Enhancing Searches for Gravitational Waves from Short Transient Bursts

On the application of Gaussian mixture modelling as a noise mitigation technique to un-modelled search algorithms

Leigh Smith

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

October 2024

Abstract

The advanced detector era of gravitational wave (GW) searches has been highly successful in the detection of compact binary coalescence (CBC) events throughout three complete observing runs, while more detections are expected with the fourth run currently underway. As the search for GWs continues with increased detector sensitivities, it is expected for sources beyond CBCs to be detected. One example of an expected yet so far undetected source are short transient GW bursts; this signal type encapsulates a wide range of astrophysical sources with varying signal morphologies. The detection of transient bursts often relies upon search algorithms which hold minimal assumptions on a given signals morphology, referred to as un-modelled searches.

One particular un-modelled search is the coherent WaveBurst (cWB) algorithm, which bases the detection of GWs upon excess coherent energy across a network of detectors. The weakly-modelled nature of cWB makes it sensitive to a wide range of transient GW signals, however also makes it highly susceptible to spurious transient noise artefacts known as glitches. Glitches directly effect the detection capabilities of searches, and the employment of noise mitigation techniques within searches is required to separate them from GW signals. The development of such noise mitigation techniques is crucial in optimising the sensitivities of searches as new detector upgrades introduce new sources of glitches into the data. The work presented here explores the enhancement of Gaussian mixture modelling (GMM) as a noise mitigation tool to the cWB algorithm in the search for short-duration GW transients. GMM allows for the populations of noise and signal to be modelled over a set of representative attributes, aiding in the classification of GW signals against glitches.

Investigations into various aspects of model training are executed in order to increase the reliability of the GMM methodology. Specifically, the analysis is made robust to a wide range of signal morphologies by removing a bias in the training set, and new approaches to optimise models are chosen to increase the accuracy of the analysis. Through initial tests, we show that the modification the GMM methodology also results in increased sensitivity to a selection of expected burst sources. Following this result, the performance of the enhanced cWB+GMM algorithm is further evaluated through extensive testing with data from the third LIGO-Virgo-KAGRA (LVK) observing run. For both 2- and 3-detector networks, we show that the GMM methodology obtains significant sensitivity improvements for Gaussian pulse and cosmic string waveforms compared to those obtained by previous cWB post-production methodologies. Thus,

we demonstrate the ability of GMM to effectively mitigate the dominant source of noise for un-modelled searches: blip glitches. Through these tests it also shown that with the GMM methodology, the 3-detector network achieves sensitivities similar to those of the 2-detector network, which has not previously been possible with the cWB algorithm due to high glitch rates. Further comparison of the GMM method with other post-production pipelines highlight that it is competitively sensitive to wide range of expected burst sources while making minimal assumptions on morphology, proving it is beneficial to apply it in the generic search for GW transients.

We employ the cWB+GMM pipeline in the offline all-sky search for short-duration, low-frequency GWs in the LVK fourth observing run, presenting details on search configuration investigations and results with increased sensitivities. The newly enhanced GMM-based search detects 13 confident CBC events. The loudest non-CBC event is observed with significance of inverse false alarm rate (iFAR) equal to 1.75 years, however initial investigations indicate that this is likely an artefact of noise. Despite no confident burst events being detected, we show that the GMM observes significant sensitivity improvements across the short-duration low-frequency parameter space compared to those obtained in O3. Additionally, the implementation as an offline search highlighted many avenues for future improvements, with insight into GMM behaviour and initial investigations into dominant sources of glitches.

We extend the use of GMM as a noise mitigation tool in GW searches by applying it in a targeted sensitivity study for parabolic radiation-driven capture systems. The approach of a targeted GMM model is introduced, and performance is tested against against standard and generic GMM post-production approaches. We demonstrate that the use of a targeted GMM has the potential to significantly increase sensitivities, however further investigations into the construction of an optimised training set will refine this application further. Through this study we also conclude that high mass radiation-driven black hole capture systems with equal mass ratio and initial angular momentum of 0.9 may be detected with sensitive distance of 1.74Gpc, however upper limits of rates we can achieve with current sensitivities are not competitive with literature.

Overall, the development and extensive testing of the GMM post-production methodology to cWB proves that it is an effective technique in the mitigation of the dominant source of transient noise in un-modelled searches, increasing search sensitivities to expected burst sources across the parameters space. Furthermore, it is shown that the application to a 3-detector network and targeted searches for transient burst sources are promising implementations for the future.

Contents

List of Tables

List of Figures

List of Abbreviations

BBH binary black hole.

BH black hole.

BIC Bayesian information criterion.

BNS binary neutron star.

CBC compact binary coalescence.

CCSN core-collapse supernovae.

CE Cosmic Explorer.

CS cosmic string.

cWB coherent WaveBurst.

DPF dominant polarisation frame.

eBBH eccentric binary black hole.

EM expectation maximisation.

ET Einstein Telescope.

FAR false alarm rate.

FFT fast Fourier transform.

FRB fast radio burst.

GA Gaussian pulse.

GMM Gaussian mixture modelling.

GR general relativity.

GRB gamma ray burst.

GW gravitational wave.

GWOSC gravitational wave open science centre.

iFAR inverse false alarm rate.

IMBH intermediate mass black hole.

LH LIGO-Livingston, LIGO-Hanford.

LHV LIGO-Livingston, LIGO-Hanford, Virgo.

LVK LIGO-Virgo-KAGRA.

MDC mock data challenge.

NR numerical relativity.

NS neutron star.

NSBH neutron star-black hole.

oLIB Omicron-LIB.

PN post-Newtonian.

PSD power spectral density.

PTA pulsar timing array.

ROC receiver-operating-characteristic.

SASI standing accretion shock instability.

SG sine-Gaussian.

SNR signal-to-noise ratio.

TF time-frequency.

WDM Wilson-Daubechies-Mayer.

WNB white noise burst.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Ik Siong Heng for his constant advice throughout my PhD. I would not be the researcher I am today if it wasn't for your knowledge and encouragement. I am also thankful to my examiners, Patrick Sutton and Craig Stark, for their helpful comments and insights on the work within this thesis. I would also like to extend my thanks to the IGR as a whole: whether it was insightful discussions and suggestions in group meetings or chats over coffee breaks and pub trips, you all created a friendly environment which shaped my experience during my PhD. In particular I want to thank my officemates in 253 for getting me through many dark days, listening to my work rants and giving endless advice on plots and emails. The time in my PhD absolutely would not have been the same without you all.

My time as a PhD student has undoubtedly benefited from being a part of the LVK collaboration and I would like to thank everyone I have come across along the way, whether it be in conferences across the world, at a detector in Louisiana or through many zoom meetings. In particular I extend a huge thanks to the GMM team (Archana Pai, Gayathri V, Sayantan Ghosh, Jiyoon Sun and Ik Siong Heng) for their constant advice, and to Shubhanshu Tiwari and Francesco Salemi for their invaluable knowledge and help in navigating the depths of cWB. I am also extremely grateful to my colleagues in Trieste for welcoming me into my first postdoc and being so understanding while I battled through the final stages of finishing my thesis.

I am grateful for my many friends for all the laughs and distractions from my work, and always being there for me no matter where they are in the world. I am grateful for my family for supporting me in everything I do and never doubting my capabilities. I wouldn't be where I am today if it wasn't for you. Of course I can't forget my dog and chinchilla, Marley and Eddie, for their emotional support as they unknowingly helped me through my PhD.

Finally, I am forever grateful to Doug, for always believing in me when even I didn't, and making the most stressful times of my life better. No matter where the journey of research may take me, it's enough to know that you'll be by my side.

My research was supported by the Science and Technologies Facilities Council (STFC).

Declaration

The work presented in this thesis was carried out by the author under the supervision of Prof. Ik Siong Heng at the University of Glasgow. It has not been submitted for any other degree at the University of Glasgow, or at any other institution. This work has been completed as part of the LIGO-Virgo-KAGRA (LVK) collaboration and has therefore has benefited from discussions with many researchers. All work in this thesis was carried out by the author unless otherwise explicitly stated. Chapters 1 and 2 contain introductory material relevant for the remainder of this thesis, written by the author with appropriate references.

The work presented in Chapter 3 builds upon the methodology of GMM as post-production to cWB, previously presented in works by V. Gayathri and D. Lopez [1, 2]. The code development, investigations and associated results in this Chapter were led by the author with input from the GMM development team. Section 3.3 is published as part of [3] in *Physical Review D*.

The work presented in Chapter 4 was led by the author with input from the GMM development team. Sections 4.1 and 4.2 are published as part of [3] in *Physical Review D*. The work in Section 4.2 was carried out in collaboration with a masters student, Jiyoon Sun, under the author's close supervision. The preparation of data and simulations for the Burst benchmark project in Section 4.3 was performed by the LVK Burst benchmarking team. This data was processed by the author to obtain the results presented in this section, and may later be included as part of a larger publication. cWB triggers throughout this section were produced by members of the LVK cWB team.

Work in Chapter 5 was completed by the author as analysis for the LVK, some of which will later contribute to a larger publication. cWB triggers were produced in collaboration with other members of the cWB team. The work was completed with input from the GMM development team. The production of the veto definer file for the work in Section 5.2.2 was produced by Amber Stuver and the detector characterisation group.

Within Chapter 6, Section 6.1 contains background material on how radiation-driven capture systems form, with appropriate references. Section 6.2 is the result of a collaboration with Michael Ebersold, Shubhanshu Tiwari and others, published in *Physical Review D.* [4]. M. Ebersold and S. Tiwari produced cWB simulations used throughout the Chapter. Sections 6.3 and 6.4 present work completed by author, and may be considered for publication in the future.

Chapter 7 reviews all work presented in this thesis, written by the author.

Chapter 1

Introduction to gravitational waves

Gravitational waves (GWs) were first predicted by Einstein's theory of general relativity (GR) in 1916 [5] as small perturbations in the gravitational field propagating from dense objects merging or rotating. The first direct detection of GWs occurred many years later, when the advanced LIGO detectors [6] observed the merging of a binary black hole system in 2015 [7]. This detection opened a new window of observation into the universe, and since then, many more detections have been made by the LIGO-Virgo-KAGRA (LVK) collaboration [8, 9, 10, 11]. With even more detections expected in the future, the observation of GWs has the potential to advance knowledge in a variety of fields. To give just a few examples, GWs may provide insight into the formation properties of black holes (BHs) [12, 13], infer cosmological parameters [14, 15, 16], give details on the equation of state of neutron stars (NSs) [17, 18], provide information on the inner mechanisms of supernovae explosions [19, 20], and test the theory of GR [21, 22, 23].

While GWs may be detected from a variety of astrophysical objects, this thesis focuses on enhancing the detection of GWs specifically arising from short duration transients. We start by providing introductory material in Chapter 1 and 2 in order to introduce the reader to required concepts and motivate the work completed in later Chapters. Specifically, Chapter 1 aims to provide an overview on the topic of GWs. A brief introduction to the propagation of GWs from GR is given in Section 1.1, while details of current detection methods are discussed in Section 1.2. Section 1.3 details sources of noise, both in terms of limiting the sensitivity of the detector and troublesome noise artefacts for data analysis techniques. An introduction to astrophysical sources which are predicted to produce GWs in the ground-based detector sensitivity range is provided in Section 1.4, and finally an overview of the current status of GW detections is given in Section 1.5. Chapter 2 extends introductory material by providing specific details on the search for GWs from short duration transients, alongside an in-depth overview on the methodologies of coherent WaveBurst (cWB) and Gaussian mixture modelling (GMM) which are crucial for an understanding of the work in this thesis. Chapters 3, 4, 5 and 6 detail the techniques and application of an enhanced search method developed by the author. Finally, conclusions on the work are given in Chapter 7 alongside discussion of possible avenues for future improvement.

1.1 Propagation of gravitational waves

In Einstein's theory of general relativity it was predicted that accelerating massive objects would disrupt space-time, causing small perturbations known as gravitational waves (GWs) to propagate from a given source at the speed of light. Hence we start by briefly introducing the formation of GWs from Einstein's Field equations, following literature from [24, 25, 26]. Einstein's Field equations link the curvature of space-time to the way in which matter moves, described by:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu} \tag{1.1}$$

where $R = g_{\mu\nu}R^{\mu\nu}$ is the Ricci scalar, $R_{\mu\nu}$ is the Ricci tensor which describes the curvature of space-time, $T_{\mu\nu}$ is the stress-energy tensor which encodes the energy-momentum and density at a given point in space-time and $g_{\mu\nu}$ is the metric tensor. The derivation of GWs can be described by considering a perturbation $h_{\mu\nu}$ to the flat space-time Minkowski metric $\eta_{\mu\nu}$. Assuming the weak-field approximation as in [24], the metric tensor is:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.2}$$

for $|h_{\mu\nu}| << 1$. In order to find solutions to the linearised Einstein equations in this regime, we first re-scale the coordinates of the metric perturbation to the trace-reversed perturbation $\overline{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$ for $h = \eta^{\mu\nu}h_{\mu\nu}$. Assuming also the Lorentz gauge, the propagation of the perturbation is described by the wave equation:

$$\Box\overline{h}_{\mu\nu} = -16\pi T_{\mu\nu} \tag{1.3}$$

where $\Box$ is the d'Alembertian operator. This can be simplified by considering that the stress-energy tensor is zero in flat-space time, resulting in

$$\Box\overline{h}_{\mu\nu} = 0 \tag{1.4}$$

Solutions to the wave equation will then have the form of

$$\overline{h}_{\mu\nu} = A_{\mu\nu}\exp\left(ik_{\alpha}x^{\alpha}\right) \tag{1.5}$$

for polarisation tensor $A_{\mu\nu}$ and wave tensor k_{α} for a wave travelling at the speed of light. To simplify $A_{\mu\nu}$ we assume the transverse-traceless (TT) gauge so that $\overline{h}_{\mu\nu} = h_{\mu\nu}$ and $\overline{h}_{0\nu} = 0$. This reduces the polarisation tensor to 2 degrees of freedom in the plane transverse to the direction of propagation. Assuming the wave propagates on the z-axis, we have:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{xx} & A_{xy} & 0 \\ 0 & A_{xy} & -A_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.6}$$

The two non-vanishing components of the wave solution translate to two GW polarisation components denoted as *plus* '+' and *cross* '×':

$$h_+ = A_{xx} \exp(-i\omega(t - z/c)) \tag{1.7}$$

$$h_\times = A_{xy} \exp(-i\omega(t - z/c)) \tag{1.8}$$

The 2 basis polarisation tensors are defined as [24, 25]:

$$\mathbf{e}_+ = \hat{e}_x \otimes \hat{e}_x - \hat{e}_y \otimes \hat{e}_y, \quad \mathbf{e}_\times = \hat{e}_x \otimes \hat{e}_y + \hat{e}_y \otimes \hat{e}_x \tag{1.9}$$

Then a general wave in the TT-gauge will be a linear combination of both polarisations:

$$h_{ij}^{TT} = h_+ \mathbf{e}_+ + h_\times \mathbf{e}_\times \tag{1.10}$$

A simplified illustration of how the + and × polarisations affect a ring of freely moving particles is shown in figure 1.1. As the wave propagates through the page the ring of particles is stretched into an ellipse, before being contracted back to a circle and shifted into an ellipse on the other axis. It can be seen that both polarisations observe the same behaviour with a $\pi/4$ rotation around the axis of propagation. Measurements of GWs may be made by quantifying the change in distance of said particles as the wave propagates, which is discussed further in section 1.2.

Finding analytical solutions to Einstein's equations can be difficult, and the post-Newtonian (PN) approximation is often used for bound systems. The lowest order of PN is the quadruple approximation, in which the amplitude of the emitted GW is [25]:

$$h_{jk} = \frac{2G}{c^4 r} \frac{d^2 Q_{jk}}{dt^2} \tag{1.11}$$

where G is the Gravitational constant, c is the speed of light, r is the distance to the source and the Quadrupole moment (second moment of mass distribution) $Q_{jk} = \int \rho x_j x_k d^3x$, for mass density ρ, coordinates x_j, x_k. Gravitational radiation may be observed in higher modes of the PN expansion, however often is not as strong as emitted in the Quadrupole moment. The method of numerical relativity (NR) may also be used to solve Einstein's equations for gravitational radiation, explained in [27]. While this may give more accurate solutions, such simulations are computationally expensive.

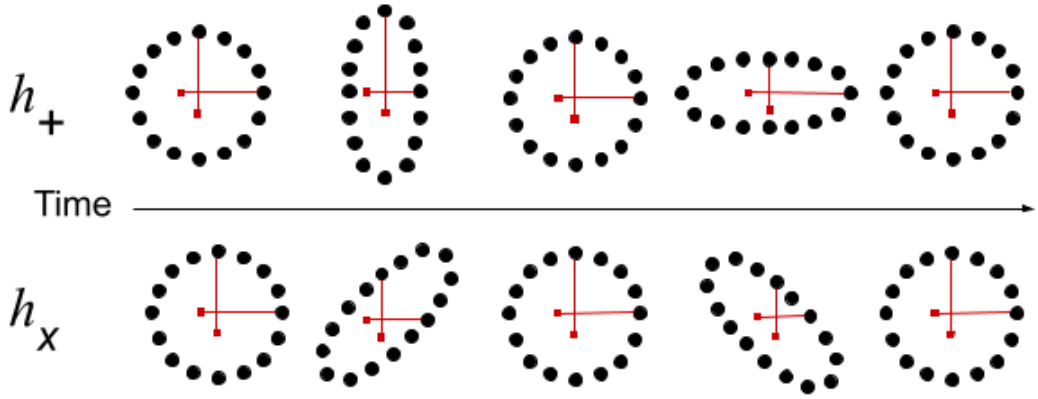


Figure 1.1: Simplified diagram showing the effect of the two polarisation components of GWs on a ring of test particles. In this case it is assumed the GWs are propagating out of the page, with time increasing along the x-axis to the left.

Eqn. 1.11 tells us that the perturbation of GWs relies on the density of the source, hence only large astrophysical systems are capable of producing GWs with an amplitude which may be detected on Earth. In fact, even considering a binary system of two $10M_\odot$ merging black holes at 100Mpc, the amplitude of GW is $< 10^{-22}$ once it reaches Earth [28].

1.2 Gravitational wave detectors

The direct detection of GWs is a very challenging task due to their extremely small amplitudes upon arrival at Earth. Many methods for such detection have been developed and tested, such as resonant bar detectors, pulsar timing arrays and both ground-based and space-based interferometer detectors. In the 1960s, Weber developed the first resonant bar GW detector [29, 30], in which a large cylindrical bar would oscillate at a resonant frequency in the presence of a GW. Dependent on the length and material of such detectors, the resonant frequency may range from $\sim 500 - 1500$Hz, however has a narrow-band sensitivity range around this value. It was difficult to reach required sensitivities with these detectors when first developed, and even though the sensitivity of bar detectors have improved over the years [31], their narrow-band frequency limits the quantity of observable GW sources. Pulsar timing arrays (PTAs) can detect GWs through the time of arrival of radio pulses from a network of known pulsars, due to the periodicity of these pulses being measured to high precision. This allows for changes in time of arrival due to the presence of a GW to be measured [32]. Such methods are sensitive to low frequency sources in the range 10^{-9} to 10^{-6}Hz, such as the stochastic GW background (see Section 1.4.4). There are groups working on the detection of GWs through PTA, such as the European PTA [33] and NanoGrav [34], which have recently found evidence of GWs in the form of a stochastic background after 15 years of observation [35, 36, 37].

Interferometers are able to detect GWs by measuring the change in proper distance of test masses to high precision as a wave propagates through. The design of ground-based interferometers has been successful, with the first detection of GWs in 2015 by the advanced LIGO detectors [7], while the advanced LIGO [6], advanced Virgo [38], and KAGRA [39] detectors have detected many GWs since [8, 9, 11]. Current ground-based interferometers target GWs in the frequency band of $\sim 10 - 10^3$Hz, however there are plans for future ground-based detectors of Einstein Telescope (ET) [40] and Cosmic Explorer (CE) [41], which aim to detect GWs over a larger frequency range of $\sim 1 - 10^4$Hz at higher sensitivities [42]. Planned space-borne interferometer detectors include LISA [43], TianQin [44] and DEICIGO [45]. Detectors based in space are able to target GW sources from much lower frequencies, approximately in the range of $\sim 10^{-4} - 1$Hz [46, 47, 48], opening up the opportunity to detect a variety of new sources.

The design of ground-based interferometers is the focus of this section, where a brief overview of the LIGO Michelson interferometer detector will be given. While the Virgo and KAGRA detectors are not the focus, their design holds similarities to the those mentioned below.

1.2.1 Laser interferometer detectors

As stated, Laser interferometers allow for very precise measurements of the change in length of optical paths, which in this case may be caused by the passing of GWs. In particular we focus on Michelson interferometers, which are the base design of current ground-based detectors. In a simple Michelson interferometer, a beam-splitter splits the light from a laser along 2 orthogonal arms of equal length, each with mirrors at the end which are free to move with the passing of GWs. The light is reflected back by the mirrors, recombining at the beam-splitter and sent to a photo-detector where an interference pattern occurs. In the absence of any signal, each arm will remain the same length and the paths of light will remain in phase when recombined. In the presence of a GW, the mirrors at the end of each arm will be affected in a similar manner as depicted in Figure 1.1, with the length of each arm changing as the GW passes. This change in arm length causes the light from each arm to be out of phase when recombined, which is observed via the interference pattern at the photo-detector.

In reality, the design of interferometer detectors is more complicated than this, with a variety of modifications in place to obtain the extreme precision required for GW detection. Some key features added to the Michelson-interferometer set-up are Fabry-Perot cavities, a power-recycling mirror and a signal-recycling mirror, which are highlighted in a simplified diagram in Figure 1.2. A brief description of these features is provided below, however they do not fully cover the many intricate techniques used in the detectors, which are detailed in [6, 38, 39].

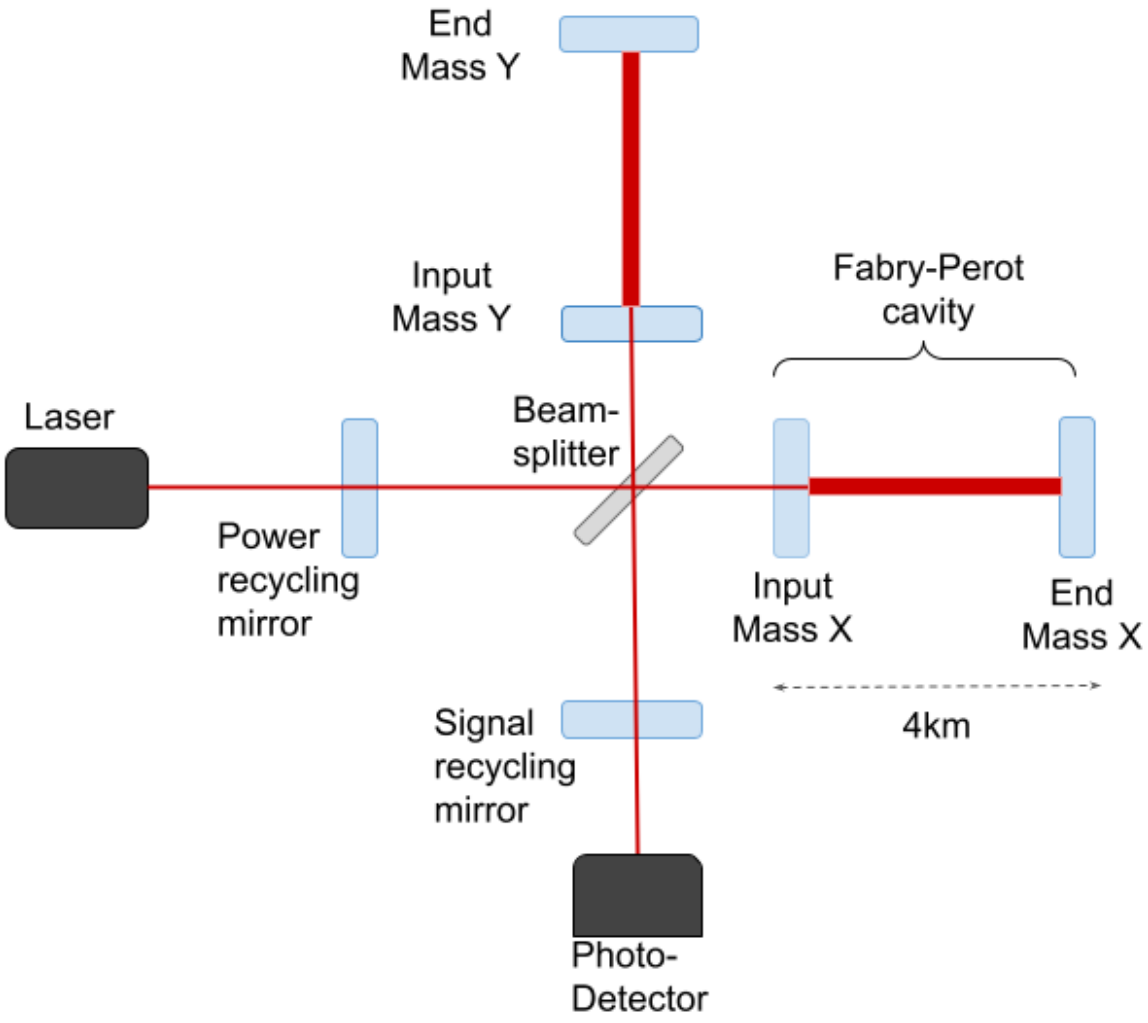


Figure 1.2: Simplified diagram of a Michelson interferometer used to detect Gravitational waves. Additions of Fabry-Perot cavities, power recycling mirror and signal recycling mirror help to increase the sensitivity of the detector. The diagram is based upon the design of LIGO, see [6] for further details.

Fabry-Perot cavities

Fabry-Perot cavities allow for the effective length of the interferometer arms to be increased through the placement of a partially transmitting input mirror and highly reflective end mirror, causing the light to be reflected many times before leaving the arm. To demonstrate the need for Fabry-Perot cavities, we will consider a simplistic detector set-up with arms running along the x- and y- axis. Suppose a GW signal with amplitude h is causing difference in arm length ΔL to the interferometer with arms of length $L_x = L_y = L$. Assuming the long wavelength approximation which states that the wave does not change as it travels along the arm, the change in arm length can be approximated by [49]

$$\Delta L \approx hL \tag{1.12}$$

From this relationship it is clear that increasing the length of the interferometer arms increases the change in arm length observed when a given GW passes. Hence extremely long interferometer arms are required in order to be able to detect GWs of $h \sim 10^{-23}\text{Hz}^{-1/2}$, which is the current sensitivity of ground-based detectors [50]. The use of Fabry-Perot cavities allows for the effective arm length of the detector to be increased without the difficulty of physically building long detector arms.

Power recycling

The sensitivity of detectors is effected by the power of the laser due to it's relation with quantum fluctuations, known as shot noise and radiation pressure noise (to be introduced in section 1.3). High power lasers are able to reduce shot noise [51], and a method known as power recycling allows for the effective power of the laser to be increased. In the absence of a GW, a fraction of the light is reflected back towards the laser. Placing a mirror cavity between the laser output and the beam-splitter allows for light to build up power before travelling back down the arms [52]. This allows for the use of a weaker laser as input into the interferometer, while still achieving the results of using a powerful laser.

Signal recycling

Signal recycling is used to increase the sensitivity of the detector over a given frequency bandwidth by placing a mirror between the beam-splitter and the photo-detector to create a resonant cavity on the detector output. When a GW signal is present, the laser will build up resonance before being output to the photo-detector, increasing the SNR of the signal. The targeted bandwidth of signal recycling is dependent upon the length and reflectivity of the mirror cavity [53], thus it is possible to target the detector response over narrow bandwidths with this technique, however in the LIGO detectors it is used to increase the signal over broadband frequencies.

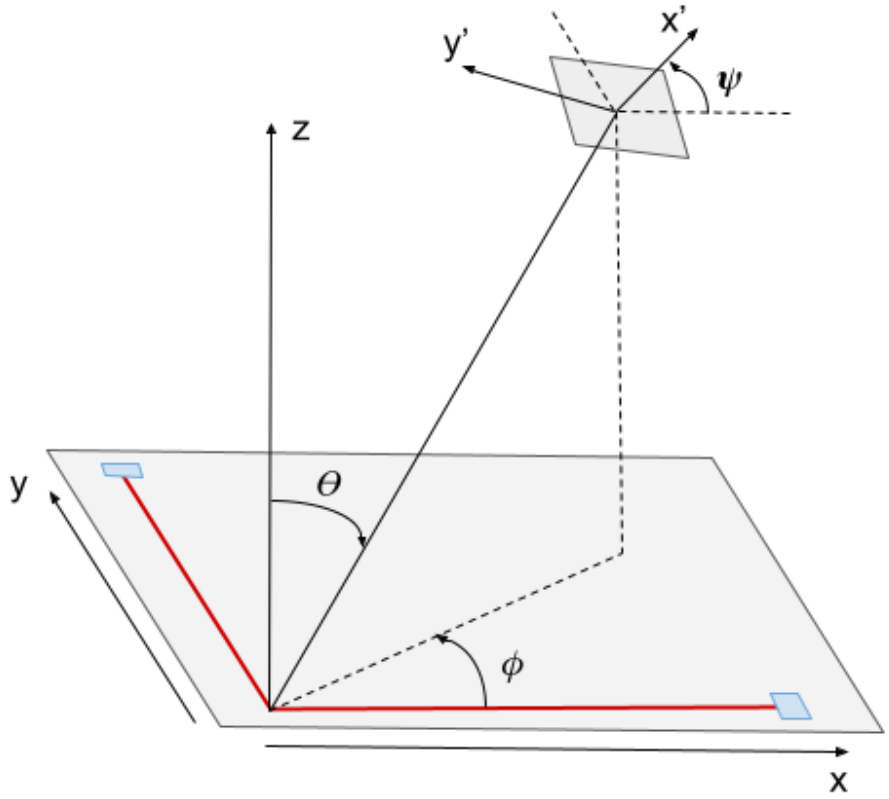


Figure 1.3: Diagram of the coordinate system of GW source with respect to a detector, used in the definition of antenna patterns in eqns. 1.14, 1.15. The detector arms stretch along the x- and y- axis, then the position of the source is given by polar angle θ and azimuthal angle ϕ. The polarisation angle ψ describes the rotation between the source plane and detector plane.

1.2.2 Detector response

Detectors are not equally sensitive to sources across the entire sky at a given instance, and instead have antenna patterns defining sensitivity based upon the polarisations and sky locations of GW signals with respect to detector orientation. The response of a detector to a given source is described by the linear combination of each polarisation component and corresponding antenna pattern functions:

$$h = h_+ F_+ + h_\times F_\times \tag{1.13}$$

for antenna patterns $F_{+,\times}$ for each polarisation varying between $[0,1]$. Considering a detector with arms along the x- and y- axis, the orientation of a source in the detector frame $((x,y,z))$ can be described by the polar angle θ and azimuthal angle ϕ. We extend this to the source frame (x',y',z') transverse to the direction of GW propagation, defining the polarisation angle ψ as a rotation of basis vectors in the source frame. A diagram of this coordinate system is seen in Figure 1.3. Assuming the above set-up, the antenna pattern functions associated to each polarisation is [25, 54]

$$F_+ = \frac{1}{2}[1+\cos^2\theta]\cos 2\phi\cos 2\psi - \cos\theta\sin 2\phi\sin 2\psi \tag{1.14}$$

$$F_\times = \frac{1}{2}[1+\cos^2\theta]\cos 2\phi\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi \tag{1.15}$$

The antenna power pattern of a detector may also be defined as

$$P(\theta,\phi) = F_+(\theta,\phi,\psi)^2 + F_\times(\theta,\phi,\psi)^2 \tag{1.16}$$

which can be useful for defining the sensitivity of a detector across the sky since it is independent of source polarisation angle ψ. This leads to detectors having regions of low sensitivity across the sky for source directions where $P(\theta,\phi) \sim 0$. For example, the detector set-up described above will have a low sensitivity to GWs propagating from $\theta = \pi/2, \phi = \pi/2$ (ie. edge on to the detector), since both arms will change length by the same amount (ΔL from eqn. 1.12 is zero). The same detector will achieve maximum sensitivity for GW propagation transverse to the detector frame (eg. $\theta = 0, \phi = 0$). Detectors with different orientations will have different antenna patterns, thus the use of multiple detectors which are not co-aligned will increase sensitivity across the sky.

1.2.3 The network of interferometric detectors

The GW interferometric detector network currently consists of 5 individual detectors, with varying sensitivity, location and orientation: advanced LIGO [6] (LIGO-Livingston and LIGO-Hanford) located in the USA, advanced Virgo [38] located in Italy, KAGRA [39] located in Japan and GEO600 [55] located in Germany. Periods of time referred to as *observing runs* are defined in which multiple detectors will enter observing mode in the search for GWs in unison. This is extremely useful for the detection of GWs, as a detection that is made in more than one detector can be better distinguished from localised detector noise. A GW event will be coincident across multiple detectors, whereas local transient noise will only occur in one detector or will have a non-physical time-shift between occurrences in different detectors. While this is useful for the detection of all GW signals, it is particularly important for the detection of burst sources, since search algorithms must rely on excess energy observed across a network of detectors rather than make assumptions on the morphology of signals within the data. This will be discussed further in Section 2.2. In addition, a network of detectors allows for better coverage of the sky due to the differing antenna patterns. This aids in the localisation of GW events; if a GW event is seen in some detectors but not others, it is directly inferred that the direction of the source must be located in the low sensitivity region of the detector which did not observe it.

During the first to third observing runs, LIGO and Virgo detectors were operating at significantly higher sensitivity than GEO and KAGRA, hence detection of GWs has been based upon what is observed by these detectors. In the case of this 3-detector network, both LIGO

detectors are nearly co-aligned so will have almost identical antenna patterns, whereas the Virgo detector has a differing antenna pattern. Thus the addition of Virgo will increase the coverage of sensitivity across the sky and aid in the localisation capabilities of detected events.

1.3 Noise sources

The high precision required by the detectors to observe GWs makes them susceptible to a wide range of noise sources arising from instrumental malfunctions and environmental disturbances. These noise sources can be categorised into two types of noise based on their form and how they affect the sensitivity of the detector: noise which occurs constantly, directly limiting the sensitivity of detectors, or spurious transient noise which may cause problems in identifying GW signals during data analysis procedures. Mitigating noise in the detectors is of high importance, and a variety of techniques are utilised to minimise their effects. Detector upgrades and commissioning take place between observing runs to reduce or remove identified sources of noise, improving the sensitivity limitations of the detector. In the case of spurious transient noise, it is also possible for some sources of noise to be mitigated subsequent to the data being obtained, via vetoing procedures and techniques employed in data analysis. An overview of the limiting sources of noise for detector sensitivity and the dominant sources of spurious transient noise is given below.

1.3.1 Detector noise limitations

Limitations to the sensitivity of detectors arise from sources of noise within detector components and surrounding environmental systems. The limiting noise sources for a detector directly depend on the frequencies it is sensitive to, and the contribution from such sources will vary as sensitivity changes. Some of the most notable noise limitations for the advanced LIGO detectors are quantum fluctuations, seismic vibrations, Newtonian gravity gradients, and thermal vibrations due to the mirror suspensions and coatings, which will be summarised below. The effect of these noise sources on the advanced LIGO sensitivity curve is seen in Figure 1.4, alongside the total limiting sensitivity, known as the noise budget.

Quantum noise

Two sources of quantum noise contribute to the LIGO sensitivity curve, arising from fluctuations in photon intensity: shot noise and radiation pressure noise. Shot noise arises from quantum fluctuations in the photon arrival rate at the photodiode. These fluctuations effect sensitivity at higher frequencies, and are directly related to the power of the laser. Increasing the power of the laser can mitigate the effect of shot noise, which is one of the motivations for the power-recycling mirror in the detector design.

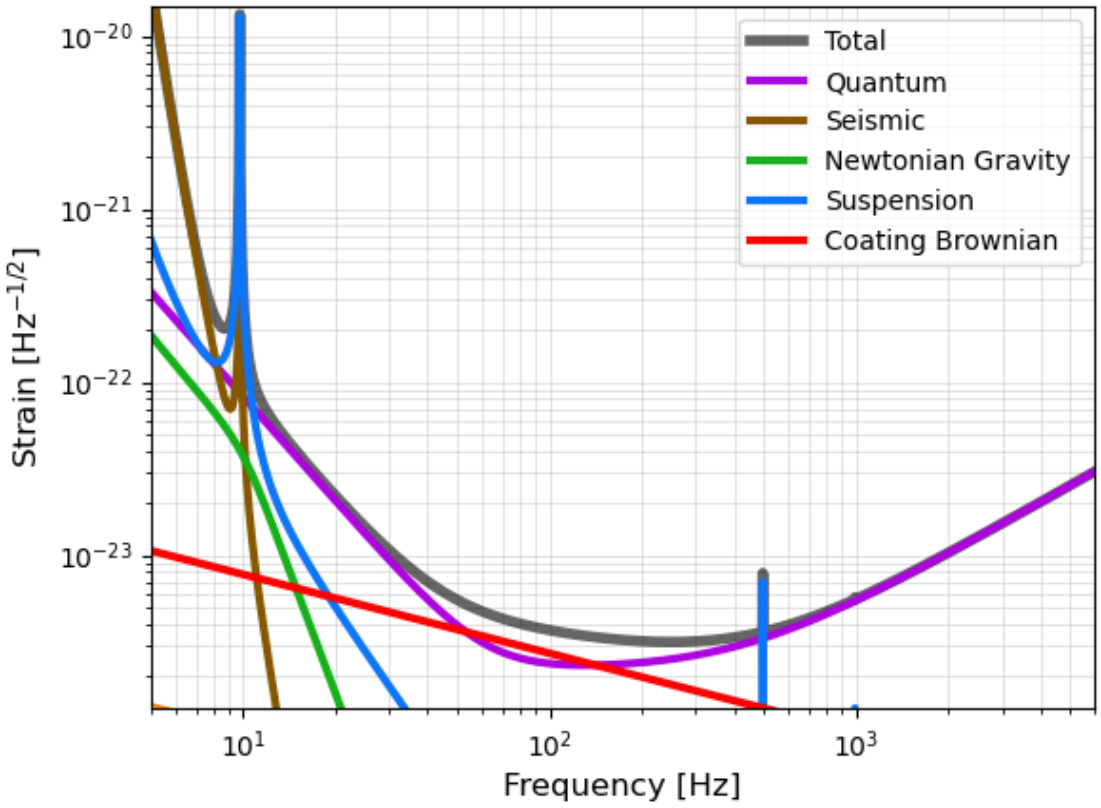


Figure 1.4: Contribution of various noise sources to the sensitivity limitation of the advanced LIGO detectors [6]. The total design sensitivity of the detectors is shown (Grey), in addition to limitations arising from quantum fluctuations (purple), seismic (brown), Newtonian gravity gradient (green), thermal suspension (blue) and Brownian noise from the mirror coatings (red). The plot was made with the GWINC package [56].

Radiation pressure is due to the variation in the number of photons arriving at the test mass at any given time, causing varying levels of pressure to be applied to the test mass [57]. Radiation pressure noise effects sensitivity at low frequencies, with dominance between 10-50Hz. The effect can be mitigated by increasing the mass of the end test mirrors, or lowering laser power at the expensive of lowering sensitivity at high frequencies due to shot noise [58].

From Figure 1.4 it is seen that quantum noise is a dominant limiting factor for the sensitivity of current detectors. Reducing this noise is not straightforward, as the use of high powered lasers to reduce shot noise increases the effect of radiation pressure noise. For a given frequency, there will exist a laser power which will optimise the sensitivity for both shot and radiation pressure noise [58], which is known as the Standard Quantum Limit. It is possible to surpass this limit through 'squeezing' light [59], which will aid in increasing sensitivity in future detector upgrades.

Seismic noise

Seismic noise occurs from vibrations in Earth, caused by seismic or anthropogenic sources such as earthquakes, tidal activity, or nearby railways and traffic. Seismic vibrations move the suspended components of the detector, such as the test masses, altering the length of the arms in absence of a GW. The effect of seismic noise is mitigated through the use of active and passive isolation systems within complex pendulum suspensions to reduce the movement of the test mass mirror due to ground motion [60]. Such isolation systems reduce the contribution of seismic noise to low frequencies of the sensitivity budget, however sudden events such as earthquakes will lead to fluctuations in this sensitivity curve.

Newtonian gravity noise

Newtonian gravity gradients are localised density fluctuations occurring on the Earth's surface from seismic surface waves or atmospheric pressure changes [24]. Such fluctuations cause displacement of the mirrors due to gravitational coupling as the surface waves propagate the ground below the detector, first identified as a source of noise to ground-based detectors in [61]. Gravity gradients affect the detector sensitivity at low frequencies, however is not currently a dominant source of noise for advanced LIGO detectors. The effect of gravity gradients are less prominent below the Earth's surface, hence this source of noise can be mitigated by building detectors underground, such as KAGRA [39] and future detectors ET [40] and CE [41].

Suspension thermal noise

Thermal vibrations arising from mechanical systems within the detector may couple into the output signal. A significant source of such noise arises from thermal vibrations in the suspension fibres attached to the mirror [62], which causes the test masses to move. This is a dominant source of noise in LIGO at low frequencies, with resonant spikes in frequency also present at 10Hz and 500Hz. Thermal vibrations can be reduced through the use of selected suspension materials and design choices which reduce the known resonant frequencies. Currently LIGO uses the low thermal noise material of fused silica for suspensions [63] to mitigate this source of noise.

Coating Brownian noise

Similar to thermal noise arising from the suspensions, coating Brownian noise arises from mechanical dissipation in the mirror coatings as the laser hits the surface of the mirror, causing thermal vibrations [64]. This effect can be mitigated through the application of multi-layer coatings of materials which have low thermal noise and high reflectivity. In the current LIGO sensitivity curve, Brownian noise is limiting at frequencies in the range $\sim 50 - 200$Hz. As

the quantum noise limit is reduced through squeezing in future detector upgrades, lowering the coating Brownian noise will become crucial.

1.3.2 Spurious transient noise

In addition to the noise sources continuously contributing to the detector sensitivity limits, GW detectors also suffer from noise events which occur in an instance of time. These transient noise sources are known as *glitches* and may occur from a wide variety of sources, often arising from environmental and human activity around the detector, or instrumental effects within the detector components.

Glitches are problematic in the identification of transient GW events such as CBC and burst type sources (introduced in section 1.4) due to their transient nature. They are also the main contributor to the estimation of background data, which ultimately effects the significance of observed GW signals (more details in sections 2.2.4). Because of this, their mitigation is extremely important for GW searches. In an ideal case, the origin of a glitch can be identified and the detector instrumentation responsible for it's cause may be modified to remove the source of the glitch, or at least reduce it's coupling to the detector output. Alternatively, if the origin of the glitch is known but it is not viable to remove it's source, periods of data in which significant glitches occur can be removed, referred to as *vetoes*. Two common categories of vetoes exist, *CAT1* and *CAT2*. CAT1 vetoes indicate periods of time which should not be analysed due to identified issues with detector components or the occurrence of known reproducible events surrounding the detector. CAT2 vetoes indicate periods of time where an identified noise source has coupling with the detector output [65], often identified through correlation in data channels set-up around the detector. Glitches with unknown cause or no data channel correlation cannot be vetoed, and instead require the development of data analysis techniques in order to reduce the effect of their presence.

In many cases, glitches must be classified into groups in order to identify their possible origin. They are often classified based upon morphology in the time-frequency space, however can also be grouped through the data auxiliary channels which are connected to sensors in various aspects of the detectors. Gravity spy [66, 67] is an example of an algorithm which attempts to identify and classify glitches, through a combination of citizen science and machine learning techniques. It is continuously used on LVK data and has been successful in identifying many sources of glitches throughout observing runs thus far.

Common types of glitch

The list of glitch types which occur in the detector is constantly changing due to a combination of known glitch sources being removed and new glitches being discovered with the ongoing upgrades to detectors. Here, only a subset of such glitches which are relevant to the work in this

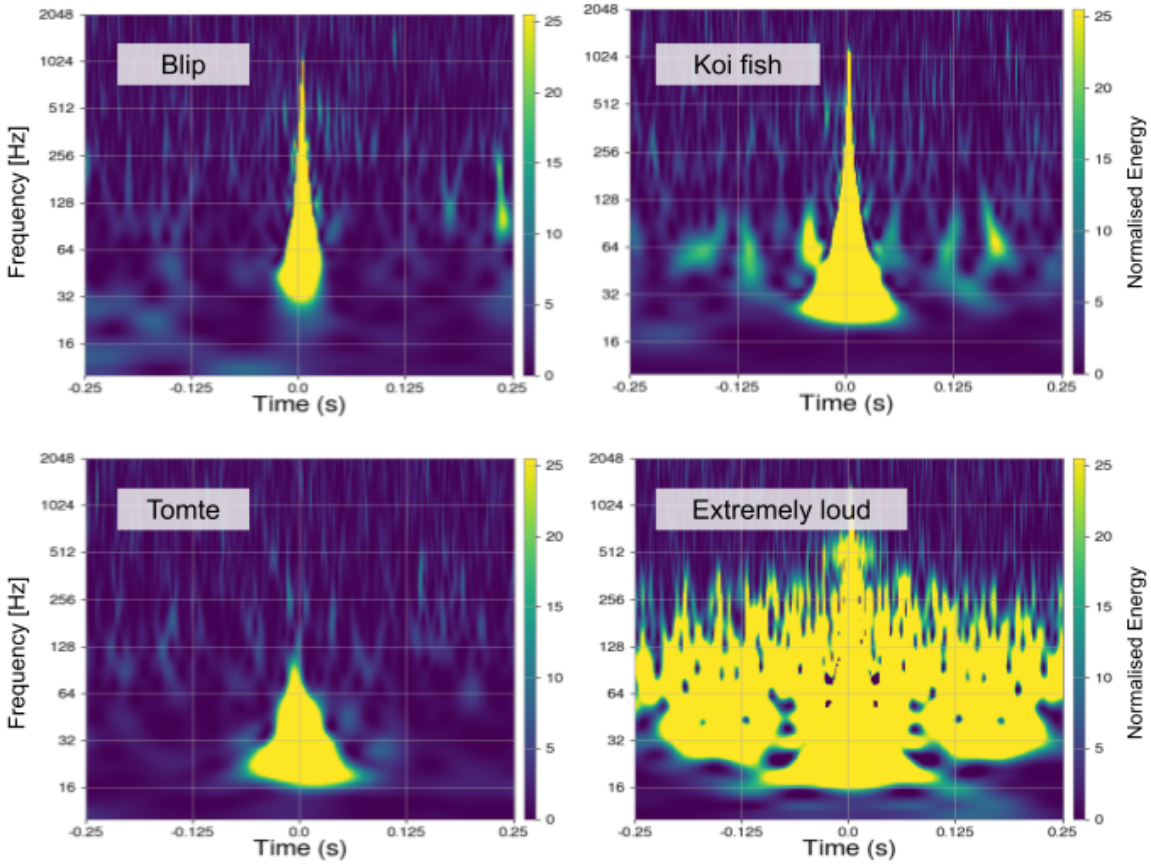


Figure 1.5: Example time-frequency spectrograms of glitches from the LIGO detectors. The above examples are common in the searches for GW transients, presenting as loud glitches from the work in Chapter 5. All plots have the same axis scales and were generated with GRAVITY SPY [66, 67]. Normalised Energy is shown on the z-axis, analogous to how loud the glitches are in the data. This value is capped at 25 so that saturation across multiple frequencies can be observed.

thesis are introduced, namely: blip glitches, koi fish, tomte and extremely loud glitches. These glitch types are the most problematic in searches for short duration GW transients due to their short, broadband morphology. Time-frequency spectrogram examples of these 4 glitch types is shown in Figure 1.5, with details of each described below.

Blip glitches are short duration events occurring over a broadband frequency range of $\sim 30-1000$Hz, seen in the top-left corner of Fig 1.5. In current detectors blip glitches are one of the most frequent glitches, occurring $\sim 2-4$ per hour in the LIGO detectors during the third LVK observing run [68]. Studies have been completed in an attempt to discover the origin of such glitches [69], however as of yet their cause remains unknown, meaning they cannot be vetoed. This makes them one of the most dominant glitches in terms of limiting the sensitivity of transient searches [70].

Looking at the top-right and bottom-left of Fig 1.5, it can be seen that the koi fish and tomte glitches have similarities to the blip glitch. Koi fish glitches have a similar short duration,

broadband morphology, except have more widespread signal at lower frequencies, known as 'pectoral fins', hence their name. Unfortunately their origin is also not understood. Tomte glitches have short duration and occur at low frequencies with a triangular shape, and have similar morphology to blips in the low frequency region. Their cause is also not fully understood.

Extremely loud glitches are a broad category of glitches which are classified by very high signal-to-noise ratio (SNR) causing saturation across the spectrogram. An example of such a glitch is seen in the bottom-right of Fig 1.5, however they can be caused by many occurrences and hence have varying morphologies in the time-frequency space. As of yet, their exact cause is uncertain, however during the third observing run some possible correlations to data channels were observed [68], with further investigations required.

As can be seen, the origin of many glitches is unknown, meaning it is often not possible to remove them from the detector data via vetoes. Instead, data analysis techniques must be developed in an attempt to mitigate their effect in the search for GW transients. Specifically, the work of this thesis develops the technique of Gaussian mixture modelling (GMM) as a tool to mitigate the effects of these glitches in the weakly-modelled searches for short duration GW transients.

1.4 Sources of gravitational waves

All physical objects which accelerate will emit GWs, however the required mass of sources in order to detect GWs is so large that it is not currently feasible to detect man-made signals. Instead, we focus on the detection of GWs from astrophysical sources. There exist a large range of astrophysical sources which cause GW perturbations, occurring over a varying range of frequencies. Below, we focus on the main categories of source which may be observed by the current ground-based detector network.

It is useful to categorise sources into groups based upon length of signal and ability to model their waveforms, resulting in 4 broad categories of GW sources: compact binary coalescence (CBC), burst, continuous, and stochastic. CBC and burst are transient sources which produce a signal lasting up to a few seconds. CBC sources are comparatively well understood allowing for well modelled and accurate waveform simulation, whereas burst sources include signals which cannot be accurately modelled or are not fully understood. Continuous and Stochastic signals are both longer in duration, expected to occur constantly throughout observing runs. Continuous waves are well-modelled since they arise from predicted sources, whereas stochastic waves arise from a combination of undetectable signals, hence cannot be well modelled. Due to the variety in signal morphology across the expected sources, each category requires different approaches in order to observe their signals.

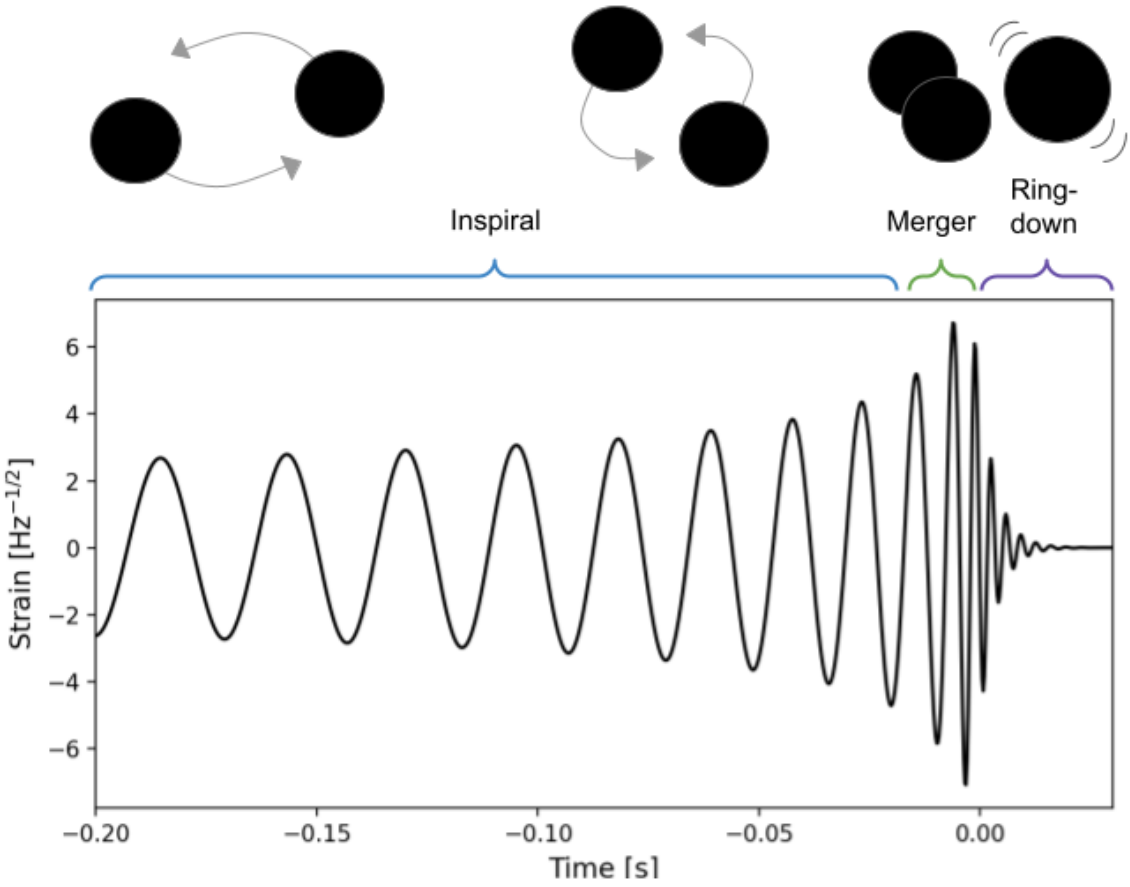


Figure 1.6: Simplified example of the evolution of a CBC system with accompanying GW signal. The portions of the signal due to the inspiral, merger and ringdown of the system are labelled alongside a visual representation of the binary system in the relevant phase. The waveform shown is a signal from a non-spinning BBH system with component masses $m_1 = m_2 = 30\mathrm{M}_\odot$ and was produced from IMRPHENOMD [71, 72]. The plot was made using PYCBC [73]

1.4.1 Compact binary coalescences

Systems of compact binary coalescence (CBC) involve two compact objects which are in orbit around a common centre of mass. As the objects inspiral towards each other, the system loses energy in the form of GWs, pulling the objects closer together until they eventually merge into a single compact object. In terms of the ground-based detector sensitivity band, the compact objects in these systems may be black holes (BHs) or neutron stars (NSs), resulting in 3 subcategories of CBC signals: binary black hole (BBH), binary neutron star (BNS) and neutron star-black hole (NSBH).

The exact signals produced from each of type of CBC system varies, however all include three phases which are visibly present: inspiral, merger and ringdown. During inspiral, the objects are in orbit and are slowly being pulled towards each other, creating a steady oscillating signal which increases in amplitude and frequency as the objects become closer. Merger is the

point in the signal where the objects pass the innermost stable circular orbit (ISCO), causing them to collide into a single remnant. Finally, the ringdown stage is where the remnant object oscillates down into an equilibrium state. An example of a standard CBC signal in time-series is seen in Figure 1.6, including a simplified diagram of the compact object system at each phase of the signal. Not all of the inspiral signal is observable to ground-based detectors since the amplitude is small and frequencies may be outwith the sensitive range. The length of the observable signal also differs dependant on the type of system: on the order of minutes for BNS, whereas on the of order of seconds for BBH. The merger phase contains the peak of the signal, at which the frequency is $f_{ISCO} \approx c^3/6^{3/2}\pi MG$ [28]. Since current ground-based detectors are sensitive to frequencies from $\sim 10 - 10^3$Hz, they are sensitive to CBC systems of total mass $\sim \mathcal{O}(1) - \mathcal{O}(100)\mathrm{M}_\odot$.

Since the expected GWs arising from CBC sources are comparatively well understood, waveforms can be simulated accurately, allowing for a technique called matched filtering to be used to detect such signals within the detector data [74]. This requires the construction of a template bank of possible signal waveforms, which are matched against detector data in order to detect a signal. The template bank includes a large variety of waveforms since the signal morphology of CBCs varies over many system characteristics, for example mass, spin, and eccentricity [28]. Template banks are also useful for inferring source parameters via parameter estimation techniques [75].

GWs from CBC systems are the most common type of signal to be observed by ground based detectors. In fact, during the first 3 advanced LVK observing runs, these types of event were the only ones confidently observed [8, 9, 10, 11]. More details on the current status of GW detections are in Section 1.5.

1.4.2 Burst sources

GW bursts are short duration transient signals which involve a 'burst' of gravitational radiation. While this definition technically encapsulates CBC systems, it is assumed that burst events arise from astrophysical phenomena which cannot be well-modelled, and hence cannot feasibly be detected via matched filtering techniques. The reasons for such sources not being well modelled often arise from either computationally expensive waveforms or lack of knowledge on the physics of such systems, introducing a requirement for weakly-modelled search techniques. The work of this thesis focuses on enhancing search algorithms for a subset of such burst searches, specifically targeting sources with short duration occurring in the low frequency range of the ground-based detector sensitivity band. Thus, an in-depth discussion of expected short duration burst sources and the weakly-modelled coherent WaveBurst (cWB) algorithm used in the search for such events is given in Chapter 2, however an overview of burst sources is given below.

Since the category of Burst events includes all transient GW signals except CBCs, there is a much broader range of predicted sources. Some examples of burst sources are core-collapse

supernovae (CCSN) [76], cosmic string (CS) [77, 78], gamma ray bursts (GRBs) [79], fast radio bursts (FRBs) [80], and NS glitches [81, 82]. There are also some CBC systems which are targeted by burst searches due to difference in signal morphology compared to 'standard' CBC events: intermediate mass black hole (IMBH) binaries, eccentric binary black holes (eBBHs), and hyperbolic encounters [83, 84]. In addition, due to their weakly-modelled nature, burst searches are also sensitive to detecting CBC systems which are targeted by the matched filter algorithms.

Due to the inability to fully model waveforms from the above sources efficiently, burst searches must use different techniques to detect GWs. Often this involves the use of coherent analysis across multiple detectors, allowing for GW signals to be identified without making restricting assumptions on their morphology. The nature of these weakly-modelled searches means that they are also able to detect signals from completely unpredicted sources, accounting for GWs which may be produced from any unknown source. A variety of all-sky [85, 86] and targeted [20, 87, 79, 88, 80, 89, 90, 91, 83] searches for GWs from burst sources have taken place to date, however as of yet the only confident detections observed by burst pipelines have been CBC candidates. The case of non-detection of many burst sources allows for upper limits to be placed on their expected rates.

1.4.3 Continuous sources

Continuous GW signals take up a completely different form to the groups of transient signals described previously, with a long duration consistently present signal. The primary source for continuous GWs is rapidly rotating neutron stars (*pulsars*) with a mass asymmetry around their rotation axis. While it is theorised for various mechanisms to cause such an asymmetry, two of the most common are permanent imperfections in the NS crust or oscillations within the NS from "r-modes" due to fluid motion [92]. As the NS rotates around it's axis, this deformation will cause the radiation of GWs.

In general the expected signals arising from continuous waves are well modelled, and often the pulsars creating such signals are well-studied thanks to radio observations. The generic waveform structure expected from continuous sources is a sinusoid with locally constant amplitude and frequency. On long duration scales, the frequency of the signals is expected to decay due to the energy lost via gravitational radiation, associated with the 'spin-down' of the NS. Due to the well-modelled nature of continuous waves, matched filtering techniques are often used to extract signals from data [92, 93]. These searches can be targeted, looking for signals arising from pulsars with known location and rotation frequency [94], or may be generic, searching for signals from unknown pulsars across the entire sky [95]. So far, both targeted and generic searches for continuous GWs have not resulted in any detections, however have allowed for upper limits to be calculated on GW amplitude and abundance of theorised sources.

1.4.4 Stochastic sources

It is theorised that in addition to discrete sources of GWs, there should exist a constant background of GWs consisting of a super-position of many signals. There are 2 predicted sources for this: an astrophysical GW background and a cosmological GW background. The astrophysical background arises from sources which produce GWs too weak to be distinguishable alone [96]. The cosmological background originates from the early universe from sources such as cosmic strings and inflation [97, 98].

In both cases, the background is expected to arise from all directions, however is predicted to be anisotropic on small scales. Stochastic signals are referred to as 'un-modelled' since the signal itself is random across the sky, making it difficult to distinguish from detector noise. It can be statistically measured by correlating signals over multiple detectors, since the noise in the detectors is un-correlated both from each other and from the stochastic background. The stochastic GW background has not yet been detected with LVK ground-based detectors despite the expectation for it to be dominant over the cosmological background [99], however these searches have allowed for upper limits to be placed on the energy density of GWs in the universe [100]. Direct evidence of the astrophysical stochastic background has been provided through results with pulsar timing arrays [35, 37, 36].

1.5 Current observations of gravitational waves

So far, the LVK advanced ground-based GW detector network (introduced in section 1.2.3) have completed 3 observing runs, referred to as O1, O2 and O3. Between each of these observing runs, the detector has undergone instrumentation upgrades to increase sensitivity. Hence, more GW detections are made with each observing run. In this section, we will give a brief overview of the detections made to date, and detail the status of observing runs planned for the future.

The first detection of GWs was made during the first observing run, and was of a stellar mass BBH [7]. A further two GW observations were made during O1, also from BBH systems [8]. During the second observing run, a further 8 GW observations were made [8]. 7 of these were from BBH systems, however also consisted of the first GW detection from a BNS system: GW170817 [101]. This detection was particularly exciting, as it was the first (and only to date) confident multi-messenger observation, since a GRB was observed electromagnetically subsequent to the observation of GWs [102]. The third observing run, split into 2 parts (O3a and O3b), observed the most GWs to date with a total of 79 confident detections [9, 10, 11]. Of these were the first detections of NSBH systems [103]. O3 also brought the observation of the heavy BBH system, GW190521 [104], with component masses of $\sim 95 M_{\odot}$ and $69 M_{\odot}$. There were a variety of investigations into whether the origin of this event could be a source other than a heavy BBH, such as an IMBH [105], BHs on eccentric orbits [106] or cosmic strings [107]. Considering all detections across all 3 observing runs, a total of 90 confident GW detections

have been made.

There are currently two future observing runs planned for the ground-based LVK network: the fourth (O4) and fifth (O5) observing runs. O4 started in May 2023 with a planned end date of June 2025, split into 3 observing epochs. O5 will take place in the future, once upgrades have been carried out on all detectors. With the increased sensitivity of detectors in these future observing runs, it is expected for even more GW detections to be made.

As can be seen, observing runs thus far have only detected GWs from CBC systems, however the increased sensitivity of detectors in future observing runs combined with improvement to data analysis search techniques presents the increased possibility of detecting a non-CBC event, such as burst sources. Hence as we move into future observing runs, the optimisation of search techniques for GW bursts is of high importance. The work presented in this thesis explores the development of such a technique to aid in the detection of burst type GW events, applied on both O3 and newly collected O4a data. This will be further motivated in Chapter 2.

Chapter 2

Un-modelled search techniques for short gravitational wave transients

The search for gravitational wave (GW)s from burst sources requires the use of search algorithms which hold minimal assumptions on the morphology of signals. This is due to a combination of lacking accurate waveforms and the widely varied morphologies of expected burst sources. One common un-modelled algorithm is coherent WaveBurst (cWB), which bases the detection of GWs upon excess coherent energy across a network of detectors. The nature of such searches makes them highly susceptible to glitches, and often a variety of techniques must be employed to reduce their effect. This thesis focuses on the application and development of such a noise mitigation technique, namely Gaussian mixture modelling (GMM).

This is an introductory Chapter which provides an overview of predicted burst sources and data analysis techniques relevant for the work presented in this thesis. Section 2.1 details a selection of GW sources which are expected to arise in the short duration all-sky search parameter space. Section 2.2 details the inner-workings of the cWB algorithm and it's current post-production approach to reduce the effect of glitches. Finally, section 2.3 introduces the machine learning technique of GMM as a detection tool for GW searches, outlining how it may be utilised to mitigate glitches based upon the construction of it's own detection statistic.

2.1 The all-sky short burst search

There are a variety of astrophysical sources that our current ground-based detectors are sensitive to aside from compact binary coalescence (CBC) systems that have been detected thus far, one of which being generic GW transients (*Bursts*), as discussed in section 1.4.2. The majority of work within this thesis concerns the search for short duration burst transients, the *all-sky short search*. Short duration GW transients are defined as signals of duration 1ms - 10s in the frequency band 16-4096 Hz. A range of astrophysical sources are expected to emit GWs within this region of the parameter space, such as core-collapse supernovae (CCSN) [108, 20], neutron

star (NS) glitches [109, 82, 81], non-linear memory effects [110, 111], cosmic strings [97, 90] and parabolic/hyperbolic encounters [4, 83, 112]. Many of these sources do not have well-known waveform structure, or are too computationally expensive to generate at an accuracy required for the matched filtering techniques utilised in CBC searches [74]. Because of this, the search for short duration transients relies on 'un-modelled' techniques which are capable of detecting GWs with minimal assumptions on signal morphology. While being sensitive to a wide array of transient sources, algorithms which adopt such techniques can also detect CBC systems, such as eccentric binary black hole (eBBH) mergers [91] and intermediate mass black hole (IMBH) mergers [113]. Burst searches cover a wide parameter space, opening up the possibility of detecting completely new sources of GWs, however also present unique challenges in differentiating such signals from detector noise transients (*glitches*).

An array of weakly-modelled search algorithms are employed for the all-sky short search in order to have optimal sensitivities to sources in all aspects of the parameter space and to independently verify any burst-type detections which may be observed. Such algorithms include coherent WaveBurst (cWB) [114, 115, 116], BayesWave [117, 118], Omicron-LIB (oLIB) [119, 120], in addition to more recent algorithms which employ machine learning techniques, such as MLy [121] and GWAK [122]. There have also been efforts to utilise machine learning techniques as tools to aid in the mitigation of glitches alongside pre-existing algorithms, such as the XGBoost decision tree method [123, 124], and Gaussian mixture modelling (GMM) [1, 2] to optimise the search sensitivity of cWB. In particular it is the cWB and GMM algorithms which are the focus of this thesis, and their methodologies will be discussed fully in Sections 2.2 and 2.3. Despite all these pipelines having varying approaches to detect short GW bursts, they all commonly assume little information about signal morphology, sky localisation and time of arrival, unless targeting a specific family of sources.

As is the case with many of the GW data analysis pipelines, there are 2 classes of analysis considered for the all-sky short search: online and offline. The online cWB search analyses the data as it is acquired from the detectors, allowing for real-time analysis of potential GW events which is important for sending out alerts to the rest of the astronomical community. In fact, it was the online version of cWB which made the first detection of GWs, GW150914 [125]. Offline searches do not have the requirement to process data immediately, often concluding results a few months after an observing run has concluded. This allows them to tune specific components of the algorithms to mitigate the effect of detector glitches more efficiently, thus enhancing sensitivities and further increasing the chances of detecting a burst-type event with high statistical significance. Searches for all-sky short GW bursts have taken place in all three previous LIGO-Virgo-KAGRA (LVK) observing runs, with results detailed in the collaboration papers [126, 127, 85]. Thus far the all-sky short GW burst searches have detected many CBC sources, and with the improved sensitivity of current detectors and enhancements to data analysis techniques, there is increasing probability that we may detect transients from non-CBC sources.

2.1.1 Short transient sources

As mentioned above, the all sky short search aims to detect an array of short duration transient sources over a wide parameter space. While many of these sources are poorly modelled or have complex structure, we can still define sets of analytic waveforms which approximate expected signals from the search to be used in sensitivity studies. Some of these waveforms are created to represent regions of the parameter space in an ad-hoc manner, while others are modelled from astrophysically motivated sources. An overview of common waveforms in the all-sky short search is provided below.

Standard ad-hoc waveforms

The standard 'ad-hoc' waveforms are a set of injections that are commonly used in the all-sky short analysis to benchmark pipeline sensitivities to the short transient parameter space, and will be referred to multiple times in this thesis. This set of injections includes 4 generic types of waveform morphology, spanned over the entire accessible frequency and duration space: Gaussian pulse (GA), sine-Gaussian (SG) wavelets, band-limited white noise burst (WNB) and Ringdowns. The definitions of strain polarisations of these waveforms in timeseries are given below, with visual representations in Figure 2.1.

The Gaussian pulse (GA) waveforms are defined by the duration of one standard deviation τ_{GA} [128], contained within a single energy oscillation:

$$h_+(t) = A\exp(-t^2/\tau_{GA}^2), \quad h_\times(t) = 0 \tag{2.1}$$

where A is some multiplicative amplitude factor [1]. Injections are often made over $\tau_{GA} = [0.1, 4]$ms. An example of a GA waveform with $\tau_{GA} = 1$ms is seen in Figure 2.1a. As can be seen, only the plus strain polarisation holds GW energy here.

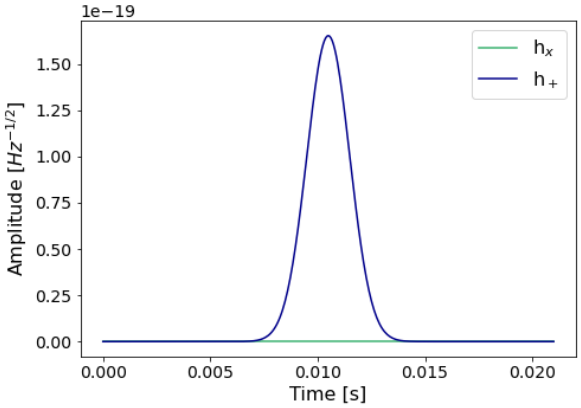
The sine-Gaussian (SG) wavelets consist of a sinusoidal signal defined by a Gaussian amplitude envelope, and are described by varying central frequency f_0 and Quality factor Q, which is the number of energy oscillations per cycle [128]:

$$h_+(t) = A\exp(-t^2/\tau^2)\sin(2\pi f_0 t), \quad h_\times(t) = A\exp(-t^2/\tau^2)\cos(2\pi f_0 t) \tag{2.2}$$

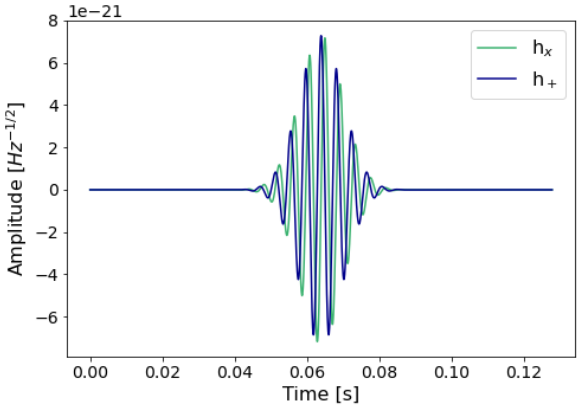
where $\tau = Q/(\sqrt{2}\pi f_0)$ and A is again an amplitude factor. Often the all sky short search considers SG waveforms with quality factors of $Q = [3, 9, 100]$, over a range of frequencies $f_0 = [70, 5000]$Hz to cover the low and high frequency parameter space. Figure 2.1b shows an example of a sine-Gaussian waveform with $Q = 9$ and $f_0 = 235$Hz.

Band-limited white noise burst (WNB) are bursts of uncorrelated white noise which is time shaped with a Gaussian profile. These are described by a lower frequency bound f_{low}, bandwidth

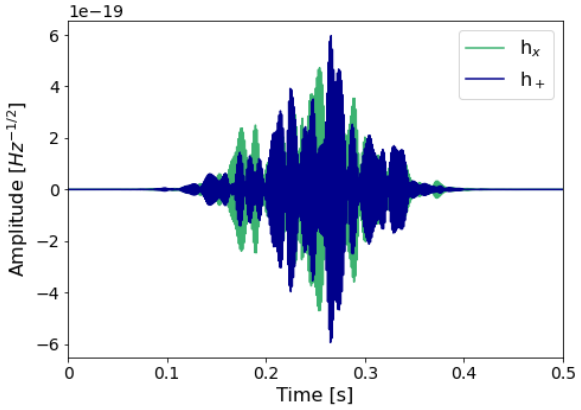
[1] for example the root-sum-square of strain, h_{rss}, introduced in equation 2.14



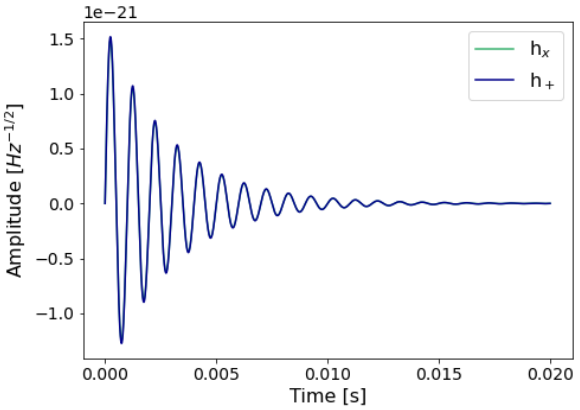
(a) Gaussian pulse: $\tau = 1ms$



(b) Sine-Gaussian: $f_0 = 235$Hz, $Q = 9$



(c) White Noise Burst: $f_{low} = 700$Hz, $\Delta f = 100$Hz, $\tau = 0.1$s



(d) Ringdown: $f_0 = 2000$Hz, $Q = 9$

Figure 2.1: Examples of ad-hoc waveforms commonly utilised in the all-sky short search. All plots show h_+ (blue) and h_x (green) polarisations displayed separately in the time domain.

Δf and duration τ_{WNB}. A WNB with $f_{low} = 150$Hz, $\Delta f = 100$Hz and $\tau_{WNB} = 0.1$s is shown in Figure 2.1c.

Ringdown waveforms show a sudden rise in amplitude with an exponential decay, defined by τ_{RD} and central frequency f_0 [128]:

$$h_+(t) = A\exp(-t/\tau_{RD})\sin(2\pi f_0 t), \quad h_\times(t) = A\exp(-t/\tau_{RD})\cos(2\pi f_0 t) \tag{2.3}$$

where $\tau = Q/(\sqrt{2}\pi f_0)$ and A is an amplitude factor as before. This waveform mimics the post-merger signal commonly seen in CBC systems, but can also closely resemble highly eccentric mergers. An example ringdown waveform with $f_0 = 2000$Hz and $Q = 9$ is seen in Figure 2.1d.

Core-Collapse Supernovae

Core-collapse supernovae (CCSN) are an astrophysical phenomena occurring from collapsing massive stars. When these explosions occur, they release gravitational energy which has the

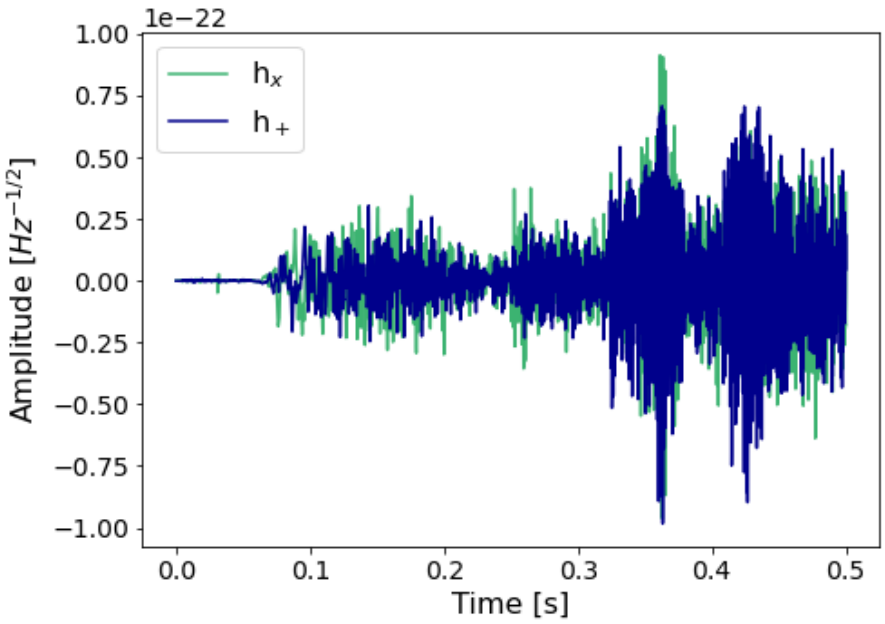


Figure 2.2: Example of a core-collapse supernovae waveform, based upon the m20 model from [132]. This model considers a non-rotating 20M$_\odot$ progenitor star with solar metallicity, containing a rise in frequency caused by g-modes. h_+ (blue) and h_x (green) polarisations displayed separately in the time domain.

possibility of being detectable with the advanced ground-based detectors [19]. There is uncertainty about the precise microphysics which occur during a CCSN event, and the emission of GWs from the core at the time of collapse would allow us to directly probe their inner dynamics [129, 130]. Due to the uncertain nature of these events, accurately representing the resulting GW signal is difficult, and requires complex, computationally expensive numerical simulations. The typical bank of CCSN waveforms used to test pipeline sensitivities in the all-sky short search [85] consists of 3 waveforms which represent a variety of CCSN (s18 [131], m20 [132], s9 [133]) and 2 of which represent extreme CCSN with higher amplitudes and lower expected rates (m39 [134] and 35OC [135]). An example of a CCSN waveform is seen in Figure 2.2, where the h_+ and $h_\times$ strain polarisations are shown for a solar-metallicity non-rotating progenitor of mass $20M_\odot$ (model m20 from [132]). A brief overview of all CCSN models used within this thesis is given in Appendix A.

