionally there will be detector noise in the measured strain. The frequency and frequency evolution of the signal over time breaks the degeneracy between d_L and $\mathcal{M}_z$, allowing measurements to be made of redshifted mass (also known as detector-frame mass) and distance. The inclination and distance of the source *are* degenerate, but the uncertainty in the inclination can be marginalised over, leading to a measurement of the luminosity distance of the source which incorporates this uncertainty.

If a GW signal is detected by multiple detectors, the source parameters can be measured to a greater degree of precision. Not only is the signal-to-noise ratio (SNR) higher, but the distance and inclination degeneracy can be partially broken, leading to tighter constraints on d_L [46]. Another parameter of importance, the sky location of the event, Ω, can also be better estimated

when the event is detected by multiple interferometers, due to the difference in the time of arrival of the signal at each detector [47]. The detected strain undergoes parameter estimation which simultaneously infers the parameters of luminosity distance, sky location, detector frame masses, inclination, polarisation, spins, and time [48].

While there is redshift information encoded within the GW signal, it is degenerate with the source-frame mass, appearing only in the $\mathcal{M}_z$ term. In order to use GW standard sirens for cosmological measurements it is therefore necessary to find a way to break that degeneracy, or measure the redshift of the source using different means.

1.4 Cosmology with gravitational waves

The use of GWs as standard sirens for cosmology was first proposed by Schutz in 1986 [8]. He outlined two different methods using BNS mergers. In the first scenario the merger is well localised on the sky, allowing its EM counterpart to be identified and unambiguously linked to its host galaxy. The redshift of the host galaxy can then be used in combination with the distance estimate from the BNS to measure H_0. The alternative, if it is not possible to identify an EM counterpart, would be to treat all the galaxies lying within the GW localisation volume as potential hosts. The true host galaxy would have a redshift that, in combination with the distance from the GW event, would give the "real" H_0 value, while the other, randomly distributed galaxies would give other potential values for H_0. By combining the results from multiple GW detections, the contributions from the true hosts would stack over time, while the contributions from non-host galaxies would statistically average out. In 2012 Del Pozzo presented a rigorous Bayesian derivation of the methods outlined by Schutz, applicable to both the EM counterpart case and the galaxy catalogue case [49]. An important aspect that was discussed, but not mathematically addressed, was the scenario in which the galaxy catalogue does not contain all the galaxies within the localisation of volume for each event (*i.e.* it is incomplete), and therefore it is probable that the true host galaxy is not present in the catalogue for every event. It was acknowledged that the only way to ensure that H_0 remained unbiased in that case was to find a way to express the likelihood of observing a GW event whose host was *not* contained within the galaxy survey. For some time, this complication remained unaddressed.

The idea of using the EM counterpart of a BNS merger to identify its redshift gained traction several years before the detection of the first BNS, GW170817. When BNSs (or neutron star - black hole pairs (NSBHs)) merge, the incredibly violent event not only produces GWs, but also a strong EM emission across a wide band of wavelengths known as a kilonova [50, 51]. There was also a growing body of evidence that these mergers are progenitors of short gamma-ray bursts (GRBs) [52], bright flashes of gamma radiation lasting less than 2 seconds. The observation of a kilonova or short GRB in association with a GW signal would allow for the GW merger's location to be pin-pointed on the sky, and the identification of its host galaxy, and

the host galaxy's redshift, which would then allow for measurements of cosmological parameters [53–56].

The detection of the BNS merger GW170817 by the Advanced LIGO and Virgo detectors happened on the 17th August 2017 [10]. Just 1.7 seconds later, a GRB was detected and an alert was sent out by the Fermi Gamma-ray Burst Monitor. Initial estimates from the GW signal put the BNS at a distance of a mere 40 Mpc away, and localised it to within 31 deg^2 on the sky. An extensive observational campaign was triggered, involving many telescopes observing all across the EM spectrum. This led to multiple observations of the merger's optical transient, and the subsequent identification of its host galaxy, NGC4993 [11,12]. The combination of this GW and EM information allowed for the first ever measurement of H_0 made using GW standard sirens, with a value of $H_0 = 70^{+12}_{-8}$ km s^{-1} Mpc^{-1} [13] (conveniently straddling the H_0 tension region described in Section 1.2). This event was a defining moment for GW cosmology. In terms of a single event demonstrating the power of standard sirens as cosmological probes, it would be hard to think of a better example.

Of course, even a measurement as clean as this (in terms of being free from potential systematics in the cosmic distance ladder), still has other sources of uncertainty and potential pitfalls. Because GW170817 was so nearby, the impact of peculiar velocities (motion of the galaxy independent of the Hubble flow) on the redshift estimate play an important role. While addressed in [13], efforts have since been made to investigate the impact of peculiar velocities and their estimation on the final H_0 inference further [57,58]. Others have examined how the observation of the EM counterpart could provide information about GW170817's inclination, thus breaking the luminosity distance-inclination degeneracy and allowing a tighter constraint on d_L, and hence a tighter constraint on H_0 [59].

Across LIGO and Virgo's first observing run (O1) and second observing run (O2), GW170817 was the only merger with an identified counterpart, but its very existence indicated a high probability that future mergers with counterparts would be detected over the coming years, which would allow H_0 to be further constrained using the same method. Forecasts predict that $O(100)$ BNS-with-counterpart events would be required for an H_0 measurement with $\sim 1\%$ uncertainty [60–62], which would be adequate to discern between early- and late-time H_0 measurements.

In the meantime, attention returned to the galaxy catalogue method, with the intention of using it to analyse the remaining O1 and O2 events. The 2018 paper [60] also included a forecast for the constraint of H_0 using BNSs in the case where they were not observed with an EM counterpart and instead galaxy catalogues provided the missing information, and considered the possibility of using well-localised binary black holes (BBHs) to do the same thing. These forecasts again assumed that the galaxy catalogues would be complete, in terms of containing all potential host galaxies, but a methodology was outlined for incorporating an incomplete galaxy catalogue, by computing a completeness fraction to weight the GW likelihoods for a

host that was inside or outside the catalogue. This methodology was practically applied in [63], a reanalysis of GW170817 under the hypothetical scenario that it had *not* been observed with an EM counterpart. Independently [64] derived a fully Bayesian methodology for accounting for galaxy catalogue incompleteness when estimating H_0 with standard sirens, by modelling the limitations of EM surveys with a magnitude threshold. This method was tested rigorously using a set of MDAs, and demonstrated that it is possible to obtain an unbiased measurement of H_0 using incomplete galaxy catalogues (see Chapters 2 and 3 for details).

The first GW transient catalogue published by Advanced LIGO and Virgo contained eleven detections: one BNS and ten BBHs [14]. One particularly well localised BBH, GW170814, landed in the footprint of the Dark Energy Survey (DES), and an H_0 measurement from this dark siren alone (informative due to a dense cluster of galaxies within GW170814's localisation volume) was produced by the DES collaboration, in conjunction with the LIGO and Virgo collaborations [65]. A paper from the LIGO and Virgo collaborations used the method from [64] to combine all the O1 and O2 events with public galaxy catalogues to give the first measurement of H_0 combining information from multiple GW detections, with $H_0 = 68.7^{+17.0}_{-7.8}\text{km s}^{-1}\,\text{Mpc}^{-1}$ [66] – a result driven by GW170817 but with a modest improvement from the BBHs (see Chapter 4).

Now LIGO and Virgo's third observing run (O3) has ended, recording dozens more GW events (39 from the first half of the run (O3a) [15], and an as-yet unconfirmed number from O3b). GW170817 remains the only event with a confirmed EM counterpart, making the desired 1% uncertainty on the H_0 measurement still far from reach, but also underlining the importance of further developing cosmological analyses that do not rely on EM counterparts. The analysis of dark sirens, of which there are orders of magnitude more, will play an important role in constraining the GW measurement of H_0. These events also tend to be detectable to higher distances, as black holes are more massive than neutron stars, opening up the possibility of using them to probe cosmological parameters beyond H_0.

There are ways of breaking the degeneracy between source-frame mass and redshift for GWs which don't rely on either counterparts or galaxy catalogues. For BNS mergers, tidal effects between the two objects contribute to the phase evolution of the merger which can break the mass-redshift degeneracy [67]. The mass distribution for BNSs is narrow, which could break the degeneracy to some extent, as the source must have a redshift which would translate its detector-frame mass into the range of accepted source-frame masses [68–70] (though the detection of GW190425 [71], which appears to have come from an unusually heavy BNS system, may impact the strength of this argument). A similar method can be applied to BBHs. If the mass distribution for BBHs turns out to contain sharp features (such as the hypothesised pair-instability supernovae mass gap) then $O(10,000)$ BBH detections could be used to constrain H_0 to a competitive level [72]. This is more in the regime of third generation GW detectors, but the point stands: events which have very little galaxy catalogue support may be relatively uninformative on their own, but the population of those events as a whole contains useful information

for constraining cosmological parameters. In a similar vein, it has long been noted that a catalogue of detected CBCs could lead to the constraint of cosmological parameters based only on the distributions of their detected SNRs and chirp masses [73, 74].

The next few years of GW cosmology promise to be exciting. Detector sensitivities are improving between every observing run, leading to higher numbers of detections and higher SNRs. The Japan-based detector KAGRA began operation in February 2020, just before the end of O3, and its addition to the GW detector network means that future observing runs can look forward to having more coincident detections between observatories, increased SNRs and more precise luminosity distance estimates, as well as better localised events. With better localisation, the chances of identifying EM counterparts are improved, and for dark sirens the number of galaxies which could be considered as potential hosts will be reduced, making their contributions to measuring H_0 more informative.

The work in this thesis addresses the gap between Schutz's initial proposal to use galaxy catalogues for dark standard siren cosmology, and practically implementing this methodology to analyse real GW detections with incomplete galaxy catalogues. This work includes the first measurement of H_0 made by combining information from multiple GW events, using detections from Advanced LIGO and Virgo's first and second observing runs, and paves the way forward for O3 and beyond.

Chapter 2

A Bayesian methodology

Declaration: the majority of the material in this chapter is based on the methodology and appendix of my first-author paper (known informally as the MDA paper) [64].

Section 2.2 is taken directly from the MDA paper with some edits for style and phrasing.

Section 2.3 is based on the appendix of the MDA paper, but expanded with additional detail. The derivations within this section are my own work.

Section 2.4 is my own unpublished work.

2.1 Introduction

This chapter introduces an overarching methodology for using GW signals from multiple standard sirens to measure H_0. Section 2.2 lays out the Bayesian framework for combining information from multiple GW events, then looks at the information from a single event. Two scenarios are considered: that in which an EM counterpart is detected and associated with the host galaxy of the GW event, and that in which it is not, and galaxy catalogues are used to fill in the missing redshift information. Section 2.3 takes an in-depth look at this second case, and explicitly derives each of its terms, indicating where and how galaxy catalogue and other prior information enters the analysis. For completeness, section 2.4 takes a closer look at the counterpart case and derives a full expression for its implementation in the same style. Section 2.5 summarises the main takeaways and concludes this chapter.

2.2 Bayesian framework

This section presents a Bayesian framework for combining information from multiple GW detections for a measurement of H_0. It then examines the case where redshift information enters the analysis through a galaxy catalogue (section 2.2.1), and the case where it enters through an associated EM counterpart (section 2.2.2). A summary of parameters which appear explicitly in this overview is given in Table 2.1.

Parameter	Definition
H_0	The Hubble constant.
$N_{\rm det}$	The number of events detected during the observation period.
$x_{\rm GW}$	The GW data associated with some GW source, s.
$D_{\rm GW}$	Denotes that a GW signal was detected, *i.e.* that $x_{\rm GW}$ passed some detection statistic threshold.
g	Denotes that a galaxy is (G), or is not ($\bar{G}$), contained within the galaxy catalogue.
$x_{\rm EM}$	The EM data associated with some EM counterpart.
$D_{\rm EM}$	Denotes that an EM counterpart was detected, *i.e.* that $x_{\rm EM}$ passed some threshold.
I	Any additional information which is not explicitly stated, *e.g.* the underlying cosmological model. This term will be expanded when the information it contains becomes pertinent.

Table 2.1: A summary of the parameters present in the methodology.

The posterior probability on H_0 from $N_{\rm det}$ GW events can be expressed as

$$p(H_0|\{x_{\rm GW}\},\{D_{\rm GW}\},I) \propto p(H_0|I)p(N_{\rm det}|H_0,I)\prod_i^{N_{\rm det}} p(x_{{\rm GW}i}|D_{{\rm GW}i},H_0,I), \tag{2.1}$$

where $\{x_{\rm GW}\}$ is a set of GW data, corresponding to $N_{\rm det}$ detections. $D_{{\rm GW}i}$ indicates that $x_{{\rm GW}i}$ passed some threshold and was deemed "detected" as an event. $p(H_0|I)$ is the prior on H_0, and the term $p(N_{\rm det}|H_0,I)$ is the probability of detecting $N_{\rm det}$ events (over the time of observation, with the associated detector configuration) for a given value of H_0. It depends on the intrinsic astrophysical rate of events in the source frame, $R = \frac{\partial N_s}{\partial V \partial T}$, where N_s is the total number of sources. The total number of expected detections is given by $N_{\rm det} = R\langle VT\rangle$, where $\langle VT\rangle$ is the average of the surveyed comoving volume multiplied by the observation time. By choosing a scale-free prior on rate, $p(R|I) \propto 1/R$, the dependence on H_0 drops out [75]. The remaining term factorises into individual likelihoods for each detected event. The expression for a single GW event i, omitting the subscript i for brevity of notation, is as follows:

$$\begin{aligned} p(x_{\rm GW}|D_{\rm GW},H_0,I) &= \frac{p(D_{\rm GW}|x_{\rm GW},H_0,I)p(x_{\rm GW}|H_0,I)}{p(D_{\rm GW}|H_0,I)}, \\ &= \frac{p(x_{\rm GW}|H_0,I)}{p(D_{\rm GW}|H_0,I)}, \end{aligned} \tag{2.2}$$

where $p(D_{\rm GW}|x_{\rm GW},H_0,I) = 1$, since the analysis is only carried out when $x_{\rm GW}$ passes some detection threshold, *i.e.* if the SNR associated with $x_{\rm GW}$ passes some SNR threshold, or its false

alarm rate (FAR) is below some minimum FAR. In this analysis it is a prerequisite that the event has been detected. Calculating $p(D_{\rm GW}|H_0, I)$ requires integrating over all detectable realisations of GW events:

$$\begin{aligned} p(D_{\rm GW}|H_0, I) &= \int_{x_{\rm GW\,det}} p(D_{\rm GW}|x_{\rm GW}, H_0, I) p(x_{\rm GW}|H_0, I)\, dx_{\rm GW}, \\ &= \int_{x_{\rm GW\,det}} p(x_{\rm GW}|H_0, I)\, dx_{\rm GW}. \end{aligned} \tag{2.3}$$

To do so requires knowledge of the GW source population, as well as detector configuration, sensitivity, and detection threshold. For practical details on the calculation of $p(D_{\rm GW}|H_0, I)$ see section 3.3.2.

2.2.1 The galaxy catalogue method

In the galaxy catalogue case, the EM information enters the analysis in the form of galaxy catalogues, which provide a set of galaxies and their associated sky locations, redshifts, and apparent magnitudes (in various observation bands). As current galaxy catalogues cannot be considered complete out to the distances to which GW events are detectable (especially in the cases of BBHs), the possibility that the host galaxy is not contained within the galaxy catalogue has to be considered. The likelihood for a single GW event is marginalised over the case where the host galaxy is, and is not, in the catalogue (denoted by G and $\bar{G}$ respectively):

$$\begin{aligned} p(x_{\rm GW}|D_{\rm GW}, H_0, I) &= \sum_{g=G,\bar{G}} p(x_{\rm GW}|g, D_{\rm GW}, H_0, I) p(g|D_{\rm GW}, H_0, I), \\ &= p(x_{\rm GW}|G, D_{\rm GW}, H_0, I) p(G|D_{\rm GW}, H_0, I) \\ &\quad + p(x_{\rm GW}|\bar{G}, D_{\rm GW}, H_0, I) p(\bar{G}|D_{\rm GW}, H_0, I)\,. \end{aligned} \tag{2.4}$$

In the likelihood where the host galaxy is in the catalogue, denoted by the conditioning on G, the EM information is used to modify the priors on galaxy redshift, sky location, and (apparent) magnitude, which are common among all GW events using the same galaxy catalogue. This differs from the counterpart case (see section 2.2.2) where the EM data enters the expression as a likelihood term, $x_{\rm EM}$, a transient which is informative for only one GW event.

In the likelihood where the host galaxy is not in the catalogue, the complementary condition $\bar{G}$ implies the inclusion of information about the limitations of the EM survey. To do so, catalogues can be modelled as having an apparent magnitude threshold, $m_{\rm th}$, since galaxy catalogues are flux-limited. This magnitude threshold, alongside the intrinsic (absolute) brightness of a galaxy and its luminosity distance, determines if a specific galaxy is inside or outside the galaxy catalogue. To determine the probability that a (host) galaxy is inside or outside the galaxy catalogue requires additional prior assumptions on the intrinsic absolute magnitude distribution of

galaxies in the universe, their distribution in redshift beyond the catalogue, and assumptions on the preferential weighting for a galaxy to host a GW event.

A detailed derivation of each of the terms in Eq. 2.4 is carried out in section 2.3.

2.2.2 The counterpart method

Now to consider the case where an EM counterpart is associated with the GW signal. The main difference between this and the method described above is the inclusion of a likelihood term for the EM counterpart data, mirroring that of the GW data.

The likelihood in this case, (the term within the product in Eq. 2.1), is given by:

$$\begin{aligned} p(x_{\rm GW}, x_{\rm EM}|D_{\rm GW}, D_{\rm EM}, H_0, I) &= \frac{p(x_{\rm GW}, x_{\rm EM}|H_0, I)p(D_{\rm GW}, D_{\rm EM}|x_{\rm GW}, x_{\rm EM}, H_0, I)}{p(D_{\rm GW}, D_{\rm EM}|H_0, I)}, \\ &= \frac{p(x_{\rm GW}|H_0, I)p(x_{\rm EM}|H_0, I)}{p(D_{\rm EM}|D_{\rm GW}, H_0, I)p(D_{\rm GW}|H_0, I)}. \end{aligned} \tag{2.5}$$

where $x_{\rm EM}$ refers to the EM counterpart data and $D_{\rm EM}$ denotes that the counterpart was detected. In the numerator it is assumed that the GW and EM data are independent of each other and so the joint GW-EM likelihood factors out. The term $p(D_{\rm GW}, D_{\rm EM}|x_{\rm GW}, x_{\rm EM}, H_0, I)$ is further factorised as $p(D_{\rm EM}|D_{\rm GW}, x_{\rm GW}, x_{\rm EM}, H_0, I)p(D_{\rm GW}|x_{\rm GW}, x_{\rm EM}, H_0, I)$. The first term is equal to 1, as this method is only used when an EM counterpart has been observed, meaning that by definition $x_{\rm EM}$ has passed some threshold for detectability set by EM telescopes. The second term also goes to 1, due to the same threshold argument as in section 2.2. If, as in [13], the detectability of EM counterparts is assumed to extend well beyond the distance to which compact binaries are detectable, a further simplification is allowed, where the term $p(D_{\rm EM}|D_{\rm GW}, H_0, I) \sim 1$. Section 2.4 takes a closer look at Eq. 2.5 and gives a detailed derivation of its expanded form.

In an ideal scenario, the observation of an EM counterpart will allow for the unique identification of one of the galaxies in the GW event localisation as the host of the GW event. In the case where the EM counterpart cannot be unambiguously linked to a host galaxy, $x_{\rm EM}$ instead picks out several potential host galaxies within the volume of the event, and this uncertainty in the host must be marginalised over. As in the galaxy catalogue case, incompleteness in the set of potential host galaxies must be considered if the host galaxy cannot be identified with certainty.

2.3 An in-depth look at the galaxy catalogue method

This section presents a detailed look into the galaxy catalogue method introduced in section 2.2.1. The approach is summarised in Fig. 2.1, a network diagram which shows how each of the parameters of this extended derivation fit together and their dependencies on each other. The parameters which appear in this diagram, and in the following subsections, are defined in Table 2.2.

Parameter	Definition
H_0	The Hubble constant
$x_{\rm GW}$	The GW data associated with some GW source, s.
$D_{\rm GW}$	Denotes that a GW signal was detected, *i.e.* that $x_{\rm GW}$ passed some detection statistic threshold.
g	Denotes that a galaxy is (G), or is not ($\bar{G}$), contained within the galaxy catalogue.
s	Denotes that a GW signal was emitted.
M	Absolute magnitude.
z	Redshift.
Ω	Sky location (right ascension and declination).
d_L	Luminosity distance.
m	Apparent magnitude.
$m_{\rm th}$	Apparent magnitude threshold of the galaxy catalogue.
$\det_{\rm th}$	Detection threshold of the GW detector network, *i.e.* an SNR or FAR threshold.
Cosmological model	The cosmological model assumed for the analysis. Typically a Friedmann-Lemaître-Robertson-Walker universe.
Schechter parameters	The parameters which characterise the assumed absolute magnitude distribution of galaxies in the universe.
GW population	The assumed underlying population of GW sources. This includes the mass distribution of the population, and the population's evolution with redshift.
Source frame parameters	Source frame parameters of a GW source, *e.g.* component masses, spins, inclination and polarisation.
Detector frame parameters	As above, but redshifted into the detector frame.
Detector configuration	The network set up, including which detectors are included in the search and their noise floors.
I	Any additional information which is not explicitly stated, *e.g.* the underlying cosmological model. This term will be expanded when the information it contains becomes pertinent.

Table 2.2: A summary of the parameters present in the network diagram, Fig. 2.1.

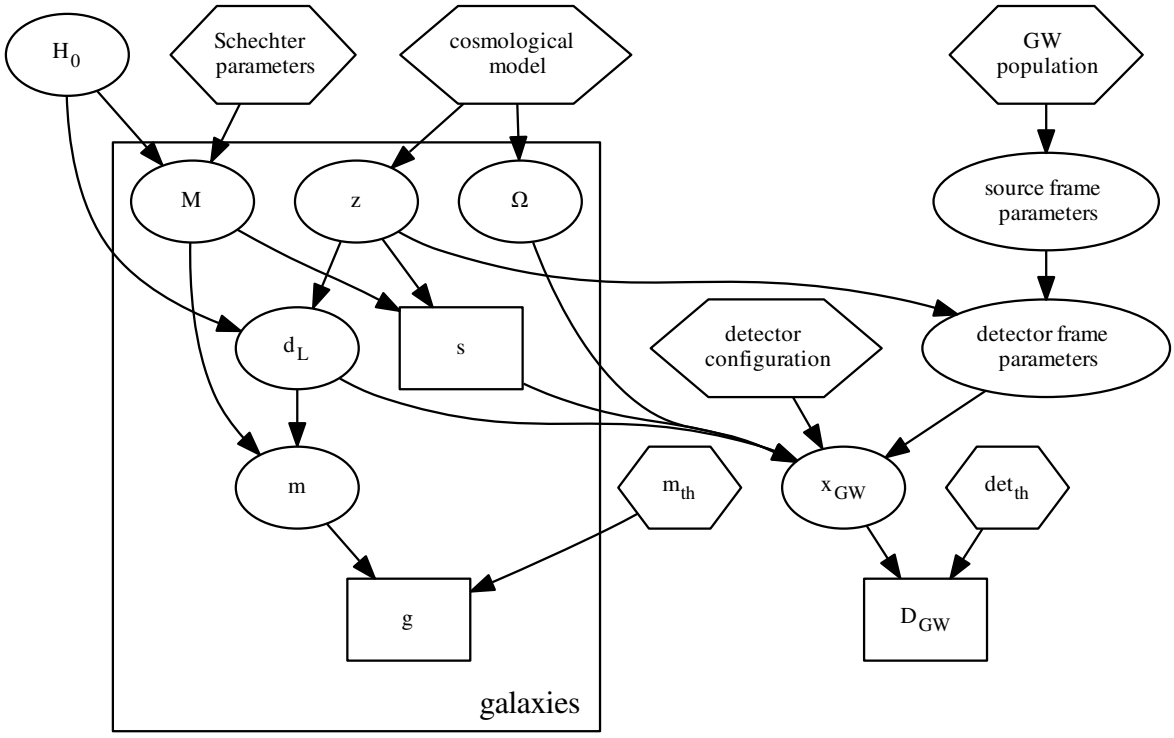

Figure 2.1: A network diagram showing how the main parameters of the methodology interlink. Circular nodes denote ordinary parameters. Hexagonal nodes denote assumed knowns. Rectangular nodes denote binary flags. The arrows indicate the dependence of each parameter on the parameters which feed into them. The parameters grouped in the "galaxies" cluster are those which can be evaluated for every galaxy in the universe.

Whenever there is ambiguity about the host of a GW event, the probability that some galaxies are more likely hosts than others must be considered. This materialises in two ways. Firstly, that more luminous galaxies could be considered to be more likely hosts, due to being a tracer for the overall mass of the galaxy or its star-formation rate. Secondly, that the rate of GW mergers may have evolved over time, and therefore be redshift-dependent.

Included in the term I on the right-hand side of all probability terms is an important assumption: that a real GW signal has been emitted (denoted from here on by the term s). When considering the terms $x_{\rm GW}$ and $D_{\rm GW}$ the implicit assumption is that a real gravitational wave has been detected, and it is not a false alarm. Up to this point, the term s has been implicit, but when considering other astrophysical parameters its existence becomes important, and it can help to include it explicitly on the right-hand side of probability terms. For example, when considering priors on the distribution of galaxies in the universe in terms of redshift (z), sky location (Ω) and absolute magnitude (M), it is not the prior on all galaxies which is important, but the prior on potential host galaxies for GW events.

The prior on redshift, sky location, absolute magnitude and apparent magnitude (m), for

galaxies which host a GW source can be expressed as

$$\begin{aligned} p(z, \Omega, M, m|s, H_0, I) &= p(m|z, \Omega, M, s, H_0, I)p(z, \Omega, M|s, H_0, I), \\ &= p(m|z, \Omega, M, s, H_0, I)p(z|s, I)p(\Omega|I)p(M|s, H_0, I), \end{aligned} \tag{2.6}$$

where it is assumed that z, Ω and M are conditionally independent given s, H_0.[1] The relationship between m and z, M and H_0 means that if the latter 3 are known, m can be calculated directly:

$$\begin{aligned} p(z, \Omega, M, m|s, H_0, I) &= \delta(m - m(z, M, H_0))p(z|s, I)p(\Omega|I)p(M|s, H_0, I), \\ &= \delta(m - m(z, M, H_0))\frac{p(s|z, I)p(z|I)}{p(s|I)}p(\Omega|I)\frac{p(s|M, I)p(M|H_0, I)}{p(s|H_0, I)}. \end{aligned} \tag{2.7}$$

The probability $p(z|I)$ is the prior distribution of galaxies in the universe, taken to be (say) uniform in comoving volume, $p(\Omega|I)$ is the prior on galaxy sky location, assumed uniform over the celestial sphere, and $p(M|H_0, I)$ is the distribution of galaxy absolute magnitudes given by *e.g.* a Schechter function. In the sections below it will be seen that the terms $p(s|I)$ and $p(s|H_0, I)$ cancel with other terms, and so their exact form does not need to be considered. Assuming the universe follows the cosmological principle and is isotropic, s retains no dependence on Ω, as it is assumed that no particular direction will be more likely to source GW signals than any other, and so no $p(s|\Omega, I)$ term appears in the equation. The probability of a galaxy with absolute magnitude M being host to a GW event, $p(s|M, I)$, can take the form

$$p(s|M, I) \propto \begin{cases} L(M) & \text{if GW hosting probability is proportional to luminosity} \\ \text{constant} & \text{if GW hosting probability is independent of luminosity.} \end{cases} \tag{2.8}$$

This is a form of luminosity weighting potential host galaxies. The term $p(s|z, I)$ represents the dependence of the merger rate on redshift, with

$$p(s|z, I) \propto \begin{cases} \dfrac{1}{1+z}R(z) & \text{if rate evolves with redshift} \\ \dfrac{1}{1+z} & \text{if rate is does not evolve with redshift.} \end{cases} \tag{2.9}$$

Here, the factor of $1/(1+z)$ comes from the time delay between a GW signal being emitted, and its detection on Earth. Even if the merger rate of CBCs is constant throughout time in the source frame, and therefore independent of redshift, this term must be included in order to convert from the source frame to the detector frame.

Now to consider the individual components of Eq. 2.4. The parameter s is included explicitly in all terms for clarity.

[1] For an absolute magnitude distribution which evolves with redshift, this assumption will need to be revisited. It is not within the scope of this thesis, but will be discussed briefly in Chapter 6.

2.3.1 Likelihood when the host is in the galaxy catalogue

A full expression for the likelihood of the GW data when the host galaxy is inside the galaxy catalogue, $p(x_{\rm GW}|G, D_{\rm GW}, s, H_0, I)$, can be obtained by marginalising over redshift, sky location, absolute magnitude and apparent magnitude. First expanding as in Eq. 2.2 gives

$$p(x_{\rm GW}|G, D_{\rm GW}, s, H_0, I) = \frac{p(x_{\rm GW}|G, s, H_0, I)}{p(D_{\rm GW}|G, s, H_0, I)}. \tag{2.10}$$

Focussing on the numerator, and noting that the GW data, $x_{\rm GW}$, is independent of G, m and M, it can be written as

$$p(x_{\rm GW}|G, s, H_0, I) = \iiiint p(x_{\rm GW}|z, \Omega, s, H_0, I)p(z, \Omega, M, m|G, s, H_0, I)\, dz d\Omega dM dm. \tag{2.11}$$

The presence of the term G on the right-hand side of $p(z, \Omega, M, m|G, s, H_0, I)$ indicates that the information on z, Ω, M and m is coming from the galaxy catalogue, with each galaxy providing a specific redshift, sky location and apparent magnitude (ignoring uncertainties on these quantities for the time being). While not obtained directly from the catalogue, the absolute magnitude for each galaxy is uniquely calculable using the relationship between m, z, H_0 and M. As such the term can be further factorised as

$$p(z, \Omega, M, m|G, s, H_0, I) = \frac{p(s|z, M(z, m, H_0), I)\delta(M - M(z, m, H_0))p(z, \Omega, m|G, I)}{p(s|G, H_0, I)}. \tag{2.12}$$

Here Ω has been dropped from the right-hand side of $p(s|z, M(z, m, H_0), I)$, and the dependence of s on m and H_0 has been recognised to only enter through their relationship with M. Equation 2.11 can now be integrated over absolute magnitude to obtain

$$\begin{aligned} p(x_{\rm GW}|G, s, H_0, I) = \frac{1}{p(s|G, H_0, I)} \iiint & p(x_{\rm GW}|z, \Omega, s, H_0, I)p(s|z, M(z, m, H_0), I) \\ & \times p(z, \Omega, m|G, I)\, dz d\Omega dm. \end{aligned} \tag{2.13}$$

Here $p(z, \Omega, m|G, I)$ is the prior on z, Ω and m of all the galaxies in the catalogue, while $p(s|z, M(z, m, H_0), I)$ is the probability that a galaxy with redshift z and absolute magnitude M would host a GW source. Assuming, in the first instance, that galaxies in the catalogue can be approximated as a series of delta functions on z, Ω and m, the integral can be converted to a sum

over the N galaxies in the catalogue:

$$\begin{aligned} p(x_{\rm GW}|G, s, H_0, I) &= \frac{1}{p(s|G, H_0, I)} \frac{1}{N} \sum_{i=1}^{N} p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) p(s|z_i, M(z_i, m_i, H_0), I), \\ &= \frac{1}{p(s|G, H_0, I)} \frac{1}{N} \sum_{i=1}^{N} p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) \frac{p(z_i, M(z_i, m_i, H_0)|s, I) p(s|I)}{p(z_i, M(z_i, m_i, H_0)|I)}, \\ &= \frac{1}{p(s|G, H_0, I)} \frac{1}{N} \sum_{i=1}^{N} p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) \frac{p(z_i|s, I) p(M(z_i, m_i, H_0)|z_i, s, I) p(s|I)}{p(z_i|I) p(M(z_i, m_i, H_0)|z_i, I)}. \end{aligned} \tag{2.14}$$

Further applying Bayes theorem to $p(z_i|s, I)$ and $p(M(z_i, m_i, H_0)|z_i, s, I)$ results in the following equation. From here on, the dependence of $p(M(z_i, m_i, H_0)|z_i, s, I)$ on z_i is encapsulated within the term $M(z_i, m_i, H_0)$ by definition, and so z_i is dropped from the right-hand side of that term.

$$\begin{aligned} p(x_{\rm GW}|G, s, H_0, I) &= \frac{1}{p(s|G, H_0, I)} \frac{1}{N} \sum_{i=1}^{N} p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) \\ &\qquad \times \frac{p(s|z_i, I) p(z_i|I) p(s|M(z_i, m_i, H_0), I) p(M(z_i, m_i, H_0)|I) p(s|I)}{p(s|I) p(s|I) p(z_i|I) p(M(z_i, m_i, H_0)|I)}, \\ &= \frac{1}{p(s|G, H_0, I) p(s|I)} \frac{1}{N} \sum_{i=1}^{N} p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) p(s|z_i, I) p(s|M(z_i, m_i, H_0, I)), \end{aligned} \tag{2.15}$$

where the majority of the terms have cancelled, and the leftover $1/p(s|I)$ term is brought outside the sum.

Returning to Eq. 2.10 and expanding the denominator, $p(D_{\rm GW}|G, s, H_0, I)$, by marginalising over z, Ω, m and M produces an equivalent term to Eq. 2.15, but with a $p(D_{\rm GW}|z_i, \Omega_i, s, H_0, I)$ term instead of $p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I)$. Substituting this and Eq. 2.15 back into Eq. 2.10 gives the likelihood for the "in catalogue" part of the galaxy catalogue method, where the factors $p(s|G, H_0, I)$ and $p(s|I)$ have cancelled:

$$p(x_{\rm GW}|G, D_{\rm GW}, s, H_0, I) = \frac{\sum_{i=1}^{N} p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) p(s|z_i, I) p(s|M(z_i, m_i, H_0), I)}{\sum_{i=1}^{N} p(D_{\rm GW}|z_i, \Omega_i, s, H_0, I) p(s|z_i, I) p(s|M(z_i, m_i, H_0), I)}. \tag{2.16}$$

While this equation is fine for the simplistic case in which galaxies in the catalogue are delta-function-like, with no associated uncertainties, this is far from the case in reality. Galaxy catalogues often provide uncertainties on the quantities quoted within them, the most significant of which tend to be redshift uncertainties (see section 4.3.1 for a more detailed discussion of redshift uncertainties). Where redshift uncertainties are provided for each galaxy in the catalogue

(such that the redshift distribution of galaxy i is given by $p(z_i|I)$), these can be incorporated as:

$$p(x_{\rm GW}|G, D_{\rm GW}, s, H_0, I) = \frac{\sum_{i=1}^{N} \int p(x_{\rm GW}|z_i, \Omega_i, s, H_0, I) p(s|z_i, I) p(s|M(z_i, m_i, H_0), I) p(z_i|I)\, dz_i}{\sum_{i=1}^{N} \int p(D_{\rm GW}|z_i, \Omega_i, s, H_0, I) p(s|z_i, I) p(s|M(z_i, m_i, H_0), I) p(z_i|I)\, dz_i}, \tag{2.17}$$

where the uncertainty in the redshift of each galaxy is now marginalised over inside the sum over galaxies.

Equation 2.17 gives an expression for the likelihood of the GW data given that the host galaxy is in the galaxy catalogue. The numerator requires evaluating the GW likelihood at the location of all the galaxies within the galaxy catalogue. Different values of H_0 will cause galaxies to pick out different parts of the GW likelihood, and the strength of the contribution from an individual galaxy will depend on how well its redshift matches the luminosity distance of the event for that value of H_0, weighted by the galaxy's probability of hosting a GW event due to its astrophysical properties. There is a great deal of symmetry in how the denominator is evaluated – instead of considering the GW likelihood for the event under consideration, it is the probability of detecting an event (any event from that population of CBCs) which is evaluated for every galaxy in the catalogue.

2.3.2 Probability the host galaxy is in the galaxy catalogue

The probability that the host galaxy is inside the galaxy catalogue, given that a GW signal was detected, can be expressed as

$$p(G|D_{\rm GW}, s, H_0, I) = \iiiint p(G|z, \Omega, M, m, D_{\rm GW}, s, H_0, I) \times p(z, \Omega, M, m|D_{\rm GW}, s, H_0, I)\, dz d\Omega dM dm. \tag{2.18}$$

Assuming that the galaxy catalogue is apparent magnitude-limited, such that only galaxies which pass some detection threshold based on their apparent magnitude are contained within it, the term $p(G|z, \Omega, M, m, D_{\rm GW}, s, H_0, I)$ depends only on m, and can be approximated by a Heaviside step function at the detection threshold $m_{\rm th}(\Omega)$ (where the inclusion of Ω allows for the magnitude threshold to be directionally-dependent, as EM surveys may have differing levels of completeness across the sky). Doing this, and expanding the remaining term using Bayes

theorem gives

$$p(G|D_{\rm GW}, s, H_0, I) = \iiiint \Theta[m_{\rm th}(\Omega) - m] p(z, \Omega, M, m|D_{\rm GW}, s, H_0, I)\, dz d\Omega dM dm,$$
$$= \iiiint \Theta[m_{\rm th}(\Omega) - m] \frac{p(D_{\rm GW}|z, \Omega, s, H_0, I) p(z, \Omega, M, m|s, H_0, I)}{p(D_{\rm GW}|s, H_0, I)}\, dz d\Omega dM dm. \tag{2.19}$$

The prior on all GW-hosting galaxies, $p(z, \Omega, M, m|s, H_0, I)$, can now be expanded as in Eq 2.7 to give

$$p(G|D_{\rm GW}, s, H_0, I) = \frac{1}{p(s|I)p(s|H_0, I)} \frac{1}{p(D_{\rm GW}|s, H_0, I)}$$
$$\times \int d\Omega \int dM \int_0^{z(M, m_{\rm th}(\Omega), H_0)} dz\; p(D_{\rm GW}|z, \Omega, s, H_0, I) p(s|z, I) p(z|I) p(\Omega|I) p(s|M, I) p(M|H_0, I), \tag{2.20}$$

where the Heaviside step function at $m_{\rm th}(\Omega)$ has been translated into an $m_{\rm th}$-dependent upper redshift limit.[2]

The term $1/p(D_{\rm GW}|s, H_0, I)$ in Eq. 2.20, which has been brought outside the integrals, should also be marginalised over z, Ω, m and M. As it does not contain the G term, no magnitude threshold enters the expression, and the limits on the integrals cover the full parameter space. Writing out the expansion explicitly produces equivalent $p(s|I)$ and $p(s|H_0, I)$ terms which cancel with those in Eq. 2.20, and the full expression for the probability that the host galaxy is inside the catalogue becomes

$$p(G|D_{\rm GW}, s, H_0, I) =$$
$$\frac{\int d\Omega \int dM \int_0^{z(M, m_{\rm th}(\Omega), H_0)} dz\; p(D_{\rm GW}|z, \Omega, s, H_0, I) p(s|z, I) p(z|I) p(\Omega|I) p(s|M, I) p(M|H_0, I)}{\iiint p(D_{\rm GW}|z, \Omega, s, H_0, I) p(s|z, I) p(z|I) p(\Omega|I) p(s|M, I) p(M|H_0, I)\, dz d\Omega dM}. \tag{2.21}$$

In summary, the probability that the host galaxy is inside the galaxy catalogue, given that a GW signal was detected, is calculated by multiplying the probability of detecting a signal from an event at redshift z and sky location Ω, by the prior distribution of GW host galaxies in redshift, sky location, and absolute magnitude. This quantity is integrated over z, Ω and M, where the upper limit on z is set by the apparent magnitude threshold of the galaxy catalogue. As $p(G|D_{\rm GW}, s, H_0, I)$ is a probability, not a probability density, it must therefore lie between 0 and 1, which dividing through by the denominator ensures. The host galaxy must be somewhere,

[2]Converting the apparent magnitude threshold to an integral limit gives two options: the first, shown here, is as a limit on the redshift integral. The second, which is mathematically equivalent, is to place the limit on absolute magnitude. While it may be more intuitive to consider the first at this stage, the second will play an important role in section 4.4.3 where the option of adding an additional redshift cut to the galaxy catalogue will be explored.

and so the corresponding probability that the host galaxy is *not* in the galaxy catalogue can be found using

$$p(\bar{G}|D_{\rm GW}, s, H_0, I) = 1 - p(G|D_{\rm GW}, s, H_0, I). \tag{2.22}$$

2.3.3 Likelihood when the host is not in the galaxy catalogue

In order to compute the likelihood of the GW data in the case where the host galaxy is outside the galaxy catalogue, the first step is to use Bayes theorem to expand $p(x_{\rm GW}|\bar{G}, D_{\rm GW}, s, H_0, I)$ (as in Eq. 2.2) to become

$$p(x_{\rm GW}|\bar{G}, D_{\rm GW}, s, H_0, I) = \frac{p(x_{\rm GW}|\bar{G}, s, H_0, I)}{p(D_{\rm GW}|\bar{G}, s, H_0, I)}. \tag{2.23}$$

Focussing on the numerator, and expanding it to explicitly marginalise over z, Ω, M and m gives

$$\begin{aligned} p(x_{\rm GW}|\bar{G}, s, H_0, I) &= \iiiint p(x_{\rm GW}|z, \Omega, M, m, \bar{G}, s, H_0, I) p(z, \Omega, M, m|\bar{G}, s, H_0, I)\, dz d\Omega dM dm, \\ &= \frac{1}{p(\bar{G}|s, H_0, I)} \iiiint p(x_{\rm GW}|z, \Omega, s, H_0, I) p(\bar{G}|z, \Omega, M, m, s, H_0, I) \\ &\qquad \times p(z, \Omega, M, m|s, H_0, I)\, dz d\Omega dM dm. \end{aligned} \tag{2.24}$$

The probability that a galaxy is outside the galaxy catalogue, $p(\bar{G}|z, \Omega, M, m, s, H_0, I)$, depends only on m and the threshold of the catalogue, $m_{\rm th}(\Omega)$, and can as such be represented as a Heaviside step function. The prior term, $p(z, \Omega, M, m|s, H_0, I)$ can be expanded as it was in Eq 2.7. Substituting this into Eq. 2.24 and converting the Heaviside step function to a $m_{\rm th}$-dependent redshift limit (in this case the *lower* limit, as this is the out-of-catalogue case) gives

$$\begin{aligned} p(x_{\rm GW}|\bar{G}, s, H_0, I) = \frac{1}{p(s|I)p(s|H_0, I)} \frac{1}{p(\bar{G}|s, H_0, I)} \int d\Omega \int dM \int_{z(M, m_{\rm th}(\Omega), H_0)}^{\infty} dz\; p(x_{\rm GW}|z, \Omega, s, H_0, I) \\ \times p(s|z, I) p(z|I) p(\Omega|I) p(s|M, I) p(M|H_0, I). \end{aligned} \tag{2.25}$$

The denominator of Eq. 2.23, $p(D_{\rm GW}|\bar{G}, s, H_0, I)$, can be expanded in the same way as the numerator to give an equivalent term:

$$\begin{aligned} p(D_{\rm GW}|\bar{G}, s, H_0, I) = \frac{1}{p(s|I)p(s|H_0, I)} \frac{1}{p(\bar{G}|s, H_0, I)} \int d\Omega \int dM \int_{z(M, m_{\rm th}(\Omega), H_0)}^{\infty} dz\; p(D_{\rm GW}|z, \Omega, s, H_0, I) \\ \times p(s|z, I) p(z|I) p(\Omega|I) p(s|M, I) p(M|H_0, I). \end{aligned} \tag{2.26}$$

Substituting this, and Eq. 2.25 back into Eq 2.23, the factors outside the integral cancel, and the

final expression becomes

$$p(x_{\rm GW}|\bar{G}, D_{\rm GW}, s, H_0, I) = \frac{\int d\Omega \int dM \int_{z(M,m_{\rm th}(\Omega),H_0)}^{\infty} dz\; p(x_{\rm GW}|z,\Omega,s,H_0,I)p(s|z,I)p(z|I)p(\Omega|I)p(s|M,I)p(M|H_0,I)}{\int d\Omega \int dM \int_{z(M,m_{\rm th}(\Omega),H_0)}^{\infty} dz\; p(D_{\rm GW}|z,\Omega,s,H_0,I)p(s|z,I)p(z|I)p(\Omega|I)p(s|M,I)p(M|H_0,I)}. \tag{2.27}$$

Equation 2.27 gives an expression for calculating the GW data likelihood when the host galaxy is not contained within the galaxy catalogue. The numerator is computed by multiplying the GW likelihood by the galaxy priors on redshift, sky location and absolute magnitude, weighted by the probabilities of galaxies with given redshifts or absolute magnitudes being host to GW sources. This quantity is integrated over redshift, sky location and absolute magnitude, and the apparent magnitude threshold of the galaxy catalogue introduces a lower limit on the redshift integral, such that only the parameter space beyond the catalogue is considered. Again there is a great deal of symmetry in the calculation of the denominator, where the only difference is that the GW likelihood is replaced with the probability of detecting any GW event from the population at a given redshift and sky location.

The "empty catalogue" case

A special case of the "out-of-catalogue" likelihood occurs when no galaxy catalogue is provided, but the catalogue method is still used to measure H_0. This is informally known as the "empty catalogue" case. The apparent magnitude threshold of the catalogue tends to $-\infty$, resulting in the integral limit $z(M, m_{\rm th}(\Omega), H_0) \to 0$, which causes $p(G|D_{\rm GW}, s, H_0, I) = 0$, and $p(\bar{G}|D_{\rm GW}, s, H_0, I) = 1$. As a result, the only contribution to the final likelihood is from Eq. 2.27. Taking it as a starting point and noting that, now the lower integral limit on redshift is zero, the correlation between redshift and absolute magnitude is removed, meaning that the integrals can be separated:

$$p(x_{\rm GW}|\bar{G}, D_{\rm GW}, s, H_0, I) = \frac{\int p(s|M,I)p(M|H_0,I)dM}{\int p(s|M,I)p(M|H_0,I)dM} \times \frac{\iint p(x_{\rm GW}|z,\Omega,s,H_0,I)p(\Omega|I)p(s|z,I)p(z|I)dzd\Omega}{\iint p(D_{\rm GW}|z,\Omega,s,H_0,I)p(\Omega|I)p(s|z,I)p(z|I)dzd\Omega}, \tag{2.28}$$

The integrals over absolute magnitude cancel, leaving the following expression, which has no dependence on galaxy catalogue information, or assumptions about the luminosity distribution

of galaxies in the universe:

$$p(x_{\rm GW}|\bar{G}, D_{\rm GW}, s, H_0, I) = \frac{\iint p(x_{\rm GW}|z, \Omega, s, H_0, I)p(\Omega|I)p(s|z, I)p(z|I)\ dzd\Omega}{\iint p(D_{\rm GW}|z, \Omega, s, H_0, I)p(\Omega|I)p(s|z, I)p(z|I)\ dzd\Omega}. \tag{2.29}$$

This might seem unexpected, given that H_0 explicitly appears in the absolute magnitude prior, $p(M|H_0, I)$, but this makes sense when it is assumed that z and M are independent (*i.e.* that the absolute magnitude distribution of galaxies does not evolve with redshift). The physical distribution of host galaxies in the universe is independent of M and of H_0. Choosing to luminosity weight the host galaxies, or not, will have no bearing on their redshift distribution, and so the empty catalogue analysis will be independent of these assumptions.

While relatively uninformative, this method is a useful tool for probing how effectively GW selection effects are being modelled, and to what extent the analysis is affected by prior assumptions on the GW population. It can also help to disentangle the EM contribution from these assumptions, and quantify, to an extent, how much useful catalogue information is entering the result.

2.4 An in-depth look at the counterpart method

For completeness it is worth taking a closer look at the EM counterpart case and how it might be evaluated in reality. Taking Eq. 2.5 as a starting point and marginalising over redshift and sky location in the numerator gives

$$\begin{aligned} p(x_{\rm GW}|s, H_0, I)p(x_{\rm EM}|s, H_0, I) &= \iint p(x_{\rm GW}|z, \Omega, s, H_0, I)p(x_{\rm EM}|z, \Omega, s, H_0, I)p(z, \Omega|s, I)\ dzd\Omega, \\ &= \frac{1}{p(s|I)} \iint p(x_{\rm GW}|z, \Omega, s, H_0, I)p(x_{\rm EM}|z, \Omega, s, H_0, I)p(\Omega|I) \\ &\qquad \times p(s|z, I)p(z|I)\ dzd\Omega, \end{aligned} \tag{2.30}$$

where s indicating a real source is included explicitly on the right-hand side of each term. Note the inclusion of the priors on galaxy sky location and redshift are included, as well as the merger rate evolution term $p(s|z, I)$. In the case that the EM counterpart can be unambiguously linked to a host galaxy, the EM likelihood, $p(x_{\rm EM}|z, \Omega, s, H_0, I)$, provides the redshift and sky location of the host. If the host galaxy's redshift and sky location are known perfectly, ($z_{\rm host}$ and $\Omega_{\rm host}$), then $p(x_{\rm EM}|z, \Omega, s, H_0, I)$ works like a delta-function on z and Ω, picking out the location of the

counterpart:

$$\begin{aligned} &p(x_{\rm GW}|s, H_0, I)p(x_{\rm EM}|s, H_0, I) \\ &\qquad = \frac{1}{p(s|I)} \iint p(x_{\rm GW}|z, \Omega, H_0, I)\delta(z - z_{\rm host})\delta(\Omega - \Omega_{\rm host})p(\Omega|I)p(s|z, I)p(z|I)\, dzd\Omega, \\ &\qquad = \frac{1}{p(s|I)} p(x_{\rm GW}|z_{\rm host}, \Omega_{\rm host}, H_0, I)p(\Omega = \Omega_{\rm host}|I)p(s|z_{\rm host}, I)p(z = z_{\rm host}|I), \\ &\qquad \propto p(x_{\rm GW}|z_{\rm host}, \Omega_{\rm host}, H_0, I). \end{aligned} \tag{2.31}$$

Here, $p(\Omega = \Omega_{\rm host}|I)$ and $p(z = z_{\rm host}|I)$ are the sky location and redshift priors evaluated at the delta-function location of the host galaxy, meaning that they are just numbers which will not impact the final result. Similarly, because $p(s|z_{\rm host}, I)$ has no H_0-dependence its inclusion will not influence the final result beyond some constant.

In reality, however, there will be some uncertainty associated with the redshift of the galaxy, due both to measurement uncertainty, and uncertainty in the peculiar velocity corrections. In such a case this uncertainty should be marginalised over. Assuming the redshift distribution of the host galaxy is no longer a delta function, but is described by $p(z_{\rm host}|I)$, this gives

$$\begin{aligned} &p(x_{\rm GW}|s, H_0, I)p(x_{\rm EM}|s, H_0, I) \\ &\qquad = \frac{1}{p(s|I)} \int p(x_{\rm GW}|z_{\rm host}, \Omega_{\rm host}, H_0, I)p(\Omega = \Omega_{\rm host}|I)p(s|z_{\rm host}, I)p(z = z_{\rm host}|I)p(z_{\rm host}|I)\, dz_{\rm host}. \end{aligned} \tag{2.32}$$

Assuming the uncertainty on the sky location of the host is negligible, $p(\Omega = \Omega_{\rm host}|I)$ remains some constant which can be ignored. The redshift terms $p(s|z_{\rm host}, I)$ and $p(z = z_{\rm host}|I)$ should be included explicitly, however, as they will have some impact on the redshift integral. The uncertainty on the host galaxy's redshift would have to be relatively large for this to have any meaningful impact on the result, but it is worth bearing in mind.

If the assumption is made (as mentioned in section 2.2.2) that $p(D_{\rm EM}|D_{\rm GW}, s, H_0, I) \sim 1$, the remaining term in the denominator of Eq. 2.5, $p(D_{\rm GW}|s, H_0, I)$, is simply the probability of detecting a GW signal from the GW population that $x_{\rm GW}$ comes from, marginalised over redshift and sky location. The redshift prior in this case would be (say) uniform in comoving volume, weighted by the probability of a galaxy at redshift z being host to a GW source. Doing so gives a final expression for the counterpart case:

$$\begin{aligned} &p(x_{\rm GW}, x_{\rm EM}|D_{\rm GW}, D_{\rm EM}, s, H_0, I) \\ &\qquad \propto \frac{\int p(x_{\rm GW}|z_{\rm host}, \Omega_{\rm host}, H_0, I)p(s|z_{\rm host}, I)p(z = z_{\rm host}|I)p(z_{\rm host}|I)\, dz_{\rm host}}{\iint p(D_{\rm GW}|z, \Omega, s, H_0, I)p(\Omega|I)p(s|z, I)p(z|I)\, dzd\Omega}. \end{aligned} \tag{2.33}$$

A more in-depth treatment of the term $p(D_{\rm EM}|D_{\rm GW}, s, H_0, I)$ is beyond the scope of this thesis,

but it will likely become of interest in the near future, when EM counterparts at higher redshifts are detected (see Chapter 6 for a discussion on this point).

2.5 Conclusion

This chapter introduced a Bayesian methodology which uses multiple GW detections to constrain H_0. If an EM counterpart is detected and can be unambiguously linked to the host galaxy of the source then the redshift of that galaxy can be used to make an informative measurement of H_0. When no EM counterpart is observed, galaxy catalogues can be used to provide prior information on the redshift distribution of potential host galaxies.

In the latter case, the incompleteness of galaxy catalogues, leading to the possibility that the host galaxy is not contained within the catalogue, must be considered. This was summarised in Eq. 2.4, which breaks the analysis for a single GW event into four constituent parts: the GW data likelihood when the host galaxy is inside the catalogue, the GW data likelihood when the host is *not* in the catalogue, and the probabilities that the host galaxy is or is not in the catalogue. In order to evaluate each of these terms, assumptions need to be made about the redshift and absolute magnitude distributions of galaxies in the universe, and their relative probabilities of being host to a GW source. It also requires modelling the EM detection threshold of the galaxy catalogue, which determines which galaxies are contained within it. In this methodology an apparent magnitude threshold is used, as it matches the flux-limited nature of EM surveys. Section 2.3 derived each of the components of Eq. 2.4 in turn and highlighted where the data and prior choices entered the analysis in each case. The end result is an explicit derivation of the galaxy catalogue method for measuring H_0 which fully marginalises over the possibility that the host galaxy is not contained within the galaxy catalogue. The method is versatile in terms of priors to describe the underlying redshift distribution of galaxies, their luminosities, and how their probabilities of hosting a GW merger relate to these properties.

This Bayesian methodology is mathematically robust and makes best use of the information available, while effectively incorporating sources of uncertainty to ensure an unbiased estimation of H_0. However, deriving these expressions theoretically, and implementing them practically are two different challenges, which leads directly into the next chapter.

Chapter 3

A Mock Data Analysis

Declaration: most of this chapter is material from the MDA paper [64].

Section 3.2 is taken directly from the MDA paper, and edited only for style and phrasing (plus a correction to the equation converting luminosity to apparent magnitude). The creation of the MDAs was done by me.

Section 3.3 is my own unpublished work, save for section 3.3.2 which is based on part of the appendix of the MDA paper. The creation of the `gwcosmo` code was co-led by me.

Section 3.4 is taken directly from the MDA paper, and only edited for style and phrasing.

3.1 Introduction

This chapter introduces a series of MDAs which are designed specifically to test the robustness of the previous chapter's method to some of the most common pitfalls: GW selection effects, which affect all measurements of H_0 using standard sirens, and EM selection effects, which are relevant in the context of galaxy catalogues. This chapter also introduces the codebase `gwcosmo`, a Python implementation of the Bayesian methodology from Chapter 2, which is able to measure H_0 using standard sirens in combination with either EM counterparts or (incomplete) galaxy catalogues.