Due to their computational cost and non-deterministic nature of the models, it is beneficial to include them in the un-modelled all-sky short search. Targeted searches for CCSN signals can also be completed, with information from known supernovae occurrences taken into account. Currently these are concluded with a null detection [108, 20].

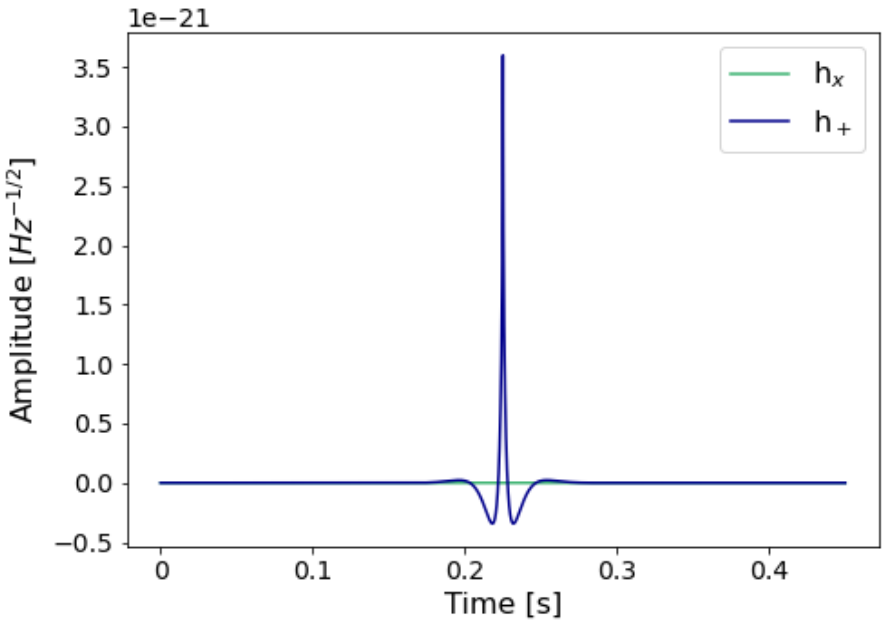


Figure 2.3: Example of a Cosmic string cusp waveform with f_{high} =150Hz. h_+ (blue) and h_x (green) polarisations displayed separately in the time domain.

Neutron star glitches

Neutron star (NS) glitches are the rapid spin-up in rotational frequency in isolated NSs, which cause a burst of gravitational radiation which may be detectable with current ground-based detectors. Depending on model parameters, these can have duration $\leq$ 0.2s and have frequencies from 1-3 kHz [82]. One of the most commonly theorised mechanisms for these glitches are excited f-modes in starquakes and super fluid crust interactions [85], although their exact cause is not certain. Recent works have explored prospects for detecting GWs from NS glitches without an electromagnetic counterpart [82] and from a known glitch in the Vela Pulsar [109], however up to this point all searches have yielded a null result. Despite this, high-frequency all-sky short searches have placed upper limits on emitted glitch sizes [85].

Cosmic strings

Cosmic strings (CSs) are hypothetical topological defects produced due to symmetry breaking phase transitions in the early universe, first proposed by Kibble in 1976 [136]. As CSs interact, there is a chance that they will form oscillating loops which decay through gravitational wave emission [77, 78]. The gravitational wave spectrum of oscillating loops is dominated by 2 string features: *cusps*, which are points in the string briefly travelling at the speed of light, and *kinks*, which are discontinuities due to collisions between strings [90, 137, 138]. Considered waveforms from cosmic strings have a lower frequency cut-off of 1Hz, while the upper-frequency cut-off varies over $f_{high} = [50, 1500]$Hz. CSs exist below this frequency band, however are outwit the sensitivity band of current ground-based detectors. Figure 2.3 shows an example

of a cosmic string cusp waveform with $f_{high} = 150$Hz. A bank of waveforms accounting for the three different cusp, kink and kink-kink interactions does exist, hence it is possible for targeted searches to be carried out via matched filtering. However the signal morphologies have low Quality factor, much like Gaussian pulses, thus are detectable within the all-sky search parameter space. Searches for GWs from cosmic strings have been carried out in previous LVK observing runs [97, 90], with null result.

Eccentric encounters

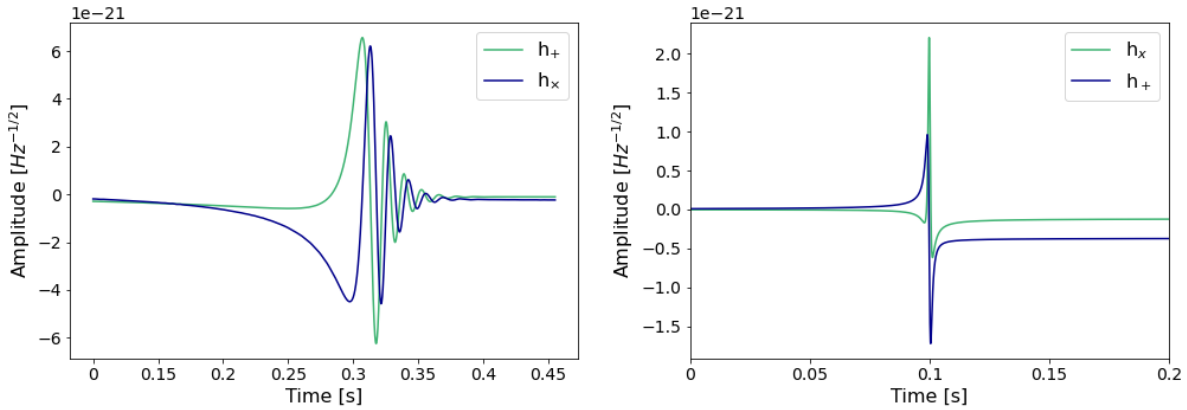
Compact objects can interact with one another via hyperbolic or parabolic trajectories in high density environments such as globular clusters and active galactic nuclei [139, 140, 141, 142]. Specifically, these trajectories can result in two phenomena which depend upon the angular momentum and energy in a given system: *radiation-driven capture* and *fly-by* encounters [84, 143]. Radiation driven capture between 2 compact objects results in a merger, however happens on a faster timescale than the standard compact binary mergers commonly detected in the CBC search. This feature causes the GW signal from such events to have a morphology similar to that of quasi-circular high mass BBH systems. Depending on the characteristics of the encounter systems, these may not be detected by matched filtering searches, or they may be mistaken for high mass quasi-circular events [144].

Fly-by encounters (sometimes referred to as hyperbolic encounters), involve 2 compact objects which scatter off each-other in unbound orbits without merging. When passing by each-other, energy is lost causing a burst of GWs [84, 145, 146]. In addition to this, as systems near the angular momentum limit between fly-by and radiation-driven capture, other signal morphologies may occur, such as head-on collisions [112] and objects interacting multiple times before merger [84]. Examples of waveforms for radiation-driven capture and hyperbolic encounter systems are seen in Figure 2.4a and Figure 2.4b respectively. The capture waveform is taken from [84], representative of a system with total mass $M_{tot} = 100M_{\odot}$, mass ratio $q = 2$, and angular momentum $L = 0.7$. The hyperbolic encounter waveform is from [147], representative of a system with total mass $M_{tot} = 4M_{\odot}$ and mass ratio $q = 1$.

Eccentric encounter signals are detectable in the all-sky short search parameter space, with various searches for GWs from radiation capture [4] and hyperbolic encounters [83, 147] with LVK data taking place in recent years. These searches have, thus far, not identified any statistically significant signals. A more thorough description of these interactions is found in Chapter 6, along with investigations into current sensitivities to these types of event.

Non-linear memory effects

Non-linear memory effects are a permanent distortion in space-time due to the passing of a GW, related to the non-linearity of GR [148, 149]. This effect may arise from any source of transient GW event, however the majority of literature details effects from CBCs . Memory effects from



(a) Parabolic BH capture from [84]

(b) Close hyperbolic encounter from [147]

Figure 2.4: Examples of waveforms from two eccentric encounter scenarios: (a) parabolic radiation driven capture with total mass $20M_\odot$, mass ratio $q = 2$, (b) close hyperbolic encounter with total mass $4M_\odot$, mass ratio $q = 1$. In both plots h_+ (blue) and h_x (green) polarisations displayed separately in the time domain

CBC events cause a GW signal which increases monotonically during inspiral, increases very suddenly during merger and saturates over ringdown. The sudden increase in signal amplitude during the associated merger phase appears as a single burst-like waveform, with duration related to the total chirp mass of the system [110]. For current detectors, such signals may be detected in the all-sky short search alongside the GW event that caused it [111]. Searches of this nature have taken place for previous LVK observing runs [110, 150].

2.2 The un-modelled search algorithm coherent WaveBurst

In the previous section, we introduced a wide range of waveform morphologies which are used to describe the short duration Burst parameter space. The requirement for burst algorithms to be sensitive across this wide parameter space, coupled with the expected signals having uncertain or complex morphology, means that they cannot rely on the typical matched filtering approaches [74] used in CBC searches. Instead they rely on the use of un-modelled techniques, in which GW signals need to be identified while making minimal assumptions on morphology. One such un-modelled algorithm is coherent WaveBurst (cWB) [114], which is commonly utilised for a variety of burst searches and has contributed to the detection of CBC events in previous observing runs [8, 9, 11].

In cWB, GW signal detection is based upon excess coherent power within a network of detectors, with little assumptions made on signal morphology or sky-localisation, making the algorithm susceptible to sources of non-stationary transient noise occurring from the detectors (*glitches*). With this in mind, cWB can be split into two main processes: *production*, which identifies significant instances of coherent excess power across the network of detectors, and *post-*

production, which aims to distinguish the coherent excess power of GW signals from glitches. During production, data from the gravitational wave detectors is processed to cluster any potential signal candidates as 'triggers': data is transformed into the time-frequency (TF) domain over different resolutions before data conditioning processes are applied to remove excess noise. Clusters of TF pixels above given energy thresholds undergo statistical attribute calculations which describe the features of the clustered pixels, and are labelled as triggers if they surpass coherent energy and correlation thresholds. The post-production stage then aims to optimise sensitivities of the search through the mitigation of detector glitches. While different techniques can be employed here, the standard cWB procedure is based upon the manual application of thresholds on triggers' summary statistics. These steps of the cWB algorithm are described in more detail below.

2.2.1 Identifying excess power in the time-frequency domain

The first step of cWB production transforms all data into the time-frequency (TF) domain through the Wilson-Daubechies-Mayer (WDM) wavelet transform [151], which is a combination of the Wilson-Daubechies transform and the Mayer wavelet method. The WDM transform allows for advantages of both standard methods above to be utilized: the energy is conserved and the inverse transform back to timeseries is well defined. The WDM transformation also comes with other benefits, namely: it uses the fast Fourier transform (FFT) for efficient transformation, uses shorter transformation filters for better time localisation, has better control over frequency localisation (spectral leakage), and uses a simple analytic expression for time-delay filters which benefits sky localisation for signals [151]. Next, the TF data undergoes data conditioning processes. A regression algorithm [152] is applied in order to remove noisy lines using information from physical environmental monitors on the detector sites. The data is then whitened through the normalisation of the TF data by an estimation of root-mean-square noise energy. This down-weights areas where noise is prominent, leaving the majority of TF pixels with energy close to zero, apart from those with the highest energy.

The algorithm must search the TF data to identify potential triggers based on excess energy in the TF pixels. These pixels are selected over multiple TF resolutions, through a set of energy threshold selection rules. The most energetic pixels are labelled as the *core* of a potential signal, and neighbouring pixels labelled as *halos*, to form a cluster. Two thresholds are applied: a lower energy threshold, E_0, and a higher energy threshold, $E_m = 2E_0$. For a pixel to be selected, it must either have energy above E_m individually, or have energy above E_0 *and* energy above E_m when taking a geometric mean over neighbouring halo pixels. Information on specific signal morphology can be considered in the pixel selection stage in order to increase sensitivity for targeted searches. Specifically, this information is detailed through using pre-defined 'wavepackets', in which the shape of coherent pixel clusters will be taken into account along with energy thresholds. Examples of this lie in the cWB targeted searches for intermediate-

and eccentric- binary black holes [113, 91]. If there is information on the sky localisation of a possible GW event from multi-messenger counterparts this can also be considered.

Once clusters of high energy pixels are formed over individual resolutions, they are combined to form a multi-resolution cluster. The coherent analysis of these clusters is completed through the maximisation of likelihood, discussed in section 2.2.2.

2.2.2 Calculation of likelihood and trigger statistics

The selection of coherent events across the network of detectors is based upon various thresholds describing the coherent energy content of the clustered pixels. Such quantities are found through the maximisation of the likelihood function over clustered TF pixels, allowing also the construction of the signal. The cWB constrained likelihood method is fully described in [153, 114, 115], however an overview of the process and how it leads to the creation of summary statistics is provided here.

The discrete series of clustered data over k detectors, $x_k[i]$ for pixel index i, is described by

$$x_k[i] = \mathscr{F} h_k[i] + n_k[i] \tag{2.4}$$

for GW signal h, network antenna pattern matrix $\mathscr{F}$ and noise n. The aim of the analysis is to construct a signal through the maximisation of a constrained likelihood based upon signal polarisations and sky location, $\Omega = (h_+, h_\times, \theta, \phi)$. To begin, the likelihood ratio is defined as:

$$\Lambda(x, \Omega) = \frac{p(x|h(\Omega))}{p(x|0)} \tag{2.5}$$

where $p(x|0)$ is the joint probability of the triggers being only instrumental noise and $p(x|h(\Omega))$ is the joint probability that a GW signal h is present in the data.

To expand eqn. 2.5 for a network of detectors, the noise-scaled data vector $\mathbf{w}[i]$ and noise-scaled network response vector $\boldsymbol{\xi}[i]$ are introduced:

$$\mathbf{w}[i] = \frac{x_1[i]}{\sqrt{S_{n1}[i]}}, ..., \frac{x_k[i]}{\sqrt{S_{nk}[i]}} \tag{2.6}$$

$$\boldsymbol{\xi}[i] = \begin{bmatrix} \frac{F_{1+}}{\sqrt{S_{n1}[i]}} & \frac{F_{1\times}}{\sqrt{S_{n1}[i]}} \\ ... & ... \\ \frac{F_{k+}}{\sqrt{S_{nk}[i]}} & \frac{F_{k\times}}{\sqrt{S_{nk}[i]}} \end{bmatrix} \mathbf{h}[i] \tag{2.7}$$

for noise power spectral density (PSD) over k-detectors S_{n1}, ..., S_{nk}, which gives an estimation of detector noise in the frequency range (see Section 2.2.6 for further details). The likelihood

function is defined as twice the logarithm of Λ, allowing it to be expressed as:

$$\mathscr{L} = 2(\mathbf{w}|\boldsymbol{\xi}) - (\boldsymbol{\xi}|\boldsymbol{\xi}) \tag{2.8}$$

where inner products of $(\mathbf{w}|\boldsymbol{\xi})$ and $(\boldsymbol{\xi}|\boldsymbol{\xi})$ are calculated over TF clusters. Finding solutions for eqn. 2.8 can be made simpler through the introduction of the dominant polarisation frame (DPF). Fully detailed in [153], this involves a basis transformation into a frame where detector antenna patterns are orthogonal and one polarisation is dominant over the other. cWB utilises the DPF where the h_+ polarisation is dominant, meaning the SNR contribution from the $h_\times$ polarisation is diminished by an alignment factor α which varies for detector networks. In the case of the nearly co-aligned LIGO-Livingston, LIGO-Hanford (LH) network, $\alpha \sim 0$ and the $h_\times$ polarisation is not detected, allowing for a fraction of TF clusters to be rejected as noise due to their un-physical solution to the likelihood function. This feature increases the LH network sensitivities compared to other detector combinations, such as the LIGO-Livingston, LIGO-Hanford, Virgo (LHV) network, which will be relevant to the work discussed in Section 4.2.

Following calculations in [153], the variation over $\mathscr{L}$ leads to a maximised likelihood statistic which may be written as

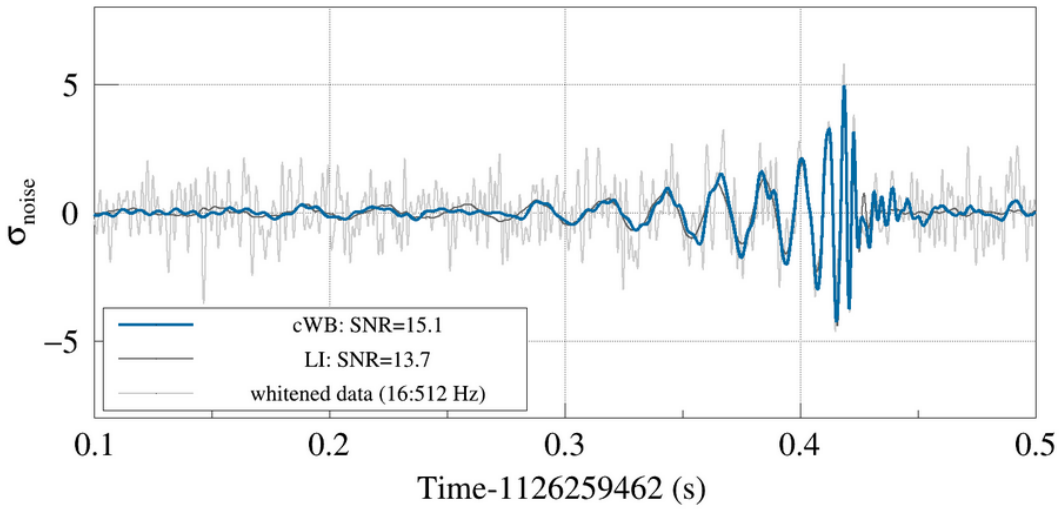
$$L_{max} = \sum_{i \in C} \sum_{n,m}^{K} \mathbf{w}_{nm}[i] P_{nm}[i] \mathbf{w}^T_{nm}[i] \tag{2.9}$$

where m, n are detector indices and P is the projection constructed from components of the unit vectors e_+ and $e_\times$:

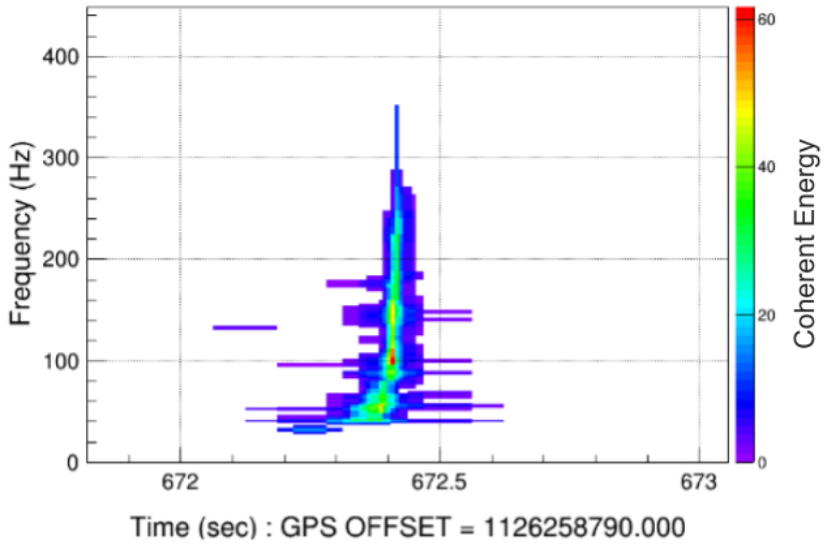
$$P_{nm}[i] = e_{n+}[i] e_{m+}[i] + e_{n\times}[i] e_{m\times}[i] \tag{2.10}$$

This L_{max} statistic has quadratic form, which can be split into two components based upon diagonal and off-diagonal terms. The coherent energy E_c statistic is constructed from off-diagonal terms, while the diagonal terms construct the incoherent energy E_{ic}. The Null energy, E_n, is defined as the residual between total energy and maximum likelihood energy, ie. $\mathbf{w}^2 - L_{max}$, and describes the minimum energy in the data which is inconsistent with the presence of a GW signal. These quantities provide insight on whether a trigger is likely to be a GW signal or glitch, since the majority of a signal's energy will be coherent. An example of a highly coherent signal in the TF domain is shown in Figure 2.5, for the first gravitational wave (GW) detection *GW150914* [154]. Here, the chirp of the signal is clearly seen through the cluster of coherent pixels. The time-series cWB construction of the signal is shown alongside for reference.

The E_c and E_n statistics alone do not provide enough information on trigger characteristics, and instead are used to construct a variety of cWB event statistics used to distinguish GW signals from noise. A large number of such statistics exist, so only those which are relevant to this thesis are described below.



(a) Signal representation in the time domain from LIGO-Livingston



(b) Signal representation in the time-frequency domain

Figure 2.5: Example of timeseries waveform (a), and cWB selected coherent pixels in the time-frequency domain (b) for GW150914. Both plots modified from the gravitational wave open science centre (GWOSC) online report of GW150914 with cWB [125].

The network correlation coefficient

$$c_{c0} = \frac{E_c}{E_c + E_n} \tag{2.11}$$

provides a measure of consistency across the detector network. The sub-network correlation coefficient is defined as $c_{c2} = E_{sub}/E_{sub} + N_{max}$, where E_{sub} is the reconstructed signal energy minus the maximum energy in a single detector, and N_{max} is the null energy in the time domain. Both network correlation coefficients will have $c_{c0,2} \sim 1$ for highly coherent signals, while triggers constructed from noise have $c_{c0,2} << 1$. With this said, it is possible for glitches to be highly coherent between detectors, so while a high c_c value indicates a probable signal, it does not equate to one.

Another important statistic is the network coherent SNR, which is the standard cWB detection statistic. For a K detector network:

$$\eta_c = \left(\frac{c_c E_c K}{K-1} \right)^{1/2} \tag{2.12}$$

Larger η_c points towards a GW signal, while lower values are more common of glitches. As will be seen in Section 2.2.4, cuts are often applied to the data based on the above c_c and η_c attributes in order to remove a large fraction of probable noise sources.

The 'penalty' statistic provides a measure of how concentrated the null energy is over pixels in the reconstructed TF cluster, defined as penalty $= E_n/N_{dof}$, where $N_{dof} = KN_{tf}$ for K detectors and N_{tf} number of TF pixels. A high value of null energy per pixel is a common feature of a glitch, thus triggers with high values of penalty may be disregarded as noise. The residual null energy, $N_{ED} = E_n - N_{dof}$, is a measure of how the null energy is spread across the TF pixels. It is expected that null energy may be spread over a higher number of pixels if the response of one of the detectors causes lower sensitivity to a GW event, due to the disbalance in detector contribution, whereas aligned detectors are more likely to have a lower pixel contribution to the null energy. Thus such a statistic can provide information on whether the distribution of null energy follows expected behaviour based upon the network of detectors. The final energy-based statistic to be introduced is $N_{norm} = E_c/E_{td}$, the ratio of coherent TF energy to reconstructed energy in the time domain. This statistic is a measure of the average number of resolutions used for event reconstruction, and will be high for GW signals, while a low value represents a reconstruction concentrated over a low number of TF resolutions - a likely artefact of noise.

As well as statistics relating to the coherence of triggers, there also exist some attributes which detail whether the selected trigger has properties common of known glitches, such as the Q_{veto} and L_{veto} statistics. $Q_{veto0,1}$ statistics are defined to identify the morphology of blip glitches [69], with Q_{veto0} estimating how much energy occurs outside the largest oscillation of the events constructed waveform, and Q_{veto1} describing the number of cycles in a waveform (ie. quality

cWB attribute	Description
E_c	Coherent energy across detectors
η_c	Coherent SNR across the detector network
c_{c0}	Network correlation coefficient
c_{c2}	Sub-network correlation coefficient
N_{ED}	Residual null energy measure
penalty	Null energy per pixel
N_{norm}	Ratio between reconstructed energy and total energy
Q_{veto0}	Energy distribution of an event over time
Q_{veto1}	Quality factor
L_{veto0}	Mean frequency of reconstructed signal
L_{veto1}	Root-mean-square frequency of reconstructed signal
L_{veto2}	Ratio of energy in the pixel to total energy of the event

Table 2.1: A list of statistical attributes calculated by the coherent WaveBurst (cWB) algorithm describing properties of reconstructed triggers. In particular the attributes here can aid in distinguishing physical GW signals from detector noise.

factor). It is expected for blips to be observed with low values of Q_{veto}. The L_{veto} attributes are used to identify narrow-band glitches frequently observed near power line frequencies. The $L_{veto0,1,2}$ attributes evaluate the mean and root-mean-square frequencies of the reconstructed triggers, and the ratio of pixel vs. total trigger energy respectively.

A summary of cWB statistics relevant for this thesis are seen in Table 2.1. For a complete list, one should refer to the cWB documentation pages [155]. As mentioned above, the creation of the clusters of coherent pixels and their calculated statistics is known as the production stage of cWB, in which reconstructed coherent wavelets are output as *triggers* with their statistical information attached. From here, the sensitivity of the pipeline is optimised during a post-production stage, in which trigger attributes are manipulated and selection cuts are applied to further distinguish GW signals from noise.

2.2.3 Background estimation

In GW searches, a data set consisting of only detector noise and glitches must be considered in order for the significance of any observed GW signals to be assessed, referred to as the *background* data. In cWB, the background is estimated by time-shifting data from the detectors so that it becomes non-physical for GW signals to be correlated in the data. The cWB algorithm performs two kinds of time-shifts known as *lags* and *super-lags*. Before time-shifts are applied, data from each detector is split into time segments. The length of these segments are defined by the user, however typically are set to be between 300-600 seconds. This lower limit is set based upon the minimum segment length where the wavelet transformation won't be affected, while the upper limit ensures computational efficiency. First the *lag* analysis takes place: within a set of single segments (ie. containing 1 segment from each detector in the network), the data

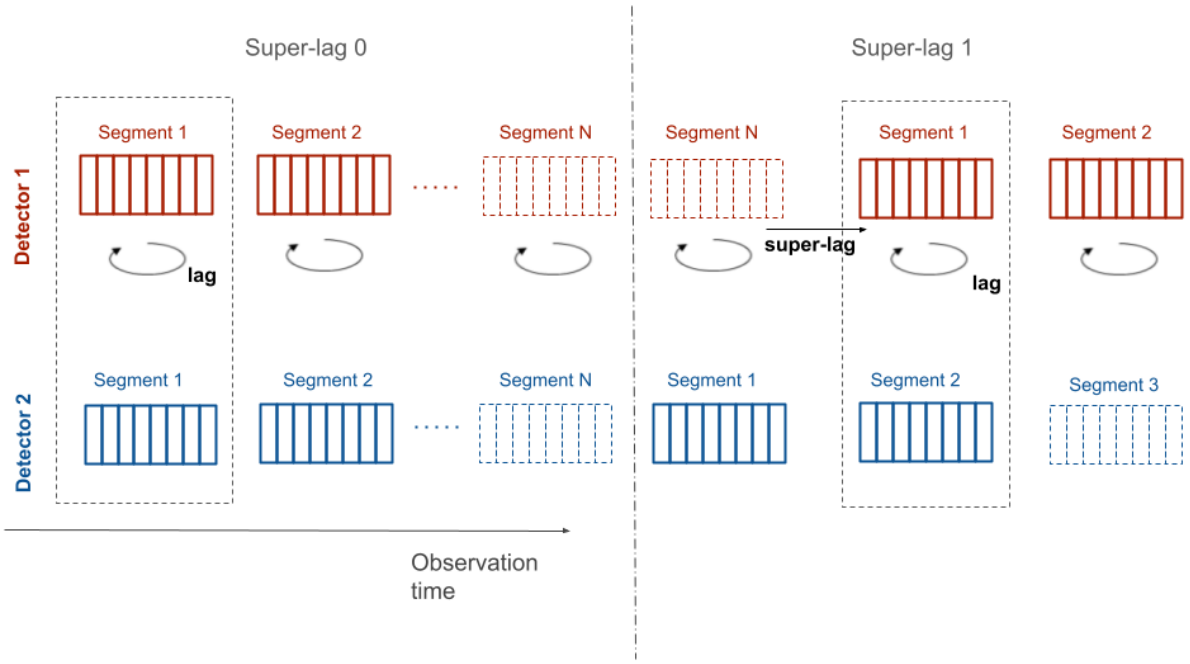


Figure 2.6: Simplified diagram of the lag and super-lag time-shifting process for the example of a 2-detector network with N data segments. Lag shifts are performed within time segments with respect to the other detector. Super-lags are applied by time-shifting entire segments of one detector with respect to the other, so that the lag time-shifting can be applied again. The entire process can be repeated for a set number of super-lags.

is shifted in a circular buffer by 1 second intervals with respect to the other detector segments in order to create a "local" background estimate, repeated for each set of segments within the observing time. Since this method of time-shifting is limited by the amount of data collected by the detectors, a secondary step of time-shifting known as *super-lags* is introduced: time-shifts are now performed between entire segments of data from one detector with respect to the other, creating *new* sets of segments in which the lag process can be applied. Since the data is shifted more in time than a physical GW could travel between detectors, it would be non-physical for any real GW event to have coherence in this data, leaving only glitches. This process allows for large amounts of background to be accumulated, usually of order a $\sim 10^2 - 10^3$ years. The time equivalent of background generated is referred to as background *livetime*, and is used to calculate the rate at which we will wrongly classify an event for significance estimates (see section 2.2.4).

A simplified diagram of the background time-shifting process can be seen in fig. 2.6 for the case of a 2-detector network. Here, segments from detector 1 are been shifted with respect to detector 2. The process shown can be repeated for a set number of super-lags until the desired amount of background livetime is created. The methodology can also be expanded to larger detector networks.

2.2.4 Standard method of cWB post-production in the all-sky short search

The next step in the search algorithm is to apply post-production methods, which involve manipulating trigger attributes to aid in the classification of events against detector glitches. Often some form of glitch mitigation technique may be applied to optimise sensitivities. In the case of standard cWB post-production, glitch mitigation occurs through the manual selection of thresholds on statistical attributes, and binning of triggers based upon loose criteria of signal properties. The background data described in Section 2.2.3 is then utilised to assign significance estimates (false alarm rate (FAR)) to triggers within each bin. Simulated waveforms are often injected into background data and run through the pipeline in order to estimate sensitivities of the search to a variety of sources, allowing for pipeline tuning. Once the search is optimised on sets of injected simulated signals, incoming data from the detectors is processed to search for GW events, known in cWB as the *zerolag* data.

The cWB offline search is split into two frequency bands in order to optimise sensitivities to expected populations in each band. The low frequency search observes data from 16-1024 Hz (this upper band was later increased to 2048Hz during O4), while the high frequency search looks at data from 512-4096 Hz. The work in this thesis concentrates on the offline low frequency (LF) search, and hence information in the remainder of this section is specific to this case.

False alarm rates

The false alarm rate (FAR) is a measure of how often a glitch is wrongly classified as a GW event. This statistic allows significance estimates to be associated with triggers of potential signals; in order for a significant detection of a GW to be made, we want FAR to be low. The FAR for an event is calculated using the background data: for a given amount of observing time that has accumulated a total background livetime of T_{bkg}, the FAR at a given value of detection statistic can be calculated by:

$$FAR = \frac{N_{th}}{T_{bkg}} \tag{2.13}$$

where N_{th} is the number of background triggers detected over the given detection statistic threshold. This gives a relationship between FAR and the detection statistic of the pipeline, allowing for significance estimates to be applied to processed triggers. In the case of standard cWB this detection statistic is η_c as seen in eq. 2.12.

From this we also introduce the inverse false alarm rate (iFAR), which is $1/\text{FAR}$. This creates a more intuitive statistic to characterise the trigger significance since iFAR increases with significance. Common iFAR thresholds for all-sky short searches are: iFAR of 1 year for significant burst candidates in online searches, and iFAR of 10 and 100 years for significant candidates in offline searches. The FAR is what directly motivates the accumulation of a large

amount of background data: if we have background data equalling less than 100 years, then we cannot reach a significance estimate of iFAR 100 years.

Post-production selection cuts and binning

The standard post-production approach of noise mitigation in cWB is to manually apply selection cuts and binning thresholds based upon trigger attributes which may be typical of noise. The first selection cuts applied is to coherent SNR η_c and network correlation coefficient c_{c0}. Since these attributes describe the coherent energy and correlation between detectors, triggers with lower η_c and c_{c0} are more likely to be glitches. For this reason, post-production starts with a manual cut on these attributes to remove a bulk of the unlikely triggers. Deciding on the threshold for these cuts is not straight forward, similarly for the majority of attribute thresholds involved in this process.

Despite the above efforts to distinguish potential GW candidates from glitches, some glitch types still contaminate the search. The most prominent and problematic glitch for the low frequency search is the blip glitch [69]. As mentioned in section 1.3.2, the source of these glitches is not known. They have low Quality factor with energy contained to a single oscillation, and often present as loud, highly correlated triggers within cWB. Their frequent occurrence and high correlation affects the FAR estimates of the search, ultimately lowering the significance of true signals. For the first 3 LVK observing runs, cWB mitigated the effect of blip glitches in the low-frequency search by applying a binning method: the data is divided into 3 classes based on the morphology of the reconstructed trigger. The definitions of the classes are largely based upon the Q_{veto} attributes: *C1* contains events with low quality factor and majority of energy contained in one oscillation, *C2* contains events with low quality factor but energy distributed over multiple oscillations, and *C3* contains events with large quality factor. By doing this, the majority of problematic blip glitches are contained to the 'dirty' *C1* class, meaning there is only a loud noise floor for this class and signals belonging to the 'clean' classes can be detected with higher significance. An example showing the benefits of the binning method is seen in Figure 2.7, where the FAR vs coherent SNR, η_c, is shown for the 3 different classes for the first-half of the third LVK observing run (O3a) [85]. It is seen that the *C1* background has a higher noise floor than the other classes, as the distribution has a long tail which levels out below FAR 1/10 years. While this is beneficial for signals contained in *C2* & *C3*, it results in short, low Quality factor signals being masked by the noisy blips and thus difficult to detect. In particular this can directly affect the detection of Gaussian pulses, cosmic strings and hyperbolic encounters.

Manually choosing the thresholds to define these class bins is delicate and time consuming, as the balance between increasing sensitivity for one class while decreasing for another needs to be optimised. Additionally, the separation of triggers into classes introduces a trials factor to the significance estimates; the false alarm rate of each trigger is penalised by a factor of 3 when quoting final significance.

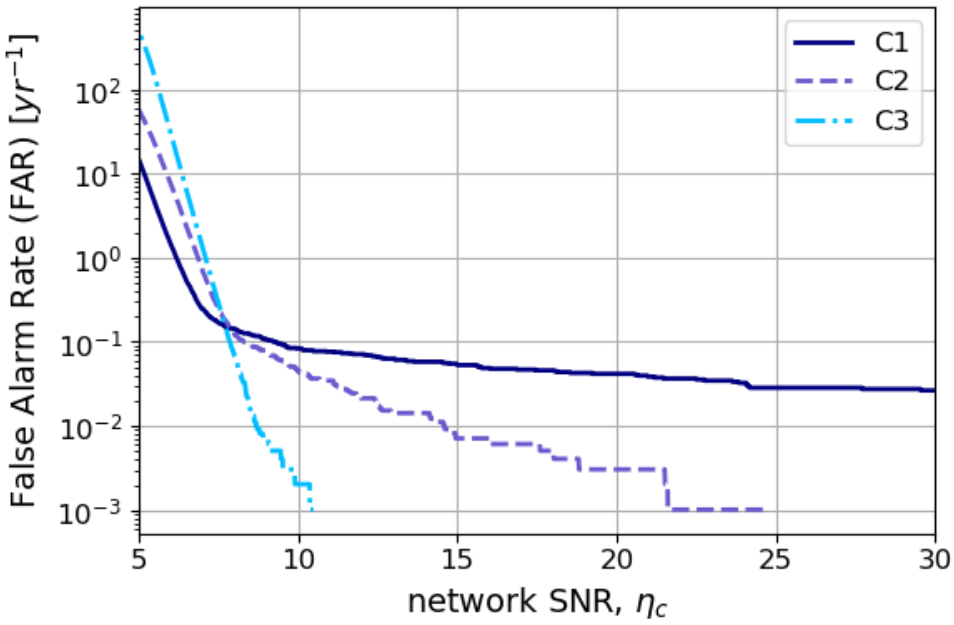


Figure 2.7: An example of FAR vs η_c distribution for the 3 cWB binned classes during the first half of the third Advanced LVK observing run (O3a), shown with increasingly lighter shades of blue. The C1 background contains a large number of problematic blip glitches, giving it a higher noise floor. This means events in clean classes C2 and C3 can be detected with higher significance, since they have competitive SNR-FAR thresholds.

2.2.5 Characterising search sensitivity with simulated data

The performance of pipelines needs to be tested on simulated data prior to analysing the real data. The main motivation of this is to ensure the analysis is working as expected, and allows for minor tweaks to be made to post-production techniques to optimise sensitivity. In the case of null detections, it also allows for upper limits on theorised burst populations to be set or for performance comparisons between different pipelines to be completed. While there are various ways to measure a pipelines overall performance to such simulated waveforms, there are a few key statistics and plots which are commonly referred to in the burst searches. These terminologies are introduced below and will be referred to throughout later chapters.

Simulated injections in cWB

cWB has a built-in mock data challenge (MDC) engine which allows for simulated waveforms to be injected into the pipeline. A set of waveforms will be defined along with chosen injection parameters to be inserted into the time-shifted background data. Some of the important injection parameters are: waveform amplitude (often root-sum-square of strain, h_{rss}, as defined in eqn. 2.14), injection rate, and sky distribution (can have uniform distribution or target a given sky location). These injection parameter settings can be tweaked to produce an array of desired burst populations.

Often the main population considered for benchmarking pipeline sensitivity in the all-sky short search is the ad-hoc set of waveforms detailed in section 2.1.1, since they represent a wide array of signals across the short transient parameter space. However, any of the populations detailed in section 2.1.1 can be injected to gain sensitivity estimates to astrophysically motivated sources.

Defining statistics to measure sensitivity

In order to measure the overall sensitivity of a pipeline to given waveforms, we can introduce the root sum square of the GW strain - ie. the h_{rss},

$$h_{rss} = \sqrt{\int_{-\infty}^{\infty} \left(h_+^2(t) + h_\times^2(t)\right) dt} \tag{2.14}$$

where h_+ and $h_\times$ are the polarisation components of the GW strain. As mentioned above, waveforms are often injected over given intervals of h_{rss}, representing different amplitudes of signal in the data.

We can also introduce the detection efficiency, ε, of a pipeline as the fraction of injections recovered, N_{rec}, over the total number of injections originally made into the pipeline, N_{tot}, ie.

$$\varepsilon = \frac{N_{rec}}{N_{tot}} \tag{2.15}$$

Often the detection efficiency is calculated for given significance thresholds or signal amplitude. Errors on detection efficiency are calculated through the binomial error, namely:

$$\delta_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{tot}}} \tag{2.16}$$

One common way to visualise pipeline performance is to plot the detection efficiency of waveforms as a function of source h_{rss}, $\varepsilon(h_{rss})$, for a given FAR threshold. From this plot, another important statistic is introduced; the h_{rss50} value. This corresponds to the h_{rss} amplitude where 50% detection efficiency is achieved. By fitting a cumulative lognormal function to the $\varepsilon(h_{rss})$ distribution, we can infer the h_{rss50} value for specific injected sources. The region of uncertainty on this distribution is found by fitting the lognormal function to upper and lower limits on detection efficiency from eqn. 2.16. The h_{rss50} statistic creates a consistent approach for comparing the sensitivity of different methodologies, and will be seen in later Chapters.

False alarm rate vs. detection statistic

Another important plot to visualise while testing pipeline performance is the relationship between false alarm rate (FAR) and the detection statistic for the given pipeline. Since FAR is calculated based upon the detection statistic associated to noise triggers, this plot directly shows

the relationship between the detection statistic and the corresponding significance estimate. If loud glitches are present in the analysis and not being dealt with efficiently, a long tail will be present in this distribution and significance thresholds will be associated with higher detection statistics. For the cWB analysis, this is the FAR vs η_c plot, and we have already seen an example in Figure 2.7. In this plot, the distribution with the long tail due to problematic glitches is the C1 background, while the other distributions hold better relationships between η_c and FAR.

False alarm rate vs. detection efficiency

The false alarm rate (FAR) vs. detection efficiency plot measures the detection efficiency of the pipeline for a given injected source population across binned FAR values. This plot is a receiver-operating-characteristic (ROC) curve, displaying the true positive vs false positive classifications of the pipeline. These plots are useful when investigating overall pipeline performance, as they show efficiency across all significance thresholds.

2.2.6 Analysing real data

Having discussed details on the cWB search algorithm, this section will provide the basic workflow of data from the detectors through cWB, including details on how 'real' data is processed in the search for GW signals. Within cWB, the real detector data is referred to as *zerolag* data, since it is the period of data with no time-shifts applied. A simplified diagram summarising this workflow is seen in Figure 2.8, with a brief description of the process provided below.

Pre-processing

First, data output from the detectors must undergo pre-processing to make it suitable for input into GW search pipelines, involving calibration, power spectral density (PSD) estimation and data quality procedures. The detector data must be calibrated and converted into time-series strain. This is done using a model of how the detector responds to passing GW signals, constructed by constantly running calibration injections into the detector in order to characterise the time dependence of the detector response [74, 156].

The power spectral density (PSD) of the noise to the detector must also be estimated for a given time period. The noise is assumed to be approximately stationary and Gaussian over small time segments, allowing the PSD to be estimated by Welch's method: the data is split into overlapping time segments and windowed, before a fast Fourier transform (FFT) is applied to each segment and the average over the squared magnitudes is computed [74]. The noise PSD estimate is used in calculations such as whitening and maximisation of likelihood.

Finally, the data undergoes data quality checks in order to remove periods of time with identified disturbances to the detector through vetoes. First introduced in Section 1.3.2, vetoes are split into 2 categories, CAT1 and CAT2. CAT1 vetoes define time periods where there are

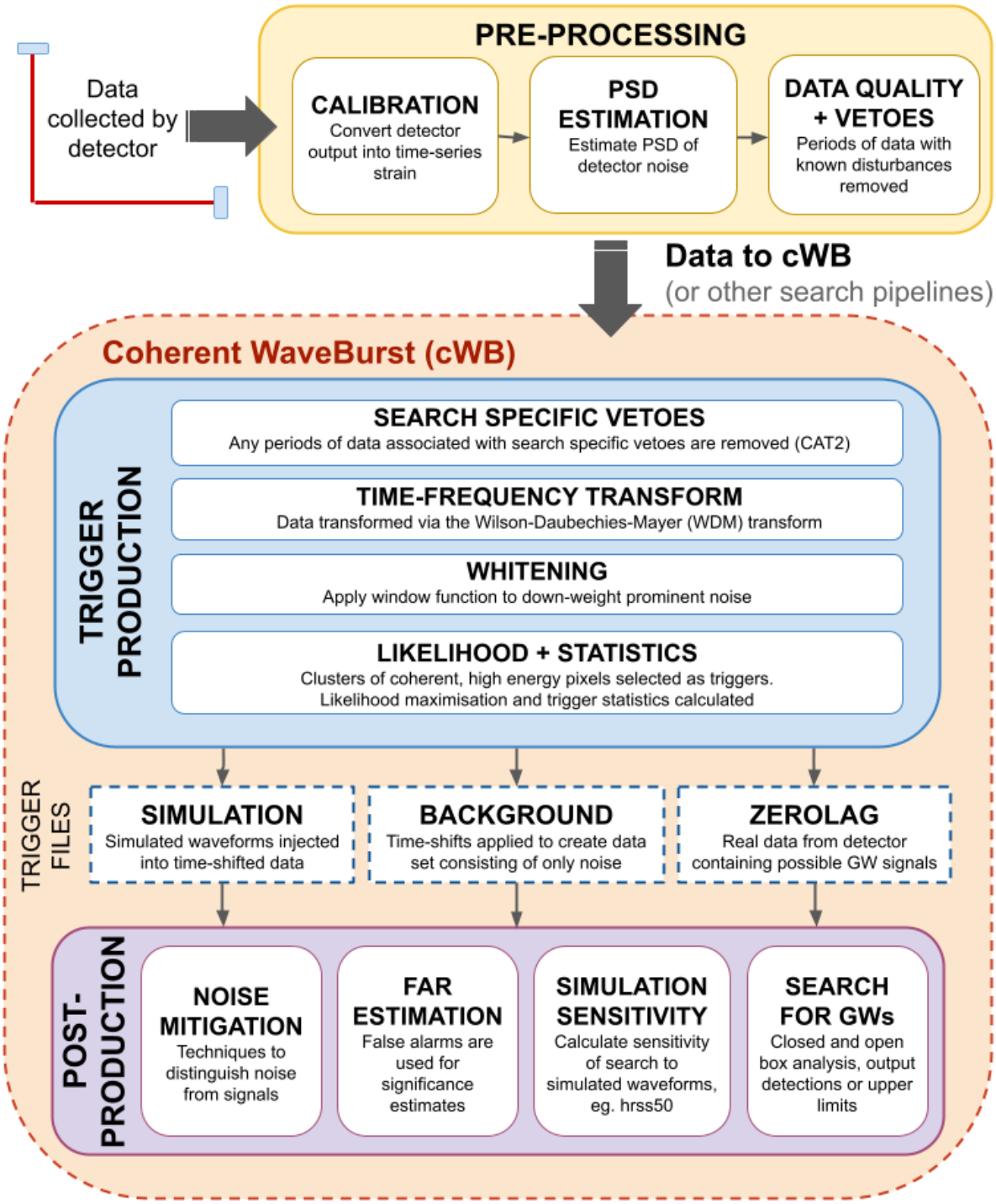


Figure 2.8: Overview of the data workflow from detectors through the cWB pipeline in the search for GW transients. Data first undergoes pre-processing procedures to prepare the data for input into search algorithms. It is then processed through search pipelines, in this case cWB. The cWB trigger production stage can be seen in blue, involving the application of search-specific vetoes, TF transform, whitening and summary statistic calculation. This produces 3 types of trigger files: simulation, background, zerolag. Each of these files are required in the post-production stage, seen in purple, involving processes such as noise mitigation, FAR estimation, estimation of sensitivity to simulated signals and the search for GW signals on zerolag data.

major known disturbances occurring in the data, eg. that would bias PSD estimation, hence these are often applied to all data as they are relevant across all transient searches. CAT2 vetoes define instances where a noise source has known coupling to the detector output, often identified through correlation in detector sensor channels. These may be defined for specific searches, eg. burst searches, and can be applied after initial data processing [68].

cWB trigger production

After calibration, data quality vetoes and data conditioning have been applied, cWB decomposes the data into multiple TF resolutions via a wavelet transform, as discussed in Section 2.2.1. High energy pixels are clustered over the multiple TF resolutions within single detector data, and are labelled as triggers if pixel clusters are coherent over the network of detectors, calculated through the likelihood maximisation. As discussed in Section 2.2.2, a range of summary statistics are calculated for each trigger, which may describe noise artifacts or GW signals.

The cWB trigger production stage outputs 3 types of data:

- Background: The background estimation is performed by correlating multiple detector data where time-shifts larger than the light travel time across the Earth are applied. Any correlations observed in the time-shifted data cannot be caused by GWs. This estimation of background noise is used for FAR estimation, as discussed in Section 2.2.3.

- Simulation: Simulated signals are injected into a background of noise based upon the estimation of the PSD to mimic the presence of GW signals. This data is used for characterising the pipeline sensitivity to selected signals before being applied to the real data, as discussed in Section 2.2.5.

- Zerolag: Zerolag data consists of constructed triggers from the 'real' detector data, ie. with no time-shifts (lags) applied. This is the data in which we search for GWs.

cWB post-production

The trigger files output from cWB production then undergo post-production procedures. The first of these is noise mitigation techniques, which attempt to down-rank the effect of glitch triggers in the data. Examples of such techniques are the statistic selection cuts and class binning of the standard cWB post-production (Section 2.2.4). Other techniques can aid this process, such as Gaussian mixture modelling (GMM), as will be seen in later sections. FAR estimates are calculated from the background triggers, which are used to assign significance to simulated or zerolag data.

Once the algorithm has been tested and optimised with the simulated data as described in Section 2.2.5, the real GW (zerolag) data from the detectors is analysed. Overall the post-production process for this is fairly straight forward, and involves the same noise mitigation

techniques and FAR estimates as detailed above applied to the zerolag data. However, first we must analyse what is known as a 'fake' zerolag. This is a randomly chosen lag, ie. a segment of data which has undergone time-shifting. By processing this through the algorithm, it can be ensured that the significance of triggers agrees with the expected background. Once we are certain the algorithm is working as expected, we can process the real GW data, and investigate whether any significant detections observed are likely to occur from CBC systems or a burst source. The above process will be completed in both Chapters 4 and 5 of this thesis.

2.3 Gaussian mixture modelling as a detection method

In this section, the method of Gaussian mixture modelling (GMM) is introduced as a tool to aid in the detection of GW signals. As has been mentioned, one of the biggest hurdles for model-agnostic search algorithms are the noise transients occurring within the detectors. Techniques can be applied to mitigate these during the post-production stage of search algorithms, which in the case of cWB involves selection cuts through the manipulation of statistical parameters. Instead, we can apply GMM as post-production to such algorithms in an attempt to improve the classification of signals and glitches. GMM and how it can be utilised as a detection technique is detailed below for generic weakly-modelled searches. The full incorporation of GMM into the cWB pipeline is described in Chapter 3 with investigations into optimising the methodology.

2.3.1 Gaussian mixture modelling

Gaussian mixture modelling (GMM) is a supervised machine learning method that allows for multi-dimensional data to be represented as superpositions of uni-modal Gaussians, constructing a probabilistic model which represents normally distributed sub-populations within one overall population. GMM has a variety of uses in science, such as density estimation, anomaly detection, data generation, sub-population recognition, and as a classification tool. Even within the GW community, GMMs have been utilized for a wide range of tasks, such as population inference of binary neutron stars [157, 158], inferring the black hole mass distribution [159], and to aid in the rapid localisation of GW events [160]. Here, we are interested in GMMs ability to classify events from 2 different populations, allowing the application as a tool to aid the detection of GW signals.

To create a GMM model from a dataset, we must assume that all data points are generated from a mixture of Gaussian distributions with unknown parameters. Given a dataset $\mathbf{x}$ with d attributes (or *dimensions*), we can construct the model representing the structure of $\mathbf{x}$ as a superposition of K Gaussians in the following way [1]:

$$p(\mathbf{x}) = \sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}|\mu_j, \Sigma_j) \tag{2.17}$$

where μ_j and Σ_j are the mean and covariance of each Gaussian within the model, while each Gaussian is weighted by w_j, normalised to satisfy $\sum_j w_j = 1$, and the Gaussian density function is given by:

$$\mathcal{N}(\mathbf{x}|\mu_j,\Sigma_j) = \frac{\exp[-\frac{1}{2}(\mathbf{x}-\mu_j)^T\Sigma_j^{-1}(\mathbf{x}-\mu_j)]}{((2\pi)^{d/2}|\Sigma_j|^{1/2})} \tag{2.18}$$

Gaussian mixture models can be created to represent a variety of distinct populations with complex multi-dimensional structure, with the individual Gaussians within each model optimised based upon a portion of data from that given population. For the application of GMM as a detection tool, we consider the noise and signal parameter space as distinct populations and construct a model for each of these. The signal parameter space can be represented by a set of simulated waveforms injected into a given search algorithm, while the noise parameter space can be represented by detector noise. The models are optimised on each population completely separately, the details of which are in the section below.

2.3.2 Optimising Gaussians in a model

Optimising Gaussian parameters and model hyper-parameters are referred to as *training* the model. It is within this process that the model learns the distribution of our specified data. While the process discussed here is general to optimising Gaussian mixture models in general, for this work the GAUSSIANMIXTURE package from SCIKITLEARN [161] is used. This allows for model initialisation parameters to be set easily and optimisation of hyper-parameters completed.

In order for the GMM model to effectively represent a set of multi-dimensional data, we must optimise the parameters of the Gaussians to the given population. A common technique for this is the maximum-likelihood approach, in which the log-likelihood is maximised with respect to the mean μ_j, covariance Σ_j and weights w_j of each Gaussian in the model. Considering the same situation as above, where we have d-dimensional data containing n points as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, and following the derivation as seen in [1], the likelihood function is given as:

$$p(\mathbf{X}|\Theta) = \prod_{i=1}^{n} p(\mathbf{x}_i|\Theta) \tag{2.19}$$

where $\Theta = w_j, \mu_j, \Sigma_j$ are the Gaussian parameters for each component $j = \{1, ..., K\}$. From this, the total log-likelihood is constructed as the sum over n individual log-likelihoods:

$$\ln \mathscr{L} = \sum_{i=1}^{n} ln(p(\mathbf{x}_i|\Theta)) = \sum_{i=1}^{n} ln\left[\sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}_i|\mu_j,\Sigma_j)\right] \tag{2.20}$$

To optimise the model parameters Θ we need to maximize $\ln \mathscr{L}$ with respect to each of these individual parameters. For example, first maximising over the mean of the kth Gaussian μ_k:

$$\frac{\partial \ln \mathscr{L}}{\partial \mu_j} = 0 \implies \hat{\mu}_k = \frac{\sum_{i=1}^{n} r_{ik}\mathbf{x}_i}{N_k}, \tag{2.21}$$

where

$$r_{ik} = \frac{w_k \mathcal{N}(\mathbf{x}_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} w_j \mathcal{N}(\mathbf{x}_i|\mu_j, \Sigma_j)} \quad \text{and} \quad N_k = \sum_{i=1}^{n} r_{ik}. \tag{2.22}$$

Thus the maximum likelihood estimate of the mean μ_k of the kth Gaussian is the weighted mean of all the data points, and all coefficients are implicit functions of μ_k itself via the normal distributions. We can see similar behaviour for the other hyper-parameters, for example maximizing $\ln \mathscr{L}$ with respect the covariance matrix Σ_k of the kth Gaussian gives

$$\hat{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{n} r_{ik} (\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)^T. \tag{2.23}$$

Lastly, maximization over w_j under the constraint that the sum of the weights adds up to unity can be obtained through the application of Lagrange multipliers [1], giving the maximum likelihood estimate for weights as

$$\hat{w}_k = \frac{\sum_{i=1}^{n} r_{ik}}{n} = \frac{N_k}{n}. \tag{2.24}$$

As is demonstrated above, the maximum likelihood estimate of each Gaussian parameter includes implicit functions of themselves, which makes this a difficult problem to solve analytically. Instead, a numerical solution is reached through the expectation maximisation (EM) algorithm [162]. This is an iterative algorithm, which first estimates trial values for each of the parameters, updating with new parameter estimates through iterative steps until converging on some maximum likelihood. While this algorithm is able to optimise the individual Gaussian parameters efficiently, it is not able to optimise on the total number of Gaussian components to include in the model. One of the most common ways to estimate the optimal number of Gaussian components is the Bayesian information criterion (BIC) [163]:

$$BIC = K \ln(n) - 2\ln(\hat{\mathscr{L}}) \tag{2.25}$$

This includes a penalty term, seen in the first term of eq. 2.25, which aids in the prevention of over-fitting Gaussians by penalising the model for having too many components for a data set of size n, which would ultimately cause the BIC score to increase. Because of this, the optimal number of Gaussian components for a given model is given by minimising the BIC score. A very simplified example of this is seen in Figure 2.9, in which the BIC score is shown against number of Gaussian components per model for some 2-dimensional data set. In this example we see that the minimum of the BIC distribution, and hence the optimal model structure, lies at 3 Gaussian components.

In theory the BIC is a good choice for selecting the optimal number of Gaussians in a model, and is effective in many applications of GMM. However, this technique starts to become less reliable for large, complex datasets, often not reaching a true minima or under-fitting the structure of the model, which can be a problem in the application to GW data. Other common methods

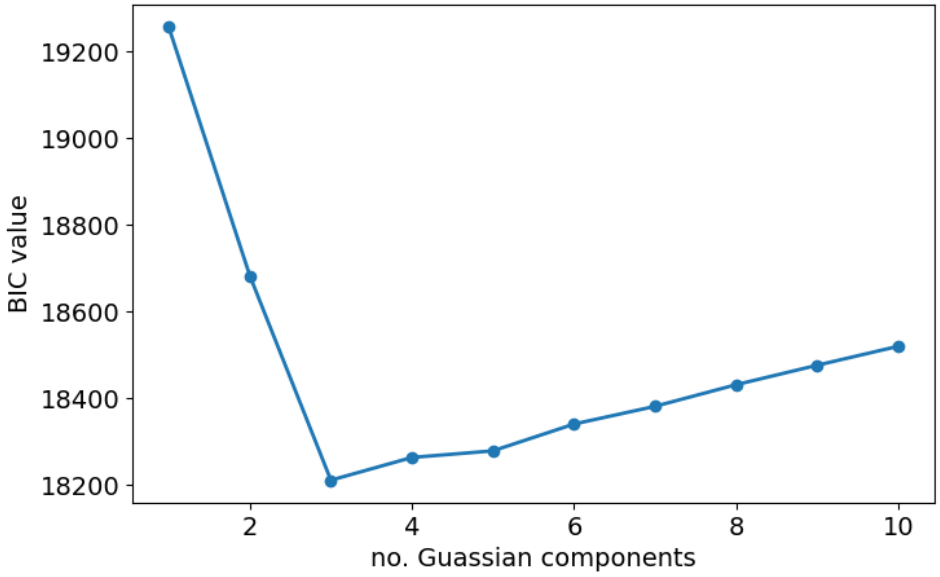


Figure 2.9: A simplified example of the Bayesian information criterion (BIC) score vs. the number of Gaussian components for some randomly generated 2-dimensional dataset. Minimised BIC score signifies optimal number of Gaussians components for a given model, in this case 3.

can be considered, such as the *Akaike Information Criterion* (AIC) [164], or *Silhouette score* [165], however these also have their downfalls. The reliability of BIC in the application to GW models is investigated in Chapter 3, and a new methodology is proposed.

In addition to selecting the number of Gaussian components for the model, an initialisation method must be chosen to first place the Gaussians before the optimisation of Gaussian parameters can begin with the EM algorithm. This choice will affect the initial placement of Gaussian centroids, and can be an important factor in reaching convergence with the EM algorithm. Various options exist for this, with the most common being randomly placed centroids or informed placings from k-means based algorithms. For k-means, the centres of K clusters are originally randomly selected, and surrounding data points are assigned to their closest cluster by euclidean distance. Centroid locations are then updated based on the mean of all data points in the given cluster, with this process repeated until optimal initialisation is reached. While this method is more trustworthy than selecting the centroids completely randomly, it is still heavily reliant on the initial selected location which can occasionally cause convergence on local minima. Other initialisation techniques can be adjusted to alter the behaviour of the models, such as the convergence threshold, the maximum number of iterations allowed in the EM algorithm, or the number of initialisations performed. The optimal settings are dependent on your data and model application, and must always be balanced with computational efficiency. Initialisation techniques will be explored further in Chapter 3.

A very simple example of a Gaussian mixture model created with the SCIKITLEARN package is seen in Figure 2.10, where the model has been fit on some 2-dimensional data. For this

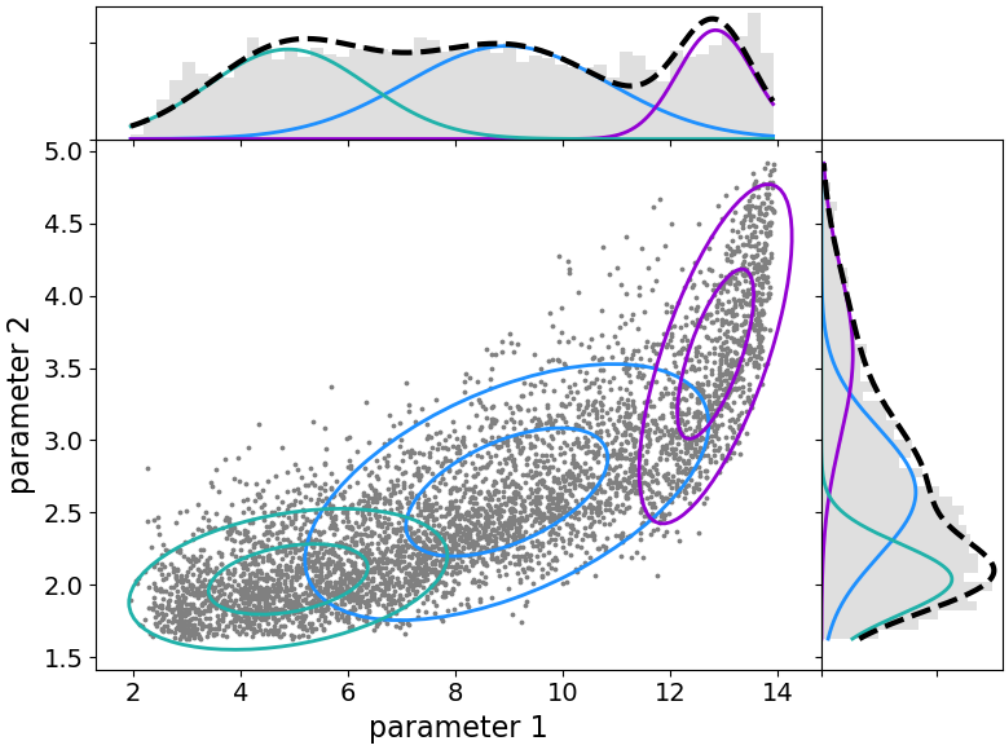


Figure 2.10: A simple 2-dimensional example of how a GMM can model multi-modal data with multiple Gaussians. Sample data is shown for 2 parameters in grey, while the GMM is constructed of 3 Gaussians in purple, blue and teal, with optimised Gaussian parameters. The superpositions of the Gaussians is shown by the black dashed line.

case, 3 Gaussians with differing mean, covariance and weights have been fit to the overall distribution of multi-modal data. By looking at the superposition of these Gaussians (represented by black dashed lines), we see that the distribution of each parameter can be well represented.

2.3.3 Defining a detection statistic

For the application as a detection tool, separate GMMs are optimised for the signal and noise populations, allowing for the construction of a log-likelihood ratio detection statistic which can distinguish which population a given data point originated from. This is known as the T statistic, formed via the maximum log-likelihoods from each model:

$$T = W_s - W_n \tag{2.26}$$

with W the log-likelihood for optimal number of Gaussians K:

$$W = ln(\hat{\mathscr{L}})|_{\hat{K}} \tag{2.27}$$

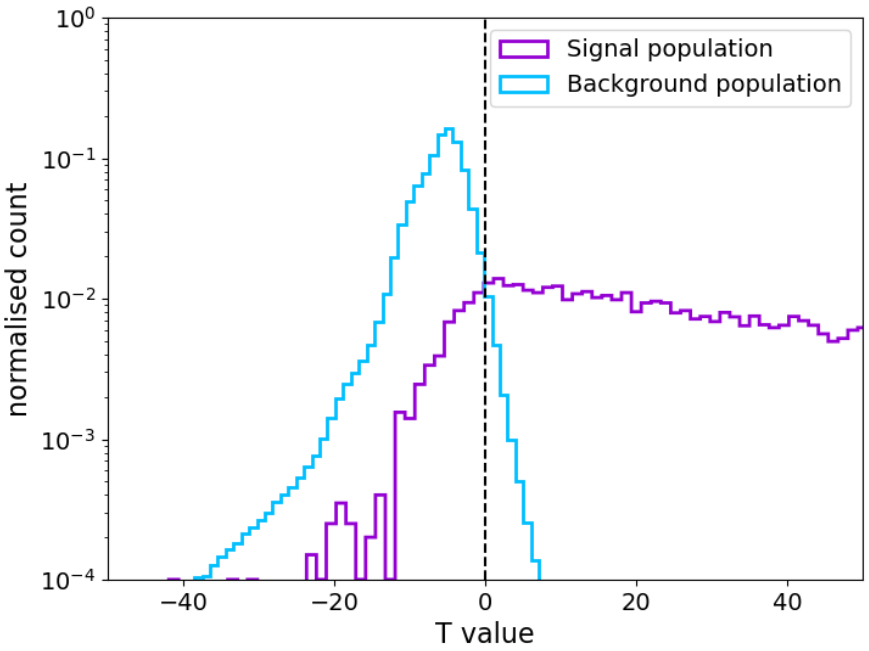


Figure 2.11: Plot showing normalised count of data over T value for a noise population (blue) vs a signal population (purple). The separation of populations in the T space demonstrates the ability of GMM to distinguish between the two populations

Here W_s and W_n refer to log-likelihood of a given data point belonging to the signal and noise populations respectively, also known as the 'scores'. From eq. 2.26 glitches are classified with low, negative T values, while likely GW signals are classified towards positive T. An example of this distribution is seen in Figure 2.11, where data from a signal and noise population are shown in a histogram over T. It is seen that the noise population has distribution over negative T values, while a large proportion of the data from the signal population have higher, positive T. The better separation the distributions have in this space, the better we can classify GW events.

Portions of data from both populations are usually reserved for calculating T after the models have been optimised, referred to as *test* data. Test data from the noise population is used to calculate a range of FAR estimates for the search, as described in eq. 2.13, allowing for the construction of significance thresholds associated with certain T values to apply to potential events. Test data from the signal population is the data we want to apply significance estimates to: the T values of this data can be estimated in the same way, with the distribution of T with respect to FAR allowing these triggers to have a significance value assigned to them, which ultimately classifies them as glitch or signal.

2.3.4 Testing, training and validation

Aside from the assumption that the overall population can be modelled as mixtures of Gaussian distributions, which prefers data attributes with Gaussian behaviour, we must also have enough data available to split each population into 3 sub-datasets: validation, training and testing data. The training data is used to fit the models and optimise the parameters of each Gaussian, so must contain enough data points to be fully representative of the populations. Validation data is used to optimise the number of Gaussians within the models, ensuring the models are not over-fitting to the training data. Lastly, test data is reserved for the estimation of significance to possible events. For all of these data sets, we want to use data which represents the noise or signal distributions well. As mentioned above, for the noise distribution a 'background' data set is used, which consists of time-shifted GW data so that only detector glitches are considered. Usually, we want to consider background data created from the same time period as being analysed, so that we have an accurate representation of glitches which are occurring in the detectors. For the signal population, we want to use data which is generic to as many short transient burst signals as possible and covers the search parameter space well in order to maintain the un-modelled nature of the search. For this, simulated waveforms can be injected into data, such as a selection of generic ad-hoc injections seen in section 2.1.1.

With this said, the analysis can be transformed into a more targeted search if desired. This can be done by selecting a chosen population of source waveforms, such as an astrophysical population detailed in section 2.1.1 as training data for the signal model. The choice for this could be to include only this population in the model training for an analysis which only targets a given source, or to reinforce the training data with injections from given populations if we still want the search to be sensitive to multiple waveforms.

Chapter 3

Enhancement of the cWB all-sky short search with GMM post-production

The coherent WaveBurst (cWB) algorithm detailed in Section 2.2 is able to successfully identify possible gravitational wave (GW) signals while making minimal assumptions on signal morphology by generating coherent coincidences across a network of detectors during the algorithm's production stage. These coincidences can either be a GW signal or a detector glitch, and so the algorithm has post-production processes which aim to separate between these types of event. As detailed in Section 2.2.4, the standard methodology for post-production in the cWB algorithm places manually chosen selection cuts and splits coherent events into search bins based on their statistical properties in order to contain the majority of problematic glitches to one bin. While this benefits the significance estimates of GW signals in the two 'clean' bins, it means that there is a large noise floor for low quality factor, single-energy cycle signals in the noisy bin due to the presence of blip glitches. This large noise floor ultimately affects the sensitivity of the search to signals such as Gaussian pulse waveforms, cosmic strings, or hyperbolic black hole encounters. It also results in significance estimates being penalised by a trials factor of 3, which can affect the significance of detected GW signals in other regions of the parameter space. In addition to this, selecting the thresholds and attributes on which to define selection cuts and class bins is a delicate and challenging task, requiring a lot of manual investigation.

In an attempt to improve the post-production efforts for cWB, we explore machine learning algorithms which can be coupled with the pipeline to aid the classification of GW signals. This has somewhat been attempted in the past, with [166] outlining a neural network method to aid in the detection of binary black hole (BBH) signals, and [167] aiding the identification of core-collapse supernovae (CCSN) signals in single-detector data through a supervised Machine Learning algorithm. While these methods are effective at enhancing cWB's efficiency in identifying targeted populations and have benefits specific to said sources, they introduce a model dependence into the pipeline which is ineffective for all-sky short burst searches. Other previous work in [168] utilised a boosted decision tree multivariate analysis in an attempt to in-

crease signal classification in cWB without making assumptions on signal morphology. Instead, the work here details use of the supervised machine learning method of Gaussian mixture modelling (GMM) as a post-production approach, which allows for enhanced classification of signals against glitches while maintaining the weakly-modelled nature of burst searches. Recent works have involved the application of other supervised machine learning methods as post-production to cWB, namely extreme gradient boosted decision trees (XGBoost) [123, 124]. While using a different concept, the analysis has the same aim of increasing signal classification while maintaining an un-modelled approach.

With the concept of GMM as a detection tool introduced previously in section 2.3, we now detail the method in the context of post-production to cWB. GMM allows for the modelling of multi-dimensional distributions of the coincident events' statistical properties, which may include relationships which were previously unseen to the human eye. This has the potential to aid in the classification of signals against generic glitches, specifically mitigating blip glitches, in order to enhance the estimated significance of signals. The application of GMM+cWB was first introduced in work by Gayathri V. *et al.* [1], while D. Lopez *et al.* presented all-sky short search results on the first half of the LIGO-Virgo-KAGRA third observing run (O3a) with the cWB+GMM methodology in [2]. While these applications were successful, it is possible to improve the classification of signals vs. noise by optimising the cWB+GMM approach. This Chapter will describe work completed by the author involving new enhancements made to the GMM methodology, through investigations on cWB data in the LVK O3a run. In section 3.1, the concept of GMM as a detection aid to cWB is introduced, building upon the generic introduction of GMM from section 2.3. Details of the new methodology enhancements are given in section 3.2, through investigations into model performance. Such investigations include the exploration of an optimised approach for the selection of number of Gaussian components, new sets of training data, alterations to the considered attribute set and model fitting methods. Subsequent to the above investigations, the sensitivity of the optimised GMM methodology to selected short duration GW sources is tested through direct comparisons to the previous methodology from [2] in section 3.3.

3.1 Application of GMM as post-production to cWB

In this section the details of applying the GMM methodology as a post-production classification tool to the cWB algorithm for the all-sky short search are discussed. The GMM analysis is applied to coherent events from the cWB production stage, ie. once coherent clusters of pixels have been collected across the detectors, with event reconstruction and statistical attribute calculations complete, but before any selection cuts or binning is applied. We refer to reconstructed events of this form as *triggers*. Before analysis can begin, a set of attributes must be chosen to represent triggers in the multi-dimensional space. Ensuring these attributes are distinguish-

able between GW signals and detector noise, training data sets containing said attributes will be selected to represent the background and signal populations, typically consisting of detector glitches and some set of simulated waveform injections respectively. Model Gaussian parameters are optimised on this data, while the number of Gaussian components within each model is optimised on a reserved portion of the same data, the validation data. A portion of noise data is reserved for the calculation of significance thresholds, in which glitches are classified with T values to construct false alarm rates, as introduced in eqn. 2.26. These FAR estimates will eventually be applied to sets of signal data to apply significance estimates, whether that be simulated injections to test sensitivity or possible GW events from real data. The following sections will explain the above process in more detail.

3.1.1 Choice of cWB data

As mentioned in section 2.3, we first must define 3 classes of data to be used in the analysis: validation data, which is used to tune the hyper-parameter number of Gaussian components in the models, training data, which is used to train the models themselves, and test data, which is used to construct significance thresholds and estimate the sensitivities to sets of generic waveforms. Since we are distinguishing between noise and signal populations, we need to select data which represents the characteristic features of signal or noise in the multi-dimensional attribute space so that we have representative models, while also being as generic as possible to particular signals in order to reserve the un-modelled nature of the search. In the application to cWB, this data is an output of the trigger production stage.

For the case of the noise model, we randomly split background data generated by cWB into validation, training and testing data. As is described in Section 2.2.4, this background data is generated by time-shifting the detector data so that only glitches are present, meaning the output of cWB production is triggers with attributes which represent only noise. We want to consider background that has been time-shifted on data from the same time period we are searching over so that the data will contain glitches that are common to the real GW data being analysed, making it a good representation of the noise space.

For the case of the signal model, we similarly randomly split simulated signals which have been injected into cWB into validation, training and testing data. These simulated signals can represent any desired population of sources, however for application to the all-sky short search we want a set of simulated waveforms which represent the wide parameter space. Previous GMM methodology in [2] used a portion of the ad-hoc waveform simulations detailed in section 2.1.1. While the morphology of this set of waveforms fit the requirement of covering a wide parameter space, they only sample a handful of waveforms within this space and have the potential to introduce bias to the search, which will be discussed in detail in section 3.2.2. It should be noted that the choice in training data is the key aspect to ensuring the GMM post-production has an un-modelled nature, hence the choice of generic simulated waveforms. If it was desired to

search for a given population of GW source, the signal training data could be altered to contain waveforms from this source.

The exact percentage splits of data sets can vary based upon the total amount of data available, with the requirement that at least a few hundred years of data is in the background test set in order to obtain reasonable FAR significance thresholds. We place selection thresholds on network coherent SNR, η_c, and network correlation coefficient, c_{c0}, for all data sets before they are used for analysis; selected triggers must have $\eta_c > 5.5$ and $c_{c0} > 0.5$. These cuts remove triggers which are likely glitches from the beginning, which aids in the computational efficiency of the analysis.

3.1.2 Choice of attributes

A set of statistical attributes must be chosen for the analysis, which will define our multi-dimensional space for noise and signal distributions. We want to consider attributes which have different distributions in the signal and noise parameter space, so that the populations can be well distinguished. In order for the analysis to have minimal assumptions on waveform type, we do not want to use attributes which contain specific information on the source of signals, such as frequency, duration or bandwidth, as this could make our models biased to given waveforms. Investigations into which cWB attributes best describe the signal and noise space were previously completed for the work by D.Lopez *et al.* [2], where the set of following attributes are considered: network coherent SNR η_c, network correlation coefficients $c_{c0,2}$, coherent energy across detectors E_c, network energy dis-balance N_{ED}, ratio of reconstructed and total energy N_{norm}, penalty on null energy per pixel, quality factor and energy distribution attributes $Q_{veto0,1}$, and frequency-based attributes describing narrow-band glitches $L_{veto0,1,2}$. The above selection contains attributes which characterise the coherence of triggers and veto attributes which describe qualities of common problematic glitches, allowing for the differences between the signal and noise parameter space to be fully modelled.

In order to optimise the fitting of the Gaussians within the models, we re-parameterise attributes to make their behaviour more Gaussian. For the case of the above cWB attributes, this mostly consists of logarithmic and inverse sigmoid re-parameterisations as selected in [2], which are detailed in Table 3.1. While making it easier for the models to converge, the re-parameterisation can also make the attribute distributions for signals and noise more distinguishable. Examples of attribute behaviours before and after re-parameterisation can be seen in Figures 3.1 and 3.2 respectively. In both plots, training data for background (BKG) and signal (generic WNBs later introduced in 3.2.2) populations are considered. Looking at Figure 3.1, the majority of distributions are clustered to attribute boundaries and are not distinguishable between background and signal. If we compare these distributions to those seen in Figure 3.2 after re-parameterisation has taken place, it is seen that all attribute distributions have better Gaussian behaviour and have clearer separation in the background and signal space. It must be noted

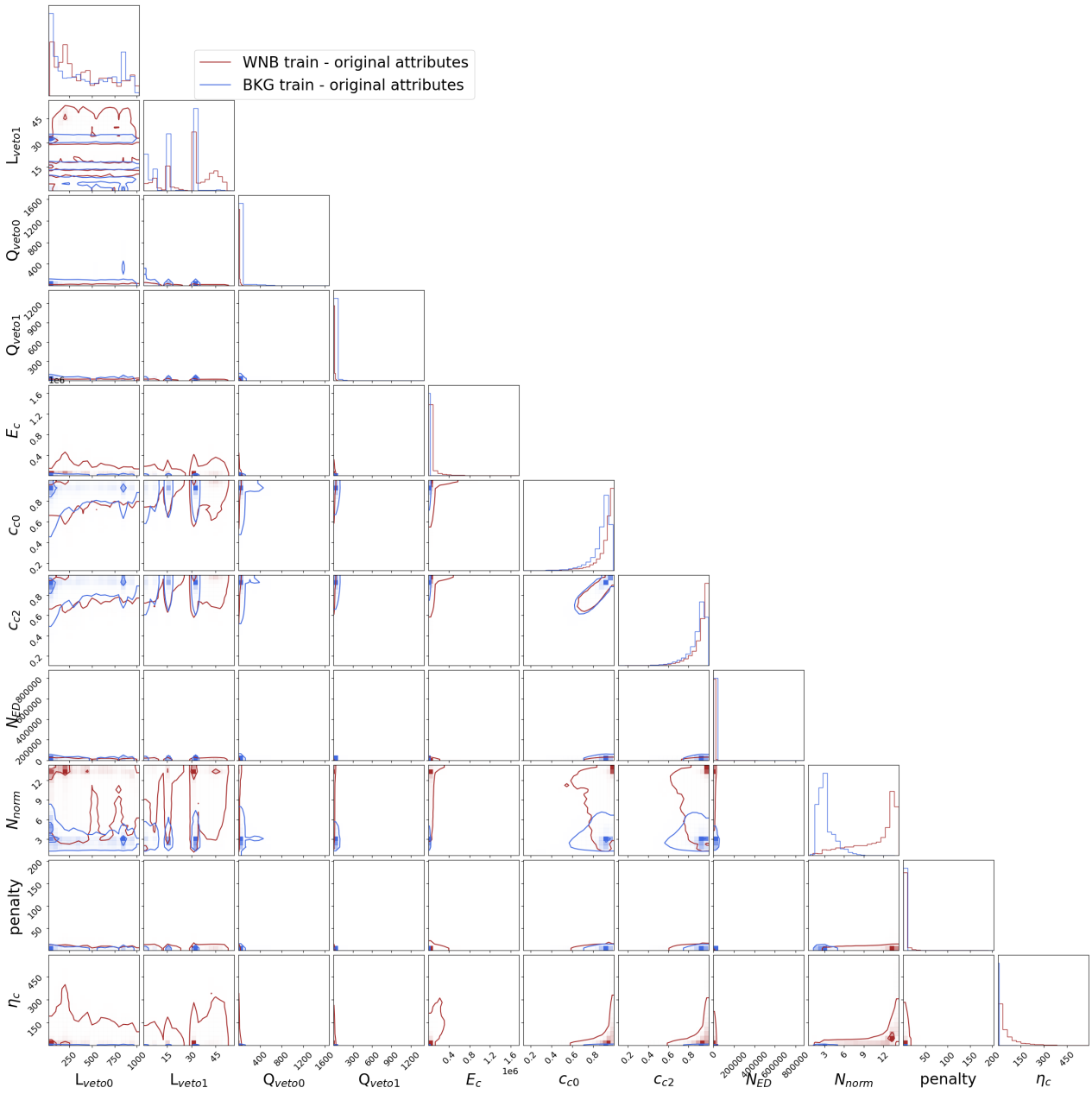


Figure 3.1: A corner plot showing the distributions of background training data (*BKG* - blue) and signal training data (*WNB simulations* - red) over original cWB attributes. The distributions of original attributes show non-Gaussian behaviour and for the majority of attributes are not distinguishable between background and signal populations.

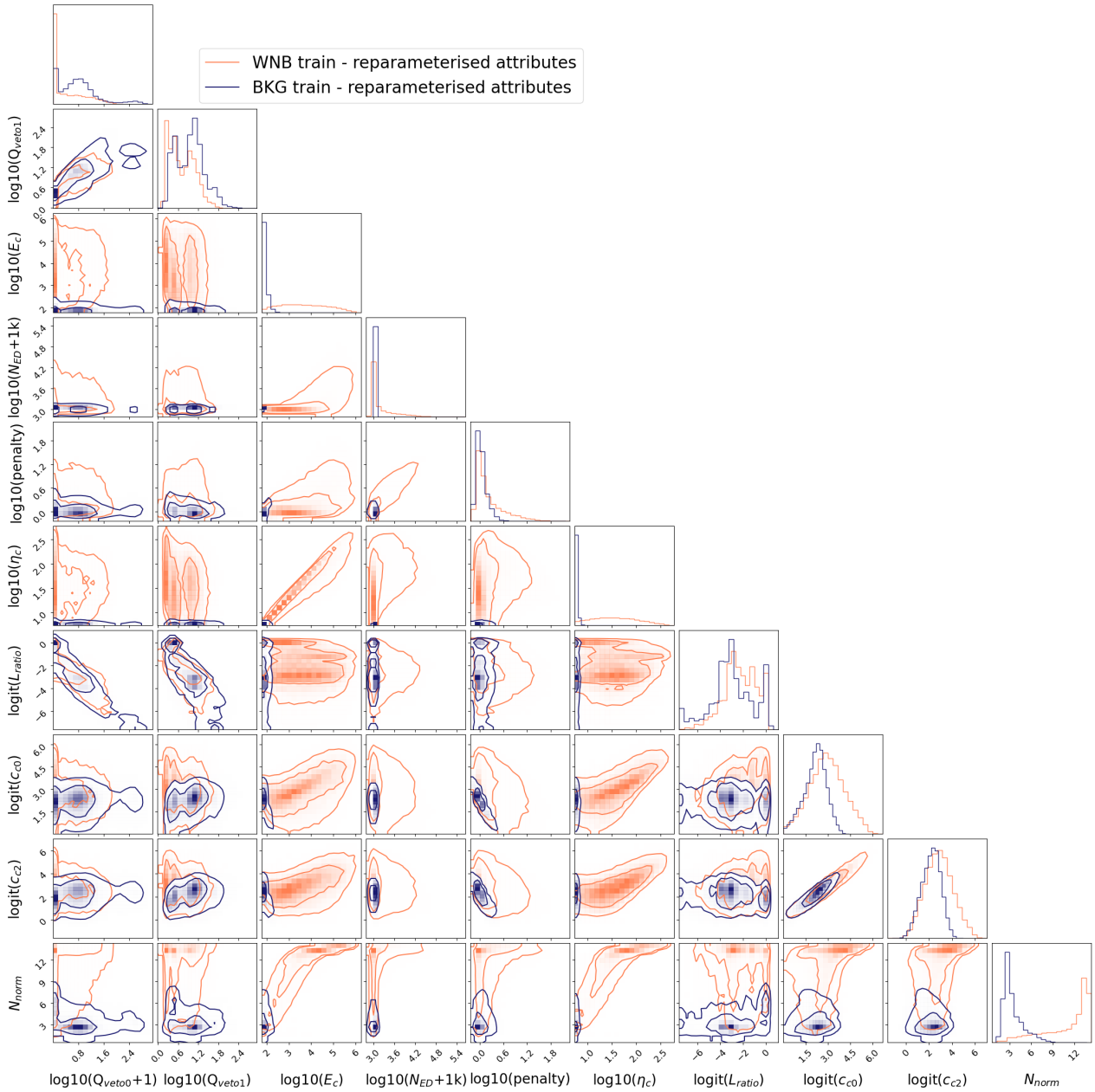


Figure 3.2: A corner plot showing the distributions of background training data (*BKG* - dark blue) and signal training data (*WNB simulations* - orange) over re-parameterised cWB attributes detailed in Table 3.1. After re-parameterisation, the attributes show better Gaussian behaviour and clearer separation between the background and signal populations.

cWB attribute	Re-parameterised attribute	Description
E_c	$log_{10}(E_c)$	Coherent energy across detectors
η_c	$log_{10}(\eta_c)$	Coherent SNR across the detector network
c_{c0}	$logit(c_{c0})$	Network correlation coefficient
c_{c2}	$logit(c_{c2})$	Sub-network correlation coefficient
N_{ED}	$log_{10}(N_{ED}+1000)$	Network energy dis-balance
penalty	$log_{10}(\text{penalty})$	Null energy per pixel
N_{norm}	N_{norm}	Ratio of reconstructed energy to total energy
Q_{veto0}	$log_{10}(Q_{\text{veto0}}+1)$	Energy distribution of an event over time
Q_{veto1}	$log_{10}(Q_{\text{veto1}})$	Quality factor
L_{veto0}	$logit(L_{\text{ratio}}) = logit(\frac{L_{\text{veto1}}}{L_{\text{veto0}}})$	Mean frequency of reconstructed signal
L_{veto1}		RMS frequency of reconstructed signal

Table 3.1: Reconstruction of Table 2.1 describing cWB attributes used in the GMM post-production, with details of the re-parameterisation applied to attributes in order to have better Gaussian behaviour.

here that we are only observing a 2-dimensional representation of attribute distributions, and that relationships between attributes in more than 2-dimensions may lead to further separation between the two populations. While it is clear that the selected logarithmic and inverse sigmoid re-parameterisations improve the Gaussian behaviour of the selected attributes, it is possible other functions may improve on this. Such work is not part of this thesis, however may be an interesting study for the future.

3.2 Investigating GMM methodology with the first half of the third observing run

In this section, we investigate possible alterations to the cWB+GMM methodology in an attempt to improve the sensitivity and robustness of the analysis. We explore such alterations on data from the first-half of the third LIGO-Virgo-KAGRA (LVK) advanced observing run (O3a), separated from the second-half by a month long commissioning break. O3a ran from April 1 2019 until October 1 2019, in which the LIGO-Livingston (L), LIGO-Hanford (H) and Virgo (V) detectors were active. For the optimisation studies here, we only consider the 2-detector LH network since the application of GMM is already known to work well on this network, which allows for direct comparisons with the previous methodology. Additionally, these detectors are nearly co-aligned which aids in the ability to disregard glitches during the cWB likelihood calculation, as discussed in section 2.2.2. This often results in the 2-detector LH network obtaining better sensitivities than other detector combinations, thus cWB burst searches are commonly completed with this network [85]. For the below studies, cWB production triggers are taken from the original O3a all-sky short analysis detailed in [85].

As part of these optimisation studies, we explore a new methodology to select the optimal number of Gaussian components, a new set of simulated waveforms as signal model training data, updates to the re-parameterisation and choice of attributes within the analysis, and finally investigate the options for initialisation within the model fitting algorithm. To observe any improvements such alterations may provide, we compare results to those seen by the original method in [2] through comparisons of ROC curves. As introduced in 2.2.5, these plots show the efficiency of the analysis over a range of FAR significance thresholds. We expect these plots to have higher detection efficiency for high values of FAR (low significance), however are particularly interested in optimising the detection efficiency for lower values of FAR (high significance, between FAR $1/10 - 1/100$ year^{-1}).

In order to maintain fair comparison, all analysis in this section uses the same test data sets as used in [2], which includes a 20% portion of injected ad-hoc waveforms. Due to the way this data was randomly split, the original number of injections into cWB associated with the 20% test set was not accurately calculated in the original methodology, leading to an overestimation of the calculated detection efficiency (see Section 3.2.2). Since this overestimation is due to a scaling factor in the number of original injections it does not effect the comparison of results, however the detection efficiencies displayed in ROC curves should not be compared to efficiencies stated in later sections. This problem did not occur after the change of signal training data in Section 3.2.2, thus is only relevant for analysis in this section.

After investigations are concluded and optimal approaches to the GMM methodology have been selected, we demonstrate the increased sensitivity and robustness of the enhanced GMM through direct comparisons with the 'old' methodology from [2] in section 3.3.

3.2.1 Selecting optimal number of Gaussian components

As detailed in section 2.3.2, the number of Gaussian components to be considered in a model must be input in order for Gaussian parameters to be estimated by the expectation maximisation (EM) algorithm. Previously, the number of Gaussian components for a set of data was optimised through the Bayesian information criterion (BIC), which is based upon minimising log-likelihood with the consideration of a penalty term for large number of Gaussian components compared to dimensions of data. The BIC is a popular method for estimating the number of Gaussians in a model, however only provides a measure of how well the Gaussians fit the training data which does not directly relate to our models' ability to distinguish between the background and signal populations. Investigations below found that it was not maximising our sensitivities.

Looking at the use of BIC in [2], it is found that the stated optimal number of Gaussian components is 104 Gaussians for the background model and 132 for the signal model. In order to test whether these are truly the optimal number of Gaussian components for our analysis, we train models with number of Gaussian components surrounding those found as optimal by

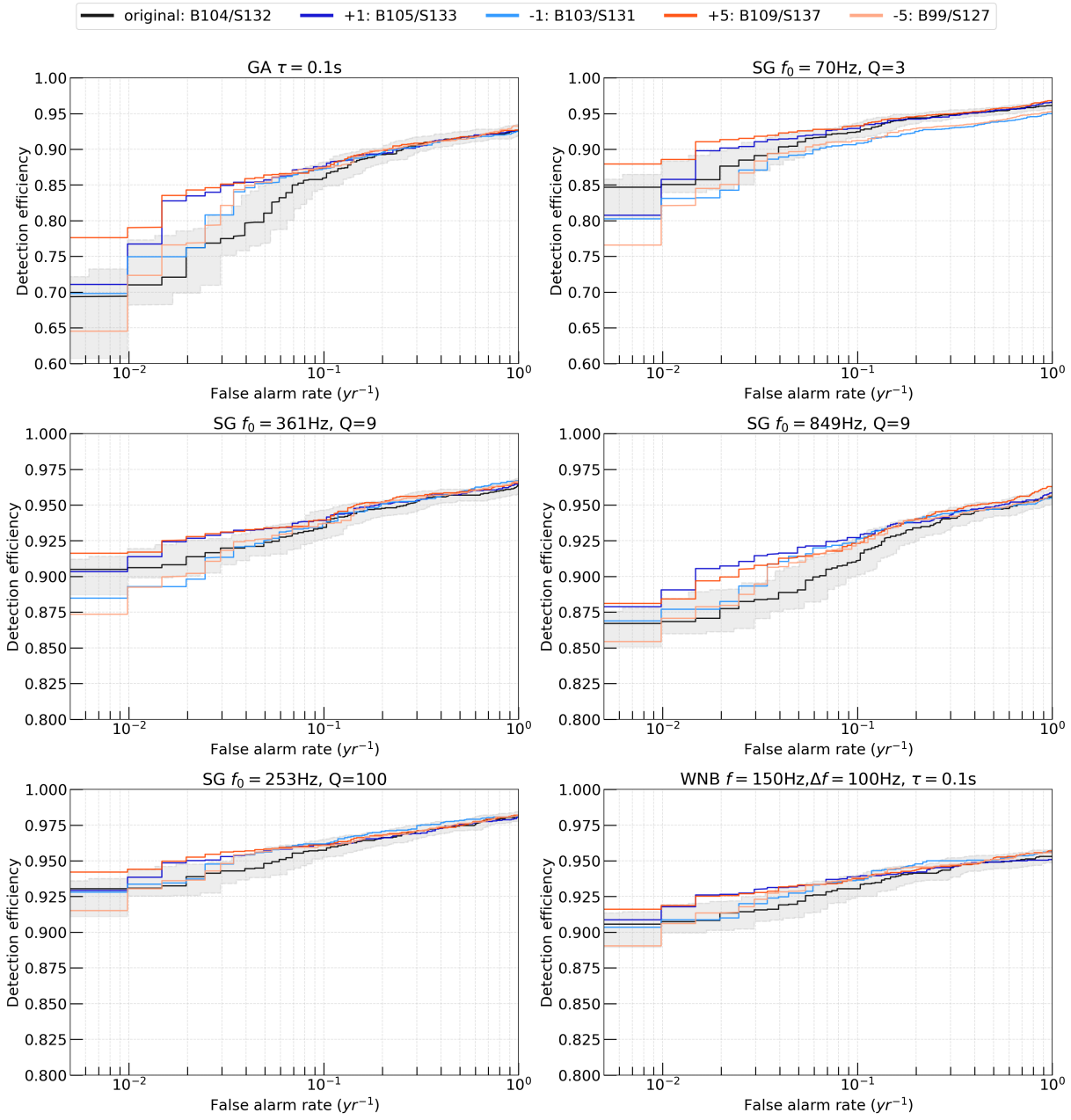


Figure 3.3: Detection efficiency vs FAR (ROC curves) for combinations of models with number of Gaussian components (*ncomp*) surrounding the optimally quoted value by BIC, over 6 injected waveforms of Gaussian pulse (GA), sine-Gaussian (SG) and white noise burst (WNB). 5 combinations of background and signal model are present; original optimal ncomp from BIC in black with associated error regions in grey, ncomp +1 (dark blue), ncomp -1 (light blue), ncomp +5 (dark red), ncomp -5 (light red), for each injected waveform. The ncomp value for background model follows 'B' in the legend, while ncomp value for signal model follows 'S'.

BIC, and compare efficiencies as a function of FAR for each combination. Specifically, we look at number of Gaussian components (*ncomp*) of $\pm 1, \pm 5$ of the quoted 'optimal', giving combinations of: background (BKG) 104 & signal (SIG) 132, BKG 105 & SIG 133, BKG 103 & SIG 131, BKG 109 & SIG 137 and BKG 99 & SIG 127. The analysis is done for a subset of injected ad-hoc waveforms, namely 1 Gaussian pulse (GA), 4 sine-Gaussian (SG) with different frequencies and quality factors and 1 white noise burst (WNB). The ROC curves are shown in Figure 3.3, with a plot for each injected waveform and coloured curves representing each combination of models with varying number of Gaussian components. From this plot, it is seen that the combination of ncomp as quoted optimal by BIC does not obtain optimal detection efficiency for any of the injected waveform types at FAR $<$ 1/10 years^{-1}. For many of the waveforms, there are a few combinations of ncomp which give significantly improved efficiencies compared to the original, while the ncomp +5 models have consistent improvement across all FARs for all injected waveforms. From these results, it is clear that the we can produce GMM models which achieve better fits than is provided through the BIC algorithm, hence we introduce a new approach to find optimal number of Gaussian components.

In this new approach, the number of Gaussian components in the models are selected by optimising detection efficiency at iFAR $\geq$ 50 years for validation data over some range of number of Gaussian components. In order to do this, we train our background and signal models over a wide range of number of Gaussian components, calculating the log-likelihood (eqn. 2.27) of background and signal validation data for each combination of ncomp in the background and signal models. This allows for FAR and T estimates to be given for the validation data, and the detection efficiency at iFAR $\geq$ 50 years to be calculated for each ncomp combination. The combination of background and signal ncomp which displays a peak in detection efficiency is selected for the remainder of the analysis. This process can be completed over a wide range of ncomp values, with training of models and calculation of log-likelihood for each combination being completed in parallel for computational efficiency. An example of the distribution of detection efficiency against combinations of number of Gaussian components is in Figure 3.4 for signal validation data consisting of Gaussian pulse (GA) injections. In this figure, number of Gaussian components in the background model are varied across the x-axis, while number of Gaussian components in the signal model are represented across shades of blue curves. It can be seen that optimal detection efficiency occurs for a combination of 114 background and 147 signal model Gaussian components.

3.2.2 New signal training data

In section 3.1.1 it was noted that the signal GMM model should be trained on a set of injected waveform simulations which cover the wide parameter space of the all-sky short search. Previously in [2], the signal model was trained on a portion of ad-hoc waveform injections from section 2.1.1, with other portions of the injections used for validation and testing sensitivities.

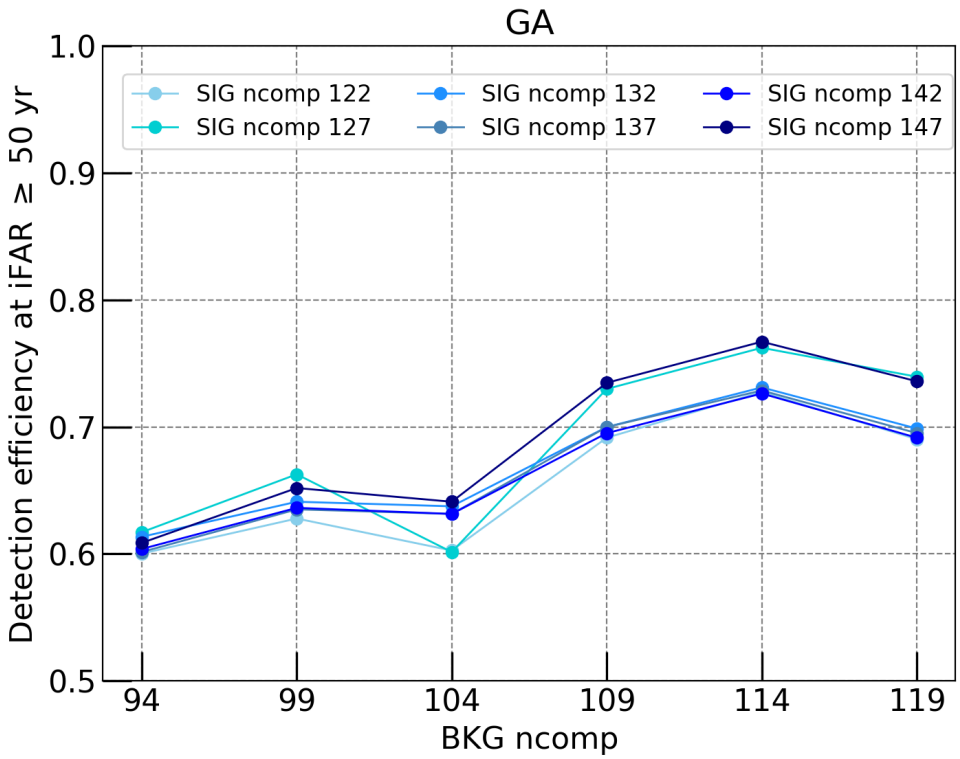


Figure 3.4: Example of new approach to optimise the number of Gaussian components (*ncomp*) in each model. Variation of detection efficiency at iFAR threshold of 50 years is shown over a range of combinations of background and signal ncomp for some signal validation data consisting of Gaussian pulse (GA) injections. Variation over background (BKG) ncomp is displayed across the x-axis, while variation over signal (SIG) ncomp is displayed over varying shades of blue.

While this covers the overall range of the parameter space, the representation of frequency and duration are sparsely sampled, with only 20 different waveforms considered. Due to this, this choice in training data may introduce a strong bias when testing the sensitivity of the pipeline, since we don't have a wide array of samples across the parameter space and the same type of waveforms are considered for training and testing. In addition to this, through investigations into the splitting of simulation data it was found that the previous work in [2] over-estimated sensitivities due to the approach of finding the total number of original injections associated to the 20% test data. This approach involved matching properties of triggers in the 20% test data to triggers in the original injection file, however this failed to account for any original injections which were not recovered as triggers by cWB production, thus leading to an over-estimation in detection efficiency. While this problem can be solved by calculating the cWB efficiency and GMM efficiency as 2 separate quantities, the choice of a separate signal training data set would reserve the full ad-hoc simulation set for testing the sensitivity of the pipeline. This would avoid any possibility of the sensitivities being over-estimated and reduces errors due to increased sample size.

Instead we introduce a generic set of simulated band-limited white noise burst (WNB) wave-

forms (white noise contained in an ad-hoc time-frequency range) to be injected into cWB, as is seen in the XGBoost post-production [124]. These simulated WNB waveforms densely cover the entire signal parameter space while the waveform morphologies are generic in order to keep the algorithm robust to all signal types in the short-duration burst sensitive range. We consider 2 distributions: First, WNBs which are uniformly distributed in central frequency [24,996]Hz and bandwidth [10,300]Hz, and logarithmically distributed in duration [0.1,500]ms, referred to as set 1 (*WNB1*). Secondly, WNBs with bandwidth 10Hz, frequency 0Hz, and duration τ randomly distributed over [0.1,1]ms, referred to as set 2 (*WNB2*). This second data set specifically targets the parameter space where shorter, low-Q waveforms lie such as GAs, since the 0Hz frequency of the WNB waveform results in a Gaussian envelope. In order to ensure the selected training data has desirable performance, we compare sensitivities to ad-hoc waveforms with a signal model trained on only WNB1 data, against a signal model trained on combined WNB1+WNB2 data. ROC curves for this are seen in Figure 3.5 over grouped types of ad-hoc waveforms Gaussian pulse (GA), white noise burst (WNB) and sine-Gaussian (SG) with $Q = 3, 9, 100$, along with curves which test differing attribute sets introduced in the section below 3.2.3. For now only looking at the WNB1 (red) and WNB1+WNB2 (navy) curves, it is seen that the addition of WNB2 in training significantly increases detection efficiency for low quality factor waveforms, especially at iFAR of 10 years. From looking at higher Q waveforms, the addition of WNB2 doesn't drastically alter sensitivities. This suggests the use of both WNB data sets as training better represents the full signal space, however further investigations involving re-parameterisations are conducted below prior to training data being selected.

3.2.3 New attribute considerations

Next, we investigate the performance of the analysis with minor alterations to the considered cWB attributes. In particular, the removal of the L_{veto2} attribute is investigated due it's distribution in the signal and noise space, and an η_c-weighted detection statistic is introduced in an attempt to improve the classification of high-SNR signals.

Removal of L_{veto2}

The $L_{veto0,1,2}$ attributes were introduced into cWB to account for narrow-band glitches occurring near power line frequencies, and describe frequency and energy properties of triggers as described in Table 2.1. Specifically, the L_{veto2} attribute evaluates the ratio of maximal pixel energy to total energy of the trigger. Despite these glitches being problematic during the second LVK observing run (O2), they were not found to be problematic in O3. Because of these glitches not being prominent, there was not a clear separation of the L_{veto2} attribute between noise and signal population, even after re-parameterisation, and it was found a lot of triggers were railing against the boundary of the distribution. It was thought that this could be causing

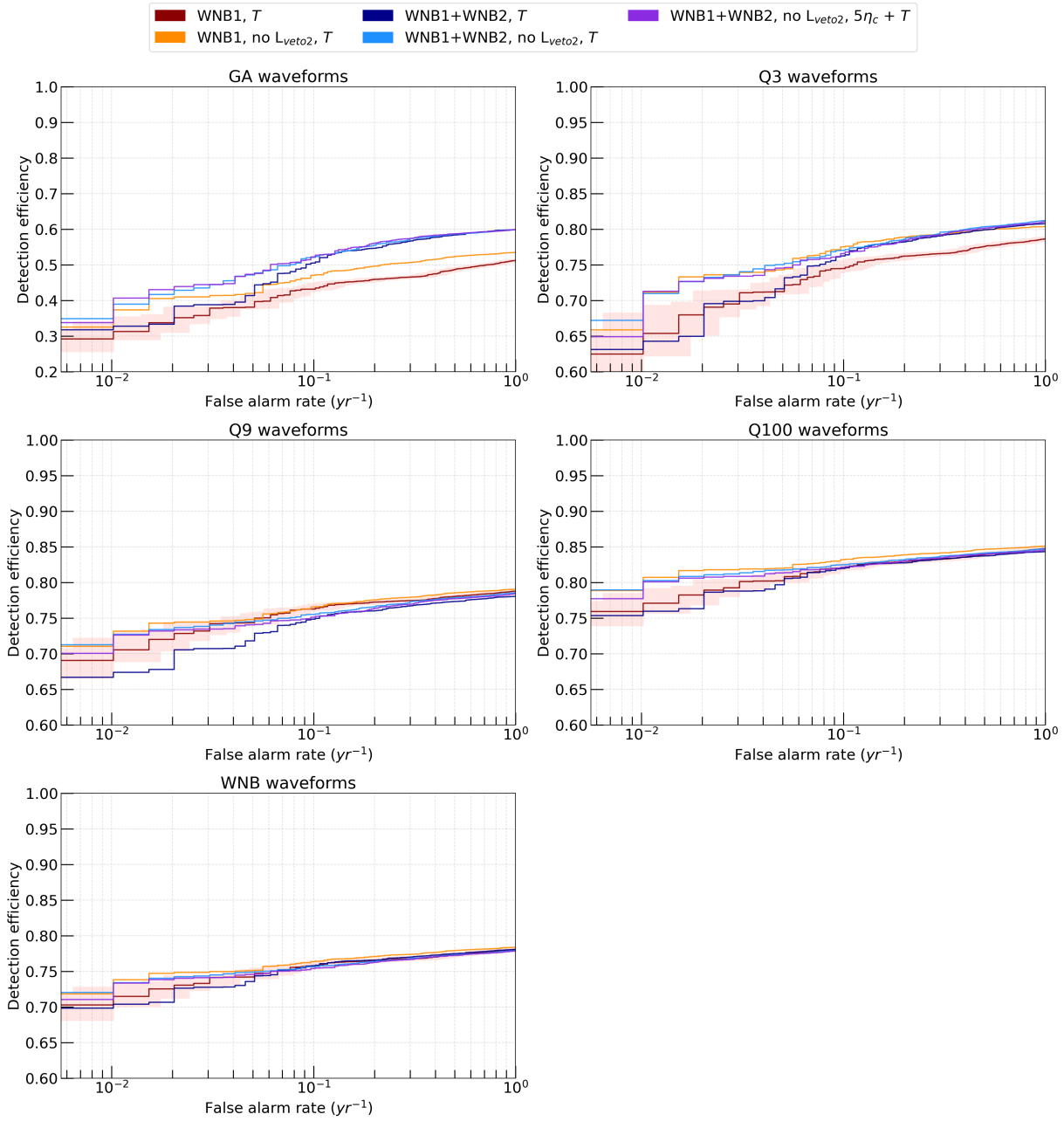


Figure 3.5: Detection efficiency vs. FAR for different variations of signal training data, attribute considerations and detection statistics, over grouped waveform types of Gaussian pulse (GA), white noise burst (WNB) and sine-Gaussian (SG) with $Q = 3, 9, 100$. Signal training set WNB1 is shown with and without L_{veto2} for detection statistic T in red and orange respectively. Signal training set WNB1+WNB2 is shown with and without L_{veto2} for detection statistic T in navy and blue respectively. Signal training set WNB1+WNB2, without L_{veto2}, but with altered detection statistic $T + 5\eta_c$ is shown in purple. Error regions corresponding to the WNB1, T, results are represented by shaded red regions.

some confusion to the models, and so we investigate removing this attribute from the analysis. GMM analysis was completed with the L_{veto2} attribute completely removed - ie. from model training, validation and test data sets. Figure 3.5 shows detection efficiency over FAR with and without L_{veto2} for signal training sets WNB1 and WNB1+WNB2 over the following grouped types of ad-hoc waveforms: Gaussian pulse (GA), white noise burst (WNB) and sine-Gaussian (SG) with $Q = 3, 9, 100$. By comparing these curves, it is seen that removing L_{veto2} from the analysis improves detection efficiency for both variations of WNB simulation training sets for all considered waveform types. Similarly to the remarks made in the previous section, the inclusion of WNB2 in signal training still brings significant sensitivity improvement for the GA waveforms without L_{veto2}.

The $L_{ratio} = L_{veto1}/L_{veto0}$ attribute is a measure of root-mean-square frequency over mean frequency and holds more distinguishing power for noise and signal populations than L_{veto2}. In theory the reliability on frequency has the potential to introduce some small bias since it will vary over different signal types, however for the new WNB signal training data, the L_{ratio} attribute spans values across the entire parameter space, hence was not thought to be problematic. For this reason, the L_{ratio} attribute remains in the analysis.

Altered detection statistic

While looking at the distribution of T vs. coherent SNR (η_c) for background and signal populations produced after completed GMM analysis, it was noted that some signal triggers with high η_c were being classified with low T values, as can be seen in Figure 3.6. While the distribution of T generally increases with η_c as desired, there are simulation triggers surrounding the iFAR 10 and 100 year thresholds which have high η_c but relatively low T, ultimately causing a fraction of high SNR injections to be classified as likely glitches. In an attempt to shift triggers with high network SNR to higher significance, we completed the analysis with a new weighted detection statistic involving both T and η_c: $T + 5\eta_c$. While this will shift high SNR triggers in both simulation and background to higher T, it should have a more significant effect on the simulation triggers since these often have much larger SNR. Drawing conclusions from the above studies on signal training data and L_{veto2}, we complete this analysis with the WNB1+WNB2 training set without consideration of L_{veto2}. The detection efficiencies over FAR can once again be seen in Figure 3.5 across the grouped ad-hoc waveforms. Comparing the results to analysis with WNB1+WNB2, no L_{veto2} and T detection statistic, it is seen that the consideration of an extra η_c term does not bring significant improvement to the analysis. Hence we proceed with only T as the detection statistic.

While the above study did not bring improvement to the analysis, it is possible that a similar approach with different weighting may have a more significant effect. For example, a statistic with a factor applied to the T statistic instead of η_c, such as $5T + \eta_c$, may have a better classification of the high SNR signal events. Further studies were not carried out within this thesis, but

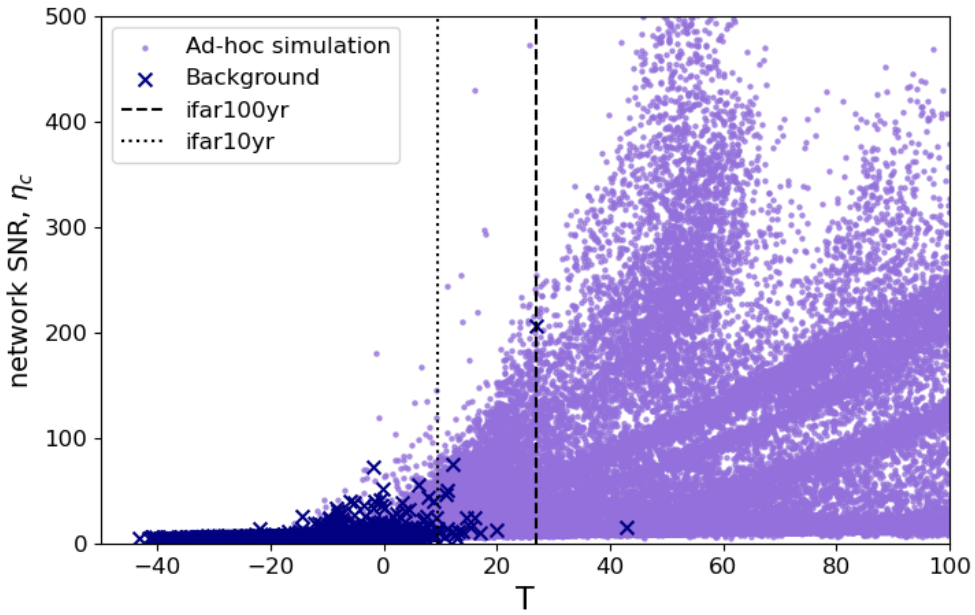


Figure 3.6: Distribution of T vs. network SNR, η_c, for GMM-processed background (blue) and simulation (purple) triggers from O3a. T associated with iFAR thresholds of 10 years and 100 years is notated by dotted and dashed lines respectively.

would be beneficial as future analysis.

Final attributes and training data

With the above studies on signal training data, attribute consideration and altered detection statistic complete, we here detail the final signal training and attribute considerations for the updated GMM methodology. From Figure 3.5, the form of analysis which most consistently gives high sensitivities across all waveform types is training set WNB1+WNB2, without L_{veto2}, and with detection statistic T. Hence, the finalised set of re-parameterised attributes used for the GMM analysis are: $\log_{10}(E_c)$, $\log_{10}(\eta_c)$, $\text{logit}(c_{c0})$, $\text{logit}(c_{c2})$, $\log_{10}(N_{ED}+1000)$, N_{norm}, $\log_{10}(\text{penalty})$, $\log_{10}(Q_{\text{veto0}}+1)$, $\log_{10}(Q_{\text{veto1}})$, $\text{logit}(L_{\text{ratio}}) = \text{logit}(\frac{L_{\text{veto1}}}{L_{\text{veto0}}})$.

An example of how the background training data differs compared to the WNB training data for the chosen re-parameterised attribute space can be seen by referring back to Figure 3.2. Here it is demonstrated that, at least in the 2-dimensional attribute space, our choice of multi-dimensional data for training our noise and signal models is distinguishable.

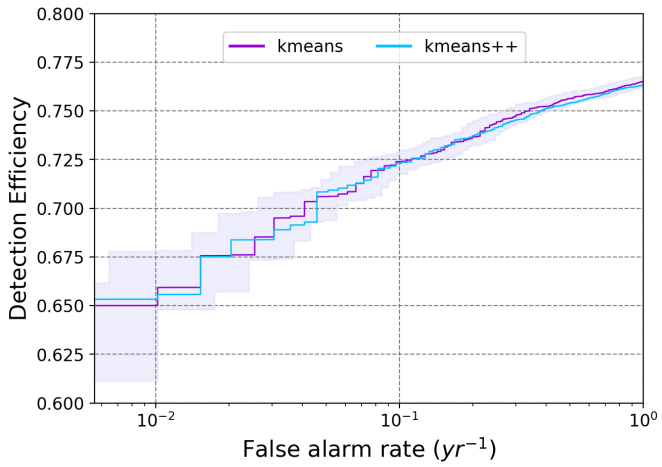
3.2.4 Model optimisation

When fitting Gaussian mixture models to specific sets of data, there may be function parameters which we want to vary to better tackle the problem being faced. For example, data sets with complex structure may require a lower convergence threshold in the EM algorithm than

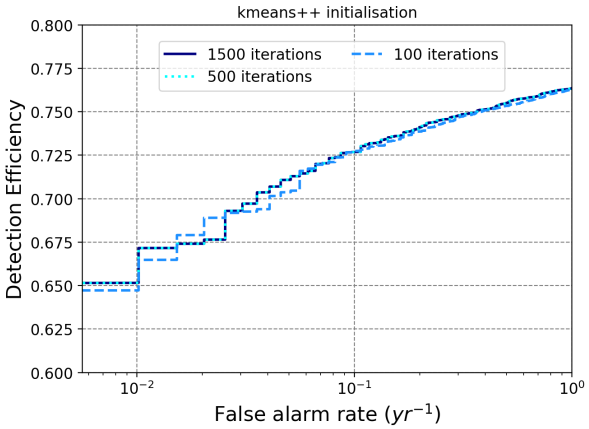
low-dimensional data with good Gaussian behaviour, to avoid convergence on local maxima. There are many parameters that can be tweaked, and the optimal choice will vary for different problems. For the models concerned here, we investigate parameters which affect initialisation and convergence of the EM algorithm to ensure our model is fitting data optimally. Specifically, this includes the method to initialise the placement of Gaussians, the number of initialisations to perform, convergence limit of the EM algorithm and the limit on iterations performed. To visualise the performance of both the signal and background model when exploring these function parameters, we again compare variations in analysis through ROC curves, this time with detection efficiency vs. FAR calculated over all ad-hoc waveform types.

While we want to ensure optimally fit models, we must also consider the effect of computational time within these processes so that analysis can be completed within a reasonable time frame, and need to find a good balance between sensitivity improvement and minimal training time. The consideration of computational runtime is a factor throughout this thesis, and will be mentioned in relation to possible alterations to the analysis. Prior to model optimisation alterations being considered, the GMM post-production analysis applied to O3a will complete within 2 days, with model training consuming the majority of computational time at $\sim$ 24 hours.

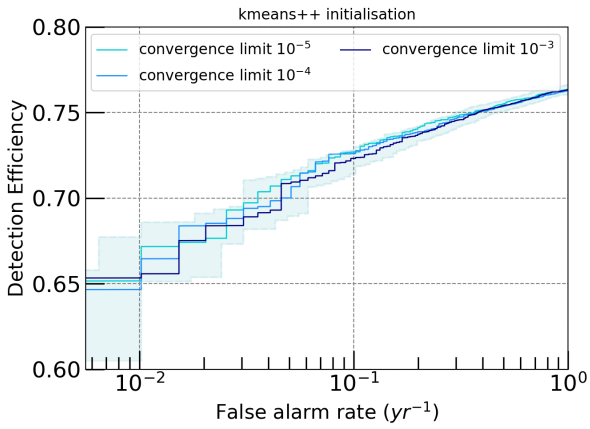
First we investigate the use of a new initialisation method, which first places the Gaussian centres before the optimisation of parameters (Gaussian properties) can begin with the EM algorithm. In section 2.3.2 we introduced the k-means method of initialisation, which was the default method used by the previous GMM methodology in [2]. During k-means initialisation, the centres of K Gaussian clusters are originally randomly selected, and surrounding data points are assigned to their closest cluster by euclidean distance. Centroid locations are then updated based on the mean of all data points in the given cluster. This can be repeated until convergence. While this method is more trustworthy than selecting the centroids completely randomly, it is still heavily reliant on the initial selected cluster locations. Instead, we test the initialisation using the k-means++ methodology, which largely follows the same approach as k-means however places the initial locations of the centroids in a more informed way: the centre of the first cluster is randomly selected, and for each point not in the cluster the distance to the nearest cluster is calculated. The next cluster centre is then placed based upon probability proportional to the distance squared to the next nearest centroid, ensuring cluster centres are well spread out across the data. Since the initial placements of Gaussians is more informed, it's been suggested that the k-means++ method is more robust to local maxima convergence. The effect of both initialisation methods is seen in Figure 3.7a, where we compare ROC curves for models trained with either the k-means or k-means++ initialisation over all ad-hoc waveforms. Overall, it seems that the change in initialisation method does not have a large effect on the final sensitivities of the analysis in this case, with results using the k-means++ initialisation within error regions of k-means initialisation. Since theory suggests that k-means++ should be more robust in the initial placements of Gaussians while bringing no significant extra computational cost, we select this



(a) Initialisation methods



(b) EM maximum iterations



(c) EM convergence limit

Figure 3.7: ROC curves comparing the performance of the analysis for variations of initialisation and EM algorithm parameters for model fitting. Figure (a) compares initialisation methods of k-means (purple) and k-means++ (blue), with k-means error regions shown. Figure (b) tests reliable maximum iteration thresholds for a convergence limit of 10^{-5}, while figure (c) compares performance of models with varied convergence limit. All plots consider all ad-hoc waveform types.

initialisation method for future training in an attempt to lessen the effect of models suffering from occasional badly placed clusters.

Having decided on the use of the k-means++ initialisation method, we now explore related parameters which may improve the fitting of our models and ensure they are robust. First we explore the convergence limit of the EM algorithm, which defines the value of gradient which must be reached to declare maximised likelihood has been obtained, thus outputting the optimal Gaussian parameters. Related to this is the defined maximum number of iterations the EM algorithm will perform, even if convergence is not reached. In general, if we have complex training data we want a low convergence limit to reduce the probability of finding local maxima in the likelihood function, however not so low that the algorithm takes a long time to converge.

Similarly with the maximum number of iterations, we want enough iterations to allow the EM algorithm to explore the parameter space without letting the optimisation run for too long if it cannot converge. The maximum number of iterations was varied over values of 100-1500 for different convergence limits. Figure 3.7b shows the results of this analysis for such variations, for a convergence limit of 10^{-5}. It is seen that above 500 iterations, the ROC curves completely agree with each other, meaning there have been enough iterations to allow the EM algorithm to converge. Below this value, the ROC curve differs, meaning the algorithm has reached maximum allowed iterations before converging. Hence we set the maximum number of iterations to 1000 to ensure the algorithm always has enough time to converge. Next we train models over varying convergence limit: 10^{-3} (the default), 10^{-4} and 10^{-5}, seen in Figure 3.7c. Here, error regions of the ROC curve for 10^{-3} are shown to more clearly demonstrate any benefit decreasing the convergence limit may have. As can be seen, while there is a slight increase in detection efficiency for 10^{-4} and 10^{-5} at FAR of 1/10 years, the curves are still within the error regions of convergence limit 10^{-3}. Hence, since it seems that a convergence limit of 10^{-3} is as efficient in reaching global maxima as lower values, we proceed with this value for model training.

The total number of initialisations performed in the training of a model was also explored. By default, this parameter is one meaning the initialisation of Gaussians and the optimisation of their parameters is performed once. However, since the initialisation methods can be dependent on the initial random centroid placement, it may be of interest to test with many initial placements in order to achieve a better model. By increasing the total number of initialisations, the initial parameters of the models are chosen multiple times, with the 'best' model selected for the EM algorithm based upon the highest lower bound value on the likelihood. While this option was desirable, we found that even performing 3 initialisations increased the computational time by an unreasonable amount. Instead, it was thought to investigate the effect of training multiple models over a variation of 'random states', which has a similar effect of being able to test different initialisations however allows for each initialisation to be run in parallel with the implementation of HTCONDOR in order to minimise computational time. Full details of these tests and why they were performed are below.

The effect of random state

Often in physics and computing, there is a need for randomness within processes and random number generators are used for a wide array of tasks, whether this is to sample from large distributions or to simulate the expected behaviour of some physical system. However since the nature of computers are deterministic, no process can be *truly* random. Instead, random number generators often use mathematical algorithms or pre-calculated tables to simulate the 'randomness' of processes and sequences. This is where random state comes into play: a random state is a parameter within a function which defines the 'random' process that is being carried out, giving the computational algorithms a starting place for the random sequence. Overall

this is useful, as it allows for random processes to be completely reproducible, which is an important aspect of data analysis. When it comes to Gaussian mixture models, a random state can be chosen when fitting the Gaussians to the training data, which will affect the placement of Gaussian centroids during the first steps of initialisation. Despite the initialisation of the Gaussians being optimised by the k-means++ method, it can still be susceptible to the first Gaussian centroid location selection since it is randomly placed. By default, the random state is set to 0, however by selecting a different random state the initial centroid will be placed differently. Hence, we can investigate the effect the random state has on our final results, and whether it could be a means to avoid 'bad' initial Gaussian placements, which ultimately lead to sub-optimal models.

We use the selected training settings from above, namely k-means++ initialisation, convergence limit of 10^{-3} and 1000 maximum iterations, to train models over a set of random state values, including random state equal to 0 since this replicates the default analysis. We assume that the number of Gaussian components for the models have already been selected, and so keep this the same for all random state models. An example of such a test with 5 variations of random state values is seen in Figure 3.8, displaying detection efficiency vs FAR for each group of ad-hoc waveform type over the models trained with different random state. Error regions are displayed for models with random state 0 so any significant deviation from this analysis can be observed. As can be seen, while different random states cause variation, the majority of models produce results within the error region of random state 0, especially for SG and WNB waveforms. However, the model trained with random state of 100 appears to be an outlier, performing significantly worse than the other models at low FARs for all waveforms. We also see that models with random state 1 and 44 lie above the error regions for GA around FAR 1/10 years. While we want to ensure we do not by chance select a low performing outlier state which has sub-optimal models, we also don't want to optimise over random state, as this could create a bias by over-fitting to the selected data. With this is mind, a new step in the analysis is introduced in which we train models on a variety of random states and observe their performance on validation data, ensuring that the originally chosen random state agrees with the majority of random states, ie. that we do not accidentally choose an outlier. Since this is a sanity check rather than an optimisation step, we perform this after the optimal number of Gaussian components have been chosen. Overall, the fact that the majority of results with differing random state lie within the error region of the original analysis suggest that the choices of GMM training parameters create robust models.

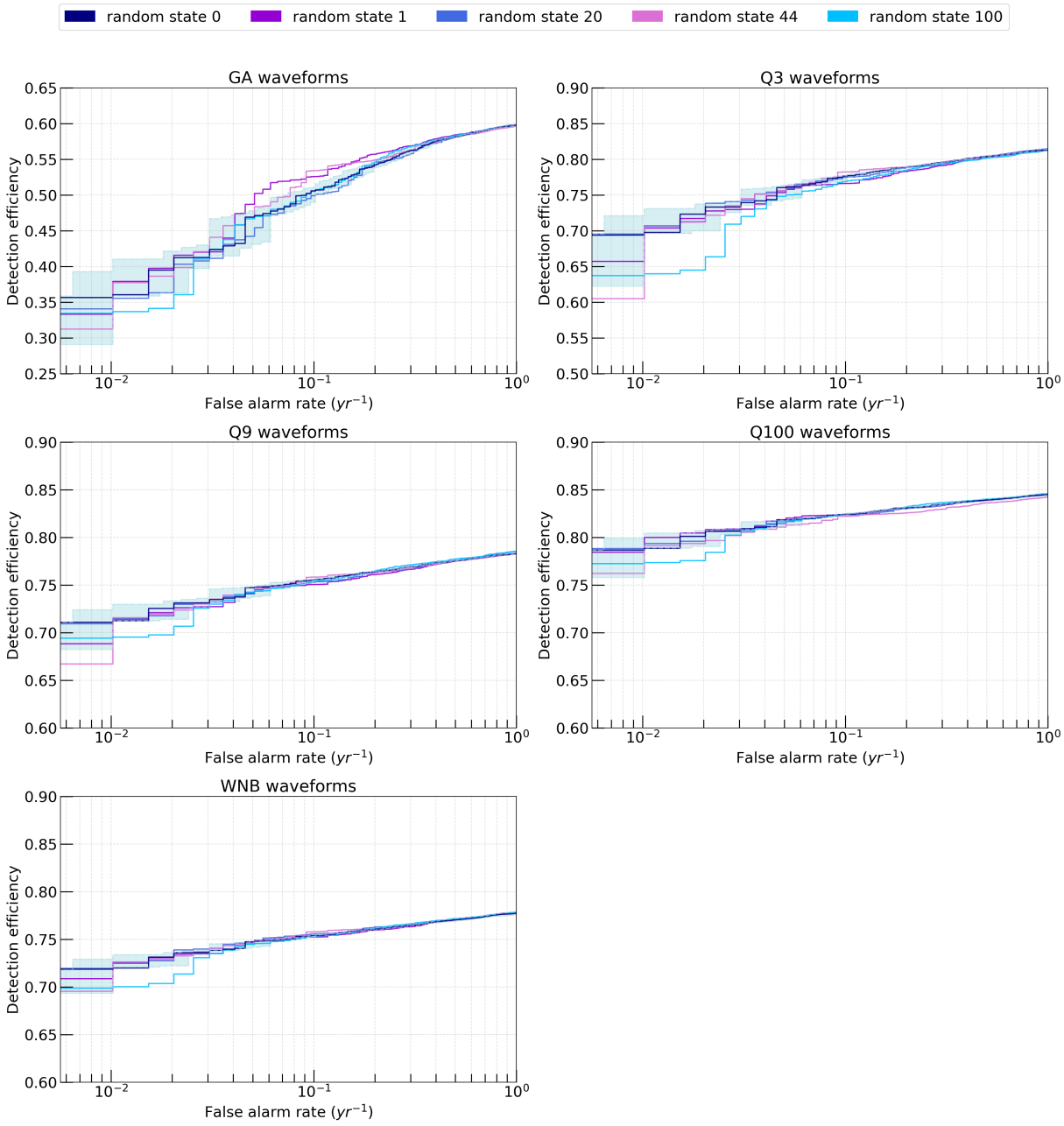


Figure 3.8: Detection efficiency vs. false alarm rate over randomly selected values of random state during training, shown separately for groups of ad-hoc waveform type. The ROC curves are shown for models trained with random state 0 (original analysis), 1, 20, 44, 100, with shaded error regions shown for random state 0. Note that subplots have different y-axis limits so that variation in ROC curves can be observed better.

Feature	previous GMM method	*GMM+* method
Signal model training data	subset of ad-hoc injections	generic WNB injections
Optimising no. Gaussian components	Bayesian Information Criterion	optimisation of detection efficiency
Attribute set	with *Lveto2*	without *Lveto2*
GMM Initialisation method	k-means	k-means++
Random state	0	sanity check on outlier initialisations

Table 3.2: Summary of feature changes in GMM methodology between previous analysis from [2] and the enhanced analysis concluded from investigations above (*GMM+*).

3.3 Sensitivity comparisons to the previous GMM methodology

The above investigations into the performance of cWB+GMM methodology were completed in an attempt to improve the creation of the models and to enhance the sensitivity of the analysis across the short duration transient parameter space. We update the GMM methodology based upon these investigations to create an enhanced version of the analysis, referred to as *GMM+* for the remainder of this section. Namely, the major updates made to the methodology in GMM+ are:

- New signal training data set consisting of generic WNB simulations that are well-sampled across the entire short duration, low-frequency parameter space.
- New method to optimise the number of Gaussian components in each model, based upon the optimisation of detection efficiency at an iFAR threshold of 50 years over combinations of models with differing number of Gaussian components with validation data.
- The removal of the L_{veto2} attribute from the analysis, since it was causing confusion within models.
- New initialisation method for Gaussian placement during training, now completed with k-means++.
- A new step in the analysis to check for outlier initialisations leading to badly fit models, completed by comparison over random states during the model training.

We test the enhanced sensitivity and robustness of the GMM+ method through sensitivity comparisons with the previous GMM methodology detailed in [2] on O3a data. A summary of the differences between the two methodologies can be seen in Table 3.2. For this comparison, we utilise population injections made into cWB, consisting of a set of generic ad-hoc waveforms commonly used to benchmark all-sky short pipeline sensitivities, and astrophysically motivated core-collapse supernovae (CCSN) waveforms. The set of generic ad-hoc waveforms used in the O3 burst all-sky short search [85] consists of Gaussian pulse (GA), sine-Gaussian (SG) and

Gaussian Pulse (GA)	Sine-Gaussian (SG)		White Noise Burst (WNB)		
τ (ms)	f_0 (Hz)	Q	f_{low} (Hz)	Δf (Hz)	τ (s)
0.1	70	3,9,100	150	100	0.1
1	100	9	300	100	0.1
2.5	153	9	700	100	0.1
4	253	3,9,100			
	361	9			
	554	9			
	849	3,9,100			

Table 3.3: Table of injected ad-hoc simulations with defining parameters for the O3 observing run. Totals to 4 GA waveforms varied over duration, 13 SG waveforms varied over central frequency and Quality factor and 3 WNB waveforms varied over lower frequency bound with set bandwidth and duration.

white noise burst (WNB) simulations injected over a variety of frequencies, bandwidths and duration in order to cover a significant portion of the search signal parameter space. The exact parameters considered for each group of waveforms are detailed in Table 3.3. These simulations are injected at sky locations drawn from a uniform distribution in solid angle over the entire sky. The GA waveforms are linearly polarised and the SG waveforms are circularly polarised so that the source is assumed to be optimally oriented [1]. The WNB waveforms are unpolarised, meaning they have equal power in both polarisations.

The injected h_{rss} of these simulations is chosen from a grid of maximum strain values given by $(\sqrt{3})^N \times 5 \times 10^{-23}$ with N ranging from 0 to 8. We take a set of CCSN waveform models which cover a variety of explosion mechanisms, progenitor star masses and progenitor rotation. This set consists of 10 neutrino explosion models: Andersen et al. 2017 [169] (And s11), Müller et al. 2012 [170] (Mul L15), Kuroda et al. 2016 [171] (Kur SFHx), O'Connor & Couch 2018 [132] (Oco mesa20), Powell & Müller 2019 [131] (Pow he3.5, s18), Radice et al. 2019 [133] (Rad s9, s13, s25), and 1 magnetorotationally driven explosion model: Abdikamalov et al. 2014 [172] (Abd A4O01.0). An overview of each of these models can be seen in Appendix A.

The previous GMM methodology used a portion of the ad-hoc waveform injections detailed in Table 3.3 for training (70%), validation (10%) and testing (20%) of the signal population. For the GMM+ methodology, we apply a 20% validation and 80% training split to the generic WNB injections detailed in Section 3.2.2. In order to keep the comparison of sensitivities accurate, we test the updated GMM+ methodology on the same 20% of ad-hoc signals used for testing in the previous methodology, despite the entire data set being available in GMM+. Both methodologies also apply a 70% training, 10% validation and 20% testing split to background triggers, reserving 196.14 years of background for FAR estimation.

[1]Note that [2] states that some of the SG waveforms considered are elliptically polarised, however this is an error in the naming convention of the waveforms. The waveforms described here are the same as utilised in [2], and are in fact circularly polarised.

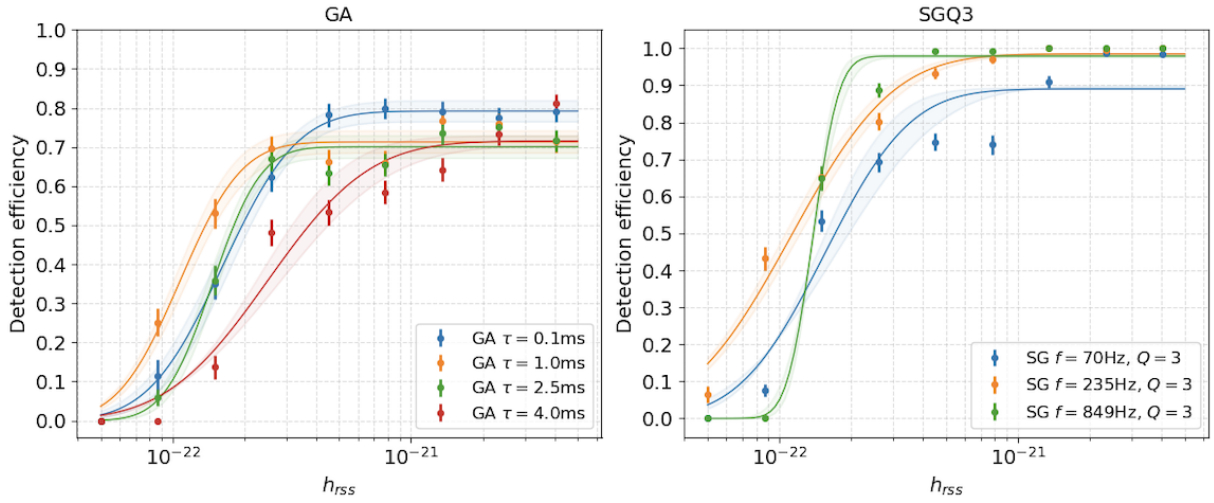
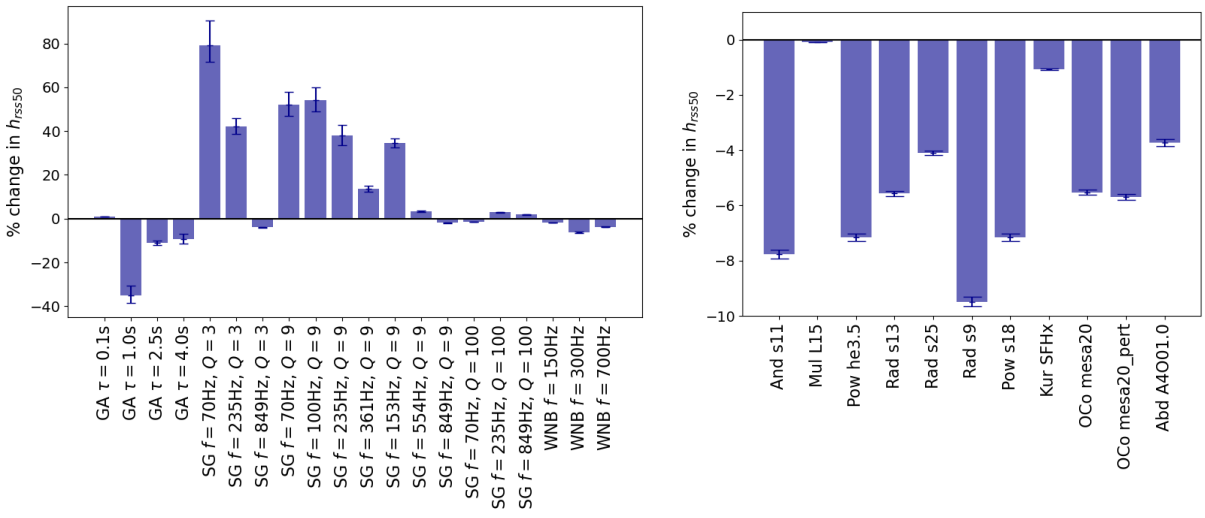


Figure 3.9: Example of detection efficiency as a function of h_{rss} for grouped GA and SG $Q = 3$ waveform injections for GMM+ in O3a. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve. Note that the detection efficiency is over-estimated here due to the definition of total number of initial injections used by previous GMM methodology [2], hence should not be compared to results in later Chapters.

To quantify the sensitivity improvement of GMM+, we compare the h_{rss50} statistic per injected waveform with the previous GMM methodology. Introduced in section 2.2.5, this is found by plotting the detection efficiency as a function of injected h_{rss}, fitting the data with a cumulative lognormal distribution and inferring the h_{rss} amplitude at which 50% of waveform injections are detected at given FAR threshold. An example of such curves for GA and SG $Q = 3$ waveforms is seen in Figure 3.9 for the GMM+ methodology, however the efficiency curves for all ad-hoc and CCSN waveforms with GMM+ methodology can be seen in Appendix B. The error regions shown are calculated by fitting a lognormal distribution to the upper and lower error on detection efficiency, computed as seen in eqn. 2.16. The error regions are large due to the consideration of only 20% test data, as the total number of considered initial injections directly affects the errors on detection efficiency. It must also be noted that the figure displays over-estimated values of h_{rss50} due to the incorrect definition of total number of initial injections associated with the 20% test data, as discussed in section 3.2.2. While a fix for this issue was discovered through investigations, we consider the old method here for fair comparison with the results in [2]. We define the percentage change in h_{rss50} at iFAR $\geq$ 100 years over the chosen waveform injections for GMM+ relative to the previous methodology:

$$\% \text{ change in } h_{rss50} = \frac{h_{rss50\text{plus}} - h_{rss50\text{old}}}{h_{rss50\text{old}}} \times 100 \tag{3.1}$$

where $h_{rss50\text{plus}}$ refers to GMM+ analysis and $h_{rss50\text{old}}$ refers to the previous analysis in [2]. With this terminology, a negative change represents a reduction in h_{rss50} and hence an improvement



(a) ad-hoc waveforms

(b) CCSN waveforms

Figure 3.10: Percentage change in h_{rss50} sensitivity at iFAR $\geq$ 100 years for the updated methodology *GMM+* compared to the old GMM methodology seen in [2]. Results are shown for a generic set of ad-hoc waveform injections in (a) and for CCSN injections in (b). A negative change infers a reduction in h_{rss50}, hence an improvement in sensitivity with GMM+.

in sensitivity for the GMM+ methodology; since h_{rss} is a measure of signal amplitude, a smaller h_{rss50} signifies the ability of GMM+ to detect smaller amplitude signals. The relative errors associated with the percentage change are calculated from the upper and lower errors on h_{rss50} from each methodology:

$$\frac{\delta_{\%\text{change\_hrss50}}}{\%\text{change\_hrss50}} = \sqrt{\left(\frac{\delta h_{rss50\text{plus}}}{h_{rss50\text{plus}}}\right)^2 + \left(\frac{\delta h_{rss50\text{old}}}{h_{rss50\text{old}}}\right)^2} \tag{3.2}$$

Figure 3.10 displays the percentage change in h_{rss50} for the set of ad-hoc injections and CCSN injections. Looking first at the results for ad-hoc waveforms in Figure 3.10a, we see that GMM+ gains sensitivity to the majority of GA waveforms, while sensitivities to SG $Q = 100$ and WNB waveforms remain comparable. The GMM+ methodology has loss in sensitivity for SG injections with $Q = 3,9$, however this is somewhat expected since we are removing bias from the signal model. Looking at the results for CCSN in Figure 3.10b, we see that the GMM+ methodology brings sensitivity improvement for all considered waveforms, which further shows the benefit of removing ad-hoc waveform bias from the models. These results prove that the enhanced methodology of GMM+ has increased sensitivity and robustness to astrophysical source populations. We will demonstrate this statement further in the next Chapter where we apply the enhanced method to a wider range of astrophysically motivated populations for the full third LVK observing run. With the above proof that the enhanced methodology of GMM+ has increased sensitivity and is robust across a variety of waveforms, we now consider it as the default GMM methodology. Hence, in the remaining Chapters we refer to the GMM+

methodology only as 'GMM'.

3.4 Summary and discussion

This Chapter has presented an enhanced approach to the method of GMM applied as post-production to the cWB algorithm in the search for short duration GW transients. GMM aids in the mitigation of glitches within the weakly-modelled cWB algorithm by modelling the noise and signal populations and calculating a log-likelihood ratio on search triggers to distinguish GW signals from glitches. Through investigations into various aspects of the GMM methodology, we have achieved an increase in the sensitivity, reliability and robustness across the expected short duration parameter space, compared to the previously proposed cWB+GMM methodology presented in [2].

A new method to optimise the number of Gaussian components in each model was developed, replacing the previously used Bayesian information criterion (BIC). The newly proposed method consists of optimising detection efficiency over a range of background and signal models trained with varying number of Gaussian components, with a validation data set. Maximal detection efficiency will give the optimal combination of Gaussian components for each model. While the BIC provided a measure of how well the Gaussians fit to the data, the newly proposed method provides a combined measure of Gaussian fit to the data with a measure on the models' ability to distinguish between the background and signal populations, which was found to reliably provide better results.

Additionally, a new signal training data set was introduced, consisting of generic WNB simulations spanning the targeted short-duration, low-frequency parameter space. The motivation behind this was to remove a bias in the signal model towards ad-hoc waveforms, which was present in previous methodology due to the choice of training data. The use of the new generic WNB training data made the search robust to a wider range of expected sources due to it's generic nature, and also allowed for the full ad-hoc simulation data set to be utilised for testing sensitivities, decreasing error margins. With this new training data set, it was discovered that the L_{veto2} attribute was not distinguishable between background and signal populations, and a sensitivity increase was observed with it's removal from the analysis. While the above methodology changes bring benefits to the analysis, it is possible that features such as training data and analysis attributes may be reconsidered in the application to future observing runs, as they are dependent on features in the data and the targeted parameter space of the analysis.

The optimisation of the GMM model training process was explored through the modification of function parameters, such as settings on the initialisation and convergence of the EM algorithm. The aim of these studies was to find the correct balance between accurate model fitting and computational run time. Overall, it was found that the k-means++ method of Gaussian initialisation did not bring visible benefits to the fitting for this data set, however is thought to

be more reliable than the k-means method alone due to informed Gaussian placement. It was also discovered that the default thresholds on convergence limit and number of total iterations were suitable for the complexity of data sets being handled. The effect of random state was also investigated, in which it was found that the majority of random states produced consistent results, with only an occasional outlier state resulting in 'bad' placement of Gaussians, affecting the final sensitivity of the search. The consistency in results proves that the GMM training procedure is reliable. However to account for the occasional outlier choice in random state, a new step was considered in the model training process, in which a small selection of random states can be tested to ensure the chosen random state is not an outlier. In order to prevent a bias from over-fitting, the choice of models is not optimised on random state.

The optimised performance of the enhanced GMM methodology was demonstrated through the direct comparison with the previous GMM methodology on a range of ad-hoc and CCSN waveforms. It was seen that the new methodology obtains significant sensitivity improvement to GA and CCSN waveforms. A loss in sensitivity was observed for low bandwidth SG waveforms, however this was a direct effect of the bias which was present in the previous methodology signal model.

To conclude, alterations made to the GMM methodology have brought improvements to the search of cWB+GMM, in the form of both sensitivity and robustness across the short duration parameter space. In addition, investigations into the model fitting parameters have increased the reliability of the analysis. The application of the enhanced methodology to O3a data have demonstrated the potential for GMM to be an effective technique in the search for short duration GW bursts, however further sensitivity tests are required in order to fully investigate it's performance. For example, testing the sensitivity to a wider range of astrophysical sources, testing on more data, and comparing to a range of search algorithms currently employed in the all-sky short search. This will be explored in further Chapters, where the enhanced cWB+GMM search is implemented with various applications in full O3 data.

Chapter 4

Application of cWB+GMM during the third LVK observing run

In Chapter 3, investigations leading to an enhanced version of the Gaussian mixture modelling (GMM) methodology applied to the coherent WaveBurst (cWB) algorithm were detailed and it was demonstrated that this new methodology increased the sensitivity of the cWB+GMM analysis to an array of waveforms in the all-sky short, low-frequency parameter space using data from the first half of the third observing run (O3a). In this Chapter, we apply this enhanced cWB+GMM methodology to data in the *full* third LIGO-Virgo-KAGRA (LVK) observing run (O3) for the first time, in the search for short gravitational wave (GW) transients. Note that since the enhanced methodology of GMM+ has now been adapted as default, we refer to this updated methodology only as 'GMM' in the following sections.

The O3 run was split into two epochs to allow for detector commissioning: O3a ran from April 1st 2019 to October 1st 2019, while O3b ran from November 1st 2019 until March 27th 2020, during which the LIGO-Livingston, LIGO-Hanford and Virgo detectors were active. Despite three detectors being available, previous burst short duration searches [85] have prioritised the 2-detector LIGO-Livingston, LIGO-Hanford (LH) network due to their similar alignment, which aids in the discrimination of glitches due to the their apparent position in the sky during likelihood calculation. Here we consider both the 2-detector and 3-detector networks in order to investigate whether GMM can mitigate the effect of glitches in the 3-detector search. First, full search results are presented for the O3 all-sky short search with the enhanced cWB+GMM pipeline for the 2-detector LH network in section 4.1. Sensitivity estimates to a range of expected burst sources are detailed, and GWs detected by the analysis are discussed. The cWB+GMM algorithm is extended to the 3-detector LIGO-Livingston, LIGO-Hanford, Virgo (LHV) network in section 4.2 for the first time, with sensitivity comparisons to the 2-detector network and GW detections again detailed. The above sections 4.1 and 4.2 follow work published by the author in Smith et al. 2024 [3], however details and plots are expanded here. Lastly, the participation of cWB+GMM in the burst benchmark project is detailed in section 4.3, in which the performance

of the algorithm is compared to other pipelines for a wide array of expected burst sources.

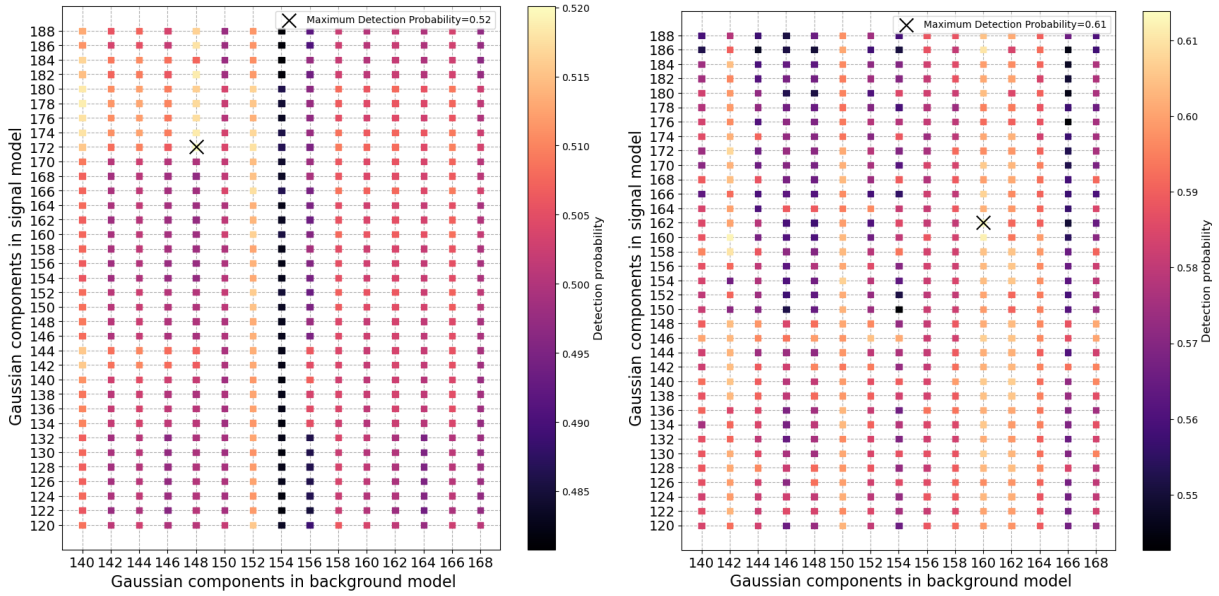
4.1 Running GMM on the full O3 all-sky short search

Considering the 2-detector LH network, a total coincidence time of 104.9 days was collected during O3a, with 980.7 years of background generated through time-shifting data segments. During O3b, there was 101.63 days of coincident data to be analysed, and 1096 years of background. Following the methodology outlined above, we train separate models for O3a and O3b to account for the different sources of noise occurring in each half of the observing run. We present sensitivities and results combined over O3a+O3b in the end stages of analyses.

4.1.1 Training optimal models for O3

The first step in the analysis is to split and re-parameterise the data. The data splits applied are the same for both O3a and O3b, with a 10% validation, 70% training and 20% testing split on background triggers, and a 20% validation, 80% training split on the simulated WNB data. This reserves a total of 196.1 years of data for false alarm rate (FAR) calculation in O3a, and 219.2 years for FAR in O3b. We apply the re-parameterisations detailed in Table 3.1.

For training of the models, we use the initialisation settings detailed in Section 3.2.4. The optimal number of Gaussian components must be selected for each signal and background model through the validation process, as detailed in section 3.2.1. Models are trained over ranges of number of Gaussian components (*ncomp*), with validation datasets used to calculate the detection efficiency at an iFAR threshold of 50 years for each combination of signal and background model. Initially the range of Gaussian components has wide increments so regions with high detection efficiency can be identified before the process is completed in smaller steps of ncomp. For both O3a and O3b, the regions of interest are 120-188 number of Gaussian components in the signal model and 140-168 in the background model, with 2 step increments. Looking at Figure 4.1, we see the variation of detection efficiency over each stated range of ncomp values for O3a and O3b separately. Detection efficiency for each background and signal model combination is shown in the z-axis with the optimal combination of models labelled with an 'X'. For O3a, the plot gives optimal detection efficiency for 148 components in the background model and 172 in the signal model, whereas for O3b we have 160 components in the background model and 162 components in the signal model. Overall we can see that the variation in detection efficiency is small for both plots, with variation of ≈ 0.05 in O3a and ≈ 0.065 in O3b. In depth studies into the effect of this variation on final sensitivities could be beneficial to the analysis in the future. In particular, whether similar sensitivities on test data can be achieved with lowered number of Gaussian components, or whether these small variations have a significant effect on final results. However for this purpose of this study, investigations in section 3.2.1 prove this method to be trustworthy.



(a) O3a (b) O3b

Figure 4.1: Variation of detection efficiency at iFAR $\geq$ 100 years over models trained with different numbers of Gaussian components, for O3a (a) and O3b (b) separately. Number of Gaussian components for the signal model is displayed on the y-axis, for background model on the x-axis, while detection efficiency calculated with combinations of the models is seen in the z-axis. The combinations of models resulting in highest detection efficiency is marked by an 'X'.

4.1.2 Sensitivities to simulated data

The above optimised GMM models are applied to sets of simulated waveform injections into respective observing run data segments in order to investigate the sensitivity of the cWB+GMM analysis across the burst short duration parameter space. Here we consider three populations of waveforms: generic ad-hoc, core-collapse supernovae (CCSN) and cosmic string (CS) waveforms, injected into cWB following the process detailed in section 2.2.5 over different waveform amplitudes (or h_{rss}). The ad-hoc injections are the same set as considered in Table 3.3, consisting of GAs varied over duration $\tau = [0.1, 1, 2.5, 4]$ms, SGs varied over frequencies $f_0 = [70, 100, 153, 253, 361, 554, 849]$Hz, quality factor $Q = [3, 9, 100]$ and WNBs varied over low frequency bound $f_{low} = [150, 300, 700]$Hz with bandwidth $\Delta f = 100$Hz and duration $\tau = 0.1$s. These waveforms were injected over h_{rss} of $(\sqrt{3})^N \times 5 \times 10^{-23}$, where N ranges from 0-8. We also utilise the same set of CCSN waveforms as described previously, namely 10 neutrino explosion models: Andersen et al. 2017 [169] (And s11), Müller et al. 2012 [170] (Mul L15), Kuroda et al. 2016 [171] (Kur SFHx), O'Connor & Couch 2018 [132] (Oco mesa20), Powell & Müller 2019 [131] (Pow he3.5,s18), Radice et al. 2019 [133] (Rad s9, s13, s25), and 1 magnetorotationally driven explosion model: Abdikamalov et al. 2014 [172] (Abd A4O01.0). Finally, we consider a set of CS cusp injections consisting of four waveforms with lower frequency cut-off of 1Hz and high frequency cut-off [50, 150, 500, 1500]Hz as seen in [97]. In past studies, it has

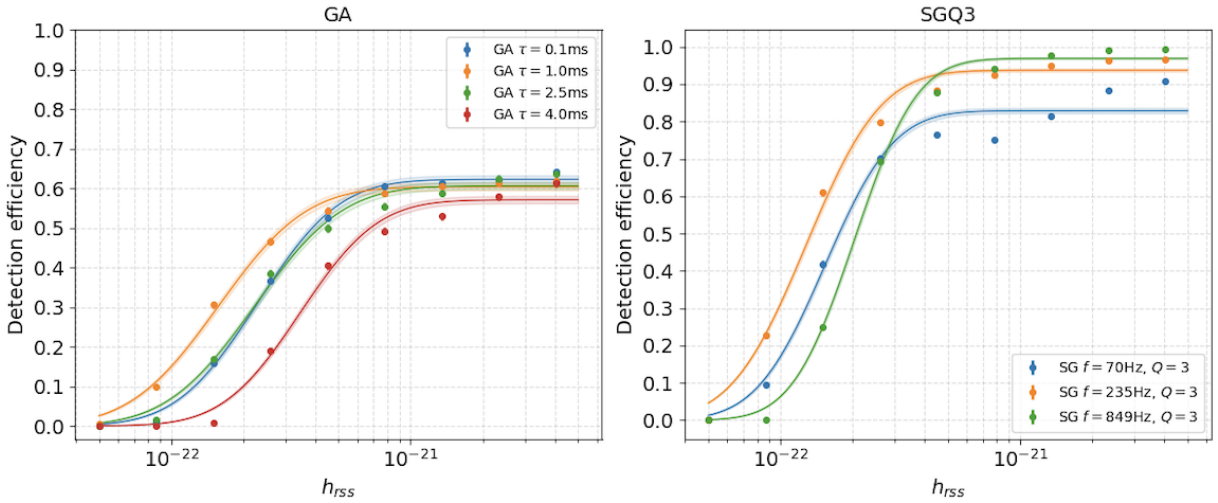


Figure 4.2: Example of detection efficiency as a function of h_{rss} for grouped GA and SG $Q = 3$ waveform injections in O3. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve.

been found that pipelines have lower sensitivities to this population of waveforms, since they are affected by blip glitches due to morphology, while only 10% of the CS injected amplitude falls inside the LVK sensitivity band. For this reason, they are injected with higher h_{rss} than other waveforms, distributed over $(\sqrt{3})^{(N+2)} \times 5 \times 10^{-23}$, where N ranges from 1-9.