The MDA paper [64], on which much of this chapter is based, is the first paper to explicitly test the robustness of the standard siren methodology through the use of incomplete galaxy catalogues which do not contain all of the host galaxies. Additionally, the GW data used in these MDAs was produced using an end-to-end simulation, including searching for "injected" signals in real detector data followed by a full parameter estimation to obtain the GW posterior samples [76, 77], making this the most realistic set of simulated GW data to be used to explore GW cosmology to date. The analyses outlined below start with the most simplistic scenario, and increase in complexity with each iteration in order to ensure that `gwcosmo` is able to pass each level satisfactorily before moving onto the next.

The rest of this chapter is structured as follows. Section 3.2 discusses the design and creation

of the MDAs. Section 3.3 introduces `gwcosmo` and describes how the methodology from Chapter 2 is implemented practically in order to address the specific challenges that arise from the MDAs. Section 3.4 presents the results of analysing the MDAs with `gwcosmo`. Section 3.5 gives a detailed discussion of results and concludes this chapter.

3.2 Creating a set of mock data analyses

This section describes the design of a series of MDAs, created to test the Bayesian formalism described in Chapter 2 and its ability to infer the posterior on H_0 under different conditions. In each case, the MDA consists of (i) simulated GW data, and (ii) a corresponding mock galaxy catalogue. In all cases, several idealised assumptions regarding both the GW and galaxy data are made. On the GW side, the detection efficiency and the source population properties are assumed to be known exactly. On the galaxy side, the luminosity function and magnitude limit are also assumed to be known, so that the incompleteness correction can be calculated precisely. Further, the effects of large-scale structure and redshift uncertainties are ignored in the mock catalogues, and the galaxies are simulated as uniformly distributed delta functions in comoving volume.

For all of the MDAs, the GW data takes the form of a set of simulated BNS events from The First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo dataset [76,77][1]. The set of BNS events comes from an end-to-end simulation of approximately 50,000 "injected" events in detector noise corresponding to a sensitivity similar to what was achieved during O2 of Advanced LIGO. Only a subset (approximately 500 events) were "detected" by a network of two or three detectors with the `GstLAL` matched filter based detection pipeline [78]. From the above detections, 249 events were randomly selected, and these events underwent full Bayesian parameter estimation using the `LALInference` software library [48] to obtain gravitational wave posterior samples and skymaps. Consistency with the First Two Years parameter estimation results in terms of sky localization areas and 3D volumes was demonstrated in [79]. It is these 249 events from the First Two Years dataset and the associated GW data which form the basis of the MDAs.

The galaxy catalogues for each iteration of the MDA described below are designed to test each part of the methodology in a cumulative fashion, starting with GW selection effects, adding in EM selection effects, and finally testing the ability to utilise the information available in the observed brightness of host galaxies, by weighting the galaxies with a function of their intrinsic luminosities.

The starting point for the galaxy catalogues was to take all 50,000 injected events from the

[1]The set of simulations in [77] are more realistic with the same injections in (recoloured) detector data as opposed to Gaussian noise used in [76]. Correspondingly, the detection criterion is in terms of a false alarm rate (FAR) rather than a threshold on the SNR. This is an important distinction, particularly affecting events marginally close to the detection threshold. Here, the simplified set of simulations in [76] are used.

First Two Years dataset and simulate a mock universe, which consists of a galaxy corresponding to each injected event's sky location and luminosity distance, where the latter is converted to a redshift using a fiducial H_0 value of 70 km s^{-1} Mpc^{-1}. The First Two Years data was originally simulated in a universe where GW events followed a d_L^2 distribution, and there was no distinction between the source frame and the (redshifted) detector frame masses. Though not ideal, this data reasonably mimics a low redshift universe ($z \ll 1$) in which the linear Hubble relation of Eq. 1.23 holds, and galaxies follow a z^2 distribution, *i.e.* $p(z|I) \propto z^2$. For consistency, the MDAs universe is created using both low redshift approximations, meaning that Eq. 1.23 is used to convert the luminosity distances of the 50,000 injections to redshifts of mock galaxies. These galaxies make up the MDA universe. It should be emphasized that the Bayesian method for estimation of H_0 outlined in Chapter 2 above is general, and can be extended to the ΛCDM cosmological model described in Section 1.1 with $\{\Omega_m, \Omega_k, \Omega_\Lambda\}$ held fixed. In particular, the method is applicable for events which are detected at higher distances, where the low redshift approximation breaks down. The restriction to a linear cosmology in this chapter comes only due to the use of the MDA dataset. It should be noted that when using a linear cosmology it is not possible to test for effects introduced by the presence of other cosmological parameters. The analysis at large redshifts may, for example, be sensitive to the values (or the assumed prior ranges) of the parameters like Ω_m and Ω_Λ.

The first four columns of Table 3.1 summarize the characteristics of each of the galaxy catalogues created and how they correspond to each MDA. A brief description is given for each case below.

3.2.1 MDA0: Known associated host galaxies

MDA0 is the simplest version of the MDAs, in which the host galaxy for each GW event is identified with certainty – equivalent to the direct counterpart case. As the galaxies are generated with no redshift uncertainties or peculiar velocities the results will be (very) optimistic, with only the uncertainty in the d_L posterior for each BNS contributing to the final uncertainty of the result. This MDA provides the "best possible" constraint on H_0 using the 249 events, which allows for a useful comparison with the other MDAs. MDA0 also tests the treatment of GW selection effects, to ensure that when given a set of GW detections and access to their known hosts, an unbiased posterior on H_0 can be returned.

3.2.2 MDA1: Complete galaxy catalogue

The initial 50,000 injections used for the First Two Years dataset were generated as uniform in distance cubed, out to a distance of ~ 435 Mpc. Once converted to redshifts, the mock universe consists of 50,000 galaxies out to $z \sim 0.1$ for $H_0 = 70$ km s^{-1} Mpc^{-1}. This gives a galaxy number density of ~ 1 per 7000 Mpc ^{3}, which is ~ 35 times more sparse when compared to the actual

density of galaxies in the local universe [80]. In order to prevent the hard edge from introducing information to the analysis, additional galaxies are generated beyond the edge of the dataset universe, uniformly across the sky and uniformly in comoving volume, thereby extending the universe out to a radius of 2000 Mpc ($z = 0.467$ for $H_0 = 70$ km s^{-1} Mpc^{-1}). This means that, even allowing H_0 to be as large as 200 km s^{-1} Mpc^{-1}, the edge of the MDA universe is more than twice the highest redshift associated with the farthest detection (which is at ~ 270 Mpc)[2]. This catalogue is *complete* in the sense that it contains every galaxy in the simulated universe, and so each of the 249 detected BNSs has a unique host galaxy contained within it. However, this galaxy is not identified, so while it is known to be present, all galaxies within an event's sky localisation must be treated as potential host galaxies. MDA1 is designed as a proof-of-principle of Schutz's original method, by ensuring that given a set of sources and access to a complete galaxy catalogue, the analysis produces a result consistent with the "real" value of H_0.

3.2.3 MDA2: Incomplete galaxy catalogue

MDA2 is designed to test the treatment of EM selection effects, by applying an apparent magnitude threshold to the MDA universe, such that a certain fraction of the host galaxies are not contained in it. This is a necessary consideration, given the current regime where GW signals are detectable beyond the distance to which current galaxy catalogues can be considered to be complete. This has been true for BBH detections since O1, and is true of BNSs as well in O3.

In order to create the galaxy catalogue for MDA2, the MDA universe is taken as a starting point, and luminosities are assigned to each of the galaxies. The luminosity distribution of the galaxy catalogue is taken to follow a Schechter function of the form [81]

$$\phi(L)\,dL = n^* \left(\frac{L}{L^*}\right)^{\alpha} e^{-L/L^*}\,\frac{dL}{L^*}\,, \tag{3.1}$$

where L denotes a given galaxy luminosity and $\phi(L)\,dL$ is the number of galaxies within the luminosity interval $[L, L + dL]$. The characteristic galaxy luminosity is given by $L^* = 1.2 \times 10^{10}\,h^{-2}\,L_\odot$ with solar luminosity $L_\odot = 3.828 \times 10^{26}$ W, and $h \equiv H_0/(100\text{km s}^{-1}\,\text{Mpc}^{-1})$[3], $\alpha = -1.07$ characterizes the power-law drop off of the luminosity function, and n^* denotes the number density of objects in the MDA universe (in practice, this only acts as a normalization constant). The integral of the Schechter function diverges at $L \to 0$, requiring a lower luminosity cut-off for the dimmest galaxies in the universe which is set to $L_{\text{lower}} = 0.001L^*$. This choice is arbitrary here, but small enough to include almost all objects classified as galaxies in real catalogues like the GLADE catalogue [80].

These luminosities are then converted to apparent magnitudes using $m \equiv 25 - 2.5\log_{10}(L/L_0) +$

[2]For MDA1 and for all subsequent MDAs, it has been tested that this artificial "edge of the universe" is high enough to have no bearing on the results.

[3]Note that the parameter L^* of the Schechter luminosity function itself depends on H_0. In the creation of the MDAs it is fixed to 70 km s^{-1} Mpc^{-1}, while in the analysis it is allowed to vary and hence is taken into account.

$5\log_{10}(d_L/\text{Mpc})$, where $L_0 = 3.0128 \times 10^{28}$ W is the zero point luminosity. An apparent magnitude threshold (m_{th}) is applied as a crude characterisation of the selection function of an optical telescope observing only objects with $m < m_{\text{th}}$. MDA2 is broken into three sub-MDAs, in order to test the method's ability to handle different levels of galaxy catalogue completeness dictated by different telescope sensitivity thresholds. In each case, the catalogue completeness is defined as the ratio of the number of galaxies inside the catalogue relative to the number of galaxies inside the MDA universe, out to a reference distance d_L,

$$f_{\text{completeness}}(d_L) = \frac{\sum_{j}^{\text{MDA2}} \Theta(d_L - d_{L_j})}{\sum_{k}^{\text{universe}} \Theta(d_L - d_{L_k})}. \tag{3.2}$$

The numerator is a sum over the galaxies contained within the MDA2 catalogue out to some reference distance d_L, and the denominator is a sum over the galaxies in the universe (*i.e.* the MDA1 catalogue), out to the same distance. Here, Θ is a Heaviside step function which evaluates to 1 when $d_{L_j} < d_L$, and 0 otherwise. This definition of completeness, using number density of galaxies, is not one commonly used in astronomy, where completeness is often defined based on the fraction of luminosity, relative to some expected total luminosity, contained within shells of the catalogue. However, the definition in Eq. 3.2 makes more sense in this case, as all galaxies within the catalogue are treated as being equally likely to host GW events, regardless of their luminosity.

Apparent magnitude thresholds of $m_{\text{th}} = 19.5$, 18, and 16 are chosen for the three sub-MDAs,[4] which correspond to integrated number density completeness fractions of 0.75, 0.5 and 0.25 respectively, evaluated at a reference distance of $d_L = 115\,\text{Mpc}$. The reference distance is chosen such that, given the luminosity distance distribution of detected BNSs, the completeness fraction for the sub-MDA to this distance is roughly indicative of the percentage of host galaxies of the 249 detections which remain inside the galaxy catalogue. The left panel of Fig. 3.1 shows how the integrated number density completeness of each of the MDA2 catalogues drop off as a function of distance.

3.2.4 MDA3: Luminosity weighting

MDA3 is designed to test the effect of weighting the likelihood of any galaxy being host to a GW event as a function of its luminosity. It is probable that the higher luminosity galaxies are more likely hosts for compact binary mergers as, in general, more luminous galaxies will contain more stars, so the chances of a binary developing and leading to a CBC are higher.

[4]These values of m_{th} cover a similar range to the estimated variation in m_{th} that is displayed by the GLADE catalogue in B-band luminosity, where the average is ~ 18 (see Section 5.3.2).

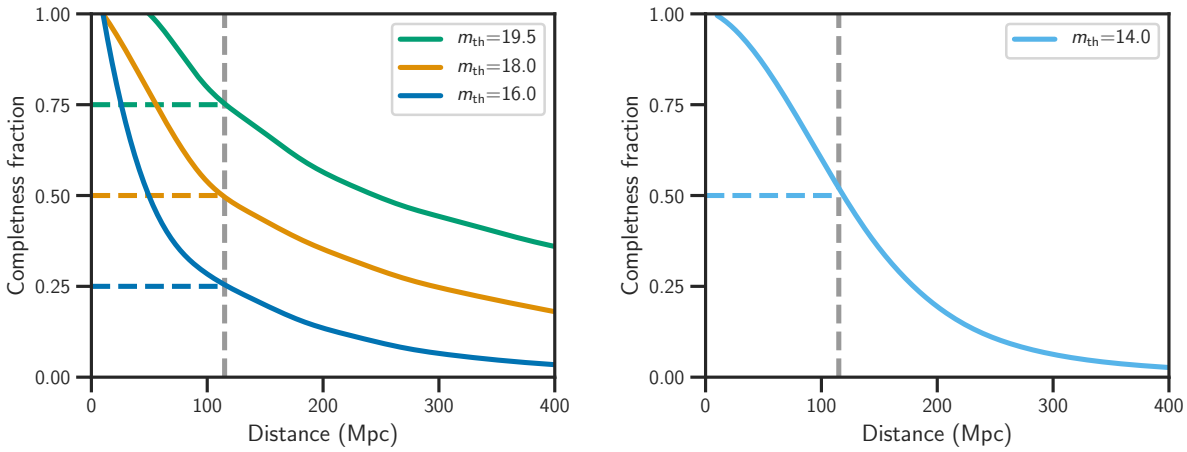

Figure 3.1: Galaxy catalogue completeness fractions for MDA2 and MDA3. *Left panel:* The integrated number density completeness fraction defined in Eq. 3.2 as a function of luminosity distance for the three MDA2 sub-catalogues. The lines in green, orange and blue correspond to the catalogues with $m_{\rm th} = 19.5$, 18, and 16 respectively; these correspond to completeness percentages of 75%, 50% and 25% out to a fiducial reference distance of 115 Mpc (shown as a vertical grey line). *Right panel:* The integrated luminosity completeness fraction defined in Eq. 3.5 as a function of luminosity distance for the MDA3 catalogue, with $m_{\rm th} = 14$. At the reference distance of 115 Mpc (vertical grey line) this corresponds to a completeness fraction of $\sim 50\%$.

More specifically, the luminosity in blue (B-band) is indicative of a galaxy's star formation rate, for example, while the luminosity in high infrared (K-band) is a tracer of the stellar mass [82–84]. The bulk of the host probability is expected to be contained within a smaller number of brighter galaxies, effectively reducing the number of galaxies which contribute meaningfully to the analysis. Additional information from luminosity is thus expected to improve the constraint on H_0.

For MDA3, the probability of a galaxy hosting a GW event is chosen to be proportional to the galaxy's luminosity.[5] Because the GW events for these MDAs were generated in advance, and the universe in which they exist is being simulated retroactively, generating the MDA3 universe required some care: luminosities have to be assigned to the host galaxies *and* the non-host galaxies in such a way that the luminosity weighting is correctly represented within the galaxy catalogue.

As with MDA2, the luminosity distribution of the galaxies in the universe is assumed to follow the Schechter luminosity function as in Eq. (3.1) (referred to from now on as $p(L)$). However, the joint probability of a single galaxy having luminosity L and hosting a GW event (which emits a signal, s) is $p(L, s) \propto L\, p(L)$, where it is assumed that the probability of a galaxy with luminosity L hosting a source is directly proportional to L. All host galaxies thus have

[5]Different luminosity bands are neglected here, as is the redshifting of the source frame luminosity that would apply to the light emitted from real galaxies.

luminosities sampled from $L\,p(L)$. In this context, all galaxies which hosted GW events must be considered, not just those from which a signal was detected. With this in mind, the overall luminosity distribution has the following form:

$$p(L) = \beta \frac{L}{\langle L \rangle} p(L) + (1 - \beta)\, x(L) \tag{3.3}$$

where β is the fraction of host galaxies to total galaxies for the observed time period ($1 \geq \beta \geq 0$), $\langle L \rangle$ is the expectation value of L, given $p(L)$, and acts as a normalisation constant for $Lp(L)$, and $x(L)$ is the unknown luminosity distribution of galaxies which did not host GW events, which can be sampled from, for a given value of β.

Rearranging to obtain the only unknown, $x(L)$, gives

$$x(L) = \frac{p(L)}{1 - \beta}\left[1 - \beta \frac{L}{\langle L \rangle}\right], \tag{3.4}$$

and from this it can be seen that there is an additional constraint on β, because the term inside the square brackets must be > 0. The maximum value that β can take which satisfies this criteria is given by $\beta_{\max} = \langle L \rangle / L_{\max}$, where $L_{\max}$ is the maximum luminosity from the Schechter function, and $\langle L \rangle$ is equivalent to the mean. From the Schechter function parameters detailed in section 3.2.3, $\beta_{\max} \approx 0.015$.

The original First Two Years data was generated by simulating 50,000 BNS events, of which 500 were detected meaning that of 50,000 original host galaxies, only 1% produced GW that were detectable. Of these 500, 249 randomly selected detections underwent parameter estimation. The number of "hosting" and "non-hosting" galaxies have to be rescaled to represent the fact that half of the detections were effectively ignored. Therefore half of the galaxies corresponding to the original injections are denoted as hosts (including those associated with the 249 detected GW events). On its own, this gives $\beta = 0.5$. In order to satisfy the requirements for β, a greater density of non-hosting galaxies has to be added to the universe before luminosities can be assigned. Thus for MDA3, the density of galaxies is increased by a factor of 100 (all uniformly distributed in distance cubed, like the originals), with the acknowledgement that this would lead to a broadening of the final posterior. MDA3 has a galaxy density of $\sim$1 galaxy per 70 Mpc^3, which is about 3 times denser than the actual density of galaxies in the local universe [80]. This also means that MDA3 is not directly comparable with the previous MDA versions, save MDA0. The 25,000 galaxies which are hosts (including the 249 associated with the detections) are assigned luminosities from $Lp(L)$, while the remainder are assigned from $x(L)$ above.

In order to include EM selection effects, an apparent magnitude cut m_{th} of 14 is applied, such that the completeness of the galaxy catalogue is $\sim$ 50% out to the same fiducial distance of 115 Mpc as in MDA2. In this case, however, completeness is defined in terms of the fractional

luminosity contained in the catalogue, rather than in terms of numbers of objects:

$$f_{\text{completeness}}(d_L) = \frac{\sum_{j}^{\text{MDA3}} L_j \Theta(d_L - d_{L_j})}{\sum_{k}^{\text{universe}} L_k \Theta(d_L - d_{L_k})}, \tag{3.5}$$

where the numerator is summed over the galaxies inside the MDA3 apparent magnitude-limited catalogue, and the denominator is summed over the galaxies in the whole MDA3 universe. This is shown as a function of luminosity distance in the right panel of Fig. 3.1. As the host galaxies are luminosity weighted, the integrated luminosity completeness out to $d_L = 115$ Mpc is representative of the percentage of detected BNS events with hosts inside the catalogue (by design).

3.3 `gwcosmo`: a Python package for standard siren cosmology

This section considers the challenges posed by the MDAs, and discusses what simplifications can be made to the methodology outlined in Chapter 2. It then presents an initial implementation in the form of `gwcosmo`, a Python package for standard siren cosmology which can handle both EM counterparts and galaxy catalogues.

The first thing to note is that there is a disconnect between the methodology as laid out in the previous chapter, and a practical implementation which would allow that methodology to be tested using the MDAs. The main reason for this lies in the EM incompleteness correction, the presence of which requires that the GW data be useable in some form that allows it both to be evaluated at delta-function-like points in right ascension (RA), declination (dec) and z, but also to be integrable over Ω (that is, RA and dec), z and M. The first condition means that the GW posterior samples cannot be used in their raw form, as the delta-function nature of both galaxies and samples requires that at least one be smoothed to a function in order to be evaluated at the location of the others. Given that galaxies are discrete, and the GW data is not, it makes sense for this smoothing to be over the GW data. This leads to two options: to make use of the GW skymaps (which contain a distance estimate along every line-of-sight), or to use the posterior samples to make some function of Ω and z, such as by applying a kernel density estimate (KDE) [85]. Before considering which path to go down, it is worth noting that both choices come with the same drawback: the 4-dimensional integrals in Eqs. 2.21 and 2.27 are computationally costly, to the point of being impractical to run for a single event, let alone for 249 of them.

Before moving forwards, it is worth taking a closer look at what the MDAs are asking for, and to consider whether there are any approximations that could be made within the method in

order to make it more tractable. One of significance is that the apparent magnitude thresholds for each MDA are uniform across the sky. This means that, in the methodology, the integration limits in 2.21 and 2.27 lose their dependence on Ω. With this change, the only thing preventing the integral over Ω from be separated from the integrals over z and M is the conditional dependence of z and Ω in the terms $p(x_{\rm GW}|z, \Omega, H_0, I)$ and $p(D_{\rm GW}|z, \Omega, H_0, I)$. Both terms are considered in closer detail below.

3.3.1 A closer look at gravitational wave data

When considering the data products associated with a GW detection, it is worth noting that $x_{\rm GW}$ is rarely used directly - instead, the event undergoes parameter estimation, which produces a set of posterior samples for the event. These posterior samples cover a range of parameters, including luminosity distance, inclination, sky location and detector-frame masses. Many of these parameters have little bearing on a cosmological analysis, but of immediate and obvious use are the luminosity distance and sky location. It is also worth noting that these are *posterior* samples – a prior has already been applied in their generation. For luminosity distance, the default prior is $\propto d_L^2$, which then must be removed before any other priors on luminosity distance (or redshift) can be applied. The dependence of the GW data on z and H_0 comes through the relationship between these parameters and d_L,[6] and so it is convenient to reframe $p(x_{\rm GW}|z, \Omega, H_0, I)$ as $p(x_{\rm GW}|d_L(z, H_0), \Omega, I)$. Taking this as a starting point and expanding gives

$$p(x_{\rm GW}|d_L(z, H_0), \Omega, I) = \frac{p(d_L, \Omega|x_{\rm GW}, I)p(x_{\rm GW}|I)}{\pi(d_L|I)\pi(\Omega|I)}. \tag{3.6}$$

Here, $p(d_L, \Omega|x_{\rm GW}, I)$ corresponds to the posterior samples on d_L and Ω for the GW event $x_{\rm GW}$. The term $p(x_{\rm GW}|I)$ has no H_0 dependence and so can be ignored as a normalisation constant, while $\pi(d_L|I)$ and $\pi(\Omega|I)$ are the priors on d_L and Ω which were used when $x_{\rm GW}$ underwent parameter estimation, *i.e.* $\propto d_L^2$ and uniform across the sky.

In order to separate Ω and d_L entirely, the expression can be further expanded:

$$\begin{aligned} p(x_{\rm GW}|d_L(z, H_0), \Omega, I) &= \frac{p(d_L|\Omega, x_{\rm GW}, I)p(\Omega|x_{\rm GW}, I)p(x_{\rm GW}|I)}{\pi(d_L|I)\pi(\Omega|I)} \\ &\approx \frac{p(d_L|x_{\rm GW}, I)p(\Omega|x_{\rm GW}, I)p(x_{\rm GW}|I)}{\pi(d_L|I)\pi(\Omega|I)}. \end{aligned} \tag{3.7}$$

Here, the assumption has been made that the GW posteriors on d_L and Ω are not correlated, *i.e.* that the line of sight d_L distribution associated with $x_{\rm GW}$ is the same for all Ω, down to a constant. This is an approximation: in reality, luminosity distance and sky location are correlated, and the distance distribution varies with Ω. However, the MDAs allows the validity of

[6]For a real GW signal, additional redshift dependence comes from the redshifting of the source-frame masses to the detector frame, but as the sources in the First Two Years dataset were not redshifted, this can be neglected for the time being. It will be revisited in detail in Chapter 4.

this approximation to be tested, insofar as if the approximation introduces no noticeable bias to MDA0 or MDA1 it can be deemed "good enough" for the degree of precision that is reachable with 249 highly informative BNSs.

The GW posterior samples on d_L have been drawn from $p(d_L|x_{\rm GW}, I)$. In order to make $p(d_L|x_{\rm GW}, I)$ a smooth function, which can be both integrated over z and M, and evaluated at delta-function-like galaxies, a KDE can be used to create a 1-dimensional function over d_L, such that $f_{\rm KDE}(d_L) \approx p(d_L|x_{\rm GW}, I)$.

Similarly, $p(\Omega|x_{\rm GW}, I)$ could be represented by a KDE on the RA and dec posterior samples. However, due to the spherical nature of these coordinates, it is easier to make use of the provided skymaps, and avoid potential issues that could arise with events that cross the prior boundary for RA and wrap around, *e.g.* $2\pi \to 0$. The skymaps do not have this issue, and instead contain the (normalised) 2D GW sky probability in a high-resolution map of a sphere gridded into $N_{\rm pix}$ equally-sized pixels, which is equivalent to $p(\Omega|x_{\rm GW}, I)$. When integrated over the full sky,

$$\int p(\Omega|x_{\rm GW}, I)\, d\Omega = \sum_{k=1}^{N_{\rm pix}} P(k|x_{\rm GW}, I) = 1, \tag{3.8}$$

where $P(k|x_{\rm GW}, I)$ is the probability contained within pixel k. The probability density at a specific location on the sky, Ω_i, is given by

$$p(\Omega_i|x_{\rm GW}, I) = P(k_{\Omega_i}|x_{\rm GW}, I)\frac{N_{\rm pix}}{4\pi}, \tag{3.9}$$

where $P(k_{\Omega_i}|x_{\rm GW}, I)$ is the probability contained within the pixel that Ω_i lies in.

A final expression for the desired term, $p(x_{\rm GW}|z, \Omega, H_0, I)$ comes from acknowledging that the normalisation of the GW likelihood is not known, but as long as this normalisation is not a function of z, Ω or H_0, this is okay, as it will be the same constant for the in-catalogue likelihood (Eq. 2.17) and the out-of-catalogue likelihood (Eq. 2.27). The term $p(x_{\rm GW}|z, \Omega, H_0, I)$ can therefore be represented by

$$p(x_{\rm GW}|z, \Omega, H_0, I) \approx C\frac{f_{\rm KDE}(d_L)}{\pi(d_L)}\frac{p(\Omega|x_{\rm GW}, I)}{\pi(\Omega|I)}, \tag{3.10}$$

where C is some function of the GW data which is constant with respect to H_0.

3.3.2 A closer look at gravitational wave selection effects

For the MDAs, the GW detection threshold is taken to be an SNR threshold, $\rho_{\rm th}$, in order to match the detection conditions of the First Two Years dataset. Probability of detection can be calculated using the first line of Eq. 2.3, where $p(D_{\rm GW}|x_{\rm GW}, H_0, I)$ is 1 if the SNR of $x_{\rm GW}$ passes $\rho_{\rm th}$, and 0 otherwise. However, looking at the individually derived components of the galaxy catalogue case (Eqs. 2.17, 2.21 and 2.27), $p(D_{\rm GW}|H_0, I)$ only appears in an expanded form, where it is

additionally conditioned on z and Ω. Calculating $p(D_{\rm GW}|z, \Omega, H_0, I)$ requires integrating over all realisations of GW events (detected and not), for ranges of z, Ω and H_0 values which span the full prior space, and applying a detection threshold ($\rho_{\rm th}$) which each event must pass in order to be deemed detected.

Probability of detection, $p(D_{\rm GW}|H_0, I)$, should also be common to all events within a given population, and therefore should be marginalised over time of detection. If probability of detection is averaged over the course of the entire simulated observation period for the MDAs, the dependence of $D_{\rm GW}$ on Ω is smeared out over the course of many days. This can be approximated to mean that $p(D_{\rm GW}|z, H_0, I)$ is uniform over the sky (ignoring the mild declination dependence which would remain after the rotation of the Earth is taken into account), and so marginalising over Ω at this stage is taken to be equivalent to also marginalising over time of detection.

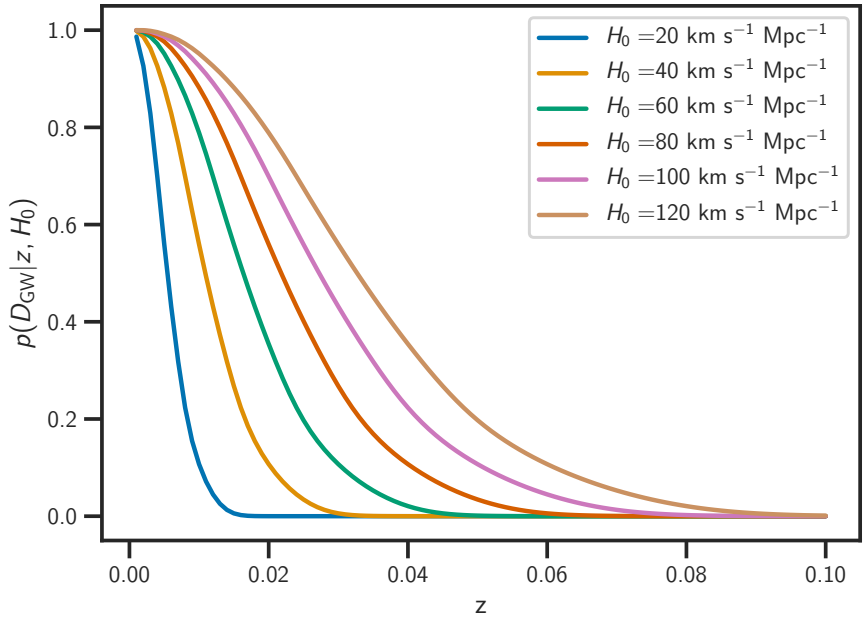

Figure 3.2: Probability of detection, $p(D_{\rm GW}|z, H_0)$, as a function of z for different values of H_0. A 2-detector network at O2-like sensitivity is assumed, for a population of BNSs.

Practically, Monte-Carlo integration can be used to compute Eq. 2.3 as a function of z:

$$p(D_{\rm GW}|z, H_0, I) = \frac{1}{N_{\rm draws}} \sum_{i=1}^{N_{\rm draws}} p(D_{{\rm GW}i}|x_{{\rm GW}i}, z, H_0, I), \tag{3.11}$$

where $x_{{\rm GW}i}$ corresponds to a potential GW event, the parameters of which have been randomly drawn from the prior distributions of parameters which affect an event's detectability (mass,[7]

[7]For the MDAs this is detector-frame mass, as the original BNS population in the First Two Years dataset was generated from a mass distribution which did not undergo redshifting for detection.

inclination, polarization, and sky location) and the event's ρ_i is calculated for specific values of z and H_0. Each potential event can be then deemed detected or not, using

$$p(D_{\mathrm{GW}i}|x_{\mathrm{GW}i}, z, H_0, I) = \begin{cases} 1, & \text{if } \rho > \rho_{\mathrm{th}} \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

This gives a smooth function for $p(D_{\mathrm{GW}}|z, H_0, I)$, which drops from 1 to 0 over a range of z and H_0 values. To match the First Two Years dataset, the SNR threshold is taken as $\rho_{\mathrm{th}} = 8\sqrt{2} \approx 11.3$ (coming from an individual detector threshold of 8, and assuming that every event was detected by two detectors). The 2016 power-spectral-density (PSD) from Fig. 1 of [76] is used, and $p(D_{\mathrm{GW}i}|x_{\mathrm{GW}i}, z, H_0, I)$ is evaluated for 5000 samples (over a range of z and H_0 values), such that the integral converges. Figure 3.2 shows how the probability of detection behaves as a function of z for different values of H_0.

Now, considering the methodology of the galaxy catalogue case again: because $p(D_{\mathrm{GW}}|z, H_0, I)$ has been pre-marginalised over Ω, the RA and dec of individual galaxies will not impact the calculated probability of detecting a signal that was emitted from that galaxy – only its redshift matters. Any Ω-dependence is no longer intrinsically tied to the probability of detection term, which allows a simplification of the galaxy catalogue case to be considered.

3.3.3 A practical form for the methodology

Using the assumption that m_{th} is uniform across the sky, and that the dependence of both x_{GW} and D_{GW} on Ω can be separated from other parameters, the expressions for each part of the galaxy catalogue case (Eq. 2.4) can be expressed as follows.

To practically compute the likelihood when the host is in the galaxy catalogue, Eq. 2.16 is taken as the starting point (because the MDAs neglect redshift uncertainties) and the expression for the likelihood of the GW data specified in Eq. 3.10 is substituted in to give

$$\frac{1}{C} p(x_{\mathrm{GW}}|G, D_{\mathrm{GW}}, s, H_0, I) = \frac{\sum_{i=1}^{N} \frac{f_{\mathrm{KDE}}(d_L(z_i, H_0))}{\pi(d_L(z_i, H_0))} \frac{p(\Omega_i|x_{\mathrm{GW}}, s, I)}{\pi(\Omega_i|I)} p(s|z_i, I) p(s|M(z_i, m_i, H_0), I)}{\sum_{i=1}^{N} p(D_{\mathrm{GW}}|z_i, s, H_0, I) p(s|z_i, I) p(s|M(z_i, m_i, H_0), I)}. \tag{3.13}$$

The probability density at the sky location of galaxy i, $p(\Omega_i|x_{\mathrm{GW}}, s, I)$, is found as described in Eq. 3.9. The term $\pi(\Omega_i|I)$ is a constant with the value of $1/4\pi$. The constant C on the left-hand side is the same constant as defined in Eq. 3.10, the value of which is irrelevant because it is the same both here and in Eq. 3.15 below.

The probability that the host galaxy is in the catalogue (previously Eq. 2.21) becomes

$$p(G|D_{\rm GW}, s, H_0, I) = \frac{\int dM \int_0^{z(M,m_{\rm th},H_0)} dz\; p(D_{\rm GW}|z,s,H_0,I)p(s|z,I)p(z|I)p(s|M,I)p(M|H_0,I)}{\iint p(D_{\rm GW}|z,s,H_0,I)p(s|z,I)p(z|I)p(s|M,I)p(M|H_0,I)\; dzdM}, \tag{3.14}$$

where the integrals over Ω, $\int p(\Omega|I)d\Omega$, have been separated from the integrals over z and M, and then cancelled in numerator and denominator. The probability that the host is not inside the galaxy catalogue can be calculated as before, using Eq. 2.22.

The likelihood when the host galaxy is not in the catalogue (previously Eq. 2.27) becomes

$$\begin{aligned}\frac{1}{C}p(x_{\rm GW}|\bar{G}, D_{\rm GW}, s, H_0, I) = &\frac{\int p(\Omega|x_{\rm GW},s,I)d\Omega}{\int p(\Omega|I)d\Omega} \\ &\times \frac{\int dM \int_{z(M,m_{\rm th},H_0)}^{\infty} dz\; \frac{f_{\rm KDE}(d_L(z,H_0))}{\pi(d_L(z,H_0))} p(s|z,I)p(z|I)p(s|M,I)p(M|H_0,I)}{\int dM \int_{z(M,m_{\rm th},H_0)}^{\infty} dz\; p(D_{\rm GW}|z,s,H_0,I)p(s|z,I)p(z|I)p(s|M,I)p(M|H_0,I)}.\end{aligned} \tag{3.15}$$

The $p(\Omega|I)$ term in the numerator of Eq. 2.27 has cancelled with $\pi(\Omega|I)$ in Eq. 3.10, as they are the same uniform on the sky prior, leaving just the integral over the 2D GW skymap. It is worth noting that even though the Ω integral has no H_0 dependence it cannot be discarded for that reason alone, as the normalisation between the in-catalogue and out-of-catalogue likelihoods must remain consistent. Here, however, the integral is over the full sky and so both remaining integrals over Ω evaluate to 1.

The simplifications outlined above have reduced the 4-dimensional integrals to 2-dimensional ones, allowing a straightforward implementation using `SciPy`'s `dblquad` [86]. This allows for a full Python implementation of the galaxy catalogue method in `gwcosmo`. The next step is to test this implementation with the MDAs.

3.4 Results

This section presents the results after analysing the MDAs described in section 3.2 using `gwcosmo`. The combined posteriors on H_0 are shown for each MDA in the following sections. The results are summarised in Table 3.1, which lists the measured values of H_0 for the combined 249-event

[8]The completeness is calculated as a number completeness using Eq. (3.2) for MDAs 1 and 2, and as a luminosity completeness using Eq. (3.5) for MDA 3, out to a fiducial distance of 115 Mpc, such that it is indicative of the fraction of host galaxies which are inside the galaxy catalogue in both cases.

MDA	Host galaxy preference	Completeness[8]	$m_{\rm th}$	Analysis assumption	H_0 (km s^{-1} Mpc^{-1})	% uncertainty
0	Known host	-	-	direct counterpart	$69.08^{+0.79}_{-0.80}$	1.13%
1	equal weights	100%	-	unweighted catalogue	$68.91^{+1.36}_{-1.22}$	1.84%
2a	equal weights	75%	19.5	unweighted catalogue	$69.97^{+1.59}_{-1.50}$	2.21%
2b	equal weights	50%	18	unweighted catalogue	$70.14^{+1.80}_{-1.67}$	2.48%
2c	equal weights	25%	16	unweighted catalogue	$70.14^{+2.29}_{-2.18}$	3.20%
3a	luminosity weighted	50%	14	weighted catalogue	$70.83^{+3.55}_{-2.72}$	4.48%
3b	luminosity weighted	50%	14	unweighted catalogue	$69.50^{+4.20}_{-3.24}$	5.31%

Table 3.1: A summary of the main MDA results. The peak value and the 68.3% highest density error region for the posterior probability on H_0 are quoted for each of the MDAs after combining all 249 events. The fractional uncertainty is defined as the half-width of the 68.3% highest density probability interval divided by the simulated value of $H_0 = 70$ km s^{-1} Mpc^{-1}.

posterior (maximum a-posteriori and 68.3% highest density posterior intervals) assuming a uniform prior on H_0 in the range of $[20, 200]$ km s^{-1} Mpc^{-1}, as well as the corresponding fractional uncertainties for each of the MDAs.

3.4.1 MDA0: Known associated host galaxies

MDA0 considers the simple case where the host galaxy for every event is correctly identified. The results of this analysis are presented in Fig. 3.3. The likelihoods for each individual GW event are shown (normalised relative to each other but scaled with respect to the combined posterior for clarity) shaded by the event's optimal SNR in the detector network, as defined in [87]. In this MDA each likelihood is informative, having a clearly-defined peak corresponding to finding the likely values of H_0 for the known galaxy redshift. Each curve traces the information in the corresponding d_L distribution, which is usually unimodal, but in some cases may have two or more peaks [76, 77]. The peaks of the individual likelihoods do not necessarily correspond to the true value $H_0 = 70$ km s^{-1} Mpc^{-1}, but there is always support for it, leading to the combined posterior, which is overlaid in thick purple. This gives a statistical estimate for the maximum a-posteriori value and 68.3% maximum-density credible interval for H_0 of $69.08^{+0.79}_{-0.80}$ km s^{-1} Mpc^{-1}. The final result combining all the 249 events have converged to a relatively symmetric "Gaussian" distribution [88].

The result of MDA0 corresponds to the best possible H_0 estimate given this set of GW detections, since this case assumes perfect knowledge of the host galaxies. This provides a benchmark against which other versions of the MDA can be compared. Since this is a best-case scenario, it has the least statistical uncertainty in the final result, making any systematic bias more apparent than for the subsequent MDAs. For the combined result with 249 events, the simulated value is contained well within the 2σ bounds of the posterior distribution on H_0.

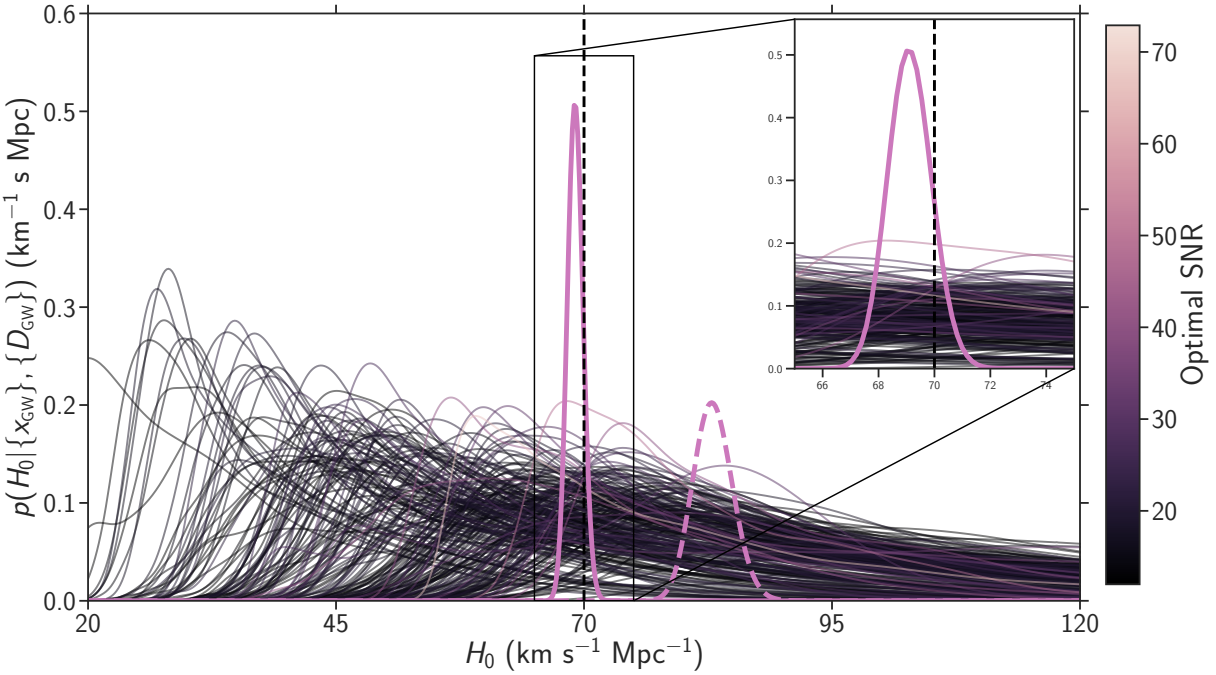

Figure 3.3: Individual and combined results for MDA0 (known host galaxy or direct counterpart case). The solid thick purple line shows the combined posterior probability density on H_0, while the dashed line shows the combined posterior when GW selection effects are neglected. Individual likelihoods (normalised and then scaled by an arbitrary value), for each of the 249 events, are shown as thin lines with shades corresponding to their optimal SNR. The simulated value of H_0 is shown as a vertical dashed line.

MDA0 demonstrates the importance of correctly accounting for GW selection effects. Finite detector sensitivity means that there is a bias towards detecting sources which are nearer-by, and which are optimally orientated (closer to face-on). If an analysis is performed without taking into consideration the denominator $p(D_{\rm GW}|H_0, I)$ of Eq. (2.2), which corrects for this, the posterior density on H_0 converges to a value different from its simulated value of 70 km s^{-1} Mpc^{-1}. This can be seen in Fig. 3.3, where the dashed purple line shows the MDA0 combined posterior for all 249 events, neglecting GW selection effects entirely. The fact that the results for MDA0 are consistent with the simulated value of H_0 indicates, firstly, that the GW selection effects are adequately modelled for this degree of accuracy, and secondly, that the decision to separate GW sky and distance information, as described in Section 3.3.1, hasn't introduced any noticeable bias.

3.4.2 MDA1: Complete galaxy catalogue

The next, more complex case is MDA1, where no EM counterpart is observed and a galaxy catalogue provides the redshifts of all potential host galaxies. This optimistic scenario assumes a *complete* galaxy catalogue containing all potential hosts, so that EM selection effects do not need to be considered. In this case, the probability of the host galaxy being in the catalogue is 1

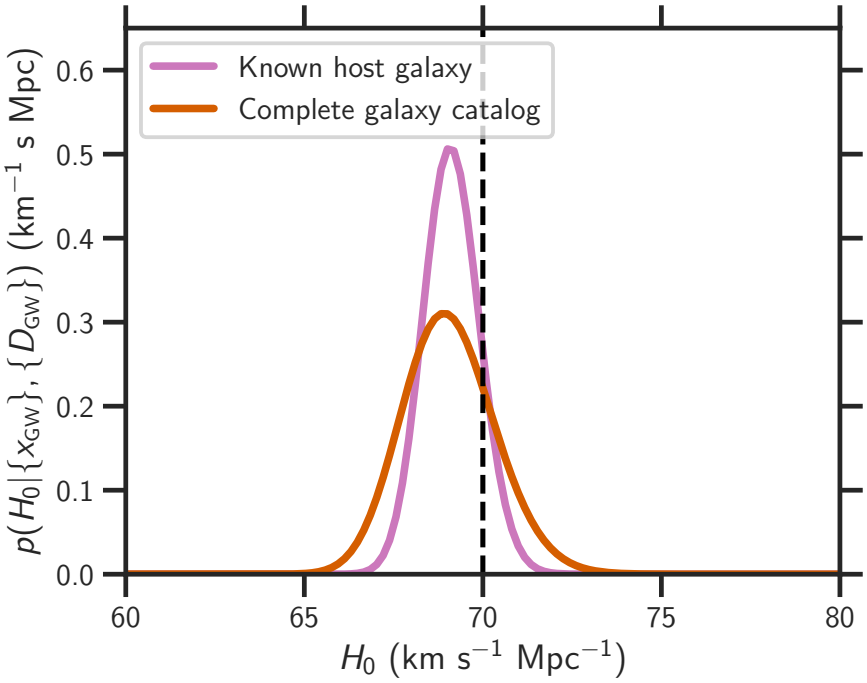

Figure 3.4: Comparison of the galaxy catalogue method with the known host galaxy case. The joint posterior probability density on H_0 using all 249 events for MDA0 (known host galaxy) and MDA1 (complete galaxy catalogue) are shown respectively in purple and red.

for all values of H_0, and the analysis is carried out by evaluating only the in-catalogue likelihood (Eq. 3.13).

The resulting posterior gives a value of $H_0 = 68.91^{+1.36}_{-1.22}$ km s^{-1} Mpc^{-1}, which is about 1.63 times broader than the result for MDA0 due to the lack of certainty about which galaxy is the true host (see Fig. 3.4). The introduction of this uncertainty means that the contributions from each event will be smoothed out, depending on the size of the event's sky localisation and the number of galaxies within it. As can be seen in Fig. 3.5, there is a far higher proportion of events for which the likelihood is relatively broad and less informative, in comparison to Fig. 3.3. However, many events clearly have a small number of galaxies in their sky-area, and hence still show clear peaks. The final result converges to the simulated value of H_0, demonstrating the applicability of the galaxy catalogue method.

3.4.3 MDA2: Incomplete galaxy catalogue

The next most complex scenario consists of incomplete galaxy catalogues, limited by an apparent magnitude threshold, meaning that EM selection effects become important for the first time. Three galaxy catalogues are used to investigate `gwcosmo`'s ability to handle varying levels of catalogue completeness. The apparent magnitude thresholds of each catalogue are $m_{\rm th} = 19.5$, 18 and 16, with respective completeness percentages of 75%, 50% and 25%. The combined 249-

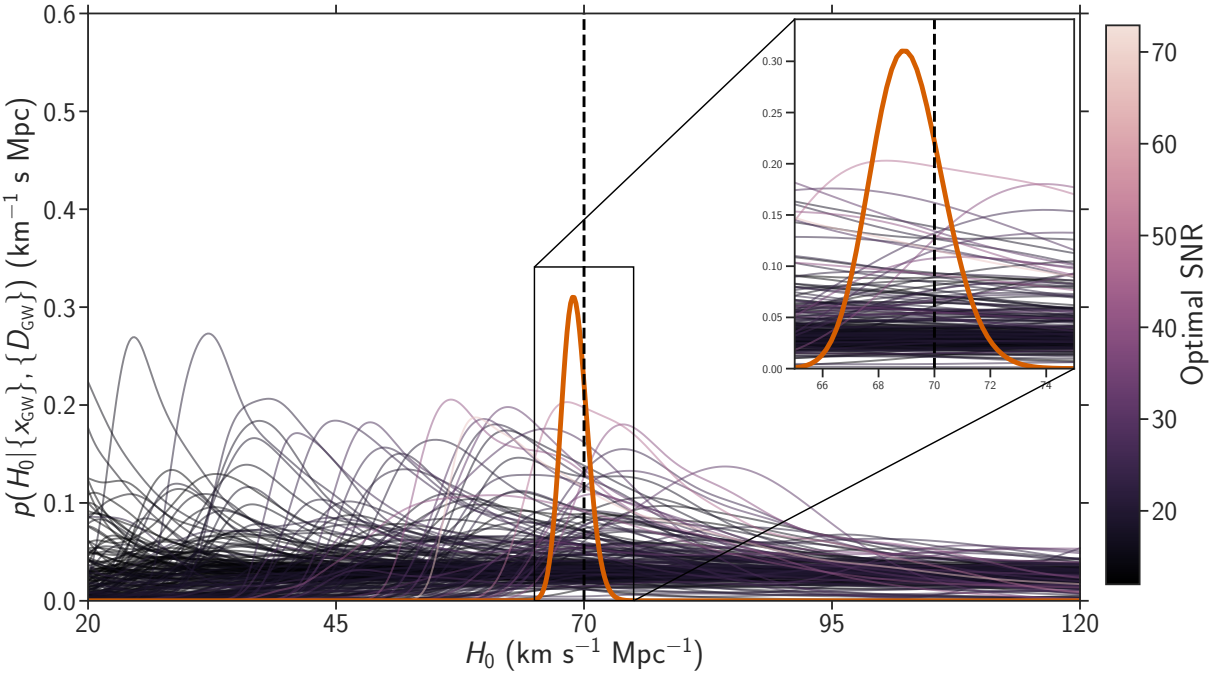

Figure 3.5: Individual and combined results for MDA1 (complete galaxy catalogue). The thick orange line shows the combined posterior probability density on H_0. Individual likelihoods (normalised and then scaled by an arbitrary value), for each of the 249 events, are shown as thin lines with shades corresponding to their optimal SNR. The simulated value of H_0 is shown as a vertical dashed line.

event posterior distributions on H_0 for these 3 cases are shown in Fig. 3.6, with the posterior from MDA1 included for reference.

As the catalogues become less complete, the combined H_0 posterior becomes wider. This is because the probability that the host galaxy is inside the catalogue decreases. The contribution from the galaxies within the catalogue is reduced, and the uninformative contribution from the out-of-catalogue term in Eq. 2.4 increases. This is visible in the individual likelihoods shown in Fig. 3.7, where instead of decreasing toward zero at high values of H_0, many of the individual likelihoods tend toward a non-zero constant. This is because, in the absence of EM data, and with the linear Hubble relation assumed in this work, the number of unobserved galaxies increases without limit as d_L^2. This behaviour is most noticeable for events at high distances (where the host has a lower probability of being in the catalogue), or for well-localised events where there is no catalogue support at the relevant redshifts within the event's sky area. However, enough events are detected at low distances, where the catalogues are more complete and so provide informative redshift information, to produce an upper constraint on H_0. Compared to MDA0 (Fig. 3.3) and MDA1 (Fig. 3.5), fewer individual likelihoods show strong peaks in Fig. 3.7. Although the final H_0 estimate is less precise, the results converge to the simulated value of H_0, demonstrating the applicability of this methodology to threshold-limited galaxy catalogues of about 25% completeness.

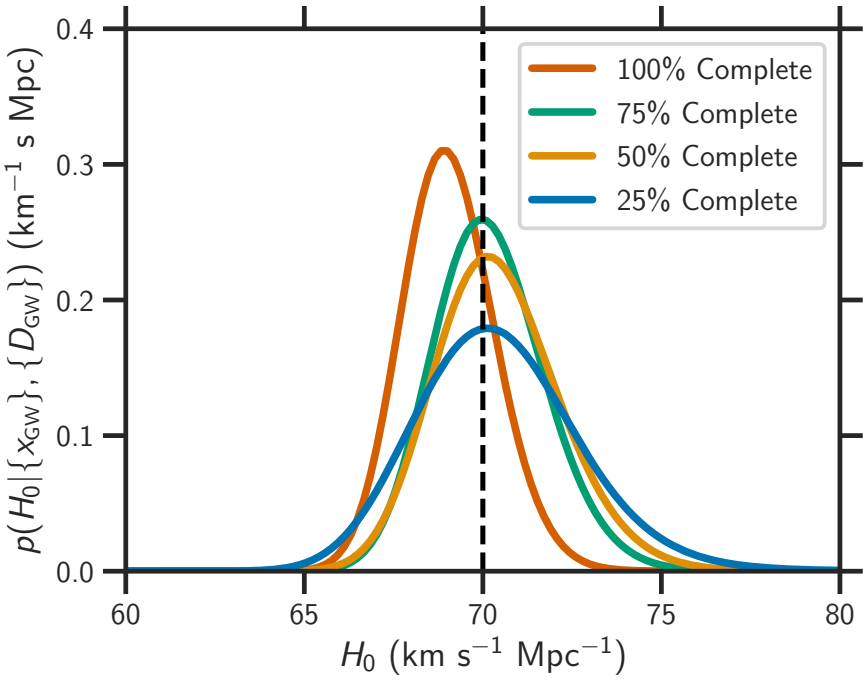

Figure 3.6: Comparison of MDA results with varying galaxy catalogue completeness. In MDA2, the simulated apparent magnitude threshold is varied to obtain galaxy catalogues of 75%, 50%, and 25% completeness. The corresponding posterior probability densities on H_0 using all 249 events are shown in green, yellow, and blue respectively. The MDA1 result is shown in red for comparison.

The results give $H_0 = 69.97^{+1.59}_{-1.50}$, $70.14^{+1.80}_{-1.67}$, and $70.14^{+2.29}_{-2.18}$ km s^{-1} Mpc^{-1} respectively for galaxy catalogues of 75%, 50%, and 25% completeness. See Section 3.4.5 for a more in-depth comparison of how galaxy catalogue completeness affects posterior width.

MDA2 demonstrates the need to know (or assess) the completeness of galaxy catalogues, and put in an appropriate out-of-catalogue term in the analysis. For any of the MDA2 catalogues, if the galaxy catalogue is assumed to be complete, when in reality it is not, the posterior distribution on H_0 becomes inconsistent with a value of 70 km s^{-1} Mpc^{-1}. This is because well-localised events for which the host is not inside the catalogue will not have support for the correct value of H_0.[9]

3.4.4 MDA3: Luminosity weighting

Up to this point, all galaxies have been considered as equally likely to host a GW source. MDA3, as described in Sec. 3.2.4, presents the case where this is no longer true by constructing a galaxy catalogue such that the probability of any single galaxy hosting a GW source is directly proportional to its luminosity. MDA3 includes the same EM selection effects as MDA2, in the sense

[9] In real catalogues, galaxy clustering might ensure that there are nearby bright galaxies in the catalogue, partially mitigating this bias.

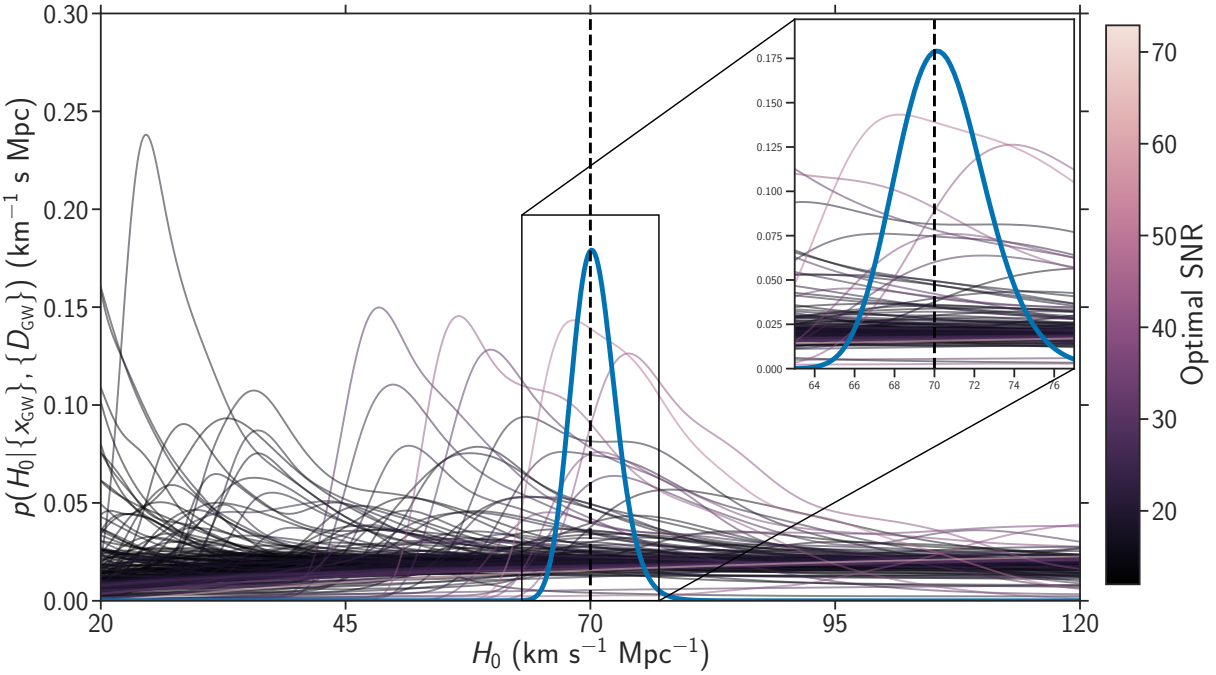

Figure 3.7: Individual and combined results for MDA2 with a 25% complete galaxy catalogue. The thick blue line shows the combined posterior probability density on H_0. Individual likelihoods (normalised and then scaled by an arbitrary value), for each of the 249 events, are shown as thin lines with shades corresponding to their optimal SNR. The simulated value of H_0 is shown as a vertical dashed line.

that the catalogue is magnitude limited. The completeness of this catalogue, defined in terms of luminosity rather than numbers of galaxies (as defined in Eq. (3.5)), is 50% out to 115 Mpc. This is indicative that approximately 50% of the detected GW events have host galaxies inside the catalogue.