The sensitivity to each injected waveform is defined by the h_{rss50} statistic, the value of h_{rss} at which we detect 50% of injected waveforms, as described in section 2.2.5. This is found by plotting the detection efficiency of each waveform type as a function of injected h_{rss}, fitting the data with a cumulative lognormal distribution and inferring the h_{rss} amplitude at which 50% of waveform injections are detected at given FAR threshold. Errors on detection efficiency are calculated as described in eqn. 2.16, leading to upper and lower regions of error using further lognormal fits and error bounds on h_{rss50} estimates. By creating this plot for combined O3a and O3b injections, we can estimate sensitivities for the full O3 observing run. An example of this for the ad-hoc GA and SG $Q = 3$ waveforms can be seen in Figure 4.2 for an iFAR threshold $\geq$ 100 years, while plots for all ad-hoc, CCSN and CS waveforms can be found in Appendix B. In Fig 4.2, it is seen that the error regions are smaller here than seen in previous efficiency curves in Fig 3.9, due to the consideration of the entire simulated injection set as test data rather than only 20% as in the previous methodology, as discussed in section 3.3.

We compare our h_{rss50} sensitivities to two other cWB post-production methods: 'standard' (STD) cWB and 'XGBoost' (XGB) cWB. STD cWB post-production follows the methodology detailed in section 2.2.4 with no machine learning present, instead separating triggers into different search bins based upon their signal properties. This is the methodology which was followed for the O3 LVK all-sky short burst search in [85]. Similarly to the GMM post-production, the XGBoost methodology replaces the binning seen in STD cWB, however instead uses a ma-

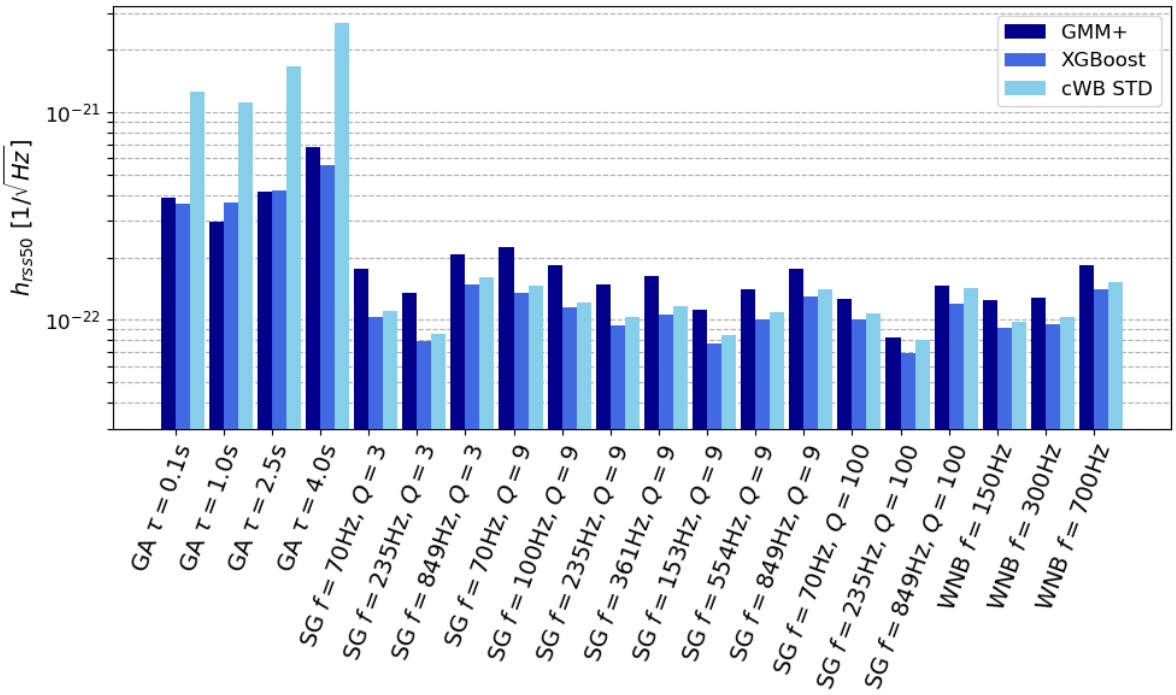
chine learning boosted decision tree methodology to aid the classification of triggers [123, 124]. For fair comparison, the 3 methods of post-productions were run on the same cWB production triggers. h_{rss50} values for both standard and XGBoost post-productions were taken from [124].

The sensitivity comparisons for all 3 populations of waveforms and post-production methods are seen in Figure 4.3, calculated at a significance threshold of iFAR $\geq$ 100 years. Exact values are quoted in Table 4.1 in Section 4.2, for comparison with the other post-production pipelines and results with a 3-detector network. First looking at the ad-hoc injections in Fig 4.3a, we see that the GMM post-production significantly enhances sensitivities for GA waveforms compared to the standard post-production, while achieving similar sensitivities to the XGBoost post-production. No sensitivity improvement is seen for the majority of SG and WNB injections, with GMM having worsened sensitivities by upto 40%. Fig 4.3b details the h_{rss50} comparisons for CCSN waveforms. Of the total 11 waveforms, the GMM methodology brings sensitivity improvements to the *Rad s25* [133] neutrino-driven explosion model and the magnetorotationally-driven explosion model *Abd A4O01.0* [172] compared to both other post-production methods. While GMM is not the most sensitive for other CCSN waveforms, it still performs well, achieving h_{rss50} within 17% of standard and 30% of XGBoost. Finally, we look at the sensitivities to CS waveforms in fig 4.3c. The GMM post-production brings improvement in sensitivity to all injected CS cusp waveforms with high frequency cut-off above 150Hz, with a 75% reduction in sensitivity compared to standard post-production and a 30% reduction compared to XGBoost. The waveform with high frequency cut-off of 50Hz achieves comparable sensitivity to the standard post-production, however has much worse sensitivity than XGBoost.

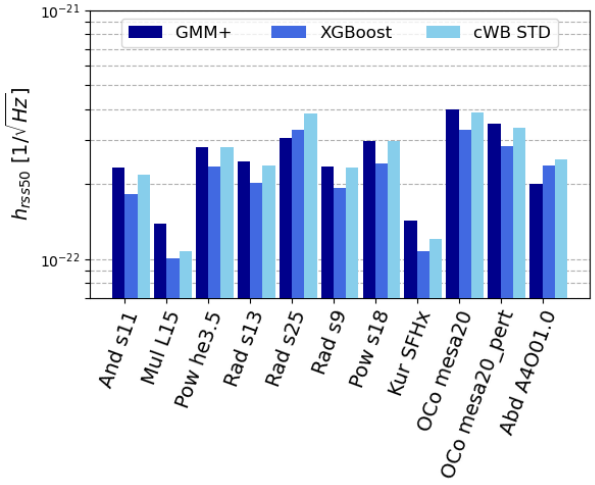
Sensitivity calculations are an important step for burst pipelines, as not only do they allow performance checks on the analysis methods, but also allow for upper limit rate estimations in the case of non-detections which can give critical insights into astrophysical populations. Overall we see that during the O3 observing run, the updated cWB+GMM methodology achieves comparable performance to the alternative post-production methods. While a loss in sensitivity is seen for SG and WNB waveforms, there is a significant improvement for GA and CS waveforms. This improvement in particular demonstrates the ability of GMM to mitigate problematic blip glitches, since they occupy the low-Q region of the parameter space where GA and CSs lie and have previously limited sensitivities to these waveforms. The ability of GMM to achieve comparable sensitivities to other post-production methods for the other considered waveforms is also a major result, as it allows for cross-checking in the case of a possible burst-type detection, or can reinforce upper-limit estimates on such populations.

4.1.3 Analysing zero-lag data

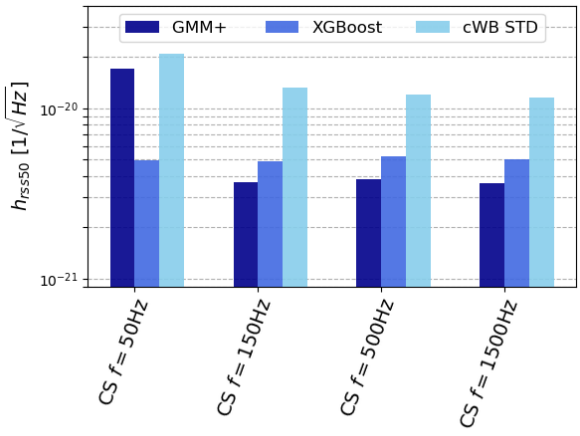
As stated above, sensitivity studies on injected populations of waveforms allow us to ensure the analysis pipeline is working to good standards with sensitivity to a wide range of sources before applying the analysis to any real data. This is important as once we analyse the incoming GW



(a) ad-hoc waveforms

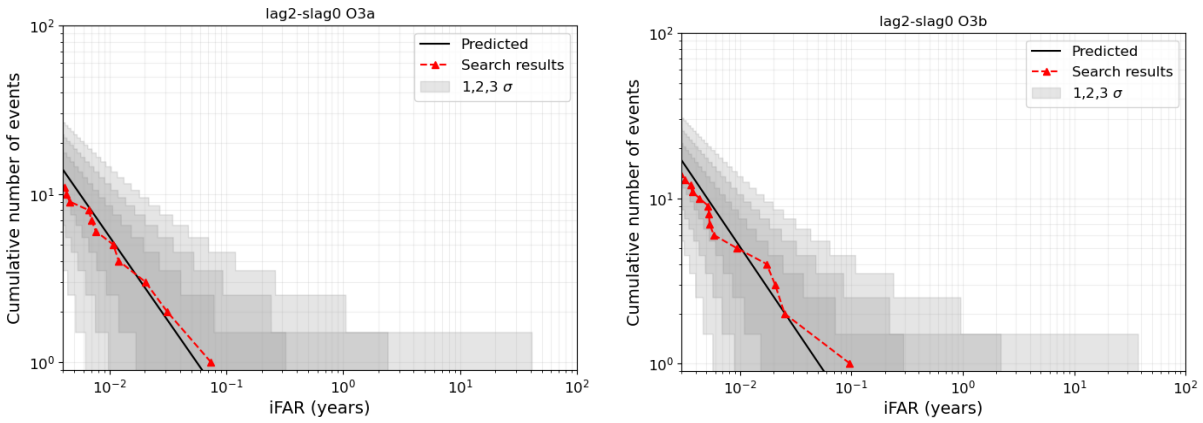


(b) CCSN waveforms



(c) CS waveforms

Figure 4.3: h_{rss50} sensitivity estimates for waveform injections in O3 data at an iFAR threshold of 100 years for cWB+GMM post-production (dark blue), cWB+XGBoost (royal blue) and the cWB standard post-production method (light blue). (a) presents sensitivities for generic ad-hoc waveforms, while core-collapse supernovae (CCSN) sensitivities are presented in (b) and cosmic string (CS) in (c).



(a) O3a closed box

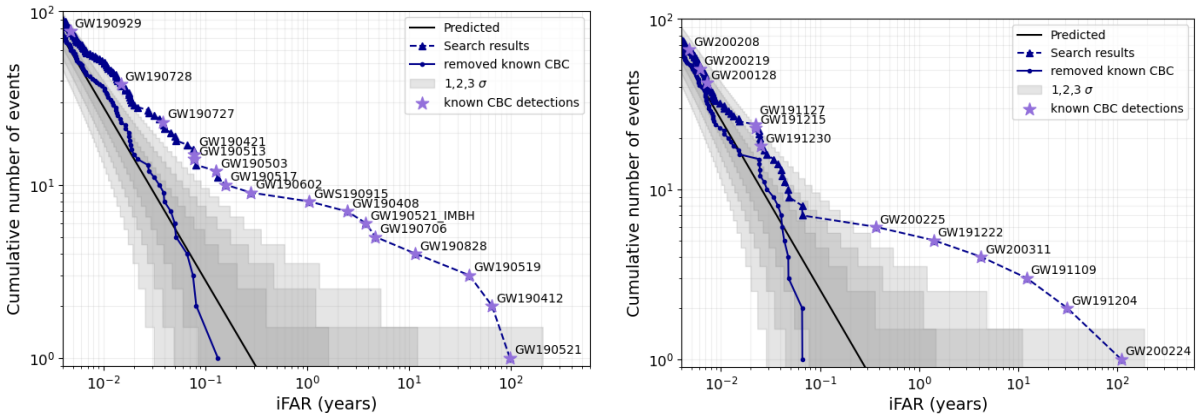
(b) O3b closed box

Figure 4.4: Closed box results for O3a (a) and O3b (b). The plots show cumulative number of events vs. iFAR for the time-shifted background segment of lag=2, super-lag=0 in red. Expected background is shown in black with grey regions representing $1, 2, 3\sigma$ errors.

data once, we should not re-analyse with different settings in order to avoid bias. The process of analysing this real data incoming from the detectors is referred to as 'opening boxes', or in cWB as analysing the 'zero-lag' data since it is the real detector data with no time-shifts applied. Since this analysis should only be completed once, we carry out a sanity check through closed box analysis, which involves analysing a time-shifted segment from the background data as a final check that the pipeline is working as expected, without bugs. In this section, we will perform the closed-box analysis for O3a and O3b separately, demonstrating that the analysis is working as expected, before going on to analyse the zero-lag data in the search for GW detections from short burst transients.

In order to complete closed box analysis, we must select a time-shifted segment to process. The choice of segment can be completely random and the process can be repeated as many times as desired. For the analysis here, we select the lag=2, super-lag=0 segment for both O3a and O3b. The segment undergoes log-likelihood and T estimation alongside the background test data to assign FARs to triggers in the segment. Since the segment is taken from the background data consisting of only detector noise, the expectation is that the assigned FARs will be consistent with the predicted background distribution. We observe these results in Figure 4.4, where cumulative number of events vs. iFAR is shown for closed box segments of O3a and O3b separately. In both plots, the predicted background is calculated by scaling the total considered background to the observed time of the segment, with calculated $1, 2, 3\sigma$ errors. For both O3a and O3b, we see that the analysed segment agrees with the predicted background to within 1σ, as expected.

As the closed box analysis demonstrates that the behaviour of the pipeline is as expected, we now go on to analyse the zero-lag data incoming from the detectors during the O3 observing run, ie. open box analysis. We again process the data from O3a and O3b separately so that



(a) O3a open box

(b) O3b open box

Figure 4.5: Open box results for the cWB+GMM LH analysis showing cumulative number of events vs. iFAR for O3a (a) and O3b (b). In both plots, full search results are notated by dashed triangular markers with known CBC detections labelled by purple stars. Search results with known CBC events removed are notated by solid circular blue markers, with predicted background in black with $1, 2, 3\sigma$ error regions shown in grey.

iFAR significance can be applied by the respective models, combining the detection results in the final stages. Known CBC detections from [9, 11] are used to label CBC GW events, so that any triggers with high significance and no CBC label are referred to as burst-type detections. Zero-lag results are shown for O3a and O3b separately in Figure 4.5, where cumulative number of events vs iFAR is shown with known CBC detections labelled. The results with known CBC detections removed are also shown, and are seen to be consistent with the predicted background. Thus, we conclude that all significant detections are known CBCs with a null result on burst-type events in both O3a and O3b. The loudest CBC events detected are GW190521 and GW200224 at iFAR of 98 years and 110 years respectively.

Analysed zero-lag is combined for the entire O3 observing run in Figure 4.6, where we compare the cumulative number of triggers vs. iFAR found with GMM post-production to the XGBoost and standard methods. Here, all search results are notated by dashed triangular markers while results with known CBC removed are notated by solid lines. The GMM analysis detects 14 CBC events for iFAR $\geq$ 1 year, similar to the standard post-production (14) and slightly less than XGBoost (16). Overall the GMM methodology detects known CBC events with lower significance, however this is not surprising as it is not a targeted CBC search and is most sensitive to waveforms with dissimilar morphologies. A full breakdown of the detected CBC events and their significance compared to the other cWB post-productions is detailed in Table 4.2 in the below section 4.2.3.

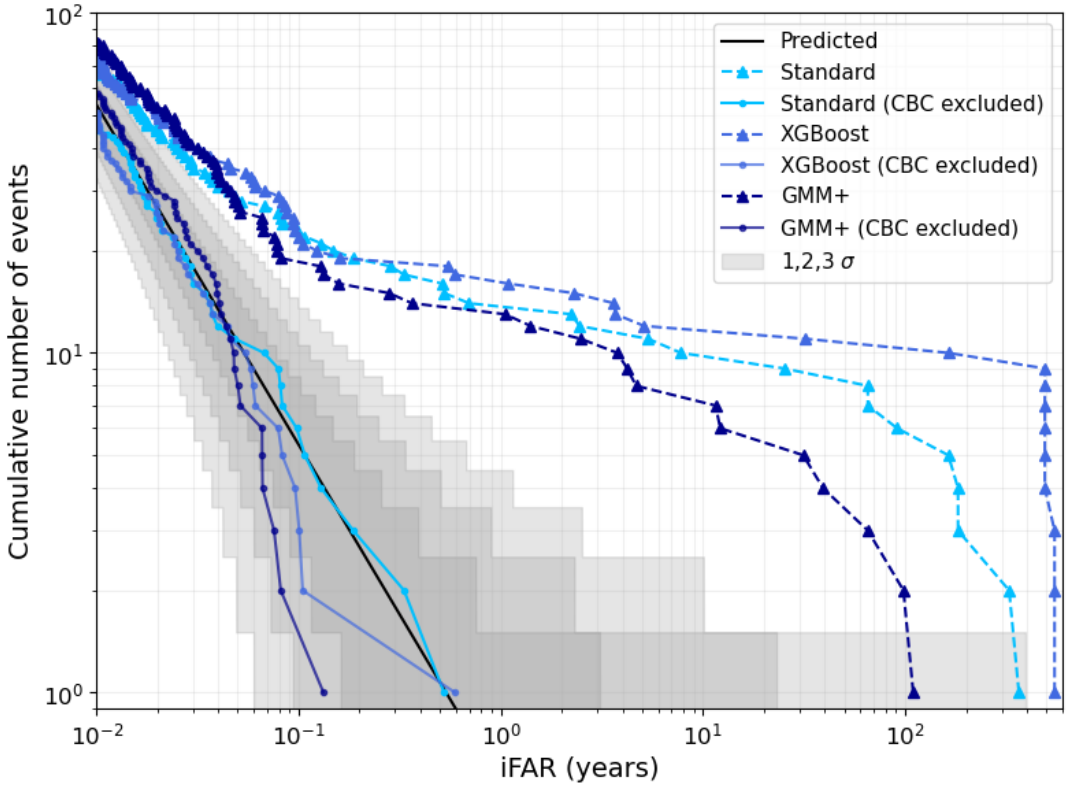


Figure 4.6: Comparison of open box results showing cumulative number of events vs. iFAR for O3 LH network between the GMM (navy), XGBoost (royal blue) and standard (light blue) post-productions. Dashed triangular lines represent search results including known CBC detections for the various post-production methods, while solid lines represent the results with known CBC events removed. Predicted background is shown in black with $1, 2, 3\sigma$ error regions shown in grey.

Analysing standard vs. extended segments

In order to optimise computational efficiency in cWB, data segments are processed in parallel during trigger production. The length of segments can vary within defined maximum/minimum lengths and are chosen to optimise observation livetime considering periods of data which have been removed due to data quality issues. These are referred to as *standard* segments and are the ones which have been presented in the figures above. However, the choice of segment lengths can sometimes lead to small portions of incoming data being missed in analysis. To fix this, cWB introduces *extended* segments, which instead consider the entire period of observing time and split the data into even segment lengths. We analyse extended segments in the same way as standard segments, searching for any GW detections which may have been missed previously. When quoting significance values, we state those found with standard segments unless the event only exists in the extended segments as the PSD estimation is more accurate in standard segments. For the cWB+GMM O3 analysis, we find one CBC event from extended analysis, which is included in the total 14 detections and is quoted in Table 4.2.

4.2 Extension to the 3-detector network in O3

The analyses thus far has concerned only the 2-detector LH network, as it has previously been known for the 2-detector network to achieve higher sensitivities than the 3-detector network, for example in the standard cWB O3 all-sky short search in [85] and the O3 search enhanced by XGBoost post-production in [124]. This is due to the 2 LIGO detectors being nearly co-aligned, meaning cWB can more efficiently mitigate glitches during the likelihood calculation (introduced in section 2.2). However, it was demonstrated above that GMM can effectively aid in the mitigation of glitches, particularly blip glitches, thus we extended the cWB+GMM analysis to the 3-detector LIGO-Livingston, LIGO-Hanford, Virgo (LHV) network during the full third observing run (O3) to test it's capabilities. After detailing some required alterations to the analysis below, we will follow the same process as detailed for the 2-detector network in section 4.1, namely optimise the models, investigate sensitivities to injected populations and search for GW detections in the LHV data through open-box analysis. While it is interesting to compare the GMM results with those found by other post-productions, we here want to concentrate on whether GMM can minimise the sensitivity gap between 2- and 3- detector networks.

Similarly to before, we initially consider the two epochs of O3 individually, obtaining optimised models for O3a and O3b separately. For the 3-detector LHV network, 79.12 days of coincident data were collected during O3a and 72.24 days for O3b. Again using the time-shifting analysis we obtain 572.29 years of background for O3a and 395.8 years in O3b.

4.2.1 Minor alterations to analysis

By default we consider the same cWB attributes as in the LH network, ie. the ones stated in Table 3.1. Some of these attribute definitions are dependent on the number of detectors in the network, hence some re-parameterisations need to be altered to avoid invalid values. Specifically, we define the updated re-parmeterisation of $\log 10(N_{ED}+2000)$ and $\log 10(Q_{veto1}+1)$ to account for increased ranges in these distributions. Additionally, the re-parameterisation of both network correlation coefficients $c_{c0,2}$ must also be altered, since the dependence on number of detectors results in some values being greater than 1 which is invalid for inverse sigmoid transformations. Since this problem only occurs for a small proportion of very highly correlated triggers, we simply set any values of $c_{c0,2} \geq 1$ to be equal to 0.9999. The rest of the attribute re-parameterisations remain the same. These new re-parameterisations could be better optimised and a full study of which attributes are best considered in 3-detector analysis would be beneficial, however the current set-up is enough for the initial investigations considered here.

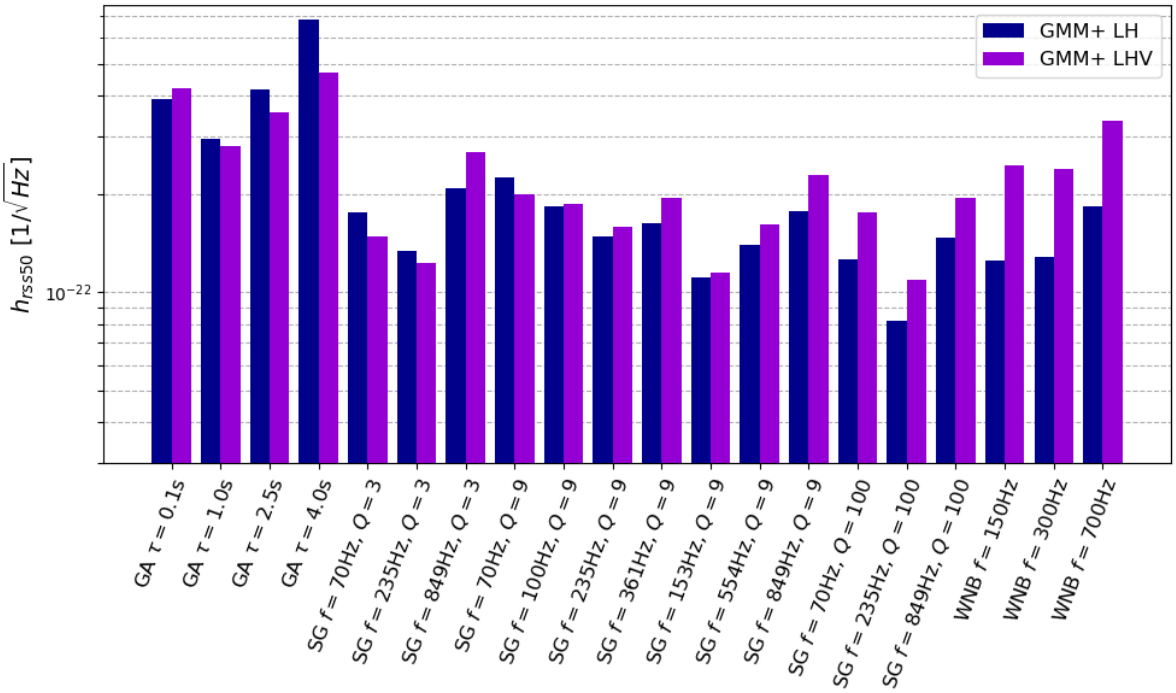
The cWB LHV analysis typically has less coincident data, leading to less accumulated background. Hence in order to increase the amount of data available for FAR estimation, we alter the background data split to 10% validation, 60% training and 30% testing data, reserving 171.87 years of test data in O3a and 118.75 years in O3b.

The process of GMM methodology itself remains the same, with models using the same initialisation methods as chosen in section 3.2.4, and undergoing the optimisation of number of Gaussian components (ncomp) through the validation process. From this, LHV models for O3a have optimal ncomp of 244 for signal model and 120 for background model, while O3b has 258 for signal and 114 for background.

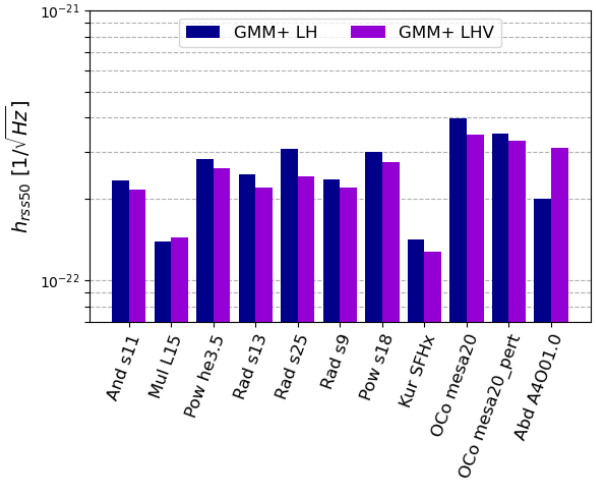
4.2.2 Sensitivity comparisons to 2-detector network

We investigate the sensitivities to the same sets of source populations as considered in the LH analysis in section 4.1.2, namely the generic ad-hoc waveforms of GA, SG and WNB with parameters outlined in Table 3.3, the set of 10 neutrino explosions and 1 magnetorotationally-driven explosion models of CCSN and the 4 CS cusp waveforms. This time we compare h_{rss50} sensitivities directly to those achieved with the 2-detector LH network, since we are interested in whether the 3-detector network can perform to the same level. If interested in direct sensitivity comparisons with the standard and XGBoost post-productions, one can refer to Smith et al. [3], or can refer to Table 4.1 for quoted h_{rss50} estimates for all 3 post-productions across all considered waveforms.

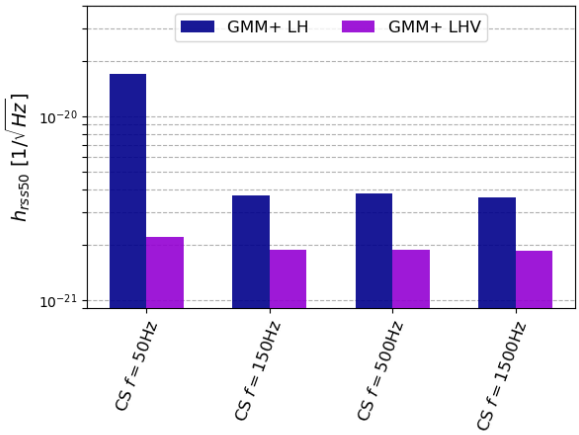
Sensitivity comparisons for all 3 injected populations are seen in Figure 4.7, where previous GMM 2-detector network h_{rss50} estimates are compared against 3-detector h_{rss50} at a threshold of iFAR $\geq$ 100 years. First looking at the ad-hoc waveform injections in Fig 4.7a, we see that the LHV GMM analysis achieves similar sensitivities to the 2-detector network for GA and low-Q



(a) ad-hoc



(b) CCSN



(c) CS

Figure 4.7: h_{rss50} sensitivity estimates for waveform injections in O3 data at an iFAR threshold $\geq$ 100 years for cWB+GMM post-production on LHV data (purple) vs LH data (dark blue). Results are shown for generic ad-hoc waveforms in (a), core-collapse supernovae waveforms in (b), and cosmic string waveforms in (c).

Waveform	h_{rss50} ($\times 10^{-22}\ 1/\sqrt{Hz}$)					
	LH network			LHV network		
	GMM	STD	XGB	GMM	STD	XGB
Gaussian Pulse						
$\tau = 0.1$ms	3.9	12.6	3.6	4.2	17.5	11.7
$\tau = 1.0$ms	2.9	11.1	3.7	2.8	13.9	8.4
$\tau = 2.5$ms	4.2	16.7	4.2	3.5	31.8	10.9
$\tau = 4.0$ms	6.8	27.0	5.5	4.7	94.5	13.8
Sine-Gaussian						
$f_0 = 70$Hz, Q=3	1.8	1.1	1.0	1.5	1.2	1.2
$f_0 = 70$Hz, Q=9	2.2	1.5	1.4	2.0	1.4	1.4
$f_0 = 70$Hz, Q=100	1.3	1.1	1.0	1.8	1.1	1.0
$f_0 = 100$Hz, Q=9	1.8	1.2	1.1	1.9	1.2	1.2
$f_0 = 153$Hz, Q=9	1.1	0.8	0.8	1.1	0.9	0.9
$f_0 = 253$Hz, Q=3	1.3	0.9	0.8	1.2	0.9	0.9
$f_0 = 253$Hz, Q=9	1.5	1.0	0.9	1.6	1.1	1.0
$f_0 = 253$Hz, Q=100	0.8	0.8	0.7	1.1	0.8	0.7
$f_0 = 361$Hz, Q=9	1.6	1.2	1.1	1.9	1.3	1.2
$f_0 = 554$Hz, Q=9	1.4	1.1	1.0	1.6	1.2	1.2
$f_0 = 849$Hz, Q=3	2.1	1.6	1.5	2.7	1.8	1.7
$f_0 = 849$Hz, Q=9	1.8	1.4	1.3	2.3	1.6	1.5
$f_0 = 849$Hz, Q=100	1.5	1.4	1.2	1.9	1.5	1.4
White Noise Burst						
$f_{low} = 150$Hz	1.2	1.0	0.9	2.4	1.1	1.0
$f_{low} = 300$Hz	1.3	1.0	1.0	2.4	1.2	1.1
$f_{low} = 700$Hz	1.8	1.5	1.4	3.3	1.8	1.5
Core-collapse Supernova						
And s11	2.3	2.2	1.8	2.2	2.9	2.2
Mul L15	1.4	1.1	1.0	1.4	1.2	1.1
Pow he3.5	2.8	2.8	2.4	2.6	3.9	2.6
Rad s13	2.5	2.4	2.0	2.2	3.1	2.2
Rad s25	3.1	3.9	3.3	2.4	5.1	3.7
Rad s9	2.4	1.3	1.9	2.2	3.3	1.3
Pow s18	3.0	3.0	2.4	2.7	4.2	2.7
Kur SFHx	1.4	1.2	1.1	1.3	1.4	1.2
Oco mesa20	4.0	3.9	3.3	3.5	5.5	4.7
Oco mesa20_pert	3.5	3.4	2.8	3.3	4.9	3.5
Abd A4O01.0	2.0	2.5	2.4	3.1	3.0	13.1
Cosmic String						
$f = 50$Hz	170.0	208.3	49.8	22.1	336.4	246.7
$f = 150$Hz	37.0	133.5	48.8	18.7	180.2	117.9
$f = 500$Hz	38.1	119.6	52.0	18.7	155.6	114.0
$f = 1500$Hz	36.3	114.6	50.1	18.5	148.4	106.5

Table 4.1: Table detailing the h_{rss50} in units of $\times 10^{-22}\ 1/\sqrt{Hz}$ achieved at an iFAR $\geq$ 100 years for each injected waveform in O3 for 2-detector LH and 3-detector LHV networks with GMM. Includes comparisons across the other cWB post-production methodologies, with values for the standard (STD) and XGBoost (XGB) taken directly from [124].

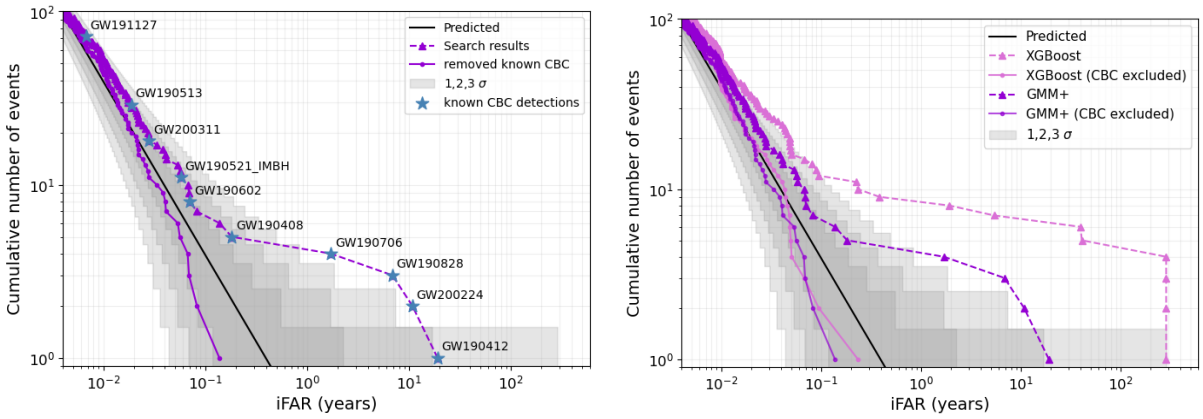
SG waveforms, with only slightly lower sensitivities for high-Q SG and WNB waveforms. For the CCSN injections in Fig 4.7b, the LHV analysis again performs similarly to the LH network, improving sensitivities up to 25%, apart from the *Abd A4O01.0* waveform. Lastly, we consider the CS injections in Fig 4.7c, where it is shown that the LHV network outperforms the 2-detector LH network in terms of sensitivity for all 4 waveforms.

This is a crucial result, since we show that the GMM methodology applied to the 3-detector LHV network can achieve similar, and in some cases better, sensitivities than the 2-detector LH network. This behaviour has not been seen before with other cWB post-production methods, especially for waveforms in the low-Q parameter space such as GA and CS which usually have much lower sensitivities for the LHV network due to glitches. The good sensitivities achieved for these waveforms with the GMM LHV analysis further reinforce the ability of the GMM methodology to mitigate problematic glitches, especially blip glitches.

4.2.3 Analysing zerolag

We open the boxes for the cWB+GMM search for LHV O3 data following the same process as detailed in section 4.1.3. Closed box analysis was again completed on the time-shifted segment lag=2, super-lag=0, however since an example of such analysis has been seen previously in Figure 4.4, we do not display the closed box plots here. The open box results of cumulative number of events vs. iFAR for LHV analysis are shown in Figure 4.8. Looking first at Fig 4.8a, we present the search results combined over O3a+O3b with known CBC detections labelled, and search results with these known CBC events removed for the GMM analysis. We again see that all significant events are known CBC detections, and once removed from the data the search results are consistent with the predicted background. Hence, the GMM LHV analysis concludes on a null result for burst-type GW detection. Turning our attention to Fig 4.8b, the GMM LHV search results are presented in comparison to the XGBoost LHV O3 search, with and without known CBC events considered. In total, the GMM analysis detects 4 significant CBC events with iFAR $\geq$ 1 year, which is less than detected by XGBoost (8). Similarly to the the conclusions drawn from the LH search results, it is not concerning that GMM is less sensitive to CBCs than XGBoost, since the LHV GMM analysis has better sensitivity to GA and CS waveforms, which have very different morphology to the known CBCs seen here. Additionally, while it can be useful for burst pipelines to be sensitive to CBC events for cross-checking purposes, it is not one of the targeted sources of the GMM methodology.

Figure 4.8 again shows search results considering the standard segments. We analyse the extended segments for the LHV data, only considering significance estimates if a new event is found. For the GMM search, we find one event from extended segment analysis. The details of this are presented alongside other detections in LH and LHV in Table 4.2 below.



(a) O3 LHV open box - detected CBC

(b) O3 LHV open box comparison

Figure 4.8: Open box results for the cWB+GMM LHV analysis showing cumulative number of events vs iFAR for full O3. (a) shows full search results notated by dashed triangular markers with known CBC detections labelled by blue stars. Search results with known CBC events removed are notated by solid circular markers. (b) shows a comparison between GMM (purple) and XGBoost (pink), with full search results again notated by dashed triangular markers and results with removed CBC by solid circular markers. In both plots predicted background is shown in black with $1, 2, 3\sigma$ error regions in grey.

Detailed results for LH and LHV analysis

Below we present the detailed results of the cWB+GMM search for short GW transients in O3, for both the 2-detector LH and 3-detector LHV networks. Table 4.1 details the exact $h_{rss}50$ estimates found by GMM across all injected waveforms, building upon Figures 4.3 and 4.7. The table also includes direct comparisons to the standard and XGBoost post-productions for both networks. Looking at the the comparisons between methodologies in the LHV network, we see that GMM brings significant improvement to the sensitivities to GA and CS waveforms by approximately an order of magnitude.

A list of GW CBC detections made by GMM in O3 is presented in Table 4.2, with comparisons of iFAR estimates across the other post-productions. The loudest event detected by the GMM LH analysis was GW200224, which was detected with higher significance in both standard and XGBoost post-productions. Similarly in the LHV analysis, the loudest event detected by GMM was GW190412, which again was detected with higher significance by XGBoost. As mentioned in above sections, we somewhat expect this behaviour since we do not target CBC sources. In this table we can also see the events which were picked up by the extended segment analysis: GW200225 for the LH analysis and GW190519 for the LHV analysis, both detected with low significance.

Event Name	LH network				LHV network		
	T	GMM+ iFAR (*yr*)	STD iFAR (*yr*)	XGB iFAR (*yr*)	T	GMM+ iFAR (*yr*)	XGB iFAR (*yr*)
GW200224	21.86	109.62	548.47	365.32	18.70	10.80	39.62
GW190521	32.37	98.09	490.79	326.88	-5.54	0.06	286.67
GW190412	23.38	65.40	490.79	15.10	12.92	19.10	286.67
GW190519 [1]	19.61	39.24	490.79	7.78	-2.04	0.45	40.95
GW191204	13.05	31.32	32.26	91.33	-	-	-
GW191109	9.63	12.18	548.47	182.66	-	-	-
GW190828	10.68	11.54	490.79	163.44	5.10	6.88	286.67
GW190706	7.55	4.67	490.79	65.38	0.68	1.70	286.67
GW200311	6.35	4.22	548.47	182.66	-0.24	0.03	0.39
GW190521	6.49	3.77	490.79	65.38	-5.54	0.06	286.67
GW190408	5.36	2.48	163.60	25.15	-3.71	0.18	5.41
GW191222	3.84	1.49	5.13	2.45	-	-	-
GW200225 [2]	3.09	1.06	5.13	2.45	-	-	-
GW190915	2.96	1.05	2.28	5.36	-10.55	0.004	0.019

Table 4.2: Table showing the search results of cWB+GMM in the O3 search for short GW transients for LH and LHV networks. Only events with iFAR $\geq$ 1 year are displayed and are arranged in decreasing order of iFAR for the LH network, with T values shown for each event. Significance estimates of the same events detected by standard and XGBoost post-productions are also included for comparison. Note that open box results were not available for standard LHV analysis, so LHV only has comparison with XGBoost.

[1] For LHV, this event was obtained from the extended segments in cWB.

[2] For LH, this event was obtained from the extended segments in cWB.

4.3 The Burst Benchmark Project

As mentioned in Section 2.1, the search for GW transients covers a large signal parameter space, often targeting a wide range of GW source populations. In fact the Burst signal parameter space covers the entire LIGO-Virgo-KAGRA (LVK) sensitive frequency range from $\approx$ 10Hz to 2kHz, in addition to searching for signals from millisecond to 10^3 second duration. In order to effectively search this parameter space, many different pipelines are needed, often targeting different regions of the frequency-duration space. The Burst Benchmark is a project set up by the LVK burst working group with the goal of assessing the performance pipelines across the whole parameter space and selecting pipelines for inclusion in official LVK search results and publications during the fourth observing run. Within this project, pipelines are directly compared over the same targeted frequency-duration regions, ensuring that sensitivity to targeted astrophysical sources is sufficient.

In order to keep the analyses completely fair to all participating pipelines, waveforms were injected into data by projects leads, with their morphology and parameters kept hidden from pipeline analysts. In this sense, the study could be treated as a real search for unknown signals, and pipelines could not knowingly create bias by increasing sensitivity to given waveform types. In the sections below, the provided dataset and considered injections by the project leads are described, before detailing the analysis completed by the cWB+GMM pipeline. Despite data being available for the entire burst parameter space, here the GMM analysis only considers the low-frequency, short duration region of the parameter space for the 2-detector LH network.

The benchmark dataset considers a 40 day period taken from the O3b observing run, from January 5th to February 14th 2020. After periods of poor data-quality have been removed, this totals approximately 24 days of coincident data for the LH detector network. Through the time-shift analysis, cWB accumulates 280.6 years of background from this data.

4.3.1 Data

An array of 40 waveform models were injected into the data over uniform sky position, with varying h_{rss} amplitudes and sequential, non-overlapping injections times. Initially unknown to participating pipelines, the 40 waveforms included a range of ad-hoc and astrophysically motivated models ranging over the entire frequency and duration space. Since the GMM analysis is only trained on the low-frequency, short duration parameter space, we detail a subset of 30 waveforms here that showcase the performance of the pipeline, seen in Table 4.3. A total of 8 generic ad-hoc waveforms are considered, consisting of sine-Gaussian with low and high frequency/Quality factor, white noise bursts with varying frequency, bandwidth and duration, and one Gaussian pulse. The rest of the waveforms are astrophysically motivated, including cosmic string cusps and kinks, core-collapse supernovae, a range of compact binary coalescence signals with varying eccentricity and mass, close hyperbolic BH encounters and models of ac-

Waveform	Details
Generic ad-hoc	
SG70Q3	Sine-Gaussian with $f_0 = 70$Hz & $Q = 3$
SG70Q100	Sine-Gaussian with $f_0 = 70$Hz & $Q = 100$
SG235Q3	Sine-Gaussian with $f_0 = 235$Hz & $Q = 3$
SG235Q100	Sine-Gaussian with $f_0 = 235$Hz & $Q = 100$
GA1	Gaussian Pulse with $\tau = 1$ms
WNB1	White noise burst with $f_0 = 150$Hz, $\Delta f = 100$Hz, $\tau = 0.100$s
WNB2	White noise burst with $f_0 = 550$Hz, $\Delta f = 900$Hz, $\tau = 0.010$s
WNB3	White noise burst with $f_0 = 295.5$Hz, $\Delta f = 123$Hz, $\tau = 9.400$s
Cosmic String	
CUSP	Cusp with power law q=-4/3, $f_{high-cut} = 300$Hz
KINK	Kink with power law q=-5/3, $f_{high-cut} = 300$Hz
KINKKINK	Double kink collision with power law q=-2, $f_{high-cut} = 300$Hz
Core-Collapse Supernova	
35OC	$35M_\odot$ sub-solar metallicity with magnetorotational collapse [135]
m20	$20M_\odot$ solar-metallicity non-rotating progenitor [132]
s18	$18M_\odot$ non-rotating red supergiant [131]
Secret SN	$12M_\odot$ non-rotating progenitor
Secret GW event	
GW170817	Binary neutron star merger, m_1=1.36-2.26$M_\odot$, m_2=0.86-1.36$M_\odot$ [101]
Intermediate BBH	
IMBH_1	m_1=50$M_\odot$, m_2=50$M_\odot$, $\tau = 0.7$s
IMBH_2	m_1=75$M_\odot$, m_2=75$M_\odot$, $\tau = 0.9$s
IMBH_3	m_1=100$M_\odot$, m_2=100$M_\odot$, $\tau = 1.8$s
IMBH_4	m_1=125$M_\odot$, m_2=125$M_\odot$, $\tau = 2.1$s
Standard BBH	
BBH_1	m_1=10$M_\odot$, m_2=10$M_\odot$, $\tau = 22.2$s
BBH_2	m_1=25$M_\odot$, m_2=25$M_\odot$, $\tau = 3.0$s
Eccentric BBH	
EBBH_1	m_1=10$M_\odot$, m_2=10$M_\odot$, eccentricity=0.4, $\tau = 2.2$s
EBBH_2	m_1=35$M_\odot$, m_2=35$M_\odot$, eccentricity=0.2, $\tau = 22.2$s
EBBH_4	m_1=100$M_\odot$, m_2=100$M_\odot$, eccentricity=0.75, $\tau = 22.2$s
Accretion Disk Instabilities	
ADI_B	$f_{min} = 110$Hz, $f_{max} = 209$Hz, τ=9.4s, BH=10$M_\odot$, disk=1.5$M_\odot$
ADI_D	$f_{min} = 119$Hz, $f_{max} = 173$Hz, τ=142s, BH=3$M_\odot$, disk=1.5$M_\odot$
Close Hyperbolic Encounters	
CHE_1	m_1=2$M_\odot$, m_2=2$M_\odot$ [147]
CHE_2	m_1=8$M_\odot$, m_2=4$M_\odot$ [147]
CHE_3	m_1=20$M_\odot$, m_2=15$M_\odot$ [147]

Table 4.3: Table detailing the subset of 30 waveform injections for the burst benchmark project considered in the GMM sensitivity search below.

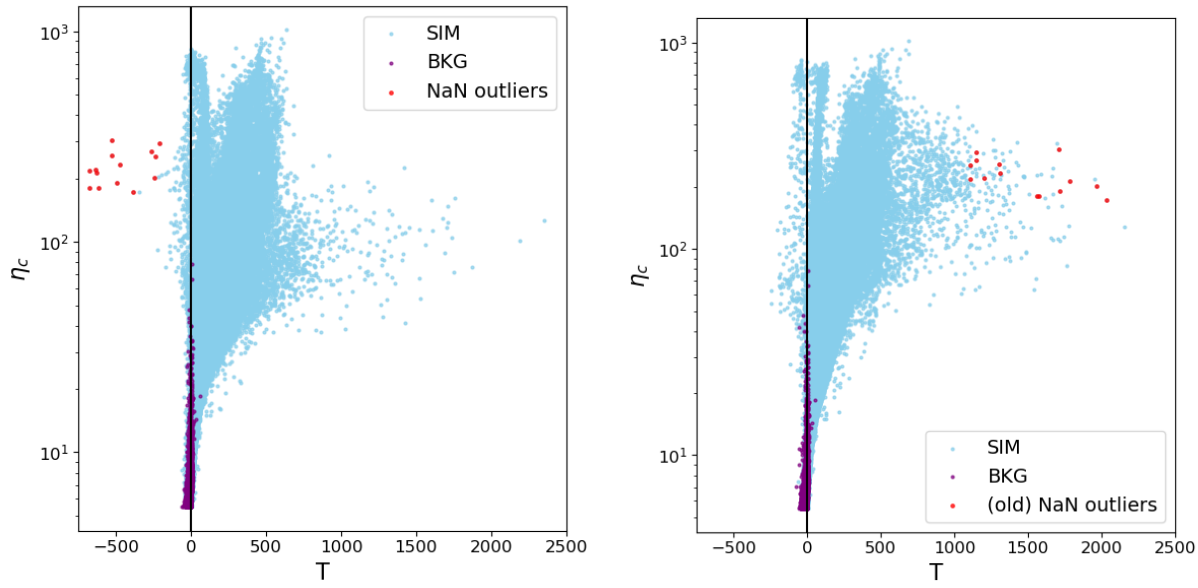
cretion disk instabilities [173, 174]. GW signals from accretion disk instabilities tend to have longer duration, hence are not usually considered in the search for short duration transients. However they are considered for the GMM analysis here as a test of robustness, to observe how the pipeline handles previously unseen types of waveforms. In addition to the above waveforms, two 'secret' waveforms were considered, consisting of the binary neutron star detection of GW170817 [101], and a previously unseen CCSN model.

4.3.2 Running on an updated version of cWB

On the lead up to the Burst Benchmark project, cWB developers made changes to internal thresholds within the production stage of the algorithm to create cWB 'subthreshold'. The idea behind this was that since triggers were now successfully being analysed by Machine Learning post-production methods (either GMM or XGBoost [124]), it was possible to process triggers with lower SNR without sacrificing significance estimates. Internal production thresholds such as η_c (network SNR), network correlation and pixel energy ratio were lowered, and the considered low-frequency search range was altered to be [16, 1280]Hz. While this cWB upgrade brought behavioural changes to the trigger distributions, it also caused a large increase in the number of triggers being processed by the GMM analysis: for example, the background distribution of O3a increased from 10^6 triggers to 10^7. This caused a large increase in computational runtime, the most noticeable being model training time increasing from $\sim$ 24 hours to $\sim$ 1 week. For this reason, minor alterations had to be made to the methodology in order for the analysis to run efficiently.

One of the key changes made was new percentage data splits for the background data. Motivated partially by the small amount of available background data in the benchmark project, and partially by the large increase in triggers causing slow training times, the splits were altered to consider a larger portion of triggers for the test data set and a lower portion for training. It was found that a new split of 5% validation, 25% training and 70% testing allowed for increased computational efficiency while the lower percentage split of training data still represented the overall noise distribution well. This split also reserved a desirable amount of data for FAR estimation of 196.42 years. In addition to reducing the amount of data considered in validation and training, computational upgrades were made to the GMM scripts, including the parallelisation of model training and log-likelihood calculations with HTCONDOR, while total memory consumption of the analysis was reduced. The above changes resulted in a significantly reduced runtime for the analysis, with model training now taking $\sim$ 1-2 hours rather than $\sim$ 1 week.

In order to investigate the behaviour of triggers in the upgraded cWB-subthreshold, the benchmark background and simulation test data containing the blind injections were processed through the GMM methodology considering original re-parameterised attributes. The background model was trained on 25% of background data estimated with the 40-day benchmark period, while the signal model was trained on the WNB1+WNB2 simulation data as detailed in



(a) old re-parameterisation of $log10(N_{ED}+1000)$ (b) new re-parameterisation of $log10(N_{ED}+2000)$

Figure 4.9: T vs network SNR η_c for background (purple) and simulation (blue) triggers from the burst benchmark analysis, with outlier NaN simulation triggers highlighted in red. (a) shows the distribution of triggers for original re-parameterisation $log10(N_{ED}+1000)$, which was causing the NaN values. (b) shows the distribution of triggers with new re-parameterisation $log10(N_{ED}+2000)$, with previous outlier NaN triggers now in high T.

Section 3.2.2. The distributions of T vs. network SNR (η_c) were observed, seen in Figure 4.9a. While looking at this plot, it was noticed that some simulation test triggers were being classified with very low T values despite having high η_c, all clustered within an outlier region compared to both background and signal distribution. Investigating these outliers, it was found that some triggers have N_{ED} < -1000, leading to NaN values during re-parameterisation. This was a feature of the benchmark data itself rather than cWB subthreshold, however in order to fix this the re-parameterisation of N_{ED} was changed from $\log_{10}(N_{ED}+1000)$ to $\log_{10}(N_{ED}+2000)$. The behaviour in T vs η_c with the new-reparameterisation is seen in Figure 4.9b, with the previous outlier triggers now lying in the high T region. Some simulation injections with high η_c are still classified with low T, lying in the region below $T=0$. However these do not appear to be outliers from the distribution and are likely waveform injections which are outside the GMM targeted parameter space, such as long duration or high-frequency waveforms.

Re-running the training and validation process with the above analysis alterations, the optimal models for the O3b benchmark period are 106 Gaussian components for the signal model and 118 Gaussian components for the background model.

4.3.3 Results

The sensitivity of pipelines to the various populations in Table 4.3 is explored by observing detection efficiency vs. injected amplitude, which in most cases is h_{rss}. In total 8 pipelines participated in the benchmark study, all targeting different regions of burst parameter space. However for the purpose of this thesis we only consider other cWB pipelines in the comparison with GMM, since the methodologies all target the same parameter region hence serve as a better benchmark of the GMM analysis. We introduce an alternative method of cWB trigger production used by another pipeline, *cWBXP*, which uses a wavescan time-frequency transformation [175] rather than Wilson-Daubechies-Meyer wavelet based transformation, and a combination of excess-power and cross-power statistics to identify triggers from incoming data [176]. With this alternative methodology, there now exists two approaches of cWB production (2G-subthreshold and XP) and two machine learning based post-productions (GMM and XGBoost). Through combinations of these alternative methodologies a total of 3 cWB-based pipelines partook in the sensitivity studies, namely:

- cWB2G-GMM: trigger production based upon details in section 2.2 (ie. cWB2G) with the implementation of the subthreshold upgrade described above. Post-production by the GMM analysis. This is the analysis developed and used for work in this thesis.
- cWB2G-XGB: Also uses the cWB2G-subthreshold approach for trigger production, however applies the boosted decision tree method of XGBoost as post-production [124].
- cWBXP-XGB: Uses the alternative approach of combined excess- and cross-power for trigger production, however also uses the the XGBoost algorithm as post-production.

The sensitivities for these pipelines are compared in the below plots for injections considered in the burst benchmark project, although it should be emphasised that only the cWB2G-GMM search was run as work considered in this thesis.

The results of the sensitivity study are seen in Figures 4.10 and 4.11 in the form of detection efficiency vs. h_{rss} at iFAR threshold $\geq$ 10 years, with comparisons to the other cWB pipelines. First we will look at Figure 4.10 which shows results for the generic ad-hoc, cosmic string and supernovae injections. The GMM analysis obtains comparable sensitivities to the other cWB methods for all ad-hoc injections. Performance on SG and WNB waveforms is comparable to the other pipelines, apart from the higher frequency *WNB2* injection which has slightly lower sensitivity. Sensitivity to the GA waveform is good for low h_{rss}, however an effect known as gating is seen as h_{rss} increases, which is due to cWB disregarding very loud signals which disguise as glitches. This behaviour isn't worrisome and is also seen for other pipelines. The GMM analysis again has good sensitivity for cosmic string waveforms, similarly to what was seen in the O3 studies in 4.1. From these O3 studies, GMM previously had the highest sensitivities to this population, however the XGBoost analysis now matches these, likely due to the upgrade

of trigger production in cWB-subthreshold. Lastly, GMM achieves comparable sensitivity to the majority of CCSN models considered, obtaining only slightly lower sensitivity to the *m20* waveform, which is consistent with previous O3 results.

Figure 4.11 contains similar sensitivity plots for various compact binary coalescence type populations, including close hyperbolic encounters and long duration signals from accretion disk instabilities. At a glance, GMM again performs well for all these waveforms, performing to similar levels as the other cWB pipelines. The GMM analysis appears to perform slightly better for high mass binaries, such as *IMBH_2, IMBH_3, IMBH_4*, while has a very minor decrease in sensitivity for *IMBH_1, BBH_2*, and a couple of *EBBH* waveforms, which have mass components $< 50M_{\odot}$. Of notable interest are the sensitivities to accretion disk instability waveforms, where the GMM analysis has highest sensitivity out of the considered pipelines. These are longer duration waveforms which are not considered to be in the training or targeted search space of the analysis. In particular the *ADI_D* waveform has a duration of 142s, however the GMM analysis outperforms the other pipelines by a significant amount. This result demonstrates the robustness and un-modelled nature of the method, with the ability to detect a wide range of GW signals.

Overall, the burst benchmark project proves the ability of the cWB+GMM analysis to detect a wide range of astrophysical sources across the burst parameter space, even proving the robustness of the methodology through good sensitivity to sources outwith the targeted search parameters such as CBC and accretion disk instability signals. Additionally, the performance of GMM in this study demonstrates the potential for the analysis to analyse data in future observing runs. Since it is capable of achieving similar or better sensitivities to the other cWB post-productions, it can be used to cross-check any burst detections that may occur or reinforce upper-limit rate estimates on astrophysical sources.

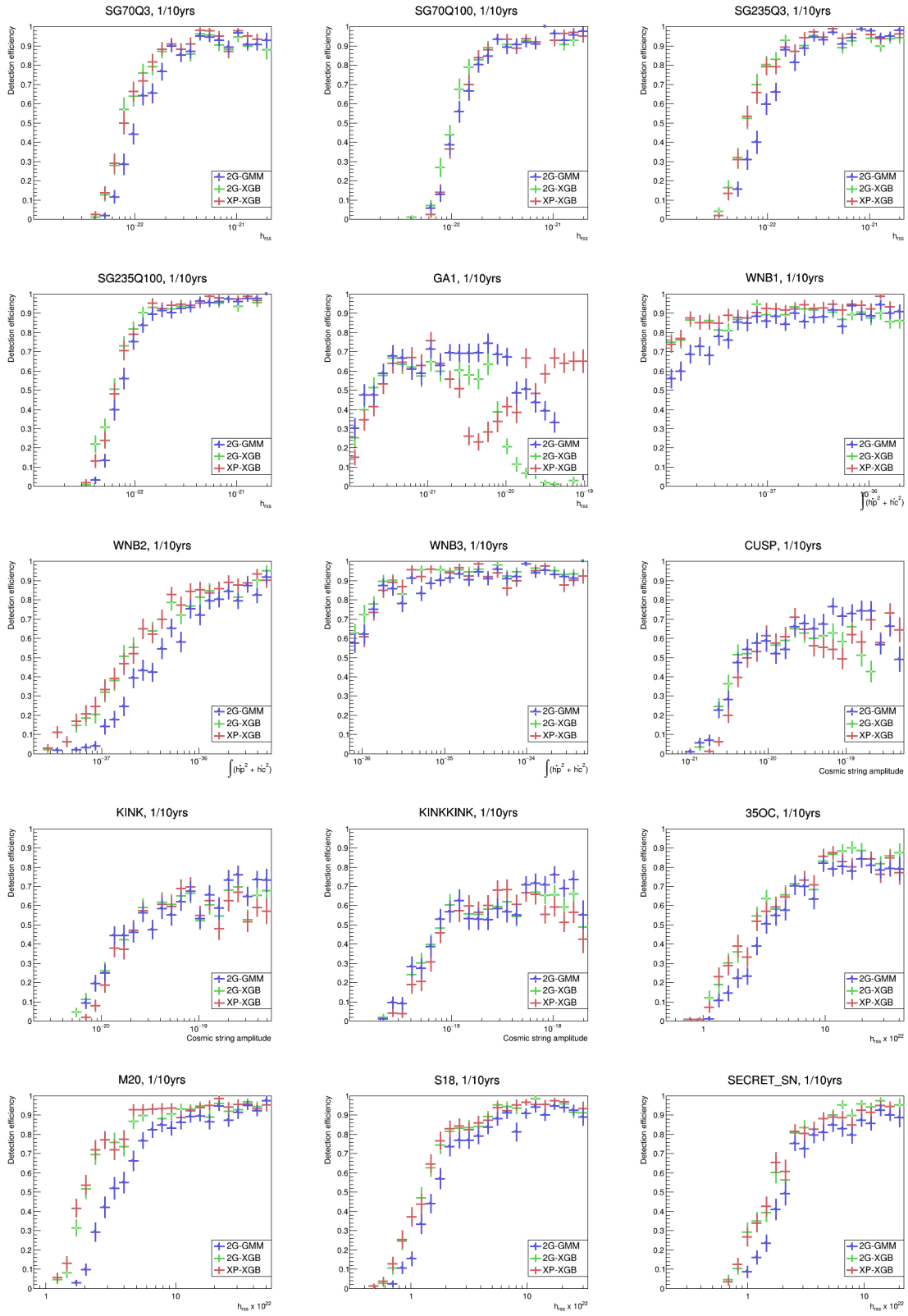


Figure 4.10: Detection efficiency vs. h_{rss} for various waveform injections in the burst benchmark project. The figure shows sensitivities for generic ad-hoc burst waveforms of sine-Gaussian (SG), Gaussian pulse (GA), white noise burst (WNB), cosmic strings and core-collapse supernovae (CCSN) for three variations of cWB; cWB2G-GMM (blue), cWB2G-XGB (green) and cWBXP-XGB (red).

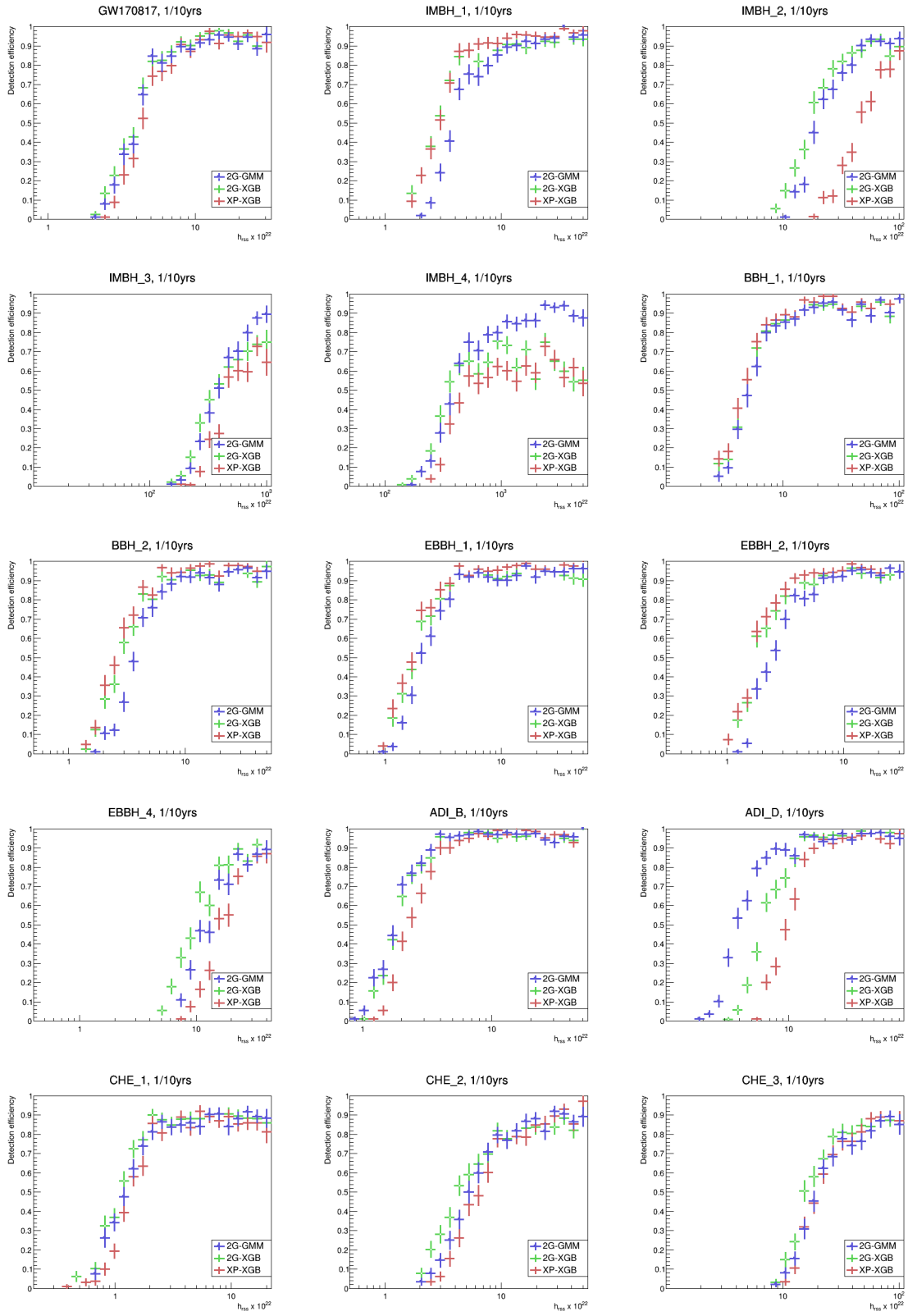


Figure 4.11: Detection efficiency vs. h_{rss} for various waveform injections in the burst benchmark project. The figure shows sensitivities for GW170817, 'vanilla' binary black hole (BBH), intermediate-mass black holes (IMBH), eccentric binary black holes (eBBH), close hyperbolic encounters (CHE) and accretion disk instabilities (ADI), for three variations of cWB; cWB2G-GMM (blue), cWB2G-XGB (green) and cWBXP-XGB (red).

4.4 Summary and discussion

This Chapter has detailed the application of the newly enhanced cWB+GMM algorithm to the search for short duration GW transients across data from the third LVK observing run. The ability of GMM to successfully down-rank blip glitches and perform as an effective search method for short GW bursts has been demonstrated through the application to O3 with the 2-detector LH network, the first application to the 3-detector LHV network, and participation in the blind MDC burst benchmarking project.

The application of the cWB+GMM pipeline to the *full* O3 observing run presented an opportunity to observe how the methodology performed when completing a full search for GWs, including sensitivity estimates to simulated sources and the opening of boxes to observe real GW detections. Through comparisons to other cWB post-production methodologies, it was demonstrated that the GMM analysis was capable of achieving similar sensitivities to a range of predicted burst sources, and in some cases obtained sensitivity improvements. In particular, the GMM-based search achieved the best sensitivity to GA and CS waveforms, which have previously been sources the most effected by blip glitches. The results here demonstrated the ability of GMM to effectively mitigate blip glitches, thus improving significance estimates. Analysing the zerolag data, the cWB+GMM search observed a total of 14 significant GW detections during O3, similarly to the original published search [85]. All of these detections were known CBC events, and no new significant detections were found. With the comparison of final search results in Figure 4.6, it was seen that the GMM-based search detected CBC events with lower significance than the other considered pipelines. While this can be considered as an area of improvement for the analysis, it is not a detrimental results for GMM since CBC events have not been our targeted source. Especially since our analysis appears to have the best sensitivity to GA and CS waveforms, which have a different morphology to standard CBC waveforms. If desired, it may be possible to develop a GMM analysis targeted for the detection of CBC systems by including such simulations in training data, however this should not be considered for minimally modelled searches.

The GMM analysis was extended to the application on a 3-detector network for the first time, which involved some minor alterations to the considered cWB attributes. The results achieved with this analysis were particularly important as in previous searches cWB observed lessened sensitivities for the 3-detector network due to the softening of regulators, which decrease the rate at which cWB can disregard glitches based upon antenna patterns. This is especially true for GA and CS waveforms due to their similarity to blip glitches. However, with the addition of GMM post-production it was seen that sensitivities agree with, or improve upon, those seen with the 2-detector network. Sensitivity with the 3-detector network is especially high for CS waveforms, which exemplifies the ability of GMM to mitigate blip-like glitches further. These results may favour the use of a 3-detector network in terms of sensitivity in future observing runs, which would bring additional benefits such as increased duty cycle and event localisa-

tion. Since this stands as an initial proof-of-principal application to the 3-detector network, the analysis has not been fully tuned. With further investigations into optimal attribute choice and re-parameterisations for the 3-detector network, it is possible for sensitivities to be improved even further.

Finally, details of the cWB+GMM participation in the Burst Benchmark project were given, in which burst pipelines analysed a blind MDC dataset to compare performance to a wide range of astrophysical sources. This project truly tested the robustness of the GMM pipeline since the majority of these waveforms had not been previously considered, and allowed for direct comparison with all other burst pipelines. In general, the GMM analysis performed at a similar sensitivity to the other cWB-based pipelines for the majority of sources, proving it to be competitive with other techniques. A particularly interesting result to come of this work was that GMM achieved good sensitivity to long duration signals arising from accretion disk instabilities. These signals were not targeted by the analysis, and in fact lie outwith the considered duration of the search. The sensitivity achieved to such waveforms without targeting their parameter space suggests that the application of GMM may be advantageous in the search for long duration burst signals with future investigations into tuning the analysis.

In this Chapter the application of GMM to the upgraded version of cWB (cWB subthreshold) was also introduced. While adapting the GMM package to process data from this version, the structure of the GMM scripts were altered to improve the computational efficiency. In particular, processes such as model training and log-likelihood calculation were parallelised with HTCONDOR, and the running of the analysis was automated. These structural upgrades reduced the total computational time and memory consumption of the analysis significantly, resulting in a more suitable structure for participating in offline searches for future LVK observing runs.

To conclude, the performance of the cWB+GMM approach was tested extensively throughout this Chapter. It's ability to mitigate blip glitches has been shown multiple times through the application to both 2- and 3- detector networks, with the analysis achieving high sensitivities to a wide range of expected astrophysical sources. In addition, it was demonstrated that the GMM-based pipeline is able to accomplish competitive performance levels with other all-sky short pipelines. With the above considerations, along with the improvement in computational efficiency, the cWB+GMM pipeline may be taken into consideration as an official pipeline in the search for short duration GW candidates in future observing runs, for example in O4.

Chapter 5

cWB+GMM in the search for short GW transients in the fourth observing run

This Chapter details the application of the cWB+GMM analysis to the offline search for all-sky short duration transients during the first-half of the LIGO-Virgo-KAGRA fourth observing run. The fourth observing run (O4) is split into 3 epochs in order to have mid-run detector commissioning breaks, with the first epoch (O4a) running from 24th May 2023 to 16th January 2024, and the remainder of the run epochs (O4b and O4c) expected to occur from 10th April 2024 until 9th June 2025. Here we consider the analysis during O4a, in which only the LIGO-Livingston, LIGO-Hanford (LH) detectors were in observing mode, accumulating 123.05 days of coincident data following the removal of periods with poor data quality. The search for short duration transients within this data was of particular interest, as upgrades to the detectors and data analysis techniques prior to O4 improved sensitivities to GW sources, thus increasing the probability of detecting a burst-type transient compared to previous observing runs.

The cWB+GMM analysis on O4a data took place while the run was ongoing, resulting in a different data processing approach than considered previously: rather than waiting until the end of the run to analyse all data simultaneously, data was processed in smaller periods of time referred to as *chunks*. In total, the 8-month long observing run was split into chunks of approximately 2 weeks, resulting in a total of 16 chunks labelled as 'K*N*' for *N* of 1-16. This allowed for investigations into glitch mitigation and optimal search configuration to occur throughout the run, and ultimately for finalised results to be delivered in a timely manner. In this set-up, the background is estimated chunk-by-chunk, so that each chunk accumulated approximately 400-800 years of background depending on the respective coincident livetime. cWB-subthreshold is utilised for trigger production before GMM post-production is applied. As introduced in section 4.3, this version of cWB has lowered attribute thresholds during trigger production allowing for the generation of quieter triggers, and is referred to as 'cWB2G'. Since the cWB2G+XGBoost analysis also ran on O4a data, trigger production was shared between the two pipelines, reducing computational cost. The cWB+GMM analysis again only considered the low-frequency search

for short transients, however for O4a this involved an increased frequency range of [16-2048]Hz.

Full details of the GMM search during O4a are given, starting with details of investigations into the optimal search configuration, including the approach to chunk-by-chunk analysis and studies on possible updates to the GMM methodology in Section 5.1. Examples of glitch investigations occurring throughout the run are given in Section 5.2, with details on the application of CAT2 vetoes and a summary of the loudest glitches for the GMM analysis. Section 5.3 describes inspection of loud outlier glitches present in K11 & K12 caused by a problem within the GMM model. Subsequent to investigations on search configuration and the mitigation of loud glitches, finalised sensitivity estimates to generic ad-hoc and CCSN waveforms throughout O4a are detailed in Section 5.4, while Section 5.5 presents full results on the cWB+GMM search for short duration GW transients, including details of the highest significance non-CBC trigger and CBC detections made by the pipeline.

5.1 Search configuration investigations

As mentioned above, data from O4a was analysed in chunks in order for glitch investigations and search optimisation to take place throughout the run, allowing for results to be finalised shortly after the end of the run. While this may increase computational cost of the GMM analysis overall due to the training of multiple sets of models, it allows for the possibility of more accurate models since considered background data is produced only on the relevant chunks. This approach to analysis required investigations into the configuration and methodology of GMM to ensure it was still performing optimally. Specifically, we investigate whether it is optimal to process data and train models through individual or combinations of chunks, and whether models should be trained on portions of the same data or surrounding data. Once this approach is decided, details of further optimisation methods on GMM are discussed. For all investigations in this section we consider the same set of ad-hoc waveforms as previously considered in Table 3.3 injected into the specified O4a chunks.

5.1.1 Analysing combined chunks

The first study to be conducted concerns whether chunks should be analysed on a completely individual basis or over a combination of chunks at one time. To investigate this we define two approaches to analysis:

- *method0*: each chunk in the run is analysed individually. Data from the individual chunk is split into validation, training and testing sets resulting in models which are trained and applied to each respective chunk. This is labelled method0 since it is the default approach following the methodology outlined in Chapter 3.

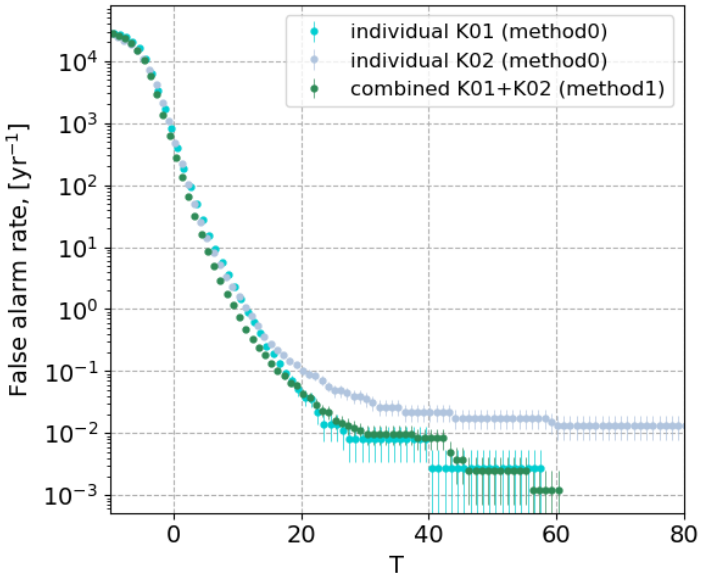
- *method1*: combine two adjacent chunks to be analysed in unison. Data from each individual chunk is split into validation, training and testing sets, then is combined with respective sets from an adjacent chunk. Models are trained over data from both chunks and applied to test data over the combination of chunks.

To explore which of these methods is optimal we apply both to the first 2 chunks of O4a, K01 & K02, comparing the T vs. FAR thresholds and h_{rss50} estimates for the ad-hoc set of waveform injections, seen in Figure 5.1. For method0, we analyse K01 and K02 separately, combining sensitivity estimates across both chunks for a direct comparison with method1. Looking first at the FAR vs. T comparisons in Fig 5.1a, we see that for method0, K02 has a long tail caused by glitches being classified with high T, while K01 is affected less by loud glitches thus achieving better T thresholds at low FAR. Comparing this to the result for method1, we see that conducting the analysis over a combination of both chunks appears to mitigate the effect of loud glitches seen in K02, with the distribution for method1 agreeing more closely with the distribution for method0-K01. This result is reinforced through the h_{rss50} comparisons in Fig 5.1b. The method0-K02 h_{rss50} estimates are significantly higher than accomplished in method0-K01, again due to the loud glitches present in the method0-K02 analysis. This is reflected in the method0 sensitivity estimates combined over both chunks, while the results for combined chunks with method1 show sensitivities which are much lower, often comparable with those of method0 - K01. While method1 occasionally increases h_{rss50} for K01+K02 compared to what would be achieved if K01 was analysed alone, the sensitivity increase achieved for K02 outweighs any disadvantage this may cause. From this result we can again conclude that analysing a combination of two chunks appears to mitigate the effect of loud glitches seen in individual chunks, making the models more robust against glitches and thus increasing the overall sensitivity. Additionally, method1 brings other benefits to the analysis, such as reducing computational time since less models need to be trained throughout the run, and increases the amount of background test data available for FAR estimation.

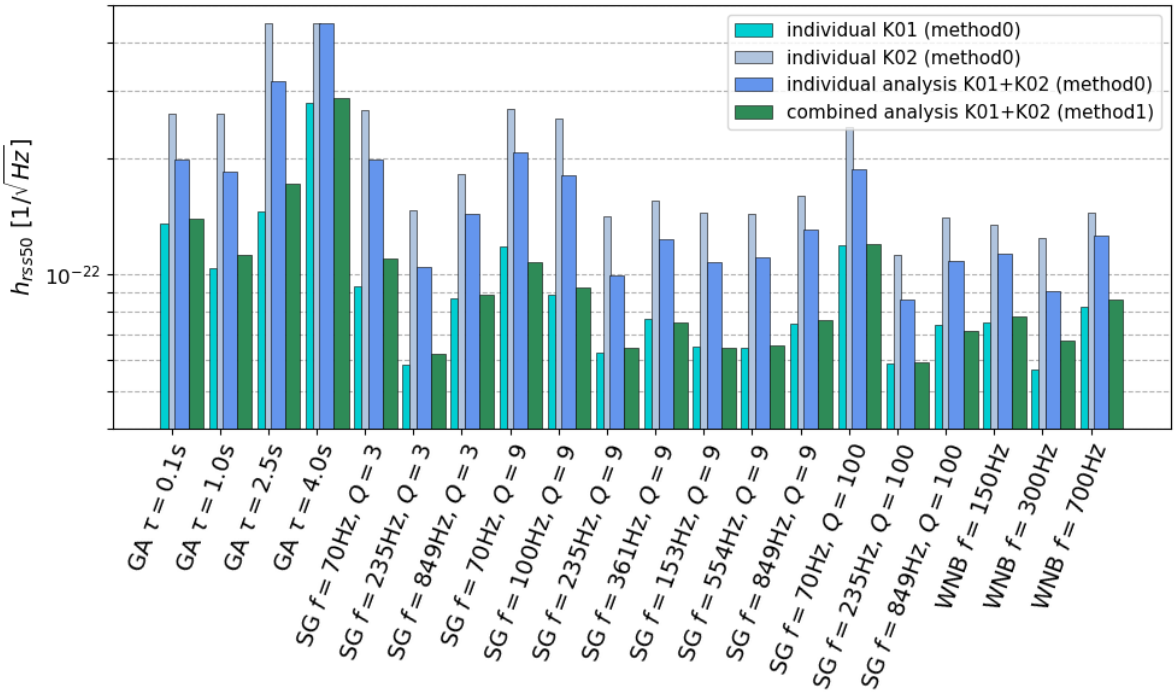
It's likely that the combined chunk models are able to better mitigate the effect of loud glitches because the background models are trained on a period of time of approximately 1 month, which is likely to include a wider variety of glitches than occurring in a single chunk, while still including noise relevant to the period of data being considered. The increased amount of available data for FAR estimation will also help, as it requires a higher number of loud glitches to significantly affect the T thresholds at iFAR of 100 years. The following investigations are all performed adopting the combined chunk analysis approach.

Combined chunk model on the same or adjacent chunk

We execute a further study on analysing combined chunks to investigate the effect of applying a model trained on a portion of data from the same combination of chunks as being analysed vs. a model trained on a portion of data from a combination of adjacent chunks. The application of a

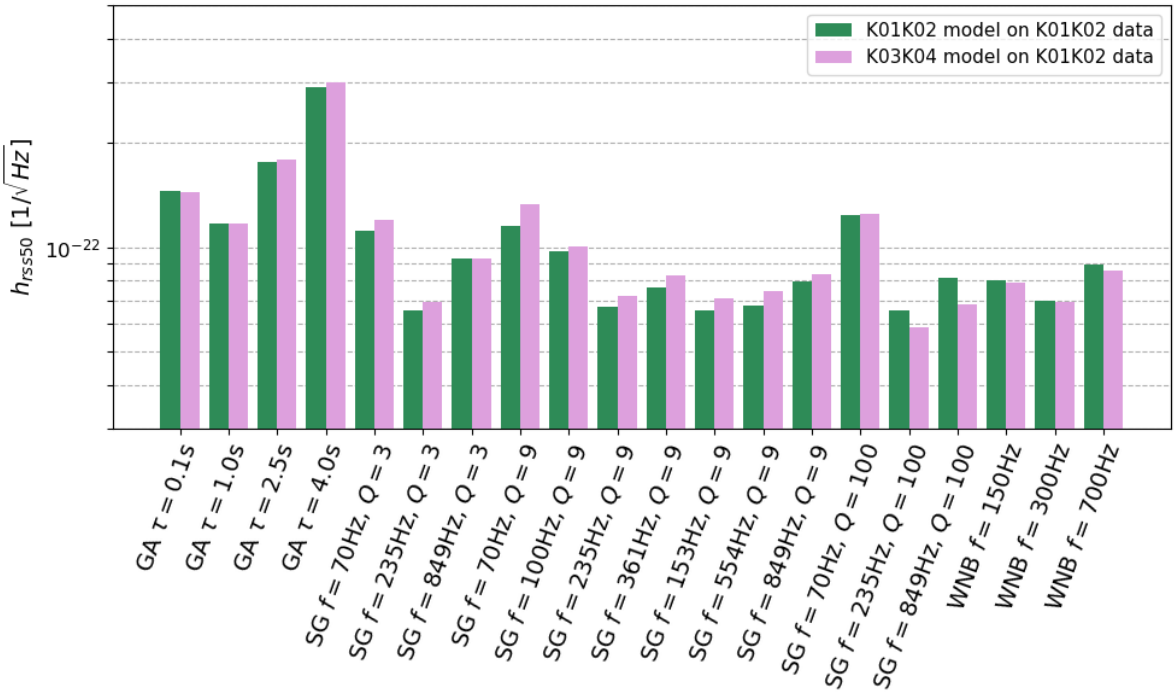


(a) FAR vs. T

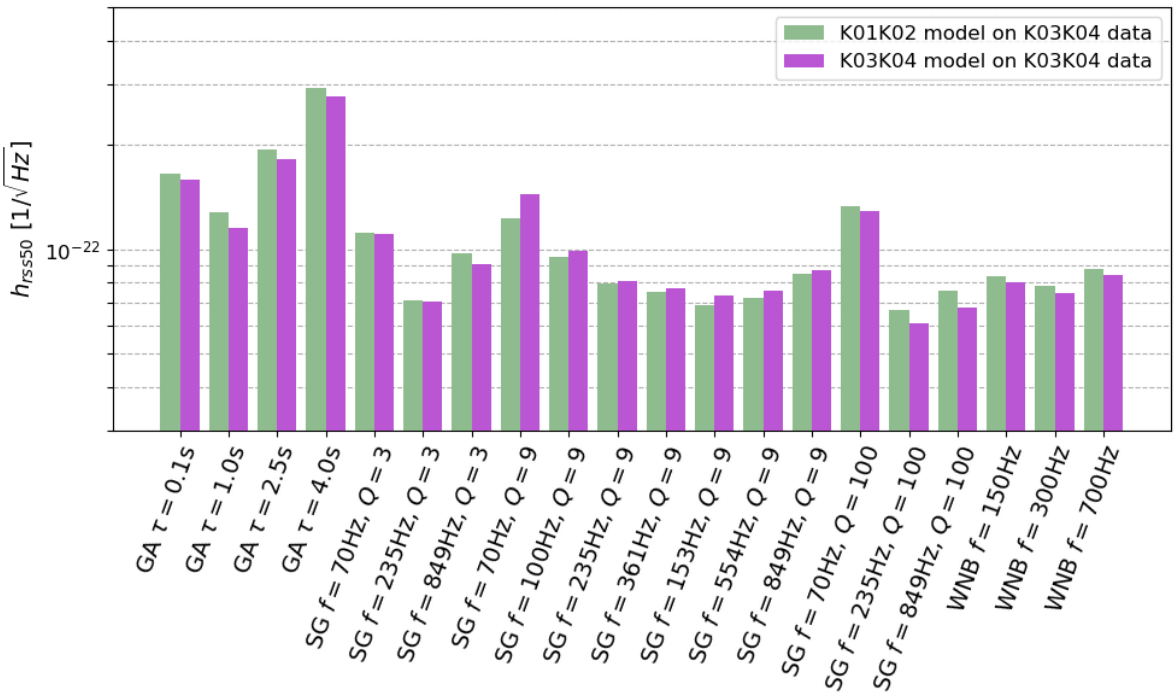


(b) h_{rss50} for ad-hoc injections

Figure 5.1: Results of the study on individual (*method0*) vs. combined chunk analysis (*method1*). (a) shows FAR vs T distribution for individual K01, individual K02, and combined analysis K01+K02. (b) shows the h_{rss50} estimates for ad-hoc waveform injections at iFAR $\geq$ 100 years for individual K01, individual K02, combined estimates of K01&K02 with method0, and combined estimates for K01+K02 with method1.



(a) Analysis on K01K02 data



(b) Analysis on K03K04 data

Figure 5.2: Comparison of h_{rss50} estimates to ad-hoc waveforms at iFAR $\geq$ 100 years for models trained on data from the same chunks vs. data from adjacent chunks. (a) shows results for K01K02 test data with K01K02 vs. K03K04 models. (b) shows results for K03K04 test data with K03K04 vs. K01K02 models.

model trained on a different chunk combination would allow for the full estimated background of the analysed chunks to be utilised in FAR estimation rather than split into training and validation sets. However we expect models trained on data from the same chunks to perform better, since they will be trained on the same types of glitches as occurring in the test data.

Sensitivity comparisons to the injected set of ad-hoc waveforms at an iFAR $\geq$ 100 years are seen in Figure 5.2, with Fig 5.2a, 5.2b showing results on K01+K02, K03+K04 data respectively for analysis with a model trained on a portion of K01K02 data vs. a model trained on a portion of K03K04 data. As expected, results on K01+K02 show slightly better performance for the K01K02 model, apart from on WNB waveforms. Somewhat similarly for K03+K04 results, the K03+K04 model performs better for the majority of waveforms other than SG $Q = 9$. Overall the difference in sensitivities between models is minimal, however since the results show slightly better performance for models trained on a portion of the same chunks, which is theoretically more robust to glitches, we continue to utilise this approach.

5.1.2 Studies to optimise GMM in O4

With the analysis approach of combining adjacent chunks determined, we go on to investigate potential updates to the GMM methodology to ensure the analysis is performing optimally. As part of these investigations, we aim to increase the distinguishability between noise and signal populations through the exploration of an updated signal training data set and a new re-parameterisation of the N_{ED} attribute. We also attempt to increase the signal sensitivity of the analysis through the application of secondary GMM models in two cases, firstly trained on triggers with low significance from the primary GMM models and also trained on triggers from the low-Q parameter space. While these adaptations to the analysis were explored, time constraints were present due to continuously incoming data, thus in-depth investigations to each adaptation were not always possible.

Updated signal training

First we explore the use of an updated signal training set to account for the increased frequency range considered by the search. The original training data introduced for the O3 search consisted of a combination of two sets of WNB simulations: *WNB1*, in which waveforms were uniformly distributed in central frequency [24,996]Hz and bandwidth [10,300]Hz, and logarithmically distributed in duration [0.1,500]ms. *WNB2*, which consisted of waveforms with bandwidth 10Hz, frequency 0Hz, and duration randomly distributed over [0.1,1]ms, as discussed in Section 3.2.2. Here we consider two new data sets consisting of WNB injections, labelled as *training set1 (TS1)* and *training set2 (TS2)*. Both of these data sets are distributed over frequency [24,1696]Hz, bandwidth [10,800]Hz and duration [0.001,0.5]s, however TS2 is more densely sampled in the Q_{veto} parameter space. This is shown in Figure 5.3, where TS2 better represents the parameter

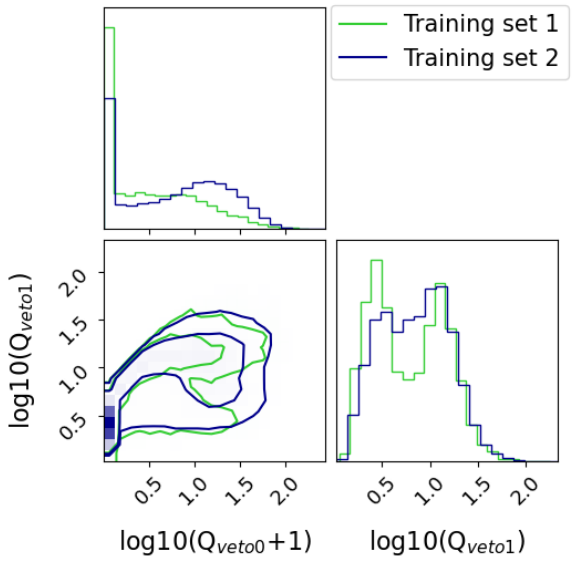
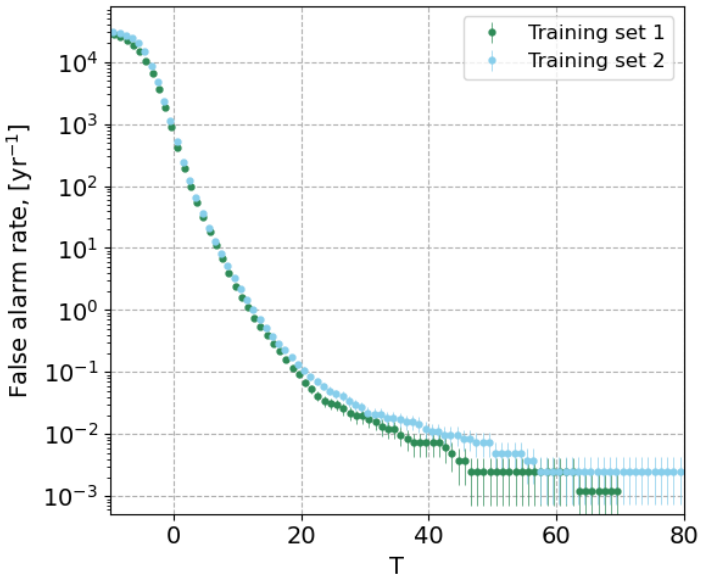


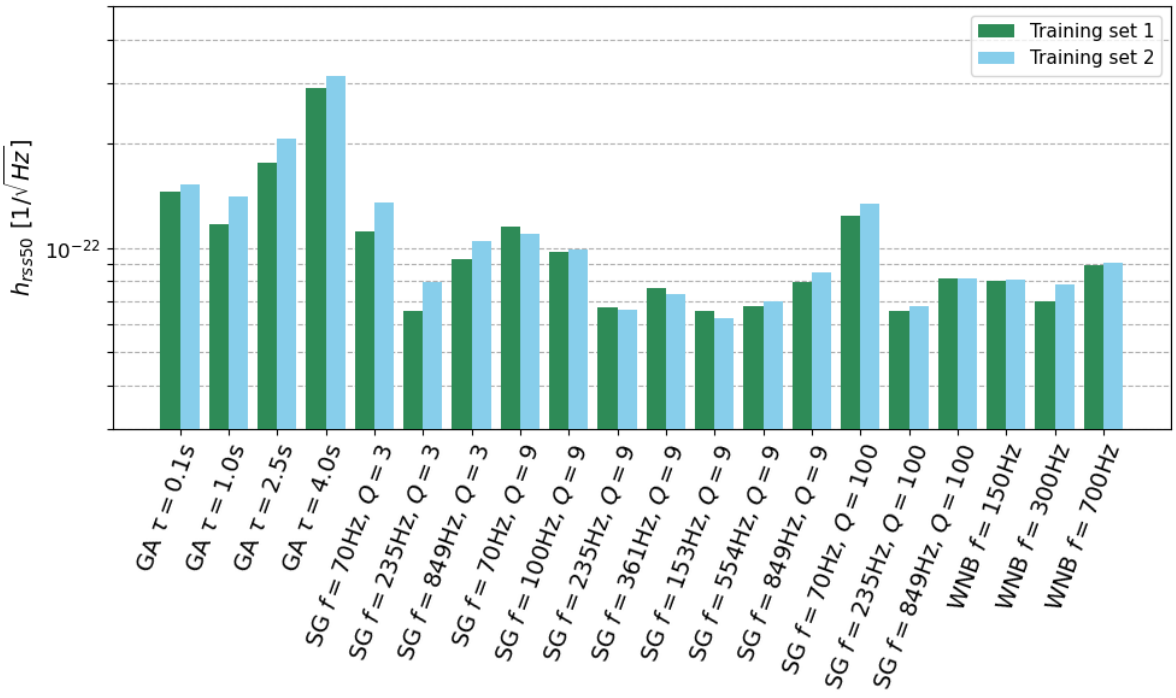
Figure 5.3: Comparison of signal training data sets 1 & 2 in the re-parameterised Q_{veto} parameter space. Training set 2 more densely samples the parameter space, covering more representation of the $log10(Q_{veto0}+1) \approx [1,2]$ and $log10(Q_{veto1}) \approx [0.5,1]$ region.

space between $log10(Q_{veto0}+1) \approx [1,2]$ and $log10(Q_{veto1}) \approx [0.5,1]$, while TS1 does not. This region is not often populated by glitches, however covers a region that may be occupied by a select SG $Q = 100$, WNB and CCSN signals, meaning the use of training set2 should aid in the ability to distinguish signals and glitches.

We investigate the performance of the GMM models trained on each of the training sets by again analysing the set of ad-hoc waveform injections into K01+K02 data, with results shown in Figure 5.4 at iFAR $\geq$ 100 years. Looking at the comparison of FAR vs. T distributions, both training sets are overall equivalent until FAR $<$ 1/10 years, with TS1 reaching slightly lower T for FAR 1/100 years. Despite this, the h_{rss50} estimates at this FAR threshold in Fig 5.1b, are comparable for the majority of considered waveforms, with TS1 showing slightly better results for low-Q waveforms. Despite TS1 showing slightly better sensitivities for the majority of waveforms, we choose training set 2 as data for the signal model in O4a since the difference in sensitivity is minimal and TS2 is theoretically more robust against glitches for a range of signal waveforms. This choice is further reinforced by the benefit to computational cost, as the the XGBoost post-production also selects TS2 for their analysis, meaning cWB trigger production can be shared between the pipelines.



(a) FAR vs. T



(b) h_{rss50} for ad-hoc injections

Figure 5.4: Results of analysis with signal training data set 1 (TS1) (green) vs. set 2 (TS2) (blue) for K01+K02. (a) shows a comparison of FAR vs. T for both training sets. (b) shows the h_{rss50} estimates of both training sets to ad-hoc waveform injections at iFAR $\geq$ 100 years.

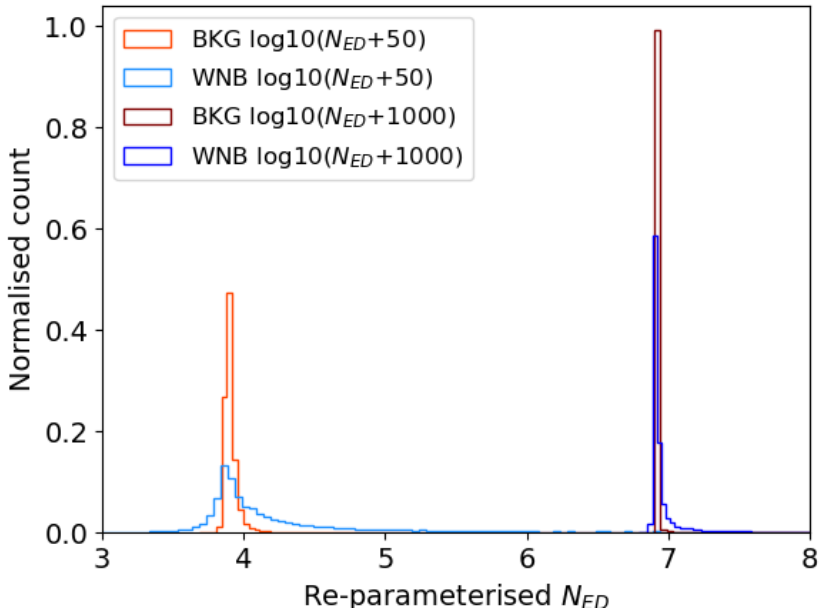


Figure 5.5: Histogram showing the updated N_{ED} re-parameterisation of $\log 10(N_{ED}+50)$ vs. the old re-parameterisation $\log 10(N_{ED}+1000)$ for background (red) and WNB signal (blue) data.

New N_{ED} re-parameterisation

From looking at the distribution of the $\log 10(N_{ED}+1000)$ attribute in Figure 3.2, it was noticed that the chosen re-parameterisation causes a sharp narrow peak in the background data, which overlapped with the peak seen in WNB signal data. While the populations appear to be distinguishable when grouped with other attributes, it was thought that a new re-parameterisation of N_{ED} could aid in the distinguishability of noise and signal further, while lessening the sharp peak in the distribution may promote better Gaussian behaviour. Since a large majority of triggers lie within $N_{ED} > -50$, we apply a test-case updated re-parameterisation of $\log 10(N_{ED}+50)$, with triggers below the -50 threshold manually set to $N_{ED} = -49$. This updated re-parameterisation is seen in Figure 5.5 for background and WNB signal data, with the old re-parameterisation present for comparison. As expected, the updated re-parameterisation removes the sharp peak from the distribution, and appears more distinguishable between background and signal at least in the 1-dimensional space.

Results of the analysis with the new N_{ED} re-parameterisation are seen in Figure 5.6 alongside original TS2 results and other investigations conducted below. Comparing only the new N_{ED} and original TS2 results, it is seen that this new re-parameterisation causes a longer tail in the FAR vs. T distribution due to glitches being classified with higher T. This is also reflected in the h_{rss50} estimates in Fig 5.6b, with sensitivities to all ad-hoc waveforms being worse for the new re-parameterisation. It is thought this decrease in performance could be caused by the removal of the sharp peak in the background distribution, which was the feature thought

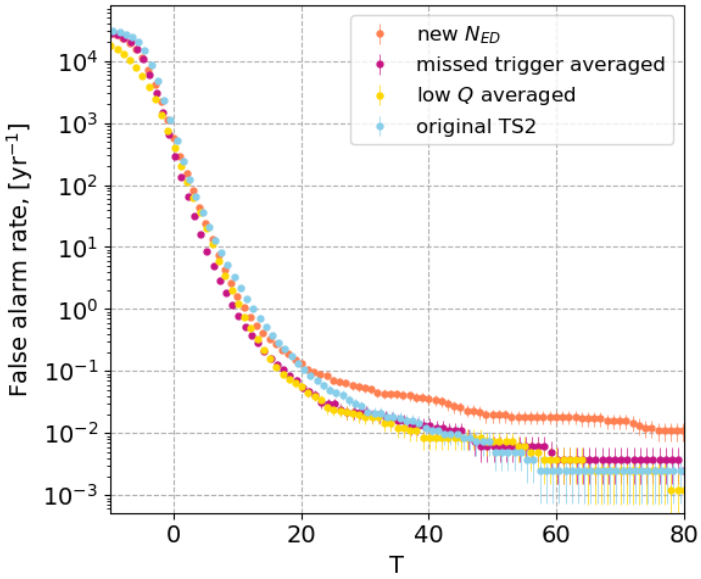
to be originally hindering performance: Since such a narrow peak exists for the original re-parameterisation, especially in background, glitches will be strongly concentrated in one region of the model, aiding in the classification of glitches and signals rather than hindering it. It is difficult to make a concrete conclusion on this behaviour while attribute relationships are only viewed in 2-dimensions.

Secondary models on missed triggers

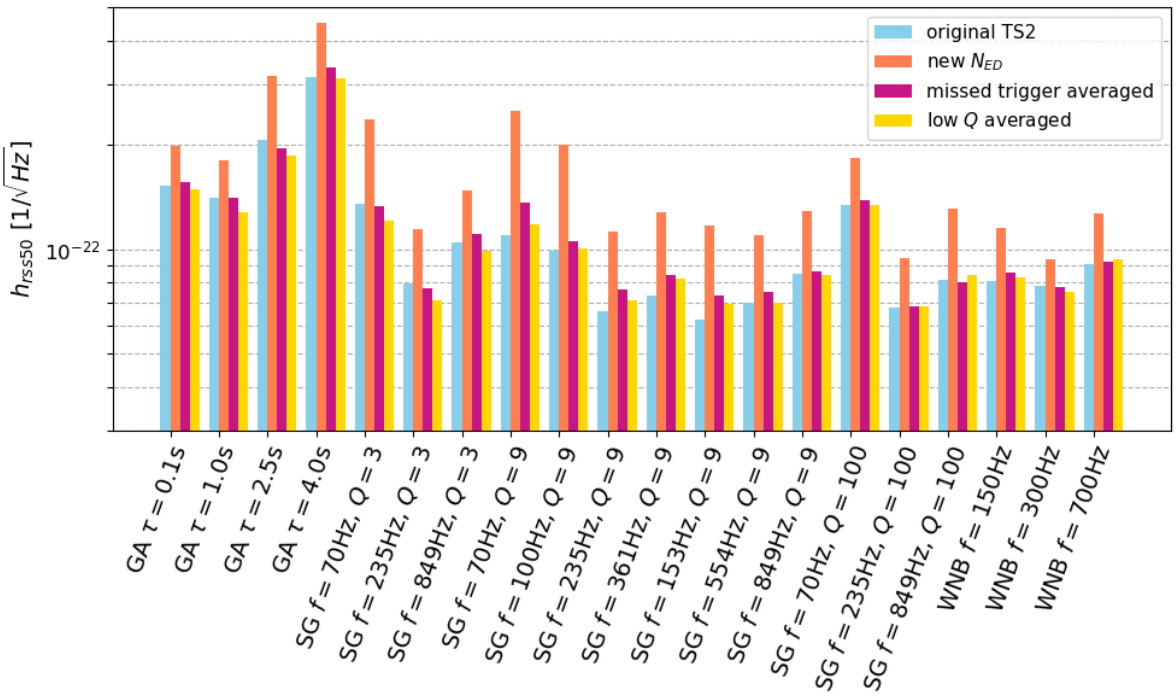
Here we introduce a new approach, in which the signal population is represented by two GMM models instead of one. Both signal and background primary models stay the same as above, being trained on WNB TS2 and a portion of background data, however the secondary signal model is trained on some sub-population of signal training data. Test data is processed by both models, and a geometric mean is taken for each trigger over the log-likelihood values from both signal models. The aim of this is to boost the log-likelihood values of triggers occurring in less distinguishable regions of the signal and background models through the reinforcement of the secondary model.

We first test this new approach by training a secondary GMM model on triggers within the signal training data which are classified with significance under some iFAR threshold after the application of the primary models (ie. 'missed' by the primary models). To do this, we calculate log-likelihood estimates on triggers in the signal training data being in the primary models, allowing for T and iFAR estimates to be calculated on the training data itself. A secondary signal training set is then formed from the original training set by taking triggers with iFAR $\leq$ 100 years, which should increase the significance estimates of triggers in this region when the mean is taken over both signal models. We label this approach as the 'missed trigger average' analysis since we average the log-likelihood of triggers in the test data over the original GMM signal model and the secondary signal model trained on 'missed' triggers from the training data. Results are seen in Figure 5.6. Comparing the results of the missed trigger average analysis to the original TS2 analysis, the averaged models appear to improve the FAR vs. T distribution at high FAR, however brings no improvement at FAR of 1/100 years. This is reinforced in Fig 5.6b, where h_{rss50} estimates at iFAR $\geq$ 100 years are comparable to the original TS2 analysis.

While not done in this thesis, there is room for further investigation into this approach. It is possible that the iFAR threshold considered above was too high, leaving too many triggers in the secondary model for a specific region to be boosted in significance. For example, considering triggers within an iFAR 10-100 year region may be more beneficial, as the secondary model would only represent triggers which are near to the significance threshold rather than modelling triggers which are confidently classified as glitches. It is also possible that these triggers spread over wide regions of the attribute space, meaning the secondary model isn't targeting a specific region. The secondary model approach is investigated on such a targeted region in the following sub-section.



(a) FAR vs. T



(b) h_{rss50} for ad-hoc injections

Figure 5.6: Comparison of results for all GMM optimisation studies. Results are shown for signal training set 2 (TS2) analysis (blue), new N_{ED} re-parameterisation (orange), averaged secondary model on missed triggers (purple) and averaged secondary model on low-Q triggers (yellow). (a) shows FAR vs. T distributions, while (b) shows h_{rss50} estimates to ad-hoc waveforms at iFAR $\geq$ 100 years.

Secondary models on low Q_{veto} triggers

We test the above secondary model approach on another sub-population of training data, namely the low-Q region of the signal parameter space. The low-Q region is of interest for this approach since it is associated with glitches, in particular blip glitches. We form a secondary training set by taking cuts of $log10(Q_{veto0}+1) < 0.6$ OR $log10(Q_{veto1}) < 0.9$ on the signal training data. Again, data is processed by both models and a geometric mean is taken over the log-likelihood values of triggers, in hope that sensitivities to signals in the low-Q region will improve.

Results are seen in Figure 5.6. We again see that this approach appears to improve the FAR vs. T distribution at higher FARs, however has a comparable distribution at FAR of 1/100 years. Despite this, we see a slight improvement in h_{rss50} estimates at iFAR $\geq$ 100 years for GA and SG $Q = 3$ waveforms, with comparable sensitivities to the majority of remaining waveforms. This result proves the newly considered approach works, however the improvement seen in this case is not large enough to merit the training of secondary models during O4a, due to an increase in computational time for only a minor improvement across select waveforms. This being said, this approach may be promising in the future, and with more investigation into it's implementation the increase in sensitivity could be more significant.

5.1.3 The final O4a configuration

Subsequent to the above studies on the analysis of combined chunks and optimising the GMM approach, the final configuration for analysing O4a data is detailed below. Data is analysed over two adjacent chunks at a time, training a single model over data from both chunks and estimating sensitivities on a combined test data set. Training set 2 is used to train the signal model, which has increased ranges of frequency compared to O3 and is more densely sampled in the Q_{veto} space. We adopt the data split of 10% validation, 25% training and 70% testing for background data, and 20% validation, 80% training for WNB simulations. Data splits are executed on individual chunks before combining data sets to ensure an equal percentage of each chunk is considered. Background is estimated per chunk, ie. the time-shifts described in section 2.2.4 are applied to a single chunk of data, accumulating between 400-800 years per chunk dependent on the coincident livetime. After splitting data and combining chunks, this results in at least 10^3 years of background for FAR estimation on combined test sets. The attributes and re-parameterisation remain the same as utilised in the O3 run, namely: $log_{10}(E_c)$, $log_{10}(\eta_c)$, $logit(c_{c0})$, $logit(c_{c2})$, $log_{10}(N_{ED}+1000)$, N_{norm}, $log_{10}(\text{penalty})$, $log_{10}(Q_{\text{veto0}}+1)$, $log_{10}(Q_{\text{veto1}})$, $logit(L_{\text{ratio}}) = logit(\frac{L_{\text{veto1}}}{L_{\text{veto0}}})$. We do not adopt any of the above approaches involving the training of secondary models, hence the training and validation stages of the GMM methodology are the same as described in Chapter 3.

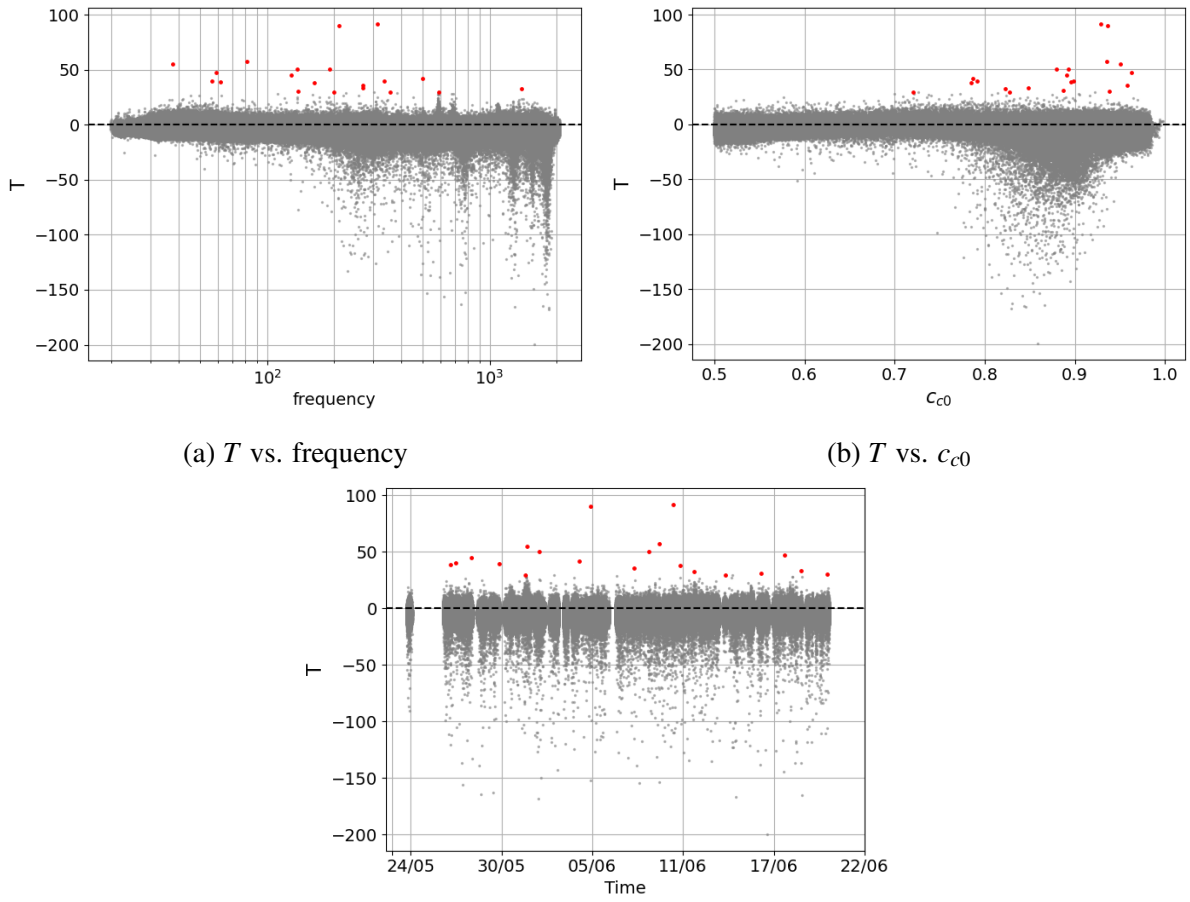
5.2 Investigations of loud glitches

One of the main motivations for offline analysis taking place while the observing run is ongoing is so that the behaviour of loud glitches in the data can be observed and investigations into their mitigation can take place before analysing final results. The mitigation of such glitches ultimately leads to improved sensitivity estimates to astrophysical sources and further increases the likelihood of detecting a burst-type GW event.

As described in section 1.3.2, the detector characterisation group aim to identify periods of data containing glitches known to be problematic, referred to as CAT1 and CAT2 vetoes, which are to be removed by all analysis pipelines. In addition to this, the group aims to identify the cause of loud glitches for individual pipelines, and investigates whether these can also be vetoed. In order for these investigations to take place, it is important for pipeline analysts to conduct their own glitch studies so that information on loud glitches can be sent to the detector characterisation group. Such investigations also offer information on the pipeline's weak points if there are certain glitches which are repeatedly ranked as significant, and can be used to improve the analysis in the future. This section will detail examples of such glitch investigations for the cWB+GMM search in O4a, leading to the creation and application of a new CAT2 veto specifically to tackle loud glitches in the GMM analysis. A summary of the noisiest chunks throughout the run is also provided.

5.2.1 Chunk-wise glitch investigations

Glitch investigations were carried out on each combination of chunks as they were analysed throughout the entirety of the O4a run. In order to investigate the properties of the loudest glitches affecting high iFAR thresholds, we observe plots of frequency, network correlation coefficient c_{c0} and time against the GMM detection statistic T for the background test data. Through these plots, we can observe whether the loudest glitches have somewhat 'expected' behaviour or whether their behaviour appears concerning and more in-depth studies are required. If loud glitches are clustered around a single frequency range, it could suggest that they all arise from the same source in the detectors. Similarly, if loud glitches are clustered around a given time period, it suggests that they arise due to some occurrence in the detectors at that given time. Both of these properties merit further investigation by the detector characterisation group. If loud glitches cluster around the low c_{c0} region, it could indicate some unexpected behaviour in other attributes of the glitches or within the models themselves, since it is expected that high network correlation will result in triggers scoring highly in T. The loudest 20 glitches are highlighted within the described plots, as they directly affect the thresholds for iFAR of 100 years. If any concerning behaviour is observed or triggers are classified with particularly high T we can investigate further, such as looking at plots of T against other attributes, exploring types of glitch through GRAVITYSPY [177], or observing the behaviour of the concerned glitches in



(a) T vs. frequency

(b) T vs. c_{c0}

(c) T vs. time

Figure 5.7: Plots of loud glitch studies in K01+K02. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

relation to the GMM models.

An example of the explained glitch investigation plots are shown for K01+K02 data in Figure 5.7. Glitch plots are shown for all combined chunks in Appendix C. As stated above, we expect the loudest glitches to be spread across frequency and time ranges, while clustering towards higher c_{c0} values, all of which is observed in the K01+K02 plots. Looking closer at T vs. frequency in Fig 5.7a, some clustering is seen at frequencies of 600Hz and 700Hz for glitches marked in grey at positive T values, suggesting a common source for said glitches. While the clustering isn't seen for the loudest 20 glitches, it can still be investigated by detector characterisation and could lead to sensitivity improvements if identified and vetoed. In the T vs. c_{c0} plot in Fig 5.7b, while the majority of the loudest glitches have c_{c0} as expected, a portion of glitches with very low T is also spread to high c_{c0}. This behaviour confirms that the GMM analyses is performing well, as it is able to classify a large portion of glitches which are correlated across the detector network.

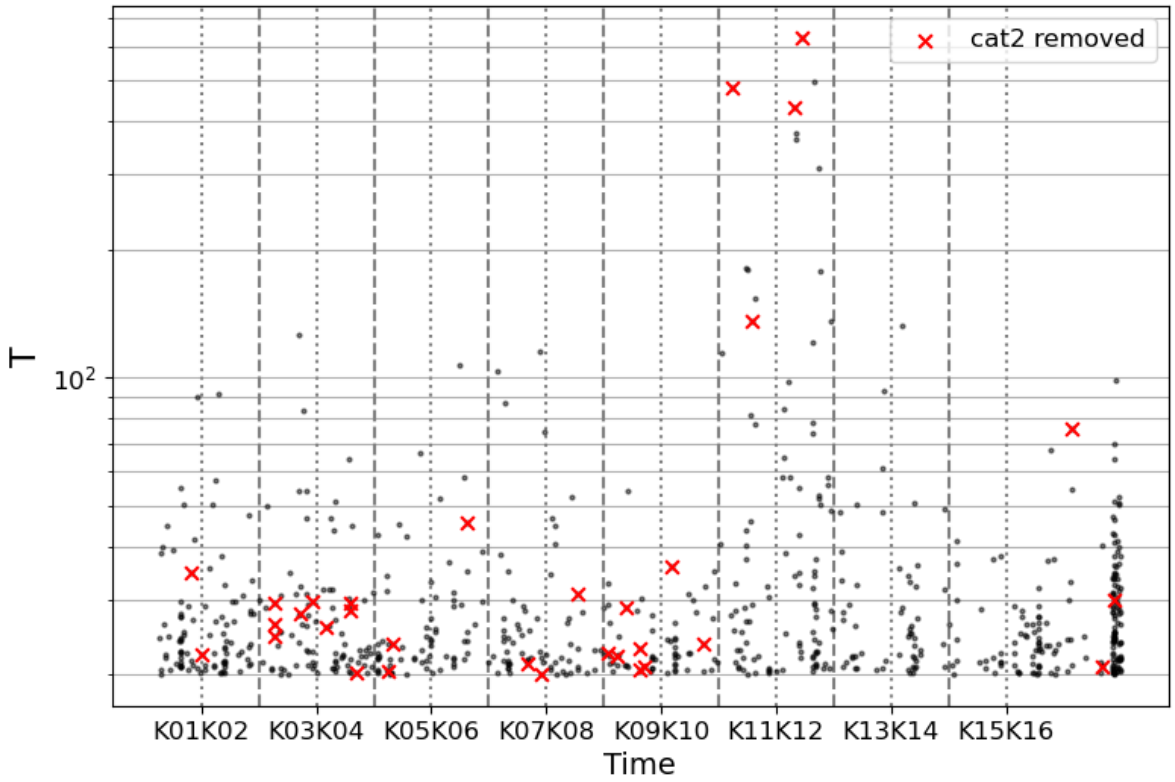


Figure 5.8: T vs. chunk for background triggers with T>20 across O4a. Glitches removed by the GMM-specific CAT2 veto are marked by red crosses. Borders between combined chunks are marked by dashed lines, while borders between chunks *within* each combined chunks are marked by dotted lines.

5.2.2 Applying a new CAT2 veto

Once the above glitch investigations were completed for all chunks in O4a, the detector characterisation group were able to identify a detector channel which caused a significant number of loud glitches for the GMM analysis, which was labelled as an extra CAT2 veto. CAT1 and CAT2 vetoes which were already defined for all pipelines were applied during the cWB production stage by removing relevant time periods of data. However since this GMM-specific CAT2 veto was defined after final trigger production was completed, we apply it during the post-production stage. The process here is the same, where time periods defined in the veto file are removed from background and zero-lag data, thus removing the labelled glitches from the analysis.

Figure 5.8 demonstrates the effect of applying this new CAT2 veto to the GMM search during O4a, with T vs. chunk shown for background triggers with T>20 and glitches removed by the GMM-specific CAT2 veto marked in red. The CAT2 veto removes many triggers with T<20, however we are interested in the T>20 region since it directly affects the iFAR thresholds at 10 & 100 years. It is seen that a significant amount of such triggers are removed, especially in K03+K04 and K09+K10, which may benefit the sensitivity to GW events in these chunks.

	K01+K02	K03+K04	K05+K06	K07+K08	K09+K10	K11+K12	K13+K14	K15+K16
T at iFAR$\geq$10yr	20.8	18.9	18.5	18.0	17.7	17.5	18.0	23.8
T at iFAR$\geq$100yr	41.7	34.6	36.7	29.8	27.5	27.7	38.8	47.0

Table 5.1: T values at inverse false alarm rate thresholds of 10 and 100 years across all chunks. These values are directly affected by loud glitches within each chunk combination.

The veto also removes some of the loudest glitches in K11+K12, which are classified with much larger T than the other chunks. Not known at this time, this behaviour was actually caused by a bug in trigger production for K12, which was later solved and is detailed further in sections 5.3.

5.2.3 Summary of loud glitches in O4a

Looking again at Figure 5.8, we can observe which chunks were most affected by loud glitches throughout the run. The distribution of glitches in T appears fairly consistent across the majority of chunks, apart from K09+K10 which have glitches classified with much lower T, and K11+K12 & K15+K16 which have features not seen in other chunks. As mentioned previously, K11+K12 had a problem with glitches being classified with particularly high T compared to other chunks due to a bug in trigger production (see section 5.3). Once this problem was solved, K11+K12 became one of the chunks with the least loud glitches, similar to K09+K10 and reflected in the T vs. iFAR distributions seen in Table 5.1. Refer to Appendix A for the corrected K11+K12 glitch investigation plots. K15+K16 contains a clustering of glitches being classified with high T at the end of K16, all occurring within the same time frame. Unfortunately the CAT2 veto from above does not account for these glitches, and investigations by detector characterisation were unable to identify the source. This group of glitches directly affects the T vs iFAR threshold for this chunk, which is demonstrated in Table 5.1 where the highest T thresholds are in K15+K16. This negatively impacts the sensitivities in these chunks, which is reflected in section 5.4. Looking across all chunks in Table 5.1, T thresholds are more consistent at iFAR $\geq$ 10 years, as expected.

In order to identify which type of glitches most affected the GMM analysis, the 10 loudest glitches from each chunk combination were processed through Gravity Spy [177]. Across the entire run, approximately half are blip glitches, while koi-fish and extremely loud glitches occurred a handful of times per chunk combination. Completing a study such as this for a larger amount of glitches would be useful for the GMM methodology in the future, as the analysis could be altered in an attempt to mitigate such glitches further.

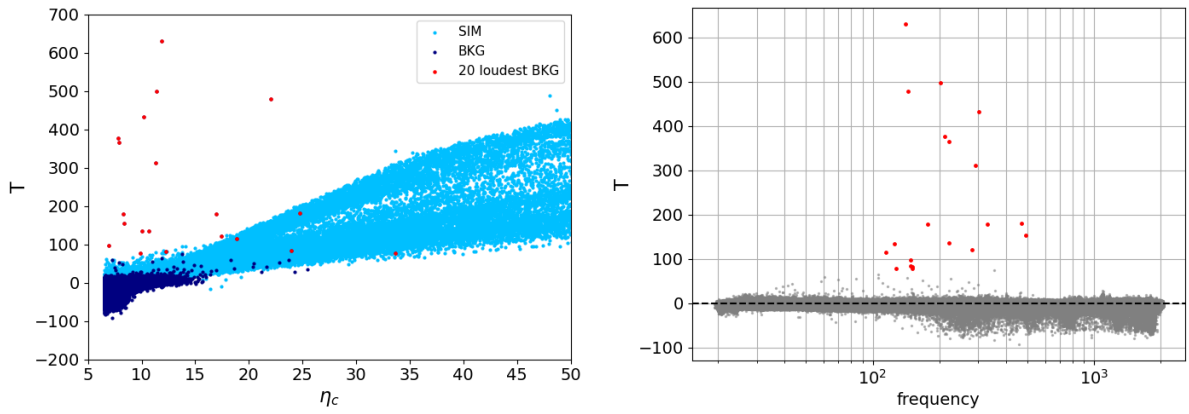
5.3 The noisy K11+K12 chunks

While analysing chunks K11+K12, it was noticed that there were ≈ 20 glitches from the background data being classified with extremely high T values (T between 100-600), which had not been seen in any of the other chunk combinations. These loud glitches caused the FAR vs. T relationship to have a very long tail out to high T values, in turn causing very high T thresholds on FAR estimates and causing poor sensitivities to astrophysical signals within these chunks. Since these glitches occurred only for K11+K12, investigations into glitch properties and GMM models took place to decipher what could cause such behaviour. It was eventually discovered that this problem arose from a bug in cWB trigger production, causing glitches from outlier regions of the attribute space to obtain extremely negative log-likelihood in the GMM background model and hence classified with very high T.

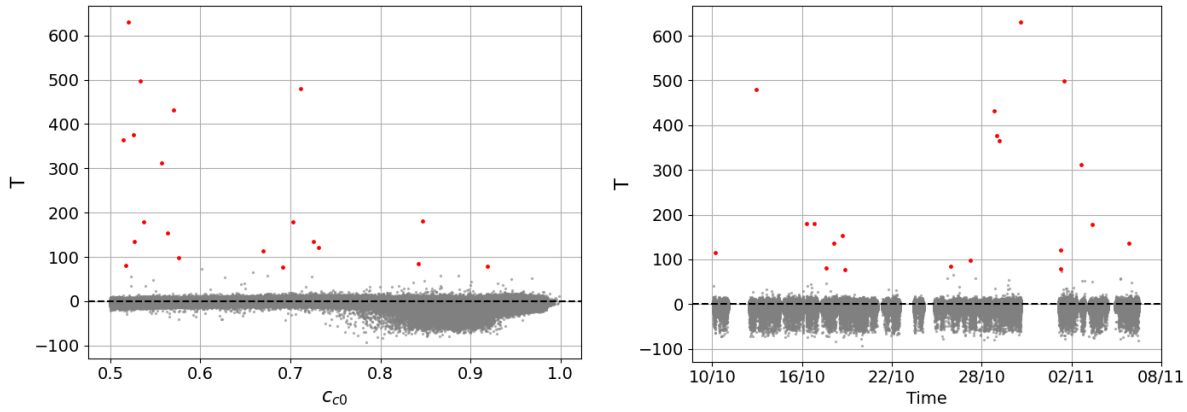
5.3.1 Investigating loud glitch attributes

As introduced in section 5.2.1, to initially look at the behaviour of glitches in K11+K12 we observe plots of T vs. network SNR η_c, frequency, network correlation coefficient c_{c0} and time in Figure 5.9. Looking first at the distribution of background and ad-hoc simulation triggers in T vs. η_c in Fig 5.9a, it can be seen that the majority of high T glitches have low η_c, which is unexpected. The majority of other triggers from both populations increase in T for higher η_c, whereas the loud glitches appear as outliers from this distribution with some being classified with higher T than any of the simulated signals. This may suggest that other properties of the glitches are causing them to lie within a region of the attribute space represented by the signal population. Fig 5.9b shows the distribution of background triggers over T vs. frequency, with all of the loudest glitches lying in the region [100,500]Hz. While they do not occur at a single frequency, there is still some clustering present compared to what is seen for other chunks, where the loudest glitches span the entire frequency range (refer back to Fig 5.7a). This behaviour of clustered frequency suggests it is a possibility for the loud glitches to have the same type (ie. similar morphology). The loud glitches also have an unusual distribution over c_{c0}, where Fig 5.9c shows that a large portion of them lie in the low c_{c0} region. We expect the GMM models to classify the majority of low c_{c0} triggers with low T since it is less likely for signals to occur here. While it is possible for a few low c_{c0} triggers to have very high T based upon other attributes, it is unexpected for it to occur for this many glitches. This behaviour suggests the some other attributes of the concerned glitches may be outliers to the background population, or there may be some problem with the model itself. Finally we look at the distribution of T vs. time in Fig 5.9d. The loud glitches do not all occur at the same time and hence are not caused by a single event on a detector site.

The above remarks on Figure 5.9 suggest that the problematic glitches may occur due to one or more of the following: a re-occurring source of glitch in the frequency range [100,500]Hz,



(a) T vs. η_c (b) T vs. frequency



(c) T vs. c_{c0} (d) T vs. time

Figure 5.9: Plots showing the distribution of loud glitches in K11K12 across a range of attributes, with the 20 loudest glitches shown in red for all plots. (a) shows network SNR η_c vs. T for background (navy) and ad-hoc simulation (blue) triggers. (b), (c), (d) show the distribution of background triggers for T vs. frequency, T vs. network correlation coefficient c_{c0}, and T vs. time respectively.

some outlier values in attributes not considered above, or a problem occurring within the models. We investigate each of these hypotheses, starting by investigating glitch type with Gravity Spy [177], which finds that *all* problematic glitches are labelled as extremely loud glitches. While it has been noted in section 5.2.3 that this type of glitch is occasionally classified as one of the loudest glitches for other chunks, it is unexpected that the K11+K12 model classifies only this type of glitch with high T. The detector characterisation group were unable to identify a single source causing these glitches, and as is seen above in Figure 5.8, the newly defined CAT2 veto only removes 4 of them. This explains the clustering of frequency seen above, however further investigation is required to learn what is causing such glitches to be classified in the high T region and whether this can be resolved.

We look into the properties of the K11+K12 extremely loud glitches further, comparing their distribution over the attribute space with loud glitches observed from other combined chunks in Figure 5.10. We take loud glitches to have $T > 75$. The K11+K12 loud glitches have similar

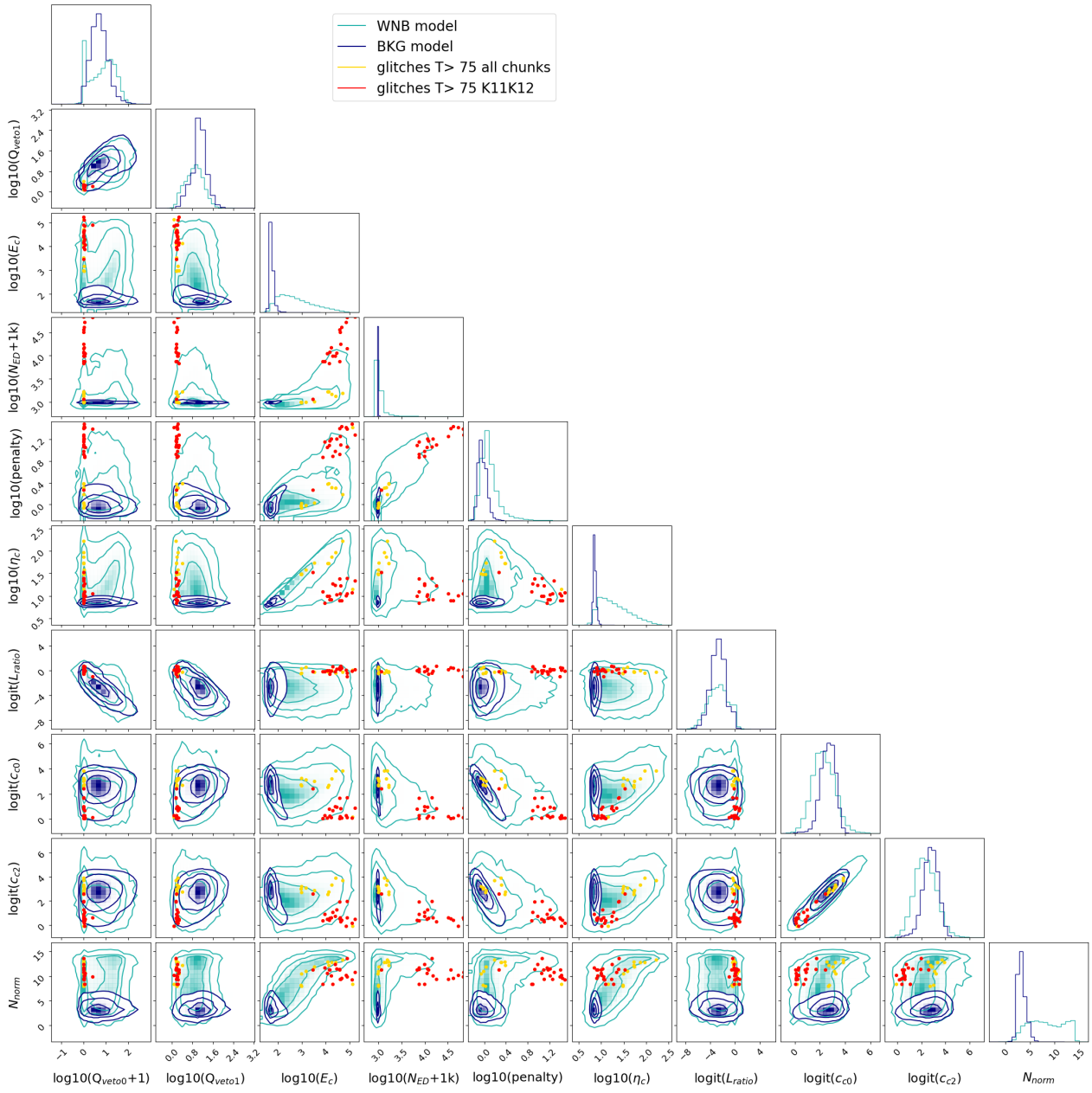


Figure 5.10: Corner plot showing the distribution of loud $T > 75$ glitches in K11+K12 (red) compared to loud glitches from all other chunks (yellow). Samples from the signal WNB model (pale blue) and background model (navy) are plotted behind for reference.

properties to loud glitches in other chunks for some attributes, however are outliers in $c_{c0,2}$, η_c, penalty, and N_{ED}. The difference in these properties makes the loud K11+K12 glitches lie in separate regions than other loud glitches for a large part of the parameter space, and coupled with the observation that all loud glitches from other chunks are grouped together with similar attribute values, it is suggested that there is some difference in the K11+K12 models which cause the classification of different glitch properties to have high T. The loud glitches do appear to lie in a region of the attribute space which is represented more by the signal population, however such glitches with low $c_{c0,2}$ and high N_{ED} & penalty exist in the other chunks but are classified with low T, reinforcing the above statement about a difference in K11+K12 models. With the inclination that there is some difference in the distribution of the K11+K12 models compared to others, the behaviour of the models are investigated below.

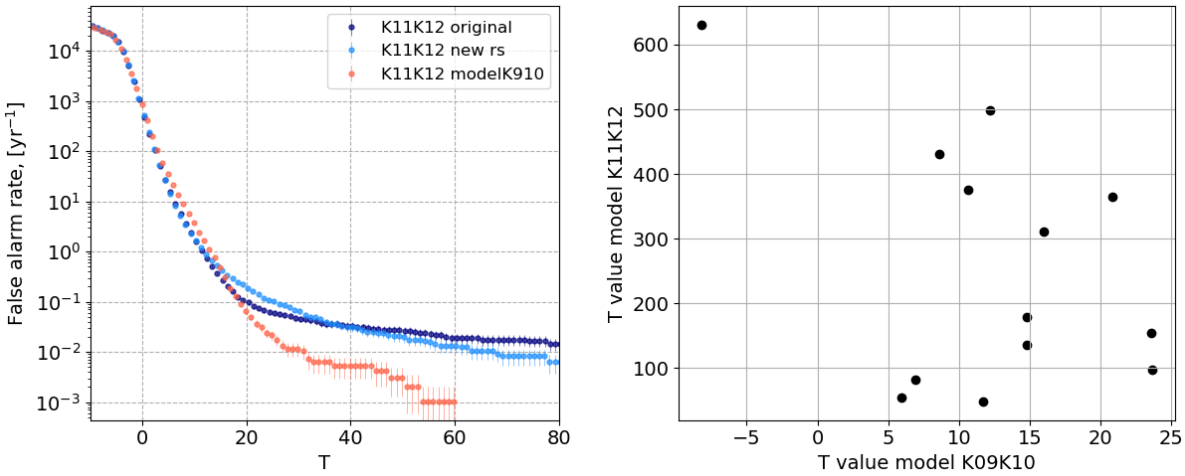
5.3.2 Application of different models

In an attempt to understand the classification of the extremely loud glitch type to high T in the K11+K12 model, we compare results on K11+K12 data completing the analysis with other models. Namely, we use models re-trained on K11+K12 data with different random states for the data split and training initialisation, comparing them to the K09+K10 trained models to observe whether these glitches are still problematic. We go on to compare samples from the K11+K12 and K09+K10 models to observe whether there is any visual difference in the 2-dimensional attribute space which could cause the high T glitch classification.

Different random state

Re-training models with different random states to avoid the potential of outlier bad Gaussian placement is not a new concept, introduced in section 3.2.4. However for this investigation we alter the random state in the data split function as well as the model training function, so that different portions of K11+K12 data is considered for validation, training and testing. The application of such a model will allow us to check whether there is a problem with the training data of K11+K12 or whether the Gaussian placements during original model training were sub-optimal.

Results on the K11+K12 data with the new random state model are seen in Figure 5.11 alongside the test with K09K10 model described below, in the form of FAR vs. T distribution comparisons. Comparing only the original K11+K12 and new random state results, it is seen that the new random state model obtains a slightly better T threshold at FAR of 1/100 years^{-1} (ie. iFAR of 100 years), however still has a long tail present in the distribution due to highly classified glitches. Investigating the distribution of loud glitches in the attribute space, the majority of glitches with $T > 75$ in the random state analysis have $c_{c0} < 0.65$ similarly to the the loud glitches in the original K11K12 model, however are clustered together in different regions of the



(a) FAR vs T

(b) T comparison for original outlier glitches

Figure 5.11: Comparison of analysis results on K11+K12 data with different GMM models. (a) shows the FAR vs. T distribution across 3 different models: the original K11K12 trained model (navy), the new random state K11K12 trained model (light blue) and the K09K10 trained model (orange). (b) shows T value comparisons between analysis with the original K11K12 data vs. with the K09K10 model for a portion of the original outlier high glitches.

attribute space in N_{ED}, penalty and E_c. While the remaining long tail of highly classified glitches suggests the problem lies within the K11+K12 data itself, the change in attribute properties of loud glitches other than c_{c0} suggest there could be some issue with the modelling of this attribute.

Model K09K10

We apply the already trained K09K10 model to the original K11K12 test data, since this model appears to perform as expected on the K09K10 data. Results for this analysis are again seen in Figure 5.11. The application of the K09K10 model appears to perform well on K11+K12 data, with a better distribution of FAR vs. T agreeing with expected behaviour compared to what is seen in other chunks. As a sanity check to ensure that better classification of glitches is due to the removal of the K11K12 model rather than K09K10 being a particularly good model, we apply the K09K10 model to all combined chunks in O4a. From this test it was concluded that the K09K10 model only brought benefit to the analysis of K11+K12, suggesting the classification of glitches with very high T is indeed a problem with the K11K12 model. For this analysis we can also compare the T values of the original highly classified glitches, seen in Fig 5.11b. Through the application of the K09K10 model, the glitches which were originally classified with high $T > 75$ with the K11K12 model are now classified with low T, which agrees with the expected behaviour from other chunks. The above results from the application of the new random state and K09K10 models confirms that there is some problem with the triggers in K11+K12 data which affect the training of the models but not the analysis of the triggers as test data.

Model comparisons

Since the problematic behaviour appears to arise from the K11K12 model, we take samples from the K09K10, the new random state K11K12 and original K11K12 models in an attempt to directly compare their structure across all attributes in the 2-dimensional space. This comparison is shown for the background models in Figure 5.12. From this visualisation no major differences are observed in the attribute space, however a few subtle differences are seen, especially in $log_{10}(E_c)$, $log_{10}(\eta_c)$ and log_{10}(penalty). It could be possible that these subtle variations are causing larger effects on the classification of triggers, or that the variations become more obvious when combined over the 10-dimensional attribute space. With this said, it must be noted that while a large pool of samples from each model are considered here, the binning definitions can affect the overall visualisation of the models. A better visualisation method is required in order to make any concrete conclusions about the variations between models.

5.3.3 The solution

While investigations into GMM model behaviour were underway, it was discovered that there was a bug within 2 jobs of cWB background trigger production for K12 in which data from Hanford was unable to be read into the algorithm, causing cWB to enter single-detector mode. In cWB, single-detector mode increases thresholds at each resolution to be much higher, resulting in these 2 jobs having a small amount of triggers. The triggers which were produced would be much louder glitches. In the interest of O4a analysis, fixing and rerunning these failed jobs solved the problem of the loud glitch classification seen above, with corrected models behaving as expected. However in terms of the future understanding of GMM, more investigation must be done on this matter as the behaviour of triggers from the failed jobs and their effect on the model is still not yet fully understood.

Only approximately 50 triggers from the failed jobs were in the training data set. By eye these triggers appeared to lie within the bulk of the distribution of the background population, however it is possible that they were somehow outliers in a region of the parameter space in more than 2-dimensions, relating to the issue of current visualisation methods mentioned above. Despite the total number of affected triggers being small in comparison to the $\sim 10^6$ trigger training set, the results thus far suggest that these triggers caused some region of the background attribute space to be under-represented, or some other region to be weighted more strongly. This statement can be reinforced by observing the distribution of background triggers log-likelihood values across background and signal models trained on the corrected data vs. data containing the trigger production bug. These distributions are seen in Figure 5.13, with T values displayed on the z-axis. In these plots, triggers with higher log-likelihoods in the background model have negative T while higher log-likelihood in the signal model results in positive T, seen directly from the definition of T in eq. 2.26. While the glitches with loud T in the 'wrong' models have

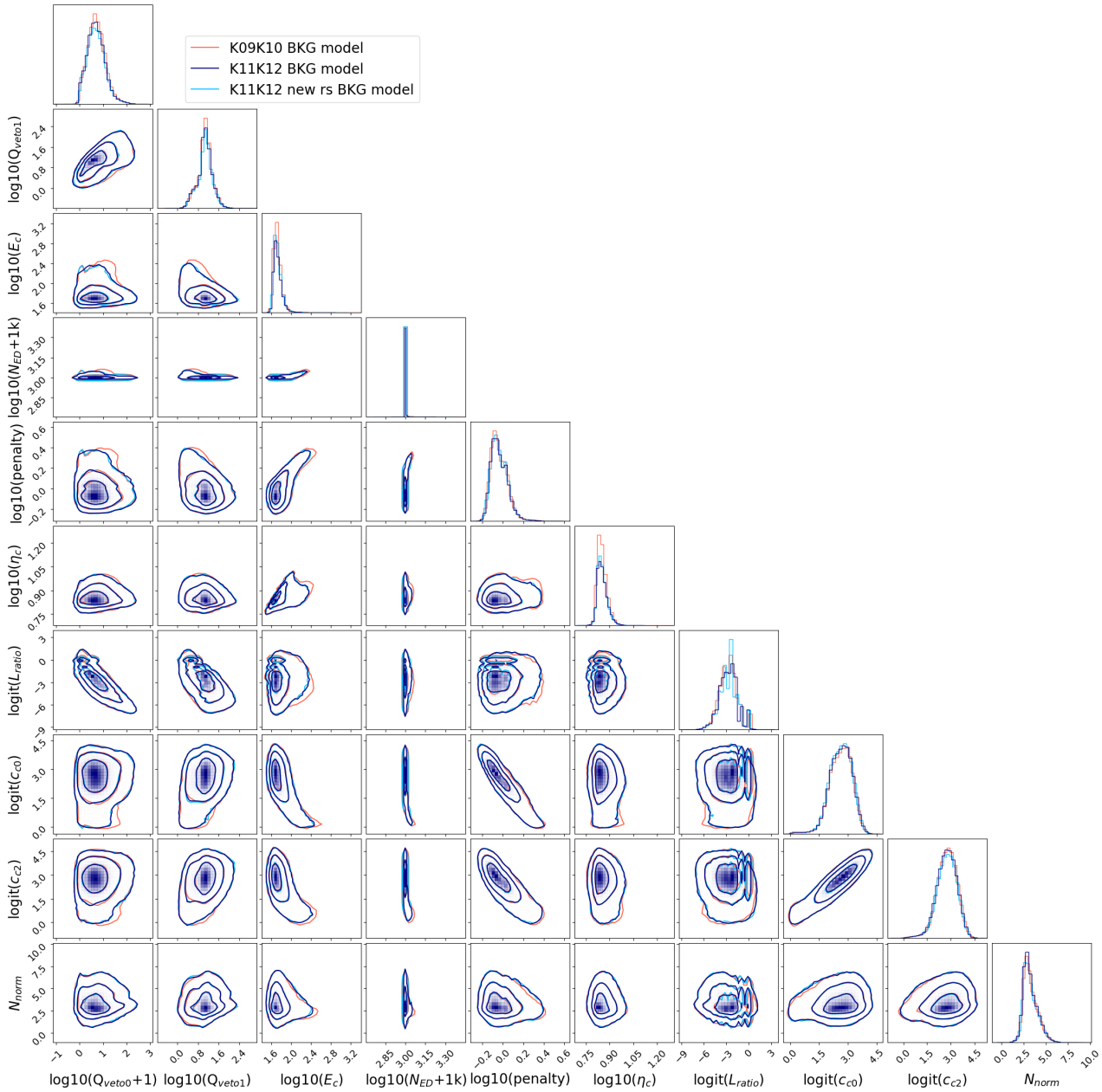
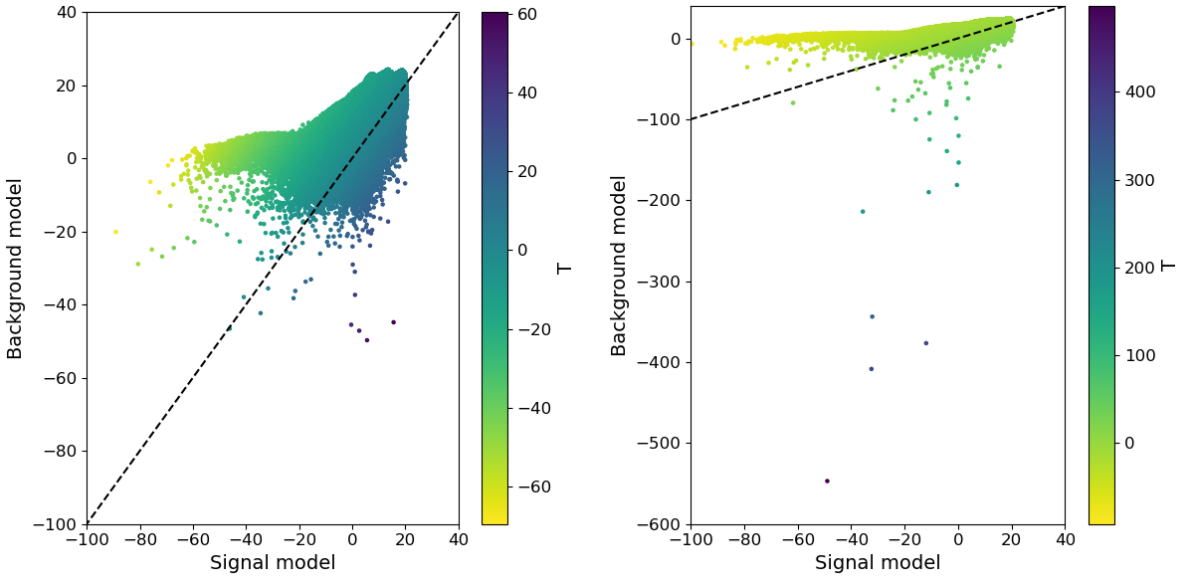


Figure 5.12: Corner plot showing the comparison of 3 different background models: the original K11K12 trained model (navy), the new random state K11K12 trained model (light blue) and the K09K10 trained model (orange). A total of 10^6 samples are taken from each model and plotted across the attribute space.



(a) Models trained on corrected K11+K12 data (b) Models trained on wrong K11+K12 data

Figure 5.13: Comparison of log-likelihood values of triggers in the background test data set with respect to background and signal models for K11+K12 data with and without the bug in cWB trigger production. (a) shows the log-likelihood distribution for models trained on the corrected K11+K12 data, while (b) shows the distribution for models trained on the 'wrong data'. Note the different scales of T on the colour-bar across both plots.

negative log-likelihood for the signal model, they have extremely negative values in relation to the background model, suggesting that the are not represented by either population. This causes a very high T classification despite the glitches not belonging to the signal attribute region. Due to time constraints within the O4a observing run, these investigations proved that our corrected K11K12 models were working as expected, and uncovering the direct cause of the above behaviour in the GMM models is put forward as a future investigation.

5.4 Finalised sensitivities to simulated signals in O4a

With glitch investigations complete and the problems with loud glitches in K11+K12 solved, we give finalised sensitivity estimates to sets of waveform injections across all chunks in O4a for the cWB+GMM pipeline. The GMM analysis is applied to the short duration, low-frequency search which in O4a covers signals within [0.001,10]s and [16,2048]Hz. We consider 2 sets of waveform injections: an updated set of ad-hoc waveforms to account for the increased frequency range of the low-frequency search, and a set of CCSN waveforms. The sensitivity to injections is considered across combinations of two adjacent chunks.

The sensitivity to given waveforms is again defined by the h_{rss50} statistic, first introduced in section 2.2.5. h_{rss50} is found in the same way as introduced in Section 4.1.2; by plotting detection efficiency as a function of h_{rss} and fitting a lognormal distribution to infer the h_{rss} at a

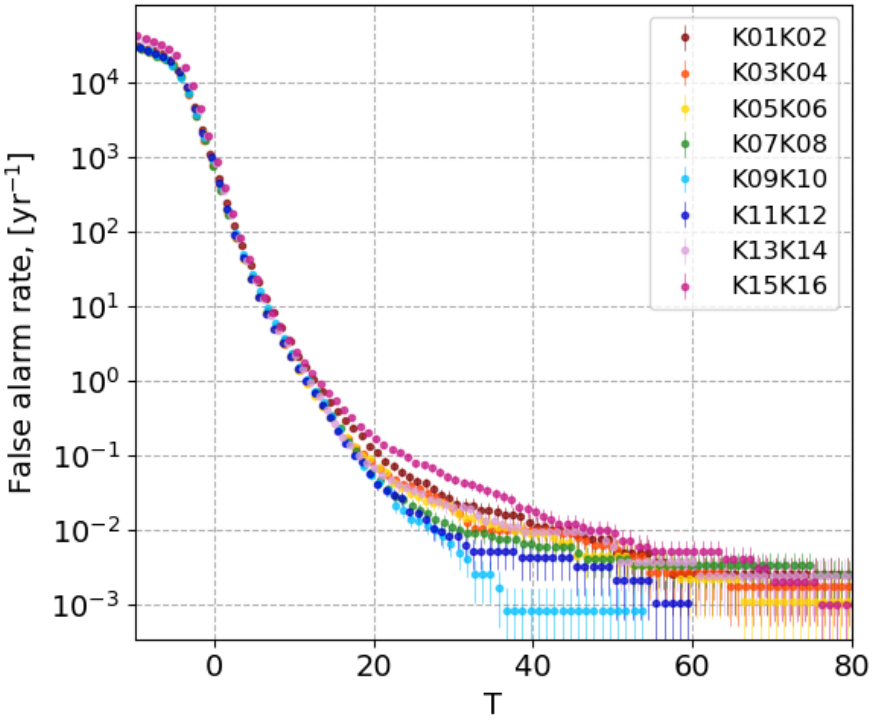


Figure 5.14: Finalised FAR vs. T distributions for all adjacent chunk combinations in O4a.

detection efficiency of 50%. Errors on h_{rss50} are computed from upper and lower lognormal fits to the efficiency vs. h_{rss} distribution based upon errors on detection efficiency from eqn. 2.16. Plots of detection efficiency vs. h_{rss} over all chunks in O4a can be seen in Appendix B. If 50% detection efficiency is not reached within the injected h_{rss} range of a given waveform, the fitting script allows for h_{rss50} to be inferred through the extension of the fitted curve, up to a value of $h_{rss} = 1 \times 10^{-21}$ Hz$^{-1/2}$. However, occasionally the curve can be fit in such a way that it levels out before this point, thus not reaching 50% detection efficiency before 1×10^{-21}Hz$^{-1/2}$, in which case the h_{rss50} is set to the maximum injected h_{rss} for the set of waveforms. This behaviour is seen in the sensitivity estimates of SGE $f_0 = 36$Hz, $Q = 3,9$ in Figure 5.15, marked by black crosses.

Before looking at the h_{rss50} of given waveforms, we first look at the variation of noise across each chunk combination by comparing FAR vs. T distributions in Figure 5.14. Since this distribution directly displays the T values associated with various FAR thresholds, it can give an indication on which chunks will have the best sensitivities to signals. Here K09+K10 and K11+K12 have the best T thresholds for both FAR 1/10 & 1/100 years, which is likely due to less loud glitches occurring within this time frame. As expected from glitch investigations in section 5.2, K15+K16 appear to be the noisiest chunks, with higher T thresholds for FAR>1/10 years. With this said, sensitivities do also rely on the distribution of T in the signal data, so such a plot is not deterministic on sensitivities in a given chunk and it is often beneficial to observe both the FAR vs T and h_{rss50} plots in unison.

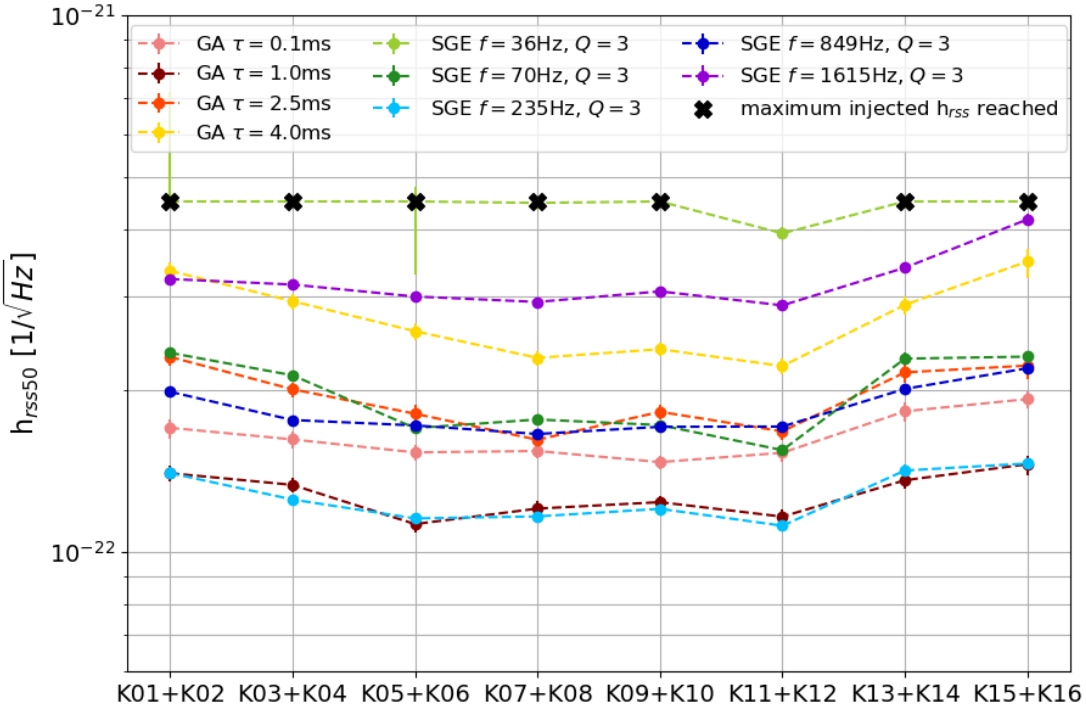
Gaussian Pulse (GA)	Sine-Gaussian Circular (SG)		Sine-Gaussian Elliptical (SGE)		White Noise Burst (WNB)		
τ (ms)	f_0 (Hz)	Q	f_0 (Hz)	Q	f_{low} (Hz)	Δf (Hz)	τ (s)
0.1	70	9	36	3,9	150	100	0.1
1	100	9	48	9,100	300	100	0.1
2.5	253	9	70	3,100	700	100	0.1
4	361	9	153	9			
			235	3,100			
			554	9			
			849	3,9,100			
			1304	9,100			
			1615	3,9,100			

Table 5.2: Table of injected ad-hoc waveforms with defining parameters for the all-sky short low-frequency search in the O4a observing run. Totals to 4 GA waveforms varied over duration, 22 SG waveforms varied over central frequency and Quality factor with circular (SG) and elliptical (SGE) polarisation and 3 WNB waveforms varied over lower frequency bound with set bandwidth and duration.