To investigate the importance of luminosity weighting, MDA3 was analysed twice under different assumptions, given in Eq. 2.8. In the first, the analysis was matched to the known properties of the galaxy catalogue, such that the probability of any galaxy hosting a GW event was proportional to its luminosity. In the second, ignorance was feigned and the analysis was run with the assumption that each galaxy was equally likely to be host to a GW event (as was true in MDAs 1 and 2). The combined H_0 posteriors for both cases are shown in Fig. 3.8. The estimated values of the Hubble constant are $70.83^{+3.55}_{-2.72}$ km s^{-1} Mpc^{-1}(assuming hosts are luminosity weighted) and $69.50^{+4.20}_{-3.24}$ km s^{-1} Mpc^{-1}(assuming equal weights). By weighting the host galaxies with the correct function of their luminosities the constraint on H_0 improves — the uncertainty narrows by a factor of 1.2 compared to the case in which equal weights are assumed. Both results are consistent with the fiducial H_0 value of 70 km s^{-1} Mpc^{-1}. In the limit of a far greater number of events, one might expect to see a bias emerge when the assumptions in the analysis do not match those with which the catalogue was simulated. The luminosity weighting of host galaxies, by its very nature, increases the probability that the host galaxy is inside the

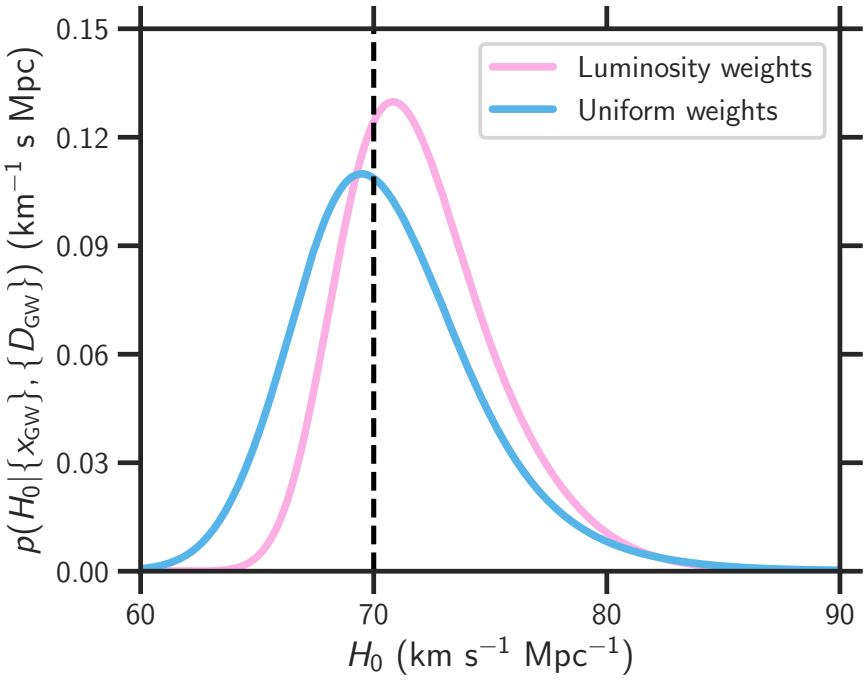

Figure 3.8: Comparison of the MDA3 results with and without luminosity weighting. In MDA3, by construction, the probability of any galaxy hosting a GW event is proportional to its luminosity. The pink curve shows the posterior probability density on H_0 when the analysis accounts for this. The blue curve shows the posterior density for the case where this extra information is ignored, and every galaxy treated as equally likely to be hosts.

galaxy catalogue; assuming equal weighting gives disproportionate weight to the contribution that comes from beyond the galaxy catalogue. However, for the 249 BNS events considered here, the final posteriors are too broad to imply any kind of bias.

3.4.5 Comparison between the MDAs

So far the results section has focused on individual event likelihoods and combined results for all 249 events. However, the dataset also allows the convergence of the combined H_0 posterior, as events are added to the analysis, to be studied. 100 different realisations of N events are randomly drawn from the 249, and the width of their combined posterior is calculated (noting that, as N increases there is more likely to be correlation between different sets of events, as the same events are likely to appear in multiple realisations). These intermediate combined posteriors are computed as a function of the number of events, and the resulting convergence is shown in Fig. 3.9. The fractional H_0 uncertainty (defined here as the half-width of the 68.3% credible interval divided by H_0, *i.e.* $\Delta^{68.3\%}_{H0}/2H_0$), is plotted against the number of events included in a randomly-selected group. The scatter between realisations of the group is indicated by the error bars, which encompass 68.3% of their range. There is a considerable variation between different realisations for the incomplete catalogues. For example, of the 100 realizations used,

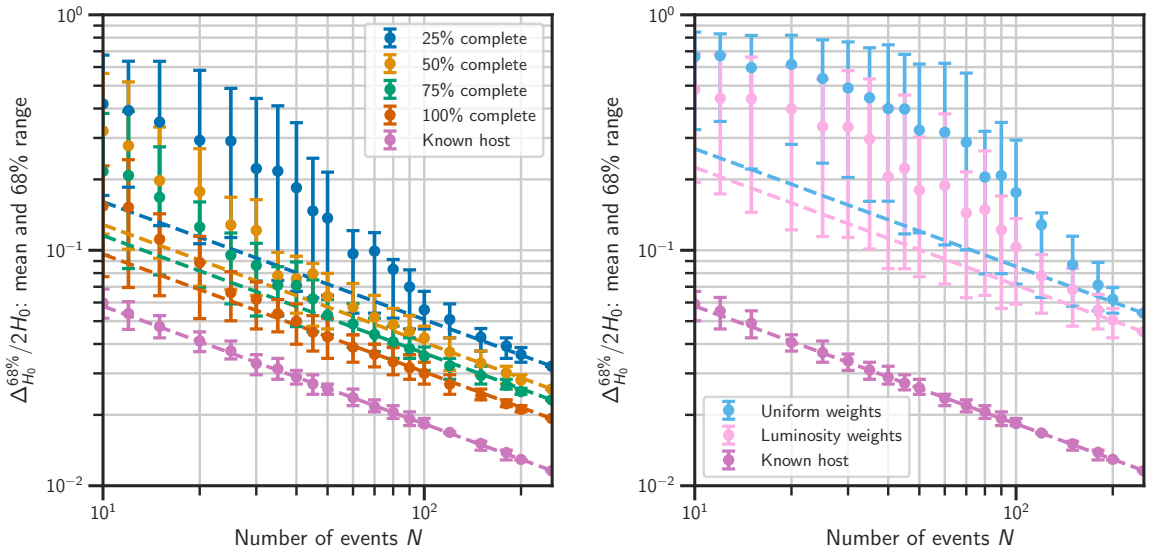

Figure 3.9: Fractional uncertainty on H_0 as a function of the number of events, N, for the combined H_0 posteriors. The fractional uncertainty in H_0 is defined as the half-width of the 68.3% highest probability density interval divided by 70 km s^{-1} Mpc^{-1}, and is shown as the plotted dots for all cases. The error bars contain 68% of the scatter arising from 100 different realisations of the events. *Left:* Purple, red, green, yellow and blue lines show show the associated host galaxy case (MDA0), complete galaxy catalogue case (MDA1), and the 75%, 50% and 25% completeness cases; fractional H_0 uncertainties of 1.13%, 1.84%, 2.21%, 2.48% and 3.20% are found respectively for the combined H_0 posterior from 249 events. *Right:* Convergence for MDA3 (event probability proportional to galaxy luminosity), analysed with luminosity-weighted likelihood (pink) or equally-weighted likelihood (light blue). Fractional H_0 uncertainties of 4.48% and 5.31% are found respectively for the combined H_0 posterior from 249 events. MDA0 (purple) is included for reference. The expected $1/\sqrt{N}$ scaling behaviour for large values of N is shown for all cases with the dashed lines. This scaling behaviour is met by all MDAs as the number of events reaches 249, but for the less informative, lower completeness MDAs the trend is slower to emerge. This is even more evident in MDA3, where the density of galaxies is 100 times greater, producing more potential hosts for each event. This is mitigated somewhat by the effect of luminosity-weighting the potential hosts (pink).

for 25% completeness and 40 events, there are groups that give $\sim$ 10% precision, but others that give $\sim$ 70% precision. This is because there is a wide range in the informativeness of a single event, and so realisations which include more informative events will constraint H_0 more than realisations which contain high proportions of relatively uninformative events. This is backed by the observation that, in general, the least complete galaxy catalogues allow for the most variation in the width of the posterior when low numbers of events are selected, whereas the variation possible for, *e.g.* MDA0 (known host galaxies) is relatively small even for low numbers of events.

With a sufficiently large number of events, a $1/\sqrt{N}$ scaling of the uncertainty with the number of events is expected [56, 60]. To check whether this behaviour is reached, the results for each MDA are fitted to the expected scaling, obtaining the coefficient of $1/\sqrt{N}$ by maximizing

its likelihood given the fractional uncertainties and their variances from the different realisations. The coefficient of the scaling is automatically dominated by the fractional uncertainties at large N where the variances are small. This scaling is shown for each MDA as a set of dashed lines in Fig. 3.9.

It can be seen that for each MDA, the results converge to the expected $1/\sqrt{N}$ scaling. The number of events required before this behaviour is reached is dependent on the amount of EM information available on average for each event, in agreement with the results of [60]. The direct counterpart case is always on the trend after $O(10)$ events, and shows a $\sim 18\%/\sqrt{N}$ convergence, comparable to and consistent with the results in [60, 61]. With the most complete galaxy catalogues, if the host galaxy is not directly identified it will take tens of events before this behaviour is reached. However, even the least complete catalogue (25%) appears to have reached this behaviour by the time all 249 events are combined. It should be noted that as the catalogues for MDAs 1 and 2 were not simulated realistically, their low density relative to the density of the universe means that these numbers should not be taken as predictions of how fast $1/\sqrt{N}$ may be reached (except, perhaps, in the counterpart case, although it should be remembered that peculiar velocities and redshift uncertainties have also been neglected). Even with a galaxy catalogue which is 25% complete, MDA2 leads to a posterior on H_0 which is only about a factor of 3 times broader than the counterpart case.

As the density of galaxies in MDA3 was increased by 2 orders of magnitude over MDAs 1 and 2, the final posteriors cannot be directly compared between MDAs. However, by plotting the equivalent convergence figure for MDA3 (including the "known host" case as a reference, see Fig. 3.9), the impact of increasing the density of galaxies in the universe on the rate at which the posterior converges on the $1/\sqrt{N}$ behaviour becomes clear. When there are more potential host galaxies, the results are overall less precise, and take longer to reach the $1/\sqrt{N}$ trend, a fact that is only partially mitigated by the additional information entering the analysis when luminosity-weighting of host galaxies is assumed.

3.4.6 Systematics studies

These results from MDAs 2 and 3 are expected to be sensitive to the luminosity distribution parameters. If the analysis uses values of the Schechter function parameters α and L^* that differ from those used to simulate the galaxy catalogues, it would be expected for the final result to exhibit a bias, if the result is precise enough. However, with variations of these parameters within their current measurement uncertainties, the resulting variation in the final result is small compared to the statistical uncertainties reached with the current set of MDAs.

It is worth noting that the choice of lower luminosity limit for the Schechter function in the analysis (taken to be $0.001L^*$ for the MDAs) can have an impact on the result. Increasing it can lead to discarding the faintest galaxies which, for MDA2, may be the host. However, in the luminosity weighted scenario of MDA3, fainter galaxies are less likely to be the hosts of GW

events, and so the results become less sensitive to the choice of lower luminosity limit (explored in greater detail in Section 4.8.3 of the next chapter).

3.5 Conclusion

Measurements of H_0 using real GW standard sirens will rely on a combination of the counterpart and galaxy catalogue methods, as appropriate for each new detected event, with catalogue incompleteness playing an important role for the more distant events. This chapter introduced and tested `gwcosmo`, a Python-based package which tackles both of these scenarios, including treatment of selection effects for both GWs (due to the limited sensitivity of GW detectors) and galaxy catalogues (due to the flux-limitations of EM observing channels). A series of MDAs were created and used to validate this implementation of the method, for up to 249 observed BNS events. For each of the MDAs, the final posterior on H_0 was found to be consistent with the value of $H_0 = 70\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ used to simulate the MDA galaxy catalogues, demonstrating that `gwcosmo` can produce unbiased results with this numbers of detections when the underlying population of GW events and galaxies are sufficiently well-known.

The MDAs were created using a set of simulated GW detections from the First Two Years dataset [76, 77], and building a mock universe for them to have come from. MDA0 consisted only of the GW data and the correctly identified host galaxy for each event. This tested `gwcosmo`'s implementation of the direct counterpart case, and its ability to account for GW selection effects, which are inherent to the simulated GW data, and so are common between all of the MDAs. Having succeeded here, the mock universe was then populated with galaxies, and no knowledge of the host galaxy was assumed for MDA1. This tested `gwcosmo`'s implementation of the most basic version of the galaxy catalogue case, in which all galaxies are contained within the catalogue and EM selection effects do not apply.

From this universe, mock galaxy catalogues were created for MDA2 by applying EM selection effects in the form of apparent magnitude thresholds. The incompleteness of these galaxy catalogues, and the subsequent possibility that the host was not contained within each catalogue, was accounted for by `gwcosmo`'s incompleteness terms. The resulting posterior on H_0 was consistent with the true value of H_0 for all catalogues, down to the case where only 25% of the host galaxies were inside the catalogue.

MDA3 introduced the weighting of host galaxies by their luminosity, and this additional information led to a tightening of the posterior on H_0 in comparison to the case where all galaxies were assumed to be equally likely hosts. The "uniform weights" analysis of MDA3 remained consistent with the simulated H_0 value, making it difficult to conclude whether an incorrect assumption would lead to a biased result with higher numbers of events, as might be expected. Future work can (and should) expand these studies to include a larger numbers of simulated GW events, to discern what level of accuracy assumptions about the universe require in order to

produce an unbiased result with greater precision.

The galaxy catalogue standard siren measurement of H_0, while less precise than the counterpart measurement, is still able to constrain H_0 though it requires at least an order of magnitude more events in order to reach a comparable accuracy (in the most realistic case of MDA3). However, in the galaxy catalogue case, the precision of results can vary by almost an order of magnitude depending on the particular realization of the detected population, before eventually converging to the expected $1/\sqrt{N}$ behaviour [56, 60]. Overall, these MDAs have validated the method outlined in Chapter 2 and its implementation in `gwcosmo` for simplified scenarios, but work is needed to improve in several directions, to test its applicability to BBHs (which are detectable out to much farther distances), realistic cosmology, and real galaxy catalogues.

Thinking first of more realistic EM data: the lack of redshift uncertainties and peculiar velocities in the MDAs makes the contributions from individual galaxies more informative than they would be in reality. Moreover, the low density of galaxies in MDAs 1 and 2, compared to the real universe, compounds this. MDA3, with a galaxy density of 1 galaxy per 70 Mpc^3, comes closest to the actual density of galaxies in the local universe of $\sim$ 1 galaxy per 200 Mpc^3 [80]. However, in comparison to real galaxy catalogues such as GLADE [80], the apparent magnitude threshold of 14 is very low, leading to a much smaller catalogue than would be available in reality. It could be expected that a real catalogue-only analysis with BNSs would fall somewhere between MDAs 2 and 3, but there are further competing factors to consider. Luminosity weighting of host galaxies, if trustworthy, will boost the probability of host galaxies being inside galaxy catalogues relative to the uniform weights scenario. Clustering of galaxies is also expected to improve the constraint on H_0 (see, *e.g.* [60,89]), since even when the host galaxy is not in the catalogue, it is likely that other nearby galaxies will be observed, and give support at the same redshifts. However, using more realistic catalogues will also require revisiting the assumption that the magnitude threshold of a catalogue is uniform across the full sky, as it can vary significantly according to the design of specific surveys. The presence of the Milky Way in the real universe means that, even for a particularly uniform survey, there will still be limits on the galaxies that are contained in the catalogue which are not well described by a uniform threshold. Chapters 4 and 5 will explore this topic in depth.

Also needing consideration is real GW data: Advanced LIGO and Virgo's first, second, and third observing runs have provided more than 50 confirmed GW detections [14, 15]. The majority of these have been from BBH systems, which are detected at a higher rate than BNSs because their greater mass allows them to be observed at much greater distances. Because BBHs are not expected to be detected with EM counterparts, and because galaxy catalogues are more likely to be incomplete for them, the galaxy catalogue method implemented and tested above will be particularly important. However, with real GW data comes real complications, most notably from GW selection effects. The unknown underlying source-frame mass distribution of BBHs, the changing sensitivity of GW detectors over time, the very definition of what is deemed

a “detection” in real data — all of these will need to be addressed (see Section 4.5 in the next chapter).

This chapter, and the MDA paper it was based on [64], have made huge progress in implementing and testing a useable methodology for GW cosmology with dark standard sirens, but there are many challenges ahead for applying this method to real data.

Chapter 4

Measuring the Hubble constant with real data

Declaration: the majority of material in this chapter is based on the O2 Hubble constant (O2-H_0) paper [66]. It should be noted that the results presented here may differ marginally from those in the published version of the paper, due to the correction of a couple of bugs in the analysis codes which had little impact on the results, but which have been corrected here for the sake of consistency.

Sections 4.2 and 4.3 are adapted from the O2-H_0 paper, edited for style and expanded with additional detail.

Section 4.4 is my own unpublished material.

Section 4.5 is mostly material from the O2-H_0 paper, edited for style and with additional details. Within it, Section 4.5.1 is my own unpublished material.

Sections 4.6 and 4.7 are from the O2-H_0 paper, edited only for style and phrasing.

Section 4.8 is my own unpublished material.

4.1 Introduction

The first and second observing runs of Advanced LIGO and Virgo took place between 2015 and 2017. In total, 10 BBHs and one BNS were detected [14]. An EM counterpart to the BNS was also observed, allowing the identification of its host galaxy, NGC4993 [11]. The redshift of this host galaxy, in combination with the luminosity distance estimate from the GW data, led to the first measurement of H_0 using standard sirens [13] – a remarkably informative result from a single event.

The 10 BBHs received, at first, markedly less attention on the cosmological front, as none had associated EM counterparts. However, the BBH GW170814 was highlighted as a reasonably well localised event, at a luminosity distance that was well within the range of the more complete galaxy surveys currently available. Combined with the fact that this event was lo-

calised within the region of the sky covered by the DES Year 3 catalogue, a galaxy catalogue with high completeness, and photometric redshift estimates for the majority of galaxies, interest was piqued in using the "dark siren" galaxy catalogue method of constraining H_0 with this event. The first result, published by the DES collaboration in association with the LIGO and Virgo collaborations in [65], showed that, due to a cluster of luminous galaxies around a redshift of 0.1, the result was reasonably informative around an H_0 of 70 km s^{-1} Mpc^{-1}, despite the large number of total galaxies within the event's localisation volume.

Of the rest of the BBHs, the majority had localisation volumes that were much larger than GW170814's, and therefore would be expected to produce relatively uninformative results. However, a couple of events had volumes that were of a similar size, leading to the possibility of a more informative H_0 measurement using all 10 BBHs than by just one alone. Not to mention, of course, that any measurement of H_0 using standard sirens should be self-consistent, and use all the GW events which pass the detection criteria modelled in the analysis. Picking and choosing events should be done with care, especially when in pursuit of a more informative measurement on H_0, as ignoring events without proper justification could introduce a bias to the result.

A measurement of H_0 by the LIGO and Virgo collaborations, based on the O1 and O2 detections, was published in [66]. The analysis which went into this paper forms the basis for this chapter, though this chapter will go into greater detail when explaining the specifics of how the analysis was carried out, the justifications for the choices made, and the additional sanity checks that went into ensuring that the final result was robust.

The rest of this chapter is laid out as follows: Section 4.2 introduces the O1 and O2 GW detections in detail, summarising their properties and discussing how they differ from the mock detections used in the MDAs of the previous chapter. Section 4.3 introduces the galaxy catalogues that are used to provide the redshift information for the dark siren analysis and discusses in which ways these catalogues are more complex than the mock catalogues used for the MDAs in the previous chapter. Section 4.4 outlines the improvements that were introduced to the `gwcosmo` implementation of the dark siren methodology in order to make it compatible with the real GW and EM data. Section 4.5 lays out the prior assumptions used for the main O2-H_0 analysis, as well as detailing some of the sanity checks that were used to get to that point. Section 4.6 gives the main results, equivalent to those reported in [66]. Section 4.7 explores how changing the various assumptions that were used in the main analysis affects the results. Section 4.8 describes an easily implementable improvement to the methodology used for the O2-H_0 result, and reanalyses the O1 and O2 detections to get an improved measurement on H_0. Finally, section 4.9 summarises this chapter and brings it to a close.

Throughout this chapter, as in [66], a ΛCDM cosmology is assumed, along with the best-fit Planck 2015 values of $\Omega_m = 0.308$ and $\Omega_\Lambda = 0.692$ for the fractional matter and dark energy densities in the present epoch respectively [90]. Although these parameters enter the redshift-

distance relationship central to the method for Bayesian inference of H_0, the results presented here are expected to be robust with regards to a variation of their values within the current measurement uncertainties.

4.2 Gravitational wave data

The confirmed GW detections following O1 and O2 consist of 10 BBHs and one BNS merger [14]. The BNS event, GW170817, was the most well-localised of all the detections and, at a distance of 40^{+10}_{-10} Mpc , was also the most nearby. The BBHs span a large range of distances from 320^{+120}_{-110} Mpc to 2840^{+1400}_{-1360} Mpc and are distributed over the sky with 90% credible regions as low as 39 deg^2 to as high as 1666 deg^2 (see Table 4.1 for an explicit list of the values these parameters take for each event). The GW data to be used in this analysis takes the form of posterior samples and skymaps associated with the data release of the first GW transient catalogue, GWTC-1 [14].[1]

One of the key parameters that determines how informative a GW dark siren is is its localisation volume. Ignoring the clustering of galaxies within the volume of an event, it is broadly true that the smaller the volume, the fewer the galaxies, and the more likely an event is to be informative on H_0. Of the BBHs, this singles out not only GW170814 (already mentioned in the introduction) but also GW150914 and GW170608. These three events have localisation volumes nearly an order of magnitude smaller than the rest of the BBHs.

It is also important to consider the BBH detections as part of a larger population. As discussed in sections 2.2 and 3.3.2, the normalisation of the GW likelihood for the measurement of H_0 requires being able to accurately model the GW selection effects, making it important to understand both the astrophysical population of BBHs and the real GW detector configuration and detection threshold.

Parameters which determine the detectability of a GW event are luminosity distance, inclination, and detector-frame chirp mass, $\mathcal{M}_z$, where $\mathcal{M}_z \equiv \mathcal{M}(1+z)$. In order to infer any information about the source-frame mass of a detected BBH, some cosmological assumptions will have to be made in order to undo the redshifting of the masses. This means, for the most distant event detected during O1 and O2, GW170729, the source-frame component masses could increase by as much as an additional 50% between the minimum and maximum prior values of H_0 assumed in this analysis. GW170729 is also the heaviest of the O1 and O2 detections, with a source-frame primary mass of $m_1 = 50.2^{+16.2}_{-10.2} M_\odot$ and a secondary mass of $m_2 = 34.0^{+9.1}_{-10.1} M_\odot$ (assuming Planck cosmology). The best-estimates of the BBH mass distribution parameters in [91] assume a cosmology in order to do so, and as such the same distributions cannot be assumed for this analysis without introducing bias. Care needs to be taken in choosing a source-frame mass prior for the BBH population in a cosmological analysis.

Beyond the source properties, GW detectability is a function of detector configuration and

[1] Available at: `https://www.gw-openscience.org/GWTC-1`.

detection threshold. Table 4.1 quotes the network SNR of each event in the search pipeline (PyCBC/GstLAL) in which the signal is the loudest, but in practice a detection is not claimed solely on the basis of the SNR. Additional data quality vetoes are applied in order to remove noise transients, and a ranking statistic is constructed, such as an inverse false alarm rate (IFAR) or a likelihood-ratio [14]. If all the O1 and O2 detections are to be used, a threshold on a ranking statistic rather than the SNR should be used as the selection criterion. However a distinction between the two does not cause an appreciable difference if the considered detections are significantly louder than transient noise artifacts (see, *e.g.* Appendix A.1 of [91]). There is therefore the option to place a more stringent threshold on the SNR and select a subset of loud events from the detected population (see section 4.5.1 for an in-depth exploration of this).

Event	SNR	$\Delta\Omega/\mathrm{deg}^2$	$d_\mathrm{L}/\mathrm{Mpc}$	z_event	V/Mpc^3	Galaxy catalogue	N_gal	m_th	$p(G\|z_\mathrm{event}, D_\mathrm{GW})$
GW150914	24.4	182	440^{+150}_{-170}	$0.09^{+0.03}_{-0.03}$	3.5×10^6	GLADE	3910	17.92	0.42
GW151012	10.0	1523	1080^{+550}_{-490}	$0.21^{+0.09}_{-0.09}$	5.8×10^8	GLADE	78195	17.97	0.01
GW151226	13.1	1033	450^{+180}_{-190}	$0.09^{+0.04}_{-0.04}$	2.4×10^7	GLADE	27677	17.93	0.41
GW170104	13.0	921	990^{+440}_{-430}	$0.20^{+0.08}_{-0.08}$	2.4×10^8	GLADE	42221	17.76	0.01
GW170608	15.4	392	320^{+120}_{-110}	$0.07^{+0.02}_{-0.02}$	3.4×10^6	GLADE	6267	17.84	0.60
GW170729	10.8	1041	2840^{+1400}_{-1360}	$0.49^{+0.19}_{-0.21}$	8.7×10^9	GLADE	77727	17.82	< 0.01
GW170809	12.4	308	1030^{+320}_{-390}	$0.20^{+0.05}_{-0.07}$	9.1×10^7	GLADE	18749	17.62	< 0.01
GW170814	16.3	87	600^{+150}_{-220}	$0.12^{+0.03}_{-0.04}$	4.0×10^6	DES-Y1	31554	23.84	> 0.99
GW170817	33.0	16	40^{+7}_{-15}	$0.01^{+0.00}_{-0.00}$	227	–	–	–	–
GW170818	11.3	39	1060^{+420}_{-380}	$0.21^{+0.07}_{-0.07}$	1.5×10^7	GLADE	1059	17.51	< 0.01
GW170823	11.5	1666	1940^{+970}_{-900}	$0.35^{+0.15}_{-0.15}$	3.5×10^9	GLADE	117680	17.98	< 0.01

Table 4.1: Relevant parameters of the O1 and O2 detections: network SNR for the search pipeline (PyCBC/GstLAL) in which the signal is the loudest, 90% sky localization region $\Delta\Omega$ (deg^2), luminosity distance d_L (Mpc, median with 90% credible intervals), and estimated redshift z_event (median with 90% range assuming Planck 2015 cosmology) from [14]. The remaining columns report the corresponding 90% 3D localisation comoving volumes, the number of galaxies within each volume for public catalogues which are the most complete in each event's sky area, and the apparent magnitude threshold, m_th, of the galaxy catalogue associated with the corresponding sky region (as described in Section 4.5.2). The final column gives the probability that the host galaxy is inside the galaxy catalogue for each event, $p(G|z_\mathrm{event}, D_\mathrm{GW})$, also evaluated at the median redshift for each event.

4.3 Galaxy catalogues

The differences between the mock catalogues described in Chapter 3 and the galaxy catalogues of the real world are many. A list of real-world factors that will need to be dealt with includes (but is not limited to) the following:

- A lack of full-sky galaxy catalogue coverage.
- Varying catalogue completeness within the bounds of the catalogue, especially for catalogues made up of different observing surveys.
- Redshift corrections due to peculiar velocities and converting from the heliocentric to the CMB frame.
- Redshift uncertainties, especially where spectroscopic redshift measurements are not available for all galaxies, and photometric redshift estimates have been made instead.
- Magnitude information in multiple bands, affected by extinction and redshifting.

Broadly, this list breaks down into two categories: things which affect the individual galaxies, and therefore will impact the "in-catalogue" part of the analysis; and bulk properties of the catalogue, which will impact how the "out-of-catalogue" corrections are applied. This section will focus on the former, as the latter is method-dependent, and will be revisited in section 4.4.2.

In order to carry out a cosmological analysis with the BBHs outlined in Section 4.2, appropriate galaxy catalogues need to be identified. The processing that raw galaxy surveys undergo to be converted into galaxy catalogues can vary between surveys and catalogues, and so each catalogue needs close inspection in order to ensure that it is fully understood which corrections have been applied in advance. First it is worth considering what information is required in order for a galaxy catalogue to be used in the dark siren analysis. Ideally it should include the following:

1. An estimate of galaxy redshifts in the CMB-frame, corrected for peculiar velocities for those at small enough redshifts for peculiar velocities to have a significant impact on the redshift estimate.
2. A quantification of the uncertainty of the redshift estimate, if it is significant (discussed in section 4.3.1 in more detail).
3. The right ascension and declination of the galaxy.
4. The (extinction-corrected) apparent magnitudes for each galaxy in a specific observation band. This point is nuanced; section 4.3.2 addresses it in more detail.

The Galaxy List for the Advanced Detector Era (GLADE) catalogue [80] is a curated galaxy catalogue made up of several different surveys, specifically put together for use with GW detections. It has close to full-sky coverage, and because of this is chosen as the default catalogue for the analysis (see section 4.3.3). As many of the GW events have large localisation areas, broad coverage is prioritised for the default catalogue, over the depth that is possible with certain catalogues which cover smaller patches of the sky. However, the DES Year 1 (Y1A1 GOLD or

simply Y1) catalogue [92, 93] covers the majority of the sky area of GW170814, and is complete to a much greater depth than the GLADE catalogue in the same patch of the sky (see section 4.3.4).[2]

The following sections discuss redshift uncertainties and observed galaxy magnitudes in more detail, before looking closely at the GLADE and DES-Y1 catalogues and describing the pre-processing which is done to them in order to convert them into comparable formats.

4.3.1 Redshift uncertainties

An important source of measurement uncertainty with galaxy catalogues is in the galaxy redshifts, which are often photometric estimates due to lack of spectroscopic measurements out to large redshifts. Spectroscopic measurements are costly, while photometric estimates are not, making them ideal for finding redshift estimates for galaxies in large surveys. However, this comes at the expense of accuracy, and photometric redshifts can have large uncertainties associated with them, up to 10s of percent. There are many ways in which photometric redshifts can be computed [94]. Modern methods include template fitting or machine learning methods (see, *e.g.* [95,96]). In general, photometric redshift estimates tend to become less reliable for galaxies at high redshifts, with low apparent magnitudes, and can introduce some artificial structure in the redshift distribution of the catalogue.

Most catalogues will quote a mean (or median) redshift value for each galaxy, and some uncertainty – usually a 1σ value, or the upper and lower 1σ bounds if the distribution is known to be asymmetric. In some cases, posterior samples may be provided for the redshift distribution of each galaxy, in which case these could be used directly for a Monte Carlo marginalisation over the redshift uncertainty. However, in the absence of this detailed information, modelling the uncertainty as, say, a Gaussian distribution allows some of the uncertainty in the redshift estimates to be taken into account, and will blur out artificial structure which the photometric redshifts may introduce.

4.3.2 Source-frame luminosities

For galaxy catalogues which are complete to high redshifts, it is important to note that the observed apparent magnitudes are redshifted from their source frame. Galaxies do not emit uniformly in all wavelengths, and galaxy surveys are only sensitive in certain bands. In order to compare the properties of galaxies at different redshifts, *e.g.* to apply luminosity weighting, or to apply a luminosity cut to the galaxy sample, the K correction, described in [97, 98], needs to be applied. In general,

$$M_a = m_b - DM - K \tag{4.1}$$

[2]It should be noted, however, that this catalogue is not the same as that used in [65], as that catalogue was not public at the time of writing the O2-H_0 paper.

where M_a is the desired absolute magnitude of a galaxy emitting in band a, m_b is the observed apparent magnitude using photometric bandpass b, DM is the distance modulus, and K is the K correction which will convert between the photometric bandpass b and band a. In general, the K correction term is redshift dependent, and so neglecting it could lead to systematic bias in the estimation of H_0. An analytical expression for its form will depend on the bands that it is converting between and can also be colour-dependent.

In the case where galaxies have high redshifts and large redshift uncertainties, this uncertainty must be accounted for in the calculation of the K corrections. Using only the mean redshift to calculate the K correction would lead to unrealistically bright or faint luminosities at the tail ends of the redshift distribution. Because of this, it is sensible for the K corrections to calculated in-situ during the "in-catalogue" part of the analysis, rather than as part of the galaxy catalogue pre-processing (see Section 4.4.3). However, because of the potential requirement of extra colour information in order to accurately model the K corrections, it is sensible to make the decision on the form of the K corrections when considering the catalogue as a whole. This is discussed for GLADE and DES in Sections 4.3.3 and 4.3.4.

4.3.3 GLADE

The GLADE version 2.4 catalogue is a composite galaxy catalogue, constructed from the GWGC [99], 2MPZ [100], 2MASS XSC [101], HyperLEDA [102] and SDSS-DR12Q [103] catalogues,[3] which gives it close to full sky coverage (see Fig. 1 of [80]). At low redshifts ($\lesssim 0.05$), galaxy redshifts are expected to be dominated by the peculiar velocity field, and GLADE reports peculiar-velocity-corrected redshifts following the reconstruction of [104]. GLADE provides apparent magnitudes in the B, J, H and K bands where the information is available. It is complete (in B-band luminosity) out to 37 Mpc and has an estimated completeness of 50% out to 91 Mpc (Fig. 2 of [80]).

The GLADE catalogue does not provide redshift uncertainties directly, although they are discussed in [80]. Galaxies with no redshift estimates, or for which a cosmological model was assumed in order to obtain a redshift estimate, are discarded for this cosmological analysis (about 17% of the total galaxies). Of the remainder, the galaxies which have photometric redshifts (approximately 39% of the sample) have an average 1σ uncertainty of $\Delta z = 1.5 \times 10^{-2}$, which is more or less independent from galaxy redshift. Galaxies with spectroscopic redshifts (around 17%) have a much smaller uncertainty of $\Delta z = 1.5 \times 10^{-4}$. Galaxies from HyperLEDA have a relative uncertainty of 36% and make up the remaining galaxies. Redshifts in GLADE are quoted in the heliocentric frame, and so before being used in the analysis a correction to the CMB frame is applied, which is of order 10^{-3} [105].

GLADE is most highly complete in the B-band, and so galaxies without B-band apparent magnitudes are discarded. Because GLADE is only complete to relatively low redshifts ($z <$

[3] GLADE is publicly available at: `http://glade.elte.hu`.

0.1), where the effect of K corrections is expected to be within the bounds of the uncertainty on the galaxy magnitudes, K corrections are not considered, and so no other magnitude information is required.

4.3.4 DES Year 1

The Dark Energy Survey is a five year survey mapping $\sim$ 300 million galaxies in five filters (*grizY*) over 5000 deg^2. The DES [92, 93] Year 3 (Y3) "gold" catalog was used for the H_0 estimate with GW170814 in [65]. Here, however, the publicly available DES-Y1 catalogue [93] is used.[4]

First, objects in the DES-Y1 catalogue that are classified as high-confidence galaxies are selected using the default classification scheme, "MODEST_CLASS" [92, 106]. These galaxies have photometric redshift estimates which were derived using the Bayesian Photometric Redshift (BPZ) template fitting method [107]. Galaxies with redshift errors larger than twice their corresponding quoted median redshift value (around 5%), are discarded. Such a choice is not expected to impact the result since the discarded galaxies are highly uninformative.

The DES *grizY* magnitudes are then converted to the SDSS *ugriz* system using the photometric transformations provided in the DES-Y1 paper [92], which requires discarding a further $\sim$ 5% of galaxies with inadequate color information. This transformation enables K corrections to be calculated using the calculator described in [108, 109].[5] These K corrections are valid out to redshifts of 0.5, which means that galaxies above $z = 0.5$ should be discarded. Because the photometric redshift uncertainties are large, this redshift cut is not done as part of the galaxy catalogue pre-processing, but in-situ as part of the `gwcosmo` analysis and the methodology is adapted to remain self-consistent (see section 4.4.3).

4.4 Adapting `gwcosmo` to handle real data

It is clear, after discussing the GW data introduced in section 4.2 and the galaxy catalogues introduced in section 4.3, that the `gwcosmo` implementation described in the previous chapter requires some improvements in order to successfully tackle the challenges posed by real data. This section details the main changes that `gwcosmo` underwent before the O2-H_0 analysis took place. Section 4.4.1 details how the GW posterior samples are re-weighted to marginalise over the same source-frame mass prior used to compute the GW selection effects. Section 4.4.2 outlines an update to the methodology that allows GW cosmology to be carried out with galaxy catalogues with non-uniform sky coverage. Section 4.4.3 presents details of the method in a mathematical framework, including how catalogue non-uniformity, redshift uncertainties, and hard redshift cuts are applied.

[4] DES-Y1 is available at: `https://des.ncsa.illinois.edu/releases/y1a1`.

[5] The K-corrections calculator is available at `http://kcor.sai.msu.ru/`.

4.4.1 Posterior samples re-weighting

The GW data, $x_{\rm GW}$, has undergone parameter estimation which provides a set of posterior samples on source parameters, including the luminosity distance d_L, sky-coordinates Ω, and the observed component masses in the detector frame $m^{\rm det}_{1,2}$. As before, it is the likelihood of the GW data $p(x_{\rm GW}|H_0, I)$ which is necessary for the analysis, which requires removing the priors used for the parameter estimation. In particular, this means that the d_L^2 prior must be removed before the priors on z and H_0 can be applied. However, in addition a more subtle effect comes into play regarding source- and detector-frame masses. The treatment of GW selection effects requires a marginalisation over the source-frame mass distribution, while the data $x_{\rm GW}$ contains information about the detector-frame masses. In order to be self-consistent, the analysis must be performed using the same prior assumptions on both the individual event data, and the normalizing probability of detection term, $p(D_{\rm GW}|H_0, I)$ (from Eq. 2.2). This requires removing the detector-frame mass prior and re-weighting the posterior samples with the source-frame mass prior.

The use of a KDE to approximate the likelihood of the GW data, $p(x_{\rm GW}|z, H_0, I)$, was outlined in Section 3.3.1. To explicitly derive how the mass samples re-weighting enters the analysis, and how it relates to the KDE, $p(x_{\rm GW}|z, H_0, I)$ is taken as a starting point. First marginalising over the necessary source-frame mass prior gives

$$p(x_{\rm GW}|z, H_0, I) = \int_{m_{\rm min}}^{m_{\rm max}} \int_{m_{\rm min}}^{m_1} p(x_{\rm GW}|z, m_1, m_2, H_0, I) p(m_1, m_2|I)\, dm_2 dm_1. \tag{4.2}$$

The source-frame mass distribution is $p(m_1, m_2|I)$ where in this case I includes the hyperparameters describing the astrophysical model (concrete details of the assumed model are discussed later in Section 4.5). It can be rewritten as

$$p(x_{\rm GW}|z, H_0, I) = \int_{m_{\rm min}}^{m_{\rm max}} \int_{m_{\rm min}}^{m_1} p(x_{\rm GW}|d_L(z, H_0), m_1^{\rm det}(m_1, z), m_2^{\rm det}(m_2, z), I) p(m_1, m_2|I)\, dm_2 dm_1, \tag{4.3}$$

where the relationship between m and $m^{\rm det}$ is $m^{\rm det} = (1 + z)m$. Grouping z, H_0, and $m_{1,2}$ into d_L and $m^{\rm det}_{1,2}$ on the right hand side of the probability distribution on $x_{\rm GW}$ requires no additional changes. The GW data is in the form of posterior samples on d_L and $m^{\rm det}_{1,2}$. The first term, $p(x_{\rm GW}|d_L, m_1^{\rm det}, m_2^{\rm det}, I)$, can be rewritten:

$$\begin{aligned} p(x_{\rm GW}|d_L, m_1^{\rm det}, m_2^{\rm det}, I) &= \frac{p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I) p(x_{\rm GW}|I)}{\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)} \\ &\propto \frac{p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I)}{\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)} \end{aligned} \tag{4.4}$$

where $p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I)$ is the distribution that the posterior samples are drawn from, and $\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)$ are the priors on d_L and $m_{1,2}^{\rm det}$ used for parameter estimation. Substituting this into Eq. 4.2 gives

$$p(x_{\rm GW}|z, H_0, I) \propto \int_{m_{\rm min}}^{m_{\rm max}} \int_{m_{\rm min}}^{m_1} \frac{p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I)}{\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)} p(m_1, m_2|I)\, dm_2 dm_1. \tag{4.5}$$

Now looking more closely at the integral over m_1, m_2: because the samples are in the detector frame it helps to reframe the integral to the detector frame too. Doing so introduces a Jacobian:

$$\begin{aligned} dm_1 dm_2 &= \left|\frac{\partial(m_1, m_2)}{\partial(m_1^{\rm det}, m_2^{\rm det})}\right| dm_1^{\rm det} dm_2^{\rm det}, \\ &= \frac{1}{(1+z)^2}\, dm_1^{\rm det} dm_2^{\rm det}, \end{aligned} \tag{4.6}$$

which leads to the following:

$$p(x_{\rm GW}|z, H_0, I) \propto \int_{m^{\rm det}(m_{\rm min}, z)}^{m^{\rm det}(m_{\rm max}, z)} \int_{m^{\rm det}(m_{\rm min}, z)}^{m^{\rm det}(m_1, z)} \frac{p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I)}{\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)} \frac{p(m_1(m_1^{\rm det}, z), m_2(m_2^{\rm det}, z)|I)}{(1+z)^2}\, dm_2^{\rm det} dm_1^{\rm det}. \tag{4.7}$$

The posterior samples on d_L, $m_1^{\rm det}$ and $m_2^{\rm det}$ for a GW event can be approximated as a sum over delta-functions:

$$p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(d_L - d_{Li}) \delta(m_1^{\rm det} - m_{1i}^{\rm det}) \delta(m_2^{\rm det} - m_{2i}^{\rm det}), \tag{4.8}$$

where the subscript i refers to the $i^{\rm th}$ sample. If converting the luminosity distance samples to a distribution in redshift space,

$$p(z, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, H_0, I) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(z - z(d_{Li}, H_0)) \delta(m_1^{\rm det} - m_{1i}^{\rm det}) \delta(m_2^{\rm det} - m_{2i}^{\rm det}), \tag{4.9}$$

where the relation between the two distributions is defined as follows:

$$p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, I) = p(z, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, H_0, I) \left|\frac{\partial z}{\partial d_L}\right|. \tag{4.10}$$

That is, the conversion between the two introduces a Jacobian $|\partial z/\partial d_L|$.

At this point it is worth taking a step back to consider what a KDE is mathematically. Assuming the KDE in question is a Gaussian KDE (as `gwcosmo`'s implementation makes use of

SciPy's gaussian_kde [86]), it can be expressed as follows:

$$f_{\rm KDE}(z; H_0) = \frac{1}{\sum_{i=1}^{N} w_i(H_0)} \sum_{i=1}^{N} w_i(H_0)\mathcal{G}(z - z(d_{Li}, H_0); \sigma). \tag{4.11}$$

Here $f_{\rm KDE}(z; H_0)$ means that the KDE is on redshift samples, but the returned function is itself a function of H_0 as well. The term $\mathcal{G}(z - z(d_{Li}, H_0); \sigma)$ is a Gaussian centred at $z(d_{Li}, H_0)$, where d_{Li} is the luminosity distance of the ith posterior sample, and σ is the standard deviation of the Gaussian which is related to the KDE smoothing parameter, and is chosen internally by SciPy. Each Gaussian has a weight, $w_i(H_0)$ which defines how important its contribution is to the final distribution. The sum over $w_i(H_0)$ in the prefactor ensures that the final KDE is properly normalised.

It should be noted here that the choice to do a KDE on redshift samples vs luminosity distance samples is purely down to preference (in this case it is for consistency with the denominator, so that the final expression for $p(x|z, H_0, I)$ can be evaluated directly on redshift, rather than converting redshift to luminosity distance before evaluating). Choosing to do the KDE on redshift samples means that the Jacobian $|\partial z/\partial d_L|$ is included here, while it wasn't present in the KDE from Chapter 3. Returning to Eq. 4.7, and substituting in the sum over samples in place of $p(d_L, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW})$ gives

$$\begin{aligned} p(x_{\rm GW}|z, H_0, I) \propto & \int_{m^{\rm det}(m_{\rm min},z)}^{m^{\rm det}(m_{\rm max},z)} \int_{m^{\rm det}(m_{\rm min},z)}^{m^{\rm det}(m_1,z)} \frac{p(z, m_1^{\rm det}, m_2^{\rm det}|x_{\rm GW}, H_0, I)}{\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right| \\ & \times \frac{p(m_1(m_1^{\rm det}, z), m_2(m_2^{\rm det}, z)|I)}{(1+z)^2}\, dm_2^{\rm det} dm_1^{\rm det}, \\ \propto & \frac{1}{N}\sum_{i=1}^{N} \int_{m^{\rm det}(m_{\rm min},z)}^{m^{\rm det}(m_{\rm max},z)} \int_{m^{\rm det}(m_{\rm min},z)}^{m^{\rm det}(m_1,z)} \frac{1}{\pi(d_L, m_1^{\rm det}, m_2^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right| \frac{p(m_1(m_1^{\rm det}, z), m_2(m_2^{\rm det}, z)|I)}{(1+z)^2} \\ & \times \delta(z - z(d_{Li}, H_0))\delta(m_1^{\rm det} - m_{1i}^{\rm det})\delta(m_2^{\rm det} - m_{2i}^{\rm det})\, dm_2^{\rm det} dm_1^{\rm det}. \end{aligned} \tag{4.12}$$

The integrals over $m_1^{\rm det}$ and $m_2^{\rm det}$ can now be carried out using Monte Carlo integration, producing the following:

$$p(x_{\rm GW}|z, H_0, I) \propto \frac{1}{N}\sum_{i=1}^{N} \frac{p(m_1(m_{1i}^{\rm det}, z), m_2(m_{2i}^{\rm det}, z)|I)}{\pi(d_L, m_{1i}^{\rm det}, m_{2i}^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right| \frac{1}{(1+z)^2} \delta(z - z(d_{Li}, H_0)). \tag{4.13}$$

The similarity between this expression and the expression for a KDE (Eq. 4.11) is clear. If the delta function on z can be approximated as a Gaussian, the rest of the terms within the sum must be analogous to the KDE weights term. The issue is, of course, that weights are a function of the continuous variable z, not the individual redshift samples $z(d_{Li}, H_0)$. The delta-functions

$\delta(z - z(d_{Li}, H_0))$ would pick out the redshifts of specific posterior samples, leading to

$$w_i(H_0) = \frac{p(m_1(m_{1i}^{\rm det}, z(d_{Li}, H_0)), m_2(m_{2i}^{\rm det}, z(d_{Li}, H_0))|I)}{\pi(d_{Li}, m_{1i}^{\rm det}, m_{2i}^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right|_{z=z(d_{Li},H_0)} \frac{1}{(1 + z(d_{Li}, H_0))^2}. \quad (4.14)$$

However, allowing $\delta(z - z(d_{Li}, H_0))$ to be represented as a Gaussian, $\mathcal{G}(z - z(d_{Li}, H_0); \sigma)$ means that z should technically remain as a continuous variable that can change over the width of the Gaussian. Given that the width of the Gaussians in the KDE is very small relative to the overall size of the distribution over z, due to the high density of posterior samples, the impact of this approximation is small. Equation 4.13 can therefore be rewritten as

$$p(x_{\rm GW}|z, H_0, I) \propto \frac{1}{N}\left(\sum_{i=1}^{N} w_i(H_0)\right) f_{\rm KDE}(z; H_0), \quad (4.15)$$

where $w_i(H_0)$ is defined as in Eq. 4.14, and the sum over weights is necessary to undo the KDE's inherent normalisation ($f_{\rm KDE}(z; H_0)$ integrates to 1 over z) and to ensure the correct relative weights for different values of H_0.

A proof using the empty catalogue case

The use of a KDE to represent $p(x_{\rm GW}|z, H_0, I)$ can be tested using the numerator of the empty catalogue case (Eq. 2.29). Ignoring the integrals over Ω, it can be written as follows:[6]

$$p(x_{\rm GW}|s, H_0, I) = \int_0^\infty p(x_{\rm GW}|z, H_0, I) p(z|I) p(s|z, I)\, dz. \quad (4.16)$$

Here $p(z|I)p(s|z, I)$ is the prior on potential GW host galaxy redshifts. As no EM information enters the empty catalogue analysis, this takes the form of some continuous function (such as uniform in comoving volume-time), rather than a discrete set of delta-functions at known galaxy redshifts. Taking Eq. 4.13 as the starting point gives

$$p(x_{\rm GW}|H_0, I) \propto \int_0^\infty \frac{1}{N}\sum_{i=1}^{N} \frac{p(m_1(m_{1i}^{\rm det}, z), m_2(m_{2i}^{\rm det}, z)|I)}{\pi(d_L, m_{1i}^{\rm det}, m_{2i}^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right| \frac{p(z|I)p(s|z, I)}{(1 + z)^2} \delta(z - z(d_{Li}, H_0))\, dz. \quad (4.17)$$

Because the redshift prior is a continuous function, it can be evaluated at specific delta-function locations, meaning that a Monte Carlo integral over redshift can be carried out directly, without

[6]The separation of $p(\Omega|x_{\rm GW}, I)$ and $p(d_L|x_{\rm GW}, I)$ as described in Section 3.3.1 means that the integral over Ω is entirely separable from the integral over z.

requiring a KDE:

$$p(x_{\rm GW}|H_0, I) \propto \frac{1}{N}\sum_{i=1}^{N} \frac{p(m_1(m_{1i}^{\rm det}, z(d_{Li}, H_0)), m_2(m_{2i}^{\rm det}, z(d_{Li}, H_0))|I)}{\pi(d_{Li}, m_{1i}^{\rm det}, m_{2i}^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right|_{z=z(d_{Li},H_0)} \frac{p(z(d_{Li}, H_0)|I)p(s|z(d_{Li}, H_0), I)}{(1+z(d_{Li}, H_0))^2}. \tag{4.18}$$

However, a KDE could still be constructed on redshift samples. If the weights are re-defined to include the $p(z|I)p(s|z, I)$ terms, *i.e.*

$$\hat{w}_i(H_0) = \frac{p(m_1(m_{1i}^{\rm det}, z(d_{Li}, H_0)), m_2(m_{2i}^{\rm det}, z(d_{Li}, H_0))|I)}{\pi(d_{Li}, m_{1i}^{\rm det}, m_{2i}^{\rm det}|I)} \left|\frac{\partial z}{\partial d_L}\right|_{z=z(d_{Li},H_0)} \frac{p(z(d_{Li}, H_0)|I)p(s|z(d_{Li}, H_0), I)}{(1+z(d_{Li}, H_0))^2}, \tag{4.19}$$

such that

$$p(x_{\rm GW}|z, H_0, I)p(z|I)p(s|z, I) \propto \frac{1}{N}\left(\sum_{i=1}^{N} \hat{w}_i(H_0)\right) \hat{f}_{\rm KDE}(z; H_0), \tag{4.20}$$

then the integral over redshift becomes

$$p(x_{\rm GW}|H_0, I) \propto \frac{1}{N}\left(\sum_{i=1}^{N} \hat{w}_i(H_0)\right) \int_0^\infty \hat{f}_{\rm KDE}(z; H_0)\, dz. \tag{4.21}$$

As `SciPy`'s `gaussian_kde` is normalised by default, the integral evaluates as 1, leaving

$$p(x_{\rm GW}|H_0, I) \propto \frac{1}{N}\left(\sum_{i=1}^{N} \hat{w}_i(H_0)\right), \tag{4.22}$$

which is identical to Eq. 4.18 if the weights are defined as in Eq. 4.19.

4.4.2 Handling inhomogeneous galaxy catalogue completeness: the GBO method

On the galaxy catalogue side of things, there are additional changes that need to be made to `gwcosmo`. The most important is that the assumption that galaxy catalogue completeness is uniform over the whole sky is clearly not applicable to either the GLADE catalogue or the DES catalogue. To step away from this assumption, first lets assume that there is a galaxy catalogue of uniform completeness (*i.e.* one which can be described by a constant $m_{\rm th}$) which only covers some fraction of the sky. The sky area covered by the catalogue can be denoted as Ω_G, and the area outside the catalogue as Ω_O, such that $\Omega_G + \Omega_O$ covers the whole sky. The likelihood for an

individual GW event can then be written as follows:

$$p(x_{\rm GW}|D_{\rm GW}, H_0, I) = \int p(x_{\rm GW}|\Omega, D_{\rm GW}, H_0, I)p(\Omega|I)\, d\Omega$$
$$= \int_{\Omega\in\Omega_G} p(x_{\rm GW}|\Omega, D_{\rm GW}, H_0, I)p(\Omega|I)\, d\Omega + \int_{\Omega\in\Omega_O} p(x_{\rm GW}|\Omega, D_{\rm GW}, H_0, I)p(\Omega|I)\, d\Omega. \tag{4.23}$$

As the first term corresponds to the part of sky covered by the galaxy catalogue, it will break down into in-catalogue and out-of-catalogue terms. It is equivalent to Eq. 2.4 with limits on the integrals over Ω. The second term has no catalogue coverage, and so covers the remainder of the sky from redshift 0 to ∞.

Mathematically, Eq. 4.23 can be written out in a way which follows the notation in Eq. 2.4:

$$\begin{aligned} p(x_{\rm GW}|D_{\rm GW}, H_0, I) =& p(x_{\rm GW}|G, D_{\rm GW}, H_0, I)p(G|D_{\rm GW}, H_0, I) \\ &+ p(x_{\rm GW}|B, D_{\rm GW}H_0, I)p(B|D_{\rm GW}, H_0, I) \\ &+ p(x_{\rm GW}|O, D_{\rm GW}, H_0, I)p(O|D_{\rm GW}, H_0, I). \end{aligned} \tag{4.24}$$

Here, G denotes that the host galaxy is contained within the galaxy catalogue (as before). The $\bar{G}$ term has been split into two further terms, B and O. B denotes that the host galaxy lies somewhere beyond the galaxy catalogue (within the patch of sky the catalogue covers, Ω_G, but at an apparent magnitude above the catalogue threshold). O denotes that the host galaxy lies outside the patch covered by the catalogue, in Ω_O. There remains the necessary constraint of

$$p(G|D_{\rm GW}, H_0, I) + p(B|D_{\rm GW}, H_0, I) + p(O|D_{\rm GW}, H_0, I) = 1, \tag{4.25}$$

as the host galaxy must be somewhere. The probability that a host galaxy is outside the (sky area) bounds of the galaxy catalogue depends only on the relative fractions of the sky covered by the catalogue, or not. As such,

$$p(G|D_{\rm GW}, H_0, I) + p(B|D_{\rm GW}, H_0, I) = \frac{\Omega_G}{4\pi}, \tag{4.26}$$

and

$$p(O|D_{\rm GW}, H_0, I) = 1 - \frac{\Omega_G}{4\pi}, \tag{4.27}$$

where $\Omega_G/4\pi$ is the fraction of the sky covered by the galaxy catalogue under this definition.