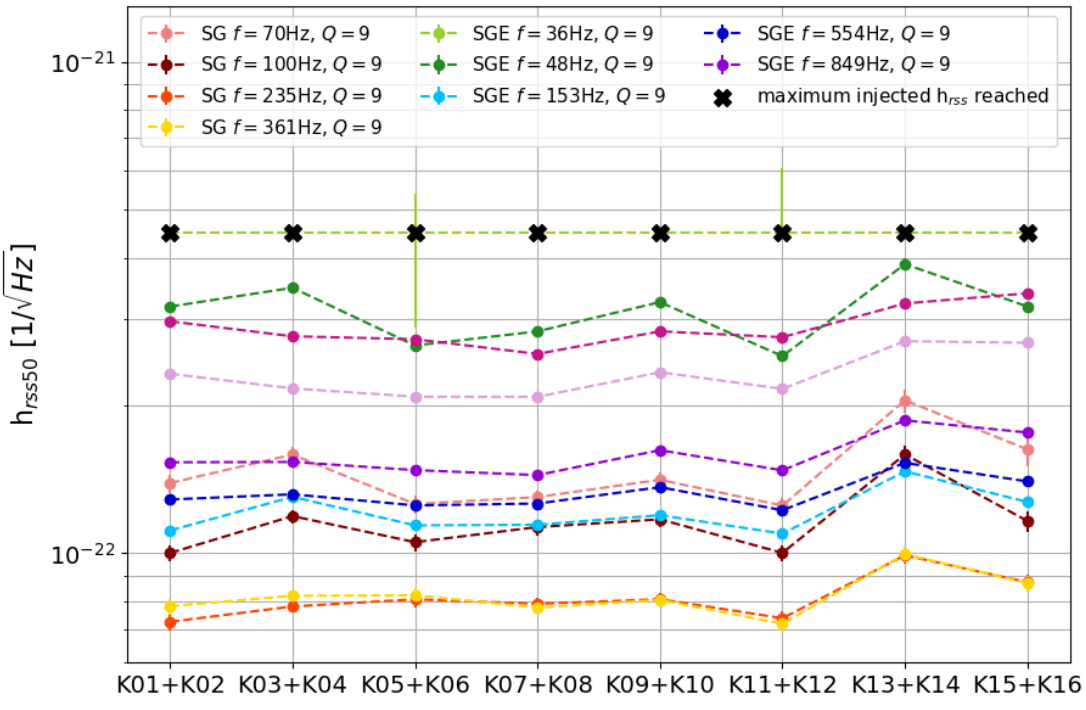
5.4.1 Ad-hoc injections

For the final sensitivity studies in O4a, we consider an updated set of ad-hoc waveforms from the ones used previously, with the aim of sampling more of the considered frequency range of the search. These include the same GA waveforms varied over duration [0.1,4]ms and the same WNB waveforms varied over lower frequency bound [150,700]Hz with bandwidth of 100Hz and duration of 0.1s. However additional SG waveforms are added to sample central frequency from [36,1615]Hz with $Q = 3, 9, 100$, this time with a mix of circular polarisation (source is assumed to be optimally oriented) and elliptical polarisation (distributed uniformly in the cosine of the inclination angle of the source). A full rundown of the considered waveforms are in Table 5.2. Injections are made over a fixed grid of h_{rss} distributed over $(\sqrt{3})^{(N-4)} \times 5 \times 10^{-23}$, where N ranges from 0-8, giving a maximum injected h_{rss} of 4.5×10^{-22} Hz$^{-1/2}$.

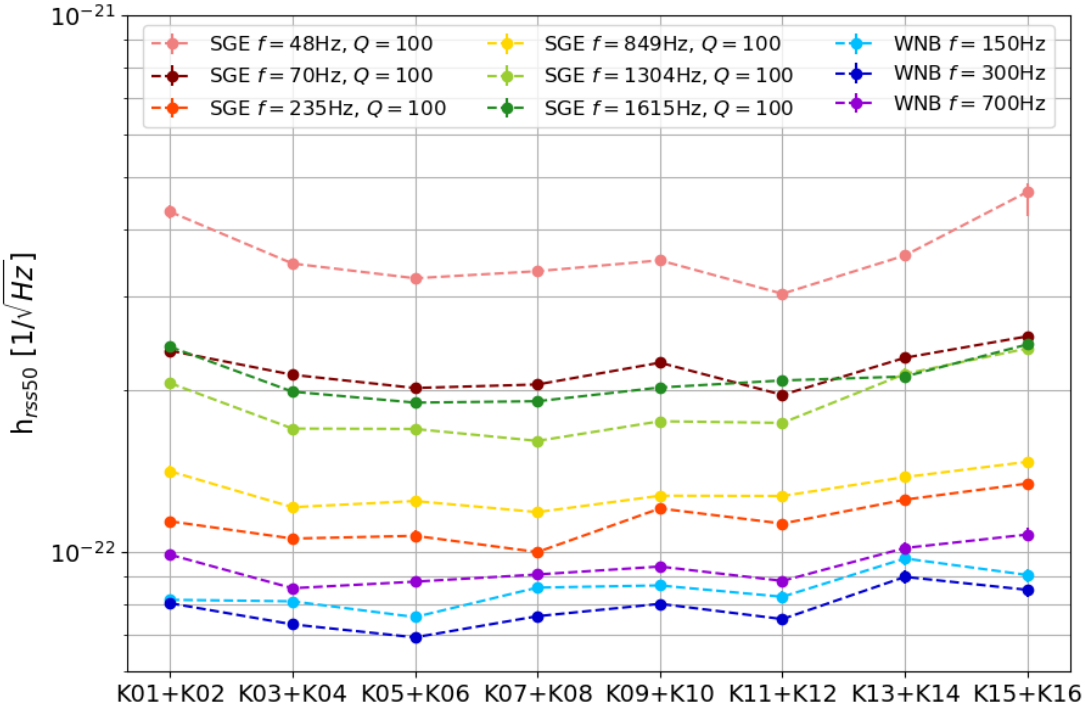
Figure 5.15 shows the final h_{rss50} estimates to all ad-hoc injections at iFAR $\geq$ 100 years, over combinations of chunks throughout O4a. Due to the large number of waveforms considered, results are split into 3 separate plots, grouped by Quality factor: GA and SG $Q = 3$ waveforms are shown in Fig 5.15a, SG $Q = 9$ waveforms are shown in Fig 5.15b, while SG $Q = 100$ and WNB waveforms are in Fig 5.15c. Looking first at the sensitivities to specific waveforms, the GMM analysis appears to be the most sensitive to WNBs and SG waveforms with frequency in the range [200,900]Hz. Better sensitivities are also achieved for SG waveforms with $Q = 9, 100$ than $Q = 3$ waveforms at the same frequencies. With this said, waveforms in the low-Q region of the parameter space such as GA and SG $Q = 3$ are typically recovered with lower h_{rss50} due to the presence of blip glitches. This means that despite having worsened sensitivity to such waveforms compared to the rest of the parameter space, the GMM analysis is able to achieve



(a) h_{rss50} estimates for GA and elliptically polarised SG $Q = 3$



(b) h_{rss50} estimates for circular and elliptically polarised SG $Q = 9$



(c) h_{rss50} estimates for elliptically polarised SG $Q = 100$ and WNB

Figure 5.15: Finalised h_{rss50} estimates to the ad-hoc waveform injections detailed in Table 5.2 in O4a at iFAR $\geq$ 100 years. The h_{rss50} of each injected waveform is shown over combinations of chunks throughout the run. Waveform sensitivities are split into 3 sub-figures based upon Quality factor, with GA and SGE $Q = 3$ shown in (a), SG and SGE $Q = 9$ shown in (b), and SGE $Q = 100$ and WNB shown in (c). All plots have the same lower-bound y-axis for sake of comparison. If the injected waveform does not reach 50% detection efficiency, h_{rss50} is set to the maximum injected h_{rss} value, notated by a black cross.

competitive sensitivities, improving on quoted h_{rss50} estimates from O3. The GMM search is least sensitive to low-frequency waveforms, with h_{rss50} estimates for SGE, $f_0 = 36$Hz, $Q = 3,9$ saturated at the maximum injected h_{rss} of 4.5×10^{-22}Hz for multiple chunks due to detection efficiency not reaching 50%. Such estimates are marked by a black 'X' on Figure 5.15. Higher h_{rss50} is also quoted for SG waveforms with $f_0 = 48, 1304, 1615$Hz, suggesting the models are also less sensitive to the high-end region of the considered frequency range. Having information on GMM's least sensitive regions is important, as it has the potential to lead to investigations into reinforcing the models in given areas of the attribute space to improve performance on such regions of the signal parameter space, with the possibility of increasing sensitivities for future observing runs.

Turning our attention to look at h_{rss50} estimates over chunk combinations throughout the run in Figure 5.15, sensitivities appear to be worse for chunks at the very start and end of the

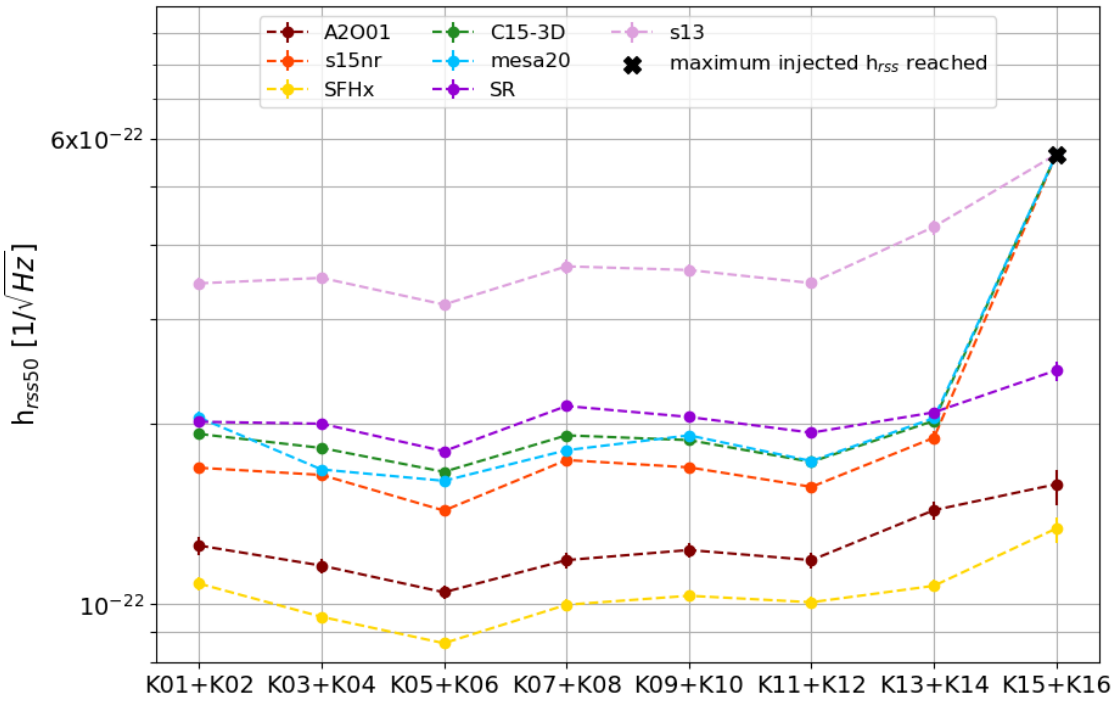


Figure 5.16: Finalised h_{rss50} estimates to selected CCSN waveforms in O4a at iFAR $\geq$ 100 years. The h_{rss50} of each injected waveform is shown over combinations of chunks throughout the run. If the injected waveform does not reach 50% detection efficiency, h_{rss50} is set to the maximum injected h_{rss} value, notated by a black cross.

run for the majority of waveforms, particularly for K13+K14, which somewhat agrees with the FAR vs. T distributions in Figure 5.14. With this said, the most sensitive chunks are highly reliant on individual waveform type, with highest h_{rss50} not always occurring in K15+K16 as was previously suggested.

5.4.2 Core-Collapse Supernovae injections

The considered set of CCSN injections varies compared to the set observed in O3, this time only consisting of 6 neutrino-driven explosion models: Andersen et al. 2018 [178] (s15nr), Kuroda et al. 2016 [171] (SFHx), Mezzacappa et al. 2020 [179] (C15-3D), O'Connor & Couch 2018 [132] (mesa20), Pan et al. 2020 [180] (SR), Radice et al. 2018 [133] (s13), and 1 magnetorotationally-driven explosion model: Abdikamalov et al. 2014 [172] (A2O01). The considered waveform models exhibit various possible features of CCSN during collapse, with an overview of such properties provided in Appendix A. These models are injected over a fixed grid of h_{rss} distributed over $(\sqrt{2})^N/2 \times 5 \times 10^{-23}$, where N ranges from 1-9, giving a maximum injected h_{rss} of 5.7×10^{-22} Hz$^{-1/2}$.

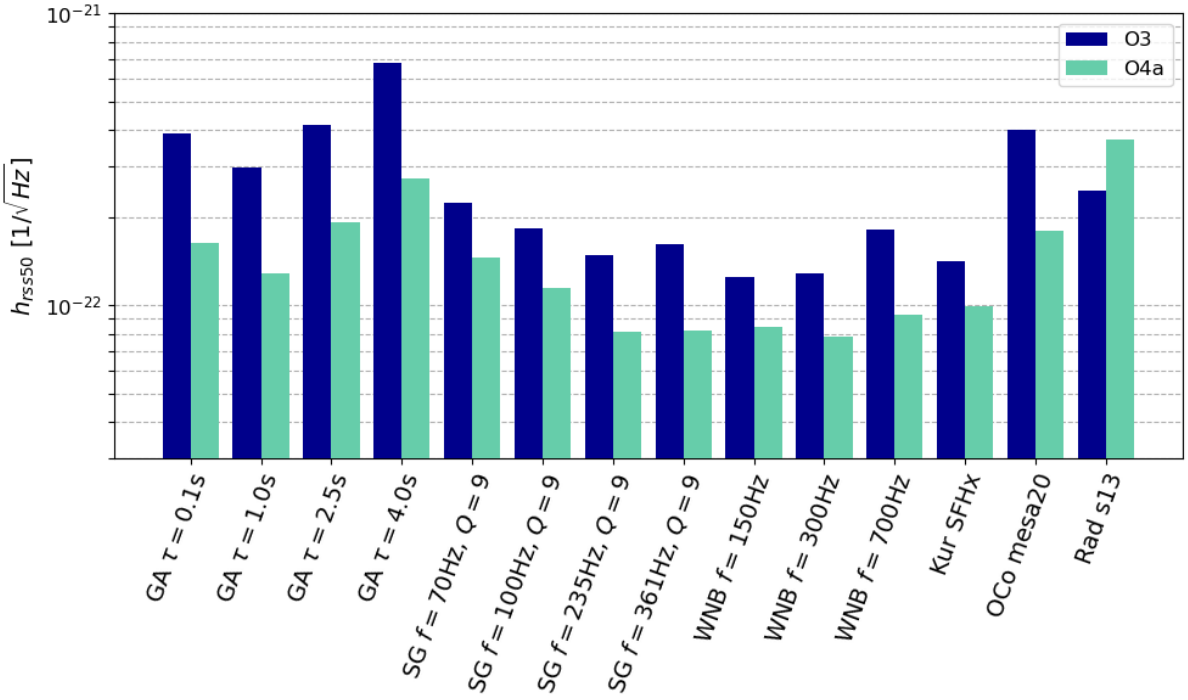


Figure 5.17: Comparison of h_{rss50} sensitivity estimates between O3 (dark blue) and O4a (pale green) for waveform injections at an iFAR threshold of 100 years. Includes estimates to GA, SG $Q = 9$, WNB and CCSN.

Final h_{rss50} sensitivities to CCSN injections are shown in Figure 5.16 at iFAR $\geq$ 100 years, over combinations of chunks throughout O4a. Looking at sensitivities to each waveform, it is seen that the GMM analysis is able to recover the majority of CCSN waveforms with comparable sensitivities. The best sensitivity is achieved for the *SFHx* and *A2O01* models, while the analysis is least sensitive to *s13*. The *s13* model is the only one which does not actually explode, which may explain the lowered sensitivity compared to other models. Turning attention to the variation of sensitivity over chunk combinations, the h_{rss50} in K15+K16 is the highest, with values being maxed out at the maximum injected h_{rss} of 5.7×10^{-22} $Hz^{-1/2}$ for *s15nr, C15-3D, mesa20, s13* waveforms, due to efficiency not reaching 50% in the h_{rss} curves. This is likely due to the spike of loud glitches occurring at the end of the run seen from Figure 5.8, however no solution could be found to remove these. For all CCSN waveforms best sensitivities are achieved during K05+K06, which is different from the most sensitive chunks for ad-hoc injections.

5.4.3 Comparison with O3

The total h_{rss50} estimates over all chunks in O4a are estimated by considering all recovered vs. injected waveforms throughout the run. These values are compared to O3 h_{rss50} estimates from Chapter 4 in Figure 5.17. Not all waveform injections can be considered here since the set of both ad-hoc and CCSN waveforms differed between observing runs. The compared waveforms

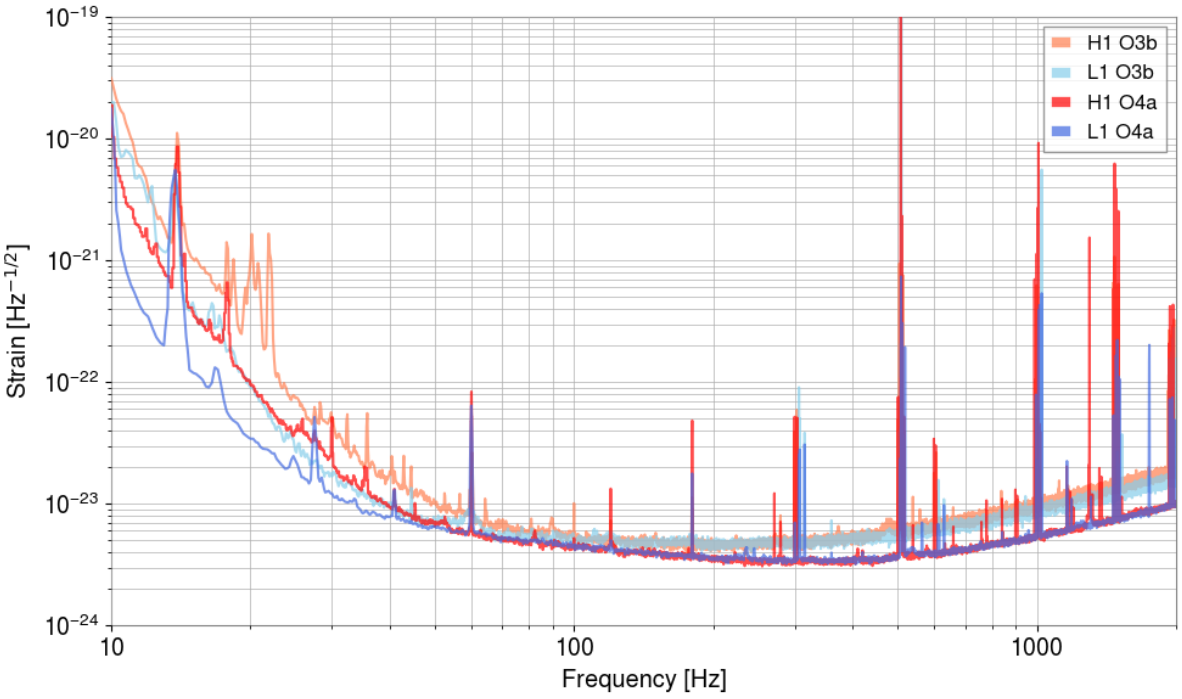


Figure 5.18: Comparison of detector sensitivities for LIGO-Hanford (H1 - red) and LIGO-Livingston (L1 - blue) across the O3b and O4a observing runs. The amplitude spectral density (strain) is shown across the LIGO sensitive frequency range. The noticeable increase in sensitivity seen in both detectors for O4a is due to a range of detector improvements made between the runs, detailed in [181].

include a selection of GA, SG $Q = 9$ and WNB waveforms, alongside 3 CCSN models: Kuroda (SFHx) [171], O'Connor & Couch (mesa20) [132] and Radice (s13) [133]. It can be seen that a significant increase in sensitivity is observed across all waveforms except from the Radice s13 CCSN. The cause of decreased sensitivity for Rad s13 during O4a is unknown, and investigations into it's position across the GMM parameter space may give insight into this in the future. Excluding the s13 CCSN, an average of $\sim 1.9\times$ sensitivity increase is observed across all considered waveforms. While a portion of the observed sensitivity improvement between O3 and O4a is due to advancements in cWB and GMM methodologies, detector upgrades have also contributed. A measure of detector sensitivity between the observing runs is seen in Figure 5.18, where the averaged amplitude spectral density (strain) associated with the LIGO-Hanford and LIGO-Livingston detectors is shown separately for O3b vs. O4a [11, 181]. By looking at the factor of improvement between the O3b and O4a curves at given frequencies, it is possible to observe how much of the h_{rss50} improvement for given given waveforms is due to detector upgrades. For example, the improvement factor at 700Hz is $\frac{O3b}{O4a} \sim \frac{7\times10^{-24}}{4\times10^{-24}} \sim 1.75$. Thus we can tell ~ 1.75 of the sensitivity improvement for the WNB $f_0 = 700$Hz waveform is due to detector upgrades.

The substantial sensitivity improvements seen between observing runs provides a promising statement on how sensitivities may continue to increase in the future. Not only does this increase the possibility of a burst detection, but allows for the placement of more stringent upper limits on expected astrophysical sources.

5.5 Zero-lag analysis in O4a

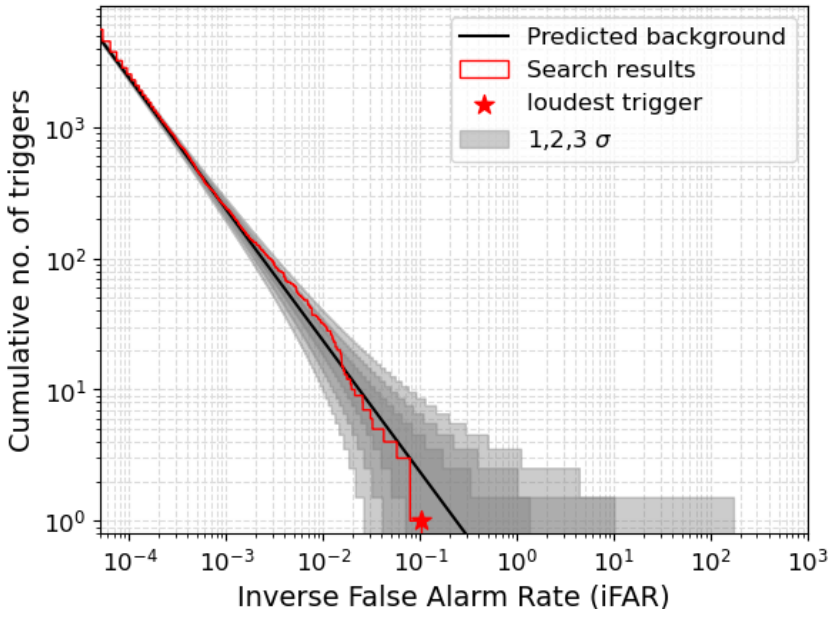
Having finalised sensitivity estimates to the considered waveform injections throughout the run, we now present the final results of the cWB+GMM search for short GW transients in the O4a observing run. Similarly to previous zero-lag analysis, we process a 'fake' lag from the time-shifted background to ensure the analysis works as expected, referred to as *closed box analysis*, before observing the real data incoming from the detectors, referred to as *open box analysis*. The fake-lag and zero-lag data are processed as individual chunks, with respective combined models being applied to each and the extra CAT2 veto from section 5.2.2 being applied to remove periods known to contain loud glitches. Triggers from all chunks are combined in the final stages of analysis for the presentation of results.

5.5.1 Closed box analysis

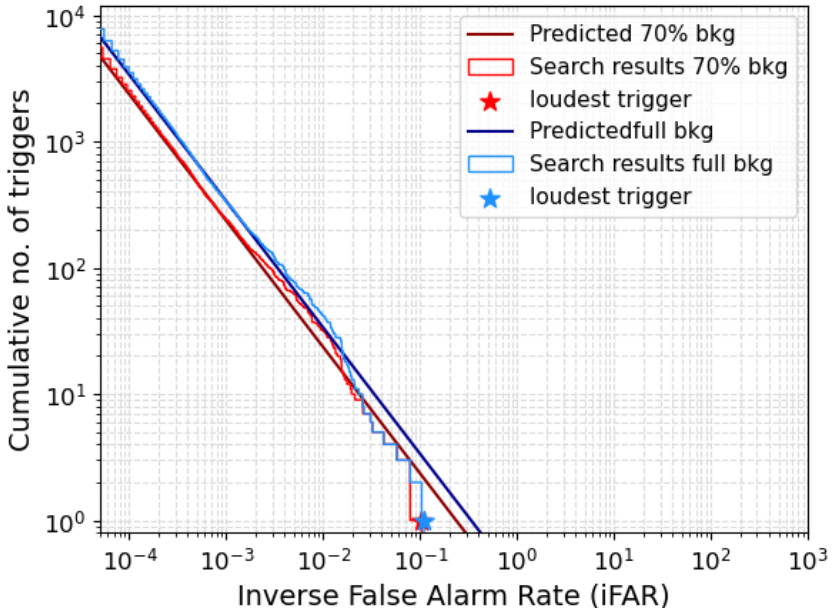
Closed box analysis is completed on the time-shifted segment of lag=2, slag=0 from the test data set for all individual chunks. As was discussed in Section 4.1.3, any number of lag and super-lag can be chosen. The time-shifted segment is processed by GMM, allowing FARs to be assigned to triggers and the distribution of trigger significance to be observed. Since we are analysing a time-shifted background segment, we expect the cumulative number of triggers per iFAR to be consistent with the predicted background. The predicted background is calculated by counting the number of expected triggers per iFAR and scaling the distribution to be consistent with the considered livetime of the segment. Results from this process are shown in Figure 5.19a with triggers combined over all chunks. We see that the significance of triggers from the segment agree with the background to within 2σ, demonstrating that our analysis is working as expected.

The effect of test percentage split of lag vs. full lag

As mentioned above, we complete closed box analysis by taking a time-shifted segment from the background test data set, which for O4a is only 70% of the total background per chunk. The background is randomly split when obtaining the test data set, meaning our considered segment may not contain *exactly* 70% of the full lag=2, slag=0 segment. However in the above result we have assumed this is the case within a statistical error, using 70% of the segment livetime to calculate the predicted background distribution. Ideally closed box analysis should be completed



(a) True closed box results of lag=2,slag=0 from test data



(b) closed box on test data vs. closed box on full segment

Figure 5.19: Closed box results for the cWB+GMM search in O4a. (a) shows the 'true' closed box results for a segment taken from the background test data in red, over predicted background with $1, 2, 3\sigma$ regions in grey. (b) shows the comparison of closed box results for a segment taken from the background test data (red) vs. results for a segment taken from the full background data (blue)

on a full segment to avoid this possible inconsistency, however doing this in O4a would cause an overlap of data between the full segment and model training data, which could potentially cause some bias.

To ensure the statement that any effect of considering a time-shifted segment from the 70% test data set is negligible is true, we compare the results using the 'true' 70% of the time-shifted segment to closed box analysis completed with the full time-shifted segment of lag=2, slag=0. The results for this are seen in Figure 5.19b, with predicted background calculated for each segment considering the respective livetime of each segment. It can be seen that the overall distribution of both analyses are very similar, with the same excess in triggers occurring between iFAR $10^{-3} - 10^{-2}$ in both cases. Both results also appear to follow the distribution of predicted background. This suggests that, as expected, considering only 70% of the time-shifted segment is trustworthy for this analysis, and the behaviour of the pipeline agrees with predictions. For future analysis, some lag=*N*,slag=0 segment should be removed from the background before undergoing data splits, as is done the for zero-lag data, to ensure data from the entire time-shifted segment is considered.

5.5.2 Open box analysis

With the confirmation that the analyses follows expected behaviour, we go on to analyse the zero-lag data in the search for GWs from short duration transients. The zero-lag data is processed in individual chunks, with iFAR estimates applied to triggers by respective combined chunk models and all triggers combined in the end stages of analysis for presentation of results. Previously in Section 4.1.3 the concept of standard and extended segments within cWB was introduced, with standard segments occasionally missing small amounts of data due to segment length definitions. For the O4a search the cWB+GMM analysis only analyses standard segments since extended segment analysis is covered by the cWB+XGBoost search and does not find any extra significant event.

Search results are shown in Figure 5.20, in the form of cumulative number of events vs. iFAR, with known CBC detections with iFAR ≥ 1 year labelled. In total the cWB+GMM search recovers 46 known CBC events across all iFARs, with 13 found at a significance threshold of iFAR ≥ 1 year. The loudest event detected by GMM is S230627c with an iFAR $\approx 10^3$ years. Figure 5.20 also shows results with known CBC detections removed, with the loudest non-CBC trigger marked by a square. Results with no known CBCs are consistent with the background to within 1σ, which suggests a null detection on short duration, low-frequency burst-type events in the LVK O4a observing run.

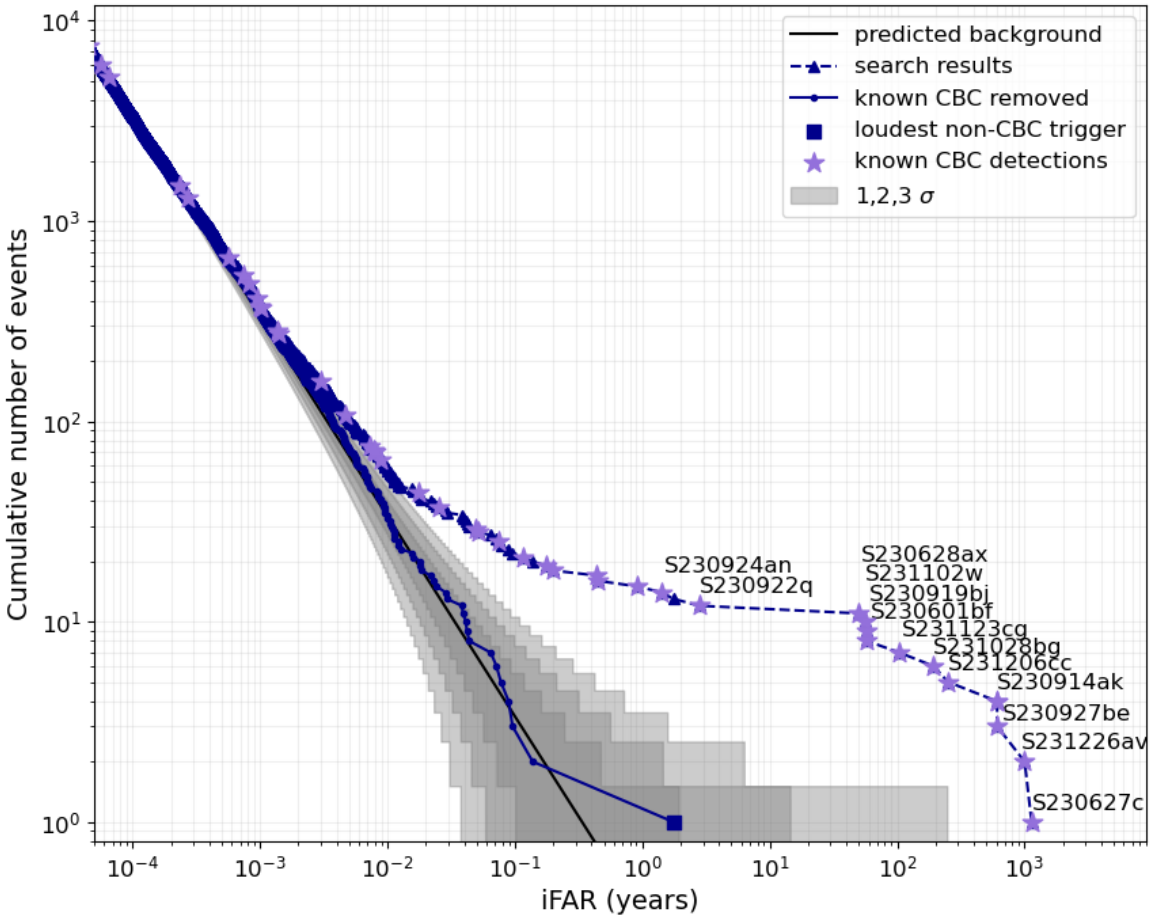
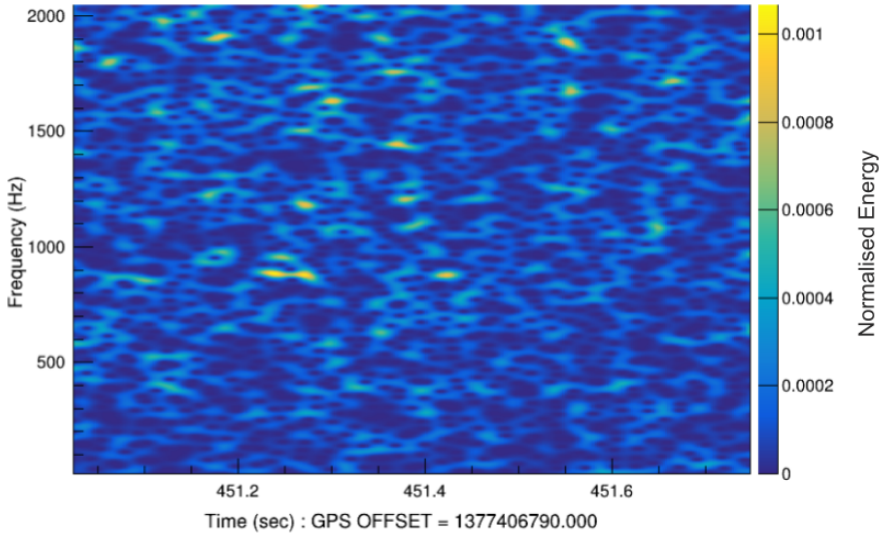


Figure 5.20: Search results of cWB+GMM across all chunks in O4a. Full search results are notated by triangular dashed markers, with known CBC detections labelled with purple stars. Search results with known CBC events removed are notated by solid circular markers, with the loudest non-CBC trigger marked with a square at iFAR 1.75 years. Predicted background is shown in black with $1, 2, 3\sigma$ error regions in grey.

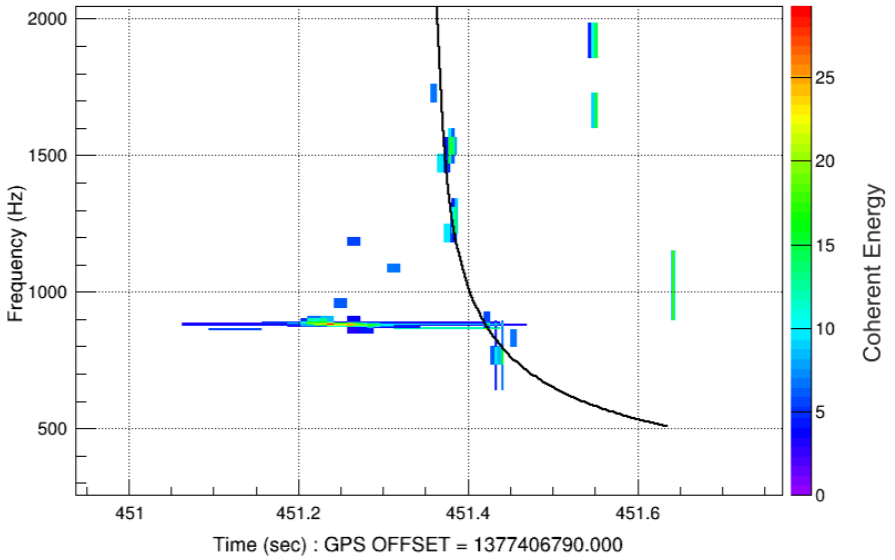
The loudest non-CBC trigger

The loudest non-CBC trigger in the GMM O4a search has a iFAR of 1.75 years. This is within the 1σ region of the predicted background, however since an iFAR of 1 year may signify an event of interest in offline burst searches, we investigate it's properties further.

The trigger occurred during K08, with η_c of 12.5 and c_{c0} of 0.76, and was also found by the cWB2G-XGBoost search at a much lower iFAR of 0.059 years, which is under the iFAR threshold for an interesting event. Representations of the trigger in time-frequency space from cWB is seen in Figure 5.21. The time-frequency spectrogram of the Hanford data surrounding the time period of the trigger is seen in Fig 5.21a. There are a few scattered peaks across the time-frequency space, however the total normalised energy of the data is rather small, suggesting there is not a loud event in the data. Fig 5.21b displays the likelihood of pixels from the trigger constructed over both detectors, showing the contribution to the total energy in the data stream



(a) Time-Frequency spectrogram in Hanford data



(b) Time-frequency representation of reconstructed trigger

Figure 5.21: Time-Frequency representation of the loudest non-CBC trigger found by the cWB+GMM search in O4a. (a) shows the time-frequency plot of Hanford data surrounding the time period concerned with the trigger, with normalised energy on the z-axis. (b) shows the time-frequency representation of the processed trigger over likelihood. The black curve shows cWBs attempt to fit a pre-defined TF template to the trigger, in this case a ringdown template. Both plots are taken from the cWB generated coherent event display.

		iFAR [year]		
Event Name	Chunk	2G-GMM	2G-XGB	XP-XGB
S230627c	K03	1135.27*	354.15	486.60*
S231226av	K16	1001.01*	870.32*	491.83*
S230927be	K10	608.75	252.78	467.77*
S230914ak	K09	608.75	201.73	502.98*
S231206cc	K15	250.25	416.62*	481.90*
S231028bg	K12	192.51	555.91*	481.43*
S231123cg	K14	104.33	488.42*	481.75*
S230601bf	K01	58.21	91.71	516.96*
S230919bj	K09	57.98	57.64	502.98*
S231102w	K12	56.62	92.65	481.43*
S230628ax	K03	49.36	354.15	243.30
S230922q	K09	2.8	6.5	0.025
S230914an	K09	1.42	31.04	502.98*

* iFAR estimate is saturated by the available background

Table 5.3: Table of CBC detections with iFAR over 1 year in the cWB+GMM search. For each event the iFAR found by the cWB2G-XGBoost and cWBXP-XGBoost are also shown, alongside the corresponding chunk where the event was found. Note that the highest available iFAR per chunk is dependent on the accumulation of available background, which varies per chunk and per pipeline. Events which have saturated the available iFAR estimation are noted by *, and such significance estimates cannot be directly compared between pipelines.

associated to the reconstructed trigger. The black curve shows cWBs attempt to fit a pre-defined TF template to the trigger, in this case a ringdown template, which does not appear to follow the distribution of pixels in the reconstructed trigger well.

Searching for known glitches which occurred within this time frame in gravity spy, it is found that a scattered light glitch with duration 1.99s occurs in Hanford 1 second earlier than the observed trigger. No glitch has been labelled for Livingston. It is possible that the observed trigger is some artefact from the remnants of the scattered light glitch.

From the above initial investigations into the properties of this trigger, the loudest non-CBC trigger found by cWB+GMM is likely to be a glitch, however further investigations into the origin of the trigger have the potential to alter this statement.

Detection comparison to other pipelines

With a null conclusion on the detection of a burst-type GW event in O4a, we turn our attention to the CBC detections made by the cWB+GMM pipeline, comparing significance with other pipelines which ran on the search. In total 3 pipelines analysed O4a data in the search for short duration low-frequency transients, all varying flavours of cWB. As introduced previously in section 4.3, there exists two versions of cWB trigger production, ‘2G’ and ‘XP’, which use different techniques to identify coherence in data across the detectors, as well as other minor

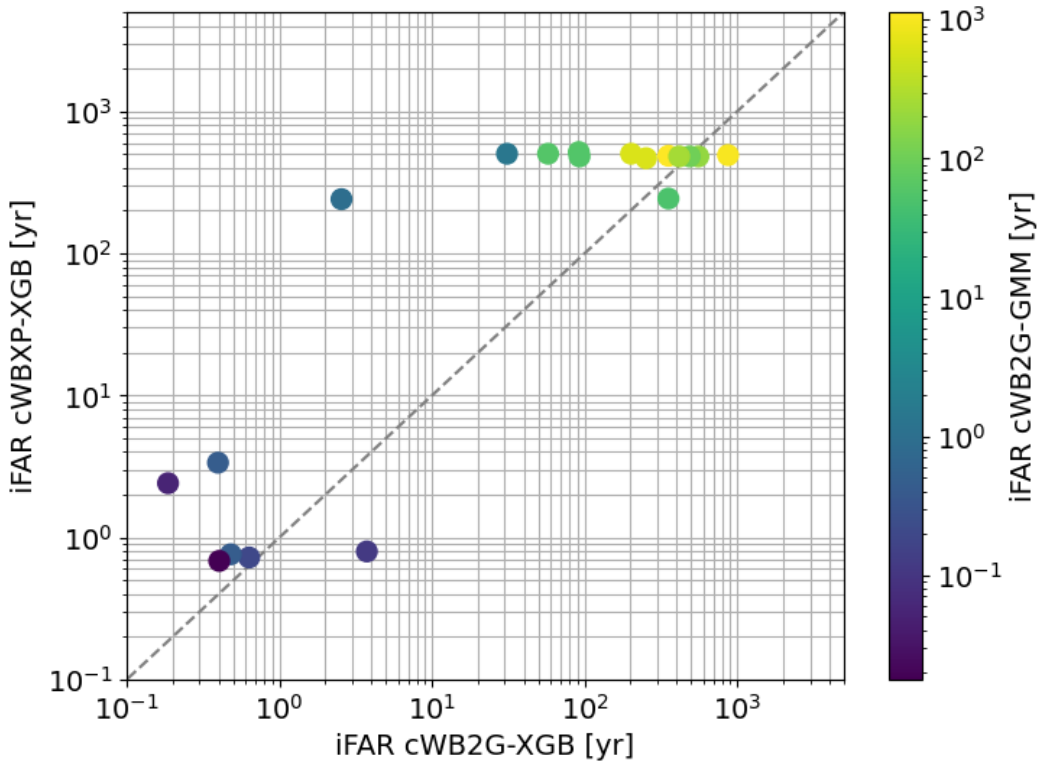


Figure 5.22: Comparison of iFAR over the three cWB pipelines in O4a for GW events with iFAR $\geq$ 0.1 years. iFAR of cWB2G+XGBoost in shown on the x-axis, iFAR of cWBXP+XGBoost on the y-axis, and iFAR of cWB+GMM in shown in the z-axis via colour bar.

differences. There also exist two versions of machine learning based post-production, 'GMM' and 'XGBoost' (XGB). Through combinations of these alternative methodologies we obtain the 3 cWB pipelines which ran in O4a: cWB2G-GMM (detailed throughout this thesis), cWB2G-XGBoost [124], and cWBXP-XGBoost [175, 176].

Of the 46 CBC events observed by cWB2G in O4a, GMM detected 13 with high significance of iFAR $\geq$ 1 year, compared to 14 for the cWB2G-XGB search and 14 for the cWBXP+XGB search. The 13 high significance detections from GMM are listed in Table 5.3 by decreasing iFAR, alongside the iFAR estimates found by cWB2G-XGB and cWBXP-XGB analyses for the same events. Note that due to differences in time-shifting, search configuration and percentage splits chosen for each of the 3 pipelines, the maximum available background will vary across the pipelines for the same chunk, affecting the maximum possible iFAR assignment of each pipeline. With this in mind, 2G-GMM and 2G-XGB detect most events with iFAR in similar orders of magnitude, while XP-XGB detects more events with higher iFAR $>$ 100 years.

The variation in iFAR can further be seen in Figure 5.22, where iFAR of the same detected CBC events is reported for 2G-XGB vs. XP-XGB vs. 2G-GMM on the z-axis. Majority of the events are detected with iFAR in the same order of magnitudes for all pipelines, however GMM

iFARs are overall slightly lower than others and cWBXP-XGB achieves the best sensitivity to CBC sources. From both Figure 5.22 and Table 5.3, it is clear that the cWB+GMM pipeline is slightly less sensitive to CBC signals than the other cWB-based pipelines. The GMM analysis detects less events at iFAR $\geq$ 1 year than both other pipelines and overall detects such events with slightly lower significance, especially compared to cWBXP-XGB. This behaviour is not surprising, as the same statement was observed during the O3 search. Similarly to the conclusions made previously, the lowered sensitivity of GMM to CBC events is somewhat expected since it is not a CBC-targeted search, and is trained towards the morphology of generic burst events. The CBC detections it is able to make can aid in cross-checking such detections made by the CBC-based pipelines, and can assist in the detection of GWs arising from less common sources of CBC which may not be well represented in matched filtering searches.

5.6 Summary and discussion

This Chapter has detailed the application of cWB+GMM as an offline search for short duration GW transients in the first half of the fourth LVK observing run (O4a). The improved sensitivities of the detector and different set-up in analysis required by an offline search lead to numerous investigations into the pipeline performance and optimisation. Additionally, running a full search in near real-time provided a lot of insight into the behaviour of GMM and highlighted improvements to be made in the future, both in terms of methodology and computational efficiency.

Due to the nature of offline searches, the configuration of the analysis was adapted to analyse data in a chunk-by-chunk manner, rather than processing the entire observing run at once. Investigation into the configuration of the analysis found that it was beneficial to analyse a combination of two chunks in unison, rather than individually, due to an increase in computational efficiency and better mitigation of loud glitches. It is possible that the increased glitch mitigation is due to a larger training set for the background model. In addition to this, optimisation studies were carried out to enhance the analysis for O4a data. Firstly, a new signal training data set was introduced, which provided denser sampling across the Q_{veto} parameter space. The region affected by this new data set contains a sub-population of high Q SG and CCSN waveforms, making it more robust across the signal space. It was also noticed that the re-parameterised N_{ED} attribute was not visibly distinguishable between the background and signal parameter space, hence a new re-parameterisation was tested. While the new re-parameterisation considered here did not provide a sensitivity improvement, it suggests that a more in-depth study on optimal re-parameterisations for all attributes may be beneficial to the analysis. Such a study could not take place prior to the O4a analysis due to time constraints, however is a possible avenue for future development. Finally, a new approach to signal training was investigated in an attempt to improve sensitivity, in which the signal population is represented by two models, one representing a sub-region of the signal parameter space. It was thought that the secondary model could

target regions of the parameter space which were not well distinguishable from noise, somewhat 'boosting' the T value in these regions. This approach was tested with the secondary model being trained on 1) missed triggers from the first model, and 2) the low Q_{veto} region. These examples did not provide significant sensitivity improvement, however gave promising results for the secondary model approach, and should be studied in further detail in the future.

Loud glitch investigations occurred throughout the run, observing T vs. frequency, time, and network correlation coefficient for each set of combined chunks. Such investigations were crucial for ensuring the behaviour of the pipeline was as expected, and provided potential avenues for further glitch mitigation. For example, providing a summary of loud glitch studies to the detector characterisation group allowed for the production of a newly defined veto, which targeted a number of loud triggers present in the GMM analysis. After the newly defined veto had been applied, we summarised the effect of glitches by observing the T values associated with iFAR thresholds across all chunks, seen in Table 5.1. It was shown that T at iFAR$\geq$10 years is consistent, however thresholds at iFAR$\geq$100 years significantly varied over chunks due to the presence of loud glitches. Despite the ability of GMM to mitigate blip glitches, these were still the most prominent glitch throughout the O4a analysis. The loudest glitches for each chunk also often involved koi fish, tomte, and extremely loud glitches. Future in-depth studies into the populations of loud glitches in GMM and where they lie within models may provide useful information for improving sensitivities further.

The loud glitch investigations lead to the discovery of a set of outlier glitches in K11+K12, which were causing poor sensitivity estimates for these chunks. These glitches had properties which were not common of loud glitches in the analysis, such as low network SNR, clustered in frequency and low network correlation coefficient. Their presence lead to various investigations into the behaviour of the GMM models, such as regions of the attribute space where loud glitches often occurred, comparisons to 'good' performing models and investigations into the distributions of log-likelihood values. Despite their cause later being identified as a problem in cWB trigger production, it is clear that there is some aspect of unreliability in the GMM models that caused the drastic change in results from a comparatively small number of faulty triggers. Undoubtedly the problems arising with K11+K12 models gave valuable insight into the behaviour of GMM, and even if not fully understood, have provided many avenues of investigation for the future.

The finalised sensitivities of the cWB+GMM search to ad-hoc and CCSN waveforms were shown over chunks throughout O4a. Overall, chunks towards the end of the run were less sensitive due to a spike in loud glitches, causing failed h_{rss50} estimates for a handful of CCSN waveforms in K15+K16. Aside from this, the sensitivity estimates give insight into areas of the parameter space where the GMM analysis may improve. For example, looking at results from the ad-hoc waveform injections, it was seen that the search is consistently the least sensitive to low frequency SGs with $f < 70$Hz, while high frequency SGs with $f > 1000$Hz were

also observed with comparatively poor sensitivity compared to others. This suggests that such waveforms may be under-represented by the models, and further updates to signal training data may assist with this in the future. With this said, sensitivity estimates combined over O4a were observed for a selection of ad-hoc and CCSN waveforms compared to those observed in O3, credited to both enhancements in the search techniques and upgrades to the detectors themselves.

Having summarised all relevant steps of the offline analysis throughout the run, the zerolag analysis was completed in the search for short duration GWs. It was realised during closed box analysis that all time-shifted segments had been considered in the background data split, resulting in only $\sim$70% of a lag being available for closed box analysis. Comparisons to closed box analysis with a full lag demonstrated that considering only $\sim 70\%$ was still reliable, however highlighted that a selection of full lags should be reserved for closed box analysis in future observing runs in the interest of complete accuracy. The cWB+GMM analysis recovered a total of 46 CBC GW events during O4a, 13 of which were detected with high significance of iFAR$\geq$1 year. Once parameter estimation has been completed on these events, it will be possible to investigate whether there is a region of the CBC parameter space which the GMM search is more sensitive to, or whether there is any possibility of the high significance events arising from an unusual CBC system. The loudest non-CBC event occurred with an iFAR of 1.75 years, however is likely to be a noise artefact based upon it's morphology in the TF-plane and it's 1 sigma relation with the expected background. Comparisons of CBC detections with other all-sky short search algorithms demonstrated that the GMM search detected a similar total number of significant events as the other pipelines. In addition, it was found that the cWB+GMM search observes CBCs with slightly less significance. This is consistent with what was found in Chapter 4, and as mentioned, the sensitivity to CBC could possibly be increased with future development if desired.

To conclude the cWB+GMM offline search for short GW transients during O4a did not confidently detect any burst-type events. 13 significant CBC detections were made, and full comparisons to targeted CBC pipelines may be made in the future. The search achieved high sensitivities to simulated waveforms, which may improve upper limit rate estimations with the injection of waveforms from astrophysical sources. Overall, the search provided valuable insight into the behaviour of GMM and highlighted various avenues for improvement, such as an approach involving a secondary GMM model, studies into the populations of loud glitches, and understanding how changes in training data can alter the overall performance of models. Additionally, aspects of analysis which may benefit from additional computational efficiency and automation were highlighted.

Chapter 6

Exploring the sensitivity to parabolic black hole encounters

Eccentric encounters of compact objects may result in gravitational wave (GW) signals with different morphology than seen from standard compact binary merger signals due to interactions in dense astrophysical environments. First introduced in section 2.1.1, such interactions will result in one of 2 phenomena: *fly-by encounters*, in which compact objects emit gravitational waves (GWs) as they pass by each other without merging, or *radiation-driven capture*, in which objects become bound through the emission of GWs and merge on extremely quick timescales. Due to their quick merger time, and hence lack of inspiral, radiation-driven capture events are not guaranteed to be detected with compact binary coalescence (CBC) matched filter searches [74]. A more reliable method for the detection of such signals is the weakly-modelled burst searches, since search algorithms are designed to make minimal assumptions on waveform morphology and the expected frequency and duration of capture events places them directly into the parameter space of the all-sky short search. In this Chapter, we will explore sensitivities to radiation-driven capture of black holes (BHs) following parabolic orbits with current coherent WaveBurst (cWB) analysis tools used in the search for short duration transients. The detection of such objects would allow insight into the formation dynamics of capture systems, and even in the case of non-detection, a search can place upper limits on the expected rates of such events.

We base our search sensitivity studies for capture events upon data from the third LIGO-Virgo-KAGRA (LVK) observing run (O3). O3 was split into 2 epochs running from April 1st 2019 to October 1st 2019 (O3a) and November 1st 2019 to March 27th 2020 (O3b). We consider only the LIGO-Livingston, LIGO-Hanford (LH) network, in which a total of 198.3 days of coincident data was accumulated throughout the entire run after periods of poor data quality are removed.

A brief overview of how radiation-driven capture events form and details of the waveforms utilised for this study are given in section 6.1. The sensitivity to parabolic capture events is then explored through 3 cWB post-production approaches. First, the sensitivity study with standard

cWB is detailed in section 6.2, exploring the sensitive distance and upper limits on rates based upon collaborative work in [4]. The work is extended further in section 6.3, where we explore the ability of Gaussian mixture modelling (GMM) to enhance sensitivities to parabolic capture events through the application of two GMM approaches as post-production to cWB: the generic all-sky short GMM approach is considered to explore whether current weakly-modelled searches are capable of detecting capture events, and for the first time a targeted GMM approach is considered with the aim to specifically enhance sensitivities to radiation-driven capture events. Finally, initial studies into the properties of parabolic capture waveforms are completed in section 6.4, in an attempt to understand downfalls of the current GMM models for future improvement.

6.1 Formation of radiation-driven capture systems

This section gives an overview of how radiation-driven capture systems form so that the reader obtains an understanding of such events before investigations are detailed in later sections.

Eccentric binaries of compact objects are often formed through dynamical evolution scenarios within dense astrophysical environments such as globular clusters and galactic nuclei [143, 139]. The exact mechanisms occurring in these scenarios can vary, however often involve the interaction of single or binary systems with another astrophysical object in the environment. These interactions may eventually lead to phenomena known as radiation-driven capture, in which 2 compact objects become gravitationally bound through the loss of gravitational energy emitted during interactions. Radiation-driven capture results in objects merging on very quick timescales with no inspiral phase, which in some cases causes the resulting GW signal to resemble a high mass quasi-circular BBH [144]. However as systems approach the boundary between capture and fly-by, the resulting signal may have a different morphology than CBC systems. The short timescale between formation and merger of the binary also results in the system maintaining high eccentricity at the time of merger, however in some cases it is possible for capture waveforms to slowly radiate away eccentricity and obtain a quasi-circular orbit.

There exists many explorations on the formation of such radiation-driven capture systems in literature, such as single-single interactions [182, 183, 140, 184], binary-single interactions [141] and binary-binary interactions [142]. Single-single interactions have the highest probability of resulting in a radiation-driven capture event, with it estimated that around 26%-50% of such systems will form above 10Hz with high eccentricity, which lies in the LVK detectable band [183]. While literature exists on the variety of possible formation mechanisms, accurate waveform models are often scarce due to complicated dynamics.

For the following studies we use parabolic radiation-driven capture waveforms from Bae et al. [84]. These are numerical relativity (NR) simulations of non-spinning BHs chosen to follow the parabolic approximation, which assumes that a weakly hyperbolic orbit emits the same amount of energy in a parabolic orbit with the same distance at the point of closest approach.

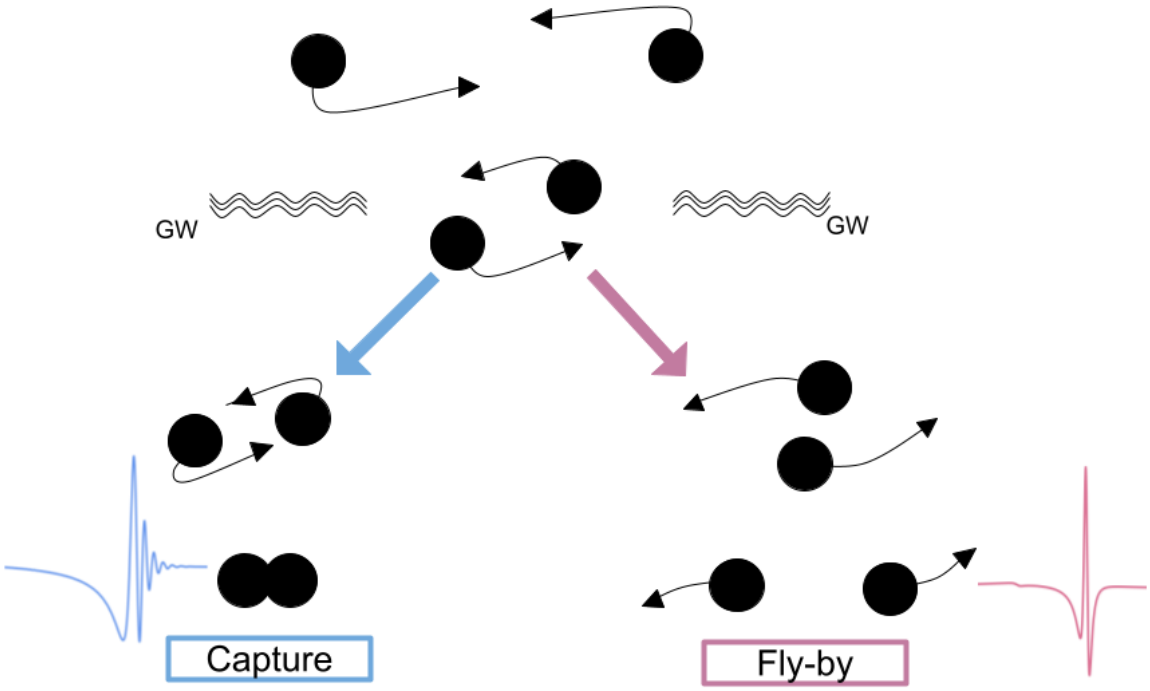


Figure 6.1: Simplified diagram showing the difference in orbits leading up to a capture event (left) vs. a fly-by event (right). The route which the compact objects take is dependent on parameters such as initial angular momentum and mass ratio.

Under these parabolic orbits systems have initial orbital energy of zero, thus simplifying the required simulations. Whether the two BHs become gravitationally bound to the point of merger is dependent on the radiated energy as the objects interact, related to the initial angular momentum L, impact parameter, reduced mass and relative velocity of a system, detailed in [84]. For the purposes of describing the waveforms in this study we state only the initial angular momentum and mass ratio of waveforms. Due to the relation with reduced mass, the initial angular momentum threshold between capture and fly-by scales with mass ratio of systems, decreasing as the mass ratio of the system increases. In systems with initial angular momentum above the thresholds, the two objects will undergo a fly-by encounter, emitting GWs as they interact, however will continue on their paths without merging. It is possible that in this case the objects are in fact still bound, however the apocenter of the systems is too large to be simulated. For initial angular momentum below the boundary, the objects will undergo radiation-driven capture. As L decreases, the BHs undergo capture on shorter timescales, almost directly merging for the smallest considered L. In regions close to and including the boundary, the objects will interact via a close encounter before merging, causing double blip waveforms.

In this study only capture waveforms which directly merge or are close to the initial angular momentum boundary are considered, since these systems radiate the most GW energy. It can be assumed that these waveforms follow the single-single interaction formation channel detailed in [183, 185]. Examples of the considered waveforms are given in section 6.2.1.

6.2 Search sensitivity of radiation-driven black hole capture events with standard cWB post-production

As mentioned above, the waveform morphology of eccentric radiation-driven capture events varies from quasi-circular BBHs due to the short merger timescales and lack of inspiral, which may make their detection in CBC matched filter pipelines less likely. However, the minimal assumptions made in burst pipelines makes them sensitive to such sources, and the short duration and low frequency properties of the GW signals arising from capture events places them directly within the detectable range of the all-sky short low-frequency search. This section details a study on the search sensitivity of parabolic radiation-driven capture waveforms with the 'standard' cWB algorithm with O3 data. The work presented was completed as part of a collaborative study and is published in [4].

Within this study, the sensitivity to capture events is estimated with the 'standard' version of cWB post-production, in which manual selection cuts and binning of events based on morphology is applied in order to minimise the effect of glitches, described in section 2.2.4. Only the 2 detector LIGO-Livingston, LIGO-Hanford (LH) detector network is considered since the detectors are nearly co-aligned and can mitigate a portion of glitches based upon their polarisation (see section 2.2.2), totalling to 198.3 days over O3a+O3b. The background is estimated through time-shifts on the coincident data, accumulating a total of 980.7 years to use for FAR estimation in O3a, and 1096 years in O3b. Since the study utilises the same cWB settings as used in the LVK search for short transient GWs in O3 [85], it is already concluded that all significant GW events detected are known CBC systems identified in [9, 10]. Sensitivity estimates are given in terms of sensitive distance through simulated injections into cWB, and the non-detection of any radiation-driven capture events stated above is used to place upper limits on rate estimates of these sources for the first time.

6.2.1 Simulations of radiation-driven capture in cWB

As described in section 6.1, this study utilises NR simulations of parabolic non-spinning BH capture waveforms from Bae et al. [84]. We consider 14 waveforms varied over mass ratio of $q = m_1/m_2 = [1,2,4,8]$, with each mass ratio including waveforms from systems with different initial angular momentum L up to approximately the capture threshold limit associated with each mass ratio. The above 14 waveforms are converted into time-series through double integration and scaled to desired total masses of $M = [20,50,100,200]\mathrm{M}_{\odot}$, yielding a total of 56 different capture waveforms for consideration in the study.

An example of each waveform is seen in Figure 6.2 in time-series for total mass of $100\mathrm{M}_{\odot}$ at distance of 100Mpc. As mentioned, for small initial angular momentum the BHs almost directly merge, causing immediate ringdown-type waveforms with no inspiral. As initial angular momentum of the systems increases, more orbits occur before merger causing additional oscil-

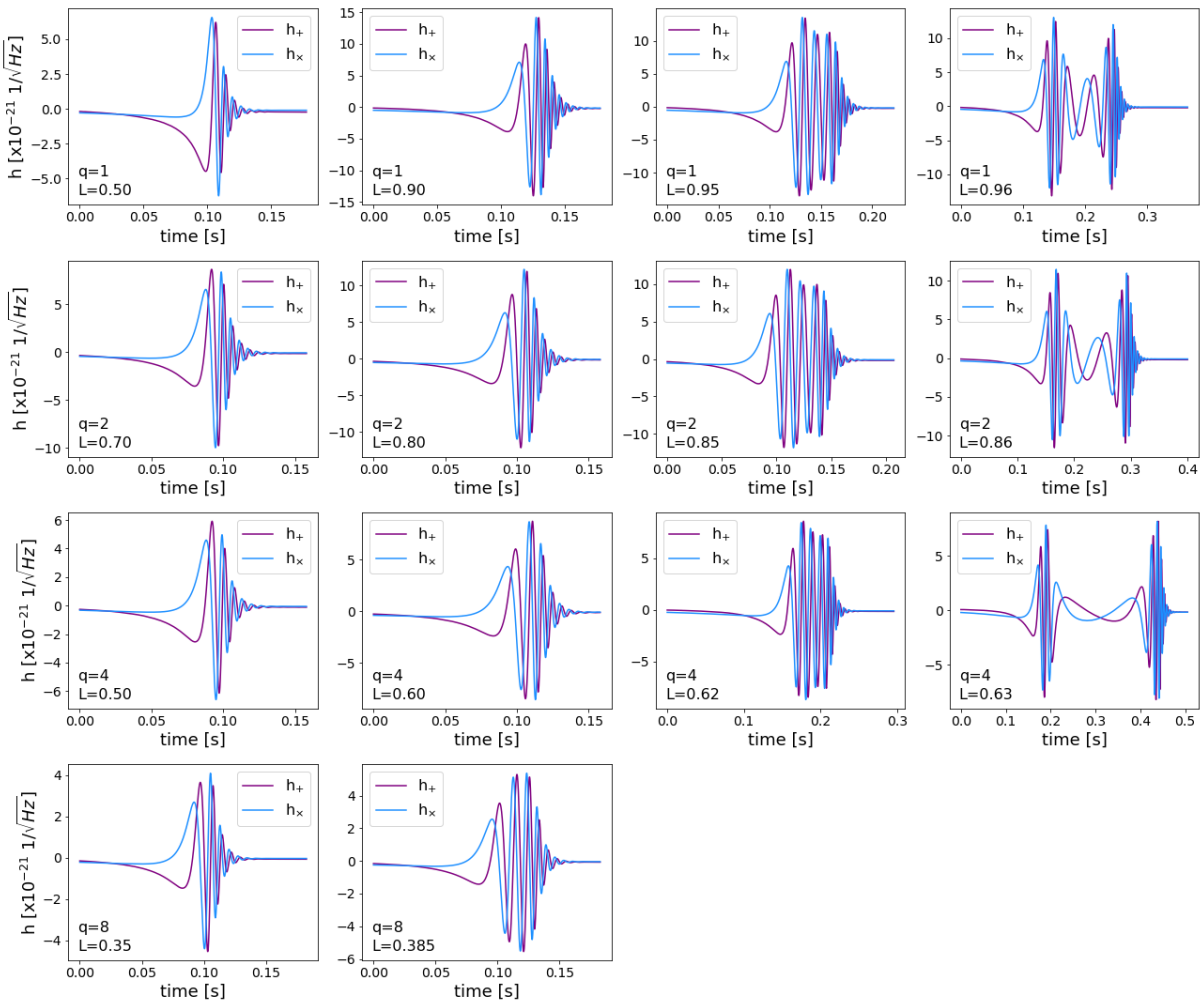


Figure 6.2: Time-series representation of the considered set of parabolic radiation-driven capture waveforms for h_+ (purple) and $h_\times$ (blue) polarisation, re-scaled to have total mass $100M_\odot$ at a distance of 100Mpc. Each row represents a different mass ratio from $q = [1, 2, 4, 8]$, with initial angular momentum L increasing from left to right.

lations before the ringdown. Lastly, looking at the highest considered initial angular momentum values near the fly-by boundary, it is seen that the first interaction of the BHs causes a blip-like structure similar to a hyperbolic encounter before the BHs are pulled back in to merge, causing ringdown. We refer to such waveforms as having 'double blip' structure. The above behaviour is seen for initial angular momentum values across all mass ratios. The waveform simulations are injected into cWB over the entirety of the O3 run, uniformly distributed in sky location and inclination angle. Each waveform is uniformly distributed in co-moving volume up to a maximum redshift value which is individually chosen for each waveform to ensure injections are made only within reasonable detection ranges.

6.2.2 Sensitive distance and upper limit rate estimates

As mentioned above, the standard version of cWB is utilised for this study, as described in section 2.2.4. Thus in order to distinguish the injected signals from detector glitches, triggers from both the background and simulation data sets were divided into 3 manually chosen classes based upon morphology in order to limit the effect of loud blip glitches. For O3 these classes were defined as *C1* containing triggers with a single cycle, *C2* containing triggers with multiple cycles and low Quality factor ($Q < 3$) and *C3* containing triggers with Quality factor > 3. The FAR is estimated for triggers within each bin separately, with a trials factor of 3 taken on final significance estimates to account for this.

In order to quantify the sensitivity to the considered parabolic BH capture events, we define the sensitive space-time volume and the sensitive distance which give measures of the current reach of detectors and pipelines to such sources.

The averaged space-time volume is a measure of the spherical volume in which we can expect to detect GW events from a given astrophysical population at a given significance threshold, and is defined as [186]:

$$\langle VT \rangle = \int dz d\theta \frac{dV_c}{dz} \frac{1}{1+z} p_{pop}(\theta) f(z,\theta) T \tag{6.1}$$

where dV_c/dz is the differential co-moving volume, p_{pop} is the distribution of binary parameters θ, $f(z,\theta)$ is the probability of recovering a signal at given redshift with parameters θ, and T is the length of the observation in the detector frame. In our case, we consider $p_{pop}(\theta)$ to be uniform for sky location and inclination angle, while sensitive distance is calculated for each considered total mass, initial angular momentum and mass ratio. $f(z,\theta)$ is the efficiency of recovering a given waveform injection at a chosen iFAR threshold over binned values of redshift z. The results below all consider a detection threshold of iFAR $\geq$ 100 years.

From sensitive volume we can calculate the corresponding sensitive distance, which represents the maximum distance at which we can expect to observe GWs from a source population at given significance. This is defined as:

$$D_{\langle VT \rangle} = \left(\frac{3\langle VT \rangle}{4\pi T_s} \right)^{1/3} \tag{6.2}$$

where T_s is the length of detector data analysed by the search, in this case the length of O3 coincident data.

The distribution of sensitive distance at iFAR $\geq$ 100 years over total mass for each waveform is seen in Figure 6.3. We see that for all waveforms we obtain the largest sensitive distance for high mass systems of $M_{tot} = 200\mathrm{M}_{\odot}$, with sensitivity decreasing as we go to lower masses. This behaviour is expected since high mass systems release the most gravitational energy, with the majority of their signal strength in the merger-ringdown region of the waveform. We can also

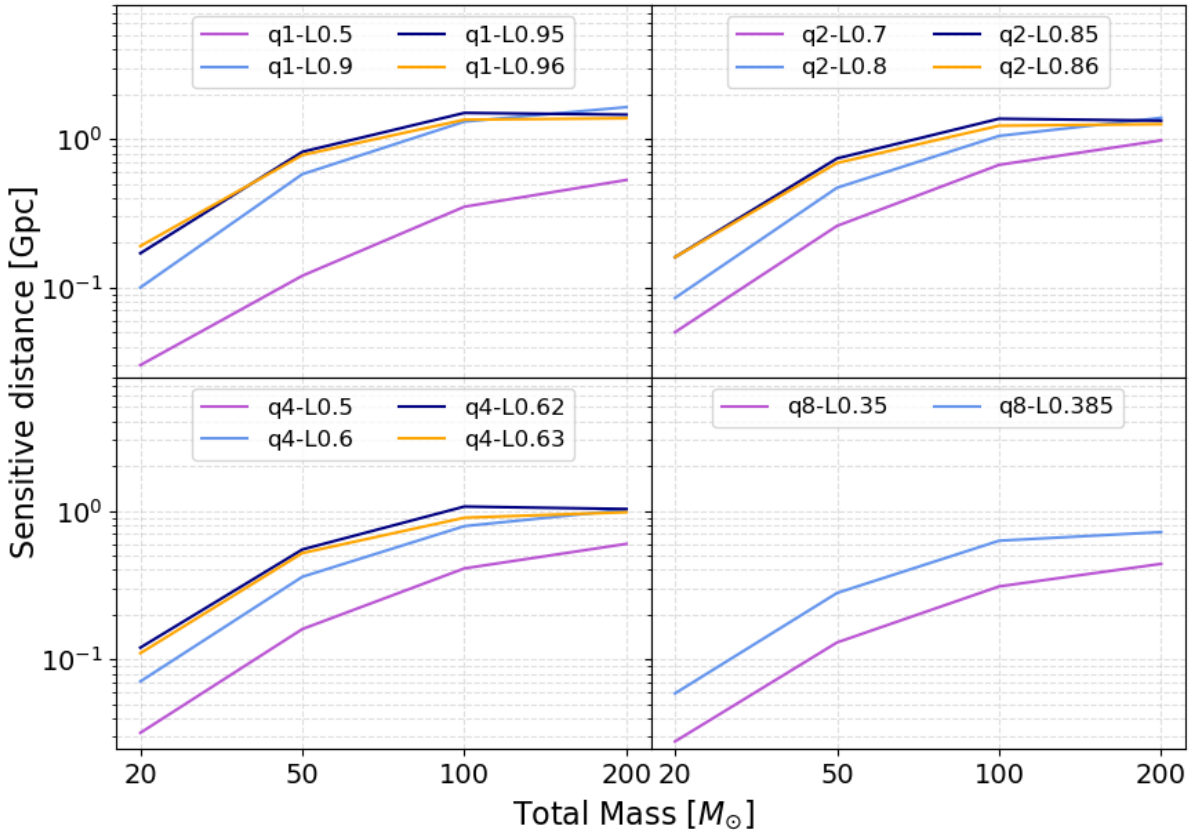


Figure 6.3: Distribution of sensitive distance at iFAR $\geq$ 100 years for the standard cWB analysis, over total mass $M = [20, 50, 100, 200]\mathrm{M}_\odot$ for each individual parabolic capture waveform. Results are split by mass ratio, with $q = 1$ waveforms in the top left, $q = 2$ waveforms in the top right, $q = 4$ waveforms in the bottom left and $q = 8$ waveforms in the bottom right. For each mass ratio, the lowest initial angular momentum L is shown in purple, with increasing L displayed in light blue, navy and orange respectively.

observe the relationship between sensitivity and initial angular momentum L by looking at the colour-coded waveforms across mass ratios. Across all total mass and mass ratios, the lowest sensitive distance is achieved for low L waveforms, which are the systems where the BHs almost directly merge. These waveforms have low Quality factor meaning they are categorised into the noisy bin in cWB post-production, hence are affected by the high glitch rate. Sensitive distance increases with L for each group of mass ratio values, with the two waveforms with largest L achieving similar sensitivities across all total masses.

The maximum sensitive distance is achieved for $200\mathrm{M}_\odot$, $q = 1$ systems with $D_{\langle VT \rangle} = 1.64$Gpc, which is a similar sensitive reach achieved to quasi-circular BBH systems during O3 [113]. During peak GW emission in the merger-ringdown phase, high mass capture waveforms have similar morphology to quasi-circular waveforms. The minimum sensitive distance is achieved for $20\mathrm{M}_\odot$, $q = 8$ waveforms with $D_{\langle VT \rangle} \approx 0.045$Gpc, however all $20\mathrm{M}_\odot$ achieve comparatively low sensitivities compared to higher masses. A full list of the sensitive distance

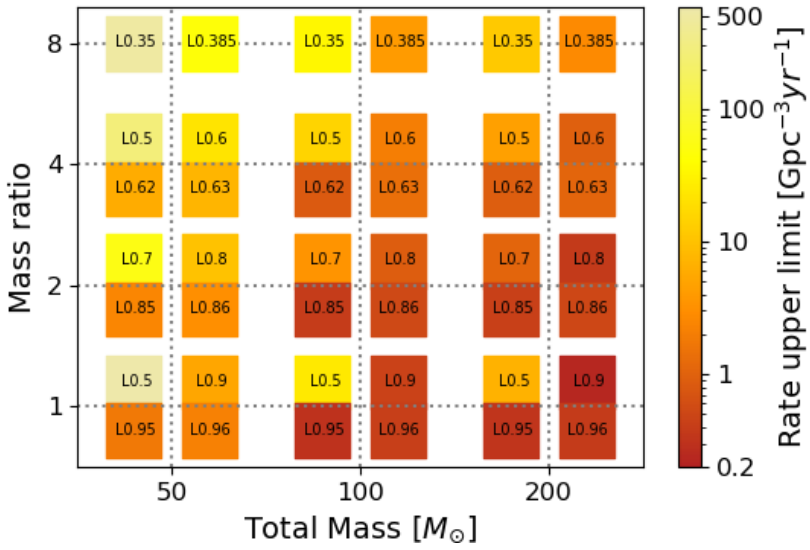


Figure 6.4: The rate upper limit estimates at 90% confidence interval for parabolic capture events given cWB standard analysis at iFAR threshold $\geq$ 100 years. Total mass of 20M$_\odot$ is excluded as the low sensitivity does not lead to competitive upper rate estimates.

estimates for each waveform type at each total mass is seen in Table 6.1.

The above definition of sensitive space-time volume also allows us to define upper limits on the astrophysical rate of parabolic BH capture events based upon their non-detection. Assuming the population of considered sources produces a Poisson distribution with rate R, the upper limit on rate at a 90% confidence interval for non-detection at a given iFAR is [187]:

$$R_{90} = \frac{2.303}{\langle VT \rangle_i^{iFAR>x}} \tag{6.3}$$

where $\langle VT \rangle_i^{iFAR>x}$ is the sensitive volume for a given injection set i assuming a detection threshold of iFAR $> x$. The value of 2.303 arises directly from the 90% confidence interval through $ln(1-p)$ for probability p.

As mentioned above, the results of the O3 all-sky search for short duration transients concluded with only known CBC detections noted in the O3 GW catalogues [9, 10, 11], with a null result for parabolic radiation-driven capture events. Hence, we calculate the upper limits on the expected rates of the considered capture waveforms across all total masses based upon the above estimates of sensitive distance at iFAR $\geq$ 100 years. The results are seen in Figure 6.4 for total masses 50-200M$_\odot$. Total mass of 20M$_\odot$ is excluded from the figure since the lowered sensitive distance did not produce competitive rate estimates. The most stringent upper limit estimate is obtained for the 200M$_\odot$, $q = 1$, $L = 0.9$ capture system at 0.23Gpc^{-3}yr^{-1}. Referring to literature surrounding the formation of radiation-driven capture through single-single interactions, the rate of such events is expected to be 0.002-0.04Gpc^{-3}yr^{-1} [185]. Hence the upper limit on

rate obtained from this study is at least an order of magnitude too large compared to theorised estimates, and sensitivities will have to improve by $\sim$ a factor of 2 before more stringent rates can be provided. Such a sensitivity increase may be observed in the observing runs of O4 or O5.

The above study into the sensitivity of the weakly-modelled cWB pipeline to radiation-driven capture events proves that it is possible to detect GWs from such sources, with standard cWB achieving similar sensitive distance estimates to high mass captures as quasi-circular high mass BBHs during O3. However with recent Machine Learning based updates to cWB post-production it is possible that sensitivity estimates during O3 can be improved. This will be explored through the application of GMM in the following sections.

6.3 Enhancing sensitivity to radiation-driven black hole capture events with GMM post-production

In this section we attempt to enhance the sensitivities of the cWB pipeline to parabolic radiation-driven capture events in O3 through the application of GMM post-production. We test the application of GMM through two approaches: firstly we utilise the generic signal model employed in the all-sky short search, holding minimal assumptions on morphologies of GW signals. Secondly, we train a targeted GMM signal model on a portion of parabolic capture waveforms in an attempt to optimise the sensitivity to this population of sources further. The application of the generic all-sky short GMM model allows us to investigate the current sensitivities to parabolic capture waveforms in the generic search for short duration transients, while the targeted GMM approach explores whether the application of a targeted GMM model is able to increase the possibility of detecting a parabolic capture event. Additionally, since this is the first application of a targeted GMM model, it also serves as a proof-of-concept study on how much benefit targeted GMM analysis may bring to searches for other astrophysical populations.

All work in this section utilises the same NR non-spinning BH parabolic capture waveforms from Bae et al. [84] injected into O3 data with cWB as described previously in section 6.2. Thus the same 14 waveforms scaled over total masses of $M = [20, 50, 100, 200]\mathrm{M}_{\odot}$ are considered, totalling to 56 individual injection sets.

6.3.1 Application of the generic vs. targeted GMM analysis

The all-sky short generic GMM model

The first part of the study concerns the application of the O3 GMM models described in section 4.1, so that the sensitivity to parabolic radiation-driven capture events within the typical all-sky short search can be estimated with enhanced cWB. The background considered in the standard cWB study above is the same as used in section 4.1, allowing for the direct use of the already trained and optimised O3 models. Similarly to previous GMM studies on O3 data, models were

trained on O3a and O3b separately and results were combined to present sensitivity across the entire O3 run. The background data was split into 10% validation, 70% training and 20% testing, reserving 980.7 years of background for FAR estimation in O3a, and 1096.0 years in O3b. The signal model was trained on the generic WNB simulations detailed in section 3.2.2, consisting of two distributions, one ranging over frequency [24,996]Hz, bandwidth [10,300]Hz, duration [0.1,500]ms, and the other over frequency [10,100]Hz, bandwidth 10Hz, duration [0.1,1]ms. 80% of these simulations were used for training of the signal model, while 20% were used for validation of optimal number of Gaussian components. The full set of capture injections were reserved for sensitivity estimations.

The post-production process follows the methodology described in section 3.1, with selection cuts of $\eta_c > 5.5$ and $c_{c0} > 0.5$ taken on all datasets. The log-likelihood values of the capture injections and background test data are computed with respect to the background and generic signal model in order to calculate T as in eqn. 2.26 and assign FAR estimates. The sensitivity to capture events found with this methodology are explored in section 6.3.2 alongside the targeted capture GMM methodology described below.

The targeted capture GMM model

The second part of the study utilises the GMM post-production methodology with a signal model specifically targeting the parabolic radiation-driven capture events. Thus far the GMM approaches considered in this thesis have concentrated on maintaining a weakly-modelled nature in order to be sensitive to a wide range of signals which may arise in the low-frequency short duration region of the GW parameter space. However, as mentioned in section 2.3, it is possible to tune the GMM methodology to target a desired population of sources by training the signal model on data representing the properties of the desired sources. Hence we construct a parabolic capture targeted GMM signal model in order to optimise the sensitivity estimates to such sources.

The signal model is trained on a portion of the considered 56 capture waveforms injected into O3 data with cWB in section 6.2. Models are again trained separately for O3a and O3b, with estimated sensitivities combined over the entire O3 run. Capture injections output from cWB production are split by total mass before being randomly split into 10% validation, 70% training and 20% testing data sets. Respective types of data set are then combined over total mass to ensure an equal percentage of injections from each total mass is considered. We take the same percentage split for background data, and perform the validation process on both background and signal models in order to obtain a new optimal pair of models compared to the generic all-sky approach. The remainder of the post-production process follows the GMM methodology outlined in section 3.1, taking the same selection cuts of $\eta_c > 5.5$ and $c_{c0} > 0.5$, and assigning T and FAR estimates to each trigger in the test data sets. The only difference in methodology is introduced in the calculation of detection efficiency: since the sensitivity estimates are calcu-

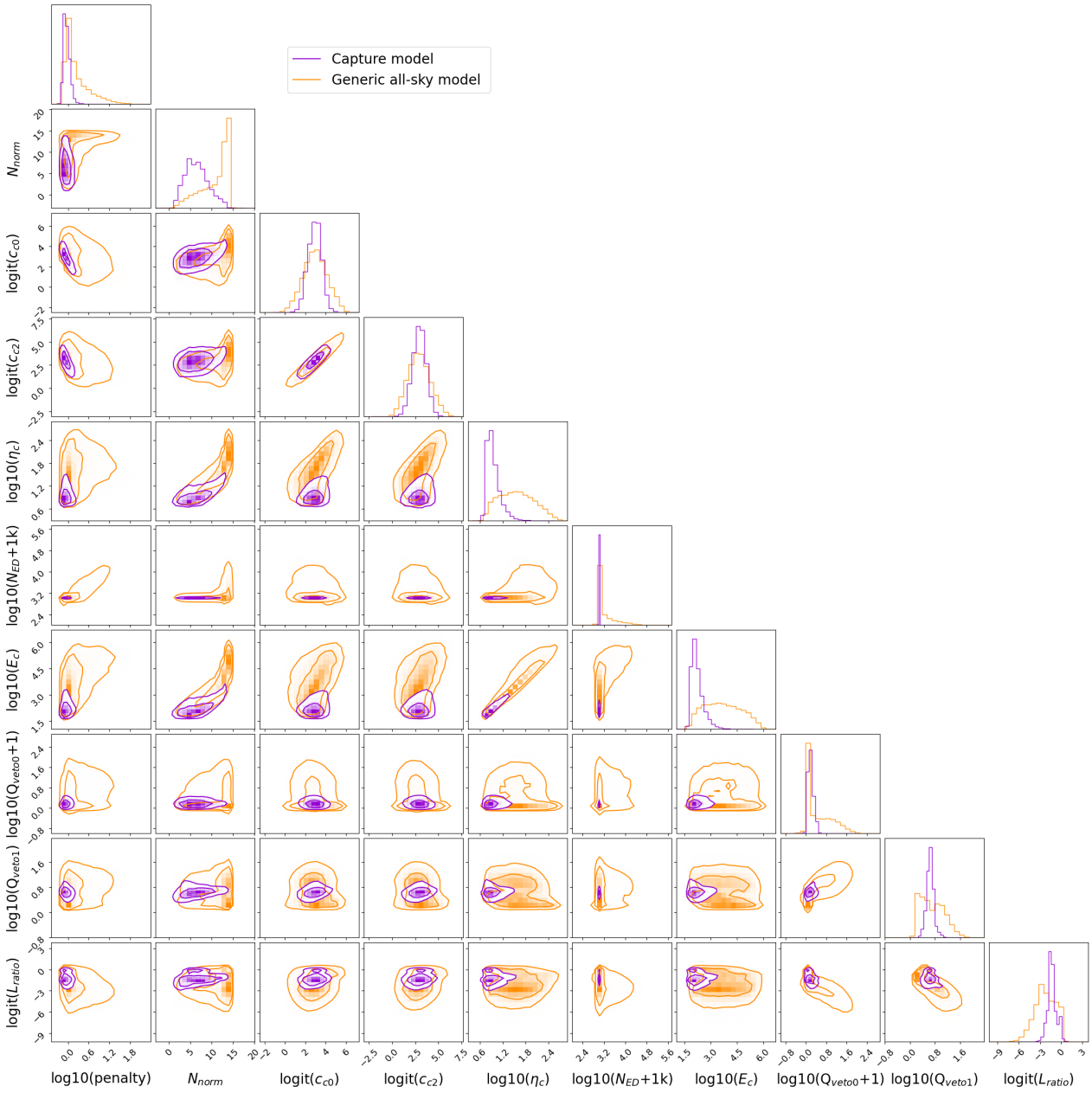


Figure 6.5: Corner plot showing the comparison of the generic all-sky short signal model (orange) and the targeted capture signal model (purple) for O3a over the considered cWB attributes in the analysis.

lated on only 20% of the full parabolic capture injection set, it is difficult to determine how many of the original injections correspond to this 20% test data since not all are recovered by cWB production. To overcome this, detection efficiency must be calculated by computing cWB efficiency and GMM efficiency separately, and multiplying together to obtain the final efficiency, eff_{final} of the 20% capture test set:

$$\text{eff}_{final} = \text{eff}_{cwb} \times \text{eff}_{gmm\_test} = \left(\frac{N_{cwb}}{N_{tot}}\right) \times \left(\frac{N^{>iFAR}_{gmm\_test}}{N_{gmm\_test}}\right) \tag{6.4}$$

for total number of injections into cWB N_{tot}, number of injections recovered by cWB N_{cwb}, number of triggers in the 20% GMM test data N_{gmm_test} and number of triggers from the GMM test data detected over given iFAR threshold $N^{>iFAR}_{gmm\_test}$. This approach to calculating detection efficiency is equivalent to that seen in eqn. 2.15, and hence ensures fair comparison of results.

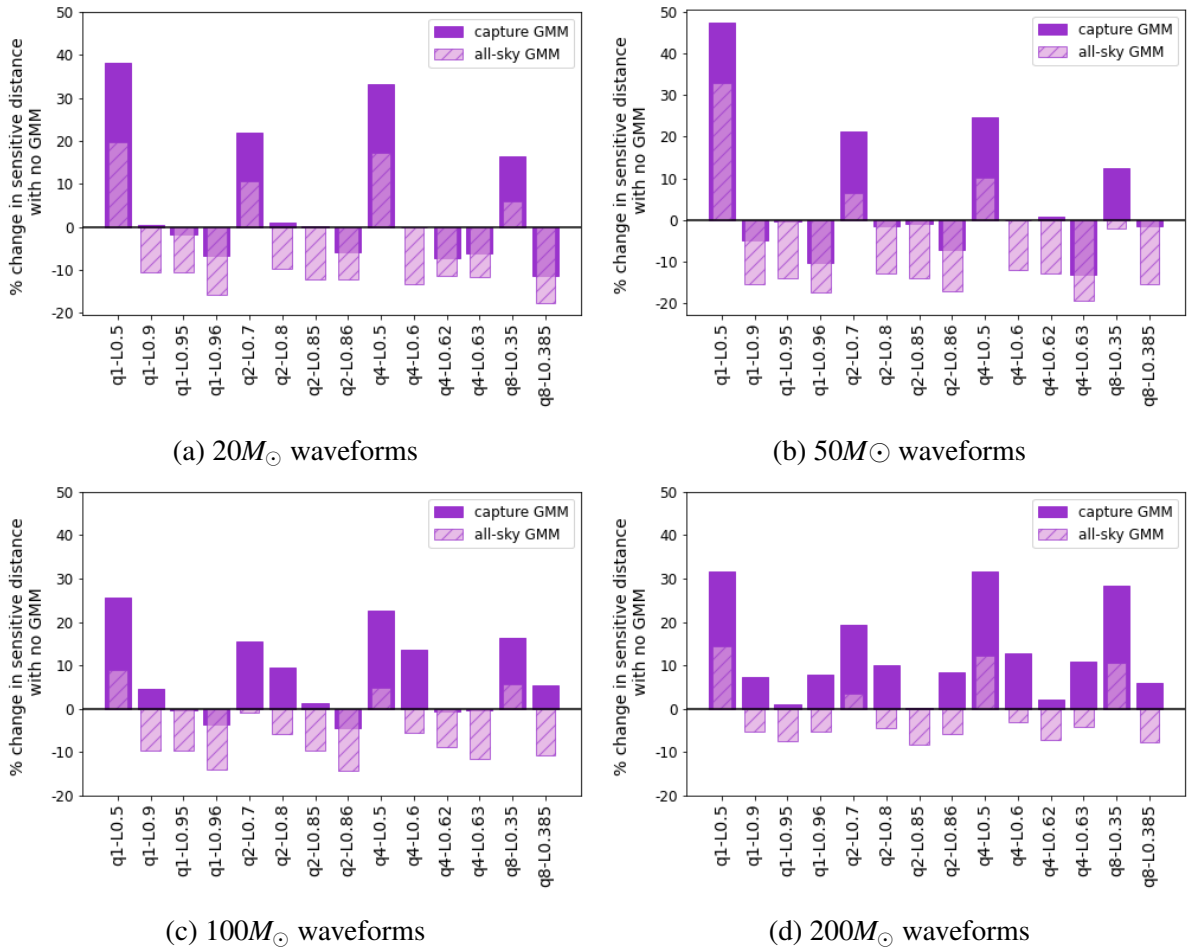
GMM signal model comparison

A comparison of the O3a generic all-sky short and targeted parabolic capture GMM signal models is shown in Figure 6.5 over re-parameterised cWB attributes. The capture model samples a much smaller area of the attribute space compared to the generic all-sky model, which is expected behaviour since the capture model only targets one source of GWs. While the generic all-sky model appears to account for regions of the attribute space associated with capture waveforms, the distribution of density over the mixture models varies which will likely cause lower sensitivities for the generic all-sky approach.

6.3.2 Comparison of sensitive distance estimates

Sensitivity estimates achieved by the two GMM-based approaches described above are quantified through the calculation of sensitive distance as described in eqn. 6.2.2. We again consider distances based upon the detection threshold of iFAR $\geq$ 100 years, estimating values for each waveform individually.

We first compare the performance of each GMM approach against the sensitivity achieved with standard cWB post-production in section 6.2. This is seen in Figure 6.6, where the percentage change in sensitive distance at iFAR $\geq$ 100 years is shown for each GMM methodology with respect to standard cWB. First comparing results across individual waveforms, it is seen that the most significant improvement in sensitivity is seen for the low initial angular momentum systems across all total mass and mass ratios for both the generic and targeted GMM analysis. It was previously noted that the morphology of the low L waveforms placed them in the noisy bin of standard cWB post-production (*C1*) associated with low Quality factor triggers. This suggests that the consistent improvement observed is due to the ability of GMM to mitigate low-Q glitches, such as blip glitches, which has been a particularly consistent result in the application



(a) $20M_{\odot}$ waveforms

(b) $50M\odot$ waveforms

(c) $100M_{\odot}$ waveforms

(d) $200M_{\odot}$ waveforms

Figure 6.6: Percentage change in sensitive distance at iFAR $\geq$ 100 years achieved with generic GMM and targeted GMM analysis with respect to standard cWB from section 6.2. Comparison with targeted capture GMM is shown in solid purple while comparison to generic all-sky GMM is shown in striped lilac. All waveforms are shown for total mass $20M_{\odot}$, $50M_{\odot}$, $100M_{\odot}$, $200M_{\odot}$ separately in (a), (b), (c), (d) respectively.

of GMM throughout this thesis. Despite the improvement seen for low L, the application of the generic all-sky GMM post-production decreases sensitivity compared to the standard cWB analysis for all other waveforms. This decrease in sensitivity suggests some region of the all-sky short model is not well-representative of the high L waveforms, which could be improved with the consideration of new signal training data in future applications to the all-sky short search. With this said, the most significant decrease in sensitivity seen is approximately 20%, which is not as significant as some of the sensitivity improvements seen for other waveforms. Looking at results across different total masses, sensitivity appears to improve more significantly with higher total mass systems, with the $200M_{\odot}$ waveforms seeing the best improvement in sensitivities for both GMM models. There does not appear to be a clear relationship between sensitivity improvement and mass ratio. While the sensitivity observed with the targeted capture analysis shows significant improvement for the majority of high total mass waveforms, the percentage change in sensitive distance decreases for high L waveforms with total mass of $20M_{\odot}$

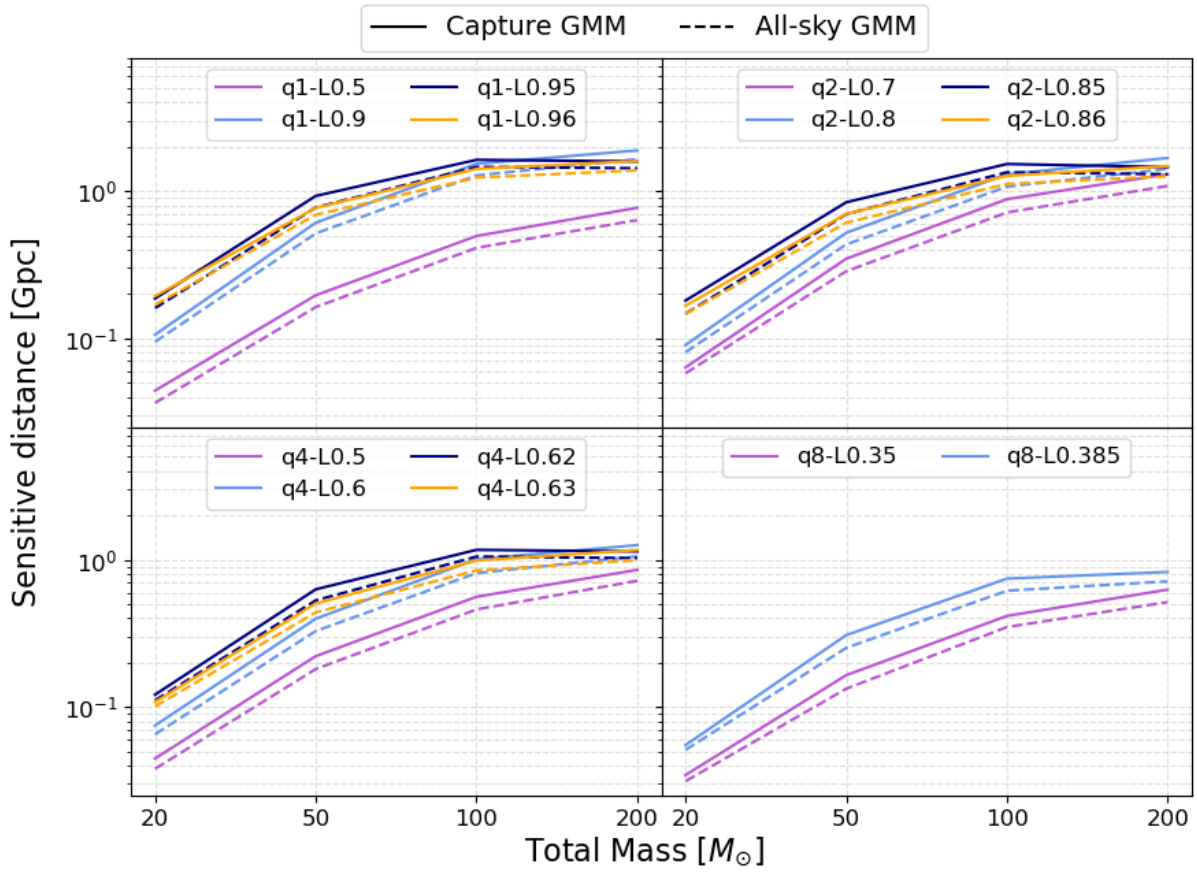


Figure 6.7: Distribution of sensitive distance at iFAR $\geq$ 100 years over total mass for each individual parabolic capture waveform, shown for both applications of GMM. Results are split by mass ratio, with $q = 1$ waveforms in the top left, $q = 2$ waveforms in the top right, $q = 4$ waveforms in the bottom left and $q = 8$ waveforms in the bottom right. For each mass ratio, the lowest initial angular momentum L is shown in purple, with increasing L displayed in light blue, navy and orange respectively. Results for the targeted capture GMM analysis are notated by solid lines, while results for the generic all-sky GMM analysis are notated by dashed lines.

and 50$M_\odot$. This sensitivity reduction is minor at $\sim$10% loss, however suggests that there is potential for further improvement in the targeted model, which is discussed in section 6.4. Overall, it is found that the application of a targeted GMM model can significantly improve the detection probability in comparison to a generic GMM model, as expected.

Having made comparisons between the three considered cWB post-production methods, we now observe the direct distribution of sensitive distance estimates at iFAR $\geq$ 100 years for all considered waveforms for each GMM approach in Figure 6.7. The relationships of sensitive distance with respect to total mass and initial angular momentum L are similar to those previously seen in Fig 6.3, where sensitivity increases with both total mass and L for both GMM approaches. While it was previously shown that the application of GMM achieved the greatest improvement in sensitive distance for the low L waveforms, we see that the analysis is still least sensitive to this type of system across all masses and mass ratios. This lessened sensitiv-

ity likely arises from the low L systems radiating less energy. The generic all-sky short model is less sensitive than the targeted model across all considered waveforms, as expected due to it's weakly-modelled nature. While the use of the generic all-sky short model caused loss in sensitivity to the majority of high L waveforms, it is here seen that competitive sensitive distance estimates are still achieved. From this we can conclude that parabolic radiation-driven capture events are detectable with the current analysis employed in the search for short duration transients.

Table 6.1 details the numerical estimates of sensitive distance for the considered parabolic radiation-driven capture waveforms across all 3 cWB post-production approaches. The highest sensitive distance is achieved for the $q = 1$, $L = 0.9$, $M = 200\mathrm{M}_{\odot}$ with the targeted capture GMM analysis at $D_{\langle VT \rangle} = 1.76$Gpc. As stated in section 6.2, this is a similar detectable distance as seen for high mass quasi-circular BBH systems. With improvements to both the detectors and data analysis techniques ahead of the fourth LVK observing run, this distance is expected to increase further.

While sensitivities to radiation-driven capture events were estimated to be the highest with the targeted GMM approach, we do not calculate upper limits on rates with these since it is expected that further improvements can be made to the model before searching for such events in the detector data. However, it can be assumed that while the increase in sensitive distance observed with this approach would allow for more stringent upper limits to be applied, they would likely not be high enough to be competitive with current literature [185]. The search for short GW transients with the generic all-sky GMM approach was detailed previously in section 4.1.3, detecting only known CBC events detailed in [9, 11]. However since the generic GMM approach brings minimal sensitivity improvements to radiation-driven capture events compared to the standard cWB analysis, expected rates will not be competitive and hence are not calculated here.

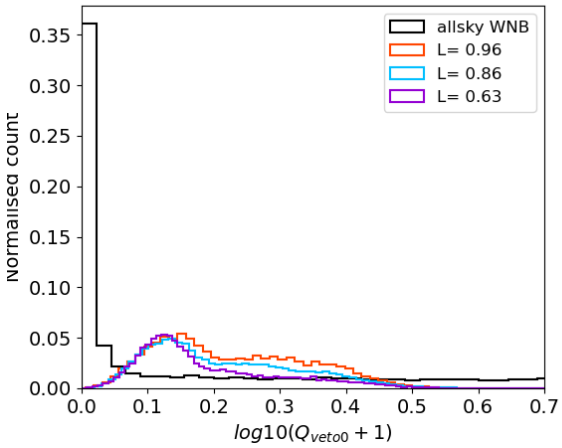
6.4 Investigating properties of parabolic capture waveforms

From the above studies, we have discovered that it is possible to detect parabolic radiation-driven capture waveforms with current weakly-modelled pipelines in the all-sky short search, and demonstrated that a targeted GMM search has the potential to achieve significant sensitivity improvement to such events. In light of these studies, we explore some of the properties of the considered capture waveforms in an attempt to gain insight into how the GMM approaches could be improved.

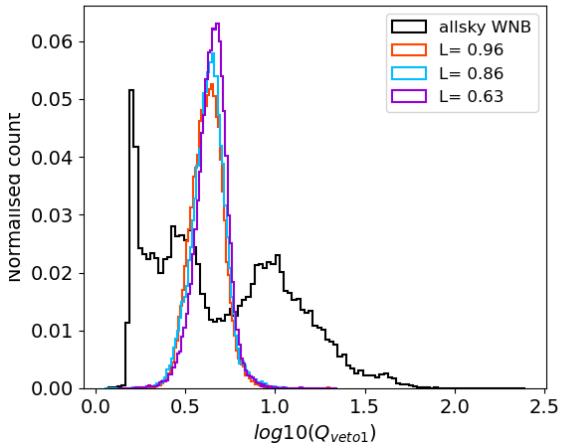
First, we investigate regions of the generic all-sky short signal model which may cause a loss in sensitivity compared to the cWB standard post-production analysis, particularly for the high L waveforms. In order to do this, the distribution of each q, L waveform was observed over all re-parameterised cWB attributes in the 2-dimensional space. From this representation, the

q	L	M_{tot} $[M_\odot]$	$D_{\langle VT\rangle}$ [Gpc]			M_{tot} $[M_\odot]$	$D_{\langle VT\rangle}$ [Gpc]		
			Capture GMM	all-sky GMM	STD cWB		Capture GMM	all-sky GMM	STD cWB
1	0.5	20	0.041	0.036	0.030	50	0.18	0.16	0.12
	0.9		0.10	0.089	0.10		0.55	0.49	0.58
	0.95		0.17	0.15	0.17		0.82	0.71	0.82
	0.96		0.18	0.16	0.19		0.70	0.65	0.78
2	0.7		0.061	0.055	0.050		0.32	0.28	0.26
	0.8		0.086	0.077	0.085		0.46	0.41	0.47
	0.85		0.16	0.14	0.16		0.73	0.64	0.74
	0.86		0.15	0.14	0.16		0.64	0.57	0.69
4	0.5		0.043	0.038	0.032		0.20	0.18	0.16
	0.6		0.071	0.062	0.071		0.36	0.32	0.36
	0.62		0.11	0.11	0.12		0.55	0.48	0.55
	0.63		0.10	0.097	0.11		0.45	0.42	0.52
8	0.35		0.033	0.030	0.028		0.15	0.13	0.13
	0.385		0.052	0.049	0.059		0.28	0.24	0.28
1	0.5	100	0.44	0.38	0.35	200	0.70	0.61	0.53
	0.9		1.37	1.18	1.31		1.76	1.55	1.64
	0.95		1.50	1.35	1.50		1.48	1.35	1.46
	0.96		1.30	1.16	1.35		1.49	1.31	1.38
2	0.7		0.77	0.66	0.67		1.17	1.01	0.98
	0.8		1.15	0.99	1.05		1.53	1.33	1.39
	0.85		1.39	1.24	1.37		1.33	1.22	1.33
	0.86		1.17	1.06	1.23		1.37	1.19	1.26
4	0.5		0.50	0.43	0.41		0.79	0.67	0.60
	0.6		0.90	0.75	0.79		1.15	0.99	1.02
	0.62		1.06	0.98	1.07		1.05	0.96	1.03
	0.63		0.90	0.80	0.90		1.09	0.94	0.98
8	0.35		0.36	0.33	0.31		0.56	0.49	0.44
	0.385		0.66	0.56	0.63		0.76	0.66	0.72

Table 6.1: Table of sensitive distance estimates at an iFAR of 100 years for parabolic radiation-driven capture waveforms over considered total masses of $M_{tot} = [20, 50, 100, 200]M_\odot$. Comparisons are made between the 3 analysis approaches, namely the targeted capture cWB+GMM analysis (*capture GMM*), the generic all-sky short cWB+GMM analysis (*all-sky GMM*) and the cWB with standard post-production analysis (*STD cWB*).



(a) distribution over $log10(Q_{veto0}+1)$

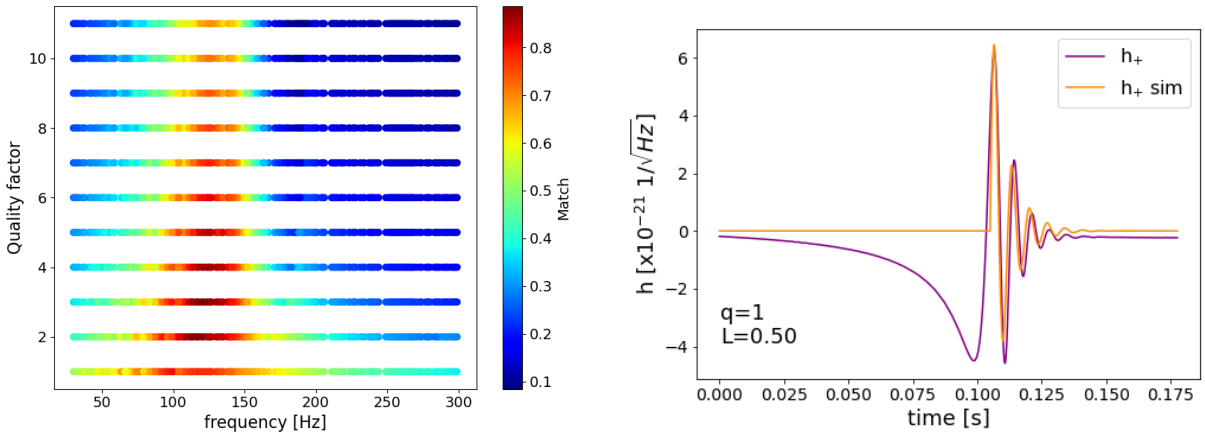


(b) distribution over $log10(Q_{veto1})$

Figure 6.8: Histograms showing the distributions of re-parameterised $log10(Q_{veto0}+1)$ (a) and $log10(Q_{veto1})$ (b) attributes for the high L 'double blip' waveforms against the generic all-sky short WNB GMM model. For both plots, the distribution of $q=1, L=0.96$ (red), $q=2, L=0.86$ (blue) and $q=4, L=0.63$ (purple) is shown against the generic all-sky short model (black)

distribution of the waveforms vs. signal model appears consistent for the majority of attributes, with differences only observed in the Q_{veto} region. The distribution of $Q_{veto0,1}$ varies over the value of L for each mass ratio, with the high L waveforms causing a peak outwith the density of the mixture model. We observe the distribution of the highest L waveforms for mass ratios $q=[1,2,4]$ in the re-parameterised $Q_{veto0,1}$ space in comparison to the distribution of the all-sky short GMM model in Figure 6.8. The distributions of $log10(Q_{veto0}+1)$ and $log10(Q_{veto1})$ are consistent across high L waveforms from all mass ratios, the bulk of which appear to lie within a region which is not well represented by all-sky short model. While the distributions are within the range covered by the all-sky short model in single dimensions, they lie in a region which has lower density in the model, which will ultimately effect the log-likelihood estimates of such triggers. Specifically, the regions in which the double blip waveforms occur are $0.1 < log10(Q_{veto0}+1) < 0.2$ and $0.5 < log10(Q_{veto1}) < 0.9$. It is possible to update the all-sky short generic WNB training data to more densely sample from this region to explore whether this would increase the sensitivity of the weakly-modelled search to radiation-driven capture waveforms. In fact referring back to section 5.1, we see the newly considered WNB training data for O4a, referred to as *TS2*, is more densely sampled in the Q_{veto} space, particularly for $0.5 < log10(Q_{veto1}) < 1$, which may increase the sensitivity to parabolic capture events further in O4.

For the targeted GMM approach, the loss in sensitivity could be accounted for by the number of injections considered in training data for each total mass. While the capture injections were split in such a way that ensured equal percentage of each mass was considered in the training data, this does not equate to an equal number of injections from each mass. In fact, less of



(a) Grid of match over Q and frequency

(b) Ringdown and Capture waveform in timeseries

Figure 6.9: Example of the frequency and Q match study between ringdown and capture waveforms for a system with $q = 1, L = 0.5$. A grid of match values over the frequency and Q space is shown in (a), while the ringdown waveform with optimal frequency of $f_0 = 122$Hz and quality factor $Q = 3$ is shown in comparison to the h_+ components of the $q = 1$, $L = 0.5$ capture waveform in timeseries in (b).

the low mass systems are recovered by cWB production, resulting in a much lower fraction being considered in the GMM training data. This could have an effect on the distribution of the models, and in the future a study should be done to correct this. The same effect is seen over initial angular momentum L, where less low L waveforms are considered in training due to their recovery rate in cWB production. However this does not account for the pattern of sensitivity increase over L, suggesting some other artefact of the model is lessening sensitivity for high L waveforms.

In an attempt to further understand the properties of capture waveforms, we conduct a simplistic match-based study between the considered set of parabolic capture waveforms and simulated ringdown waveforms (introduced in section 2.1.1). Such a study will provide insight into the distributions of frequency and Quality factor over system properties such as total mass or initial angular momentum, which may aid in reinforcing the targeted GMM model for parabolic capture events. If it is found that ringdown waveforms can accurately represent the morphology of parabolic capture waveforms, there is scope for constructing a targeted parabolic capture training set out of simulated ringdown waveforms, which would allow for more control over the parameter space of the GMM model.

Ringdown waveforms are simulated following Eqn. 2.3 over frequencies $f_0 = [20, 300]$Hz and Quality factor $Q = [2, 7]$. The match between a selected parabolic capture waveform and each simulated ringdown of given frequency and Quality factor is computed by taking a weighted inner product over both waveforms h_1 and h_2:

$$\text{match} = \frac{(h_1|h_2)}{\sqrt{(h_1|h_1)(h_2|h_2)}} \tag{6.5}$$

where the weighted inner product is given by [188]:

$$(h_1|h_2) = 4\mathrm{Re}\left[\int_o^\infty \frac{\widetilde{h}_1^*(f)\widetilde{h}_2(f)}{S_n(f)}df\right] \tag{6.6}$$

for noise PSD $S_n(f)$. The match is normalised to vary between 0-1, resulting in ringdowns with optimal frequency and Q having match ≈ 1, and is calculated only within the considered frequency range of the O3 all-sky short search, $f = [16, 1024]$Hz. This process is repeated for all parabolic capture waveforms for total mass of $100M_\odot$ as an initial study. An example of the match study is shown in Figure 6.9 for capture waveform $q = 1$, $L = 0.5$. 6.9a displays the distribution of match across frequency and Quality factor Q, with the optimal region clearly shown around $f_0 \approx 120$, $Q = 2,3$. A comparison of the $q = 1$, $L = 0.5$ parabolic waveform and a ringdown with optimally chosen f_0, q is shown in 6.9b. The match reaches a maximum of 0.89, suggesting that the simulated ringdown is a reasonable representation of the capture waveform, along with visual agreement over the sensitive frequency range of the search. Despite this, the simulated ringdowns could be improved further through the consideration of a variable amplitude factor: while the given example obtains an amplitude factor which agrees well with the capture waveform, this is not taken into account and so this is not always the case. The low L waveforms have a better resemblance to the simplistic ringdown structure used here, however as L increases it is difficult to accurately represent the waveforms. This is especially true for the high L double blips waveforms, which would require the consideration of a sine-Gaussian type waveform attached to the ringdown signal. A lot more work is required before a reliable training set could be created in this way, however as an initial study, we find that the optimally matched frequencies range from 90-160Hz, while Quality factor ranges over 2-9. Central frequency increases with L for all mass ratios as expected, however this relationship failed at the highest L waveforms, likely due to poor fitting as discussed. No clear relationship was seen with Q.

It is possible that a more reliable training set could be created from tapered NRSur7dq4 BBH simulations [189], since it has been noted that capture waveforms have similar morphology to quasi-circular BBH at high masses [4]. The behaviour of such models at low masses would need to be investigated further, and the consideration of a sine-Gaussian-like waveform would still be required for accurate representation of the double blip signals.

The above studies provide an initial insight on selected properties of parabolic capture waveforms, and suggest possible avenues which may increase the sensitivities to such events within the GMM approach. However, more detailed investigations need to be completed before conclusions can be made.

6.5 Summary and discussion

Throughout this Chapter, the sensitivity of the all-sky short search to parabolic radiation-driven BH capture systems has been investigated. To do this, a set NR waveforms varied over initial angular momentum, mass ratio and total mass from [84] were injected into O3 cWB data, and the sensitivity to these waveforms has been tested with 3 cWB post-production methodologies: cWB standard post-production, generic GMM post-production and a capture-targeted GMM post-production. The sensitive distance was calculated across all 3 post-productions for comparison.

The application of the standard cWB post-production involved binning the processed simulations into 3 classes based upon morphology in order to lessen the effect of blip glitches on significance estimates. It was demonstrated that sensitive distance increased with total mass of the systems, which was expected behaviour since the gravitational energy released increases with higher mass systems. It was also found that the search is least sensitive to low angular momentum systems. This is likely due to their placement in the noisy bin of standard post-production, since their signals have low Quality factor, similar to frequent loud glitches. The maximum value of sensitive distance with the standard post-production was calculated as 1.64Gpc for a system with M_{tot}=200M$_\odot$, q=1 which is similar to the sensitive distance of high mass quasi-circular BH systems. In general, it was concluded that it is possible to detect GWs from high mass parabolic capture systems with current sensitivities, however the sensitive reach of lower mass systems makes their detection less probable. Since the zerolag data of the standard O3 search had previously been analysed in [85] with no confident detections of parabolic capture, upper limits on estimated rates were calculated. It was found that these were not competitive with literature at current sensitivities, with even the most stringent rates calculated being an order of magnitude larger than theoretically predicted rates.

The GMM post-production involved the application of 2 separate signal models: the O3-trained all-sky short signal model from Chapter 4 for a ‘generic’ approach, and a capture-based signal model for a ‘targeted’ approach. The capture-based model was trained on a portion of triggers from the injected waveforms. Calculating the sensitivity improvement for each GMM approach compared to the standard cWB post-production, a consistent increase in sensitivity is observed for low initial angular momentum systems for both approaches. The significant improvement seen for these systems is due to GMM’s ability to mitigate blip glitches, since previously these signals fell into the noisy bins of cWB post-production. While it has been shown multiple times throughout this thesis, this result strengthens the statement that GMM is effective in glitch mitigation. In addition to the low initial angular momentum waveforms, the targeted approach observed a more significant sensitivity increase for higher total mass waveforms.

The targeted GMM approach consistently gains sensitivity over the generic approach, which demonstrates the advantage a targeted GMM analysis may bring to GW searches. The generic GMM approach decreases sensitivity to systems with L above the lowest value, while the targeted GMM occasionally decreases sensitivity to double blip systems. Overall, the proportion

of decrease observed with the targeted model is low compared to the increase observed in other areas, however further investigation is required to understand why the higher L waveforms are not well-represented by the models. Overall, the targeted GMM achieves similar or improved sensitivities compared to the other 2 post-productions, and observes a maximum sensitive reach of 1.76Gpc.

Initial investigations into the properties of parabolic BH waveforms were completed in an attempt to gain insight on why the GMM models decreased sensitivity to selected waveforms. It was thought that these waveforms were under-represented in the training data set, specifically in the Q_{veto} attribute space. In addition, the total number of each waveform-type considered in the training data was not equal, which may have contributed to this issue further. One possibility to improve the robustness of the targeted signal model would be to create a new signal training set consisting of a separate set of simulations. It was briefly investigated whether this could be done by simply using ringdown waveforms, however is thought that using tapered high mass BBH models may obtain better accuracy.

To conclude, it has been found that the current all-sky short searches are sensitive to detecting GWs from parabolic radiation-driven capture systems. A proof-of-concept application of a targeted GMM analysis has also shown to have promising potential in increasing the sensitivity of searches to astrophysical sources. Further investigations into the targeted GMM analysis may improve observed sensitivities to parabolic radiation-capture further, while the probability of detection will increase with future LVK observing runs.

Chapter 7

Conclusions

The work in this thesis has explored the use of Gaussian mixture modelling (GMM) as a glitch mitigation tool in un-modelled searches for short GW transients, with a particular focus as post-production to the cWB algorithm. cWB was introduced in Chapter 2, and bases the detection of GW events upon excess coherent energy across a network of detectors, with little assumption on signal morphology. Due to it's weakly-modelled nature, it is highly susceptible to glitches, which can negatively effect the detection of GW events. Hence, the need for techniques to mitigate the effect of such glitches is crucial, especially as we move into a new era of GW detection with increased detector sensitivity.

First introduced in [1, 2], GMM distinguishes glitches from GW events in the cWB data by modelling the noise and signal populations and constructing a log-likelihood ratio to quantify where cWB triggers lie within the populations. The work presented in this thesis has enhanced the cWB+GMM search through in-depth investigations into GMM methodology, and has demonstrated it's effectiveness as a mitigation tool to cWB multiple times through various applications in the search for short duration GW transients.

Investigations into the enhancement of the GMM methodology as post-production to the cWB search were detailed in Chapter 3. A new approach to optimising the number of Gaussians in the models was presented, where detection efficiency is optimised over a range of models trained with differing number of Gaussian components rather than the Bayesian information criterion (BIC). The new approach provides a combined measure of how well the Gaussians in the model fit the training data and how well these models can distinguish between populations, increasing the performance and reliability of the analysis. Another significant development to the methodology was a new signal training data set, consisting of generic WNB simulations across the low-frequency, short-duration parameter space. Use of the new training data removed a bias which was discovered in the previous analysis due to training on a portion of the same injection population as used for testing sensitivities. Combining sets of WNB simulations allowed for increased sensitivity to short-duration, low quality factor (Q) waveforms without compromising

the sensitivity to other waveform types, suggesting multiple populations of WNB simulations may be utilised in the future to construct a fuller training set. Investigations into GMM model training parameters enhanced the reliability of the methodology by ensuring training processes were optimised for computational efficiency and complexity of the data. The use of k-means++ as an initialisation procedure was believed to provide a more reliable placement of Gaussians, while checks on convergence limits and number of total iterations were complete. Assessing parameters such as this helped minimise the possibility of under or over-fitting models to training data, hence increased reliability. The consistency of results over random state was also investigated, in which it was found that an occasional outlier result may occur due to assumed 'bad' initial placement of Gaussians. Despite this, the majority of random states produced results within statistical error margins, suggesting the methodology is robust to initialisation variations.

Direct sensitivity comparisons with the previous GMM methodology confirmed the above investigations had lead to an enhancement in the application to cWB post-production. An increase in sensitivity was observed to short-duration, low-Q waveforms such as GA, while the increase in sensitivity to all considered CCSN waveforms demonstrated robustness to a wider range of sources. The loss of sensitivity to SG $Q = 3,9$ waveforms arose due to removal of bias compared to the previous methodology, hence was not of great concern. The above enhancements provided a base methodology on the application of GMM to cWB which has been demonstrated to be reliable and robust to a variety of expected sources, however further minor alterations are expected to continually occur as the methodology undergoes additional testing. For example, further updates to training data and the considered re-parametrised attributes may arise when the search is run on different data.

The enhanced GMM post-production was thoroughly tested in Chapter 4, where the cWB+GMM search for short GW transients was applied to data throughout the third LVK observing run in order to ensure it was capable of performing at the level required of search pipelines, investigating it's potential use in future observing runs. The application of the pipeline as a full search in O3 with the 2-detector LH network allowed for direct comparison with other all-sky short cWB-based pipelines: the standard cWB post-production and XGBoost post-production [124]. Through this, it was demonstrated that GMM post-production is capable of achieving similar sensitivities as other methods for the majority of ad-hoc, CCSN and CS waveforms. In fact, GMM was the most sensitive post-production method for the majority of GA and CS waveforms, which was the the first direct evidence of GMM's ability to mitigate blip glitches due to their similar morphology to these waveforms. The GMM search concluded with a null detection on burst events in O3, however detected 14 CBC events, similarly to the other post-productions (all previously reported in GWTC-3 [11]). This, coupled with the observed sensitivity estimates, proves the potential of GMM to perform at the level required to run as a pipeline in the search for short durations GWs.

The first-time application of GMM post-production to a 3-detector network further exemplified the ability to mitigate blip glitches and potential to achieve competitive sensitivities to predicted burst sources. It was demonstrated that GMM on the 3-detector LHV network could achieve similar sensitivities as the 2-detector network, performing significantly better for CS waveforms. This is a major result and a first of it's kind with the cWB algorithm, since sensitivities with the 3-detector network are often lower than the LH network due to the mis-alignment of Virgo in comparison to LIGO detectors causing a lower rate of glitch rejection. With the addition of GMM post-production allowing for similar sensitivities to be achieved with the 3-detector network, it may be possible to run burst searches on LHV data in future observing runs, gaining the benefits of increased duty cycle and better sky localisation for detected events. It is likely for performance on the 3-detector network to improve with further investigations into optimal attributes for LHV analysis, since section 4.2 detailed a proof-of-principal application of GMM.

The robustness of the cWB+GMM search to a wide range of expected short-duration burst sources was tested through participation in the burst benchmark project. By running the search on a blind MDC data set, it was found that GMM was capable of detecting the majority of predicted astrophysical sources considered from the short-duration, low-frequency parameter space at similar levels to other cWB post-production techniques. This further demonstrated that the cWB+GMM pipeline was capable of effectively searching for short-duration GW transients with minimal assumptions. Additionally, it was found that the GMM search was sensitive to sources outwith it's targeted parameter space, such as the long-duration accretion disk instability waveforms, demonstrating the potential of application to searches in other regions of the parameter space.

Through extensive testing in Chapters 3 and 4, it was shown that the cWB+GMM algorithm was able to perform at a competitive level in the search for short duration GW transients. Hence, the application as an offline search for such signals in the first epoch of the fourth LVK observing run (O4a) was detailed in Chapter 5. No confident burst detections were made, however 13 significant detections of GWs from CBC systems were observed, all also identified by dedicated CBC searches. The loudest non-CBC event was found with an iFAR of 1.75 years, however was consistent with the expected background distribution to within 1σ and initial investigations into the morphology of the event suggested that it was consistent with an artefact of noise. Despite no burst event being detected, it was found that the GMM search achieved significantly increased sensitivities to ad-hoc and CCSN waveforms in comparison to those obtained during O3. It is likely that a large majority of these improvements are due to detector upgrades, however some improvement must be credited to the enhanced cWB and GMM methodologies. The targeted frequency space of the GMM search increased for O4a, hence more ad-hoc waveforms were used for benchmarking sensitivity at both lower and higher frequencies. Comparing the

GMM sensitivity across waveforms, it was found that the GMM search is least sensitive to SG waveforms with $f < 70$Hz and $f > 1000$Hz. Further investigations are required to make confident conclusions on the cause of this, however it is possible that these waveforms lie in low density regions of the signal model and that a more densely populated training set could aid sensitivities to these regions.

The application to O4a was extremely valuable in terms of insight into the GMM methodology, and highlighted a variety of avenues for future improvement and full understanding of GMM behaviour. While the capability of GMM to mitigate blip glitches has been demonstrated numerously throughout this work, initial glitch studies have shown that these are still the dominant source of noise in the search. Full investigations into the population of loudest glitches present in the analysis may provide insight into areas of the attribute space which need reinforced in the background model, which has the potential to aid in glitch mitigation further. The abnormal glitch behaviour initially observed in K11+K12 raises some concern on the reliability of the GMM models. The drastic change in background log-likelihood values due to a small number of problematic triggers in the training data indicates in-depth investigations must occur in order to better understand model behaviour and ensure such issues are avoided in the future. Additionally, the lack of visual difference between problematic and correct models suggests that the 2D visualisation does not effectively represent the distribution of the models. The development of a new visualisation method for the multi-dimensional models would assist in sensitivity and robustness investigations in several aspects. Finally, the secondary signal model approach introduced in Section 5.1 demonstrated potential to increase the sensitivity of the search in the future, with further investigations required in order to optimise which region of the signal parameter space to target.

The approach of a targeted GMM analysis was explored in Chapter 6, in relation to a search sensitivity study on parabolic radiation-driven BH capture systems. The sensitivity to parabolic capture systems was investigated with three variations of post-production to the cWB algorithm: standard post-production, generic all-sky short GMM and the newly constructed targeted-capture GMM. Across all post-productions, it is seen that highest sensitivity is achieved towards high mass and high initial angular momentum systems due to the total gravitational energy emitted in such systems. Comparing the performance of each post-production approach, it was demonstrated that both GMM models increase sensitivity towards low initial angular momentum systems. This is due to these waveforms having low-Q, placing them within the noisy bins in standard cWB post-production, reinforcing the ability of GMM to mitigate low-Q glitches further. The targeted GMM was found to decrease sensitivity to high initial angular momentum (double blip) systems, and achieved less sensitivity improvement for lower mass systems. This is likely due to an imbalance in the waveforms considered in the training data, and may be improved with more careful construction of the signal training set in the future. Of the three

approaches, the targeted GMM was found to achieve the highest sensitivities for the majority of considered waveforms, proving that the approach of a targeted GMM has the potential to improve search sensitivities to predicted astrophysical sources. While this is not desired for the all-sky short search, it may be a useful application in targeted searches.

The above investigations have demonstrated that GWs from parabolic capture systems are detectable with current detectors and search techniques. A maximum sensitive distance of 1.76Gpc was found for high mass systems with the targeted GMM approach, similar to that of high mass quasi-circular BBHs. Open box analysis did not take place with the targeted GMM approach, as it is believed that better sensitivities may be achievable with an optimised signal training data set. However, upper limits on the rates of parabolic capture systems were calculated based upon the non-detection with the standard cWB post-production. The most stringent upper limit rate estimate was found to be an order of magnitude larger than predicted in literature, meaning sensitivities will need to increase before such calculations become useful to the astronomical community. With this said, it is likely that the sensitivity increase observed during O4a will provide more completive rate estimates.

The development of burst search methods is of high importance in the advanced detector era of gravitational waves and beyond, as it presents the opportunity to detect a wide array of predicted GW sources that as of yet have not been observed. In addition, these searches are important in providing confidence on CBC detections and present the opportunity to detect GW from unpredicted sources. In particular, the development of noise mitigation techniques is crucial for refining sensitivities of searches and optimising the possibility to detect GWs from burst sources, especially as the sensitivity of detectors reaches new levels. The work presented in this thesis has done just that, through the development of the GMM technique as a post-production noise mitigation tool to the cWB algorithm. The enhancement and extensive testing of the cWB+GMM methodology has demonstrated it's capability in effectively running as a search pipeline for short GW transients and mitigating the limiting sources of noise. This has shown particular promise for short duration, single-cycle waveforms such as GA and CSs, which have previously obtained low sensitivities within burst searches due to effects from blip glitches. The benefit of a targeted GMM post-production approach has also been shown, which will be able to aid in the search for predicted astrophysical searches in the future. The effectiveness of the cWB+GMM search has been demonstrated further through the application as an offline search during O4a, providing increased sensitivities compared to previous observing runs and a range of CBC detections. Of course there is scope for sensitivities to improve further as analysis of O4 data continues, especially with possible future improvements to the GMM technique. This, alongside other future prospects from work detailed in this thesis, are discussed in section 7.1.

7.1 Future prospects

With the increased sensitivities that come with detector upgrades in future observing runs comes the necessity of improved data analysis and glitch mitigation techniques within GW searches. There are a range of developments which may be made to the cWB+GMM pipeline to improve it's performance, including enhancements to the all-sky short search methodology as well as the possible application to other regions of the parameter space.

While the work presented in this thesis has extensively demonstrated the success of the cWB+GMM methodology, continued development will improve the pipeline further, especially as we move into future observing runs. Numerous avenues of required investigation have been highlighted throughout this work, such as studies into the selection and re-parameterisation of attributes, a better visualisation method for the models in order to quantify their distinguishability, and population studies of dominant glitches. It is also important for development on the computational efficiency of the analysis to take place as we move into future observing runs to ensure an optimisation on computational resources and ability to deliver results in appropriate time frames.

Future development of the GMM post-production method may also see it applied to other searches. With the results of the burst benchmark project in Section 4.3 demonstrating the ability of GMM to detect long-duration signals, an obvious choice may be to test on un-modelled searches covering the long-duration or high-frequency parameter space. It has also been demonstrated that a targeted GMM analysis can provide sensitivity improvement for a targeted set of astrophysical sources. While further development is required in the search for eccentric radiation-driven capture systems, a natural development may be to target a GMM search on eccentric 'fly-by' encounters. Investigations may also be made into combinations of glitch mitigation techniques, for example the use of the cWB-XGBoost post-production [124] detection statistic as an attribute in the GMM analysis. A combination such as this may allow for the benefits of both techniques to be utilised. Finally, it is possible for the GMM noise mitigation technique to be applied to search pipelines other than cWB, which may be beneficial as we move into new generations of detectors.

To conclude, the development of techniques to be applied in the search for GW bursts is crucial as we move into future advanced detector observing runs and next-generation detector networks. As of yet, a confident detection of GWs from a burst source is yet to be seen, however with the continual sensitivity improvements achieved thanks to detector upgrades and improvements to data analysis techniques, it is only a matter of time before this occurs.

Appendix A

Details of core-collapse supernovae models

A range of predicted CCSN models are utilised as cWB injections throughout this thesis, with various features and mechanisms considered. While there are many processes which may be considered within a CCSN explosion, the details of which are out of scope of this thesis, some processes which are important for the understanding of the considered CCSN models are detailed below.

Typically during the collapse of a star, the inner core will collapse upon itself until it approaches the Chandrasekhar mass limit, at which the core bounces, launching a shock wave. This is referred to as the core bounce, and often GW CCSN waveforms will be modelled from this point. As the shock expels outward and loses energy, there is a chance it may 'revive', in which the infalling outer shell is expelled. This is referred to as shock revival, and is likely to result in the formation of a NS. If shock revival does not take place, the outer shell continues to accrete onto the progenitor star, leading the the formation of a BH [76]. There are a range of processes which may effect the shock revival and explosion mechanism, such as progenitor star mass, rotation, or the presence of magnetic fields.

Two broad categories of explosion mechanism are considered for the CCSN models here: neutrino-driven explosion and magnetorotational-driven explosion. In neutrino-driven explosion, neutrino heating drives convection in the post-shock region. In some cases, neutrino-driven convection may lead to standing accretion shock instability (SASI), which drives large-scale oscillations of the shock, further energising the explosion [190]. In magnetorotational-driven explosion, magnetic fields of rapidly-rotating progenitors transfer rotational kinetic energy to the shock.

The exact details of these models and the physics involved are complicated, and it is not yet understood how the above features effect the explosion of CCSN, however many of them directly effect the GW signature. A brief description of the key features in each model is given below alongside corresponding references for further information. For a more complete overview of CCSN physics in relation to the emission of GWs, one can refer to [76].

- Abdikamalov A4O01.0: Model consists of the magnetorotational-driven explosion of a $12M_{\odot}$ progenitor with solar metallicity. A differential rotation parameter of 1268km is considered, which results in weaker rotational than the similar model A2O01.0 detailed below. Further details can be seen in [172].

- Abdikamalov A2O01.0: Model consists of the magnetorotational-driven explosion of a $12M_{\odot}$ progenitor with solar metallicity. A differential rotation parameter of 417km is considered. Further details can be seen in [172].

- Andersen s11: Model consists of a non-rotating $11M_{\odot}$ progenitor star with solar metallicity. No shock revival or SASI growth is achieved within the simulation. Further details can be seen in [169].

- Andersen s15nr: Model consists of a non-rotating $15M_{\odot}$ progenitor star with solar metallicity. Neutrino-driven convection develops strong SASI activity and shock revival is achieved. Further details can be seen in [178].

- Kuroda SFHx: Model consists of a non-rotating $15M_{\odot}$ progenitor star. SASI is achieved. Model has a shorter timescale than is typically considered. Further details can be seen in [171].

- Mezzacappa C15-3D: Model consists of a non-rotating $15M_{\odot}$ progenitor star with solar metallicity. Includes low-frequency emission from neutrino convection and high-frequency emission due to SASI. Further details can be seen in [179].

- Muller L15: Model consists of a non-rotating $15M_{\odot}$ progenitor star. Majority of GW emission at low frequencies due to vanishing asymmetry factor as shock expands. Further details can be seen in [170].

- O'connor and Couch mesa20: Model consists of a non-rotating $20M_{\odot}$ progenitor star with solar metallicity. A strong SASI development is observed. The 'mesto20_pert' model is the same expect includes progenitor perturbations in the shell, which result in an increased shock radius. Further details can be seen in [132].

- Pan SR: Model consists of a slowly-rotating $40M_{\odot}$ progenitor star with a rotational speed of 0.5 radians per second. Accretion is slowed down by the explosion causing features at high-frequency. Further details can be seen in [180].

- Powell he3.5: Model consists of a strongly rotating progenitor star of $3.5M_{\odot}$. Shock revival is achieved and has low explosion energy. Shows high-frequency signature from the excitation of surface g-modes. Further details can be seen in [131].

- Powell s18: Model consists of a non-rotating 18$M_\odot$ progenitor star with solar metallicity. Shock revival is achieved and the model includes strong low-frequency emission due to SASI growth and high-frequency signature from the excitation of surface g-modes. Further details can be seen in [131].

- Radice s9: Model consists of a non-rotating 9$M_\odot$ progenitor star with solar metallicity. Dominated by f-modes and g-modes. Low mass progenitor causes explosion on shorter timescales. Further details can be seen in [133].

- Radice s13: Model consists of a non-rotating 13$M_\odot$ progenitor star with solar metallicity. Dominated by f-modes and g-modes. Explosion doesn't occur. Further details can be seen in [133].

- Radice s25: Model consists of anon-rotating 25$M_\odot$ progenitor star with solar metallicity. Dominated by f-modes and g-modes however has a low-frequency signature due to SASI. Further details can be seen in [133].

Appendix B

Detection efficiency vs. hrss curves

Throughout this thesis, the sensitivity of pipelines to simulated waveforms has been evaluated through the h_{rss50} statistic, which is the value of h_{rss} at which we detect 50% of injected waveforms, as described in section 2.2.5. The h_{rss50} statistic is found by plotting the detection efficiency of each waveform type as a function of injected h_{rss}, fitting the data with a cumulative lognormal distribution and inferring the h_{rss} amplitude at which 50% of waveform injections are detected at given FAR threshold. Errors on detection efficiency are calculated as described in eqn. 2.16, leading to upper and lower regions of error using further lognormal fits and error bounds on h_{rss50} estimates. This appendix provides detection efficiency vs. h_{rss} plots for all injected waveforms considered by the GMM pipeline, specifically relating to work in Chapters 3, 4 and 5.

Figures B.1 and B.2 show the detection efficiency vs. h_{rss} curves at iFAR $\geq$ 100 years for all ad-hoc and CCSN waveforms injected into O3a data, as described in Section 3.3. These curves relate to the results displayed in Figure 3.10. It must be noted that the curves for ad-hoc injections shown in Figure B.1 display over-estimated values of h_{rss50} due to the incorrect definition of total number of initial injections associated with the 20% test data, as discussed in section 3.2.2. The use of 20% test data also results in larger error regions in Figure B.1, as the total number of considered initial injections directly affects the errors on detection efficiency.

Figures B.3, B.4 and B.5 show the detection efficiency vs. h_{rss} curves at iFAR $\geq$ 100 years for all ad-hoc, CCSN and CS waveforms injected into O3a+O3b data, as described in Section 4.1.2. These curves relate to the results displayed in Figures 4.3.

Figures B.6 and B.7 show detection efficiency vs. h_{rss} at iFAR $\geq$ 100 years for injected waveforms throughout O4a, as described in Section 5.4. Note that in Chapter 5, Figures 5.15 and 5.16 show h_{rss50} estimates over each individual chunk combination, whereas the curves shown below consider injections combined over all chunks K01-K16 within O4a, to reduce the number of plots required. Note also that the range of injected h_{rss} has changed in these plots to account for sensitivity improvements observed in O4a.

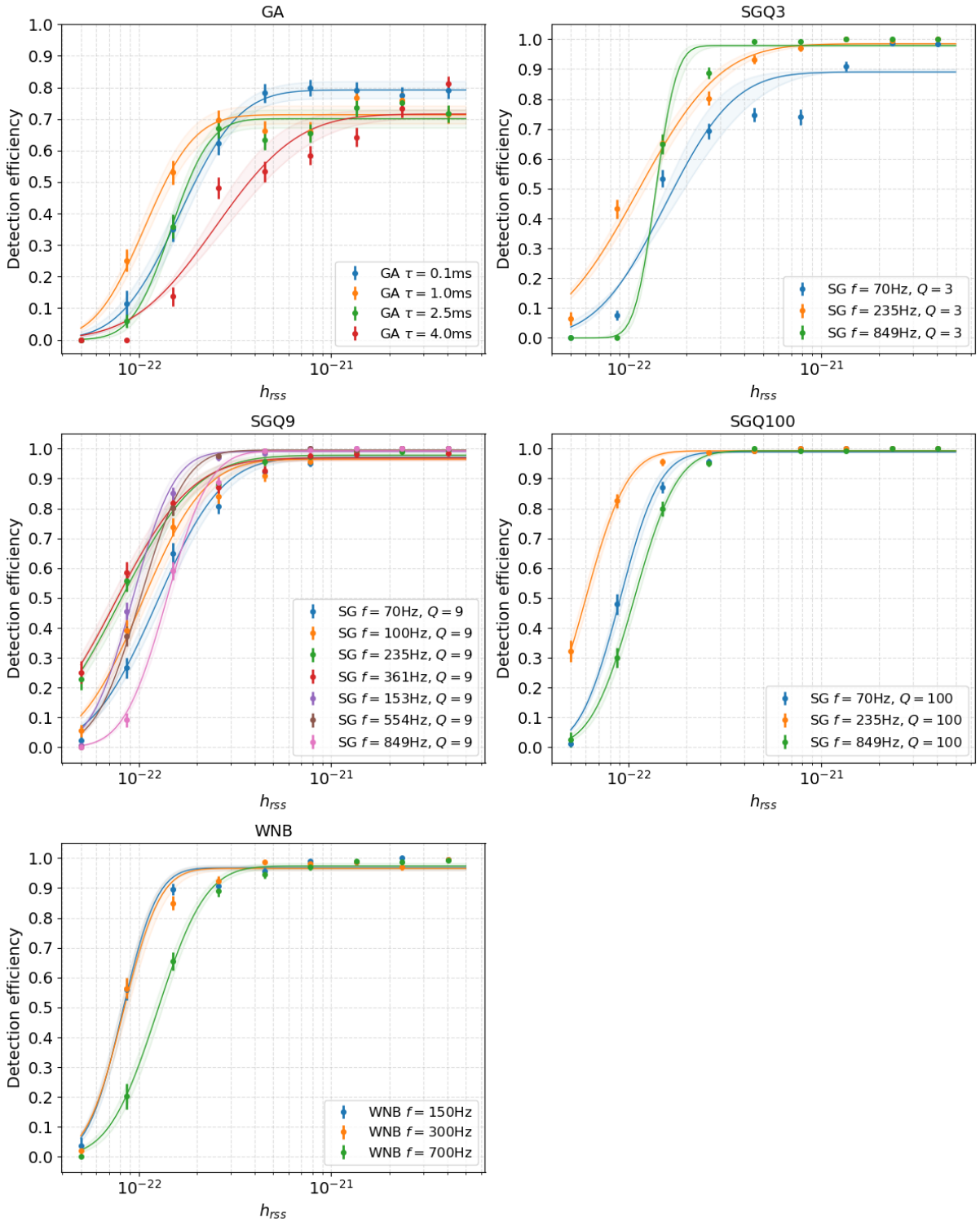


Figure B.1: Detection efficiency as a function of h_{rss} for ad-hoc waveform injections for GMM+ in O3a, using the 20% test data for comparison with previous GMM methodology. Within each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve. Note that the detection efficiency is over-estimated here due to the definition of total number of initial injections used by previous GMM methodology [2].

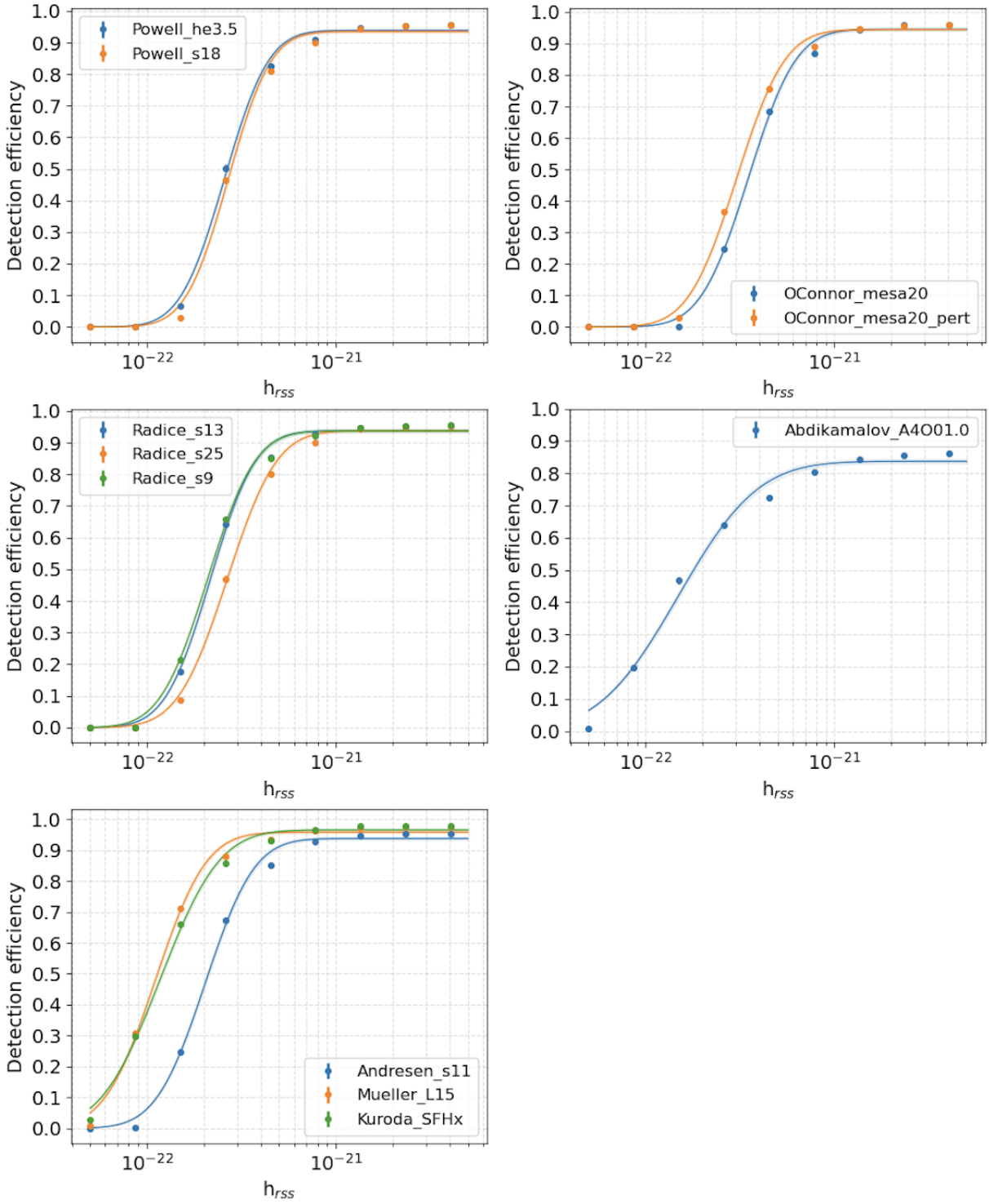


Figure B.2: Detection efficiency as a function of h_{rss} for CCSN waveform injections for GMM+ in O3a. Within each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve.

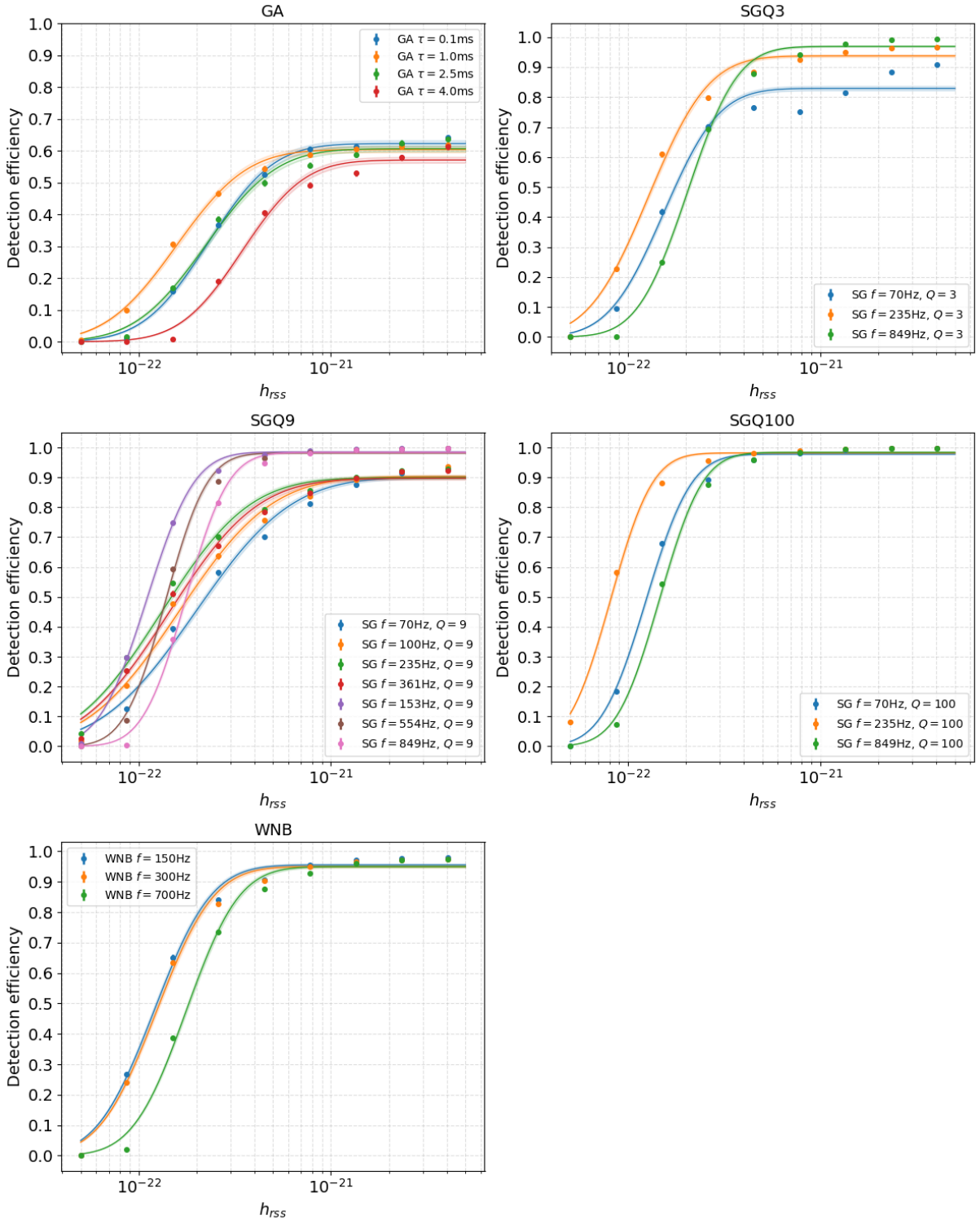


Figure B.3: Detection efficiency as a function of h_{rss} for ad-hoc waveform injections for GMM+ in O3a+O3b. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve.

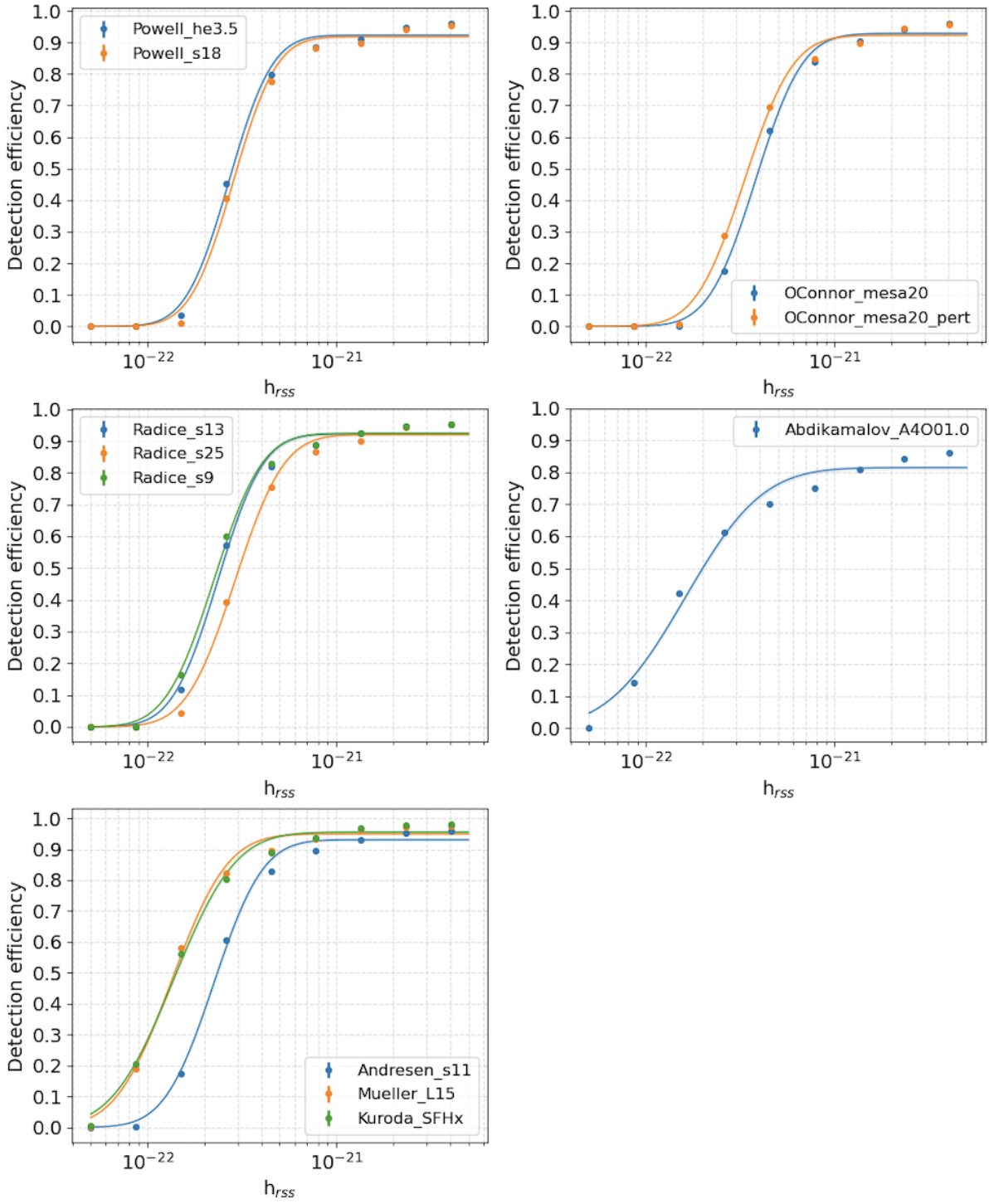


Figure B.4: Detection efficiency as a function of h_{rss} for CCSN waveform injections for GMM+ in O3a+O3b. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve.