This method can be applied to both the DES and GLADE catalogues, but requires slightly different assumptions about how Ω_G is defined in both cases. The specifics are outlined below.

Applying the method to the DES catalogue

The DES catalogue has a relatively uniform $m_{\rm th}$ over the patch of sky that it covers, but it does not cover the full sky area of GW170814, for which only approximately 88% has catalogue support. Thus the integrals over Ω are particularly important, and in order to easily implement this, rectilinear limits on RA (α) and dec (δ) were placed on the DES catalogue around the patch covered by GW170814. The chosen boundary (in degrees) was defined by $35.0 \leq \alpha \leq 55.0$ and $-55.0 \leq \delta \leq -40.0$. All galaxies outside this area were discarded, and the integral limits were defined by those bounds.

Applying the method to the GLADE catalogue

The problematic aspect of the GLADE catalogue is not that it has hard edges, but that it is made up of many different surveys with different magnitude depths, which cannot be accurately represented by a single apparent magnitude threshold. In order to make this approximation a little less egregious, the patch of sky defined as covered by the catalogue, Ω_G, was adapted for each event, such that the limits on the galaxy catalogue are defined to follow the bounds of the GW event's sky-localisation region. The boundary of this region can be defined by a confidence interval, such that it contains some fraction of the GW sky probability. If this threshold is chosen to contain 100% of the GW sky probability,[7] this leads to $p(x_{\rm GW}|O, D_{\rm GW}, H_0, I) = 0$, and Eq. 4.24 simplifies to

$$\begin{aligned} p(x_{\rm GW}|D_{\rm GW}, H_0, I) = p(x_{\rm GW}|G, D_{\rm GW}, H_0, I)p(G|D_{\rm GW}, H_0, I) \\ + p(x_{\rm GW}|B, D_{\rm GW}H_0, I)p(B|D_{\rm GW}, H_0, I). \end{aligned} \tag{4.28}$$

The validity of choosing the bounds of the catalogue like this is debatable, as technically the redshift prior should be common to all GW events – but this approximation was deemed less problematic than the assumption that $m_{\rm th}$ is uniform across the full sky for GLADE (and Chapter 5 will do away with this approximation altogether).

4.4.3 A(nother) practical form for the methodology

To be explicit, the following equations lay out the form each part of Eq. 4.24 takes, having made the same approximations as outlined in section 3.3 (the separability of Ω and d_L for GW data, and the uniformity of $D_{\rm GW}$ across the sky). It also covers the ability to include a hard redshift cut on the galaxy catalogue, such that all galaxies above some redshift limit $z_{\rm cut}$ are excluded from

[7] Technically it is impossible to draw a confidence interval that includes 100% of the probability. In reality this would be the full sky. However, because skymaps are used in practice, and each pixel of the skymap has a finite probability associated with it, it is possible to identify all the pixels with non-zero probabilities and use this to define Ω_G, which is what was done in this case. This is 100% of the GW sky probability, to within a rounding error, and is referred to as "100% probability" throughout the rest of this chapter.

the analysis – necessary in some cases where the redshift or luminosity estimates of galaxies are unreliable above some redshift, due to K corrections or other sources of uncertainty. The redshift cut will need to be accounted for within the methodology, in order to avoid biasing the results, unless it is at such a high redshift that all GW data has no support above it, for any value of H_0, and the probability of detecting any event from the population has dropped to zero. Given that the K corrections considered for the DES catalogue are only reliable up to $z = 0.5$, and during O2 there was support for detecting GWs as far as a redshift of 1, this adaption of the methodology is necessary.

The exact separation of GW sky and distance information differs from Section 3.3.1 due to the differences in the way the KDE was constructed (described in Section 4.4.1). This time, $p(x_{\rm GW}|z, \Omega, s, H_0, I)$ is represented by

$$p(x_{\rm GW}|z, \Omega, s, H_0, I) \approx C\kappa(z, H_0)\frac{p(\Omega|x_{\rm GW}, s, I)}{\pi(\Omega|I)}, \tag{4.29}$$

where $\kappa(z, H_0)$ is defined to be the right-hand side of Eq. 4.15, $p(\Omega|x_{\rm GW}, I)$ comes from the GW skymap, $\pi(\Omega|I)$ is a uniform prior on the sky, and C is a constant which has no bearing on the results, as it appears identically in Eqs. 4.30, 4.34 and 4.35 below.

The likelihood of the GW data when the GW host is inside the galaxy catalogue is found by taking Eq. 2.17 as a starting point and substituting in Eq. 4.29, to give

$$\frac{1}{C}p(x_{\rm GW}|G,D_{\rm GW}, s, H_0, I) = \frac{\sum_{i=1}^{N_{\rm gal}} \frac{1}{N_{{\rm samp}_i}} \sum_{j=1}^{N_{{\rm samp}_i}} \kappa(z_j, H_0)\frac{p(\Omega_i|x_{\rm GW}, I)}{\pi(\Omega_i|I)} p(s|z_j, I)p(s|M(z_j, m_i, H_0), I)}{\sum_{i=1}^{N_{\rm gal}} \frac{1}{N_{{\rm samp}_i}} \sum_{j=1}^{N_{{\rm samp}_i}} p(D_{\rm GW}|z_j, s, H_0, I)p(s|z_j, I)p(s|M(z_j, m_i, H_0), I)}, \tag{4.30}$$

where the redshift uncertainty of individual galaxies are taken into account with a Monte-Carlo integration using redshift samples, z_j, drawn from the redshift distribution of each galaxy in the galaxy catalogue. The absolute magnitude, $M(z_j, m_i, H_0)$ is calculated on a per-sample basis and includes the K correction, if applicable to the catalogue being used. Calculating this integral by a sum is fast, relative to the speed of explicitly integrating over z for each galaxy, but for very high numbers of samples (total), it will slow down (a concern for galaxy catalogues that contain O(millions) of galaxies). It is therefore worth noting that the number of samples drawn from each galaxy, $N_{{\rm samp}_i}$, can be chosen based on properties of the galaxy to speed up the efficiency of the implementation. Galaxies with particularly small redshift uncertainties, or lower luminosities (where luminosity weighting is in play), will generally need fewer samples to accurately represent their uncertainty distributions.[8] In order to incorporate a hard redshift cut here, every

[8]For GLADE and DES it was found that drawing 10,000 samples from 0.01% galaxies with the lowest apparent magnitudes, and 20 samples from the rest was enough to make statistical fluctuations negligible.

redshift *sample* above $z_{\rm cut}$ (not galaxy, as some galaxy redshift distributions will have tails above $z_{\rm cut}$) must be discarded, or counted as having zero contribution to the sum.

In the derivations in Sections 2.3 and 3.3.3 the apparent magnitude threshold was displayed as an absolute magnitude-dependent limit on redshift, but in order to apply a hard redshift cut to the analysis (if necessary, as discussed in Section 4.3.4), the limit on apparent magnitude can be converted to a (completely equivalent) redshift-dependent limit on absolute magnitude, thus leaving the redshift integrals available for upper or lower limits of $z_{\rm cut}$. For notational convenience, the following shorthand will be used for the rest of this section:

$$f_x(z, M, H_0) = \kappa(z, H_0)p(s|z, I)p(z|I)p(s|M, I)p(M|H_0, I), \tag{4.31}$$

and

$$f_D(z, M, H_0) = p(D_{\rm GW}|z, s, H_0, I)p(s|z, I)p(z|I)p(s|M, I)p(M|H_0, I). \tag{4.32}$$

Looking now at the probability that the host is inside the galaxy catalogue: the host can be considered inside the catalogue if it is both below the redshift cut, and below the redshift-dependent limit on absolute magnitude. Recall $M_{\rm min}$ refers to the brightest end of the luminosity function, and $M_{\rm max}$ refers to the dimmest end. The probability that the host galaxy is inside the galaxy catalogue can therefore be calculated as

$$p(G|D_{\rm GW}, s, H_0, I) = \int\limits_{\Omega\in\Omega_G} p(\Omega|I)\, d\Omega \times \frac{\displaystyle\int_0^{z_{\rm cut}} \int_{M_{\rm min}}^{M(z,m_{\rm th},H_0)} f_D(z, M, H_0)\, dM\, dz}{\displaystyle\int_0^{\infty} \int_{M_{\rm min}}^{M_{\rm max}} f_D(z, M, H_0)\, dM\, dz}, \tag{4.33}$$

where the prefactor (the integral over Ω) is simply the fraction of the sky covered by the galaxy catalogue. The probability that the host galaxy is beyond the galaxy catalogue, $p(B|D_{\rm GW}, s, H_0, I)$, can then be calculated using Eq. 4.26.

The expression for the likelihood of the GW data when the host galaxy is beyond the catalogue is marginally more complicated, because it must now account for not only the galaxies which lie above the redshift-dependent limit on absolute magnitude, but also all galaxies above

the redshift cut. It can be written as

$$\frac{1}{C}p(x_{\rm GW}|B, D_{\rm GW}, s, H_0, I) = \frac{\int\limits_{\Omega\in\Omega_G} p(\Omega|x_{\rm GW}, s, I)\, d\Omega \times \left(\int\limits_0^{z_{\rm cut}} \int\limits_{M(z,m_{\rm th},H_0)}^{M_{\rm max}} f_x(z, M, H_0)\, dM\, dz + \int\limits_{z_{\rm cut}}^{\infty} \int\limits_{M_{\rm min}}^{M_{\rm max}} f_x(z, M, H_0)\, dM\, dz \right)}{\int\limits_{\Omega\in\Omega_G} p(\Omega|I)\, d\Omega \times \left(\int\limits_0^{z_{\rm cut}} \int\limits_{M(z,m_{\rm th},H_0)}^{M_{\rm max}} f_D(z, M, H_0)\, dM\, dz + \int\limits_{z_{\rm cut}}^{\infty} \int\limits_{M_{\rm min}}^{M_{\rm max}} f_D(z, M, H_0)\, dM\, dz \right)}. \tag{4.34}$$

The likelihood of the GW data when the host is outside (the bounds of) the galaxy catalogue is equivalent to the empty catalogue case (see Section 2.3.3), but integrated only over the part of the sky which is not covered by the galaxy catalogue:

$$\frac{1}{C}p(x_{\rm GW}|O, D_{\rm GW}, s, H_0, I) = \frac{\int\limits_{\Omega\in\Omega_O} p(\Omega|x_{\rm GW}, s, I)\, d\Omega \times \int\limits_0^{\infty} \kappa(z, H_0) p(s|z, I) p(z|I)\, dz}{\int\limits_{\Omega\in\Omega_O} p(\Omega|I)\, d\Omega \times \int\limits_0^{\infty} p(D_{\rm GW}|z, s, H_0, I) p(s|z, I) p(z|I)\, dz}. \tag{4.35}$$

The integrals over absolute magnitude and redshift are separable because z and M are no longer linked through the integral limits, and the integrals over M have cancelled in the numerator and denominator.

The probability that the host is outside the galaxy catalogue, $p(O|D_{\rm GW}, s, H_0, I)$ is as defined in Eq. 4.27, *i.e.* the fraction of the sky not covered by the galaxy catalogue, which can be calculated explicitly using

$$p(O|D_{\rm GW}, s, H_0, I) = \int\limits_{\Omega\in\Omega_O} p(\Omega|I)\, d\Omega. \tag{4.36}$$

With the six parts of Eq. 4.24 calculated as laid out in this section, it is now possible to analyse the real GW data summarised in section 4.2 with the real galaxy catalogues described in section 4.3.

4.5 Details of the main analysis

In Eqs. 4.31 and 4.32, $p(s|z, I)p(z|I)$ is the prior on the redshift of potential host galaxies of GW events, taken to be of the form

$$p(s|z, I)p(z|I) \propto \frac{1}{1+z}\frac{dV_{\rm c}(z)}{dz}R(z)\,. \tag{4.37}$$

Here $V_c(z)$ is the comoving volume as a function of redshift and the factor $(1+z)^{-1}$ converts the merger rate from the source frame to the detector frame. The merger rate density may in general be a function of redshift; however here $R(z) =$ constant throughout (other than in Section 4.7, where an alternative redshift-dependent rate model is considered). The prior on the GW sky location $p(\Omega|I)$ is taken to be uniform across the sky. The term $p(M|H_0, I)$ is the prior on absolute magnitudes for all the galaxies in the universe (not just those inside the galaxy catalogue), which is set to follow the Schechter luminosity function:[9]

$$p(M|H_0) \propto 10^{-0.4(\alpha+1)(M-M^*(H_0))} \exp\left[-10^{-0.4(M-M^*(H_0))}\right]. \tag{4.38}$$

Following [112], B-band luminosity function parameters of $\alpha = -1.07$ for the slope of the Schechter function and $M^*(H_0) = -19.7 + 5\log_{10} h$ for its characteristic absolute magnitude (with $h \equiv H_0/100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) are used in conjunction with the GLADE catalogue.[10] For the upper limits of integration over M, the magnitude of the dimmest galaxies is chosen to be $-12.2+5\log_{10} h$ (equivalent to $0.001L^*$, where $L^* = L_0 10^{-0.4M^*}$, and $L_0 = 3.0128\times10^{28}$ W is the zero point luminosity).[11] The integrals are not sensitive to the choice of their lower limits, *i.e.* the magnitudes of the brightest galaxies. When matching the DES catalogue, and its SDSS g-band luminosities, the following parameters are used: $\alpha = -0.89$ and $M^*(H_0) = -19.39 + 5\log_{10} h$ based on [113]. In this case, the upper (fainter) magnitude limit is set to be $-16.1 + 5\log_{10} h$ based on the same paper, which corresponds approximately to $0.05L^*_g$. It is worth noting that this choice is not entirely consistent with the choice of limit for the B-band, as it neglects more of the faint galaxies. However, if luminosity weighting is applied to the analysis, the results become less sensitive to this choice (see section 4.8.3 for a detailed investigation of this point). Luminosity weighting is applied as default, with $p(s|M, I) \propto L(M)$.

For the BBH selection effects and posterior samples reweighting, the source frame mass prior is chosen as a power-law mass distribution with $p(m_1) \propto m_1^{-\alpha}$ and m_2 uniform in its range with $5M_\odot < m_2 < m_1 < 100M_\odot$. The power-law index α, is chosen to be $\alpha = 1.6$ (which is supported by Model B of [91]). For BNSs, a Gaussian mass distribution with a mean of $1.35M_\odot$ and a standard deviation of $0.15M_\odot$ is used [114].[12]

For the GW selection effects, the remaining GW parameters are marginalized over their natural distributions: uniform in the sky, uniform on the sphere for orientation, and uniform in polarisation angle. The time-averaged PSD of detector noise for the corresponding observation

[9] More complex models for $p(M|H_0, I)$ can be used; in fact, it is expected that the luminosity distribution of galaxies will also evolve with redshift [110], as well as to depend on galaxy type and color [111]. However, these considerations are not within the scope of this work and are left for a future investigation.

[10] The absolute magnitude is related to the intrinsic luminosity of a galaxy by the relation, $M - M^* \equiv -2.5\log_{10}(L/L^*)$. The parameter M^* of the Schechter function itself depends on H_0, which is taken into account.

[11] The conversion between absolute magnitude M and luminosity, L, should actually be band-specific (not bolometric, which is the conversion L_0 is valid for). However, within the `gwcosmo` analysis, and in this conversion of the limits, this constant cancels.

[12] While this does not significantly affect the current results, the BNS mass distribution will need to be revisited in light of GW190425 [71], which is potentially a BNS merger with heavier components.

run from [115] is used, meaning that O1 and O2 BBHs are treated with different selection effects. GW selection effects should, in theory, be common for all events within one population of GW binaries (*i.e.* all BBHs should be treated with the same selection effects). However, this requires taking into account the time each detector spent observing at different levels of sensitivity and computing an overall "sensitive volume" for the detector network. The difficulty here is that sensitive volume depends on both the source-frame mass prior of GWs, and a cosmology to convert those to detector-frame. As such, the sensitive volume used for other LIGO and Virgo analyses cannot be used here, and updating it to include H_0 as a changeable parameter is not a trivial analysis. With this in mind, the decision was made to separate the GW selection effects calculated for O1 and O2, such that there were unique BBH selection effects for O1 and O2, and the BNS selection effects assumed O2 sensitivity. The selection criterion is chosen to be a network SNR threshold of 12 for a 2-detector network,[13] which must be found in at least one search pipeline. Section 4.5.1 examines how well the GW selection effects calculated in this way match the detected GW data.

Table 4.1 summarises the galaxy catalogues used to analyse each of the GW detections, along with the number of galaxies in each event's 90% localisation volume. It includes estimates of the probability that the host galaxy is in the catalogue, evaluated at the median redshift for each detection assuming a Planck 2015 cosmology, which is calculated in section 4.5.2.

4.5.1 The impact of population priors and detection threshold choices

In order to confirm that the mass samples re-weighting outlined in section 4.4.1 is carried out correctly, and that the GW selection effects are adequately modelling the detection of the events, the empty-catalogue method is used. This means that no EM information is used – it is as if no attempt to observe the galaxies has been made, and so the redshift prior takes the form of Eq. 4.37. Carrying out the empty catalogue analysis is equivalent to evaluating Eq. 4.35, but where the integrals over Ω cover the full sky, such that they go to 1 and can be ignored, and the integral over z is all that remains. The GW posterior samples are re-weighted using the source frame mass prior described in Section 4.5, before marginalising over it to create the $\kappa(z, H_0)$ term. The detection probability term, $p(D_{\rm GW}|z, s, H_0, I)$ is calculated using the same prior. Three different selection criteria are investigated: network SNR thresholds of 12, 11 and 10.

The empty catalogue analysis is useful for a number of reasons, mainly that it allows the user to estimate contribution of information to the measurement of H_0 that does not come from EM galaxy catalogues. One might expect that, if the prior on redshift is taken to be uniform in comoving volume, that the posterior on H_0 should be entirely uninformative. This is not

[13] This is taken to be representative for the runs as an average, as Virgo did not start observing until near the end of O2. While it played an important role in localising GW170817 on the sky (all 3 detectors being in observing mode when the event occurred) it did not impact the detection of the event, which was loud enough to easily pass the detection threshold without its contribution.

true. While the posterior is significantly *less* informative, there is information about cosmology present in the observed source distribution [69, 70, 72].

Two ways in which this can enter the analysis are: information from the mass prior, and information from the detected distance distribution of events. In the former, sharp features in the source-frame mass distribution are informative, as there are certain cosmologies at which the redshifted masses of detected events will be most consistent with the source-frame prior. Take GW170729, for example. It is the heaviest BBH from O1 and O2, with a primary mass that is, for some values of H_0, higher than the theoretical mass gap around $50M_\odot$ motivated by the pair instability supernova process [116]. So an upper mass limit of $50M_\odot$ in this case would disfavour low values of H_0 as it pushes the masses of detected events to be inconsistent with the prior (see section 4.7.2 for results which investigate this further). The choice of a relatively uninformative power-law mass distribution, with an upper mass limit of $100M_\odot$ (consistent with the masses of all O1 and O2 events for all prior values of H_0) is motivated to ensure the final result on H_0 is not driven by this prior choice.

However, even in the presence of an uninformative mass prior, there is still information to be gleaned from the luminosity distance distribution of detected events. The more distant events (such as GW170729, GW170823, and to a lesser extent GW151012, GW170809 and GW170818) lead to H_0 estimates favouring the lower end of the prior. This is because in a universe where host galaxies are distributed uniformly in comoving volume, for lower values of H_0, the expected distribution of detected GW events favours relatively higher luminosity distances. Thus events at high luminosity distances have more support for smaller values of H_0, while relatively nearby events, namely GW150914, GW151226, GW170608 and GW170814, correspondingly have lower support at smaller values of H_0. The information present in the observed luminosity distance distribution would thereby potentially contribute to the H_0 measurement, independent of, or even in absence of information in galaxy catalogues, if the underlying distributions of source parameters were known, or marginalised over.

The combined contribution from the mass and distance distributions of detected events is captured in the empty catalogue analysis, which can be used to ensure that the prior choices are not being overly informative. This is particularly important at this stage in the analysis, where it is not yet possible to marginalise over, or jointly estimate, the hyper-parameters of the mass prior and merger rate evolution with redshift. Instead the astrophysical population is set to a fiducial distribution, and the analysis is later repeated in section 4.7 with different choices for the mass distribution and binary merger rate evolution with redshift in order to quantify possible systematic effects resulting from this assumption. First, however, the empty catalogue (or "population") contribution is quantified for the fiducial choice.

Figure 4.1 shows the empty catalogue results for three different detection thresholds: a network SNR threshold of 12, 11 and 10 respectively. Individual (normalised) event likelihoods are shown in coloured lines (solid when the event passed the detection threshold, dashed otherwise).

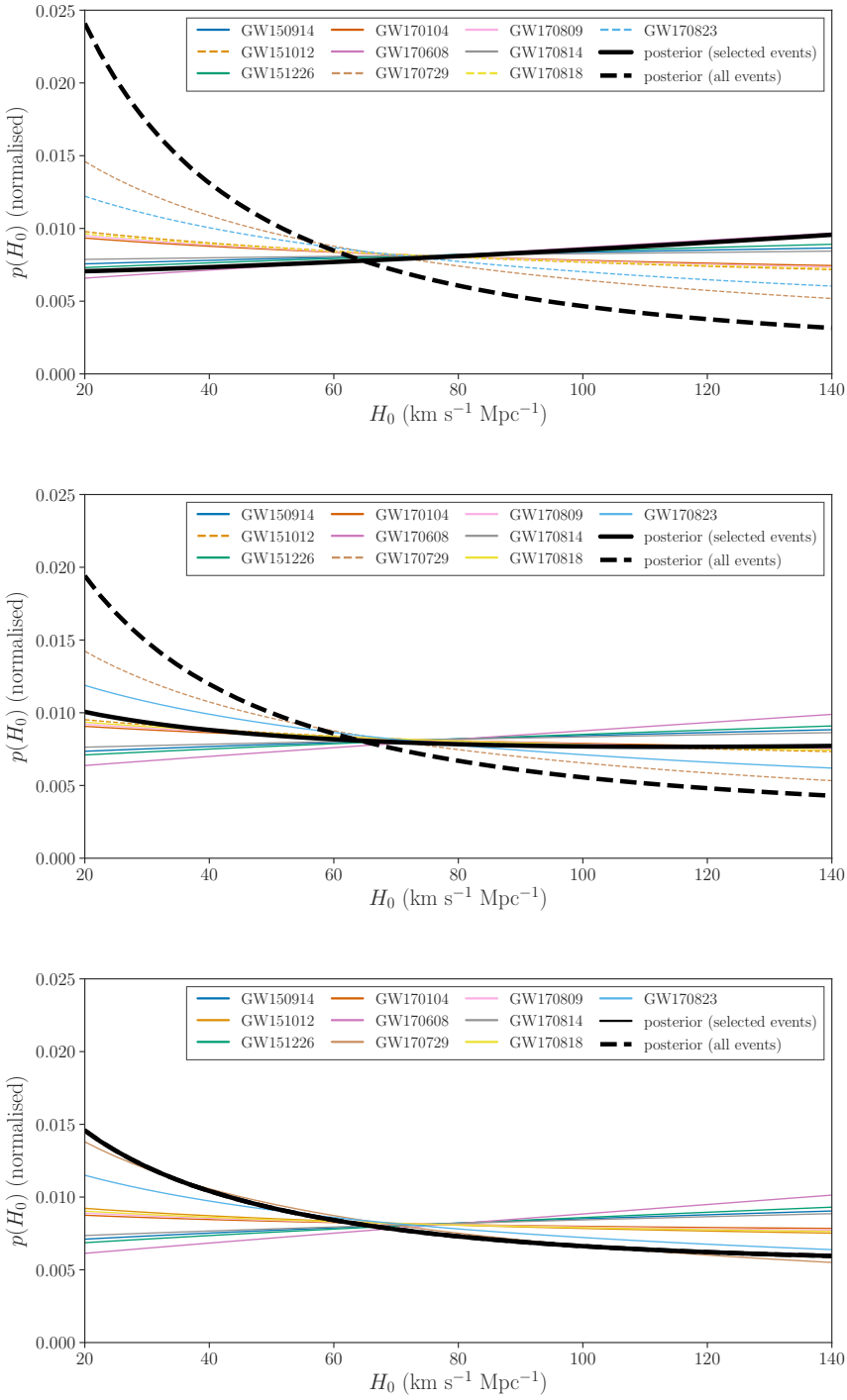

Figure 4.1: Empty catalogue likelihoods for the O1 and O2 BBHs, for a range of network SNR thresholds. The thin coloured lines show individual (normalised) likelihoods. Dashed lines indicate that the event was below the SNR threshold, while solid lines were above it. The thick dashed black line shows the combined posterior for all events (assuming a uniform prior on H_0). The thick solid black line shows the combined posterior for all events which passed the SNR threshold. *Top panel:* a network SNR threshold of 12 is assumed. *Middle panel:* a network SNR threshold of 11 is assumed. *Bottom panel:* a network SNR threshold of 10 is assumed. Here the combined posterior for all events which passed the threshold includes all of the events, and so overlays the black solid line exactly.

The black solid lines show the combined posterior in each case, when the selected events are consistent with the detection threshold, and the dashed black line shows the combined posterior from all events, *i.e.* when the chosen events are mismatched with the GW selection effects.

The most conservative choice, a threshold of 12 (top panel), is consistent with the SNRs of 6 of the BBHs from O1 and O2 (GW150914, GW151226, GW170104, GW170608, GW170809 and GW170814). It is immediately obvious that including all the events, even those inconsistent with the detection threshold, causes a large degree of support for low values of H_0. Discarding events which are inconsistent with the GW selection effects leads to a final posterior which is much less informative, and less biased. That said, the low number of events means that small number statistics play a large role in the shape of the final posterior - it is easy to see that the strong low-H_0 support from the combined posterior of all events is mainly driven by GW170729 and GW170823.

Looking at the SNRs of detected events quoted in Table 4.1, the actual SNRs go down to 10.0, with GW151012 being the quietest event. Lowering the network SNR to 11 (middle panel of Fig. 4.1) allows the inclusion of an addition two events in the final posterior: GW170818 and GW170823. Given the small localisation region of GW170818, which makes it a potentially good candidate for the galaxy catalogue analysis, it is worth considering how plausible it is to use this threshold. The individual likelihoods for every event favour higher values of H_0 than they did when the threshold was set to 12, due purely to the changes in the computed GW selection effects. The posterior which uses all 10 events still favours low H_0, though to a lesser extent than previously. The posterior from events with SNRs above 11 is relatively flat, indicating that no strong information is coming from either the GW population on its own, or the GW selection effects modelled here.

Finally it is worth lowering the threshold even further, to a network SNR of 10, so that all the O1 and O2 detections can be included in the analysis. The results are shown in the bottom panel of Fig. 4.1. Again, there is a slight shift in the likelihoods of each event to prefer slightly higher values of H_0 than before. The final posterior using all events, which are fully consistent with the GW selection effects, still prefers lower values of H_0, but to a lesser extent than the previous cases (the dashed black lines on the upper panels). However, with these results comes a word of warning. By lowering the SNR threshold to this point, there is the possibility that there may be real events with SNRs above the threshold, but for which other detection statistics were not so favourable, *i.e.* they may have a high FAR, making it impossible to determine if the signal is real. In GWTC-1 there are several marginal triggers listed which would pass the SNR threshold of 10, but which were not confident enough to be deemed detections. This mismatch between the real data and the modelled selection effects could lead to a bias in the results. The fact that the empty catalogue posterior for the SNR threshold of 10 still favours low H_0 could be down to small number statistics, or it could be an early indication that there is a mismatch between the data and the model.

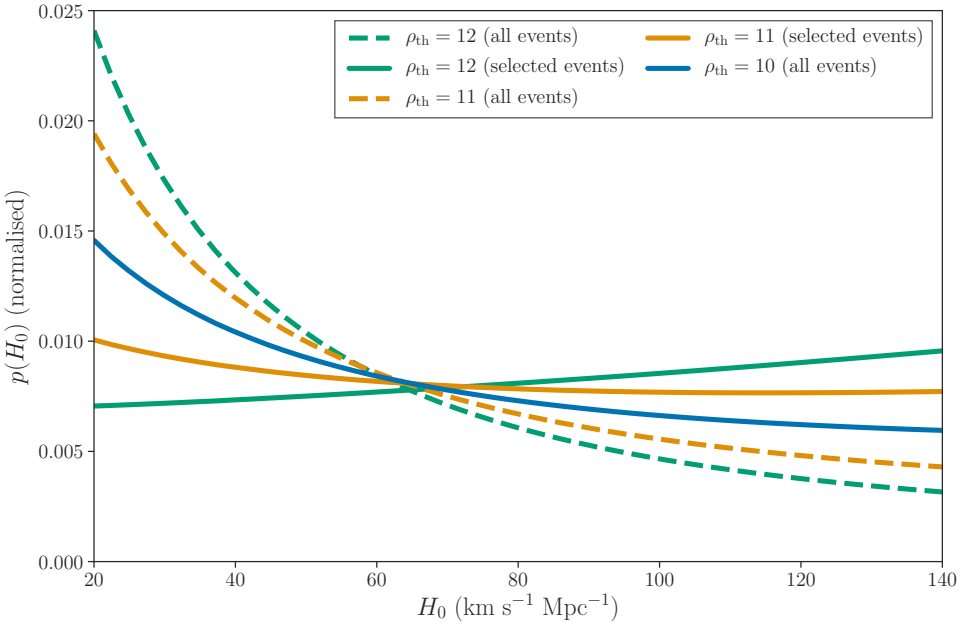

Figure 4.2: Empty catalogue posteriors for the O1 and O2 BBHs, assuming different network SNR thresholds. The green lines correspond to a detection probability calculated assuming a network SNR threshold of 12, while orange and blue correspond to 11 and 10 respectively. Dashed lines indicate the posterior was calculated using all 10 BBHs from O1 and O2, while solid lines indicate that only the events which passed the SNR threshold are included. For $\rho_{\rm th} = 10$, these sets of events are the same.

A final comparison is shown in Fig. 4.2, which compares the empty catalogue posteriors for the three different SNR thresholds. It demonstrates that, when all BBHs from O1 and O2 are used in the analysis, lowering the SNR threshold reduces low H_0 support. However, equally effective (and a sounder choice) is to discard events which fall below the chosen detection threshold, while taking care to set the threshold to be conservative enough that it is unlikely to include real events which were discarded based on other detection statistics.

In conclusion, the choice of SNR threshold to model GW detectability is a trade off: a higher, more conservative threshold is a safer assumption, which can more easily be made to match the real data with little risk. However, it involves discarding quieter, more marginal events, which will involve discarding potentially useful information, especially if it turns out that a particularly well-localised event was only a marginal trigger. There is no one "right way" to make this choice (though modifying the GW selection effects to use a FAR statistic to determine detection would be a nice improvement for the future) – instead the benefits and drawbacks should be weighed up. Here it should be concluded that network SNR thresholds of 11 or 12 is relatively safe, while one of 10 risks introducing bias.

4.5.2 Probability that the host galaxy is in the catalogue

As in Chapter 3, here the completeness of a galaxy catalogue is estimated using an apparent magnitude threshold (limiting magnitude) $m_{\rm th}$, along with an assumption about the underlying redshift and luminosity distributions of galaxies in the universe. Previously $m_{\rm th}$ was assumed to be uniform across the sky – now it is found by calculating the median value from the apparent magnitude distribution of all the galaxies within the sky localization of each event. The choice of median apparent magnitude is likely to be an underestimate of the true limiting magnitude, but, as with SNR threshold, it is safer to be more conservative. An overestimate of $m_{\rm th}$ would lead to the assumption that the catalogue is more complete than it is in reality, which would introduce the bias to the result. By underestimating $m_{\rm th}$, galaxies fainter than this limit can be discarded and the analysis remains self-consistent.[14] For GLADE, estimating $m_{\rm th}$ this way takes into account some of the larger changes in completeness across the sky, which come from it being a composite catalogue, comprised of many surveys of differing depths. Assuming $m_{\rm th}$ is constant across the sky area of an event is revisited both in section 4.8 and, in much greater detail, in Chapter 5. That the completeness of a galaxy catalogue is modelled by a set of limiting magnitude thresholds, can by itself be a non-trivial assumption, especially for photometric catalogues, since galaxies may be missing for various reasons other than them being too faint. This will also need to be revisited in the future in a catalogue-specific manner.

For now, $m_{\rm th}$ is estimated as described above, and show in Fig.4.3 is the probability of a host galaxy being inside the catalogue $p(G|z, D_{\rm GW})$ as a function of redshift z, for each of the galaxy catalogues under consideration. These probabilities are calculated by adapting Eq. 4.33 to be a function of z, along the line-of-sight of the event:

$$p(G|z, D_{\rm GW}, s, H_0, I) = \frac{\int_{M_{\rm min}}^{M(z,m_{\rm th},H_0)} p(D_{\rm GW}|z, s, H_0, I)p(s|z, I)p(z|I)p(s|M, I)p(M|H_0, I)\, dM}{\int_{M_{\rm min}}^{M_{\rm max}} p(D_{\rm GW}|z, s, H_0, I)p(s|z, I)p(z|I)p(s|M, I)p(M|H_0, I)\, dM}. \tag{4.39}$$

For the DES catalogue, Eq. 4.39 would additionally include a Heaviside step function such that $p(G|z, D_{\rm GW}, H_0, I) = 0$ above $z = 0.5$, due to the redshift cut that is applied to the analysis. While still written as a function of H_0 in Eq. 4.39, in practice the curves in Fig. 4.3 are found to be insensitive to the chosen value of H_0.

Figure 4.3 additionally shows the median redshift for each event $z_{\rm event}$ (which *is* H_0-dependent and has been calculated assuming a Planck 2015 cosmology in this case) as the vertical lines. These lines are solid until they intersect the curve of the catalogue they are associated with,

[14] There are more rigorous ways in which the limiting magnitude of a survey can be calculated, but this is left for future work.

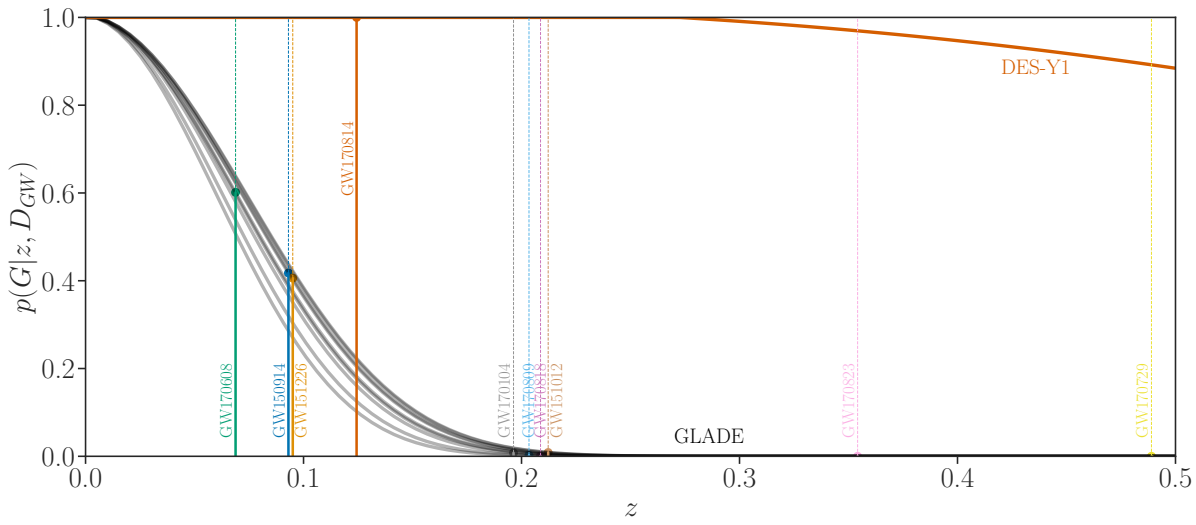

Figure 4.3: The probability that the host galaxy is inside the galaxy catalogue, shown for GLADE (grey curves) and DES-Y1 (orange curve), as a function of redshift. For GLADE this quantity is calculated for each individual event, using the $m_{\rm th}$ estimated within each event's sky localisation. For DES-Y1 the curve is only valid in the patch of sky covering GW170814. Each curve is independent of the value of H_0. The vertical lines show the median redshift (assuming a Planck 2015 cosmology) for each event as in Table 4.1. These lines are thick and solid up to the intercept with the galaxy catalogue they are used with, and thin and dashed above.

and dashed above. The final two columns of Table 4.1 report the $m_{\rm th}$ of the relevant catalogue within the sky localisation of each event, and the probability of the host galaxy being in the catalogue at the median redshift of each event. This final value gives an indication of how informative the galaxy catalogue result might expect to be for each event (caveated by the fact that larger localisation volumes will cause results to be less informative even for a catalogue with a high level of completeness, due to the larger number of potential host galaxies contained within it). It is immediately clear that GW170814 has the highest level of catalogue support with the DES catalogue. For the GLADE catalogue, events of interest are GW150914, GW151226 and GW170608, all of which have above 40% probability of containing the host galaxy (at the median redshift associated with the Planck 2015 value of H_0). The remainder of the events are expected to be relatively uninformative, with likelihoods similar to those shown in the empty catalogue plots (Fig. 4.1) – though for low values of H_0, which correspond to low redshifts for a given luminosity distance, the catalogue may introduce some additional information.

Detailed investigation of DES-Y1 for GW170814

The high completeness of DES-Y1 within the GW170814 sky localisation is apparent from Fig. 4.3. As the result for GW170814 will be driven by the galaxy catalogue information, it is worth examining the catalogue in question in greater detail – especially as this event has been analysed, under slightly different assumptions, in [65]. It is helpful to have an assessment

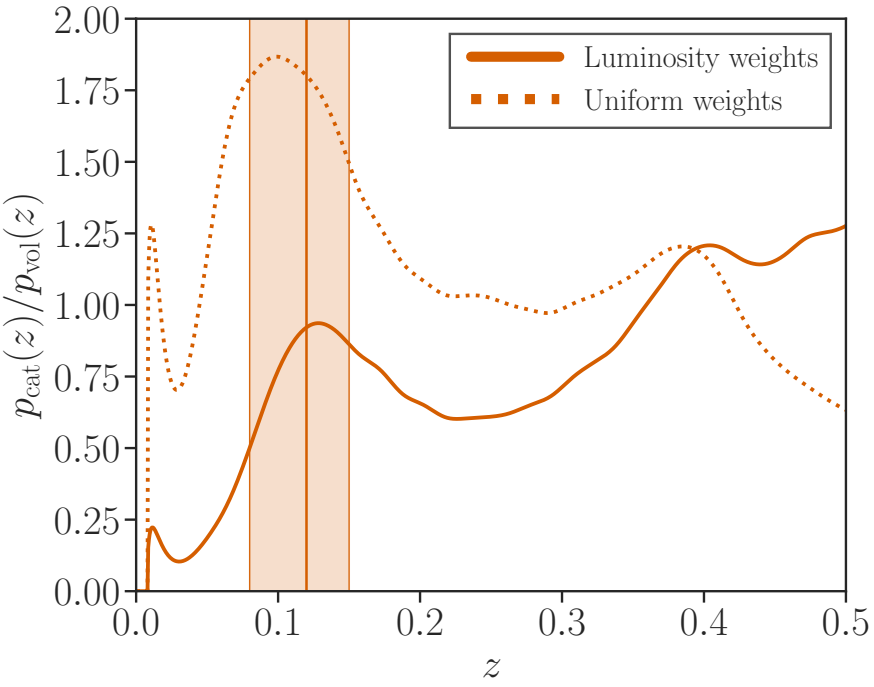

Figure 4.4: Probability distribution for the redshifts of potential host galaxies $p_{\rm cat}(z)$, with redshift uncertainties taken into account, divided by a uniform in comoving volume distribution $p_{\rm vol}(z)$ of galaxies. When computing $p_{\rm cat}(z)$ all galaxies from the DES-Y1 catalogue brighter than $0.05L_g^*$ within GW170814's 99% sky localisation region are included, and each galaxy is weighted proportional to its g-band luminosity (solid line) as well as with uniform weights (dotted line). Also shown is the median and 90% highest probability density interval estimated redshift range for GW170814 (calculated assuming a Planck 2015 cosmology) as the vertical band for reference.

of the contribution from potential host galaxies as a function of redshift for this event. This contribution is quantified using a treatment analogous to that in [63], by computing the ratio $p_{\rm cat}(z)/p_{\rm vol}(z)$ between the probability distribution for the redshifts of potential host galaxies $p_{\rm cat}(z)$ and of a uniform in comoving volume distribution of galaxies $p_{\rm vol}(z)$. When computing $p_{\rm cat}(z)$ all galaxies brighter than $0.05L_g^*$ within the event's 99% sky localization region are included. Specifically, $p_{\rm cat}(z)$ is the redshift distribution of galaxies in the catalogue (marginalised over redshift uncertainties), multiplied by the 2D GW skymap probability, and additionally luminosity weighted. Two cases of luminosity weighting are assumed: the first, where probability of a galaxy hosting a GW merger is proportional to its g-band luminosity (assuming a Planck 2015 cosmology for the required magnitude conversion), and the second of uniform weights.

Figure 4.4 shows the distributions $p_{\rm cat}(z)/p_{\rm vol}(z)$ for the DES-Y1 galaxies within GW170814's sky localisation region, for the redshift range $0 < z < 0.5$. The unweighted curve traces the over/under-density of galaxies, and then falls off at larger redshift due to incompleteness in the catalogue. The luminosity-weighted redshift distribution is driven partially by the overdensity of galaxies at $z \approx 0.4$, and partially by bright high-redshift galaxies. The host galaxy for

GW170814 is more likely to be located near the higher galaxy density regions in the DES-Y1 catalogue – these features in the redshift prior are expected to drive the inferred H_0 posterior. It is worth noting, however, that the choice to exclude galaxies below $0.05L_g^*$ has a significant impact upon the plot. Including lower luminosity galaxies without luminosity weighting causes a large over-density around a redshift of zero and obscures much of the structure seen in the unweighted curve. The luminosity weighted curve does not display the same behaviour. This is important when considering the final result from GW170814 – the informativeness (or not) of this event can be strongly influenced by the choice to apply a strict luminosity cut, if luminosity weighting is not applied. Section 4.8.3 studies this matter in detail.

The features seen here in the DES-Y1 catalogue differ from the over-density in the DES-Y3 data seen in [65] – in particular the redshift of the peak. A part of the difference may be driven by the difference in the photometric redshift estimation algorithms, namely, template fitting methods such as BPZ [107] and machine learning based methods such as the `ANNz2` algorithm [96]. Only the former of the two has been used for the DES-Y1 catalogue and a combination of both for the DES-Y3 catalog. The different selection criteria for choosing galaxies from the two catalogs, such as the stringent redshift cut placed in [65] versus a more relaxed redshift prior used in this work, is another potential source of difference between the corresponding redshift distributions.

4.6 Results

The analysis, as described in Section 4.5, is applied to the O1 and O2 detections to obtain a measurement of H_0 using GW standard sirens. The analysis is carried out with a prior on H_0 uniform in the interval of [20, 140] km s^{-1} Mpc^{-1}; final results are also reported using a flat-in-log prior $p(H_0) \propto H_0^{-1}$ in the same interval for ease of comparison with previous studies. The marginalised distance likelihood and skymaps constructed from the posterior samples of [14] are used. Six BBHs pass the GW selection criteria of a network SNR threshold of 12 in at least one search pipeline (GW150914, GW151226, GW170104, GW170608, GW170809, and GW170814), as well as the BNS GW170817.

The result for the O1 and O2 BBHs is shown in Fig. 4.5. Of the six BBHs, GW150914, GW151226, GW170104, GW170608 and GW170809 are analysed with the GLADE catalogue, while GW170814 is analysed with the DES-Y1 catalogue. Of these events, GW170814's contribution dominates due to the extremely high completeness of the DES catalogue at all relevant redshifts for the event. This highlights the importance of deeper surveys and of dedicated EM follow-up of sky regions following GW triggers for a better H_0 measurement. The GW170814 estimate is qualitatively similar to the result in [65] with analogous peaks in the posterior distribution. The differences in peak locations can be attributed partially to the difference in the redshift distribution for the DES-Y3 catalogue used in [65] versus that for the public DES-Y1

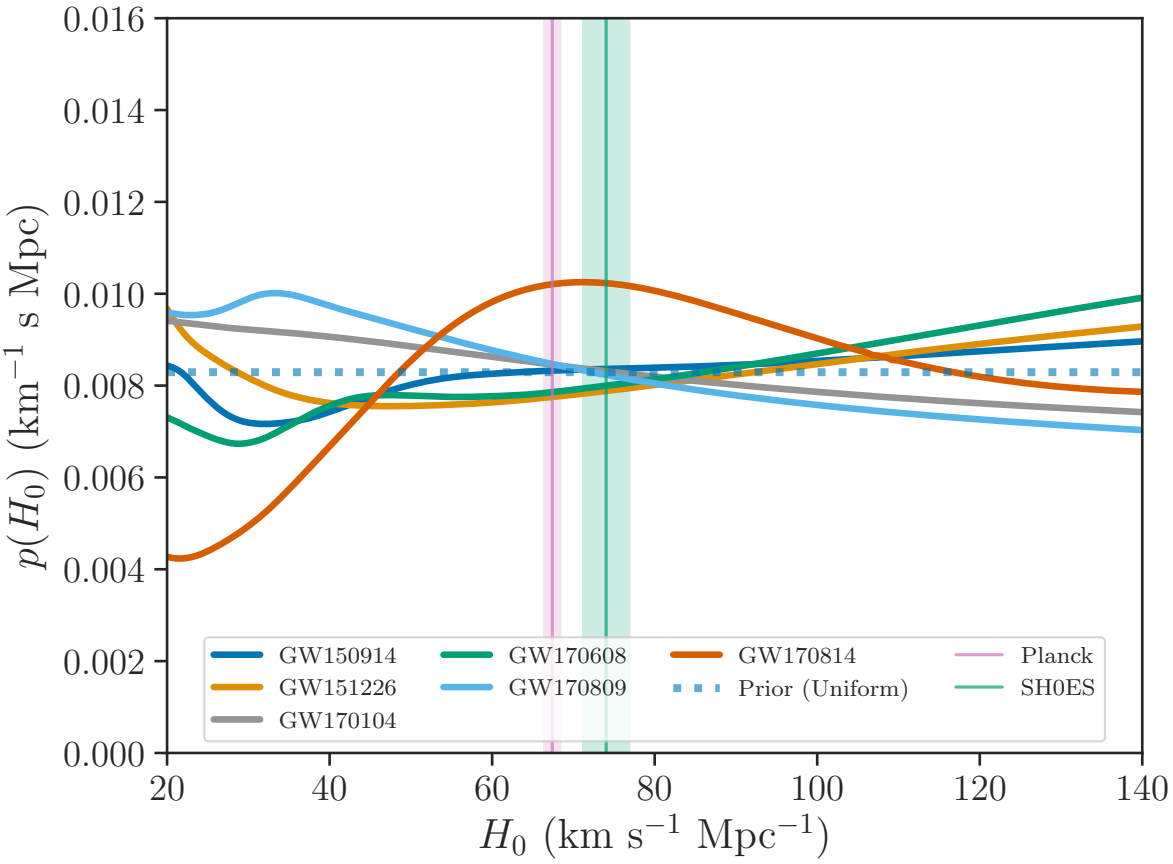

Figure 4.5: Individual estimates of H_0 from the six BBH detections which satisfy the selection criterion of network SNR $>$ 12 in at least one search pipeline. These results assume an $m^{-1.6}$ power-law distribution on masses and a non-evolving rate model. All results assume a prior on H_0 uniform in the interval [20, 140] km s^{-1} Mpc^{-1}(dotted blue). Also shown are the estimates of H_0 from CMB (Planck: [4]) and supernova observations (SH0ES: [30]).

catalogue used in this work. Additionally, the choice of uniform weighting of galaxies, and the application of luminosity and redshift cuts in [65] will also drive the difference.

For the final result the contribution of the BBHs is combined with the result from GW170817 obtained using the low spin prior samples from [14] and an estimated Hubble velocity of $v_H \equiv cz = 3017 \pm 166\,\mathrm{kms}^{-1}$ (where c is the speed of light) for NGC4993 from [13]. The final combined result is shown in Fig. 4.6, with the posterior distribution plotted assuming a uniform H_0 prior. It gives $H_0 = 69.6^{+20.4}_{-8.6}$ km s^{-1} Mpc^{-1} (68.3% highest density posterior interval). To compare with values in the literature, a flat-in-log prior, $p(H_0) \propto H_0^{-1}$, is also used, and gives $H_0 = 68.7^{+17.0}_{-7.8}$ km s^{-1} Mpc^{-1}, which corresponds to an improvement by a factor of 1.04 (about 4%) over the GW170817-only value of $68.7^{+17.5}_{-8.3}$ km s^{-1} Mpc^{-1}. Quoting the median and symmetric 90% credible interval for this measurement (with flat-in-log prior) gives $74.9^{+39.2}_{-13.9}$ km s^{-1} Mpc^{-1}.

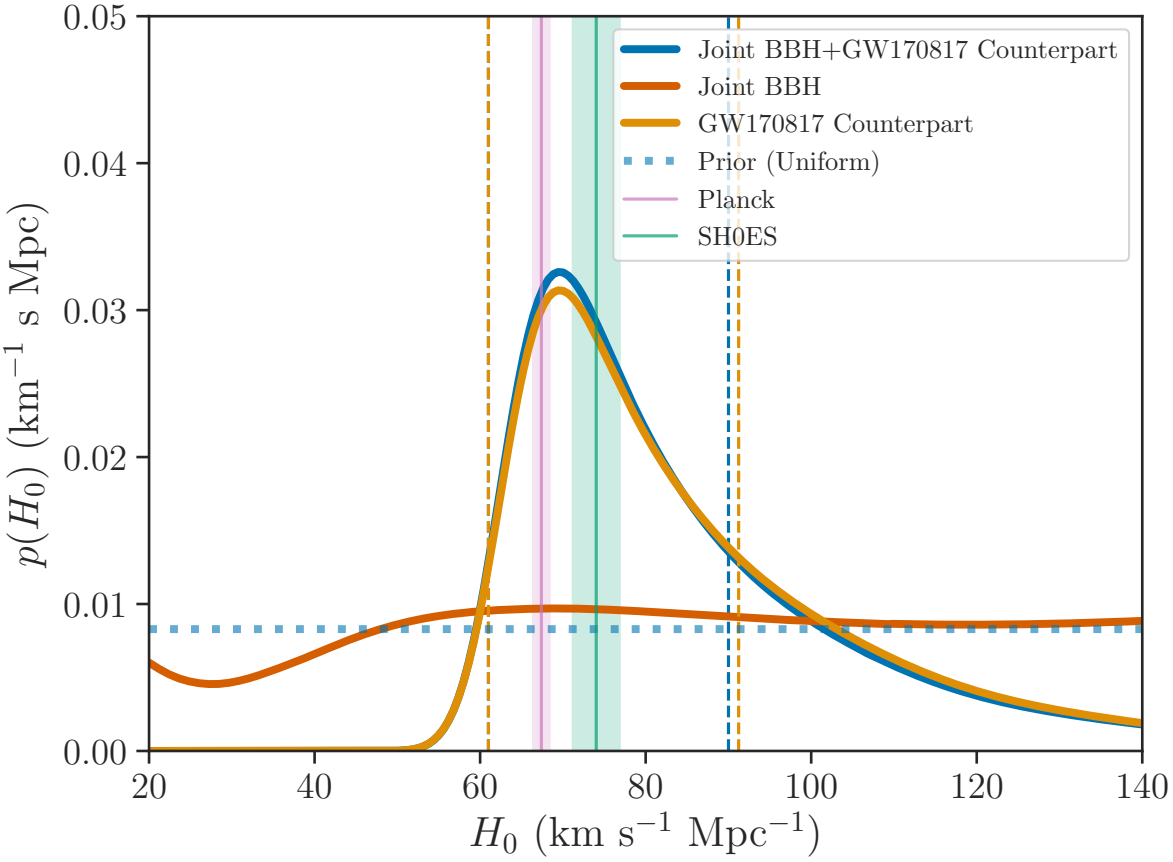

Figure 4.6: The GW measurement of H_0 (dark blue) from the detections in the first two observing runs of Advanced LIGO and Virgo. The GW170817 estimate (orange) is made using the redshift of its host galaxy NGC4993 [13]. The additional contribution comes from BBHs in association with appropriate galaxy catalogues; for GW170814 the DES-Y1 galaxy catalogue is used, while for the remaining five BBHs, GW150914, GW151226, GW170104, GW170608, and GW170809, the GLADE catalogue is used. The 68% maximum a-posteriori intervals are indicated with the vertical dashed lines. All results assume a prior on H_0 uniform in the interval $[20, 140]\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$(dotted blue). Also shown are the estimates of H_0 from CMB (Planck: [4]) and supernova observations (SH0ES: [30]).

4.7 Systematic effects

In this section the analysis is repeated with a few alternative assumptions for the GW selection criterion, the GW population model, and source host properties. This is followed by a discussion of other sources of systematic effects, a detailed study of which is beyond the scope of this work.

4.7.1 Selection criterion

First examining the sensitivity of the results to the GW selection criterion, the threshold on the network SNR is reduced from 12 to 11 in least one search pipeline. The computation of the GW selection effects is adjusted accordingly. Reducing the SNR threshold to 11 introduces two

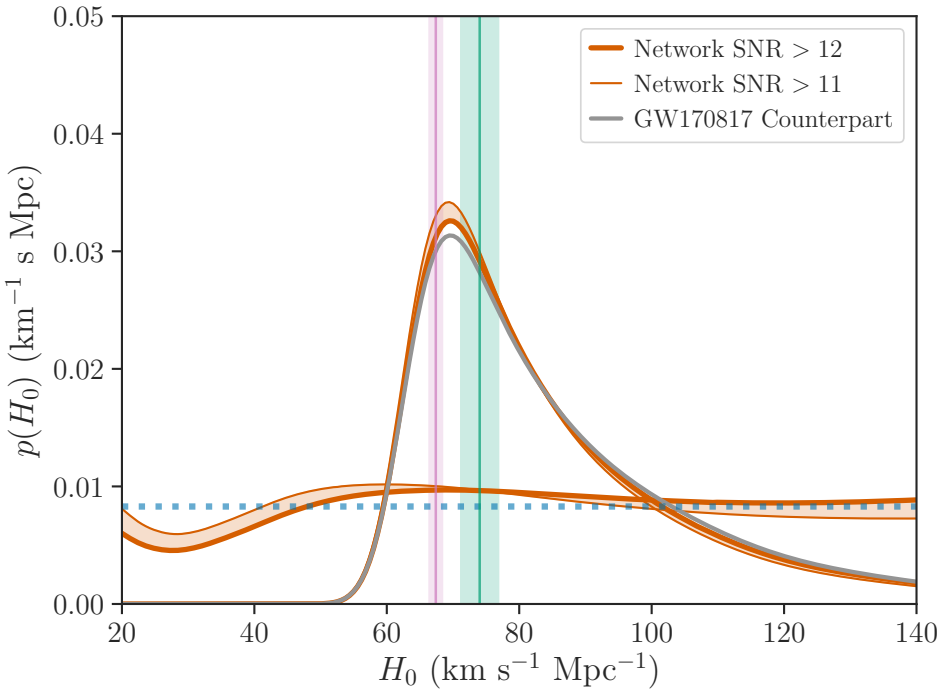

Figure 4.7: Sensitivity of the H_0 posterior to the SNR detection criterion. The final joint result as well as the contribution from all the BBHs which satisfy the selection criterion are shown for a threshold network SNR of 12 (thick orange) and 11 (thin orange); the variation is shown as a shaded band. The GW170817 counterpart result (grey) is added to guide the eye. Six BBHs (GW150914, GW151226, GW170104, GW170608, GW170809, and GW170814) pass the selection criterion with SNR $>$ 12. Two additional BBHs (GW170818 and GW170823) are included with SNR $>$ 11. These results assume the default $m^{-1.6}$ power-law distribution on masses and a non-evolving rate model.

additional events in the analysis, GW170818 and GW170823, neither of which have a significant in-catalogue contribution. Figure 4.7 shows the results with the two sets of assumptions. Differences are expected due to the fact that additional low-SNR events are introduced, and also because the individual likelihoods change slightly with a different SNR threshold used in the GW selection term (as demonstrated in Section 4.5.1). In the regime of a large number of events, these two effects are expected to effectively cancel, provided that the additional low-SNR events do not have significant in-catalogue support which could provide additional constraint on H_0. For this result, the difference is not significant; however the small variation is a reminder that this is still the regime of a low number of events.