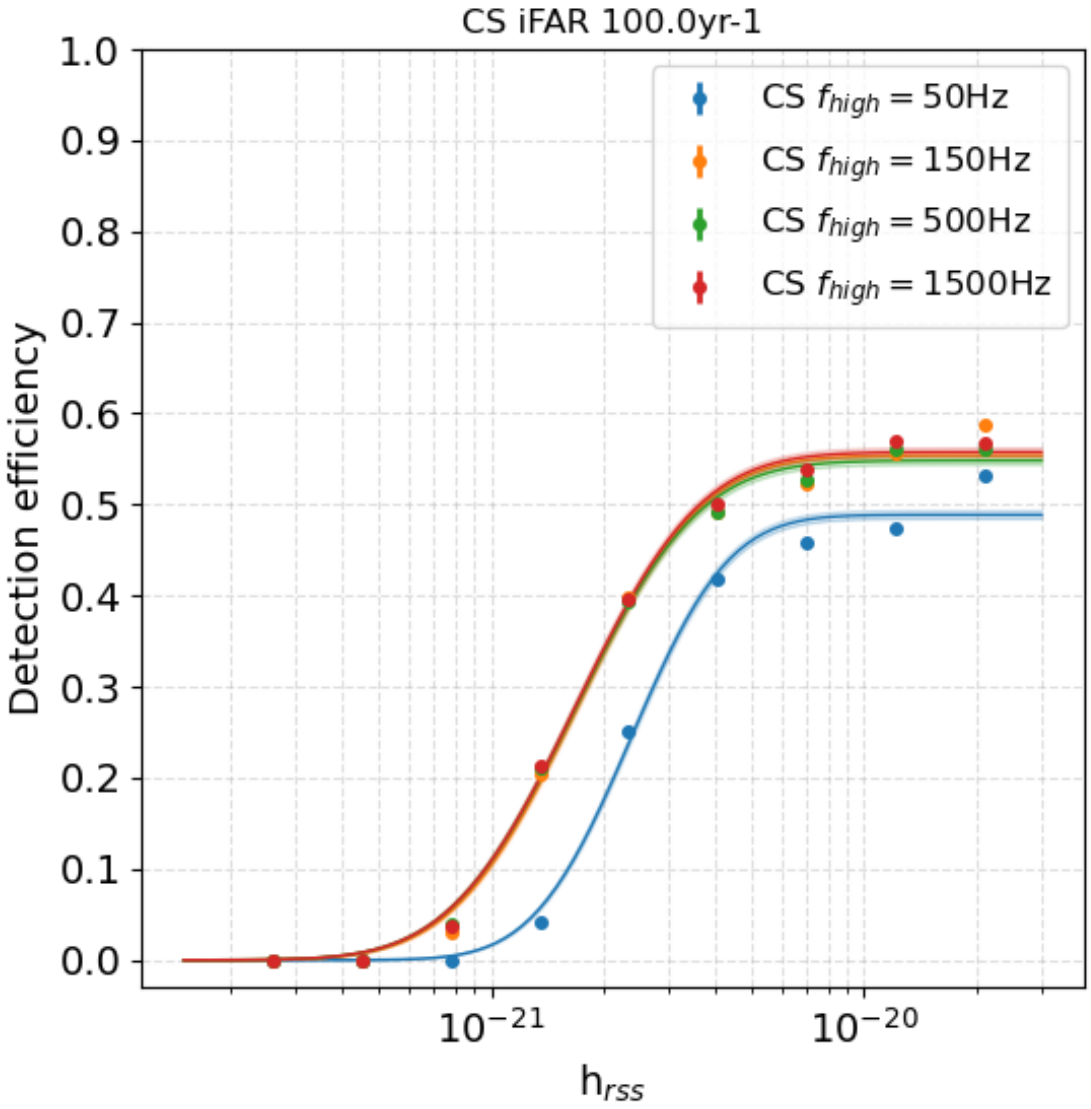


Figure B.5: Detection efficiency as a function of h_{rss} for CS waveform injections for GMM+ in O3a+O3b. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve.

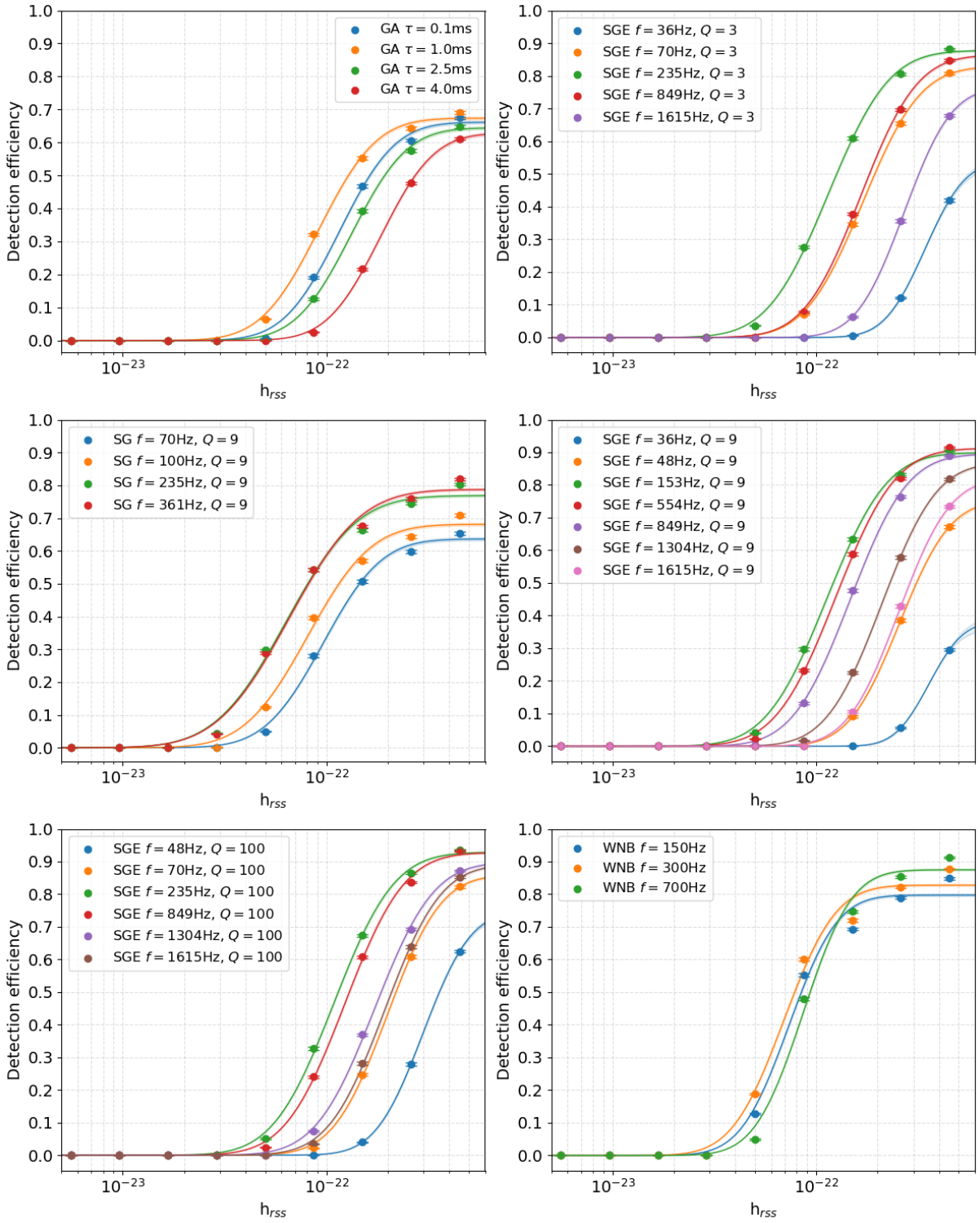


Figure B.6: Detection efficiency as a function of h_{rss} for ad-hoc waveform injections for GMM+ in O4a. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve. Curves are shown for injections combined over all chunks K01-K16 throughout O4a.

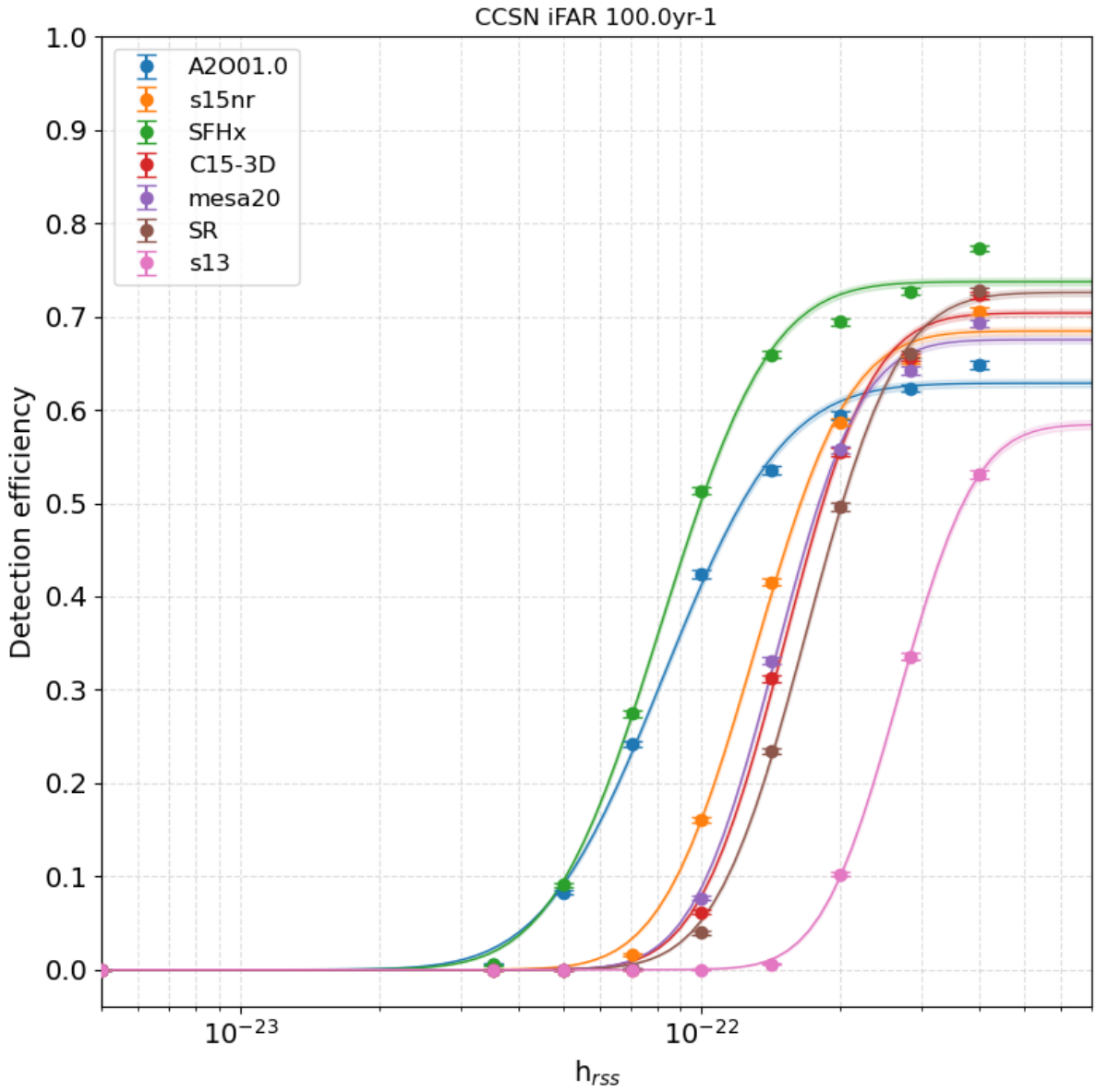


Figure B.7: Detection efficiency as a function of h_{rss} for CCSN waveform injections for GMM+ in O4a. For each plot, a different waveform injection is shown by a different colour, with lognormal functions fit to each set of points. Shaded error regions are shown for each curve. Curves are shown for injections combined over all chunks K01-K16 throughout O4a.

Appendix C

Loud glitch investigations throughout the O4a run

This section details all chunk-by-chunk loud glitch investigations referred to in Section 5.2. For each combination of chunks, the background test data distribution of T vs. frequency, network correlation coefficient c_{c0}, and time were observed. Such investigations are crucial in ensuring the pipeline is performing as expected and give valuable insight into potential glitch correlation. As mentioned before, if loud glitches are clustered around a single range of frequencies, it could suggest that these glitches could all arise from the same source in the detectors. Similarly, if the loudest glitches are clustered around a single instance in time, it may suggest that they were caused by a given disturbance in the detector. The presence of such features in the distributions merit further investigations by the detector characterisation group, and may lead to the definition of a new veto. The clustering of glitches around the low c_{c0} region could indicate a problem with the analysis or models, since low c_{c0} is a common feature of noise and should be down-ranked by the GMM analysis.

The figures below show the distributions of T vs. frequency, network correlation coefficient c_{c0}, and time for each set of combined chunks. The loudest 20 glitches are highlighted as they directly affect the thresholds for iFAR of 100 years. The K11+K12 plots show the observed distributions *after* the discovered trigger production problem from Section 5.3 had been fixed (seen in Figure C.6). With this considered, the distribution of loud glitches did not show any concerning behaviour, however a spike of loud glitches in K15+K16 is present in an instance of time, seen in Figure C.8. The clustering of these loud glitches in time lead to further investigations by detector characterisation, however a common cause could not be identified, hence a veto could not be created for these glitches. This contributed to higher iFAR thresholds in K15+K16 and ultimately affected sensitivity and significance estimates in this chunk.

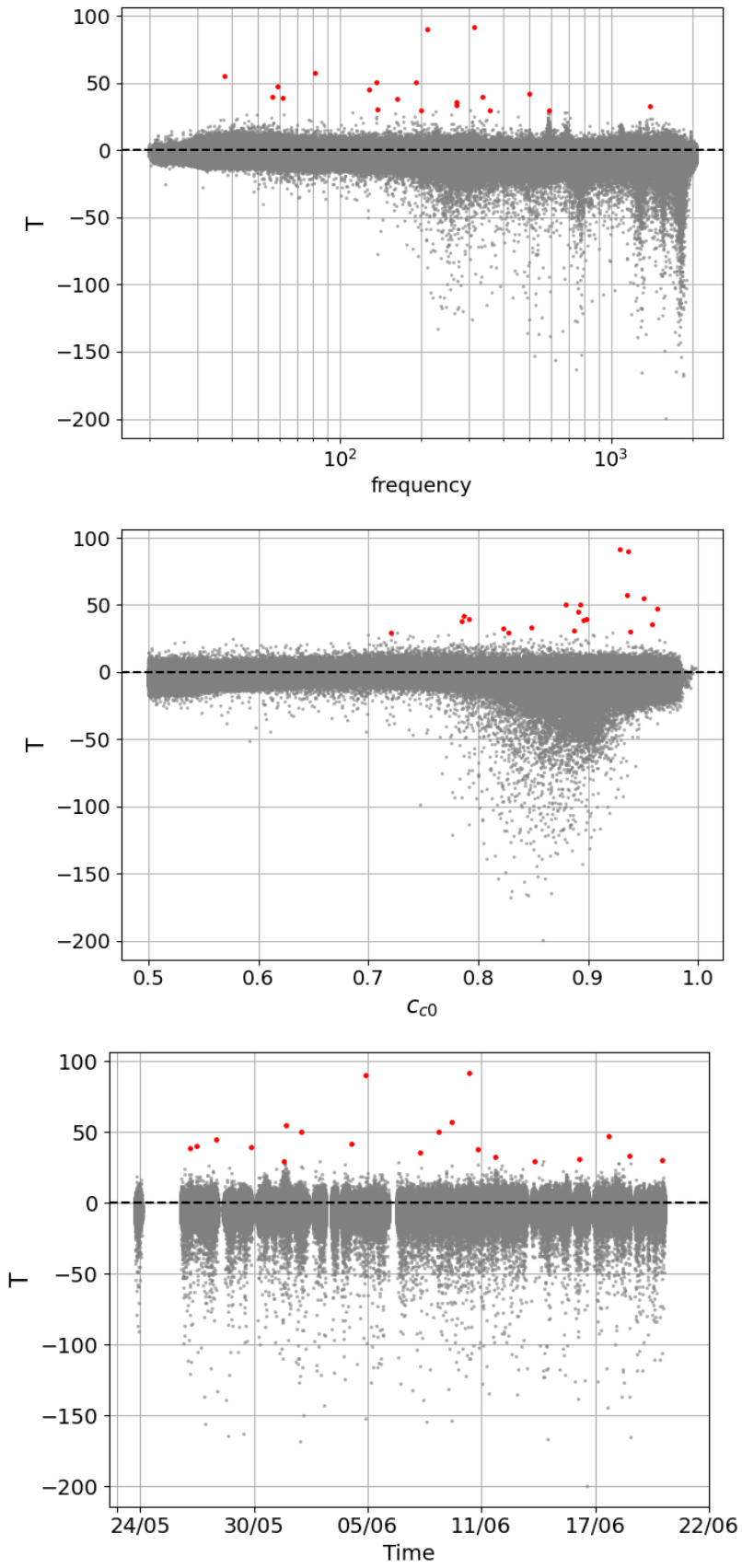


Figure C.1: Plots of loud glitch studies in K01+K02. Distributions of T vs. frequency, network correlation coefficient c_{c0} and time are shown for background test data, with the 20 loudest glitches marked in red.

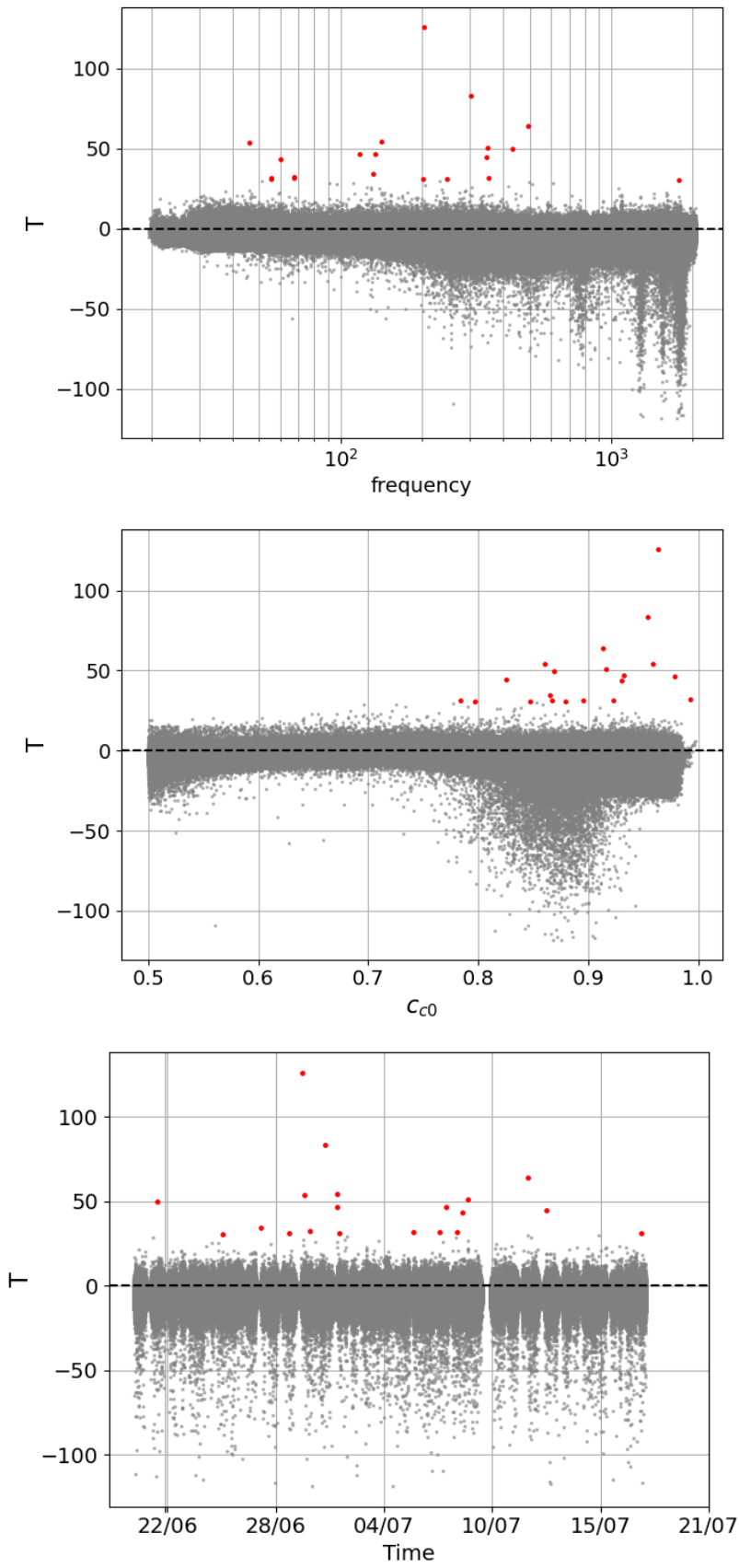


Figure C.2: Plots of loud glitch studies in K03+K04. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

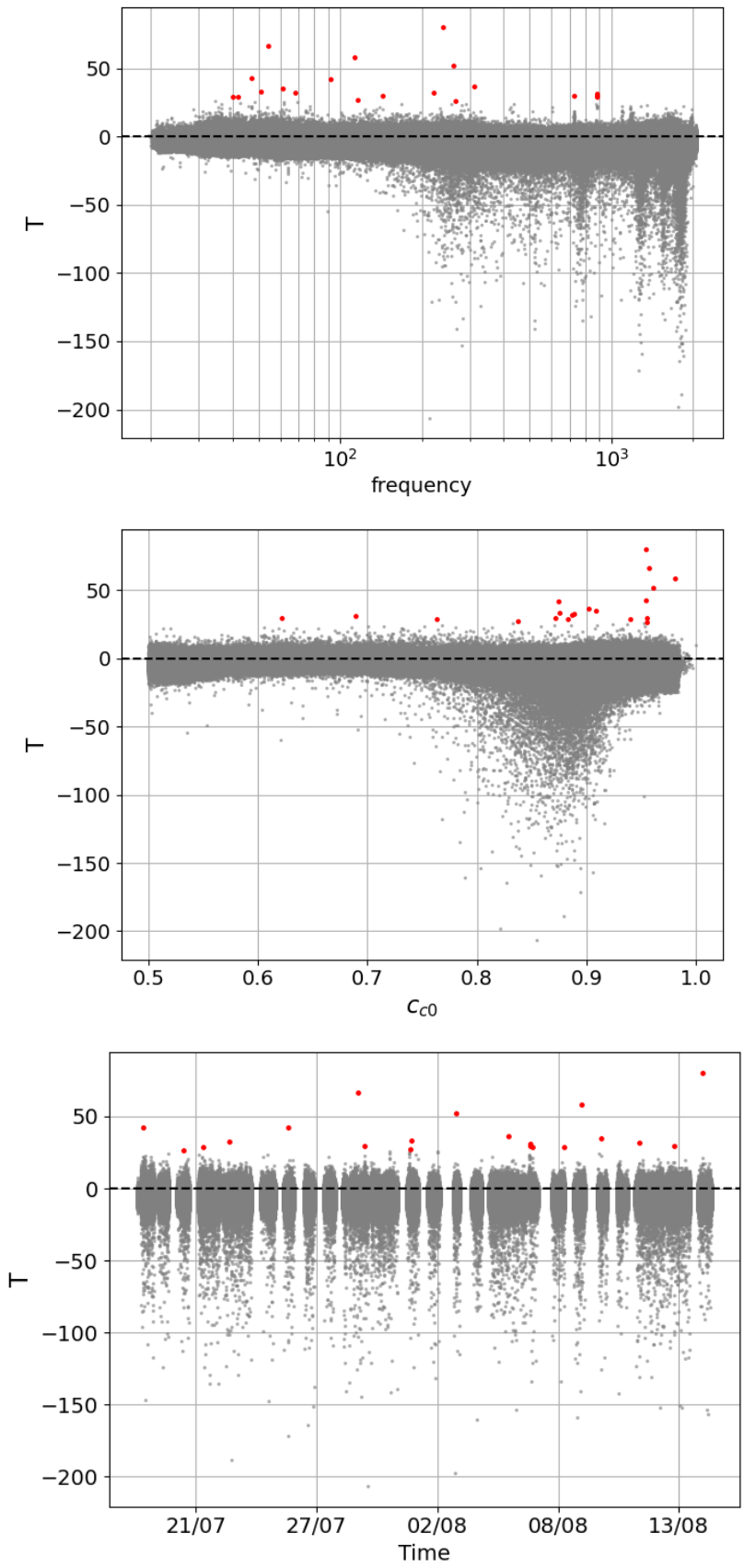


Figure C.3: Plots of loud glitch studies in K05+K06. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

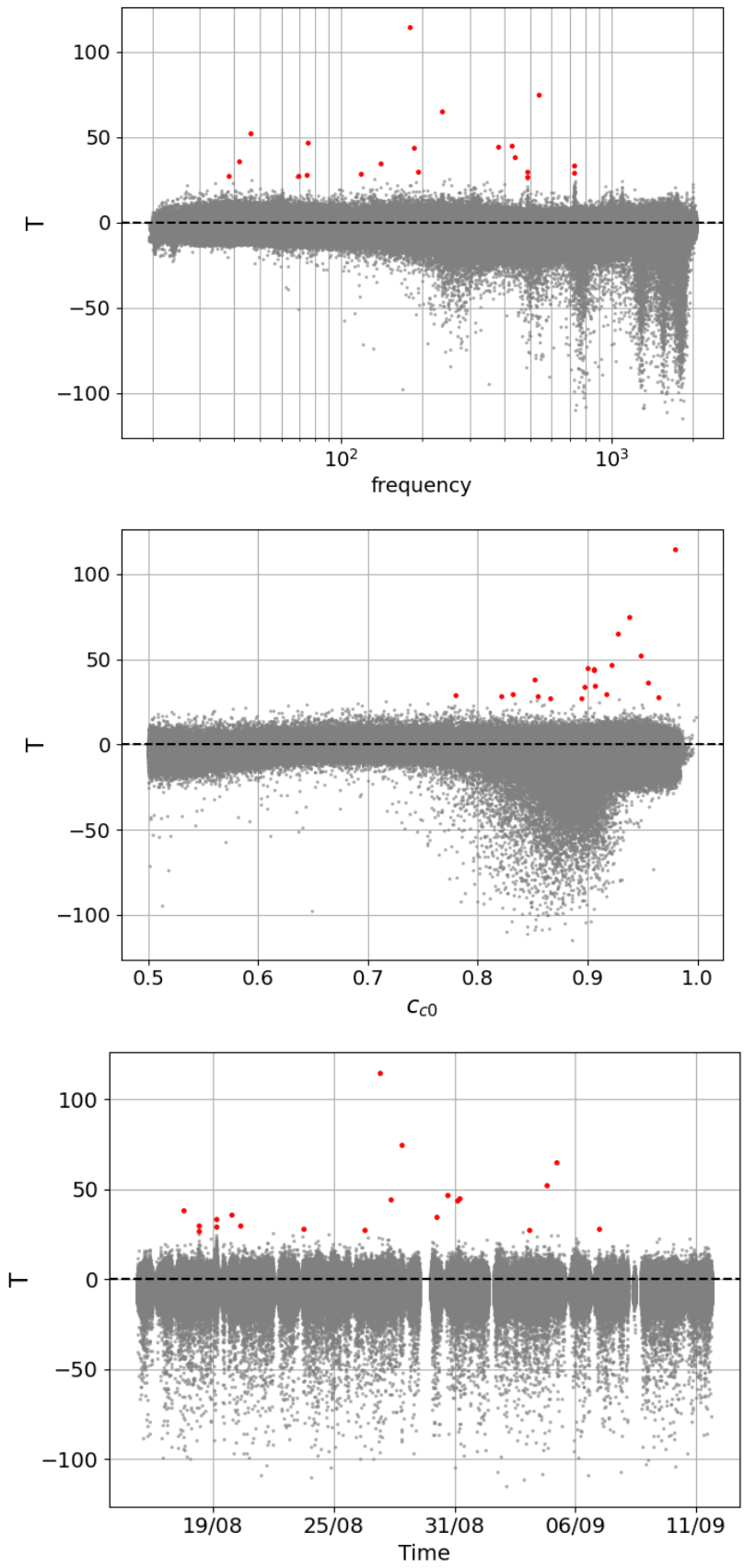


Figure C.4: Plots of loud glitch studies in K07+K08. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

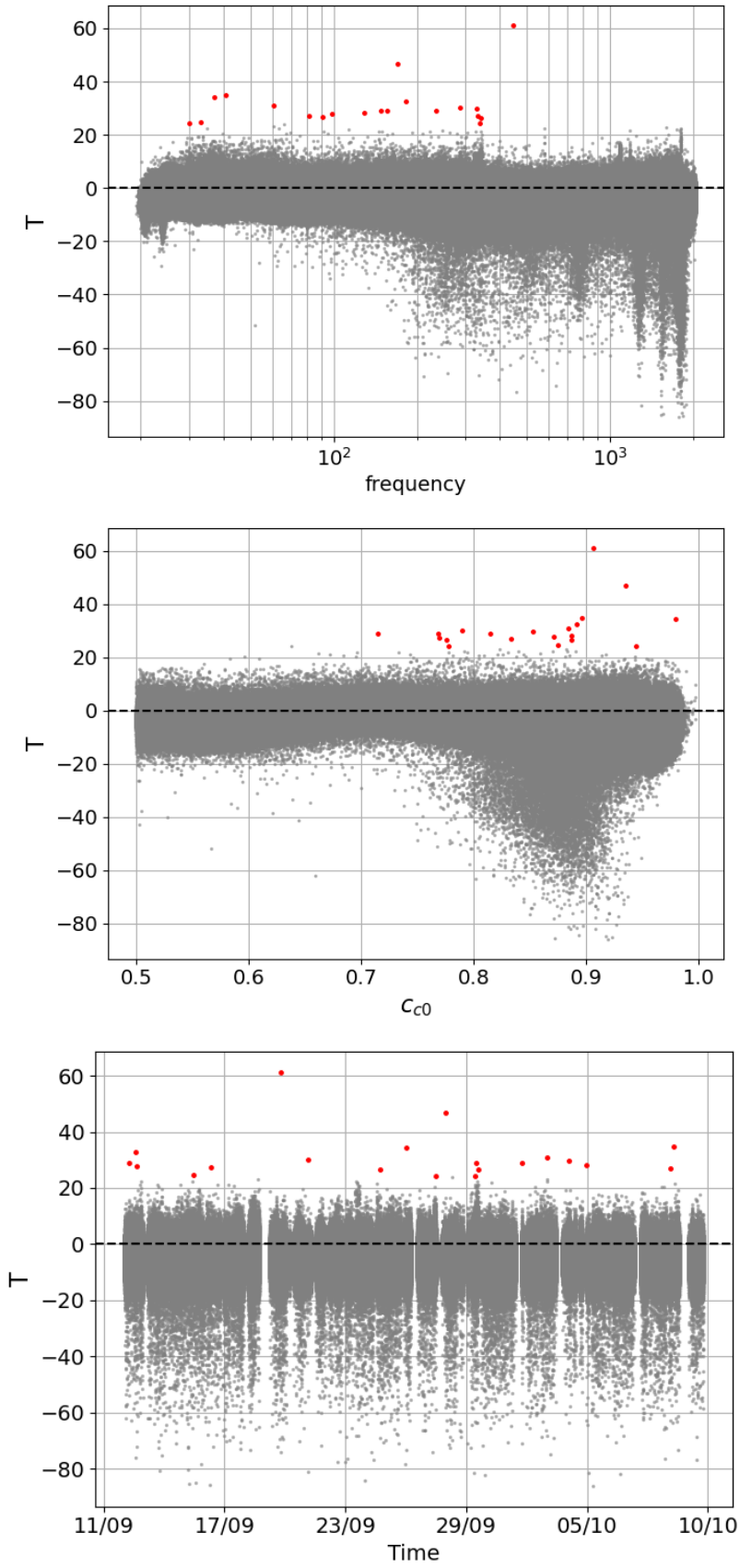


Figure C.5: Plots of loud glitch studies in K09+K10. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

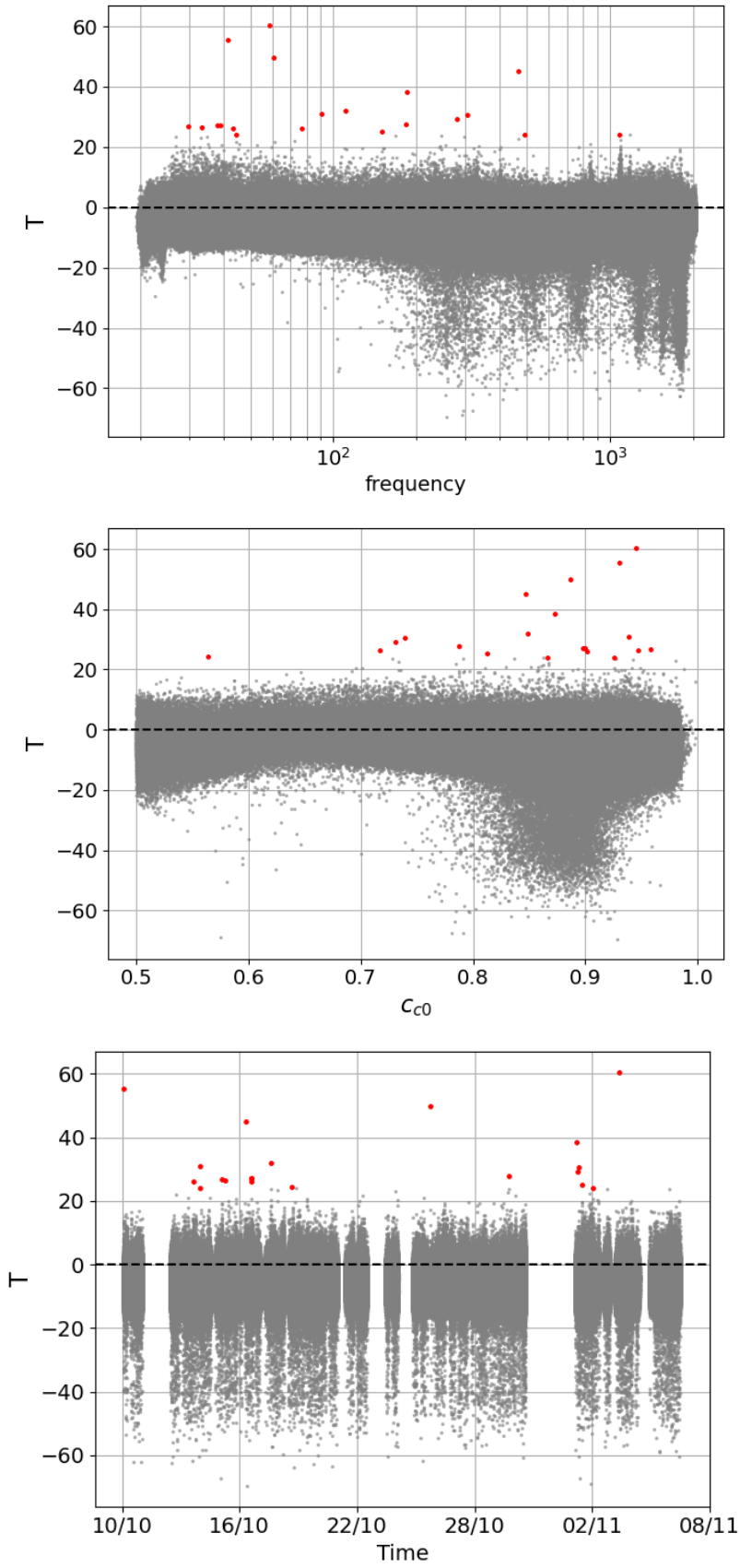


Figure C.6: Plots of loud glitch studies in K11+K12. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

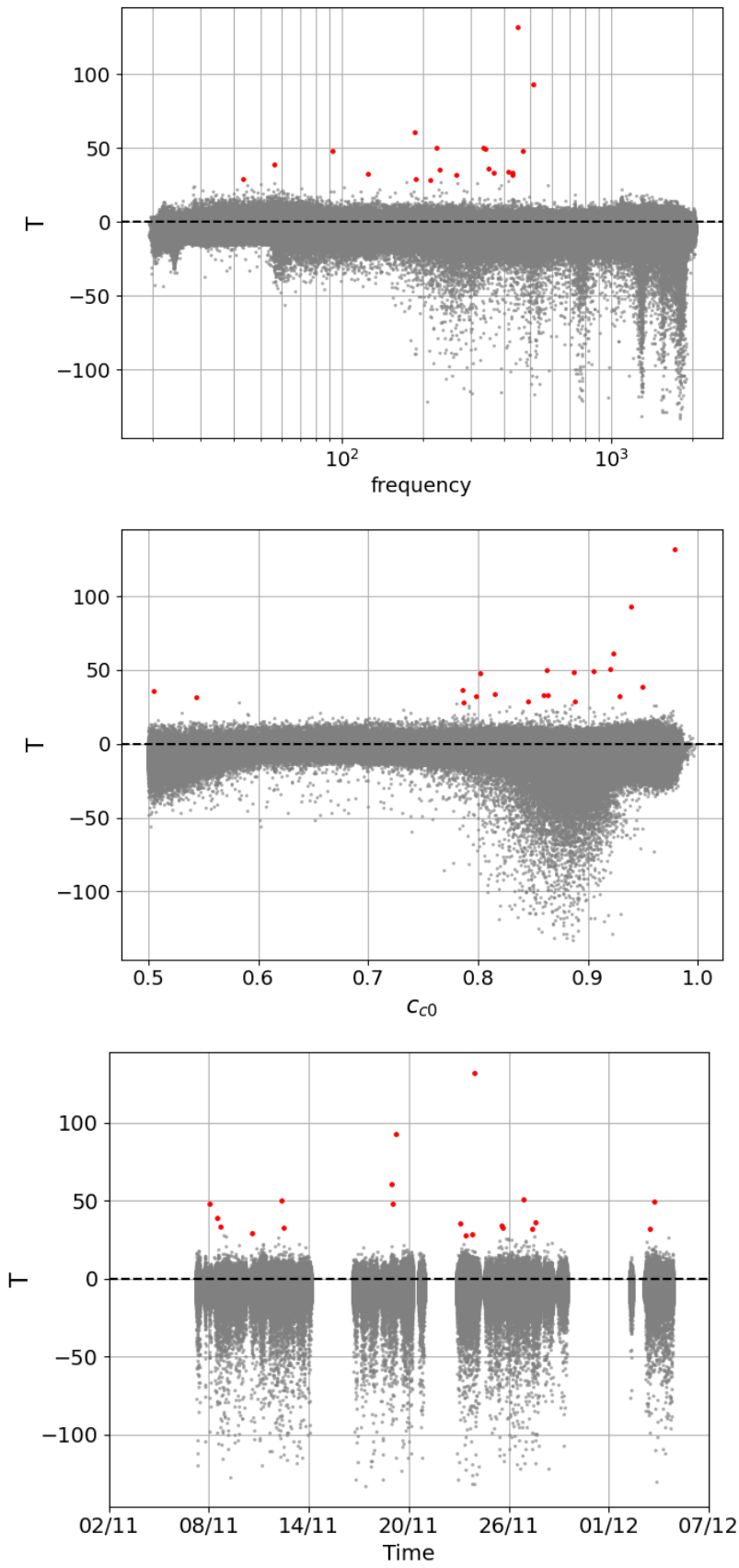


Figure C.7: Plots of loud glitch studies in K13+K14. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

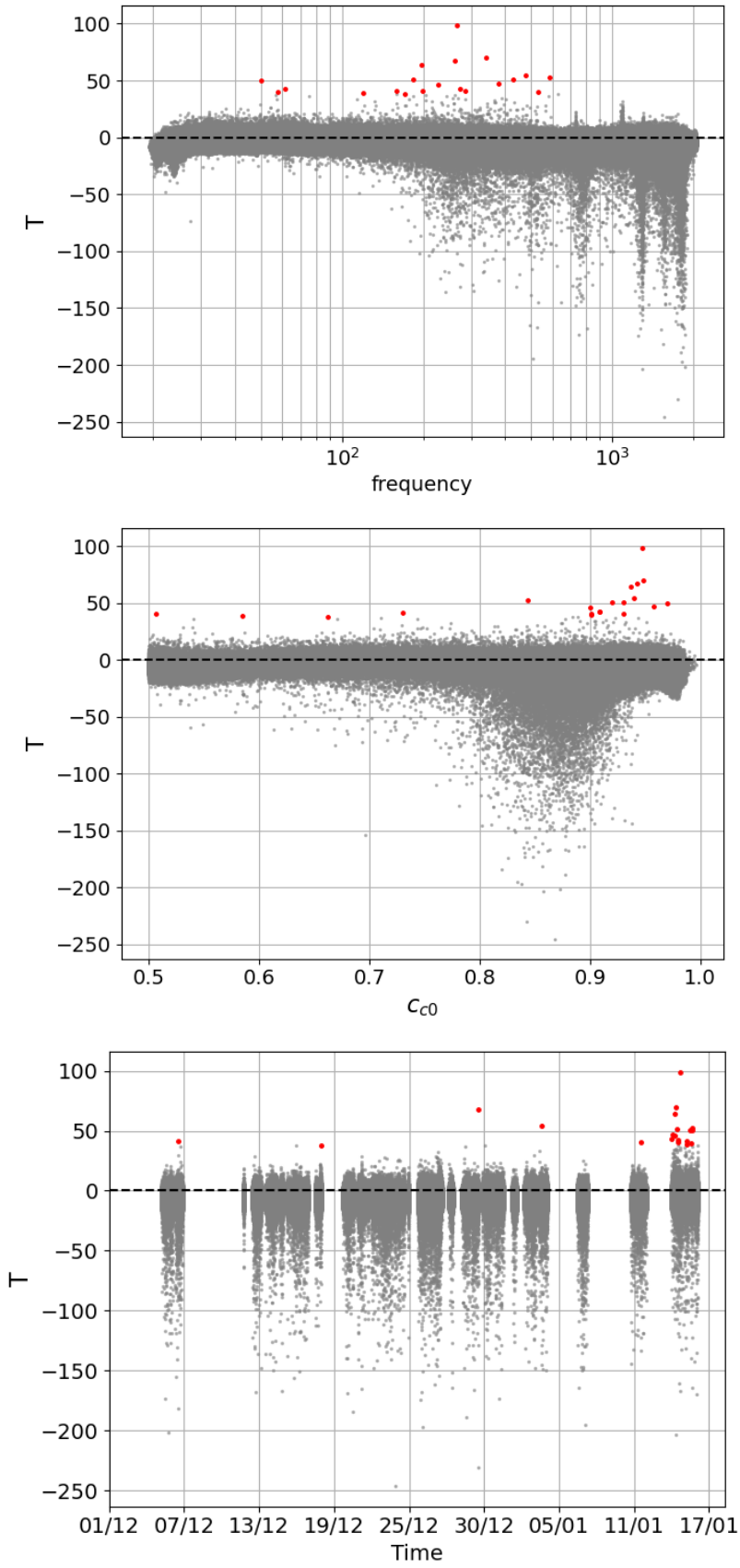


Figure C.8: Plots of loud glitch studies in K15+K16. Distributions of T vs. frequency (a), T vs. network correlation coefficient c_{c0} (b) and T vs. time (c) are shown for background test data, with the 20 loudest glitches marked in red.

Bibliography

[1] V. Gayathri, D. Lopez, R. S. Pranjal, I. S. Heng, A. Pai, and C. Messenger, "Enhancing the sensitivity of transient gravitational wave searches with Gaussian mixture models," *Phys. Rev. D*, vol. 102, no. 10, p. 104023, 2020.

[2] D. Lopez, V. Gayathri, A. Pai, I. S. Heng, C. Messenger, and S. K. Gupta, "Utilizing Gaussian mixture models in all-sky searches for short-duration gravitational wave bursts," *Phys. Rev. D*, vol. 105, no. 6, p. 063024, 2022.

[3] L. Smith, S. Ghosh, J. Sun, V. Gayathri, I. S. Heng, and A. Pai, "Enhancing search pipelines for short gravitational-wave transients with Gaussian mixture modeling," *Phys. Rev. D*, vol. 110, no. 8, p. 083032, 2024.

[4] M. Ebersold, S. Tiwari, L. Smith, Y.-B. Bae, G. Kang, D. Williams, A. Gopakumar, I. S. Heng, and M. Haney, "Observational limits on the rate of radiation-driven binary black hole capture events," *Phys. Rev. D*, vol. 106, no. 10, p. 104014, 2022.

[5] A. Einstein, "Näherungsweise integration der feldgleichungen der gravitation," *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, vol. 1916, pp. 688–696, 1916.

[6] J. Aasi *et al.*, "Advanced LIGO," *Class. Quant. Grav.*, vol. 32, p. 074001, 2015.

[7] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.*, vol. 116, no. 6, p. 061102, 2016.

[8] 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*, vol. 9, no. 3, p. 031040, 2019.

[9] 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*, vol. 11, p. 021053, 2021.

[10] R. Abbott *et al.*, "GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run," *Phys. Rev. D*, vol. 109, no. 2, p. 022001, 2024.

[11] R. Abbott *et al.*, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run," *Phys. Rev. X*, vol. 13, no. 4, p. 041039, 2023.

[12] 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.*, vol. 882, no. 2, p. L24, 2019.

[13] R. Abbott *et al.*, "Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3," *Phys. Rev. X*, vol. 13, no. 1, p. 011048, 2023.

[14] B. P. Abbott *et al.*, "A gravitational-wave standard siren measurement of the Hubble constant," *Nature*, vol. 551, no. 7678, pp. 85–88, 2017.

[15] 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.*, vol. 909, no. 2, p. 218, 2021.

[16] R. Abbott *et al.*, "Constraints on the Cosmic Expansion History from GWTC–3," *Astrophys. J.*, vol. 949, no. 2, p. 76, 2023.

[17] B. P. Abbott *et al.*, "GW170817: Measurements of neutron star radii and equation of state," *Phys. Rev. Lett.*, vol. 121, no. 16, p. 161101, 2018.

[18] B. P. Abbott *et al.*, "Constraining the p-Mode–g-Mode Tidal Instability with GW170817," *Phys. Rev. Lett.*, vol. 122, no. 6, p. 061104, 2019.

[19] M. Szczepanczyk *et al.*, "Detecting and reconstructing gravitational waves from the next galactic core-collapse supernova in the advanced detector era," *Phys. Rev. D*, vol. 104, no. 10, p. 102002, 2021.

[20] M. J. Szczepańczyk *et al.*, "An Optically Targeted Search for Gravitational Waves emitted by Core-Collapse Supernovae during the Third Observing Run of Advanced LIGO and Advanced Virgo," 5 2023.

[21] B. P. Abbott *et al.*, "Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1," *Phys. Rev. D*, vol. 100, no. 10, p. 104036, 2019.

[22] R. Abbott *et al.*, "Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog," *Phys. Rev. D*, vol. 103, no. 12, p. 122002, 2021.

[23] R. Abbott *et al.*, "Tests of General Relativity with GWTC-3," 12 2021.

[24] B. F. Schutz and R. Ricci, "Gravitational waves, sources and detectors," pp. 11–87, 2001.

[25] B. S. Sathyaprakash and B. F. Schutz, "Physics, Astrophysics and Cosmology with Gravitational Waves," *Living Rev. Rel.*, vol. 12, p. 2, 2009.

[26] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007.

[27] L. Lehner and F. Pretorius, "Numerical Relativity and Astrophysics," *Ann. Rev. Astron. Astrophys.*, vol. 52, pp. 661–694, 2014.

[28] C. Cutler and E. E. Flanagan, "Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form?," *Phys. Rev. D*, vol. 49, pp. 2658–2697, 1994.

[29] J. Weber, "Observation of the Thermal Fluctuations of a Gravitational-Wave Detector," *Phys. Rev. Lett.*, vol. 17, pp. 1228–1230, 1966.

[30] J. L. Levine, "Early gravity-wave detection experiments, 1960-1975," *Phys. Perspect.*, vol. 6, pp. 42–75, 2004.

[31] O. D. Aguiar, "The Past, Present and Future of the Resonant-Mass Gravitational Wave Detectors," *Res. Astron. Astrophys.*, vol. 11, pp. 1–42, 2011.

[32] G. Hobbs and S. Dai, "Gravitational wave research using pulsar timing arrays," *Natl. Sci. Rev.*, vol. 4, no. 5, pp. 707–717, 2017.

[33] R. D. Ferdman *et al.*, "The European Pulsar Timing Array: current efforts and a LEAP toward the future," *Class. Quant. Grav.*, vol. 27, p. 084014, 2010.

[34] A. Brazier *et al.*, "The NANOGrav Program for Gravitational Waves and Fundamental Physics," 8 2019.

[35] J. Antoniadis *et al.*, "The second data release from the European Pulsar Timing Array - III. Search for gravitational wave signals," *Astron. Astrophys.*, vol. 678, p. A50, 2023.

[36] G. Agazie *et al.*, "The NANOGrav 15 yr Data Set: Running of the Spectral Index," 8 2024.

[37] D. J. Reardon *et al.*, "Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array," *Astrophys. J. Lett.*, vol. 951, no. 1, p. L6, 2023.

[38] F. Acernese *et al.*, "Advanced Virgo: a second-generation interferometric gravitational wave detector," *Class. Quant. Grav.*, vol. 32, no. 2, p. 024001, 2015.

[39] B. P. Abbott *et al.*, "Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA," *Living Rev. Rel.*, vol. 19, p. 1, 2016.

[40] M. Maggiore *et al.*, "Science Case for the Einstein Telescope," *JCAP*, vol. 03, p. 050, 2020.

[41] D. Reitze *et al.*, "Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO," *Bull. Am. Astron. Soc.*, vol. 51, no. 7, p. 035, 2019.

[42] E. D. Hall and M. Evans, "Metrics for next-generation gravitational-wave detectors," *Class. Quant. Grav.*, vol. 36, no. 22, p. 225002, 2019.

[43] P. Amaro-Seoane *et al.*, "Laser Interferometer Space Antenna," 2 2017.

[44] J. Luo *et al.*, "TianQin: a space-borne gravitational wave detector," *Class. Quant. Grav.*, vol. 33, no. 3, p. 035010, 2016.

[45] S. Kawamura *et al.*, "The Japanese space gravitational wave antenna: DECIGO," *Class. Quant. Grav.*, vol. 28, p. 094011, 2011.

[46] S. Babak, A. Petiteau, and M. Hewitson, "LISA Sensitivity and SNR Calculations," 8 2021.

[47] J. Mei *et al.*, "The TianQin project: current progress on science and technology," *PTEP*, vol. 2021, no. 5, p. 05A107, 2021.

[48] S. Kawamura, "Space gravitational wave antenna DECIGO and B-DECIGO," *PoS*, vol. ICRC2023, p. 1516, 2023.

[49] J. D. E. Creighton and W. G. Anderson, *Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis*. 2011.

[50] B. P. Abbott *et al.*, "Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA," *Living Rev. Rel.*, vol. 19, p. 1, 2016.

[51] B. Abbot *et al.*, "Ligo: the laser interferometer gravitational-wave observatory," *Reports on Progress in Physics*, vol. 72, no. 7, 2009.

[52] B. J. Meers, "Recycling in Laser Interferometric Gravitational Wave Detectors," *Phys. Rev. D*, vol. 38, pp. 2317–2326, 1988.

[53] S. Hild, H. Grote, M. Hewitson, H. Lueck, J. R. Smith, B. Willke, K. Danzmann, and K. A. Strain, "Demonstration and comparison of tuned and detuned signal recycling in a large-scale gravitational wave detector," *Class. Quant. Grav.*, vol. 24, pp. 1513–1523, 2007.

[54] B. F. Schutz, "Networks of gravitational wave detectors and three figures of merit," *Class. Quant. Grav.*, vol. 28, p. 125023, 2011.

[55] H. Luck *et al.*, "The upgrade of GEO600," *J. Phys. Conf. Ser.*, vol. 228, p. 012012, 2010.

[56] C. W. Jameson Rollins and R. Adhikari, "pygwinc," Sep 2024.

[57] W. A. Edelstein, J. Hough, J. R. Pugh, and W. Martin, "Limits to the measurement of displacement in an interferometric gravitational radiation detector," *Journal of Physics E: Scientific Instruments*, vol. 11, p. 710, jul 1978.

[58] M. Pitkin, S. Reid, S. Rowan, and J. Hough, "Gravitational Wave Detection by Interferometry (Ground and Space)," *Living Rev. Rel.*, vol. 14, p. 5, 2011.

[59] J. Aasi *et al.*, "Enhancing the sensitivity of the LIGO gravitational wave detector by using squeezed states of light," *Nature Photon.*, vol. 7, pp. 613–619, 2013.

[60] F. Matichard *et al.*, "Seismic isolation of Advanced LIGO: Review of strategy, instrumentation and performance," *Class. Quant. Grav.*, vol. 32, no. 18, p. 185003, 2015.

[61] P. R. Saulson, "TERRESTRIAL GRAVITATIONAL NOISE ON A GRAVITATIONAL WAVE ANTENNA," *Phys. Rev. D*, vol. 30, pp. 732–736, 1984.

[62] G. Gonzalez, "Suspensions thermal noise in LIGO gravitational wave detector," *Class. Quant. Grav.*, vol. 17, pp. 4409–4436, 2000.

[63] G. D. Hammond, A. V. Cumming, J. Hough, R. Kumar, K. Tokmakov, S. Reid, and S. Rowan, "Reducing the suspension thermal noise of advanced gravitational wave detectors," *Class. Quant. Grav.*, vol. 29, p. 124009, 2012.

[64] G. M. Harry *et al.*, "Thermal noise in interferometric gravitational wave detectors due to dielectric optical coatings," *Class. Quant. Grav.*, vol. 19, pp. 897–918, 2002.

[65] B. P. Abbott *et al.*, "Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914," *Class. Quant. Grav.*, vol. 33, no. 13, p. 134001, 2016.

[66] M. Zevin *et al.*, "Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science," *Class. Quant. Grav.*, vol. 34, no. 6, p. 064003, 2017.

[67] M. Zevin *et al.*, "Gravity Spy: lessons learned and a path forward," *Eur. Phys. J. Plus*, vol. 139, no. 1, p. 100, 2024.

[68] D. Davis *et al.*, "LIGO detector characterization in the second and third observing runs," *Class. Quant. Grav.*, vol. 38, no. 13, p. 135014, 2021.

[69] M. Cabero *et al.*, "Blip glitches in Advanced LIGO data," *Class. Quant. Grav.*, vol. 36, no. 15, p. 15, 2019.

[70] A. H. Nitz, "Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers," *Class. Quant. Grav.*, vol. 35, no. 3, p. 035016, 2018.

[71] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, "Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal," *Phys. Rev. D*, vol. 93, no. 4, p. 044006, 2016.

[72] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, "Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era," *Phys. Rev. D*, vol. 93, no. 4, p. 044007, 2016.

[73] A. Nitz *et al.*, "gwastro/pycbc: v2.3.3 release of pycbc," Jan. 2024.

[74] B. P. Abbott *et al.*, "A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals," *Class. Quant. Grav.*, vol. 37, no. 5, p. 055002, 2020.

[75] N. Christensen and R. Meyer, "Parameter estimation with gravitational waves," *Rev. Mod. Phys.*, vol. 94, no. 2, p. 025001, 2022.

[76] E. Abdikamalov, G. Pagliaroli, and D. Radice, "Gravitational Waves from Core-Collapse Supernovae," 10 2020.

[77] C. J. Hogan and M. J. Rees, "Gravitational interactions of cosmic strings," *Nature*, vol. 311, pp. 109–113, 1984.

[78] A. Vilenkin, "Gravitational radiation from cosmic strings," *Phys. Lett. B*, vol. 107, pp. 47–50, 1981.

[79] R. Abbott *et al.*, "Search for Gravitational Waves Associated with Gamma-Ray Bursts Detected by Fermi and Swift during the LIGO–Virgo Run O3b," *Astrophys. J.*, vol. 928, no. 2, p. 186, 2022.

[80] R. Abbott *et al.*, "Search for Gravitational Waves Associated with Gamma-Ray Bursts Detected by Fermi and Swift during the LIGO–Virgo Run O3b," *Astrophys. J.*, vol. 928, no. 2, p. 186, 2022.

[81] G. Yim and D. I. Jones, "Transient gravitational waves from pulsar post-glitch recoveries," *Mon. Not. Roy. Astron. Soc.*, vol. 498, no. 3, pp. 3138–3152, 2020.

[82] D. Lopez, S. Tiwari, M. Drago, D. Keitel, C. Lazzaro, and G. A. Prodi, "Prospects for detecting and localizing short-duration transient gravitational waves from glitching neutron stars without electromagnetic counterparts," *Phys. Rev. D*, vol. 106, no. 10, p. 103037, 2022.

[83] S. Bini, S. Tiwari, Y. Xu, L. Smith, M. Ebersold, G. Principe, M. Haney, P. Jetzer, and G. A. Prodi, "Search for hyperbolic encounters of compact objects in the third LIGO-Virgo-KAGRA observing run," *Phys. Rev. D*, vol. 109, no. 4, p. 042009, 2024.

[84] Y.-B. Bae, H. M. Lee, G. Kang, and J. Hansen, "Gravitational radiation driven capture in unequal mass black hole encounters," *Phys. Rev. D*, vol. 96, no. 8, p. 084009, 2017.

[85] R. Abbott *et al.*, "All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run," *Phys. Rev. D*, vol. 104, no. 12, p. 122004, 2021.

[86] R. Abbott *et al.*, "All-sky search for long-duration gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run," *Phys. Rev. D*, vol. 104, no. 10, p. 102001, 2021.

[87] R. Abbott *et al.*, "Search for Gravitational Waves Associated with Gamma-Ray Bursts Detected by Fermi and Swift During the LIGO-Virgo Run O3a," *Astrophys. J.*, vol. 915, no. 2, p. 86, 2021.

[88] R. Abbott *et al.*, "Search for Gravitational Waves Associated with Fast Radio Bursts Detected by CHIME/FRB during the LIGO–Virgo Observing Run O3a," *Astrophys. J.*, vol. 955, no. 2, p. 155, 2023.

[89] R. Abbott *et al.*, "Search for Gravitational-wave Transients Associated with Magnetar Bursts in Advanced LIGO and Advanced Virgo Data from the Third Observing Run," *Astrophys. J.*, vol. 966, no. 1, p. 137, 2024.

[90] R. Abbott *et al.*, "Constraints on Cosmic Strings Using Data from the Third Advanced LIGO–Virgo Observing Run," *Phys. Rev. Lett.*, vol. 126, no. 24, p. 241102, 2021.

[91] A. G. Abac *et al.*, "Search for Eccentric Black Hole Coalescences during the Third Observing Run of LIGO and Virgo," 8 2023.

[92] K. Riles, "Searches for continuous-wave gravitational radiation," *Living Rev. Rel.*, vol. 26, no. 1, p. 3, 2023.

[93] R. Tenorio, D. Keitel, and A. M. Sintes, "Search Methods for Continuous Gravitational-Wave Signals from Unknown Sources in the Advanced-Detector Era," *Universe*, vol. 7, no. 12, p. 474, 2021.

[94] R. Abbott *et al.*, "Narrowband Searches for Continuous and Long-duration Transient Gravitational Waves from Known Pulsars in the LIGO-Virgo Third Observing Run," *Astrophys. J.*, vol. 932, no. 2, p. 133, 2022.

[95] R. Abbott *et al.*, "All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data," *Phys. Rev. D*, vol. 106, no. 10, p. 102008, 2022.

[96] T. Regimbau, "The astrophysical gravitational wave stochastic background," *Res. Astron. Astrophys.*, vol. 11, pp. 369–390, 2011.

[97] B. P. Abbott *et al.*, "Constraints on cosmic strings using data from the first Advanced LIGO observing run," *Phys. Rev. D*, vol. 97, no. 10, p. 102002, 2018.

[98] R. Easther, J. T. Giblin, Jr., and E. A. Lim, "Gravitational Wave Production At The End Of Inflation," *Phys. Rev. Lett.*, vol. 99, p. 221301, 2007.

[99] R. Abbott *et al.*, "All-sky, all-frequency directional search for persistent gravitational waves from Advanced LIGO's and Advanced Virgo's first three observing runs," *Phys. Rev. D*, vol. 105, no. 12, p. 122001, 2022.

[100] R. Abbott *et al.*, "Upper limits on the isotropic gravitational-wave background from Advanced LIGO and Advanced Virgo's third observing run," *Phys. Rev. D*, vol. 104, no. 2, p. 022004, 2021.

[101] B. P. Abbott *et al.*, "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," *Phys. Rev. Lett.*, vol. 119, no. 16, p. 161101, 2017.

[102] B. P. Abbott *et al.*, "Multi-messenger Observations of a Binary Neutron Star Merger," *Astrophys. J. Lett.*, vol. 848, no. 2, p. L12, 2017.

[103] R. Abbott *et al.*, "Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences," *Astrophys. J. Lett.*, vol. 915, no. 1, p. L5, 2021.

[104] R. Abbott *et al.*, "GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$," *Phys. Rev. Lett.*, vol. 125, no. 10, p. 101102, 2020.

[105] M. Szczepańczyk *et al.*, "Observing an intermediate-mass black hole GW190521 with minimal assumptions," *Phys. Rev. D*, vol. 103, no. 8, p. 082002, 2021.

[106] R. Gamba, M. Breschi, G. Carullo, S. Albanesi, P. Rettegno, S. Bernuzzi, and A. Nagar, "GW190521 as a dynamical capture of two nonspinning black holes," *Nature Astron.*, vol. 7, no. 1, pp. 11–17, 2023.

[107] J. C. Aurrekoetxea, C. Hoy, and M. Hannam, "Revisiting the Cosmic String Origin of GW190521," *Phys. Rev. Lett.*, vol. 132, no. 18, p. 181401, 2024.

[108] B. P. Abbott *et al.*, "Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of advanced LIGO and advanced Virgo," *Phys. Rev. D*, vol. 101, no. 8, p. 084002, 2020.

[109] J. Abadie *et al.*, "A search for gravitational waves associated with the August 2006 timing glitch of the Vela pulsar," *Phys. Rev. D*, vol. 83, p. 042001, 2011.

[110] M. Ebersold and S. Tiwari, "Search for nonlinear memory from subsolar mass compact binary mergers," *Phys. Rev. D*, vol. 101, no. 10, p. 104041, 2020.

[111] D. Lopez, S. Tiwari, and M. Ebersold, "Gravitational wave memory of compact binary coalescence in the presence of matter effects," *Phys. Rev. D*, vol. 109, no. 4, p. 043039, 2024.

[112] J. Calderón Bustillo, N. Sanchis-Gual, A. Torres-Forné, and J. A. Font, "Confusing Head-On Collisions with Precessing Intermediate-Mass Binary Black Hole Mergers," *Phys. Rev. Lett.*, vol. 126, no. 20, p. 201101, 2021.

[113] R. Abbott *et al.*, "Search for intermediate-mass black hole binaries in the third observing run of Advanced LIGO and Advanced Virgo," *Astron. Astrophys.*, vol. 659, p. A84, 2022.

[114] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, "A coherent method for detection of gravitational wave bursts," *Classical and Quantum Gravity*, vol. 25, p. 114029, may 2008.

[115] S. Klimenko *et al.*, "Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors," *Phys. Rev. D*, vol. 93, no. 4, p. 042004, 2016.

[116] M. Drago *et al.*, "Coherent WaveBurst, a pipeline for unmodeled gravitational-wave data analysis," 6 2020.

[117] N. J. Cornish and T. B. Littenberg, "BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches," *Class. Quant. Grav.*, vol. 32, no. 13, p. 135012, 2015.

[118] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, "BayesWave analysis pipeline in the era of gravitational wave observations," *Phys. Rev. D*, vol. 103, no. 4, p. 044006, 2021.

[119] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, "Omicron: a tool to characterize transient noise in gravitational-wave detectors," *SoftwareX*, vol. 12, p. 100620, 2020.

[120] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, "Information-theoretic approach to the gravitational-wave burst detection problem," *Phys. Rev. D*, vol. 95, no. 10, p. 104046, 2017.

[121] V. Skliris, M. R. K. Norman, and P. J. Sutton, "Toward real-time detection of unmodeled gravitational wave transients using convolutional neural networks," *Phys. Rev. D*, vol. 110, no. 10, p. 104034, 2024.

[122] R. Raikman *et al.*, "GWAK: gravitational-wave anomalous knowledge with recurrent autoencoders," *Mach. Learn. Sci. Tech.*, vol. 5, no. 2, p. 025020, 2024.

[123] T. Mishra, B. O'Brien, V. Gayathri, M. Szczepanczyk, S. Bhaumik, I. Bartos, and S. Klimenko, "Optimization of model independent gravitational wave search for binary black hole mergers using machine learning," *Phys. Rev. D*, vol. 104, no. 2, p. 023014, 2021.

[124] M. J. Szczepańczyk *et al.*, "Search for gravitational-wave bursts in the third Advanced LIGO-Virgo run with coherent WaveBurst enhanced by machine learning," *Phys. Rev. D*, vol. 107, no. 6, p. 062002, 2023.

[125] GWOSC online documentation on detection of GW150914 with cWB, https://gwburst.gitlab.io/info/gw150914/.

[126] B. P. Abbott *et al.*, "All-sky search for short gravitational-wave bursts in the first Advanced LIGO run," *Phys. Rev. D*, vol. 95, no. 4, p. 042003, 2017.

[127] B. P. Abbott *et al.*, "All-Sky Search for Short Gravitational-Wave Bursts in the Second Advanced LIGO and Advanced Virgo Run," *Phys. Rev. D*, vol. 100, no. 2, p. 024017, 2019.

[128] J. Abadie *et al.*, "All-sky search for gravitational-wave bursts in the second joint LIGO-Virgo run," *Phys. Rev. D*, vol. 85, p. 122007, 2012.

[129] K. Kotake, K. Sato, and K. Takahashi, "Explosion mechanism, neutrino burst, and gravitational wave in core-collapse supernovae," *Rept. Prog. Phys.*, vol. 69, pp. 971–1144, 2006.

[130] C. Ott, "The Gravitational Wave Signature of Core-Collapse Supernovae," *Class. Quant. Grav.*, vol. 26, p. 063001, 2009.

[131] J. Powell and B. Müller, "Gravitational Wave Emission from 3D Explosion Models of Core-Collapse Supernovae with Low and Normal Explosion Energies," *Mon. Not. Roy. Astron. Soc.*, vol. 487, no. 1, pp. 1178–1190, 2019.

[132] E. P. O'Connor and S. M. Couch, "Exploring Fundamentally Three-dimensional Phenomena in High-fidelity Simulations of Core-collapse Supernovae," *Astrophys. J.*, vol. 865, no. 2, p. 81, 2018.

[133] D. Radice, V. Morozova, A. Burrows, D. Vartanyan, and H. Nagakura, "Characterizing the Gravitational Wave Signal from Core-Collapse Supernovae," *Astrophys. J. Lett.*, vol. 876, no. 1, p. L9, 2019.

[134] J. Powell and B. Müller, "Three-dimensional core-collapse supernova simulations of massive and rotating progenitors," *Mon. Not. Roy. Astron. Soc.*, vol. 494, no. 4, pp. 4665–4675, 2020.

[135] M. Obergaulinger and M. A. Aloy, "Magnetorotational core collapse of possible GRB progenitors – I. Explosion mechanisms," *Mon. Not. Roy. Astron. Soc.*, vol. 492, no. 4, pp. 4613–4634, 2020.

[136] T. W. B. Kibble, "Topology of Cosmic Domains and Strings," *J. Phys. A*, vol. 9, pp. 1387–1398, 1976.

[137] T. Damour and A. Vilenkin, "Gravitational wave bursts from cosmic strings," *Phys. Rev. Lett.*, vol. 85, pp. 3761–3764, 2000.

[138] T. Damour and A. Vilenkin, "Gravitational wave bursts from cusps and kinks on cosmic strings," *Phys. Rev. D*, vol. 64, p. 064008, 2001.

[139] J. Hong and H. M. Lee, "Black Hole Binaries in Galactic Nuclei and Gravitational Wave Sources," *Mon. Not. Roy. Astron. Soc.*, vol. 448, no. 1, pp. 754–770, 2015.

[140] J. Samsing, I. Bartos, D. J. D'Orazio, Z. Haiman, B. Kocsis, N. W. C. Leigh, B. Liu, M. E. Pessah, and H. Tagawa, "AGN as potential factories for eccentric black hole mergers," *Nature*, vol. 603, no. 7900, pp. 237–240, 2022.

[141] J. Samsing, "Eccentric Black Hole Mergers Forming in Globular Clusters," *Phys. Rev. D*, vol. 97, no. 10, p. 103014, 2018.

[142] M. Zevin, J. Samsing, C. Rodriguez, C.-J. Haster, and E. Ramirez-Ruiz, "Eccentric Black Hole Mergers in Dense Star Clusters: The Role of Binary–Binary Encounters," *Astrophys. J.*, vol. 871, no. 1, p. 91, 2019.

[143] R. M. O'Leary, B. Kocsis, and A. Loeb, "Gravitational waves from scattering of stellar-mass black holes in galactic nuclei," *Mon. Not. Roy. Astron. Soc.*, vol. 395, no. 4, pp. 2127–2146, 2009.

[144] W. Guo, D. Williams, I. S. Heng, H. Gabbard, Y.-B. Bae, G. Kang, and Z.-H. Zhu, "Mimicking mergers: mistaking black hole captures as mergers," *Mon. Not. Roy. Astron. Soc.*, vol. 516, no. 3, pp. 3847–3860, 2022.

[145] G. Cho, A. Gopakumar, M. Haney, and H. M. Lee, "Gravitational waves from compact binaries in post-Newtonian accurate hyperbolic orbits," *Phys. Rev. D*, vol. 98, no. 2, p. 024039, 2018.

[146] S. Dandapat, M. Ebersold, A. Susobhanan, P. Rana, A. Gopakumar, S. Tiwari, M. Haney, H. M. Lee, and N. Kolhe, "Gravitational waves from black-hole encounters: Prospects for ground and galaxy-based observatories," *Phys. Rev. D*, vol. 108, no. 2, p. 024013, 2023.

[147] G. Morrás, J. García-Bellido, and S. Nesseris, "Search for black hole hyperbolic encounters with gravitational wave detectors," *Phys. Dark Univ.*, vol. 35, p. 100932, 2022.

[148] D. Christodoulou, "Nonlinear nature of gravitation and gravitational wave experiments," *Phys. Rev. Lett.*, vol. 67, pp. 1486–1489, 1991.

[149] K. S. Thorne, "Gravitational-wave bursts with memory: The Christodoulou effect," *Phys. Rev. D*, vol. 45, no. 2, pp. 520–524, 1992.

[150] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, "Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog," *Phys. Rev. D*, vol. 101, no. 2, p. 023011, 2020.

[151] V. Necula, S. Klimenko, and G. Mitselmakher, "Transient analysis with fast Wilson-Daubechies time-frequency transform," *J. Phys. Conf. Ser.*, vol. 363, p. 012032, 2012.

[152] V. Tiwari *et al.*, "Regression of Environmental Noise in LIGO Data," *Class. Quant. Grav.*, vol. 32, no. 16, p. 165014, 2015.

[153] S. Klimenko, S. Mohanty, M. Rakhmanov, and G. Mitselmakher, "Constraint likelihood analysis for a network of gravitational wave detectors," *Phys. Rev. D*, vol. 72, p. 122002, 2005.

[154] B. P. Abbott *et al.*, "Observing gravitational-wave transient GW150914 with minimal assumptions," *Phys. Rev. D*, vol. 93, no. 12, p. 122004, 2016. [Addendum: Phys.Rev.D 94, 069903 (2016)].

[155] "coherent waveburst documentation pages." https://gwburst.gitlab.io/documentation/latest/html/index.html. Accessed: 2024-06-24.

[156] B. P. Abbott *et al.*, "Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914," *Phys. Rev. D*, vol. 95, no. 6, p. 062003, 2017.

[157] J. Golomb and C. Talbot, "Hierarchical Inference of Binary Neutron Star Mass Distribution and Equation of State with Gravitational Waves," *Astrophys. J.*, vol. 926, no. 1, p. 79, 2022.

[158] D. Keitel, "Galactic Double Neutron Star total masses and Gaussian mixture model selection," *Mon. Not. Roy. Astron. Soc.*, vol. 485, no. 2, pp. 1665–1674, 2019.

[159] S. Rinaldi and W. Del Pozzo, "(H)DPGMM: a hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function," *Mon. Not. Roy. Astron. Soc.*, vol. 509, no. 4, pp. 5454–5466, 2021.

[160] W. Del Pozzo, C. P. Berry, A. Ghosh, T. S. F. Haines, L. P. Singer, and A. Vecchio, "Dirichlet Process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations," *Mon. Not. Roy. Astron. Soc.*, vol. 479, no. 1, pp. 601–614, 2018.

[161] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[162] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[163] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[164] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[165] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[166] S. Vinciguerra, M. Drago, G. A. Prodi, S. Klimenko, C. Lazzaro, V. Necula, F. Salemi, V. Tiwari, M. C. Tringali, and G. Vedovato, "Enhancing the significance of gravitational wave bursts through signal classification," *Class. Quant. Grav.*, vol. 34, no. 9, p. 094003, 2017.

[167] M. Cavaglia, S. Gaudio, T. Hansen, K. Staats, M. Szczepanczyk, and M. Zanolin, "Improving the background of gravitational-wave searches for core collapse supernovae: A machine learning approach," *Mach. Learn. Sci. Tech.*, vol. 1, p. 015005, 2020.

[168] T. S. Adams, D. Meacher, J. Clark, P. J. Sutton, G. Jones, and A. Minot, "Gravitational-Wave Detection using Multivariate Analysis," *Phys. Rev. D*, vol. 88, p. 062006, 2013.

[169] H. Andresen, B. Muller, E. Muller, and H.-T. Janka, "Gravitational Wave Signals from 3D Neutrino Hydrodynamics Simulations of Core-Collapse Supernovae," *Mon. Not. Roy. Astron. Soc.*, vol. 468, no. 2, pp. 2032–2051, 2017.

[170] E. Muller, H. T. Janka, and A. Wongwathanarat, "Parametrized 3D models of neutrino-driven supernova explosions: Neutrino emission asymmetries and gravitational-wave signals," *Astron. Astrophys.*, vol. 537, p. A63, 2012.

[171] T. Kuroda, K. Kotake, and T. Takiwaki, "A new Gravitational-wave Signature From Standing Accretion Shock Instability in Supernovae," *Astrophys. J. Lett.*, vol. 829, no. 1, p. L14, 2016.

[172] E. Abdikamalov, S. Gossan, A. M. DeMaio, and C. D. Ott, "Measuring the Angular Momentum Distribution in Core-Collapse Supernova Progenitors with Gravitational Waves," *Phys. Rev. D*, vol. 90, no. 4, p. 044001, 2014.

[173] A. L. Piro and E. Pfahl, "Fragmentation of Collapsar Disks and the Production of Gravitational Waves," *Astrophys. J.*, vol. 658, p. 1173, 2007.

[174] M. H. P. M. van Putten, "Proposed source of gravitational radiation from a torus around a black hole," *Phys. Rev. Lett.*, vol. 87, p. 091101, 2001.

[175] S. Klimenko, "Wavescan: multiresolution regression of gravitational-wave data," 1 2022.

[176] T. Mishra *et al.*, "Search for binary black hole mergers in the third observing run of Advanced LIGO-Virgo using coherent WaveBurst enhanced with machine learning," *Phys. Rev. D*, vol. 105, no. 8, p. 083018, 2022.

[177] M. Zevin *et al.*, "Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science," *Class. Quant. Grav.*, vol. 34, no. 6, p. 064003, 2017.

[178] H. Andresen, E. Müller, H. T. Janka, A. Summa, K. Gill, and M. Zanolin, "Gravitational waves from 3D core-collapse supernova models: The impact of moderate progenitor rotation," *Mon. Not. Roy. Astron. Soc.*, vol. 486, no. 2, pp. 2238–2253, 2019.

[179] A. Mezzacappa *et al.*, "Gravitational-wave signal of a core-collapse supernova explosion of a 15 $M_\odot$ star," *Phys. Rev. D*, vol. 102, no. 2, p. 023027, 2020.

[180] K.-C. Pan, M. Liebendörfer, S. M. Couch, and F.-K. Thielemann, "Stellar Mass Black Hole Formation and Multimessenger Signals from Three-dimensional Rotating Core-collapse Supernova Simulations," *Astrophys. J.*, vol. 914, no. 2, p. 140, 2021.

[181] E. Capote *et al.*, "Advanced LIGO detector performance in the fourth observing run," 11 2024.

[182] L. Gondán, B. Kocsis, P. Raffai, and Z. Frei, "Eccentric Black Hole Gravitational-Wave Capture Sources in Galactic Nuclei: Distribution of Binary Parameters," *Astrophys. J.*, vol. 860, no. 1, p. 5, 2018.

[183] L. Gondán and B. Kocsis, "High eccentricities and high masses characterize gravitational-wave captures in galactic nuclei as seen by Earth-based detectors," *Mon. Not. Roy. Astron. Soc.*, vol. 506, no. 2, pp. 1665–1696, 2021.

[184] J. Samsing, D. J. D'Orazio, K. Kremer, C. L. Rodriguez, and A. Askar, "Single-single gravitational-wave captures in globular clusters: Eccentric deci-Hertz sources observable by DECIGO and Tian-Qin," *Phys. Rev. D*, vol. 101, no. 12, p. 123010, 2020.

[185] A. Rasskazov and B. Kocsis, "The rate of stellar mass black hole scattering in galactic nuclei," *Astrophys. J.*, vol. 881, no. 1, p. 20, 2019.

[186] V. Tiwari, "Estimation of the Sensitive Volume for Gravitational-wave Source Populations Using Weighted Monte Carlo Integration," *Class. Quant. Grav.*, vol. 35, no. 14, p. 145009, 2018.

[187] P. R. Brady, J. D. E. Creighton, and A. G. Wiseman, "Upper limits on gravitational-wave signals based on loudest events," *Class. Quant. Grav.*, vol. 21, pp. S1775–S1782, 2004.

[188] B. J. Owen and B. S. Sathyaprakash, "Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement," *Phys. Rev. D*, vol. 60, p. 022002, 1999.

[189] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, "Surrogate models for precessing binary black hole simulations with unequal masses," *Phys. Rev. Research.*, vol. 1, p. 033015, 2019.

[190] J. M. Blondin, A. Mezzacappa, and C. DeMarino, "Stability of standing accretion shocks, with an eye toward core collapse supernovae," *Astrophys. J.*, vol. 584, pp. 971–980, 2003.