4.7.2 Population model

Going back to the default assumption of a network SNR threshold of 12, the next test is on the sensitivity of the results to the chosen GW population model, *i.e.* the mass distribution and the distribution of binary merger rate with redshift. In addition to the power-law mass distribution

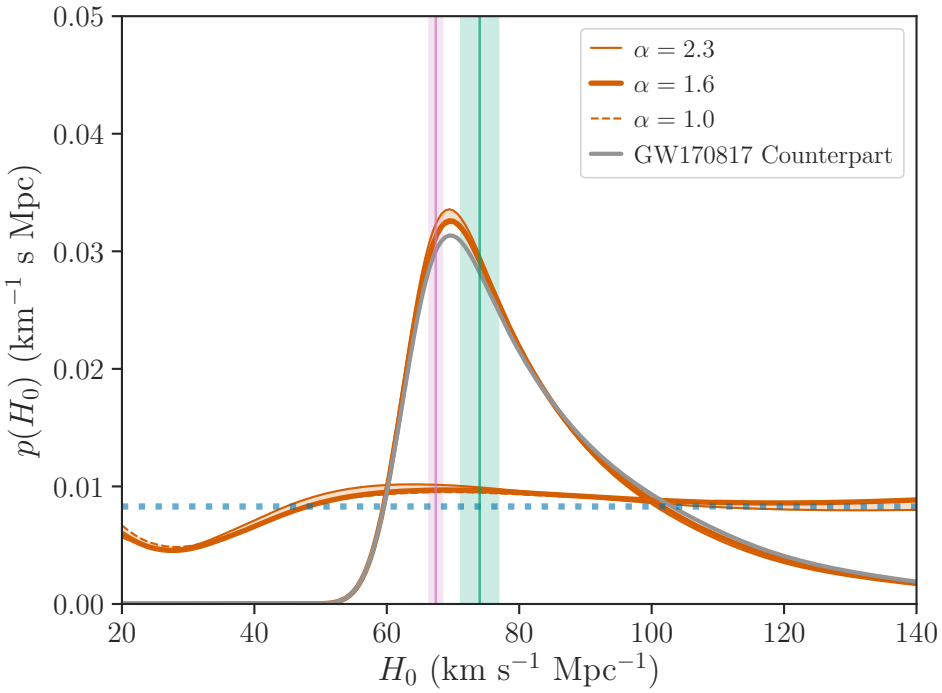

Figure 4.8: Sensitivity of the H_0 posterior to the slope of the assumed mass distribution model. Variation of the results with three different choices of the power-law index for the mass distribution, $\alpha = 1.6$ (thick solid), $\alpha = 2.3$ (thin solid) and $\alpha = 1$ (thin dashed) assuming a constant intrinsic astrophysical merger rate, $R(z) =$ constant and $M_{\rm max} = 100\,M_\odot$.

with $\alpha = 1.6$ (median inferred value using Model B of [91]), a shallower flat-in-log mass distribution with $\alpha = 1$, and a steeper distribution with $\alpha = 2.3$ are chosen (both supported by the inferred range in [91]). Results are shown in Fig. 4.8, and they demonstrate that the systematic differences due to the choice of power-law slope α are insignificant.

Next the robustness of the result to the choice of upper cut-off for the mass distribution, $M_{\rm max}$, is tested. For the default analysis, $M_{\rm max}$ was chosen to be $100\,M_\odot$, consistent with all the considered BBHs for all values of H_0 within the prior range. Reducing this cut-off to a slightly restrictive $M_{\rm max} = 50\,M_\odot$ (motivated by e.g. the pair instability supernova process, [116]), causes a significant difference (Fig. 4.9). Lowering $M_{\rm max}$ corresponds to a closer GW detection horizon. This systematically leads each event to prefer slightly lower values of H_0 than in the main result, for the reasons outlined in Section 4.6 – namely the relationship between the predicted event distribution (from our GW selection effects) and the real detected event distribution.

Finally the assumption of constant binary merger rate with redshift is relaxed. A constant merger rate density, $R(z) =$ constant, assumed that the merger rate did not evolve with redshift. The analysis is repeated using a merger rate $R(z) \propto (1+z)^3$, which traces the star formation rate at low redshifts ($z < 2.5$) [117], as well as a merger rate $R(z) \propto (1+z)^{-3}$ which could arise if typical delay times between binary formation and merger are very long, as may be expected from the chemically homogeneous evolution formation channel [118]. These relaxed assumptions thus

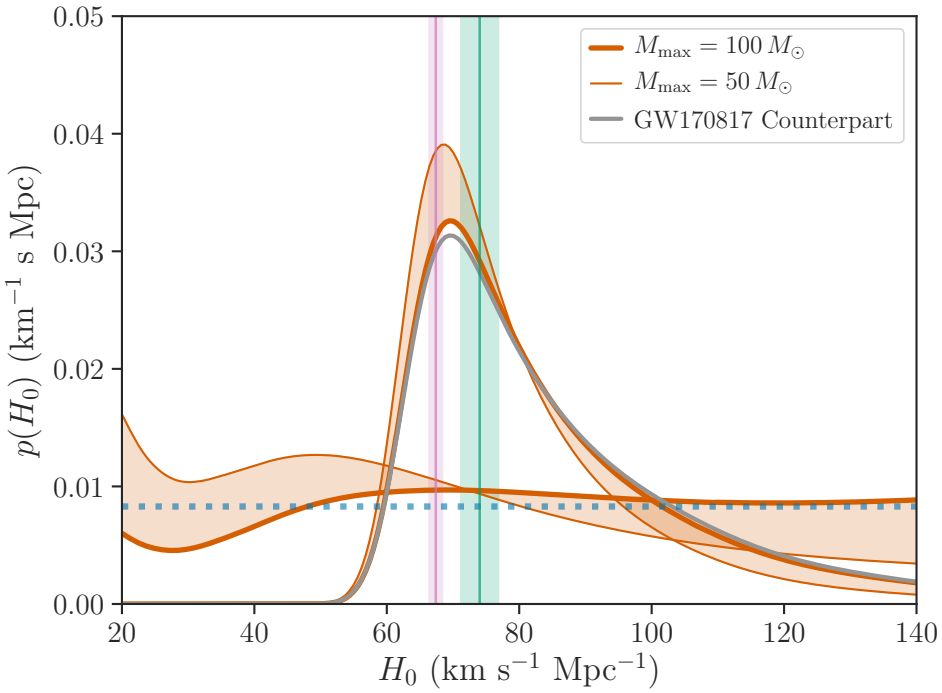

Figure 4.9: Sensitivity of the H_0 posterior to the upper mass limit of the assumed mass distribution model. Variation of the results with two different choices for the allowed black hole maximum mass, $M_{\max} = 100\,M_\odot$ (thick solid), and $M_{\max} = 50\,M_\odot$ (thin solid), both assuming $R(z) =$ constant and $\alpha = 1.6$.

cover a large fraction of physically viable and inferred population models [91]. Results are shown for the different assumed redshift evolution models in Fig. 4.10. The model in which the merger rate traces star formation ($R(z) \propto (1+z)^3$, where no cut off is applied in practice) shows a significant difference, with a tendency to prefer lower values of H_0, compared to the other two models which are quite similar.

4.7.3 Luminosity weighting

The results in section 4.6 assumed a weighting of galaxies by their luminosities in the *B*- or *g*-band (depending on the galaxy catalogue), which are indicative of galaxies' star formation rates. In order to quantify the difference likely to be caused by alternate ways of weighting the galaxies, the analysis is repeated with no luminosity weighting. These results are shown in Fig. 4.11.

With uniform luminosity weights, the joint binary black hole estimate is close to flat (thin orange line in Fig. 4.11). This can be understood as follows: 1) The beyond catalogue term in Eq. 4.24 assumes that galaxies which are not inside the galaxy catalogue are distributed uniformly in comoving volume. When galaxies are unweighted, the probability of the host galaxy being inside the catalogue is reduced (relative to the luminosity-weighted case) meaning more weight is given to the uninformative out-of-catalogue contribution, and the events become less

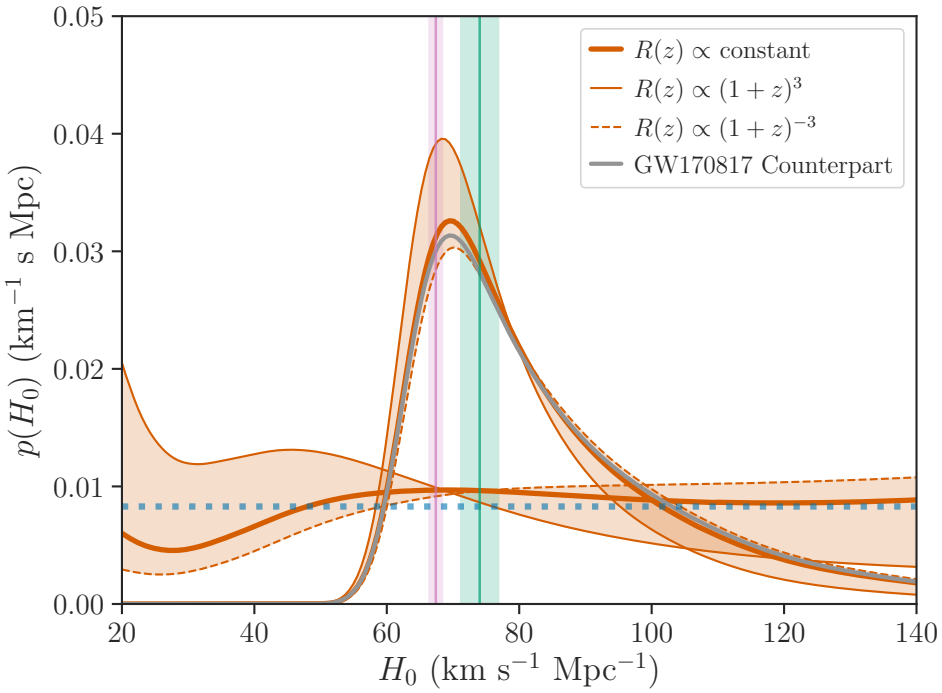

Figure 4.10: Variation of the H_0 posterior with three different choices for the rate evolution, $R(z)$ = constant (thick solid), and $R(z) \propto (1+z)^3$ (thin solid) and $R(z) \propto (1+z)^{-3}$ (thin dashed) for $\alpha = 1.6$ and $M_{max} = 100M_\odot$.

informative. 2) The catalogues used in this analysis contain high numbers of low-luminosity galaxies. When all galaxies are treated as equally likely hosts the overall in-catalogue contribution becomes less informative, leading to a flattening of the final result. This is also in agreement with expectations from [63] and [64], where weighting by luminosities enhance the features in the posterior distribution coming from the galaxy catalogue. The unweighted curves in Fig. 4.11 should not be trusted implicitly, as they are heavily influenced by prior choices on the luminosity function which will be discussed in more detail in Section 4.8.3.

4.7.4 Waveform models

The posterior samples of [14] used for the results in this analysis have been obtained by combining the results of gravitational waveform models which incorporate spin and precession effects to different extents [119–124]. These models are restricted to quasi-circular orbits (i.e., they do not include orbital eccentricity) and neglect higher-order harmonics. Systematic differences in GW parameter estimation results with the employed waveform models constitute only a small fraction of the total uncertainty budget (see, e.g., [125, 126]), and given the large statistical uncertainties, the ignored effects in waveform modelling are not expected to cause a significant difference to the GW measurement of H_0 presented in Section 4.6. However, cumulative systematic effects arising from limitations of waveform models will become increasingly important

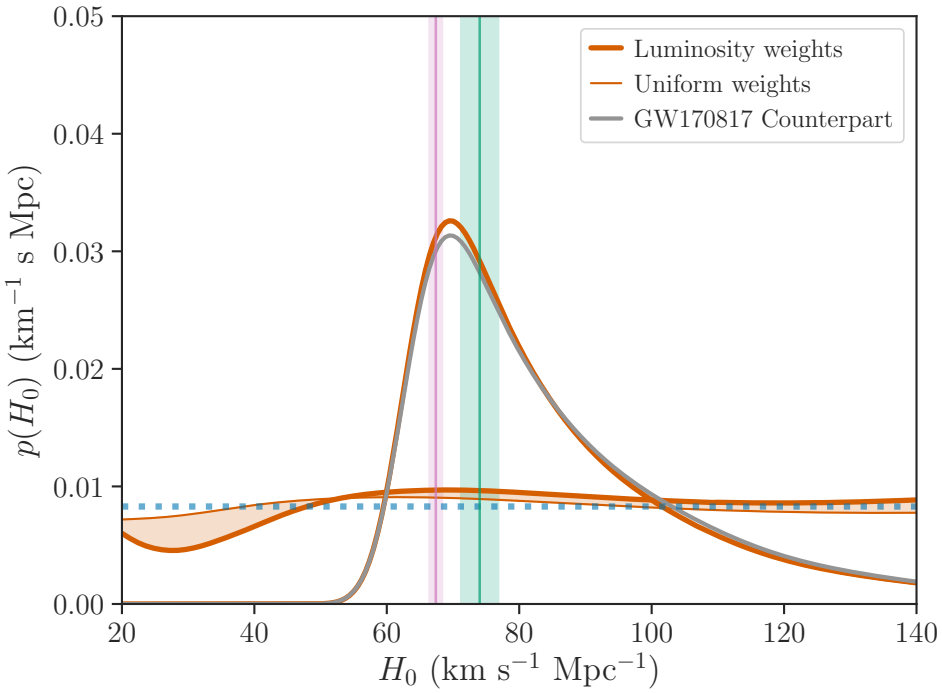

Figure 4.11: Sensitivity of the H_0 posterior to luminosity weighting. Variation of the results when two different kinds of luminosity weighting of host galaxies are considered: B- or g-band luminosity weights (thick solid) as well as constant (uniform) weights (thin solid), both assuming a power-law index for the mass distribution, $\alpha = 1.6$ and constant intrinsic astrophysical merger rate, $R(z)$ = constant.

as the statistical uncertainties become smaller and, in particular, features that can lead to biases in the GW estimation of distance will need to be incorporated.

4.7.5 Detector calibration

An independent effect to be considered is the calibration of the GW detectors. Currently, the GW parameter estimation results are marginalized over the detector calibration uncertainties ($\lesssim$ 4% in amplitude in O1 and O2), which accounts for both the statistical uncertainty and the systematic error correlated between detections [14]. Both the statistical uncertainty and the systematic error in GW detector calibration are much smaller than the other measurement uncertainties [127], and thus negligible for H_0 estimates from a small number of detections. However, the impact of correlated systematic calibration errors between detections will become more important in the long term, with an increasing number of detections driving down the statistical uncertainties, and requiring an improved understanding of other systematic effects that possibly govern the current uncertainty budget.

4.8 Improving upon the O2-H_0 result

While the result presented in Section 4.6 (henceforth referred to as the O2-H_0 result) marks an important point in the quest to measure H_0 using standard sirens, the analysis outlined up to this point is far from perfect. For instance, the assumption that the galaxy catalogue within the sky area of each GW event has an approximately uniform apparent magnitude threshold is one that is clearly untrue if studied in detail. Figure 4.12 shows the galaxy number density of the GLADE version 2.4 catalogue (per pixel, for the full sky gridded up into 49152 equally-sized pixels).[15] While galaxy number per pixel is not a direct measure of the depth to which a survey observed, it stands as a proxy. Galaxy number density varies between 0 and around 1000 galaxies in one pixel. The Milky Way band can be easily identified as the empty band (black) which sweeps across the figure. The rest of the sky breaks down into approximately two different sections (when viewed on a log scale) – areas with O(tens) of galaxies per pixel, and areas with O(hundreds).

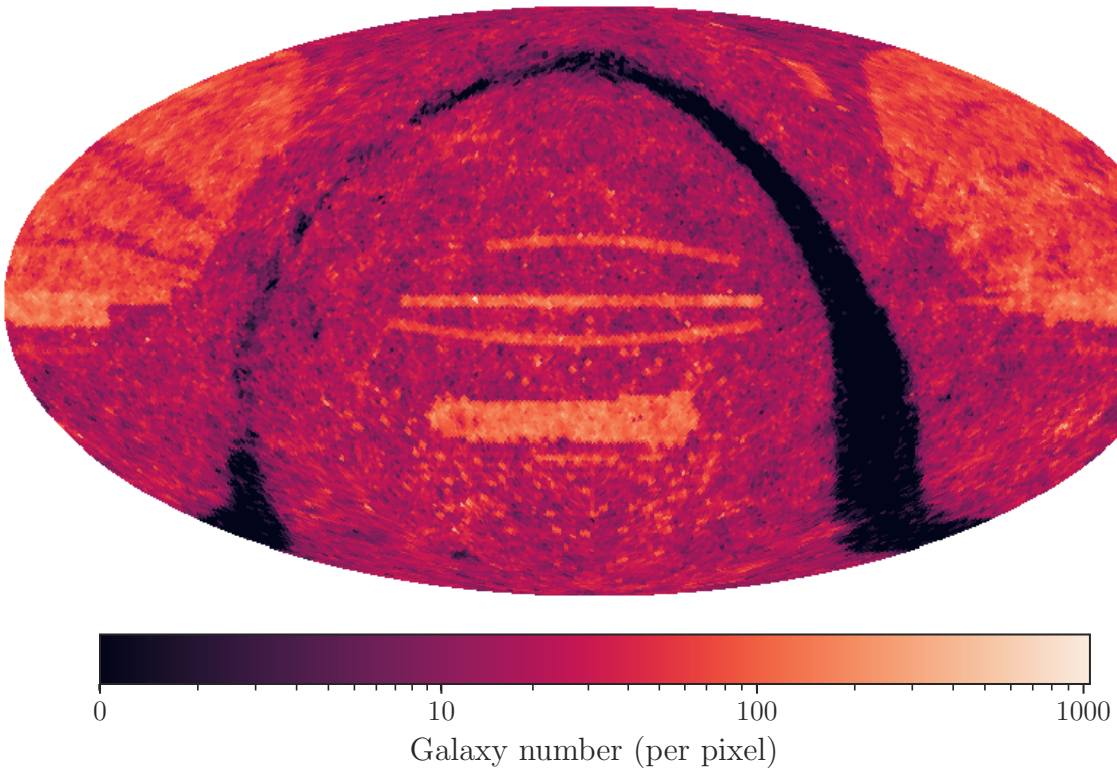

Figure 4.12: Number density of galaxies in the GLADE 2.4 catalogue [80] shown on a `healpy` map with 49152 pixels. Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities.

While it is possible that a GW event may happen to lie with the bulk of its probability in an area of GLADE which is relatively uniform, in practice, given the large sky areas of most O1 and O2 events, this did not happen. Furthermore, as stated in section 4.4.2, for the O2-H_0

[15] Produced using `healpy` [128, 129], available at `http://healpix.sf.net`.

analysis the assumption was made that 100% of the GW sky area was covered by the catalogue (to within computational rounding error). This was calculated in reality by identifying the pixels with non-zero probability from each event's 2D skymap, and using those to define the catalogue on an event-by-event basis. This is shown in the top two panels of Fig. 4.13, which takes GW150914 as an example. The top left panel shows the number density of galaxies within all the pixels covered by the "100%" sky area of GW150914, while the top right panel shows how the sky is separated into "in-catalogue" (Ω_G) and "outside (the bounds of the) catalogue" (Ω_O). For the O2-H_0 analysis, the grey area on the left-hand plot is identical to the black area on the right-hand plot. What is immediately obvious, looking at the top left hand panel, is that the sky area deemed as "in-catalogue" includes a huge range of galaxy densities – from patches of the sky which are entirely empty, to areas which include hundreds of galaxies. To assume that the apparent magnitude threshold of the galaxy catalogue is uniform across this area is a crude approximation at best.

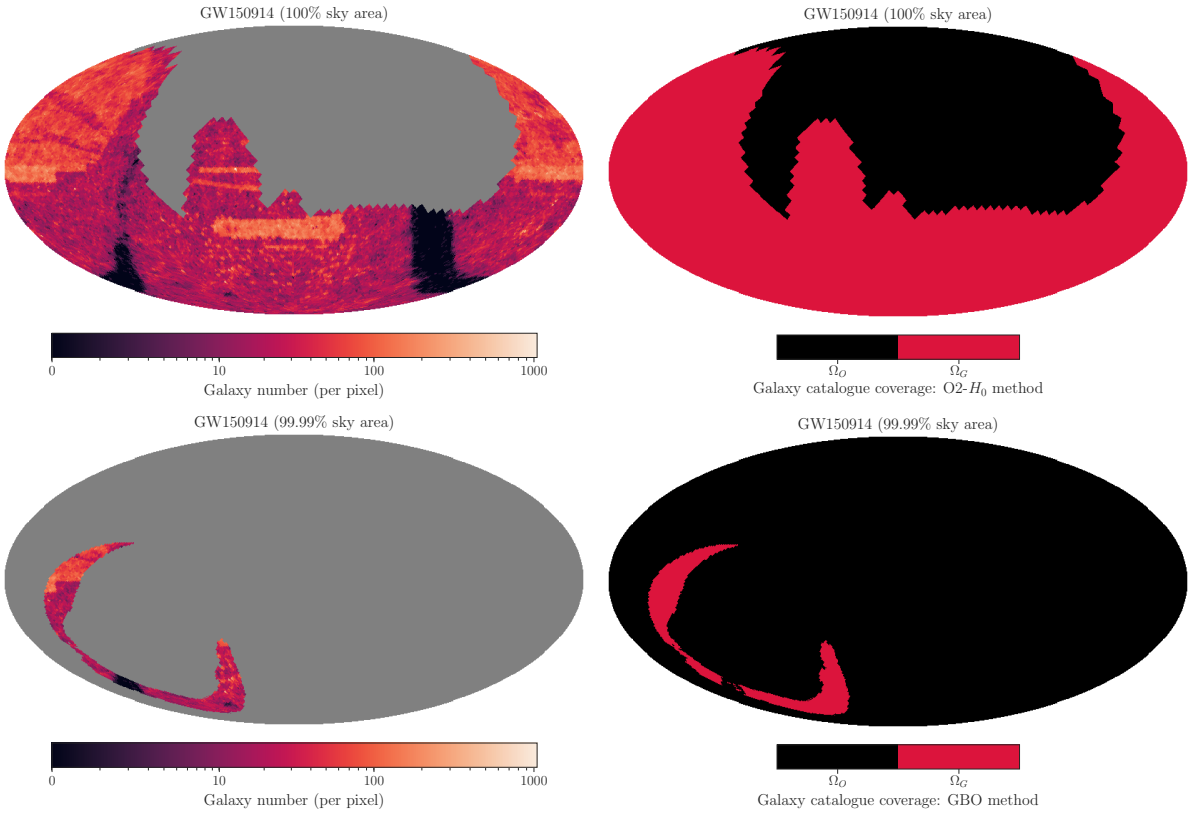

Figure 4.13: A comparison of the galaxy catalogue handling for GW150914 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW150914 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

This approximation seemed necessary due to the framing of the issue in section 4.4.2: either the entire GW event has to be covered by the catalogue (*i.e.* the GLADE catalogue case), or the boundaries of the catalogue have to be identifiable using rectilinear limits in RA and dec, which did not have to include the entirety of the GW sky area (*i.e.* the DES and GW170814 case). With hindsight, it is easy to move away from this separation, and instead use Eq. 4.24 in full to analyse GW data with either catalogue. Including a non-zero $p(x_{\rm GW}|O, D_{\rm GW}, H_0, I)p(O|D_{\rm GW}, H_0, I)$ term simply requires a re-definition of what is in-catalogue and what is not. This time Ω_G can be defined as the intersection of the galaxy catalogue with the $i\%$ GW sky area, where i can be chosen to cover any percentage up to 100. Any patch of sky with GW support, but not within the boundary of the catalogue, is accounted for with the O term. Whatever the choice for the threshold of the sky area, this method will return a mathematically "correct" result – with the caveat that the result will become less informative as the percentage decreases, and tends towards the empty catalogue case when 0% of the GW sky area is said to have catalogue support. This way of applying the method will be referred to as the GBO method henceforth. Not only does this method mean that there no longer has to be a distinction between full-sky galaxy catalogues, and catalogues which only cover a small patch of the sky, it also means that 1) the apparent magnitude threshold for each event can be computed for a much smaller area of the sky, making it more likely to be representative of the real threshold, and 2) if the full-sky catalogue has empty patches (*e.g.* due to the Milky Way band) these areas can be classed as outside the catalogue bounds, and the lack of knowledge of the distribution of galaxies in those areas can be properly accounted for.

Figure 4.14 demonstrates the difference between the O2-H_0 method of defining Ω_G and Ω_O for a full-sky galaxy catalogue, compared to the GBO method of doing so. The top panel shows the division as done in the O2-H_0 case, where Ω_G contains the entirety of the GW sky area, but also contains a patch of sky with no galaxy catalogue coverage due to its obscuration by the Milky Way band. The bottom panel shows the improved division as done in the GBO case, where this empty patch of sky is classed as outside the catalogue bounds. A practical application of the GBO method produces the bottom panels of Fig. 4.13. The 99.99% sky area of GW150914 is considered (shown by the coloured part of the left-hand panel), which isolates a much smaller patch of sky than in the upper panels. Examining the lower panels more closely, it should be noted that the red patch of the right-hand panel does not, this time, correspond to the same area shown on the left-hand panel. Instead, the part of GW150914's sky area which is obscured by the Milky Way band is defined as part of Ω_O. The contribution to the likelihood on H_0 which comes from that part of the sky will be less informative, and rightly so. Equivalent plots for the other O1 and O2 BBHs can be found in Appendix A.

The GBO method works equally well with galaxy catalogues that only cover part of the sky, which may have irregular edges and not cover the full GW sky area (such as DES). This is shown in Fig. 4.15. The top panel shows the division between Ω_G and Ω_O as made in the O2-H_0

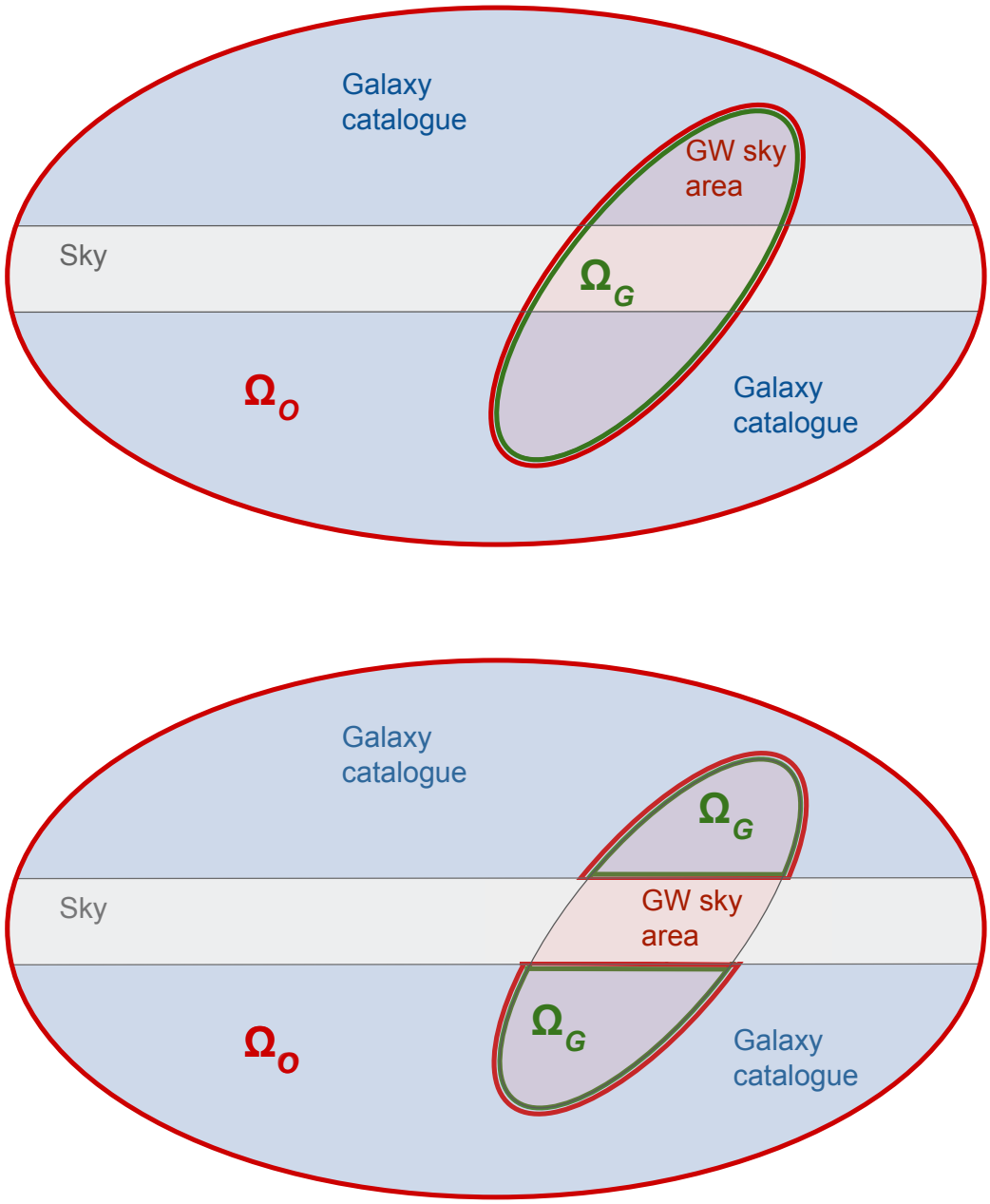

Figure 4.14: Demonstration of how the sky is divided into "in-catalogue-bounds" and "outside-catalogue-bounds" for a galaxy catalogue which covers the full sky in principle, but has areas with little-to-no coverage due to, say, the Milky Way band (*e.g.* the GLADE case). The full sky is shown in grey, and the area covered by the galaxy catalogue is shaded blue – the horizontal grey stripe across the middle is a representation of the Milky Way band. The sky area of the GW data is shaded pink. The part of the sky classed as "in-catalogue-bounds" is outlined in green, and labelled Ω_G, while the part of the sky classed as "outside-catalogue-bounds" is outlined in red and labelled Ω_O. *Top panel:* The division as made for the O2-H_0 analysis. *Bottom panel:* The division as made by the GBO method.

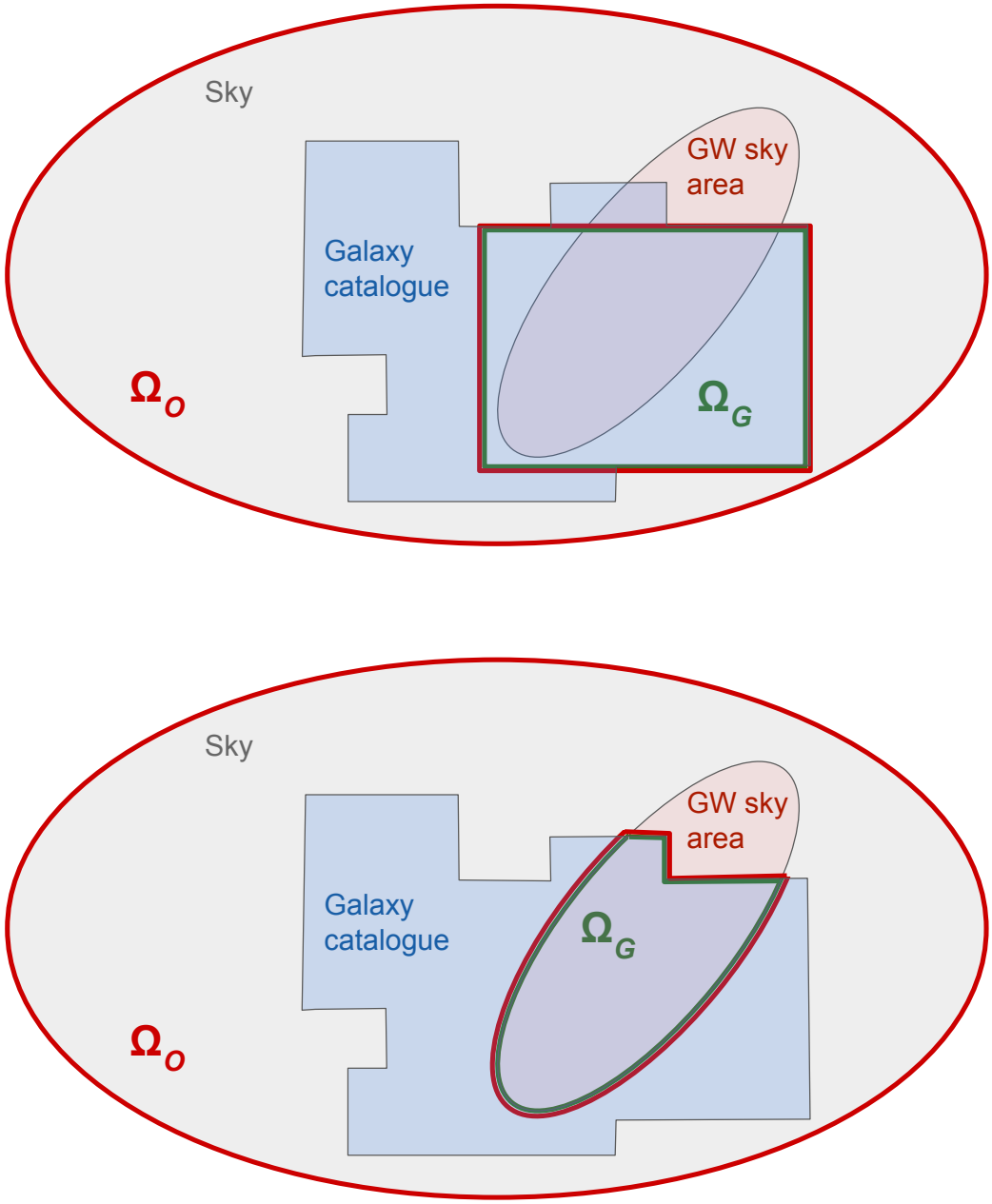

Figure 4.15: Demonstration of how the sky is divided into "in-catalogue-bounds" and "outside-catalogue-bounds" for a galaxy catalogue which only covers part of the sky (*e.g.* the DES case). The full sky is shown in grey, and the area covered by the galaxy catalogue is shaded blue (the Milky Way band is excluded in this case for clarity). The sky area of the GW data is shaded pink. The part of the sky classed as "in-catalogue-bounds" is outlined in green, and labelled Ω_G, while the part of the sky classed as "outside-catalogue-bounds" is outlined in red and labelled Ω_O. *Top panel:* The division as made for the O2-H_0 analysis. *Bottom panel:* The division as made by the GBO method.

analysis. In this case it is defined by rectilinear boundaries on RA and dec, and so does not follow the contour of the GW sky area. In the bottom panel, the GBO method of separation is shown, where Ω_G traces the GW contour until it reaches the hard edge of the galaxy catalogue, at which point that is used to define the edge of Ω_G. It can be seen from Fig. 4.15 that the GBO method allows for more of the galaxy catalogue information to be used, as the boundary of the catalogue overlap with the GW sky area no longer has to be specified by the RA and dec limits. Irregular shapes and patches can be accounted for.

In order to apply the GBO method in practice it is necessary to compute Ω_G, and the amount of GW sky probability contained within Ω_G (that is, $\int_{\Omega_G} p(\Omega|x_{\rm GW})d\Omega$). To do so, the galaxies in the catalogue are split into pixels using `healpy`, as was done for Fig. 4.12, and the number of galaxies which fall in each pixel is counted. The resolution of the pixels is set by changing the parameter "nside", which determines how many equally-sized pixels the full sky is divided into. The GW skymap is downgraded to have the same resolution, and all the pixels which make up the $i\%$ sky area are identified. Of these pixels, those which contain galaxies are defined to be part of Ω_G, and those which are empty are combined with the pixels outside the $i\%$ sky area to define Ω_O. By counting the fraction of pixels which are defined as part of Ω_G, and summing the 2D GW skymap probability contained within them, the integrals over Ω in Eqs. 4.34 and 4.35 are easily calculated.

The choice of nside, and hence the size of each pixel, will have an impact on the calculated values of Ω_G and Ω_O. Too high a resolution will be problematic if the pixels are so small that they can land between galaxies and evaluate as empty. This would artificially inflate the "out-of-catalogue" contribution, when in reality the galaxy survey *did* look in that patch of sky, and just did not see anything. Conversely, too low a resolution will mean that pixels do not accurately capture the hard edges of any empty patches of the galaxy catalogue and lead to an over-estimate of the "in-catalogue" contribution.

4.8.1 Reanalysis of the O1 and O2 BBHs

In order to rerun the O1 and O2 GW events, a choice of nside=64 was made, as it was the highest resolution useable with the GLADE catalogue.[16] For the DES-Y1 catalogue this was increased to nside=256 – this was possible due to the much deeper nature of the survey, meaning that pixels could be much smaller before being found artificially empty. This higher resolution also allowed a relatively accurate approximation of the hard edge of the DES catalogue which passes directly through GW170814's sky area.

The 99.99% sky area was chosen to define the boundary of each event, and the apparent magnitude threshold of the catalogue within that area was calculated. Table 4.2 summarises the main differences between the O2-H_0 analysis and the GBO analysis for each event. The fraction

[16] An nside of 32 was also investigated and would have given similar results, but with slightly less accurate identification of the Milky Way band.

Event name	$\Omega_G/4\pi$		$\int_{\Omega_G} p(\Omega\|x_{\rm GW})d\Omega$		$\langle n_G\rangle$/pixel		$m_{\rm th}$		$p(G\|z_{\rm event}, D_{\rm GW})$	
	O2-H_0	GBO	O2-H_0	GBO	O2-H_0	GBO	O2-H_0	GBO	O2-H_0	GBO
GW150914	0.585	0.039	1.0	0.9778	37.00	36.50	17.92	17.81	0.42	0.38
GW151012	0.327	0.107	1.0	0.9020	39.04	40.58	17.97	17.95	0.01	0.01
GW151226	0.229	0.080	1.0	0.9612	41.28	42.89	17.93	17.92	0.41	0.40
GW170104	0.517	0.074	1.0	0.9999	31.82	37.11	17.76	17.91	0.01	0.01
GW170608	0.392	0.032	1.0	0.9993	33.16	33.34	17.84	17.94	0.60	0.63
GW170729	0.766	0.099	1.0	0.9485	33.06	41.46	17.82	18.08	<0.01	<0.01
GW170809	0.272	0.036	1.0	0.9999	31.57	52.57	17.62	18.29	<0.01	0.04
GW170814	0.005	0.004	0.8754	0.8778	2467	2486	23.84	23.83	>0.99	>0.99
GW170818	0.527	0.016	1.0	0.9958	24.31	17.91	17.51	17.31	<0.01	0.01
GW170823	0.410	0.119	1.0	0.8587	38.81	41.77	17.98	17.94	<0.01	<0.01

Table 4.2: Comparison between the areas defined as in-catalogue for the O2-H_0 analysis and the GBO analysis. The fraction of the sky defined as covered by the galaxy catalogue ($\Omega_G/4\pi$) is given, as is the integrated GW sky probability within this area ($\int_{\Omega_G} p(\Omega|x_{\rm GW})d\Omega$), the average number of galaxies in the pixels included in Ω_G ($\langle n_G\rangle$/pixel), the apparent magnitude threshold estimated for that patch of sky ($m_{\rm th}$), and the probability that the host galaxy is inside the galaxy catalogue, within the patch of sky, for the median redshift of each event.

of the sky defined as in-catalogue, Ω_G, reduces dramatically between the O2-H_0 analysis and the GBO analysis (unsurprisingly). While the 99.99% sky area of each event is under consideration with the GBO method, this does not mean that 99.99% of the GW sky area has galaxy catalogue support. After identifying which pixels do not have catalogue support, and hence are defined as part of Ω_O, the GW sky probability contained within Ω_G (that is, $\int_{\Omega_G} p(\Omega|x_{\rm GW})d\Omega$) is less than 99.99% for most events (with exceptions for GW170104 and GW170809 which are particularly well covered). Most noticeable is GW170823, which has catalogue support for just 86% of its sky probability, due to a significant portion of its sky area being obscured by the Milky Way band.

In order to assess how much the switch from the O2-H_0 method to the GBO method affects the estimate of completeness within the sky area of the event, it is worth considering two quantities: $m_{\rm th}$, the approximate apparent magnitude threshold within Ω_G, and $\langle n_G\rangle$, the average number of galaxies per pixel within Ω_G. For comparison it is worth noting that for the whole GLADE catalogue these numbers are estimated as $m_{\rm th} = 17.81$ and $\langle n_G\rangle = 32.81$ per pixel. In both the O2-H_0 and the GBO method, $m_{\rm th}$ is estimated as the median apparent magnitude of all the galaxies within the footprint of Ω_G, and as such is not impacted by the redefinition of the Milky Way band to be out-of-catalogue. The galaxy number density, $\langle n_G\rangle$, is impacted by this, however, and the majority of events have an increased value in the GBO version as empty pixels

are not being included in the average.

The apparent magnitude thresholds calculated using the GBO method vary between being higher and lower than those from the O2-H_0 analysis. Noticeable improvements can be seen for GW170729 and (more dramatically) GW170809, where the difference is ~ 0.7. The final columns of Table 4.2 state the probability that a host galaxy would be inside the galaxy catalogue at the median redshift of each event (assuming Planck 2015 cosmology), $p(G|z_{\rm event}, D_{\rm GW})$. These were calculated using the apparent magnitude thresholds also quoted in the table, and comparing the values between O2-H_0 and GBO shows that the difference is not particularly dramatic for any event.

4.8.2 Results

The individual event contributions to H_0 for all 10 BBHs are shown in Fig. 4.16, where the blue curves show the previous (O2-H_0) result, and the orange curves show the new (GBO) result. The changes between the two are not particularly significant, but in general show the expected trends: events with very little in-catalogue support (such as GW170729 and GW170823) show almost no change in their contribution. Events which have high in-catalogue probability and relatively smaller sky areas, such as GW150914 and GW170608 show a more noticeable change.[17] The other event of interest is GW170809, which shows some difference between the blue and organge curves. GW170809's change is mostly driven by the increased weight given to the "in-catalogue" contribution, due to the large change in $m_{\rm th}$ estimated within its sky area (see Table 4.2).

The final posterior on H_0 can be seen in Fig. 4.17. It is calculated using the same selection of events as the O2-H_0 result is section 4.6 (that is, GW150914, GW151226, GW170104, GW170608 and GW170809 analysed with the GLADE catalogue, GW170814 analysed with the DES Y1 catalogue, and GW170817 analysed with its EM counterpart). Identical prior choices are made, including the luminosity weighting of host galaxies, and non-evolving merger rates. Overall, the GBO method gives a slight improvement over the O2-H_0 result, with $H_0 = 68.8^{+16.7}_{-7.8}$ km s^{-1} Mpc^{-1}using a flat-in-log prior, compared to $68.7^{+17.0}_{-7.8}$ km s^{-1} Mpc^{-1}from the O2-H_0 analysis. This represents an improvement of 1-2%.

4.8.3 In support of luminosity weighting

The minor change seen to GW170814's likelihood in Fig. 4.16 is due to two further improvements to `gwcosmo` that were made between the O2-H_0 and GBO analyses, rather than the differences outlined in Section 4.8, the magnitude threshold already being fairly uniform over the

[17] While GW170814 falls into the category of "high in-catalogue probability" and "small sky area" it is not mentioned explicitly here, because the changes to its posterior do not come from the changes to the method described in Section 4.8. Additional improvements were made to `gwcosmo` between the O2-H_0 and GBO analyses and this is discussed in Section 4.8.3.

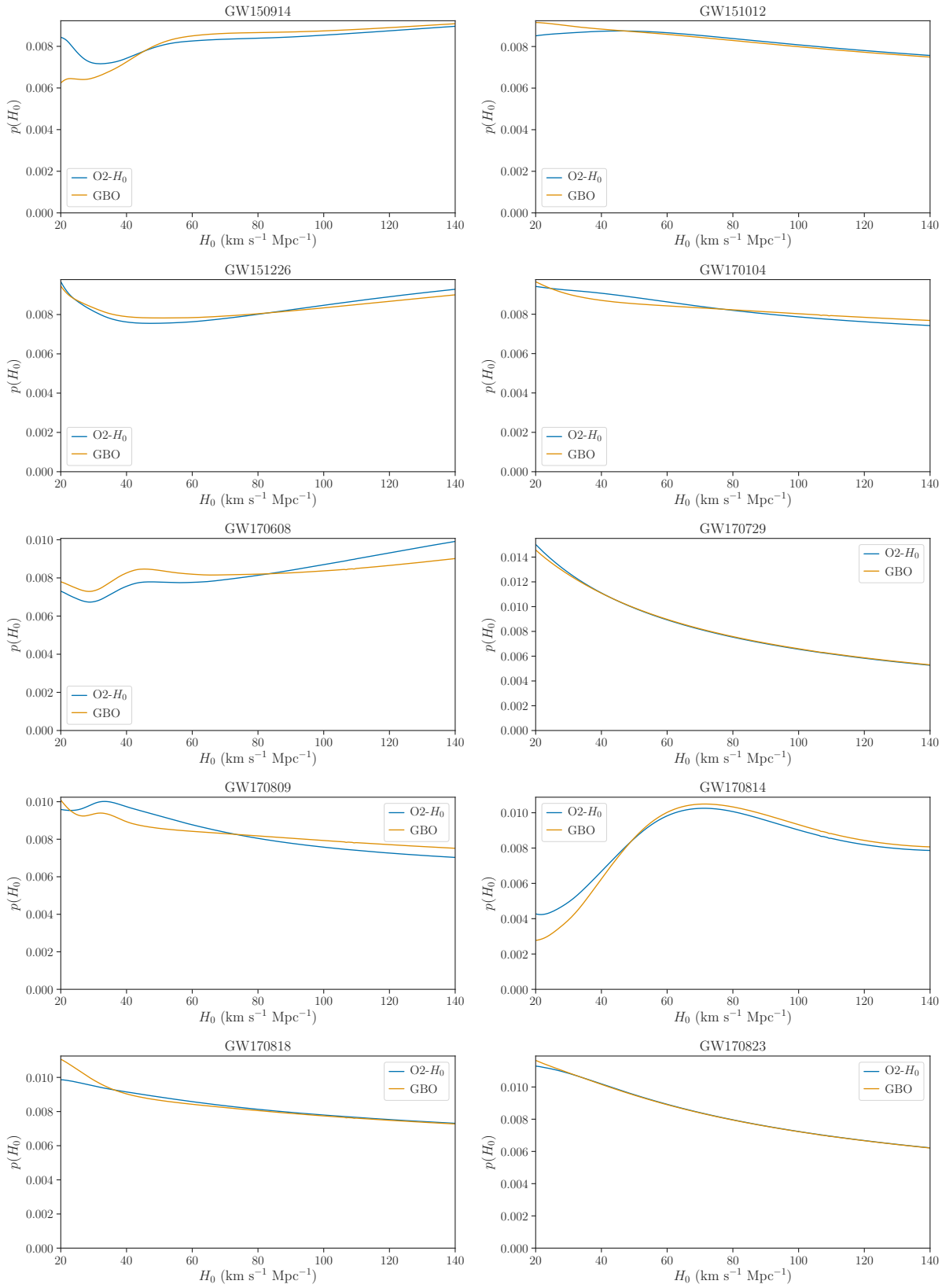

Figure 4.16: A comparison of the individual GW event likelihoods on H_0 when analysed using the O2-H_0 method vs the GBO method. Each panel shows the normalised likelihoods on H_0 for one of the ten BBH events from O1 and O2. The likelihoods from the O2-H_0 analysis are in blue, while the likelihoods from the GBO analysis are in orange.

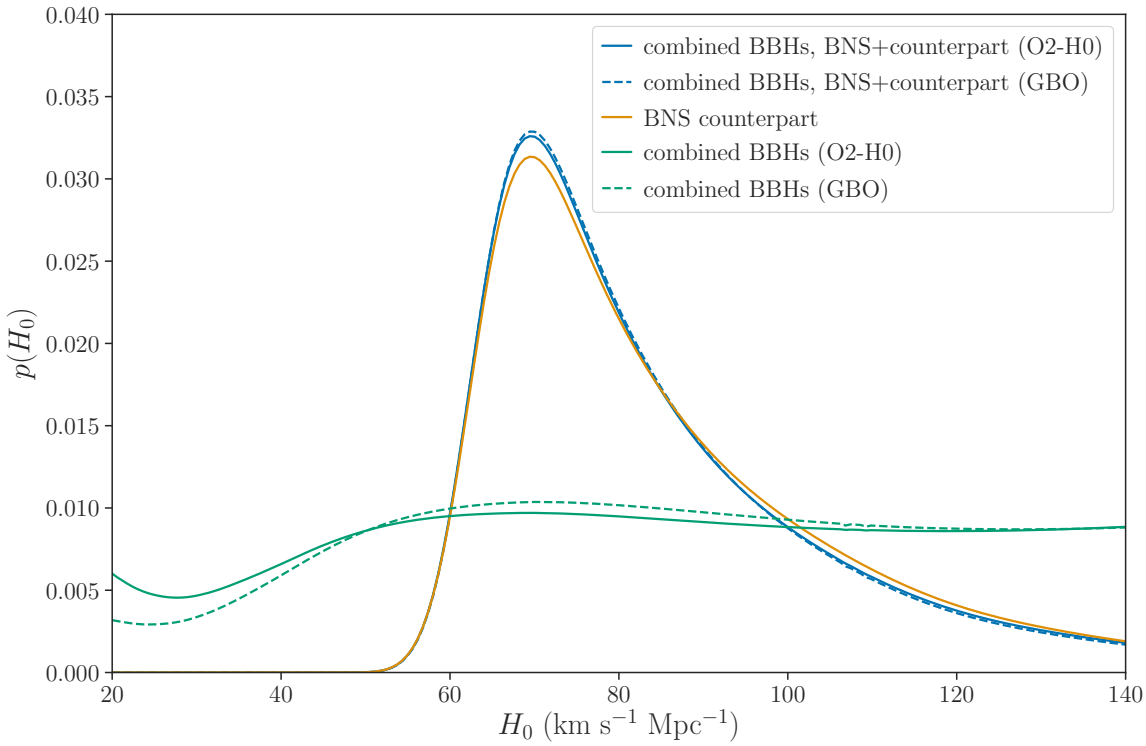

Figure 4.17: Comparison of the final posterior on H_0 for the O2-H_0 result and the GBO result. The solid green line shows the O2-H_0 contribution from the six selected BBHs. The green dashed line shows the same contribution from the GBO analysis. The orange curve shows the contribution from GW170817 and its counterpart. The solid blue line shows the final combined O2-H_0 result, while the dashed blue line shows the final result using the GBO analysis.

patch of DES considered. In order to ensure that galaxy catalogue data used for the in-catalogue part of the computation matched the model for computing the out-of-catalogue parts, two cuts were applied to the galaxies in the catalogue. The first was to explicitly discard galaxies above the apparent magnitude threshold, which don't contribute strongly to analysis in most cases, as they tend to be faint (and disfavoured by luminosity weighting) or far away (such that the out-of-catalogue contribution dominates at those redshifts). The second cut is the discarding of galaxies from the catalogue that are fainter than the low luminosity cut-off assumed for the Schechter function. Again, as these galaxies are faint, they are unlikely to contribute much to the final result. The slight change to GW170814's likelihood in Fig. 4.16 reflects both of these changes.

The main O2-H_0 and GBO results used luminosity weighting by default. However, the option of uniform weighting was considered in Section 4.7.3, where it made the results less informative. The GBO analysis was rerun for GW170814, for both the weighted and unweighted case, while varying the lower luminosity limit of the Schechter function between $0.001L^*$ (equivalent to the limit used for the GLADE catalogue in the O2-H_0 analysis), and $0.1L^*$ (stricter than the $0.05L^*$ cut that was assumed for the DES catalogue in the O2-H_0 analysis). Results for this can be seen in Fig. 4.18. The left-hand panel shows the luminosity-weighted likelihood on H_0,

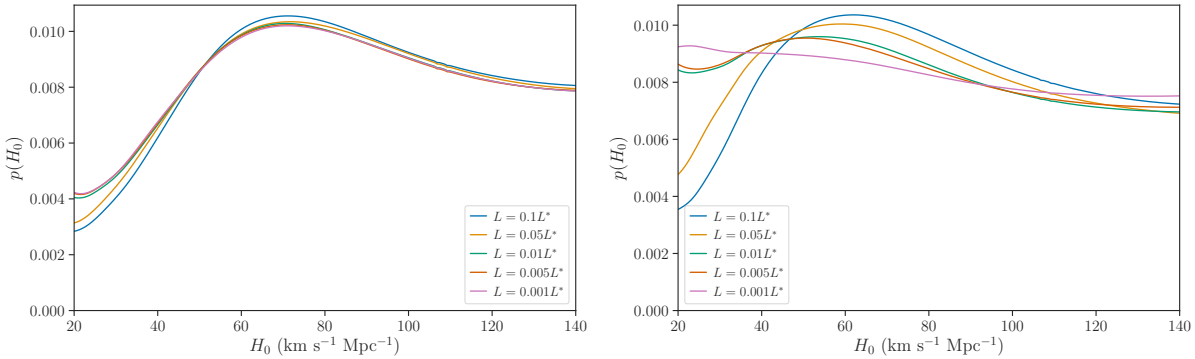

Figure 4.18: Comparison of the normalised likelihood on H_0 for GW170814 analysed with the DES-Y1 catalogue assuming different lower luminosity limits for the Schechter function. *Left-hand panel:* Probability of a galaxy hosting a GW merger is proportional to g-band luminosity. *Right-hand panel:* Probability of a galaxy hosting a GW merger is uniform for all galaxies.

which varies slightly with a changing luminosity limit. It becomes marginally more informative when a stricter cut is applied, but the overall effect is small, and the position of the peak remains the same. The right-hand plot shows the result when uniform weighting is applied. It is clear that the shape of the likelihood in this case is highly driven by the choice of luminosity limit, becoming as informative as the luminosity weighted case for a stricter cut, and becoming increasingly uninformative as the luminosity limit decreases. Not only that, the value of H_0 at which the posterior peaks changes as the luminosity limit changes.

It is worth noting that the unweighted result presented in Fig. 4.11 includes a less informative contribution from GW170814 than would have been produced using the GBO analysis, as the discarding of faint galaxies plays a much larger role in the uniform weighting case (and galaxies fainter than the dim end of the Schechter function were not discarded in the O2-H_0 analysis). If anything, this underlines the take-home point: choosing not to luminosity weight potential host galaxies is not an uninformative choice. Previous analyses have justified its use claiming that it encapsulates the lack of current knowledge of how GW mergers trace galaxy luminosity. This is not true, as uniform weighting is, at its heart, just a binary weighting where some galaxies are considered as being incapable of hosting a GW source. There is nothing uninformative about that.

4.9 Conclusion

Measuring H_0 with real GW data and real galaxy catalogues introduces a host of new challenges for the dark siren analysis.[18] This chapter introduced the GW detections from Advanced LIGO and Virgo's first and second observing runs. The H_0 analysis of the BNS and its EM counterpart carried out in this chapter is broadly the same as that carried out in [13]. However, in order to

[18] Pun intended.

analyse the BBHs using the "galaxy catalogue" method, appropriate galaxy catalogues needed to be identified. The GLADE catalogue was chosen for its full-sky coverage, and the DES-Y1 catalogue was chosen for its high completeness across the sky area of GW170814. The main result of this chapter – presented in section 4.6 and equivalent to that published in [66] – combined the counterpart analysis of GW170817 with the dark siren analysis of six BBHs which satisfied the detection criterion of a network SNR > 12. GW170817 dominated the result, with a modest improvement coming from the BBH contribution, driven largely by GW170814 due to the high completeness of the DES catalogue within its sky area.

The updates made to `gwcosmo` in order to allow it to handle real GW data were centred around removing the detector-frame mass prior that was used for event parameter estimation, and applying, then marginalising over, the source-frame mass prior desired for the cosmological analysis. This was derived explicitly in Section 4.4.1. Future complications regarding the handling of GW data will come from a higher level – the choice of population priors, such as the source-frame mass distribution and the merger rate evolution with redshift. It is clear, looking at section 4.7, that these choices can have a large impact on H_0 inference. Current estimates for the source-frame mass distribution of BBHs which use the real GW detections (as opposed to those based on astrophysical assumptions) make cosmological assumptions in order to do so, and so cannot then be used in a cosmological analysis without biasing the result. While it is possible to jointly estimate the hyper-parameters of the mass distribution and rate evolution simultaneously with cosmological parameter [130], the current methods cannot also use EM information in the form of incomplete galaxy catalogues. An estimate of H_0 which either jointly estimates these hyper-parameters, or marginalises over them, whilst also making full use of the EM information available, will be one of the next great challenges in this field.

The use of real galaxy catalogues in the dark siren analysis introduced a different set of challenges. It was immediately obvious that the mock catalogues generated in Chapter 3 were not representative of many of the challenges posed by real galaxy catalogues. Some of these issues were addressable in the pre-processing of the galaxy catalogues, but the major challenge – that of a galaxy catalogue which does not have uniform completeness across the full sky – required a re-think of the way the methodology was implemented within `gwcosmo`. The initial way this was handled (by taking two approaches: applying rectilinear limits in RA and dec to the analysis with the DES catalogue, and treating each patch of GLADE covered by a GW event as a separate catalogue) was rather simplistic. The first was not problematic in itself, as the DES catalogue is uniformly complete within the chosen boundaries, but the second approach is fundamentally flawed. Not only did it assume the variations in depth between the surveys that made up GLADE were non-existent, it also applied the same logic to completely empty patches of the sky.

Some improvements to `gwcosmo`, described in Section 4.8, removed the necessity of choosing between these two different (but very similar) analyses. The GLADE catalogue and DES

catalogue could now be treated using the same analysis, and any empty patches of sky would be evaluated as such. This improvement also opens up the possibility of using galaxy catalogues with hard edges within the GW sky area, without having to throw away the additional information by cutting them to a specific shape. While this was a significant improvement over the earlier analysis, one major sticking point remains: that of assuming the $m_{\rm th}$ of the catalogue is uniform across the patch of sky with GW support. This, however, shall be addressed in the next chapter.

Chapter 5

A pixelated approach

5.1 Introduction

The aim of this thesis is to present a methodology, alongside a practical implementation, for constraining H_0 using multiple detections of dark standard sirens in combination with (potentially incomplete) galaxy catalogues. So far, the methodology outlined in Chapter 2 has been implemented into a Python software package, `gwcosmo`, which was rigorously tested using a series of MDAs in Chapter 3, including on its ability to handle catalogues which do not contain the host galaxy of the GW merger. Having passed each of these tests, `gwcosmo` was used in Chapter 4 to analyse the GW detections from Advanced LIGO and Virgo's first and second observing runs, which required several additional improvements to address the challenges which arose when using real GW data and real galaxy catalogues. At this point, two clear limitations remain in the latest implementation of the method, which was described in Section 4.8. These are:

1. the assumption that the apparent magnitude threshold of a galaxy catalogue is uniform within the sky area of an event (excepting areas which are entirely empty of catalogue information).

2. the approximation that sky and distance information from a GW event can be separated, as outlined in Section 3.3.1, rather than treating it as fully 3D.

The first of these limitations is problematic, as taking an average $m_{\rm th}$ to represent a variable patch of the sky means that there will be areas for which the real threshold is over-estimated, and areas for which it is under-estimated. Where it is over-estimated, more weight will be given to the "in-catalogue" contribution than should be, which risks biasing the result if the host galaxy is not inside the galaxy catalogue. Where it is under-estimated, useful information about the redshift distribution of galaxies at higher redshifts will be diluted unnecessarily by the uniform "out-of-catalogue" contribution.

The second limitation, that of separating GW sky and distance information, did not introduce any noticeable bias when tested in the MDAs of Chapter 3. However, making this approximation requires discarding some useful GW information, meaning that the results produced by `gwcosmo` so far are less informative than they could, or should, be.

Fortunately there is an extension to the method presented thus far which addresses both of these limitations at once. This extension entails pixelating the sky into equally sized pieces, and analysing each one independently, using a line-of-sight distance distribution for the GW event within each pixel, and an $m_{\rm th}$ calculated using only the galaxies in that section of the sky. Both of the approximations described above are removed, making the analysis more robust and, theoretically, more informative.

Section 5.2 outlines the Bayesian implementation of the pixelated method. Section 5.3 discusses the practicalities of implementing it and applying it to real data. Section 5.4 reanalyses the O1 and O2 detections using the pixelated method, while keeping all other assumptions the same as in the O2-H_0 analysis, and quantifies the improvement to the results. Section 5.5 discusses potential extensions for the pixelated method, and concludes this chapter.

5.2 Methodology

In order to take into account the fact that galaxy catalogues have varying levels of completeness across the sky (see Section 4.8), consider a method in which the sky is split into $N_{\rm pix}$ equally-sized pieces, which are later summed. In this case, the likelihood for a single GW event (which was the starting point for the initial galaxy catalogue methodology in Section 2.2.1) is given by

$$\begin{aligned} p(x_{\rm GW}|D_{\rm GW}, H_0, I) &= \sum_i^{N_{\rm pix}} p(x_{\rm GW}|\Omega_i, D_{\rm GW}, H_0, I) p(\Omega_i|D_{\rm GW}, H_0, I), \\ &= \sum_i^{N_{\rm pix}} p(x_{\rm GW}|\Omega_i, D_{\rm GW}, H_0, I) \frac{p(D_{\rm GW}|\Omega_i, H_0, I) p(\Omega_i|H_0, I)}{p(D_{\rm GW}|H_0, I)}. \end{aligned} \tag{5.1}$$

If the probability of detection is assumed to be uniform over the sky, as it was in previous chapters (a reasonable assumption, as it is averaged over the full length of an observing run, and so the rotation of the Earth blurs out much of the sky-dependence), then $p(D_{\rm GW}|\Omega_i, H_0, I)$ loses its dependence on Ω_i and cancels with the denominator. This is not a necessary approximation to make as the variability of $D_{\rm GW}$ over the sky could be easily taken into account by simply not making the approximation, and following the maths through to the end.[1] Acknowledging that

[1] This would, however, require changing the way that the probability of detection as a function of redshift is calculated (see Section 3.3.2) in order to include sky-dependence at that point, rather than pre-marginalising over RA and dec.

the other term in the numerator, $p(\Omega_i|H_0, I)$, is also independent of H_0 gives the following:

$$\begin{aligned} p(x_{\rm GW}|D_{\rm GW}, H_0, I) &= \sum_i^{N_{\rm pix}} p(x_{\rm GW}|\Omega_i, D_{\rm GW}, H_0, I) p(\Omega_i|I), \\ &= \frac{1}{N_{\rm pix}} \sum_i^{N_{\rm pix}} p(x_{\rm GW}|\Omega_i, D_{\rm GW}, H_0, I), \end{aligned} \tag{5.2}$$

where the equally-sized pixels mean that $p(\Omega_i|I)$ becomes a constant corresponding to the fraction of the surface area of a sphere which one pixel covers. It is worth noting that any pixels with zero GW support will evaluate to zero, and so the sum over $N_{\rm pix}$ can be reduced to a sum over $N_{\rm GWpix}$ without impacting the final result.

The similarity between this expression and the starting point of the original methodology, in Eq. 2.2, is clear. At its simplest, the pixelated method is simply the original method carried out along the line-of-sight of each pixel, then summed. For each pixel it is necessary to marginalise over the possibility that the host galaxy is in, or outside, the galaxy catalogue. Looking specifically at $p(x_{\rm GW}|\Omega_i, D_{\rm GW}, H_0, I)$ and expanding as in Eq. 2.4 leads to

$$\begin{aligned} p(x_{\rm GW}|\Omega_i, D_{\rm GW}, H_0, I) = p(x_{\rm GW}|\Omega_i, &G, D_{\rm GW}, H_0, I) p(G|\Omega_i, D_{\rm GW}, H_0, I) \\ &+ p(x_{\rm GW}|\Omega_i, \bar{G}, D_{\rm GW}, H_0, I) p(\bar{G}|\Omega_i, D_{\rm GW}, H_0, I). \end{aligned} \tag{5.3}$$

The value of $m_{\rm th}$, which determines the probability that galaxies are inside (G) or outside ($\bar{G}$) the catalogue, can now be chosen on a by-pixel basis, allowing varying catalogue incompleteness to be accurately taken into account, modulo the choice of pixel size. There is also the possibility that some pixels will be empty (contain no galaxies due to *e.g.* obscuration by the Milky Way band), in which case Eq. 5.3 simplifies to the empty catalogue case (see Section 2.3.3).

Additionally, for each pixel $p(x_{\rm GW}|\Omega_i, H_0, I)$ can be approximated as the GW information corresponding to the patch of sky covered by pixel i, meaning there is no longer a requirement to separate GW sky and distance information. The distance posterior along the line-of-sight of each pixel can be estimated, making this method inherently "3D".

5.3 Pixelating `gwcosmo`: practicalities

In order to implement the pixel-based method described in Section 5.2, `healpy`,[2] a Python package to handle pixelated data on a sphere [128, 129], is used. `healpy` allows the user to split the sky into equally sized pixels, for a choice of resolutions, where the resolution is set by a parameter called nside. The total number of pixels the sky is divided into is set by $12 \times \text{nside}^2$, where nside is a power of 2. At its lowest resolution, the sky is divided into 12 pixels of equal area. An incremental increase in resolution divides each pixel into 4 further pixels.

[2] `http://healpix.sf.net`

In `gwcosmo`'s case, `healpy` allows data points with known RA and dec (*e.g.* GW posterior samples, or galaxies) to be uniquely associated to a specific pixel. As the pixel boundaries do not move when increasing or decreasing resolution, this opens up the possibility of combining different resolution `healpy` maps within the same analysis, without the danger of double-counting data. This is useful because variations in the line-of-sight distance estimate for a GW across its sky area, and variations in the $m_{\rm th}$ of a galaxy catalogue across the same area, are not (necessarily) on the same scale.

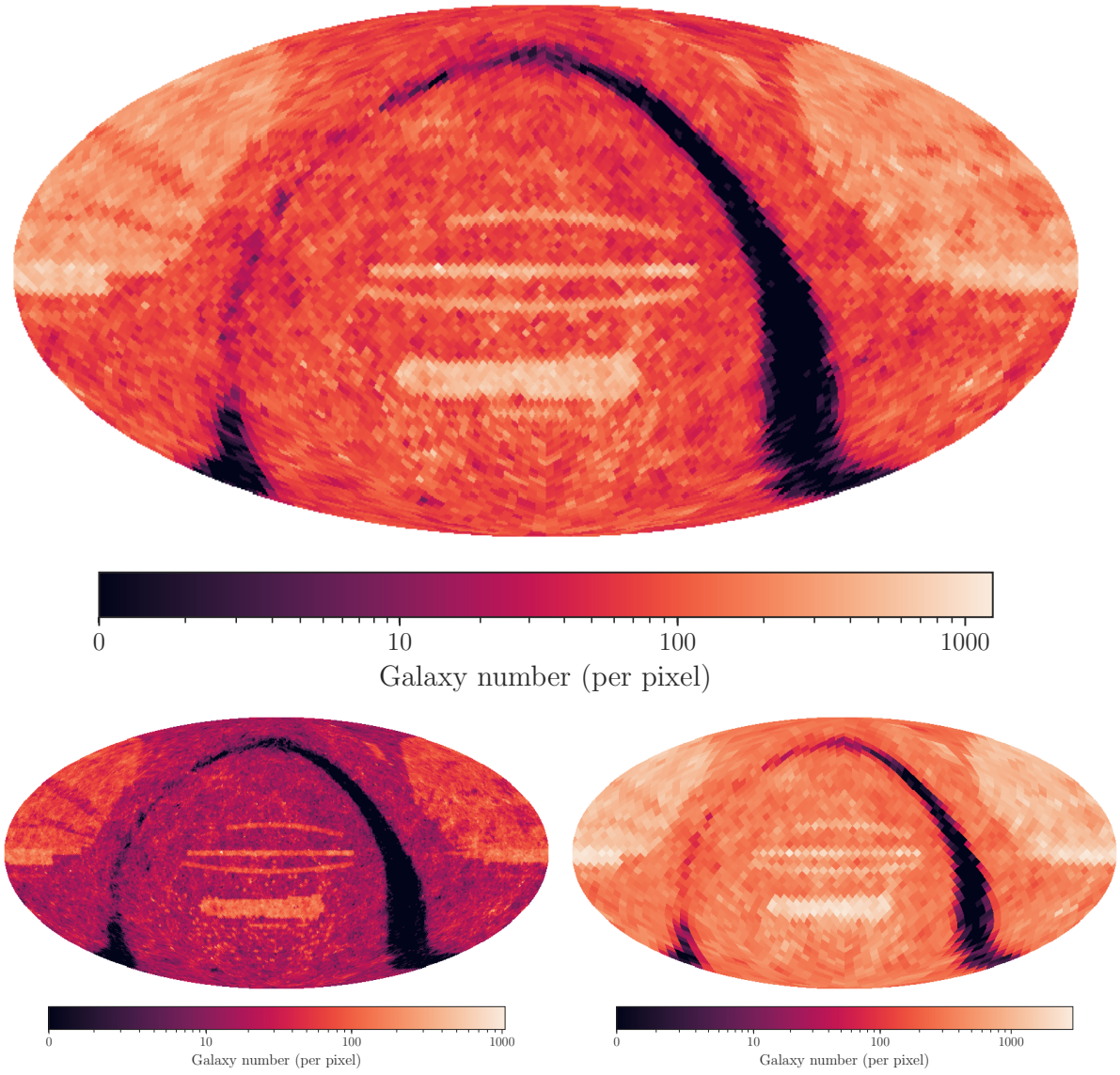

Figure 5.1: Different resolutions of the galaxy number density map for GLADE 2.4 [80]. Brighter colours correspond to a higher number density, darker colours to a lower one. Black pixels are empty, covered by the Milky Way band. *Top panel:* nside=32. The sky is divided into 12,288 pixels, each covering an area of 3.36 deg^2. *Bottom left panel:* nside=64. The sky is divided into 49,152 pixels, each covering an area of 0.84 deg^2. *Bottom right panel:* nside=16. The sky is divided into 3072 pixels, each covering an area of 13.4 deg^2.

Take, for example, the GLADE catalogue [80]. It is a composite catalogue, made up of

many different surveys. Those surveys cover different patches of the sky, and have overlapped in some areas. Taking a `healpy` nside of 32 divides the catalogue into 12,288 pixels, each covering approximately 3.36 deg^2. Figure 5.1 visualises this, and shows how the number density of galaxies varies on that scale. Many sharp features can be seen, and are adequately represented by this resolution, though they could be better represented with a higher one (*e.g.* nside=64, the lower left panel), while a lower resolution starts to blur out useful information (*e.g.* nside=16, the lower right panel). However, looking at Table 4.1 which summarises the O1 and O2 GW detections, it can be seen that GW151226, a relatively nearby event with reasonable catalogue support, has a 90% sky area of 1033 deg^2. So even a conservative analysis which only considered the 90% most probable sky area would still require 308 pixels to represent it. To cover the 99.9% sky area would push this to $O(1000)$. The computational time for this analysis increases approximately linearly with the number of pixels, meaning that it could take 1000 times longer to analyse the same event as it did when using the GBO method, or would require 1000 computer cores to analyse it on a similar timescale, if the analysis was parallelised. However, dividing the GW sky area into 1000 pixels is an unnecessarily high resolution for adequately representing the changes in the event's distance distribution.

The method of choosing which resolution to treat the GW data with, and computing the line-of-sight distance distributions for it, is discussed in Section 5.3.1. The method of combining GW data with a galaxy catalogue represented by a higher resolution is presented in Section 5.3.2.

5.3.1 Line-of-sight luminosity distance estimates

As was discussed in section 4.4.1, an important aspect of this cosmological analysis is re-weighting the GW posterior samples to remove the detector-frame mass priors used during parameter estimation, and apply the desired source-frame mass priors – not only does this affect the shape of the distribution, but also its normalisation as a function of H_0, as samples may become inconsistent with the source-frame mass prior for certain values of H_0. As such, when implementing the pixelated method into `gwcosmo`, the GW mass and distance information needs to be preserved for each pixel. This means that, even though 3D skymaps exist which contain the line-of-sight (LOS) GW distance information for each pixel, these cannot be used here because they include the marginalised-over priors applied during parameter estimation.

In order to retain the necessary information, the posterior samples are used directly to compute the LOS distance estimate for each pixel. A crude way to do this would be to choose some resolution for which to grid up the sky, then divide up the posterior samples between pixels, and create the LOS distance estimate for each pixel from the samples within it. However, this attempt fails at the first hurdle, where the finite number of samples comes into play. Most events have $O(100,000)$ samples, of which $O(100)$ are required to get a reasonable distance estimate (and more samples is preferable, as the estimate will be more reliable). As the samples cover the full parameter space, including RA and dec, the edges of the event's sky area are particularly

poorly sampled. For a relatively low-resolution skymap, this leaves pixels in the event's 99.9% sky area with as few as $O(1)$ samples.

In order to avoid this, while still remaining compatible with a `healpy` implementation, the LOS distance estimate for each pixel is found by selecting all the samples within a certain angular radius of the centre of a pixel (determined by the resolution of the pixels). If the number of samples exceeds some threshold to be deemed "enough samples" (say 100), then a KDE on these samples is used to create the LOS distance estimate. If there are not enough samples within the selected area, the angular radius is incrementally increased until the number of samples passes the threshold. KDEs are normalised by default, so each pixel's LOS distance distribution needs to be additionally weighted by the sky probability within that pixel of the GW skymap. By selecting samples in this way, the necessary information required to re-weight the samples by their source-frame mass priors is retained. Pixels in the most probable sky areas will have $O(10,000)$ samples, while those in the lowest probability regions will only have 100, taken from a larger sky patch. The decrease in reliability of these low-probability distance estimates will be compensated for by the down-weighting of their contribution to the final result.

The choice of resolution used to represent the GW data should depend on the area of the sky covered by the GW within some probability threshold. Events with small sky areas will need smaller pixels to adequately represent the changing of their distance distribution across the sky, and vice versa. As such, for `gwcosmo` the resolution is chosen by defining a threshold on the sky area of the event and determining the nside resolution that would be necessary to split that patch of the sky into (at least) a minimum number of pixels. Choosing a minimum pixel number of 30 to cover the 99.9% sky area of each event produces the LOS distance (redshift) estimates shown in Fig. 5.2. The left-hand panels shows the breakdown of the distance/redshift distribution along the line-of-sight of each pixel for the four nearest events from O1 and O2 (GW150914, GW151226, GW170608, and GW170814 – the remaining events can be found in Appendices B.1 and B.2).[3] The redshift at which the distribution peaks varies from pixel to pixel, indicating that galaxies at different redshifts will be favoured for different lines of sight. The right-hand panel of Fig. 5.2 demonstrates that the overall distribution on distance/redshift when summing the contribution from individual pixels is very close to the original distance distribution, estimated from all of the samples.

Table 5.1 summarises the `healpy` resolution ($\text{nside}_{\text{low}}$) and number of pixels ($N_{\text{pix}}$) required for each event in order to meet the criteria of at least 30 pixels to cover the event's 99.9% sky area. Additionally, the number of posterior samples in the most probable pixel is recorded (N_{samples}) which shows that for every event the most probable pixels, which will be contributing most strongly to the result, have at least 1000 samples, and in some cases as high as tens of thousands. The impact of changing the sky area under consideration, or the minimum number

[3] Figure 5.2 is shown as a function of redshift, as opposed to luminosity distance, because the KDE is done on redshift samples, as discussed in Section 4.4.1.

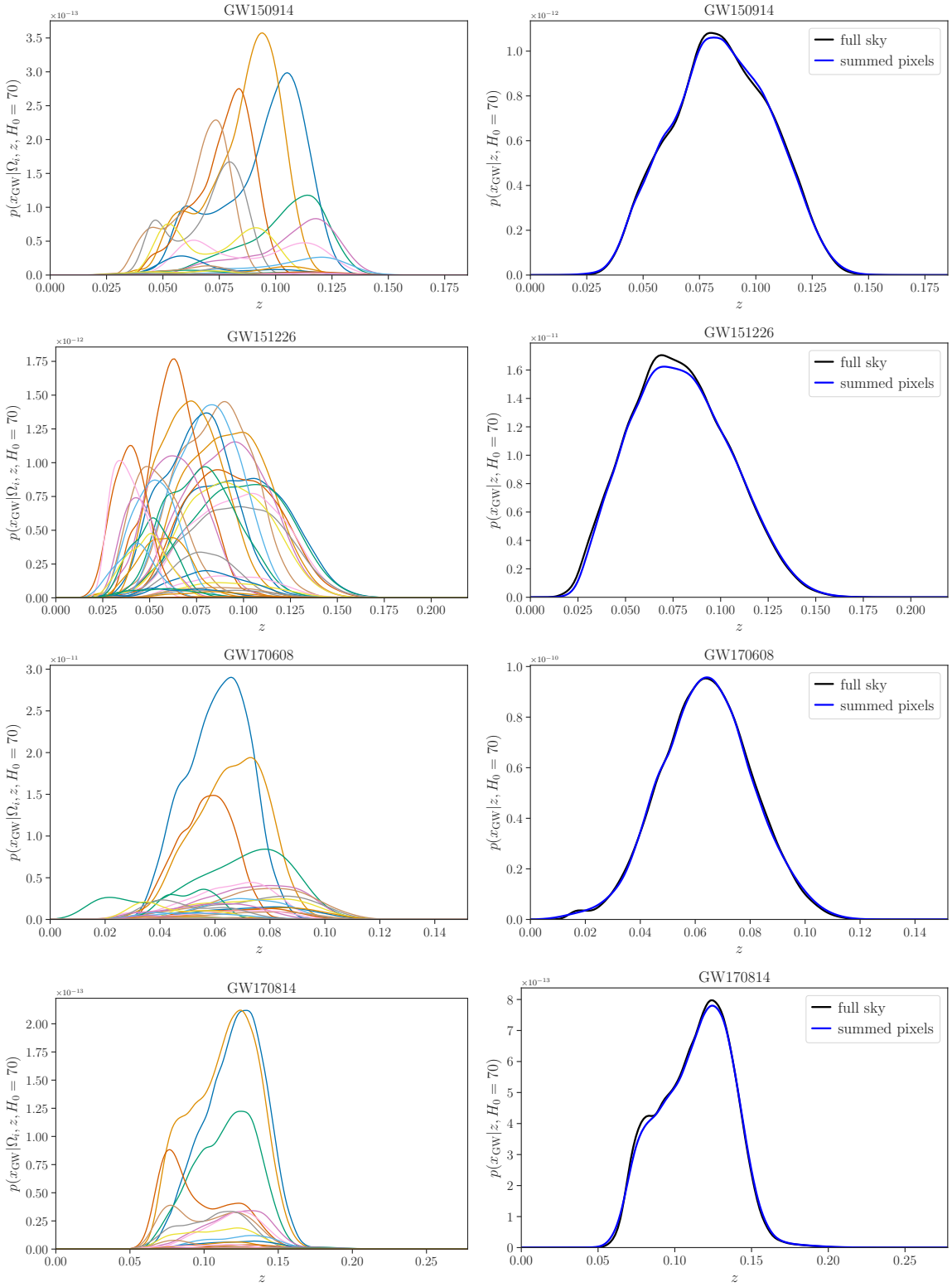

Figure 5.2: Line-of-sight redshift estimates for GW150914, GW151226, GW170608 and GW170814. *Left-hand panels:* The line-of-sight redshift distribution of the event within each pixel, where the number of pixels and pixel nside for each event are specified in the $N_{\rm pix}$ (low-res) and $n_{\rm low}$ columns of Table 5.1. *Right-hand panels:* The full-sky redshift distribution for the event. The black curve shows the estimate from doing a KDE on all the samples, while the blue curve shows the summed curves from the left-hand panel.

of pixels required to cover it (and hence the resolution with which the GW data is treated) is examined in greater detail in Section 5.4.1.

Event name	$\text{nside}_{\text{low}}$	N_{pix}	N_{samples}	$N_{\text{sub-pix}}$/pixel	$N_{\text{sub-pix}}$ (total)
GW150914	8	35	7759	16	560
GW151012	4	40	1941	64	2560
GW151226	4	34	3737	64	2176
GW170104	4	30	4804	64	1920
GW170608	8	38	4010	16	608
GW170729	4	32	22582	64	2048
GW170809	8	34	30882	16	544
GW170814	16	54	12750	4	216
GW170818	16	72	6417	4	288
GW170823	4	43	6130	64	2752

Table 5.1: A summary of the resolutions used for the pixelated analysis for the O1 and O2 BBHs. The second and third column show the lowest nside which satisfies the criteria that at least 30 pixels must cover the 99.9% sky area of the event, $\text{nside}_{\text{low}}$, and the number of pixels, N_{pix}, required to do so. The most probable pixel for each event contains N_{samples} samples. Each of the pixels is broken into $N_{\text{sub-pix}}$ sub-pixels in order to reach a galaxy catalogue resolution of nside=32, and the total number of sub-pixels which cover the event's 99.9% sky area is $N_{\text{sub-pix}}$ (total).

Sanity check: the empty catalogue case

To further prove that this method of estimating the LOS distance is "good enough", and to ensure that the H_0-dependence of the distribution has been correctly propagated, it is worth coming back to the empty catalogue analysis. The pixelated empty catalogue analysis should return results equivalent to those shown in Fig. 4.1, as Ω is marginalised over in both. Results for GW150914, GW151226, GW170608, and GW170814 can be seen in Fig. 5.3 (the remaining events can be seen in Appendices B.3 and B.4). The left-hand panels show a comparison between the pixelated empty catalogue likelihood for each event (orange) and the O2-H_0 empty catalogue result (blue). The right-hand panels show the contribution to the pixelated empty catalogue likelihood from each pixel covering the 99.9% sky area of the event. The orange curve on the left-hand panels is the combined sum of the curves from the right-hand panels. A close examination of the right hand panels shows that individual pixel contributions do not have the same H_0 dependence, independent of the scale (GW150914 in particular shows a good example of this). This difference in the slope is due to the change in the GW LOS distance distribution

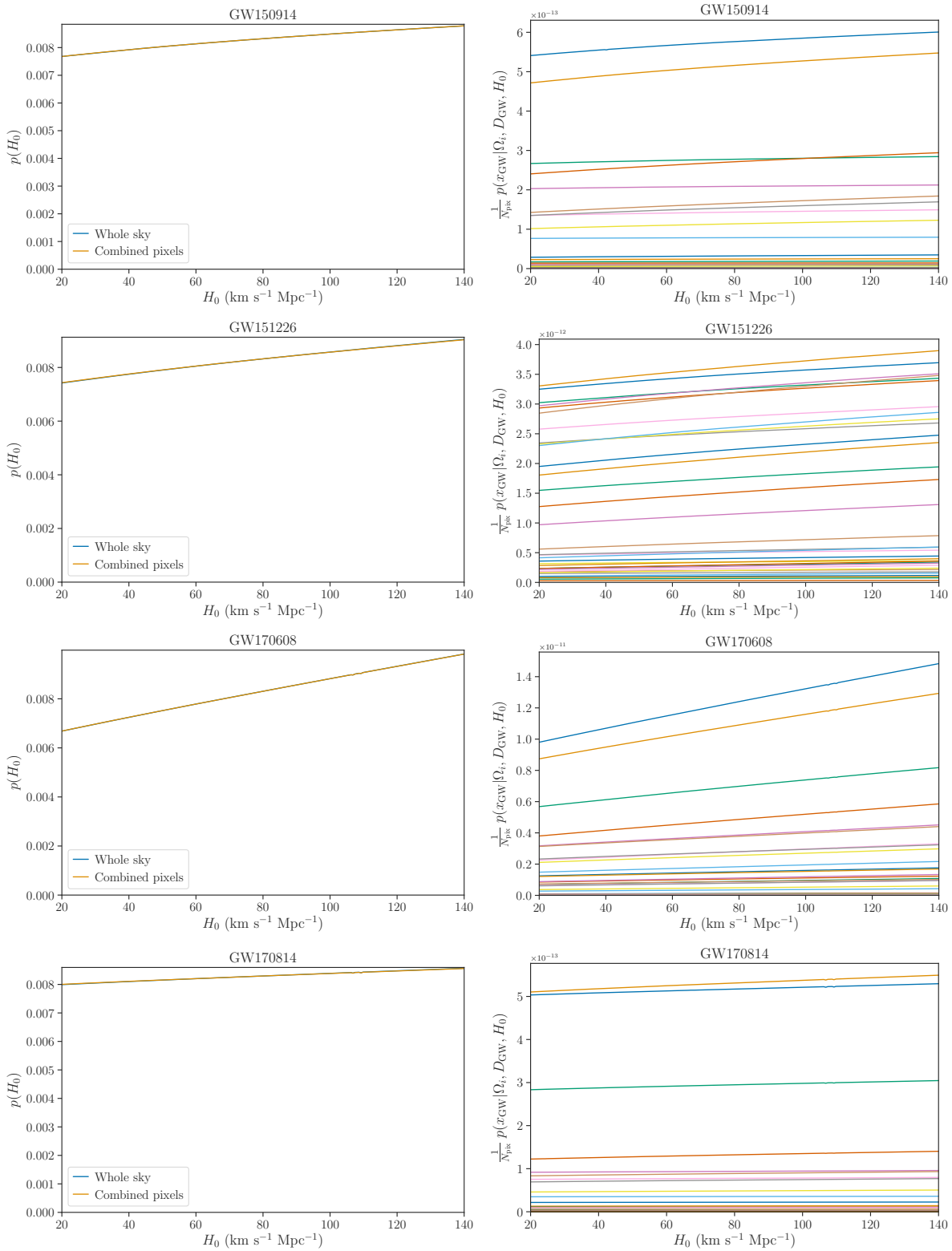

Figure 5.3: Likelihoods on H_0 for the empty catalogue analysis with GW150914, GW151226, GW170608, and GW170814. *Left-hand panels:* Comparison of the (normalised) empty catalogue likelihood on H_0 between the pixelated and non-pixelated (whole sky) cases (orange and blue curves respectively). *Right-hand panels:* A breakdown of the pixelated empty catalogue likelihood by pixel. The pixelated likelihood in the left-hand panel is the sum of the curves in the right-hand panel (then normalised).

for different pixels. The two curves on the left-hand plot overlay almost exactly, demonstrating that the GW data is well-represented by this pixelated approach.

5.3.2 Varying m_{th} within an event's sky area

The second major improvement the pixelated method offers is the ability to treat the apparent magnitude threshold, m_{th}, of a galaxy catalogue, as a function of sky direction.

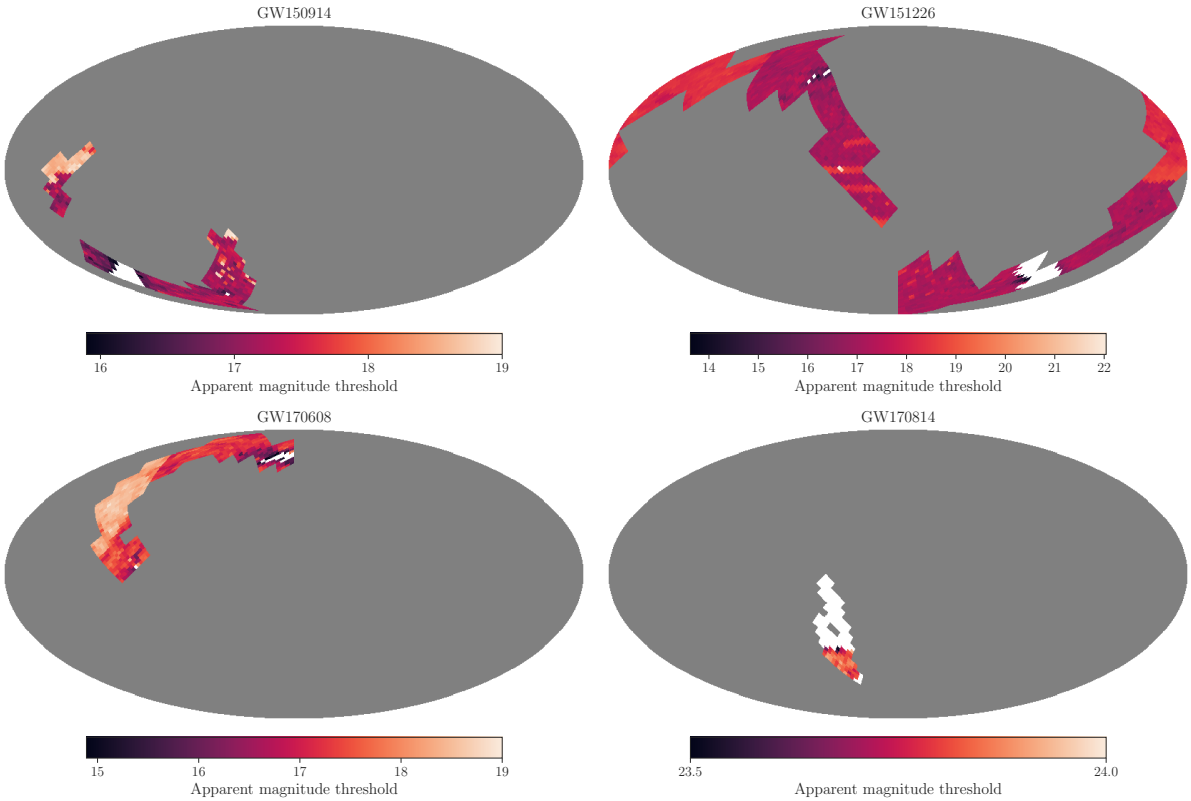

Figure 5.4: Variation of the apparent magnitude threshold within the 99.9% sky area of GW150914, GW151226 and GW170608 (using the GLADE catalogue), and GW170814 (using the DES-Y1 catalogue). The galaxy catalogue resolution (small pixels) is set with an nside of 32. The grey indicates a part of the sky which is not in the 99.9% sky area of the event. The white pixels contain fewer than 10 galaxies, and are taken to be empty.

Figure 5.4 shows the variation in apparent magnitude threshold across the 99.9% sky area of GW150914, GW151226, GW170608, and GW170814 (the remaining events can be seen in Appendix B.5). The resolution of the larger pixels, which determine which patch of the sky is considered for the analysis (in colour, where the grey is what is excluded), is determined by the resolution required to cover the 99.9% sky area with at least 30 pixels. For GW150914 and GW170608 this was an nside of 8. For GW151226, which is particularly poorly localised, this reduces to an nside of 4, while GW170814, which is much better-localised, adopts an nside of 16. An nside of 32 is chosen to represent the galaxy catalogue information, meaning that the pixels for each event needs to be split into a number of sub-pixels in order to reach the required

resolution. If $\text{nside}_{\text{low}}$ is the low-resolution nside used to represent the GW data, and $\text{nside}_{\text{high}}$ is the nside used to represent the galaxy catalogue, the number of sub-pixels that a single pixel must be divided into, $N_{\text{sub-pix}}$, is given by 4^k, where $k = \log_2(\text{nside}_{\text{high}}/\text{nside}_{\text{low}})$. This information is summarised for each event in the last columns of Table 5.1.

The method of determining m_{th} within each sub-pixel is the same as in the O2-H_0 analysis – by taking the median apparent magnitude of galaxies within the pixel.[4] Again, looking at Fig. 5.4, it is clear that this method captures a large variation in the apparent magnitude threshold within each event's sky-area. The variation of m_{th} over the sky translates directly into a variation in the probability that the host galaxy of the event is inside the catalogue, with lower thresholds corresponding to lower in-catalogue probabilities. Figure 5.5 shows how the probability that the host is in the catalogue varies within each event's sky area (events not shown here can be found in Appendices B.6 and B.7). The orange curves correspond to the pixels which contain 50% of the event's sky probability, with darker curves being more probable. The yellow shaded area shows the full range covered by the pixels which make up the 99.9% sky area. This extends to a probability of 0 at $z = 0$ if there are pixels within the event's 99.9% sky area which are empty. In general, these plots show how the probability that the host is in the catalogue compares between the pixelated case and the O2-H_0 case (Section 4.5.2) which is shown by the dashed blue line. This curve always lies within the extremes of the pixelated case (within the yellow area). However, the orange curves indicate where the bulk of the GW probability is lying. For GW150914, for example, the bulk of the GW probability corresponds to lower in-catalogue probability than previously, no doubt driven by the fact that part of the event is obscured by the Milky Way band, which will cause a low m_{th} in the adjacent pixels (pixels which are entirely empty do not show up on this plot). Conversely, GW170608's probability is clustered in an area of higher in-catalogue completeness than the average, boosting the in-catalogue support. For GW170814, the most probable pixels overlap with the O2-H_0 curve, as the m_{th} estimation for the DES-Y1 catalogue varies very little over the sky.

5.4 Results

Taking the method outlined in Section 5.2 and applying it to the O1 and O2 GW BBHs produces the individual event likelihoods on H_0 shown in Fig. 5.6 (GW150914, GW151226, GW170608, and GW170814 are shown explicitly as they have the highest in-catalogue contribution – the same information for the rest of the events can be found in Appendices B.8 and B.9).

The most interesting event is GW170608, due to its high in-catalogue probability. Looking

[4] As mentioned in the Section 4.5.2, picking the median apparent magnitude is a conservative choice, and so results which use this method will give less support to the in-catalogue part of the analysis than a method which calculates m_{th} more robustly. However, estimating m_{th} using, say, the 90^{th} percentile apparent magnitude of the galaxies is far more likely to produce a reliable result using the pixelated method than the previous O2-H_0 or GBO methods. However, for consistency of comparisons between methods, the same choice is made here.

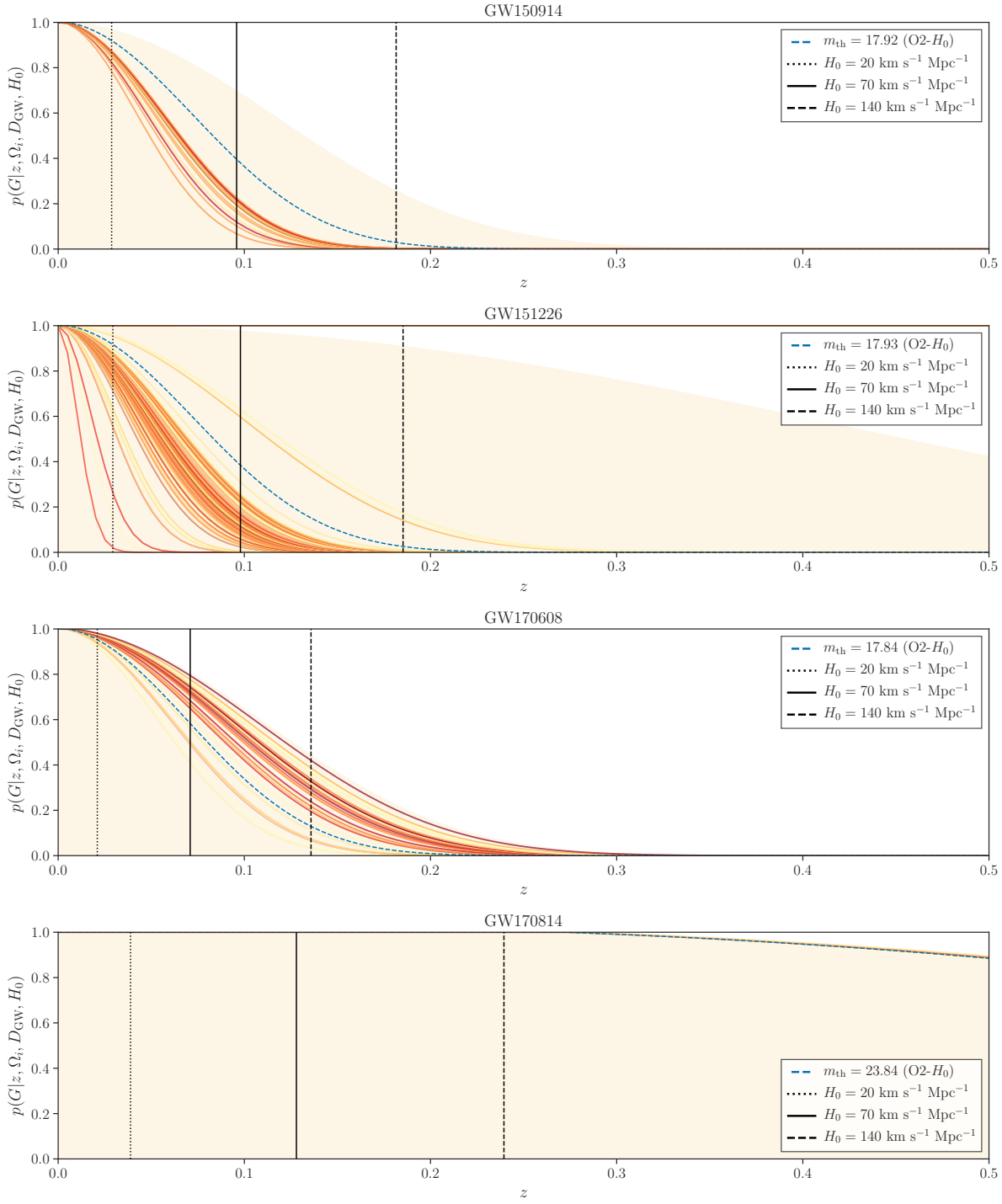

Figure 5.5: Probability that the host galaxy is in the galaxy catalogue as a function of redshift, for the pixelated catalogue case. Solid orange curves give probability that the host is inside the galaxy catalogue along the line-of-sight of the pixels which cover the 50% sky-area of the event (darker orange refers to more probable). The shaded yellow area covers the range between the minimum and maximum apparent magnitude threshold within the 99.9% sky-area of the event. Vertical black lines show the median redshift of the event for different values of H_0. The blue dashed line gives the probability that the host is in the galaxy catalogue as it was determined during the O2-H_0 analysis.

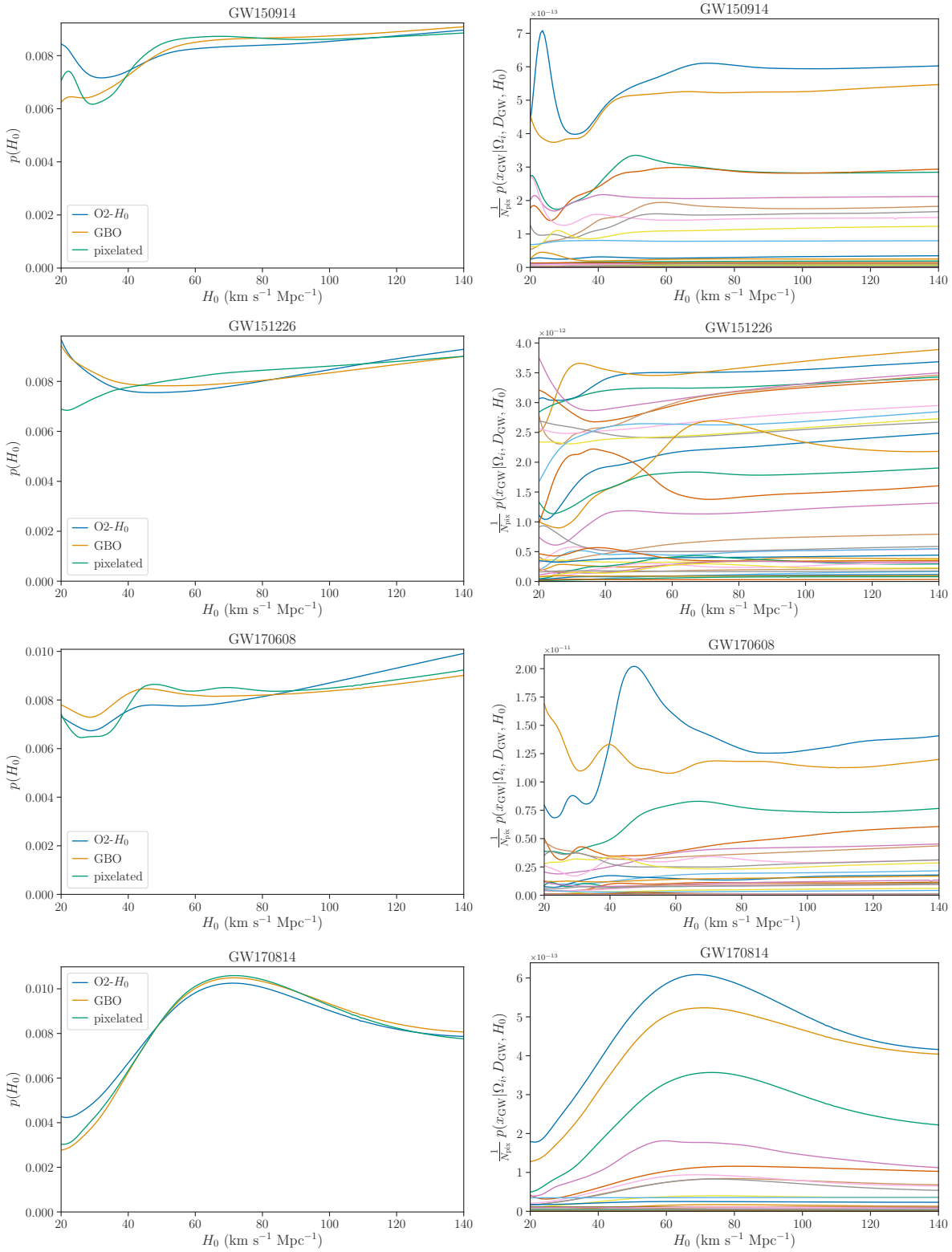

Figure 5.6: Likelihoods on H_0 for the pixelated analysis with GW150914, GW151226, GW170608, and GW170814. *Left-hand panels:* Comparison of the (normalised) likelihoods between the O2-H_0 (blue), GBO (orange) and pixelated (green) methods. *Right-hand panels:* A breakdown of the pixelated likelihood by low resolution pixel. The pixelated likelihood in the left-hand panel is the sum of the curves in the right-hand panel (then normalised).

at its right-hand panel in Fig. 5.6, which shows the contribution to the final likelihood on H_0 from each individual pixel, it is clear that for low values of H_0 the in-catalogue contribution is dominating. The final likelihood shows more structure than in either the O2-H_0 or the GBO case, and this structure intriguingly includes a bump around $H_0 \sim 70$ km s^{-1} Mpc^{-1}.

While the likelihoods on H_0 from GW150914 and GW151226 do not show much structure, both have more support around central values of H_0, and have reduced support at high H_0, compared to the O2-H_0 and GBO cases. For these events, the contributions from the individual pixels are more interesting than the combined likelihood as they show, especially at the low-H_0 end (which corresponds to low redshifts and therefore increased catalogue support) structure which corresponds to real redshift and luminosity information from the GLADE catalogue galaxies. It is interesting that GW150914's most probable pixel peaks around $H_0 \sim 70$ km s^{-1} Mpc^{-1} which could indicate an over-density of galaxies at the relevant redshift, although it remains much more likely that the host galaxy for this event is not contained within the catalogue.

The other event of interest is GW170814, which was analysed with the DES-Y1 catalogue, and for which the likelihood remains relatively unchanged with the pixelated method, though with a fraction more support around central values of H_0. Looking at the line-of-sight estimates for this event (Fig. 5.2), it is clear that the bulk of the information is coming from a small number of pixels, which peak at approximately the same redshifts. The apparent magnitude threshold within the event's sky area remains relatively uniform (see Fig. 5.4). From the likelihood breakdown of the event in Fig. 5.6, it seems likely that the most probable pixels overlap with the same over-density of galaxies in the catalogue, leading to a relatively unchanged result.

Taking the same 6 BBHs (GW150914, GW151226, GW170104, GW170608, GW170809 and GW170814) as selected in the main O2-H_0 result from Section 4.6, and combining them with the result from GW170817 and its counterpart gives the posterior on H_0 shown in Fig. 5.7. The pixelated method gives a result of $H_0 = 68.8^{+15.9}_{-7.8}$ km s^{-1} Mpc^{-1}, (MAP and 68.3% highest-density interval, with a flat-in-log prior on H_0). This is approximately a 5% improvement over the O2-H_0 result of $H_0 = 68.7^{+17.0}_{-7.8}$ km s^{-1} Mpc^{-1}, and a 3% improvement over the GBO result of $H_0 = 68.8^{+16.7}_{-7.8}$ km s^{-1} Mpc^{-1}. The improvement is driven by the additional information from GW170608, which has greater support around an H_0 of 70 km s^{-1} Mpc^{-1}due to a more informative catalogue contribution, as well as by minor improvements from GW150914, GW151226 and GW170814, all of which have marginally reduced support at higher values of H_0 under this new method.

5.4.1 The impact of resolution choices on the result

The pixelated method introduces several choices to the analysis: the threshold for the sky area that will be analysed, the number of pixels that will be used to cover that sky area (and hence the size of those pixels) and the resolution of the galaxy catalogue. This section investigates the impact of varying those choices.

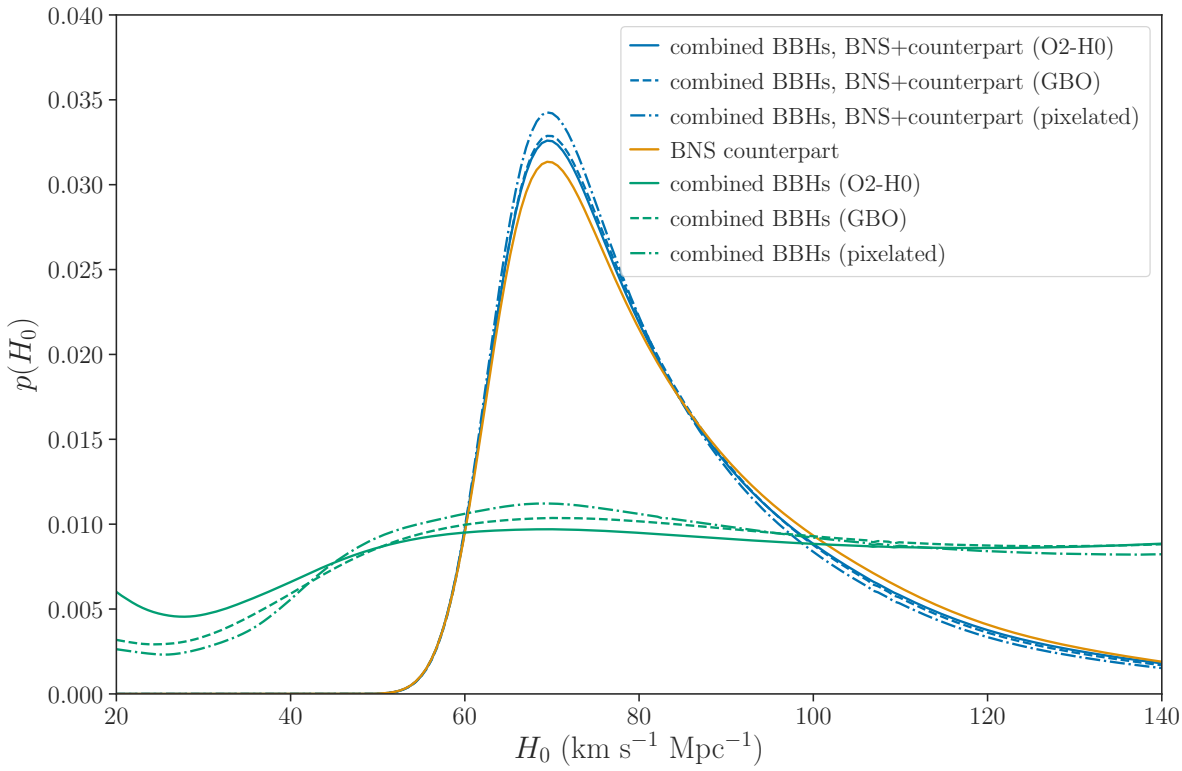

Figure 5.7: Comparison of the posterior on H_0 between the pixelated, GBO, and O2-H_0 analyses. The green lines show the combined posterior from 6 BBHs (GW150914, GW151226, GW170104, GW170608 and GW170809 analysed with the GLADE 2.4 catalogue, and GW170814 analysed with the DES-Y1 catalogue). The orange line shows the contribution from GW170817 and its counterpart. Blue lines show the combined posterior on H_0 from the 6 BBHs and the BNS with counterpart. Solid lines correspond to the O2-H_0 results, dashed lines are for the GBO results, and dot-dashed lines show the pixelated results.

In the first instance, the threshold on the GW sky area is varied. The results in Section 5.4 made use of the 99.9% sky area. Figure 5.8 demonstrates the change to the final posterior on H_0 when the area is reduced to 99% and 90% for the BBHs under consideration. The change to the final posterior is marginal, with a slight increase in the height of the peak corresponding to the smaller sky areas. This is expected as the less informative edges of the GW sky distribution are being discarded. In this case, more informative is not necessarily a good thing, as there is the possibility that discarding this additional information could eventually introduce some level of bias to the result. While not important here, that impact should be reassessed in the future, when results from large numbers of GW events are being combined.

Next, the threshold is reverted to the 99.9% sky area, but the number of pixels which covers it is increased by one factor. Effectively, the nside which was used for each event described in Table 5.1 was doubled for each event. The impact on the H_0 posterior, shown in Fig. 5.9, is a marginal increase in the peak when the higher resolution is applied to the GW data. The fact that the difference is so small should be taken as additional confirmation that a relatively low number of pixels can adequately represent the variation in the GW LOS distance distribution.

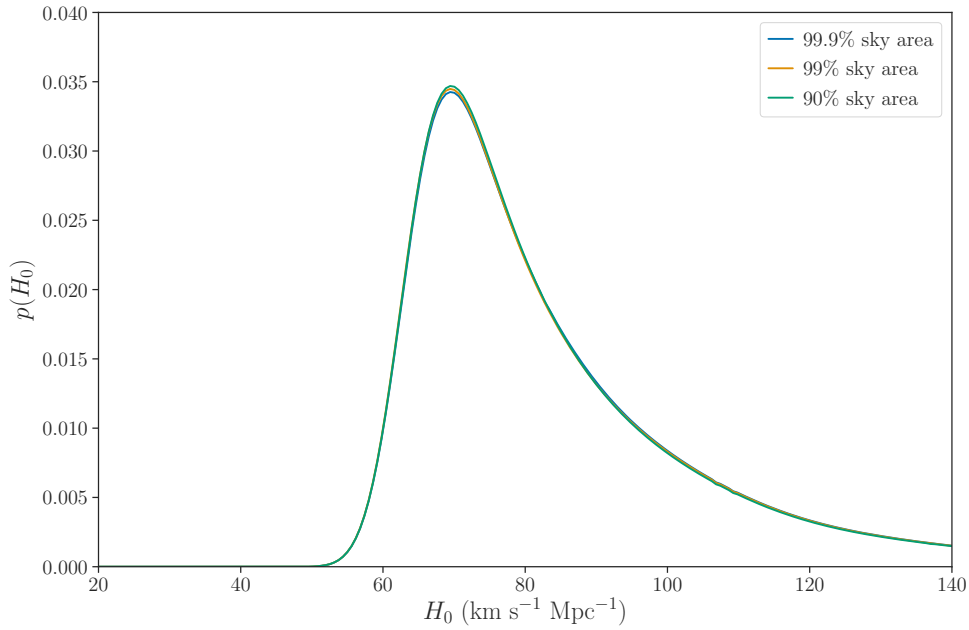

Figure 5.8: Comparison of the posterior on H_0 when the threshold on the GW sky area used in the pixelated analysis is reduced. The blue line shows the main result for 6 BBHs and the BNS, where the 99.9% sky area for each BBH is analysed. The orange line corresponds to the 99% sky area, and the green line to the 90% sky area.

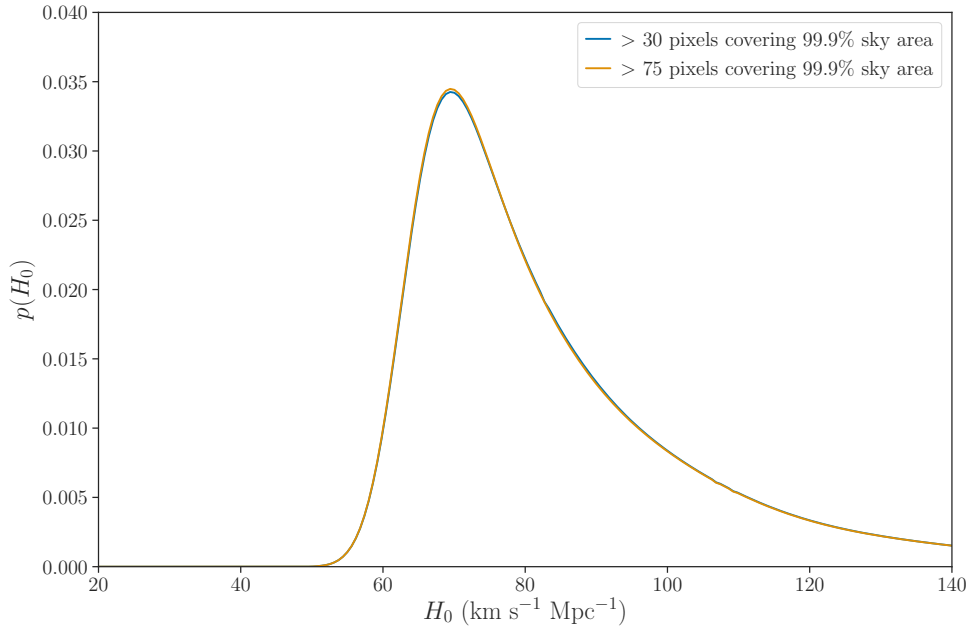

Figure 5.9: Comparison of the posterior on H_0 when the minimum number of pixels covering the 99.9% GW sky area is increased. The blue line shows the main result for 6 BBHs and the BNS, where at least 30 pixels cover the 99.9% sky area for each BBH is analysed. The orange line corresponds to an analysis where the GW data is analysed at one resolution step higher, where at least 75 pixels are used to cover the 99.9% sky area.

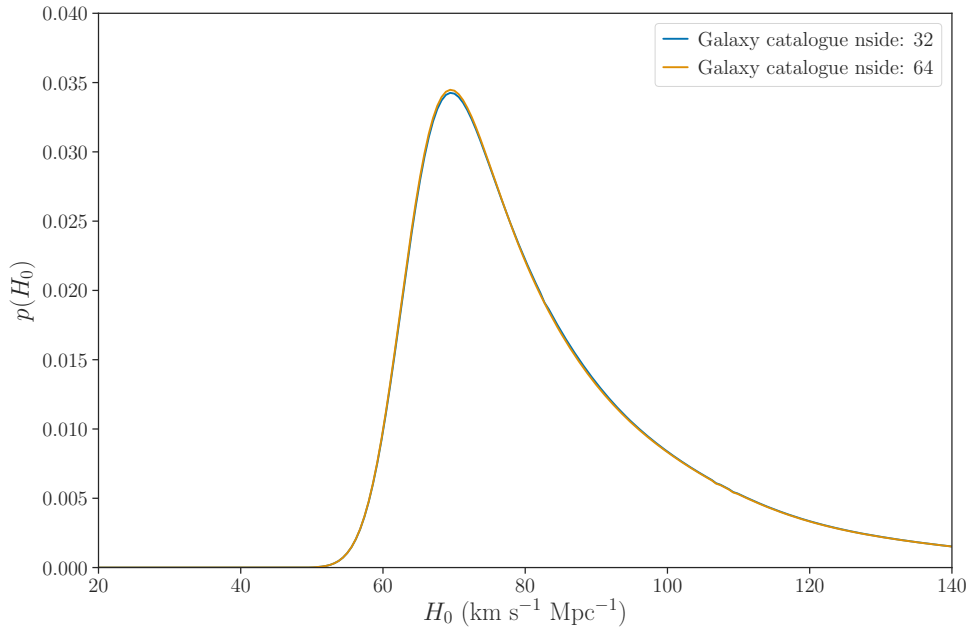

Figure 5.10: Comparison of the posterior on H_0 when the galaxy catalogue resolution is increased. The blue line shows the main result for 6 BBHs and the BNS, where an nside of 32 is assumed for both the GLADE and DES catalogue. The orange line shows the same, but where the nside of the both galaxy catalogues has been increased to 64.

Finally, the resolution of the galaxy catalogue is investigated, by increasing the nside of the sub-pixels from 32 to 64. This allows for a better representation of the hard edges of the GLADE catalogue where it is intersected by the Milky Way band, as well as better representation of the variation of $m_{\rm th}$ across different parts of the survey (see Fig. 5.1 for a reminder of the difference between nside=32 and nside=64). The results are shown in Fig. 5.10, which shows a very small increase in the height of the peak for the higher resolution results. As the difference is again very small, it is safe to assume that the resolution from nside=32 is adequate to represent the variation in galaxy catalogue completeness over the sky.

5.5 Conclusion

There are two major benefits of using the pixelated method presented in this chapter for the measurement of H_0 using standard sirens and galaxy catalogues. The first is that full use is now made of the GW data by estimating a separate distance distribution for each pixel which makes up the event's sky area. The fact that these distributions will peak at different distances for different lines of sight – and will therefore pick out galaxies at different redshifts depending on how they align – increases the information available for the dark siren analysis. The second benefit is that the pixelated method allows for an accurate estimation of how the apparent magnitude threshold – and hence the completeness – of a galaxy catalogue changes across the sky. This is particularly important around the Milky Way band, where telescope visibility is limited, and

at any boundaries between different observing surveys, between which telescope sensitivities may have changed. Given how dominant the out-of-catalogue contribution is to the dark siren result for the majority of O1 and O2 events, and will continue to be in O3 and beyond, this is an important milestone.

The results presented in Section 5.4 show a clear improvement over the results presented in Chapter 4 – around that of 5%. Not only does the pixelated method make use of the data in a way that is more technically correct, but this leads to a direct increase in the informativeness of the results it produces. These results are both more robust and have more to say.

An investigation into how various choices for the resolution of the pixelated analysis impacts the result revealed that, for the most part, the results are insensitive to reasonable changes. Variations from increasing the resolution of pixels used to represent the GW data were small, as were those from increasing the resolution of the galaxy catalogue. These changes may become more important as more events are analysed and the goal to reach 1% measurement uncertainty on H_0 comes into reach. In the future it may be worth improving the analysis outlined in this chapter to use multi-resolution maps to represent the GW data: this would allow for well-localised highly probable areas to be treated with high resolution pixels in order to capture the fine detail, and the tail ends of the distribution – the large areas with very little probability – could be treated on a lower resolution. This would allow for the most information to be gained from the GW data, without having to discard the low probability areas, which it may be necessary to include in order to avoid introducing bias to the result.

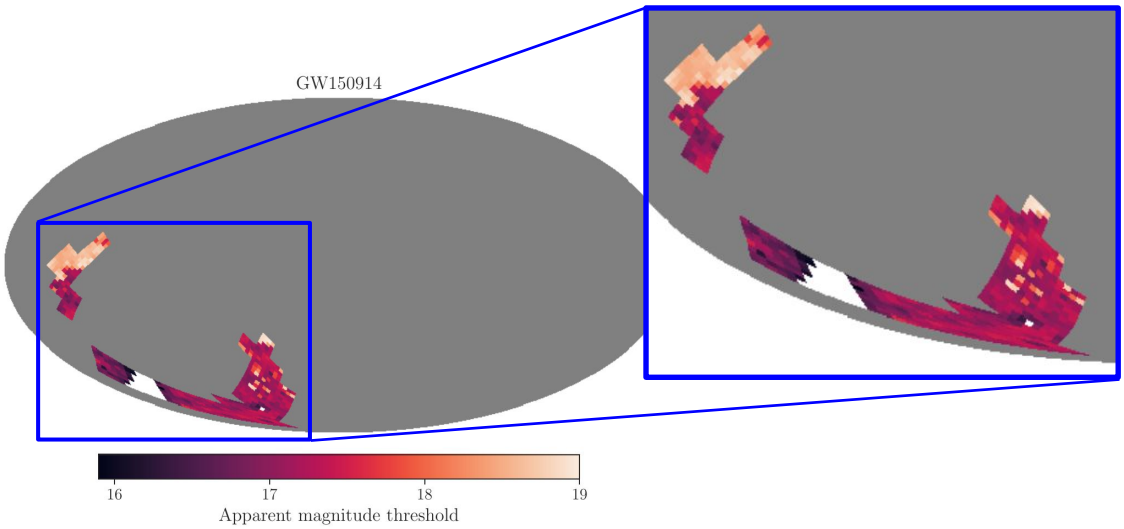

Figure 5.11: Zoomed in examination of the apparent magnitude threshold variation within the sky area of GW150914.

Of course, increasing the size of pixels means that, in order to reach the resolution set by the galaxy catalogue, each pixel needs to be split into more sub-pixels, which increases the compu-

tation time. Currently the computational cost of each pixel is approximately proportional to the number of sub-pixels which it is divided into (which means that the pixels from events with large sky areas have a longer run-time than those which are well-localised, even though these events are unlikely to contribute much information to the final result). The solution to this, which will radically reduce computation time for large pixels, is to merge sub-pixels which have similar apparent magnitude thresholds. Looking at Fig. 5.11: within the sky-area of GW150914, $m_{\rm th}$ varies between approximately 16 and 19 magnitudes - a huge variation in terms of completeness. However it is also clear that the majority of pixels have a threshold around 17.3 (reddish-pink). Pixels in the top left have higher thresholds, around 18.8. Pixels on the edge of the Milky Way band dip below 17. Overall, the variation of $m_{\rm th}$ within GW150914's sky-area could be adequately represented by a handful of different thresholds. The costly out-of-catalogue contribution within each pixel, which has to be calculated for every value of $m_{\rm th}$, would then be reduced from 16, in this case, down to only several. And if the resolution of the catalogue was increased at some point in the future, the number of out-of-catalogue calculations required would not increase, allowing incredibly high resolution of features such as empty patches and boundaries between surveys, with no extra computational cost.

In summary, the pixelated analysis demonstrated in this chapter shows a clear improvement on the analysis from Chapter 4. It improves the final constraint on H_0 through better, more effective use of the GW and galaxy catalogue data available. The implementation demonstrated here has clear avenues for future development, both in terms of further improving the accuracy of the analysis, and improving its efficiency. This pixelated approach has potential to become the main method for any future H_0 analysis with dark standard sirens and galaxy catalogues.

Chapter 6

Conclusion

The fact that the universe is expanding has long been agreed upon, but the rate at which it is currently expanding, H_0, is causing disagreement across the field of cosmology. Early-time measurements, such as the Planck CMB measurement tend to prefer lower values [4], while local measurements, such as the SH0ES supernovae measurement tend to prefer higher values [5]. This tension, at the time of writing, stands above 4σ, making the probability that it will resolve itself increasingly unlikely. Either some of the measurements have underlying systematics which have yet to be identified, or the ΛCDM model of the universe has reached its limitations, and new physics is required to describe the universe that we live in.

The idea of using GW standard sirens for cosmology has been around for decades. GW signals from CBCs have the luminosity distance to the source directly encoded within them, requiring no calibration from the EM distance ladder. Measurements of H_0 made using GWs would therefore be entirely independent of all other measurements of H_0 so far. As such, whether the GW measurement of H_0 ultimately ends up favouring the early- or late-time measurements will indicate which solution to the H_0 tension is correct. Agreeing with the early-time measurements would indicate the source of the H_0 tension is systematics in the local H_0 measurements. Agreeing with the late-time measurements would imply that the tension is real, and that its cause is as-yet unknown physics.

The detection of BNS GW170817 and its EM counterpart, and the subsequent measurement of H_0 [13], were a perfect example of how GWs might arbitrate in this case. It was incredibly informative for a single event, though not informative enough to prefer one set of measurements over the other. A few dozen more like it would be enough to make the distinction. However, GW170817 was an exceptional event, and since then, with the entirety of the third observing run of Advanced LIGO and Virgo wrapped up, no other events have confirmed EM counterparts. The majority of detections have been of BBHs, which are not expected to be detected with EM counterparts. This does not mean that these events will not be useful for GW cosmology. The use of galaxy catalogues to provide redshift information of all the potential host galaxies within the localisation volume of an event was proposed by Schutz in 1986 [8].

This thesis has focussed on using GW standard sirens to constrain H_0, with particular emphasis on darks sirens – GW signals from CBCs which have been observed without an EM counterpart. The methodology, derived in Chapter 2, takes into account both GW selection effects (from the finite sensitivity of GW detectors) and EM selection effects (the fact that galaxy surveys are flux-limited, and therefore will not contain all potential host galaxies for GW mergers). Compensating for both of these effects is required in order to obtain a measurement of H_0 which is unbiased. The incompleteness of galaxy catalogues was discussed in [49] and a methodology for addressing it was presented in [60, 63]. However, the methodology in Chapter 2 was derived independently from these, and sticks to a rigorous Bayesian approach. Quantifying the limitations of the galaxy survey using an apparent magnitude threshold mimics reality, and also allows for results that are less sensitive to the choice of galaxy luminosity function than in previous studies.

Deriving a methodology is just the first part of tackling the problem. A practical implementation is required in order to apply it. Chapter 3 introduced this in the form of the software package `gwcosmo`. For the first implementation, some simplifications were required: mainly the assumption of uniformity of galaxy catalogue incompleteness across the sky (such that it could be modelled by a single apparent magnitude threshold) and the assumption that the line-of-sight distance estimate for a GW event is the same in every direction (weighted by the GW sky probability along each line-of-sight). This implementation was then tested rigorously using a series of MDAs of increasing complexity. The MDAs made use of 250 highly realistic simulated BNS detections, and sets of mock galaxy catalogues with varying levels of incompleteness. These MDAs were the first time any GW standard siren analysis had been rigorously tested on its ability to handle the incompleteness of galaxy catalogues, and `gwcosmo` passed every test.

The first MDA confirmed that an unbiased measurement of H_0 was produced when the host galaxy of each BNS merger was known exactly. This was then extended to the case in which the host galaxy was not identified, but the galaxy catalogue was complete, so the host was definitely included within the galaxy sample. Following the success in this case, three galaxy catalogues of varying levels of incompleteness were then tested, and in each case `gwcosmo` produced an unbiased H_0 estimate, even when only as few as 25% of the BNS host galaxies were contained within the catalogue. The final MDA introduced the concept of luminosity weighting, where the host galaxies for GW mergers are more likely to come from more luminous (and therefore more massive) galaxies. Again, an unbiased measurement of H_0 was made, and the result was more informative than when assuming uniform weighting of host galaxies.

Following this success, it was decided that `gwcosmo` would be used to analyse the GW detections from LIGO and Virgo's first and second observing runs. The process of adapting `gwcosmo` to handle real data, and the results produced, make up the main body of Chapter 4. The adaptations required fall into two groups. On the GW side, `gwcosmo` needed to be updated to handle BBHs, and model GW selection effects for the BBH population. This required ensuring

that the same source-frame mass prior was used both for the selection effects and the GW data (not required for the MDAs due to the way the BNS data was generated). The outcome of not doing so would have been a strong systematic bias in the inferred value of H_0. On the galaxy catalogue side it quickly became clear that the mock catalogues used for the MDAs didn't come close to accurately modelling some of the very real challenges that come with real galaxy catalogues, including large redshift uncertainties, the redshifting of galaxy luminosities, and the variations in galaxy catalogue completeness across the sky. Each of these complications was addressed – redshift uncertainties were marginalised over, K corrections were applied to galaxy magnitudes to convert them to source-frame values, and catalogue completeness was assumed to only be uniform within the sky area of each event (excluding patches that were empty due to being covered by the Milky Way band). A measurement of H_0 was made using all the GW detections from O1 and O2, including GW170817, which is the first measurement of H_0 made by combining multiple GW events. While the improvements from the BBH detections was modest, this is a proof of principle of what will be possible in the future, with better localised GW events and deeper galaxy surveys.

At this point there were two outstanding issues in the `gwcosmo` analysis which would hamper its use going forwards. Firstly, the assumption that the completeness of a galaxy catalogue was uniform within the sky area of an event. For events with larger sky areas this remained clearly untrue. Secondly, the assumption that the distance posterior of a GW event had the same shape along every line-of-sight. While this approximation had not introduced any noticeable bias when tested with the MDAs, it was preventing the GW data from being used to its full potential, leading to results that were less informative than they could be. Both of these issues were addressed in Chapter 5. By splitting the sky into equally sized pieces (pixels), the completeness of the galaxy catalogue could be estimated for each pixel separately, and a different line-of-sight distance estimate could be used in each one, thereby solving both issues at once. A reanalysis of the O1 and O2 results produced a more informative result and, importantly, paves the way for using this analysis on the GW detections from O3 and the observing runs beyond. The pixelated version of `gwcosmo` opens up the possibility of using any galaxy catalogue for these analyses in the future, regardless of how much or how quickly its completeness varies across the sky, as long as its completeness along a given line-of-sight can be adequately modelled by an apparent magnitude threshold.

Overall the work presented in this thesis bridges the gap between the proposal outlined by Schutz in 1986 for using dark standard sirens to measure H_0, and a practical implementation which has been tested thoroughly for its ability to accurately account for galaxy catalogue incompleteness, and is well-adapted to handle the challenges of real data.

Of course there is always room for improvement, and over the course of the research undertaken in this thesis, some limitations have become apparent which will warrant further research in the years to come. As with many things in this sub-field, they can be broken into the categories

of "challenges involving GW data", and "challenges involving galaxy catalogues".

On the GW data side, one major concern that has arisen over the last year is the extent to which prior assumptions about a population of compact objects can impact cosmological inference. In particular, the source-frame mass prior for black holes is not well known. The analysis as it currently stands requires the choice of a source-frame mass distribution, and the measurement of H_0 is sensitive to this choice (as explored in Section 4.7.2). Using the detected population of BBHs to improve the source-frame mass prior comes with complications, because it is inherently linked to cosmology. There are several options. The easiest, but perhaps least "correct", would be the use of a scale-free mass prior, such that it doesn't contain any information about H_0 and has no bearing on the final result (similar to choosing a $1/R$ prior on rates as described in Section 2.2 – which, in fairness, should be revisited too). This may not be a physically realistic prior, but would allow some separation of the issues, and is implementable on relatively short timescales. Even better would be the ability to marginalise over rates and population hyper-parameters for the dark siren measurement, which would certainly be a step in the right direction. All of this is just paving the way for the ultimate goal – that of jointly estimating population parameters and cosmological ones. Currently this is doable when disregarding EM information, but if the desire is to continue to use galaxy catalogues to inform the redshift prior then the calculation becomes computationally intractable as it currently stands.

On the galaxy catalogue side, the areas for future research are numerous. The first to consider, hinted at in Section 2.3, is that the galaxy luminosity function is not truly independent of redshift. While the approximation may hold well in the local universe, there comes a point when it will break down, and as sensitivities of GW detectors improves, this will become more important. Even for redshifts below 1, the luminosity function for some bands is currently best described by an evolving Schechter function. The separation of priors on redshift and absolute magnitude won't be valid in that case, which means that the probability of a GW merger being hosted in a galaxy with redshift z and absolute magnitude M can no longer be separated into the product of those two things.

There is also much to investigate in terms of determining how the population of CBC mergers traces the population of galaxies, and identifying which galaxy properties (such as mass, star formation rate, galaxy type, or something else) are most appropriate for weighting host galaxy probability. It would then be necessary to find a way to calculate those galaxy properties using observables within a galaxy catalogue (and either ensure that they are independent of cosmological parameters, or that the cosmological dependence is accounted for within the main analysis). Even if luminosity in a certain band remains the best option for galaxy weighting, for a self-consistent analysis using galaxy catalogues that are complete to high redshifts, K corrections will have to be applied which are valid beyond the redshift of 0.5 that was considered in this thesis. K corrections are usually dependent on galaxy colour, so work will be needed in order to determine what level of accuracy is required in order to avoid introducing bias. Can the same

form of K corrections be applied to all galaxies, or will different galaxy types pose a problem?

Currently the deepest galaxy surveys come with photometric redshift estimates, as spectroscopic ones are costly. However, the reliability of photometric redshifts is not absolute, and they become less trustworthy for galaxies that are distant and faint. What level of photometric redshift reliability is required in order to avoid biasing the measurement of H_0, and how can it be quantified? Given that the galaxy catalogue is expected to be used as a prior for all GW events, a systematic shift in the redshift of faint galaxies could be catastrophic for measuring H_0. Of course, the large spectroscopic galaxy surveys that will launch this decade will be hugely important for improving both galaxy catalogue completeness, and the redshift estimates of individual galaxies, which will be a massive step forward for dark standard siren measurements of the future [131].

Of course the dark siren method is not the only tool at our disposal for measuring H_0. It is expected that over the course of the next observing runs, further BNSs or NSBHs will be observed with EM counterparts. The ideal scenario for measuring H_0 with GWs is to combine all detections, using the best EM information available, be it a counterpart or a galaxy catalogue, or dedicated EM follow-up in the sky area of an event [82]. The EM counterpart method was addressed briefly at the beginning of this thesis (see Section 2.2.2) but was not examined in any great detail as the only event which it could be applied to at the time was GW170817. What was not considered was the case in which the EM counterpart is less cut-and-dry – for example the observation of an EM counterpart that cannot be 100% confirmed to be associated with the GW merger (think GW190521 and its potential, but unconfirmed, counterpart [132]). Thus far most counterpart analyses make the assumption that if a GW is detected, and the merger creates a kilonova, then an EM counterpart will definitely be detected, or the probability that it will be detected at least has no H_0-dependence. This may not be the case, and would be worth considering in greater detail (ideally *before* the next counterpart detection).

Overall, the following decade promises to be an exciting one for GW cosmology. There is much work to be done, but with the fourth observing run approaching, more detectors set to join the GW detector network, and large spectroscopic galaxy surveys on the horizon, GWs may soon provide an answer to the H_0 tension.

One final note to end on is that, though the H_0 tension dominates headlines, and often is used to give context when discussing the use of GWs for cosmology, it is worth remembering that the GW contribution to cosmology is not limited to this one parameter, or this one question. The inherent property of GWs as standard sirens, allowing them to form a cosmic distance ladder of their own, has so much potential beyond H_0. This is more in the context of the third generation of GW detectors, and that time will come soon enough.

Appendix A

Chapter 4 supplementary material

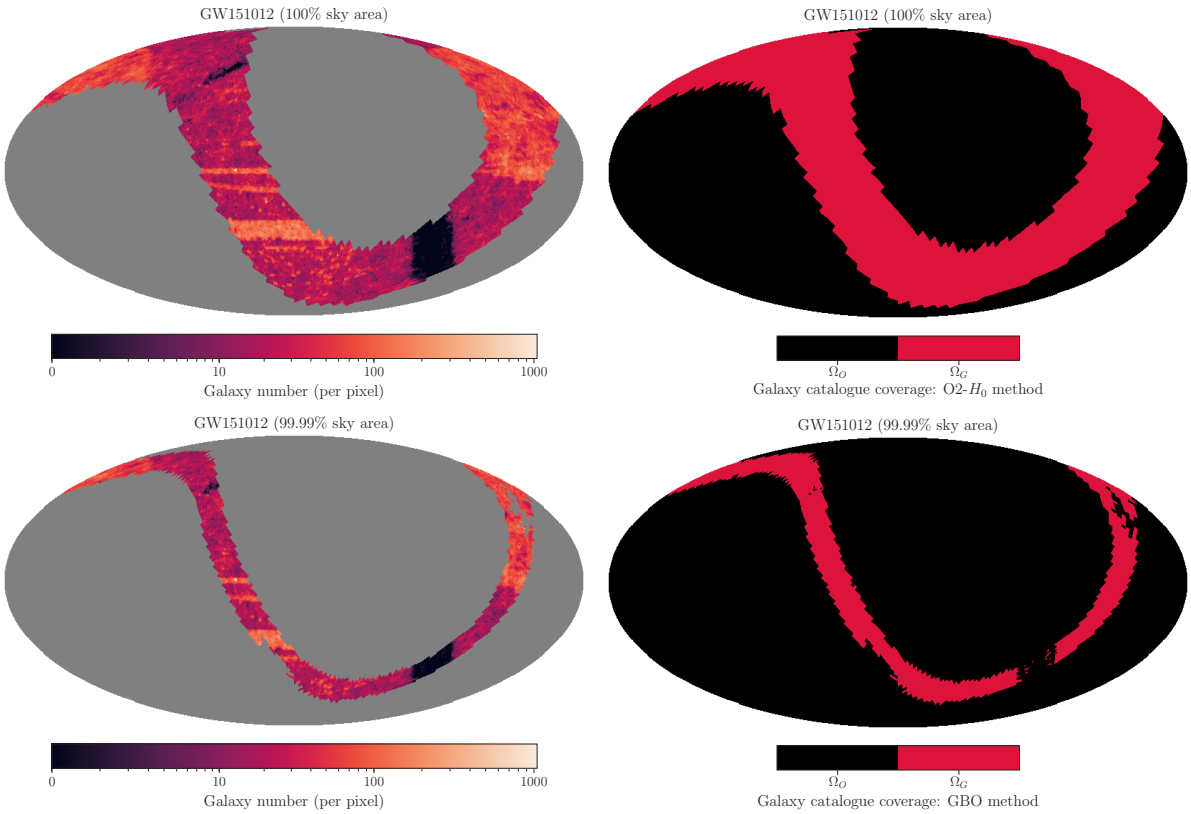

Figure A.1: A comparison of galaxy catalogue handling for GW151012 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW151012 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

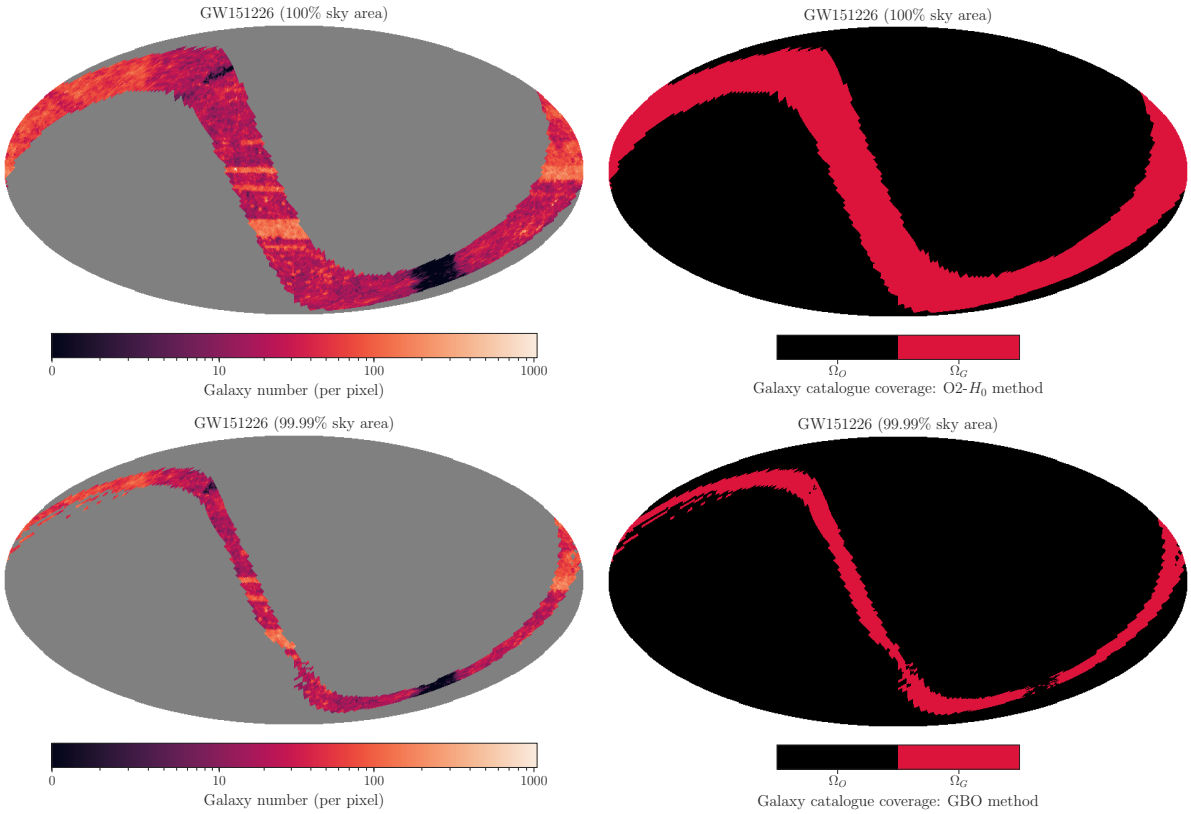

Figure A.2: A comparison of galaxy catalogue handling for GW151226 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW151226 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

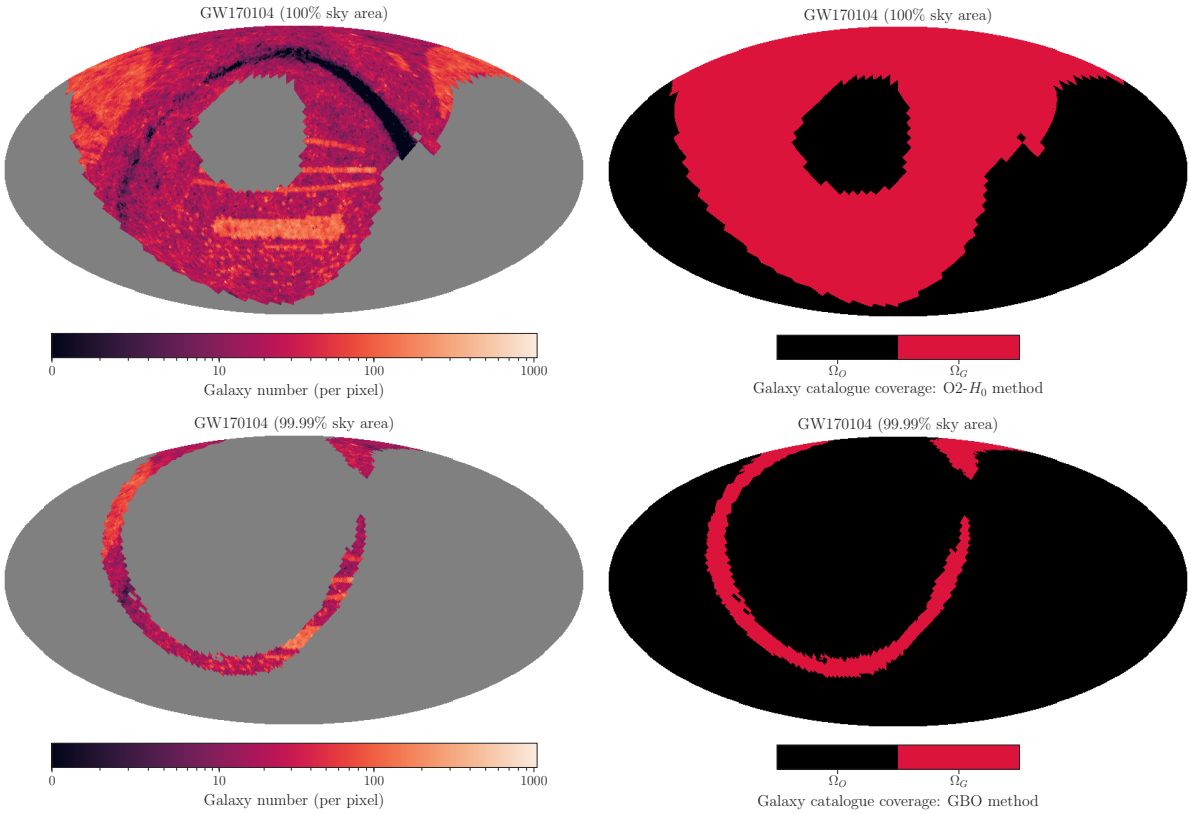

Figure A.3: A comparison of galaxy catalogue handling for GW170104 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW170104 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

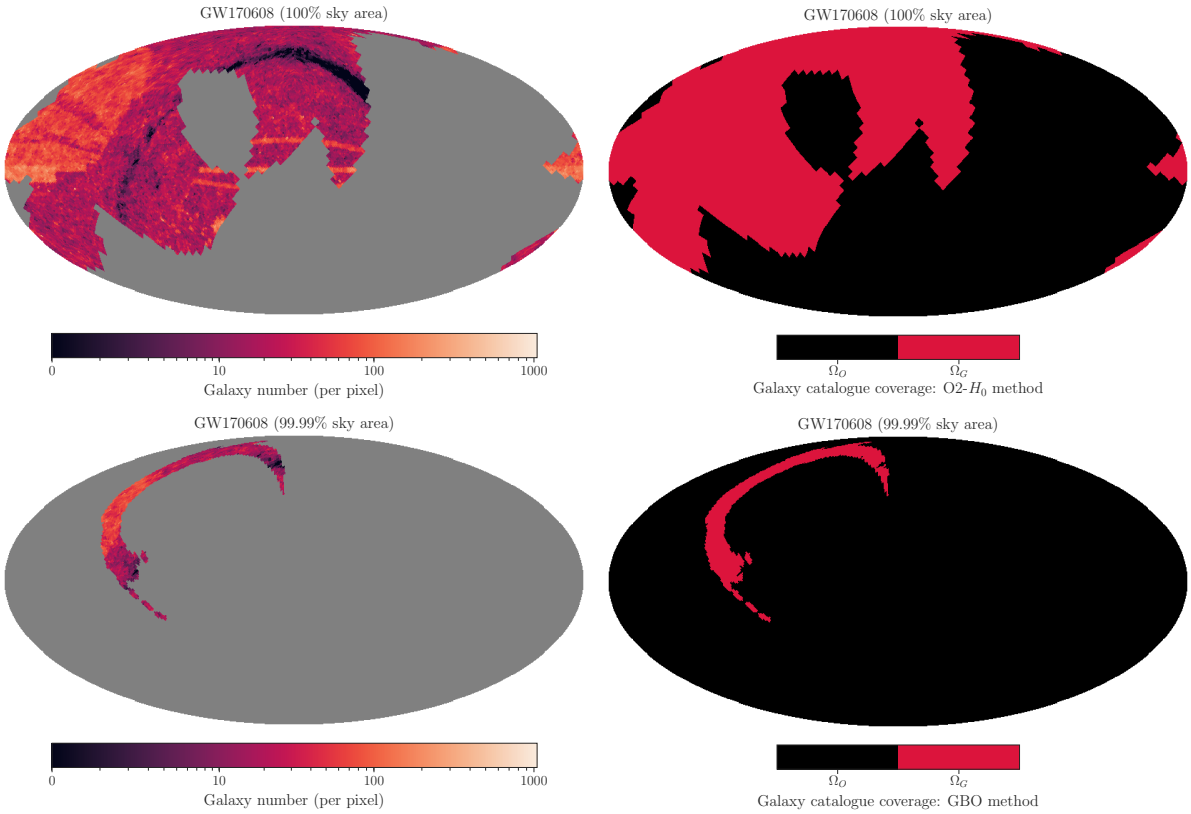

Figure A.4: A comparison of galaxy catalogue handling for GW170608 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW170608 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

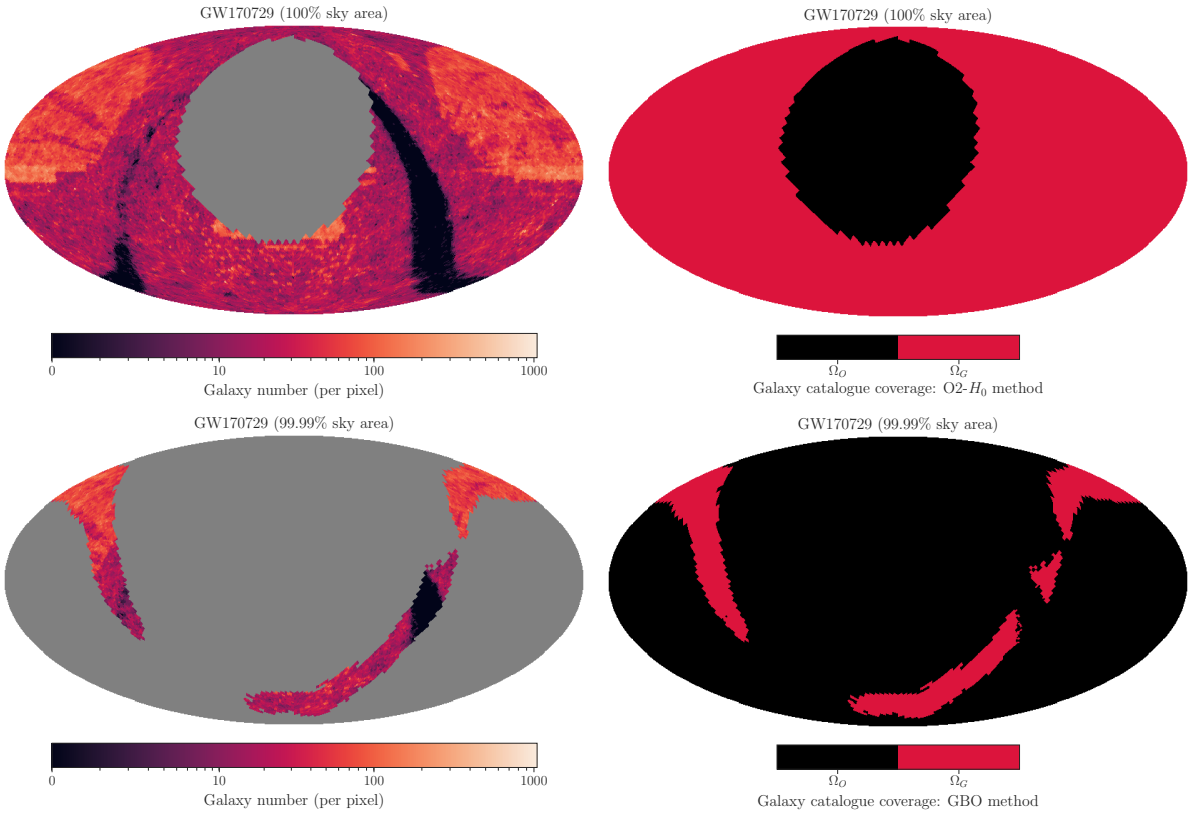

Figure A.5: A comparison of galaxy catalogue handling for GW170729 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW170729 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

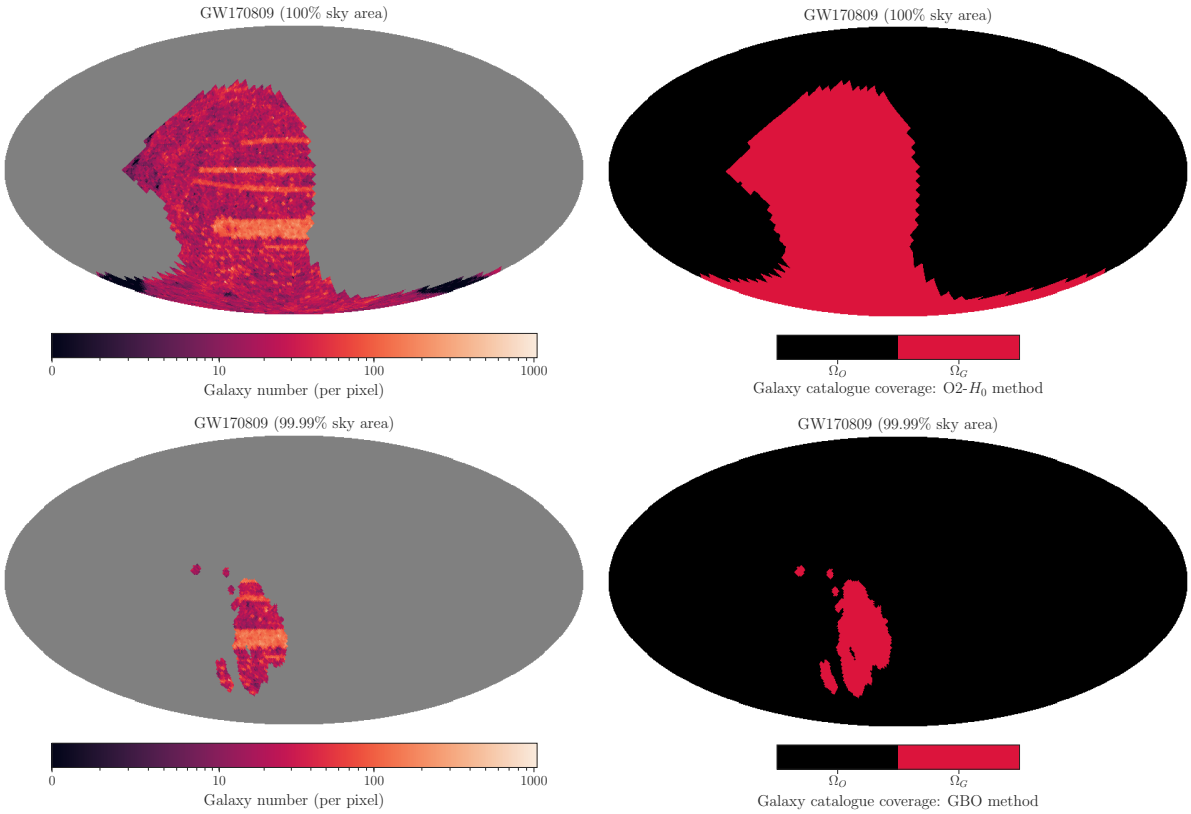

Figure A.6: A comparison of galaxy catalogue handling for GW170809 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW170809 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

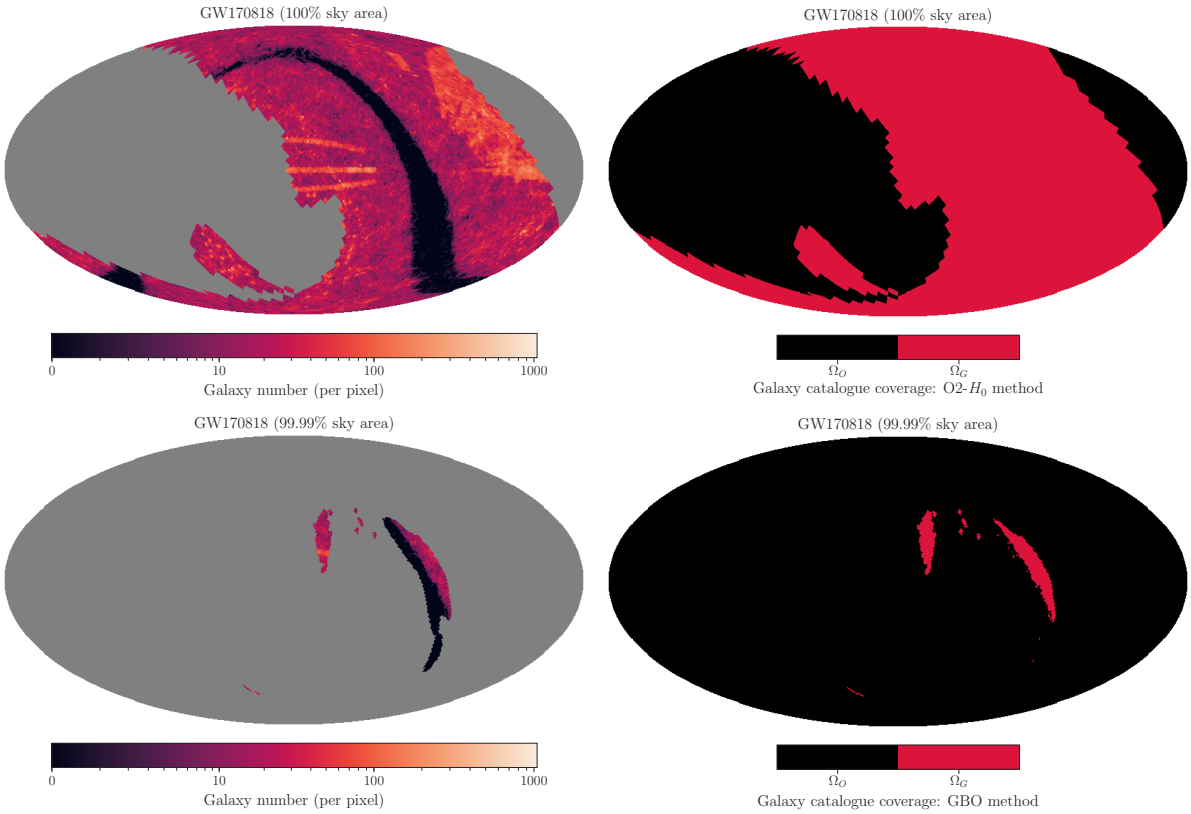

Figure A.7: A comparison of galaxy catalogue handling for GW170818 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW170818 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

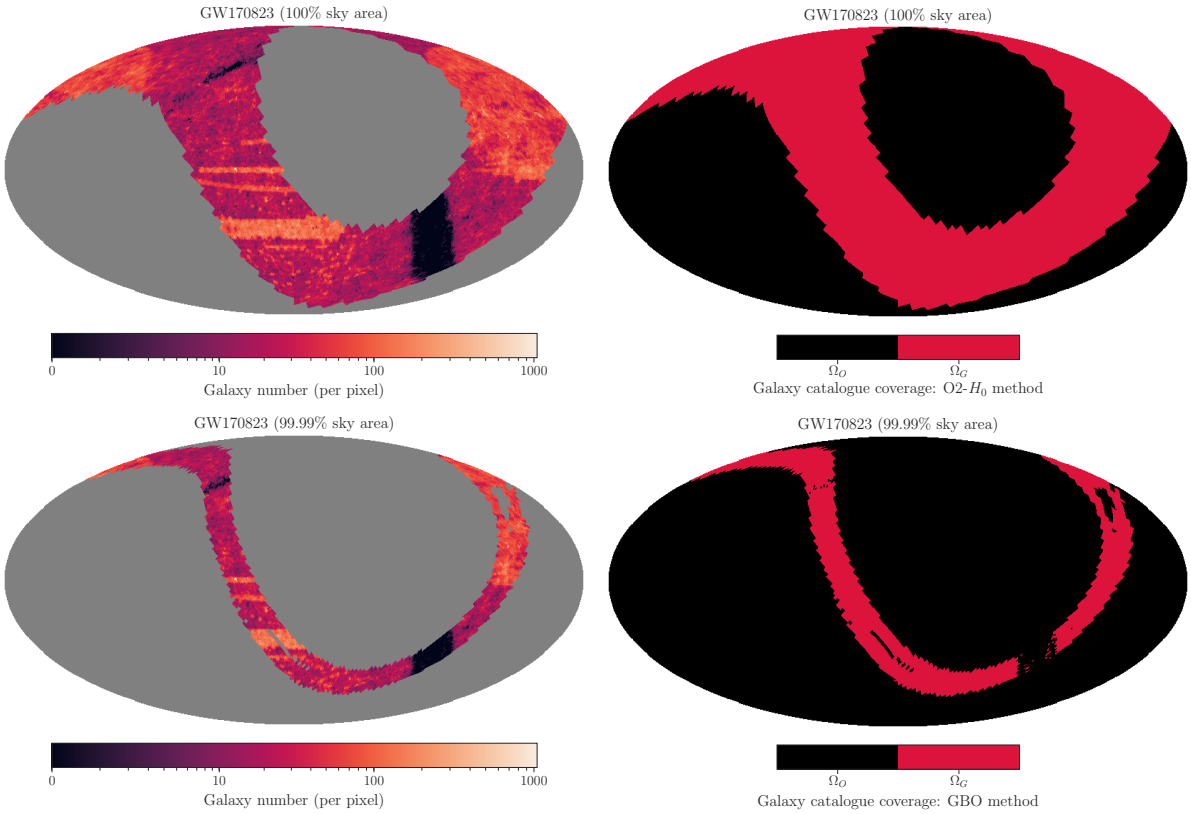

Figure A.8: A comparison of galaxy catalogue handling for GW170823 between the O2-H_0 analysis and the GBO analysis. *Left-hand panels:* Galaxy number per pixel for the GLADE catalogue, for a sky gridded into 49152 equally-sized pixels. The coloured area shows the localisation of GW170823 for the threshold indicated – 100% sky are (top panel) and 99.99% sky area (bottom panel). Darker areas correspond to low galaxy densities, while lighter areas correspond to higher densities. Black patches correspond to empty parts of the catalogue (obscured by the Milky Way band). *Right-hand panels:* In red, the area of the sky defined as being covered by the galaxy catalogue (Ω_G). In black, the area of the sky outside the bounds of the galaxy catalogue (Ω_O). Shown for the O2-H_0 analysis (top) and the GBO analysis (bottom).

Appendix B

Chapter 5 supplementary material

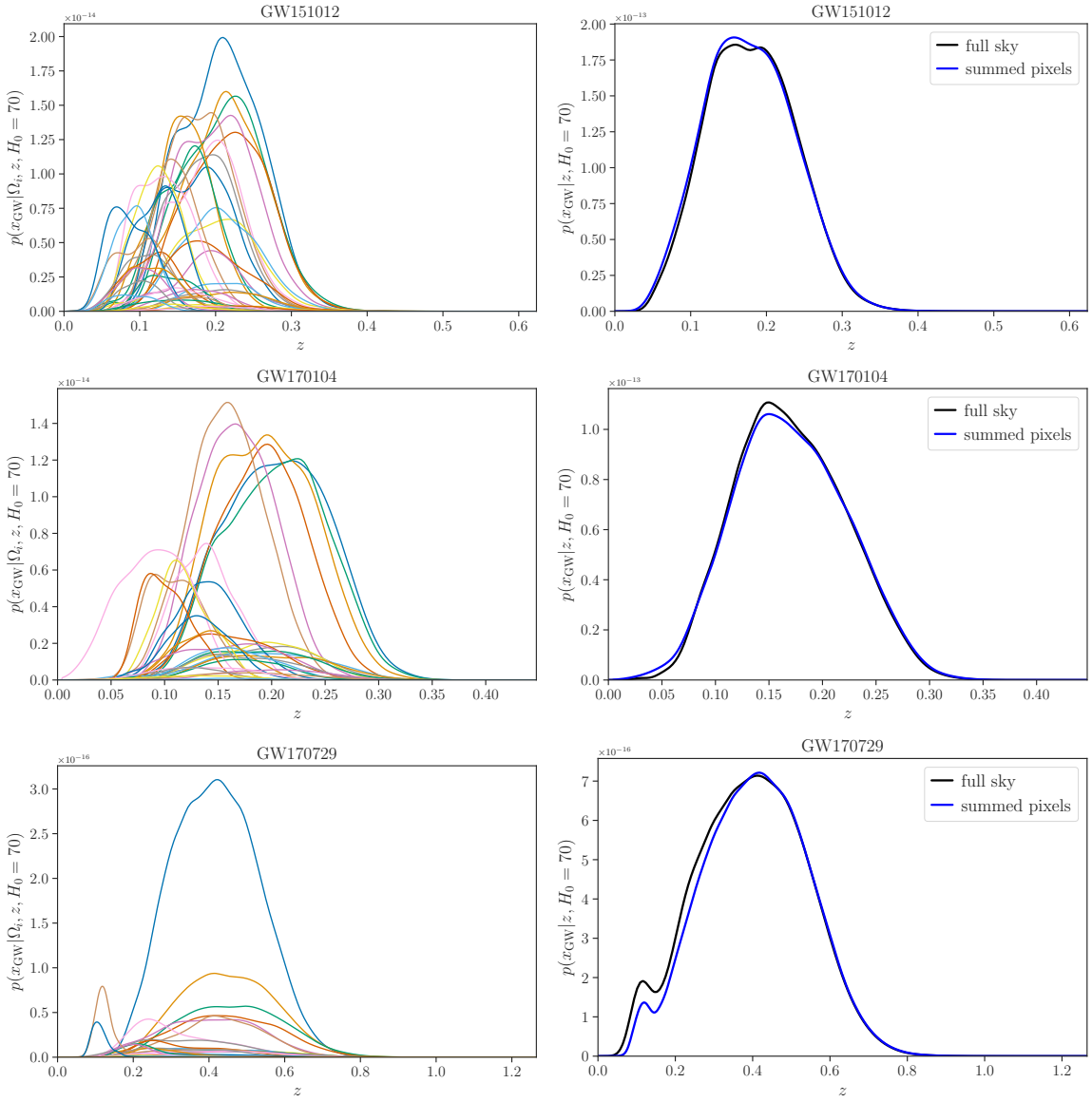

Figure B.1: Line-of-sight redshift estimates for GW151012, GW170104 and GW170729. *Left-hand panels:* The line-of-sight redshift distribution of the event within each pixel. *Right-hand panels:* The full-sky redshift distribution for the event. The black curve shows the estimate from doing a KDE on all the samples, while the blue curve shows the summed curves from the left-hand panel.

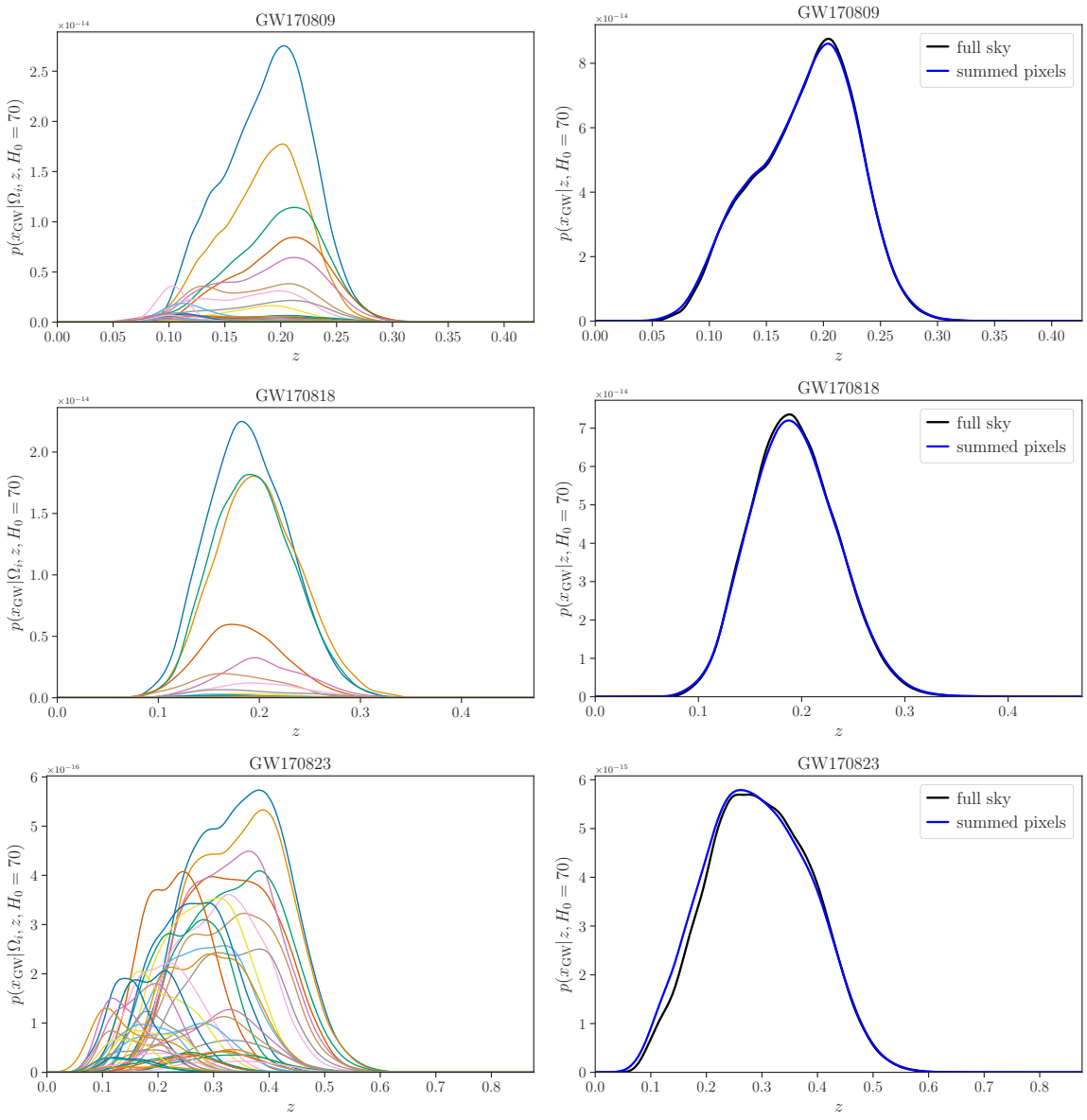

Figure B.2: Line-of-sight redshift estimates for GW170809, GW170818 and GW170823. *Left-hand panels:* The line-of-sight redshift distribution of the event within each pixel. *Right-hand panels:* The full-sky redshift distribution for the event. The black curve shows the estimate from doing a KDE on all the samples, while the blue curve shows the summed curves from the left-hand panel.

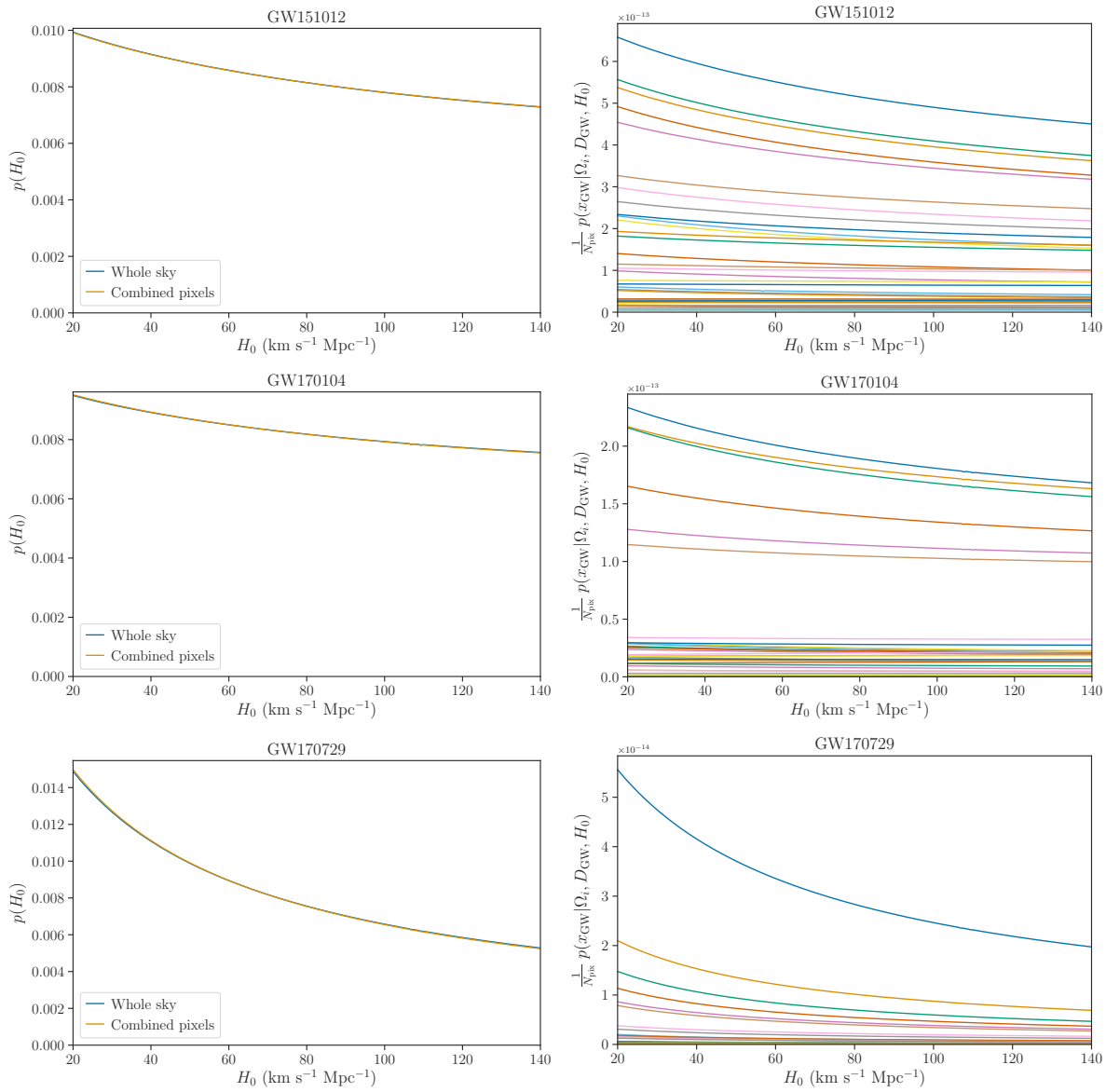

Figure B.3: Likelihoods on H_0 for the empty catalogue analysis with GW151012, GW170104, and GW170729. *Left-hand panels:* Comparison of the (normalised) empty catalogue likelihood on H_0 between the pixelated and non-pixelated (whole sky) cases (orange and blue curves respectively). *Right-hand panels:* A breakdown of the pixelated empty catalogue likelihood by pixel. The pixelated likelihood in the left-hand panel is the sum of the curves in the right-hand panel (then normalised).

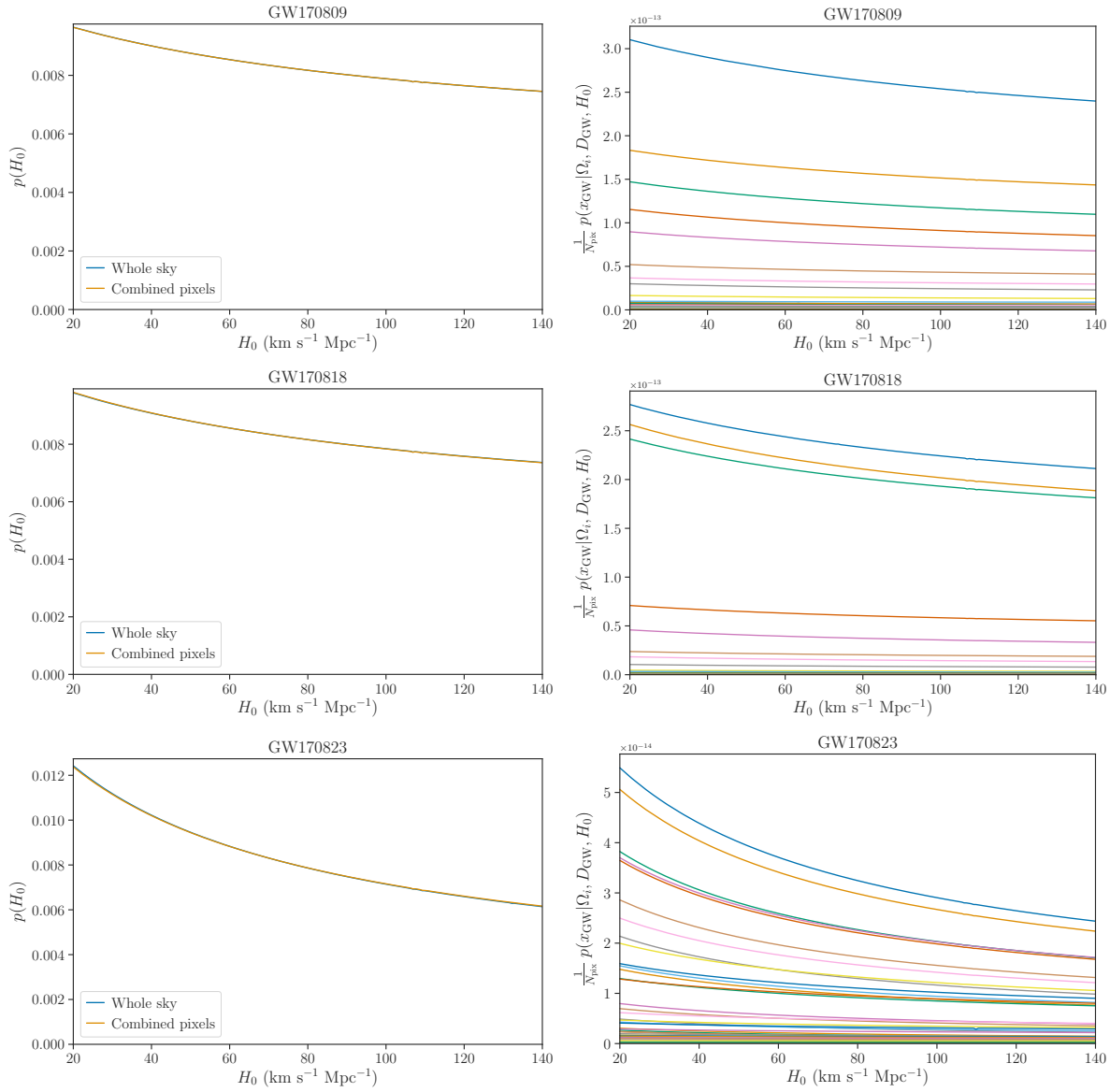

Figure B.4: Likelihoods on H_0 for the empty catalogue analysis with GW170809, GW170818, and GW170823. *Left-hand panels:* Comparison of the (normalised) empty catalogue likelihood on H_0 between the pixelated and non-pixelated (whole sky) cases (orange and blue curves respectively). *Right-hand panels:* A breakdown of the pixelated empty catalogue likelihood by pixel. The pixelated likelihood in the left-hand panel is the sum of the curves in the right-hand panel (then normalised).

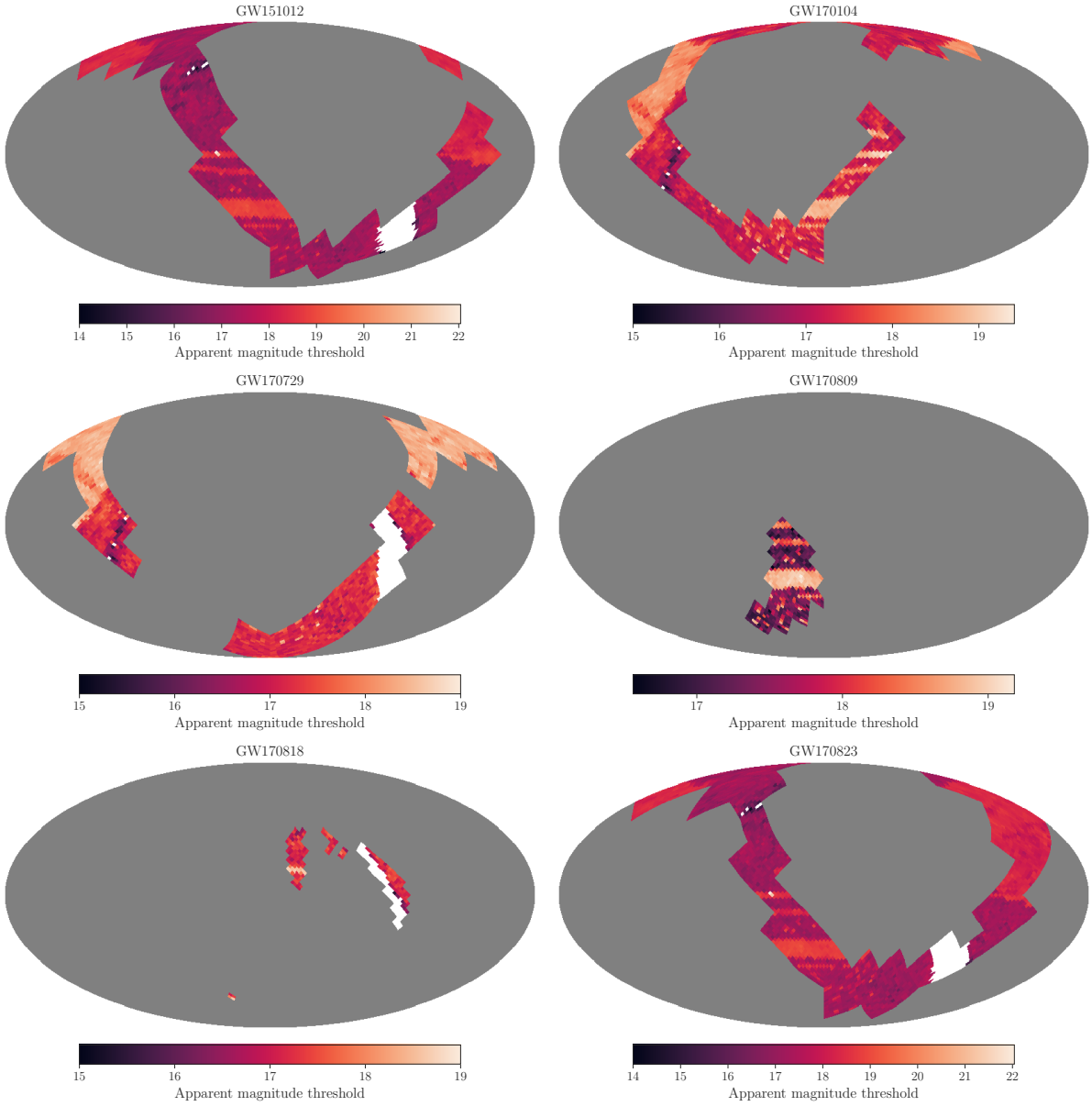

Figure B.5: Variation of the apparent magnitude threshold within the 99.9% sky area of GW151012, GW170104, GW170729, GW170809, GW170818 and GW170823 using the GLADE catalogue. The grey indicates a part of the sky which is not in the 99.9% sky area of the event. The white pixels contain fewer than 10 galaxies, and are taken to be empty.

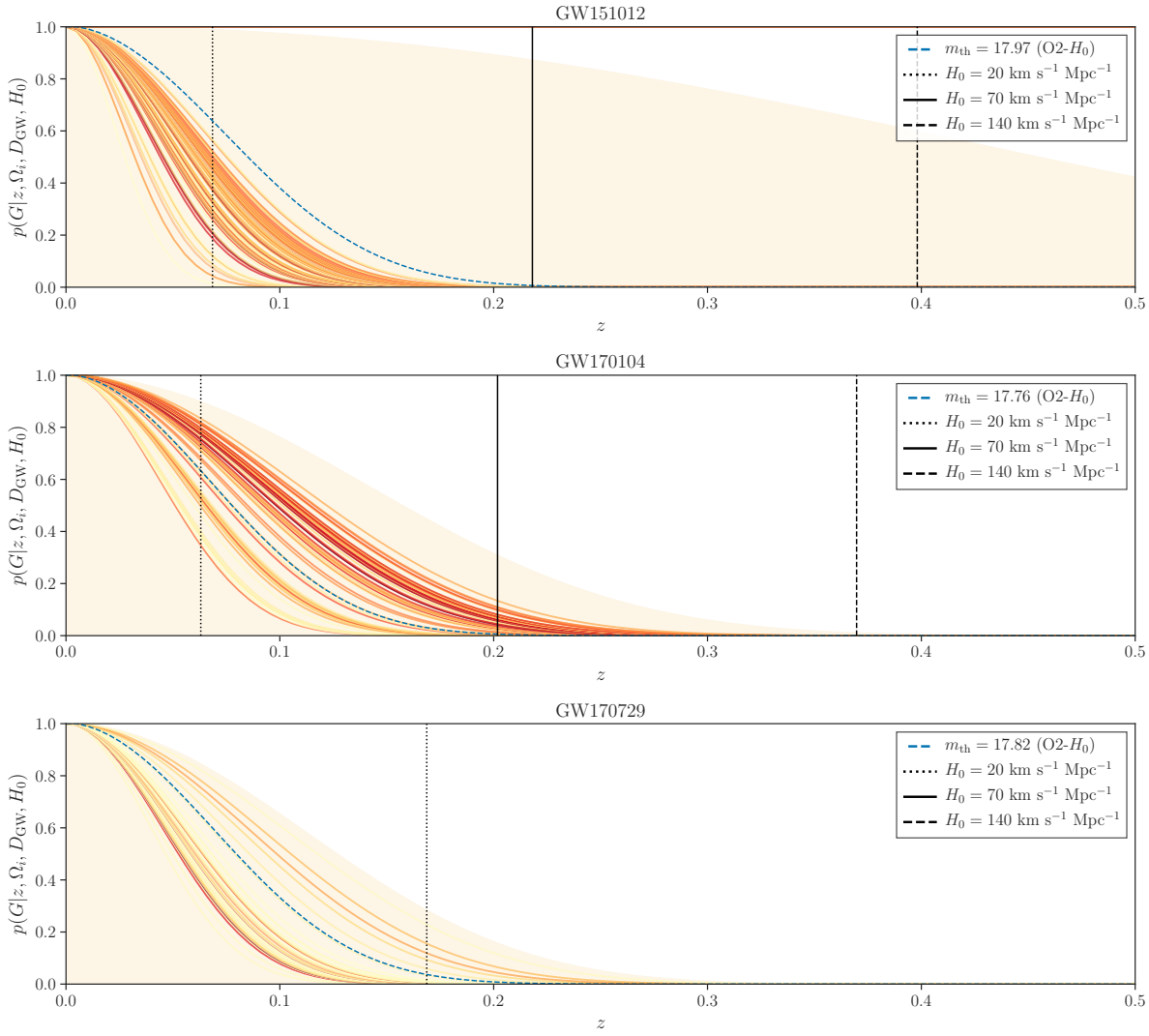

Figure B.6: Probability that the host galaxy is in the galaxy catalogue as a function of redshift, for the pixelated catalogue case. Solid orange curves give probability that the host is inside the galaxy catalogue along the line-of-sight of the pixels which cover the 50% sky-area of the event (darker orange refers to more probable). The shaded yellow area covers the range between the minimum and maximum apparent magnitude threshold within the 99.9% sky-area of the event. Vertical black lines show the median redshift of the event for different values of H_0. The blue dashed line gives the probability that the host is in the galaxy catalogue as it was determined during the O2-H_0 analysis.

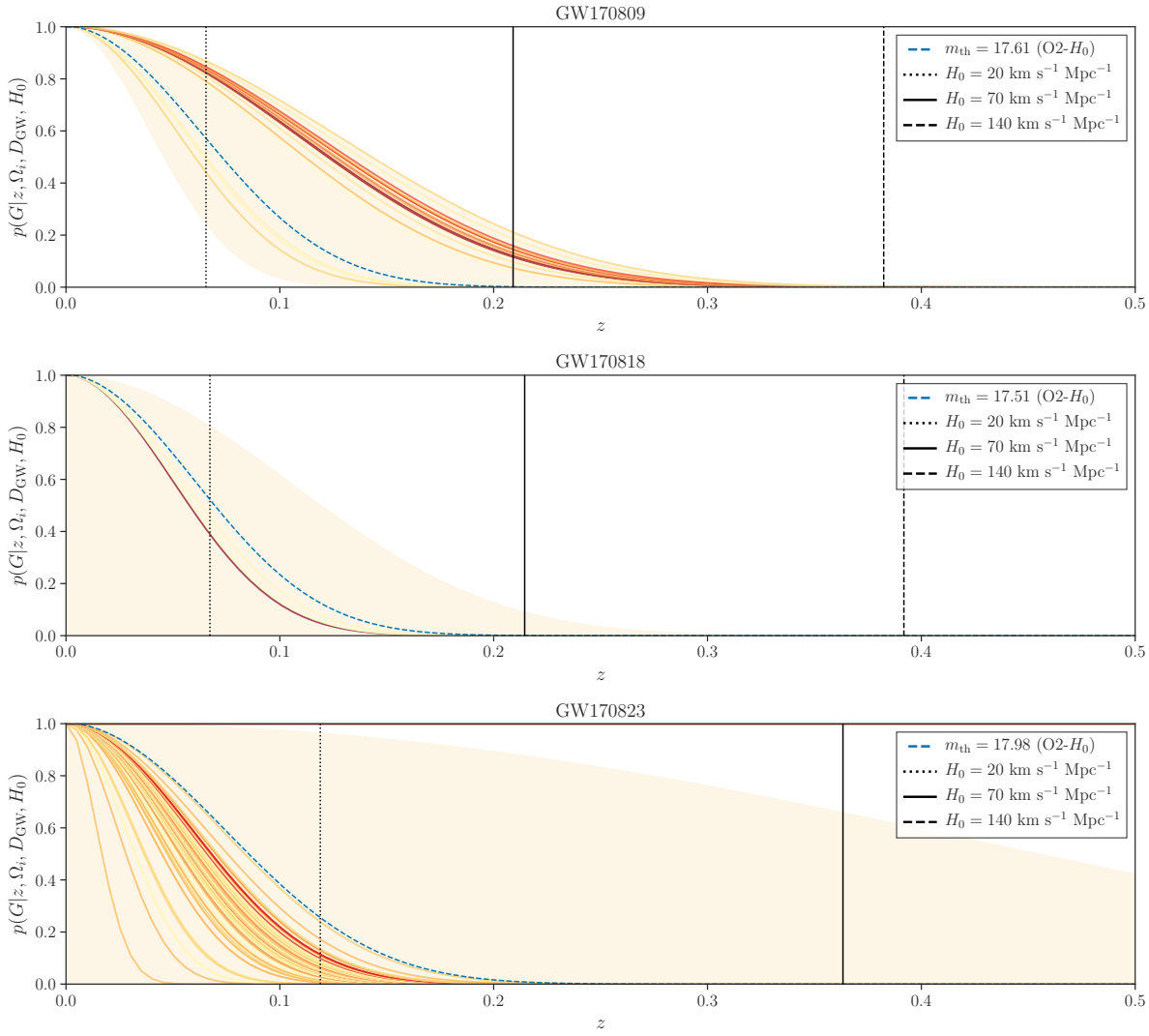

Figure B.7: Probability that the host galaxy is in the galaxy catalogue as a function of redshift, for the pixelated catalogue case. Solid orange curves give probability that the host is inside the galaxy catalogue along the line-of-sight of the pixels which cover the 50% sky-area of the event (darker orange refers to more probable). The shaded yellow area covers the range between the minimum and maximum apparent magnitude threshold within the 99.9% sky-area of the event. Vertical black lines show the median redshift of the event for different values of H_0. The blue dashed line gives the probability that the host is in the galaxy catalogue as it was determined during the O2-H_0 analysis.

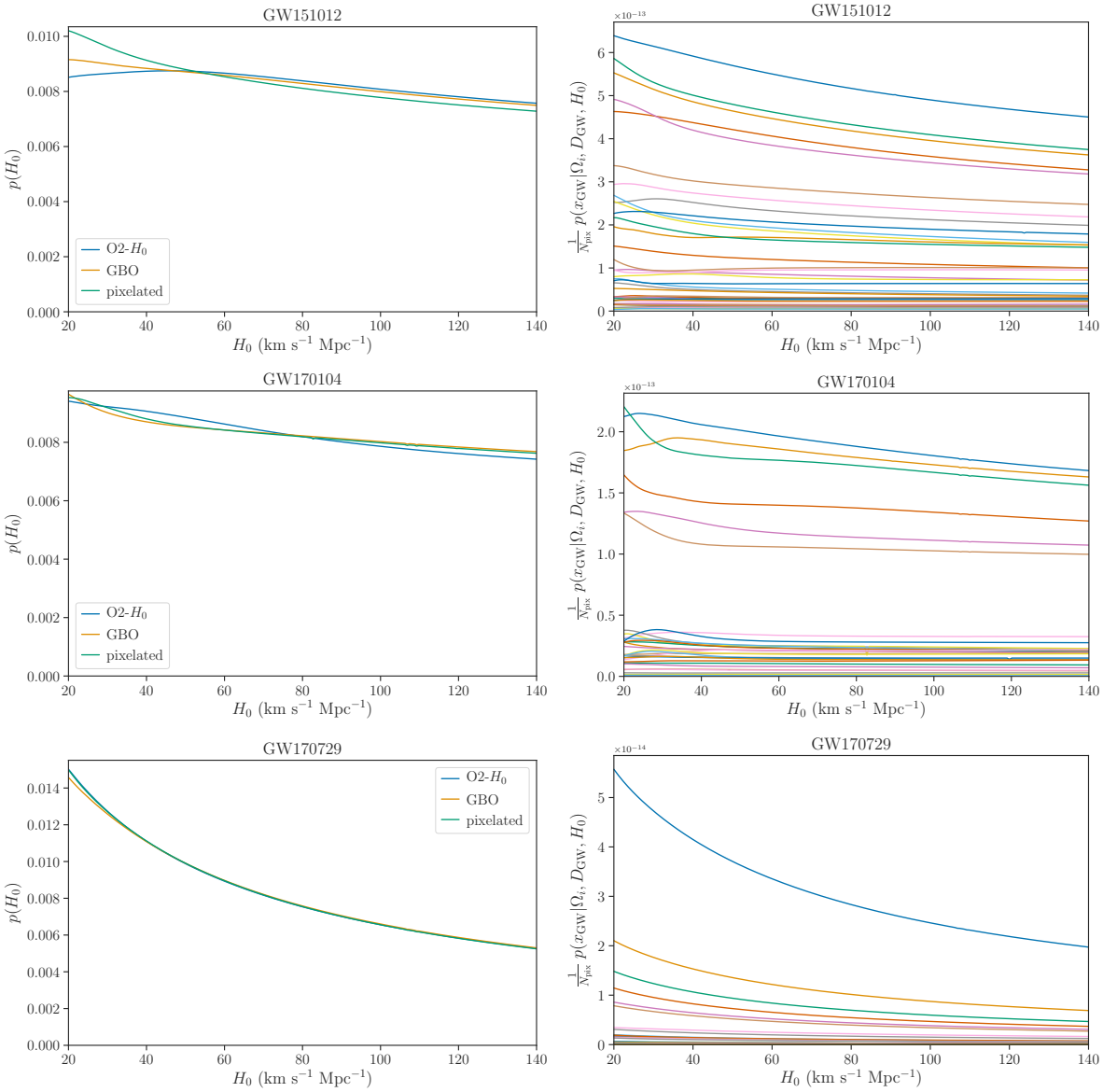

Figure B.8: Likelihoods on H_0 for the pixelated analysis with GW151012, GW170104 and GW170729. *Left-hand panels:* Comparison of the (normalised) likelihoods between the O2-H_0 (blue), GBO (orange) and pixelated (green) methods. *Right-hand panels:* A breakdown of the pixelated likelihood by low resolution pixel. The pixelated likelihood in the left-hand panel is the sum of the curves in the right-hand panel (then normalised).

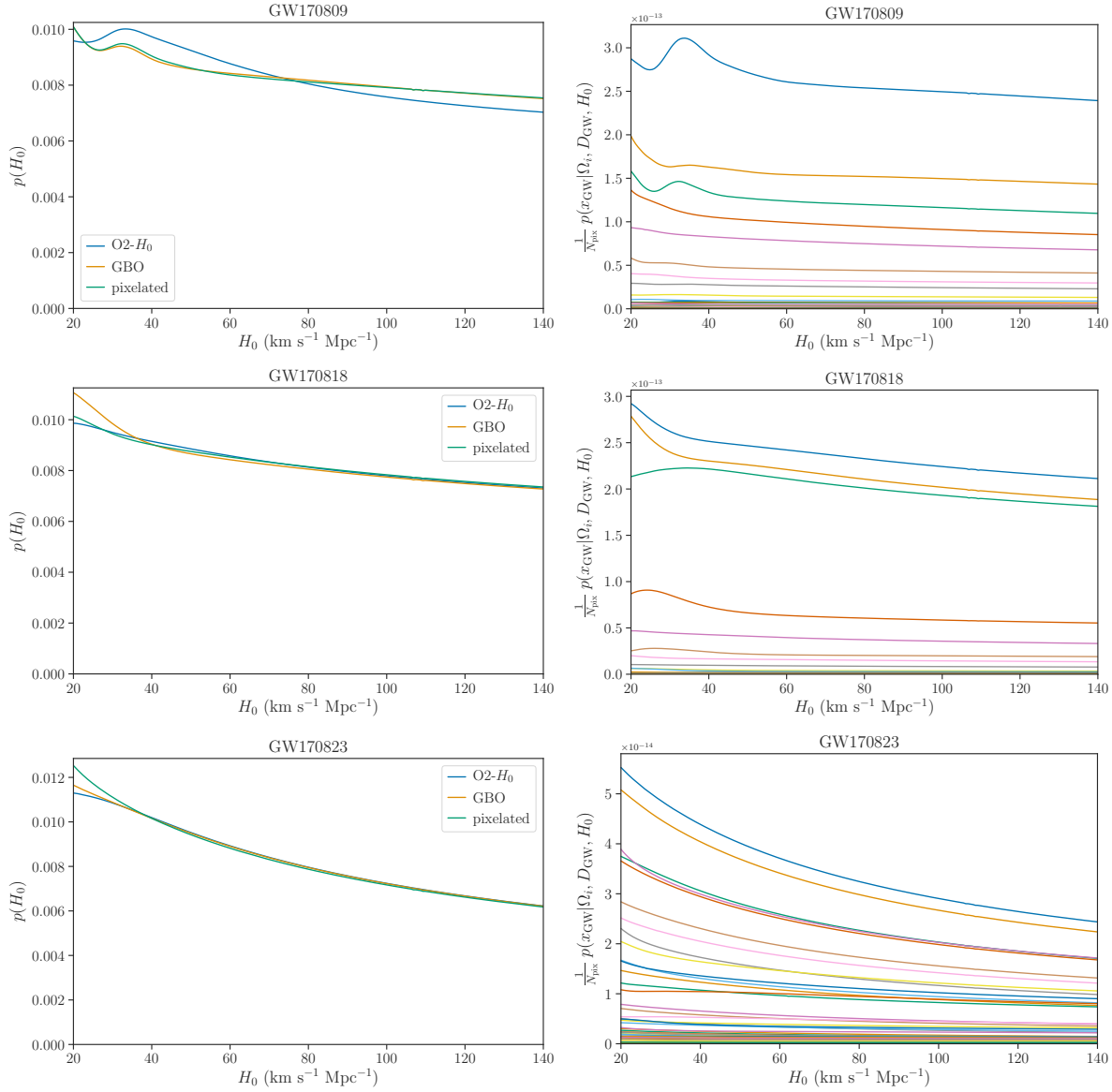

Figure B.9: Likelihoods on H_0 for the pixelated analysis with GW170809, GW170818 and GW170823. *Left-hand panels:* Comparison of the (normalised) likelihoods between the O2-H_0 (blue), GBO (orange) and pixelated (green) methods. *Right-hand panels:* A breakdown of the pixelated likelihood by low resolution pixel. The pixelated likelihood in the left-hand panel is the sum of the curves in the right-hand panel (then normalised).

Bibliography

[1] A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, January 1922. doi:10.1007/BF01332580.

[2] G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59, January 1927.

[3] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. doi:10.1073/pnas.15.3.168.

[4] Planck Collaboration, et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, September 2020. doi:10.1051/0004-6361/201833910.

[5] Adam G. Riess, et al. Cosmic distances calibrated to 1% precision with gaia EDR3 parallaxes and hubble space telescope photometry of 75 milky way cepheids confirm tension with ΛCDM. *Astrophys. J.*, 908(1):L6, feb 2021. doi:10.3847/2041-8213/abdbaf.

[6] Licia Verde, Tommaso Treu, and Adam G. Riess. Tensions between the early and late Universe. *Nature Astronomy*, 3:891–895, September 2019. doi:10.1038/s41550-019-0902-0.

[7] Eleonora Di Valentino, et al. Snowmass2021 - letter of interest cosmology intertwined II: The Hubble constant tension. *Astroparticle Physics*, 131:102605, 2021. doi:https://doi.org/10.1016/j.astropartphys.2021.102605.

[8] B. F. Schutz. Determining the Hubble constant from gravitational wave observations. *Nature*, 323:310, September 1986. doi:10.1038/323310a0.

[9] B. P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi:10.1103/PhysRevLett.116.061102.

[10] B. P. Abbott, et al. GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017. doi:10.1103/PhysRevLett.119.161101.

[11] B. P. Abbott, et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophys. J.*, 848:L12, October 2017. doi:10.3847/2041-8213/aa91c9.

[12] M. Soares-Santos, et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. I. Discovery of the Optical Counterpart Using the Dark Energy Camera. *Astrophys. J. Lett.*, 848:L16, October 2017. doi:10.3847/2041-8213/aa9059.

[13] B. P. Abbott, et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551:85–88, November 2017. doi:10.1038/nature24471.

[14] B. P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9(3):031040, 2019. doi:10.1103/PhysRevX.9.031040.

[15] R. Abbott et al. GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Phys. Rev. X*, 11:021053, Jun 2021. doi:10.1103/PhysRevX.11.021053.

[16] A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, 21(1):326–332, December 1924. doi:10.1007/BF01328280.

[17] G. Lemaître. L'Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, 53:51, January 1933.

[18] H. P. Robertson. Kinematics and World-Structure. *Astrophys. J.*, 82:284, November 1935. doi:10.1086/143681.

[19] H. P. Robertson. Kinematics and World-Structure II. *Astrophys. J.*, 83:187, April 1936. doi:10.1086/143716.

[20] H. P. Robertson. Kinematics and World-Structure III. *Astrophys. J.*, 83:257, May 1936. doi:10.1086/143726.

[21] A. G. Walker. On Milne's Theory of World-Structure. *Proceedings of the London Mathematical Society*, 42:90–127, January 1937. doi:10.1112/plms/s2-42.1.90.

[22] David W. Hogg. Distance measures in cosmology. 1999.

[23] Vivien Bonvin and Martin Millon. Source code available at `https://github.com/vbonvin/H0_tension`.

[24] Simone Aiola, et al. The atacama cosmology telescope: DR4 maps and cosmological parameters. *Journal of Cosmology and Astroparticle Physics*, 2020(12):047–047, dec 2020. doi:10.1088/1475-7516/2020/12/047.

[25] Florian Beutler, et al. The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *Mon. Not. R. Astron. Soc.*, 416(4):3017–3032, 09 2011. doi:10.1111/j.1365-2966.2011.19250.x.

[26] Ashley J. Ross, et al. The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at z = 0.15. *Mon. Not. R. Astron. Soc.*, 449(1):835–847, 03 2015. doi:10.1093/mnras/stv154.

[27] Shadab Alam, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Mon. Not. R. Astron. Soc.*, 470(3):2617–2652, 03 2017. doi:10.1093/mnras/stx721.

[28] G. E. Addison, et al. Elucidating ΛCDM: Impact of baryon acoustic oscillation measurements on the Hubble constant discrepancy. *Astrophys. J.*, 853(2):119, jan 2018. doi:10.3847/1538-4357/aaa1ed.

[29] Andrei Cuceu, James Farr, Pablo Lemos, and Andreu Font-Ribera. Baryon acoustic oscillations and the Hubble constant: past, present and future. *Journal of Cosmology and Astroparticle Physics*, 2019(10):044–044, oct 2019. doi:10.1088/1475-7516/2019/10/044.

[30] Adam G. Riess, et al. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. *Astrophys. J.*, 876(1):85, 2019. doi:10.3847/1538-4357/ab1422.

[31] Kenneth C Wong, et al. H0LiCOW – XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-Universe probes. *Mon. Not. R. Astron. Soc.*, 498(1):1420–1439, 09 2019. doi:10.1093/mnras/stz3094.

[32] D. W. Pesce, et al. The megamaser cosmology project. XIII. combined Hubble constant constraints. *Astrophys. J.*, 891(1):L1, feb 2020. doi:10.3847/2041-8213/ab75f0.

[33] Wendy L. Freedman, et al. The Carnegie-Chicago Hubble Program. VIII. an independent determination of the Hubble constant based on the tip of the red giant branch. *Astrophys. J.*, 882(1):34, aug 2019. doi:10.3847/1538-4357/ab2f73.

[34] Wenlong Yuan, et al. Consistent calibration of the tip of the red giant branch in the large magellanic cloud on the Hubble space telescope photometric system and a redetermination of the Hubble constant. *Astrophys. J.*, 886(1):61, nov 2019. doi:10.3847/1538-4357/ab4bc9.

[35] Planck Collaboration, et al. Planck 2015 results. XIV. Dark energy and modified gravity. *Astron. Astrophys.*, 594:A14, September 2016. doi:10.1051/0004-6361/201525814.

[36] Eleonora Di Valentino, Alessandro Melchiorri, and Joseph Silk. Cosmological hints of modified gravity? *Phys. Rev. D*, 93(2):023513, January 2016. doi:10.1103/PhysRevD.93.023513.

[37] Eleonora Di Valentino, Alessandro Melchiorri, Olga Mena, and Sunny Vagnozzi. Interacting dark energy in the early 2020s: A promising solution to the H_0 and cosmic shear tensions. *Physics of the Dark Universe*, 30:100666, December 2020. doi:10.1016/j.dark.2020.100666.

[38] Christina D. Kreisch, Francis-Yan Cyr-Racine, and Olivier Doré. Neutrino puzzle: Anomalies, interactions, and cosmological tensions. *Phys. Rev. D*, 101:123505, Jun 2020. doi:10.1103/PhysRevD.101.123505.

[39] Eleonora Di Valentino, Alessandro Melchiorri, and Joseph Silk. Planck evidence for a closed Universe and a possible crisis for cosmology. *Nature Astronomy*, 4:196–203, February 2020. doi:10.1038/s41550-019-0906-9.

[40] Will Handley. Curvature tension: Evidence for a closed universe. *Phys. Rev. D*, 103:L041301, Feb 2021. doi:10.1103/PhysRevD.103.L041301.

[41] Albert Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, pages 688–696, January 1916.

[42] Albert Einstein. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, pages 844–847, January 1915.

[43] F. A. Pirani. Invariant Formulation of Gravitational Radiation Theory. *Phys. Rev.*, 105(3):1089–1099, February 1957. doi:10.1103/PhysRev.105.1089.

[44] Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev.*, D47:2198–2219, 1993. doi:10.1103/PhysRevD.47.2198.

[45] Warren G. Anderson, Patrick R. Brady, Jolien D. Creighton, and Éanna É. Flanagan. Excess power statistic for detection of burst sources of gravitational radiation. *Phys. Rev. D*, 63(4):042003, February 2001. doi:10.1103/PhysRevD.63.042003.

[46] S. A. Usman, J. C. Mills, and S. Fairhurst. Constraining the Inclinations of Binary Mergers from Gravitational-wave Observations. *Astrophys. J.*, 877(2):82, June 2019. doi:10.3847/1538-4357/ab0b3e.

[47] Stephen Fairhurst. Triangulation of gravitational wave sources with a network of detectors. *New Journal of Physics*, 11(12):123006, December 2009. doi:10.1088/1367-2630/11/12/123006.

[48] J. Veitch, et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4):042003, February 2015. doi:10.1103/PhysRevD.91.042003.

[49] W. Del Pozzo. Inference of cosmological parameters from gravitational waves: Applications to second generation interferometers. *Phys. Rev. D*, 86(4):043011, August 2012. doi:10.1103/PhysRevD.86.043011.

[50] Li-Xin Li and Bohdan Paczyński. Transient Events from Neutron Star Mergers. *Astrophys. J. Lett.*, 507(1):L59–L62, November 1998. doi:10.1086/311680.

[51] B. D. Metzger, et al. Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei. *Mon. Not. R. Astron. Soc.*, 406(4):2650–2662, 08 2010. doi:10.1111/j.1365-2966.2010.16864.x.

[52] W. Fong and E. Berger. The Locations of Short Gamma-Ray Bursts as Evidence for Compact Object Binary Progenitors. *Astrophys. J.*, 776(1):18, October 2013. doi:10.1088/0004-637X/776/1/18.

[53] Neal Dalal, Daniel E. Holz, Scott A. Hughes, and Bhuvnesh Jain. Short grb and binary black hole standard sirens as a probe of dark energy. *Phys. Rev.*, D74:063006, 2006. doi:10.1103/PhysRevD.74.063006.

[54] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck. Cosmography with the Einstein Telescope. *Class. Quant. Grav.*, 27:215006, 2010. doi:10.1088/0264-9381/27/21/215006.

[55] Samaya Nissanke, et al. Exploring short gamma-ray bursts as gravitational-wave standard sirens. *Astrophys. J.*, 725:496–514, 2010. doi:10.1088/0004-637X/725/1/496.

[56] Samaya Nissanke, et al. Determining the Hubble constant from gravitational wave observations of merging compact binaries. 2013.

[57] Suvodip Mukherjee, et al. Velocity correction for Hubble constant measurements from standard sirens. *Astron. Astrophys.*, 646:A65, 2021. doi:10.1051/0004-6361/201936724.

[58] Constantina Nicolaou, et al. The impact of peculiar velocities on the estimation of the Hubble constant from gravitational wave standard sirens. *Mon. Not. R. Astron. Soc.*, 495(1):90–97, 04 2020. doi:10.1093/mnras/staa1120.

[59] C. Guidorzi et al. Improved Constraints on H_0 from a Combined Analysis of Gravitational-wave and Electromagnetic Emission from GW170817. *Astrophys. J.*, 851(2):L36, 2017. doi:10.3847/2041-8213/aaa009.

[60] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562(7728):545–547, October 2018. doi:10.1038/s41586-018-0606-0.

[61] Stephen M. Feeney, et al. Prospects for resolving the Hubble constant tension with standard sirens. *Phys. Rev. Lett.*, 122(6):061105, 2019. doi:10.1103/PhysRevLett.122.061105.

[62] Daniel J. Mortlock, et al. Unbiased hubble constant estimation from binary neutron star mergers. *Phys. Rev. D*, 100:103523, Nov 2019. doi:10.1103/PhysRevD.100.103523.

[63] M. Fishbach, et al. A standard siren measurement of the Hubble constant from GW170817 without the electromagnetic counterpart. *Astrophys. J.*, 871(1):L13, 2019. doi:10.3847/2041-8213/aaf96e.

[64] Rachel Gray et al. Cosmological inference using gravitational wave standard sirens: A mock data analysis. *Phys. Rev. D*, 101(12):122001, 2020. doi:10.1103/PhysRevD.101.122001.

[65] M. Soares-Santos et al. First measurement of the Hubble constant from a dark standard siren using the Dark Energy Survey galaxies and the LIGO/Virgo binary-black-hole merger GW170814. *Astrophys. J.*, 876(1):L7, 2019. doi:10.3847/2041-8213/ab14f1.

[66] B. P. Abbott et al. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *Astrophys. J.*, 909(2):218, March 2021. doi:10.3847/1538-4357/abdcb7.

[67] C. Messenger and J. Read. Measuring a cosmological distance-redshift relationship using only gravitational wave observations of binary neutron star coalescences. *Phys. Rev. Lett.*, 108:091101, 2012. doi:10.1103/PhysRevLett.108.091101.

[68] Lee Samuel Finn. Observational constraints on the neutron star mass distribution. *Phys. Rev. Lett.*, 73:1878–1881, 1994. doi:10.1103/PhysRevLett.73.1878.

[69] Stephen R. Taylor, Jonathan R. Gair, and Ilya Mandel. Hubble without the Hubble: Cosmology using advanced gravitational-wave detectors alone. *Phys. Rev.*, D85:023535, 2012. doi:10.1103/PhysRevD.85.023535.

[70] Stephen R. Taylor and Jonathan R. Gair. Cosmology with the lights off: standard sirens in the Einstein Telescope era. *Phys. Rev.*, D86:023502, 2012. doi:10.1103/PhysRevD.86.023502.

[71] B. P. Abbott et al. GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4\ M_{\odot}$. *Astrophys. J. Lett.*, 892(1):L3, March 2020. doi:10.3847/2041-8213/ab75f5.

[72] Will M. Farr, Maya Fishbach, Jiani Ye, and Daniel Holz. A Future Percent-Level Measurement of the Hubble Expansion at Redshift 0.8 With Advanced LIGO. *Astrophys. J. Lett.*, 883:L42, 2019. doi:10.3847/2041-8213/ab4284.

[73] David F. Chernoff and Lee S. Finn. Gravitational Radiation, Inspiraling Binaries, and Cosmology. *Astrophys. J. Lett.*, 411:L5, July 1993. doi:10.1086/186898.

[74] Lee Samuel Finn. Binary inspiral, gravitational radiation, and cosmology. *Phys. Rev.*, D53:2878–2894, 1996. doi:10.1103/PhysRevD.53.2878.

[75] Maya Fishbach, Daniel E. Holz, and Will M. Farr. Does the Black Hole Merger Rate Evolve with Redshift? *Astrophys. J.*, 863(2):L41, 2018. doi:10.3847/2041-8213/aad800.

[76] L. P. Singer, et al. The First Two Years of Electromagnetic Follow-up with Advanced LIGO and Virgo. *Astrophys. J.*, 795:105, November 2014. doi:10.1088/0004-637X/795/2/105.

[77] Christopher P. L. Berry et al. Parameter estimation for binary neutron-star coalescences with realistic noise during the Advanced LIGO era. *Astrophys. J.*, 804(2):114, 2015. doi:10.1088/0004-637X/804/2/114.

[78] Cody Messick, et al. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95:042001, February 2017. doi:10.1103/PhysRevD.95.042001.

[79] W. Del Pozzo, et al. Dirichlet Process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations. *Mon. Not. R. Astron. Soc.*, 479(1):601–614, 2018. doi:10.1093/mnras/sty1485.

[80] Gergely Dálya, et al. GLADE: A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era. *Mon. Not. R. Astron. Soc.*, 479(2):2374–2381, 2018. doi:10.1093/mnras/sty1703.

[81] P. Schechter. An analytic expression for the luminosity function for galaxies. *Astrophys. J.*, 203:297–306, January 1976. doi:10.1086/154079.

[82] Leo P. Singer et al. Going the Distance: Mapping Host Galaxies of LIGO and Virgo Sources in Three Dimensions Using Local Cosmography and Targeted Follow-up. *Astrophys. J.*, 829(1):L15, 2016. doi:10.3847/2041-8205/829/1/L15.

[83] C. N. Leibler and E. Berger. The Stellar Ages and Masses of Short Gamma-ray Burst Host Galaxies: Investigating the Progenitor Delay Time Distribution and the Role of Mass and Star Formation in the Short Gamma-ray Burst Rate. *Astrophys. J.*, 725(1):1202–1214, December 2010. doi:10.1088/0004-637X/725/1/1202.

[84] W. Fong, et al. Demographics of the Galaxies Hosting Short-duration Gamma-Ray Bursts. *Astrophys. J.*, 769(1):56, May 2013. doi:10.1088/0004-637X/769/1/56.

[85] Stanislaw Weglarczyk. Kernel density estimation and its application. *ITM Web of Conferences*, 23:00037, 01 2018. doi:10.1051/itmconf/20182300037.

[86] Pauli Virtanen, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi:10.1038/s41592-019-0686-2.

[87] Curt Cutler and Éanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D*, 49(6):2658–2697, Mar 1994. doi:10.1103/physrevd.49.2658.

[88] A. M. Walker. On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(1):80–88, 1969.

[89] Chelsea L. MacLeod and Craig J. Hogan. Precision of Hubble constant derived using black hole binary absolute distances and statistical redshift information. *Phys. Rev.*, D77:043512, 2008. doi:10.1103/PhysRevD.77.043512.

[90] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016. doi:10.1051/0004-6361/201525830.

[91] B. P. Abbott et al. Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo. *Astrophys. J. Lett.*, 882(2):L24, 2019. doi:10.3847/2041-8213/ab3800.

[92] A. Drlica-Wagner et al. Dark Energy Survey Year 1 Results: Photometric Data Set for Cosmology. *Astrophys. J. Suppl.*, 235(2):33, 2018. doi:10.3847/1538-4365/aab4f5.

[93] T. M. C. Abbott et al. The Dark Energy Survey Data Release 1. *Astrophys. J. Suppl.*, 239(2):18, 2018. doi:10.3847/1538-4365/aae9f0.

[94] Mara Salvato, Olivier Ilbert, and Ben Hoyle. The many flavours of photometric redshifts. *Nature Astronomy*, 3:212–222, June 2019. doi:10.1038/s41550-018-0478-0.

[95] Juan De Vicente, Eusebio Sánchez, and Ignacio Sevilla-Noarbe. DNF – Galaxy photometric redshift by Directional Neighbourhood Fitting. *Mon. Not. R. Astron. Soc.*, 459(3):3078–3088, 2016. doi:10.1093/mnras/stw857.

[96] Iftach Sadeh, Filipe B. Abdalla, and Ofer Lahav. ANNz2 - photometric redshift and probability distribution function estimation using machine learning. *Publ. Astron. Soc. Pac.*, 128(968):104502, 2016. doi:10.1088/1538-3873/128/968/104502.

[97] J. B. Oke and Allan Sandage. Energy Distributions, K Corrections, and the Stebbins-Whitford Effect for Giant Elliptical Galaxies. *Astrophys. J.*, 154:21, October 1968. doi:10.1086/149737.

[98] David W. Hogg, Ivan K. Baldry, Michael R. Blanton, and Daniel J. Eisenstein. The K correction. October 2002.

[99] Darren J. White, E. J. Daw, and V. S. Dhillon. A List of Galaxies for Gravitational Wave Searches. *Class. Quant. Grav.*, 28:085016, 2011. doi:10.1088/0264-9381/28/8/085016.

[100] Maciej Bilicki, et al. 2MASS Photometric Redshift catalog: a comprehensive three-dimensional census of the whole sky. *Astrophys. J. Suppl.*, 210:9, 2014. doi:10.1088/0067-0049/210/1/9.

[101] M. F. Skrutskie et al. The Two Micron All Sky Survey (2MASS). *Astron. J.*, 131:1163–1183, 2006. doi:10.1086/498708.

[102] Dmitry Makarov, et al. HyperLEDA. III. The catalogue of extragalactic distances. *Astron. Astrophys.*, 570:A13, 2014. doi:10.1051/0004-6361/201423496.

[103] Isabelle Pâris et al. The Sloan Digital Sky Survey Quasar Catalog: twelfth data release. *Astron. Astrophys.*, 597:A79, 2017. doi:10.1051/0004-6361/201527999.

[104] Jonathan Carrick, Stephen J. Turnbull, Guilhem Lavaux, and Michael J. Hudson. Cosmological parameters from the comparison of peculiar velocities with predictions from the 2M++ density field. *Mon. Not. R. Astron. Soc.*, 450(1):317–332, 2015. doi:10.1093/mnras/stv547.

[105] Planck Collaboration, et al. Planck 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove. *Astron. Astrophys.*, 571:A27, November 2014. doi:10.1051/0004-6361/201321556.

[106] I. Sevilla-Noarbe et al. Star–galaxy classification in the Dark Energy Survey Y1 data set. *Mon. Not. Roy. Astron. Soc.*, 481(4):5451–5469, 2018. doi:10.1093/mnras/sty2579.

[107] B. Hoyle et al. Dark Energy Survey Year 1 Results: Redshift distributions of the weak lensing source galaxies. *Mon. Not. Roy. Astron. Soc.*, 478(1):592–610, 2018. doi:10.1093/mnras/sty957.

[108] Igor V. Chilingarian, Anne-Laure Melchior, and Ivan Yu. Zolotukhin. Analytical approximations of K-corrections in optical and near-infrared bands. *Mon. Not. R. Astron. Soc.*, 405(3):1409–1420, 06 2010. doi:10.1111/j.1365-2966.2010.16506.x.

[109] Igor V. Chilingarian and Ivan Yu. Zolotukhin. A universal ultraviolet-optical colour-colour-magnitude relation of galaxies. *Mon. Not. R. Astron. Soc.*, 419(2):1727–1739, January 2012. doi:10.1111/j.1365-2966.2011.19837.x.

[110] David Caditz and Vahe Petrosian. Cosmological Parameters and Evolution of the Galaxy Luminosity Function. *Astrophys. J. Lett.*, 337:L65, Feb 1989. doi:10.1086/185379.

[111] Darren S. Madgwick, et al. The 2dF Galaxy Redshift Survey: galaxy luminosity functions per spectral type. *Mon. Not. R. Astron. Soc.*, 333(1):133–144, Jun 2002. doi:10.1046/j.1365-8711.2002.05393.x.

[112] Neil Gehrels, et al. Galaxy Strategy for LIGO-Virgo Gravitational Wave Counterpart Searches. *Astrophys. J.*, 820(2):136, 2016. doi:10.3847/0004-637X/820/2/136.

[113] Michael R. Blanton, et al. The galaxy luminosity function and luminosity density at redshiftz= 0.1. *Astrophys. J.*, 592(2):819–838, aug 2003. doi:10.1086/375776.

[114] Bülent Kiziltan, Athanasios Kottas, Maria De Yoreo, and Stephen E. Thorsett. The Neutron Star Mass Distribution. *Astrophys. J.*, 778:66, 2013. doi:10.1088/0004-637X/778/1/66.

[115] B. P. Abbott et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 21(1):3, 2018. doi:10.1007/s41114-018-0012-9.

[116] William A. Fowler and F. Hoyle. Neutrino Processes and Pair Formation in Massive Stars and Supernovae. *Astrophys. J. Suppl.*, 9:201–319, 1964. doi:10.1086/190103.

[117] W. Saunders, et al. The 60-μ and far-infrared luminosity functions of IRAS galaxies. *Mon. Not. R. Astron. Soc.*, 242(3):318–337, 06 1990. doi:10.1093/mnras/242.3.318.

[118] Ilya Mandel and Selma E. de Mink. Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries. *Mon. Not. Roy. Astron. Soc.*, 458(3):2634–2647, 2016. doi:10.1093/mnras/stw379.

[119] Sascha Husa, et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev.*, D93(4):044006, 2016. doi:10.1103/PhysRevD.93.044006.

[120] Sebastian Khan, et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev.*, D93(4):044007, 2016. doi:10.1103/PhysRevD.93.044007.

[121] Mark Hannam, et al. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.*, 113(15):151101, 2014. doi:10.1103/PhysRevLett.113.151101.

[122] Yi Pan, et al. Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism. *Phys. Rev.*, D89(8):084006, 2014. doi:10.1103/PhysRevD.89.084006.

[123] Andrea Taracchini et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev.*, D89(6):061502, 2014. doi:10.1103/PhysRevD.89.061502.

[124] Stanislav Babak, Andrea Taracchini, and Alessandra Buonanno. Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity. *Phys. Rev.*, D95(2):024010, 2017. doi:10.1103/PhysRevD.95.024010.

[125] B. P. Abbott et al. Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *Phys. Rev. X*, 6(4):041015, 2016. doi:10.1103/PhysRevX.6.041015. [Erratum: Phys.Rev.X 8, 039903 (2018)].

[126] Benjamin P. Abbott et al. Effects of waveform model systematics on the interpretation of GW150914. *Class. Quant. Grav.*, 34(10):104002, 2017. doi:10.1088/1361-6382/aa6854.

[127] Craig Cahillane, et al. Calibration uncertainty for advanced ligo's first and second observing runs. *Phys. Rev. D*, 96:102001, Nov 2017. doi:10.1103/PhysRevD.96.102001.

[128] Andrea Zonca, et al. healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python. *Journal of Open Source Software*, 4(35):1298, March 2019. doi:10.21105/joss.01298.

[129] K. M. Górski, et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *Astrophys. J.*, 622:759–771, April 2005. doi:10.1086/427976.

[130] S. Mastrogiovanni, et al. Cosmology in the dark: On the importance of source population models for gravitational-wave cosmology. *arXiv e-prints*, page arXiv:2103.14663, March 2021.

[131] Željko Ivezić, et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.*, 873(2):111, March 2019. doi:10.3847/1538-4357/ab042c.

[132] M. J. Graham, et al. Candidate electromagnetic counterpart to the binary black hole merger gravitational-wave event S190521g. *Phys. Rev. Lett.*, 124:251102, Jun 2020. doi:10.1103/PhysRevLett.124.251102.