



MONASH University

**Bayesian Inference in Astrophysics:
Applications to Gravitational-Wave
Memory and Quasi-Periodic Oscillations**

Moritz Thomas Hübner

Master of Science

A Thesis Submitted for the Degree of Doctor of Philosophy at
Monash University in 2021
School of Physics and Astronomy

Copyright notice

This thesis must be used only under the normal conditions of “fair dealing” under the Copyright Act. It should not be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis. I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner’s permission.

©[Moritz Thomas Hübner](#) (2021)

Abstract

In this thesis, I explore how Bayesian inference is used in the analysis of gravitational waves, and in the analysis of astrophysical time series more broadly. I present the BILBY project, a collaborative software package that has now become a standard analysis tool for gravitational-wave signals. One specific application of Bayesian inference is the search for nonlinear gravitational-wave memory, a hereditary effect in which gravitational-wave strain builds up during an accompanying burst and remains at a constant value different from zero after the wave has passed. This thesis presents the results of measuring memory in the first and second LIGO/Virgo gravitational-wave transient catalogue. This catalogue contains 50 events that have been associated with the coalescence of compact binary objects, with most of them being black holes. Combining the results from these events I find that the hypothesis that memory is present in the catalogue is very slightly favoured but not statistically significant. While memory will not be detectable in a single gravitational-wave signal with current generation detectors, we provide the infrastructure with which to do it in $\mathcal{O}(2000)$ events, a milestone likely reached by the end of the decade.

The generality and versatility of BILBY allows for the easy deployment of inference methods in wide contexts such as in astrophysical x-ray time series. I present the discovery of a new bias in the search for quasi-periodic oscillations using frequency-domain methods that arises in non-stationary time series. Not accounting for this bias can lead to a strong overestimation of the significance of quasi-periodic oscillations, especially when periodograms and Whittle likelihoods are used. I show that one way of mitigating this effect is to employ Gaussian process modelling. Gaussian processes are in principle superior as they can model deterministic trends, non-stationary behaviour, and heteroscedastic data, but are more computationally expensive and involve more complex modelling procedures. Nevertheless, I show that using a narrow class of kernel functions Gaussian processes are a viable method for finding and characterising quasi-periodic oscillations.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: _____

Print Name: Moritz Thomas Hübner _____

Date: August 27, 2021 _____

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes three original papers published in peer-reviewed journals in Chaps. 2, 4, and 5 and one submitted publication in Chap. 6. The core theme of the thesis is Bayesian inference in astrophysics. The ideas, development and writing up of Chaps. 1, 3, 4, 5, 6, 7, 8, and App. A in the thesis were the principal responsibility of myself, the student, working within the School of Physics and Astronomy under the supervision of Paul Lasky. Chap. 2 is the result of substantial code development by myself, Gregory Ashton, and Colm Talbot, initiated by Paul Lasky, with many other contributors. Due to the collaborative nature of this project, it is not examinable under the university rules, however, it is included to preserve the narrative of the thesis.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

My contribution to each chapter published separately is summarised in a table on the page following the next.

I have renumbered sections of submitted or published papers to generate a consistent presentation within the thesis.

Signature: _____

Date: August 27, 2021 _____

I, Paul Lasky, hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Signature: _____

Date: August 27, 2021 _____

Publications during enrolment

1. G. Ashton, **M. Hübner**, P. Lasky, C. Talbot, et al. 2019, *Bilby: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy*, [ApJS](#), **241**, 27, [arxiv:1811.02042](#)
2. **M. Hübner**, C. Talbot, P. Lasky, E. Thrane 2020, *Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog*, [Phys. Rev. D](#), **101**, 023011, [arxiv:1911.12496v1](#)
3. I. M. Romero-Show et al., incl. **M. Hübner** 2020, *Bayesian inference for compact binary coalescences with BILBY: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue*, [MNRAS](#) **499**, 3295–3319 [arxiv:2006.00714](#)
4. **M. Hübner**, P. Lasky, E. Thrane 2021, *Memory remains undetected: Updates from the second LIGO/Virgo gravitational-wave transient catalog*, [Phys. Rev. D](#), **104**, 023004, [arxiv:2105.02879](#)
5. **M. Hübner**, D. Huppenkothen, P. Lasky, A. Inglis 2021, *Pitfalls of periodograms: The non-stationarity bias in the analysis of quasi-periodic oscillations*, submitted to [ApJS](#), [arxiv:2105.02879](#)

I am also a co-author on many LIGO Scientific Collaboration papers which are not listed here and for which my contributions vary.

Thesis chapter	Publication Title	Status	Nature and % of student contribution	Co-author names, Nature and % of Co-authors' contribution	Co-author(s), Monash student
4	Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog	published in Phys. Rev. D , 101 , 023011	75%, analysis, code development, method, writing	C. Talbot/P. Lasky/E. Thrane, 25% concept, method, writing feedback	Yes, C. Talbot
5	Memory remains undetected: Updates from the second LIGO/Virgo gravitational-wave transient catalog	published in Phys. Rev. D , 104 , 023004	75%, analysis, code development, method, writing	P. Lasky/E. Thrane, 25% concept, method, writing feedback	No
6	Pitfalls of peridograms: The non-stationarity bias in the analysis of quasi-periodic oscillations	Submitted to ApJS	75%, analysis, code development, method, writing	D. Huppenkothen/P. Lasky/A. Inglis, 25% concept, method, writing feedback	No

Acknowledgements First of all, I would like to thank my supervisor Paul Lasky for providing me with the opportunity to work on gravitational waves during this exciting time. His supervision and guidance throughout my candidature allowed me to become an independent researcher. Paul encouraged me to take leading roles in projects early on, which enabled me to do the best I could within the research community. I would also like to thank my co-supervisor, Eric Thrane, who always provided great feedback on my results and paper drafts. My gratitude also goes to Daniela Huppenkothen, whom I have worked with over the last year, for being a great collaborator from whom I could learn so much.

I am grateful for the support from my friends and family during my candidature. Most of all, I would like to thank my wonderful girlfriend Rosie (Dương Quỳnh Hoa), who has always supported me despite my erratic work schedule and my pedantic habits. On our trips, Rosie made me appreciate so many things that I would have barely noticed otherwise. I would also like to thank my friends and colleagues at the School of Physics and Astronomy, specifically, Nikhil Sarin and Shanika Galaudage, who have made life and research during lockdown bearable. Going on vacation in Tasmania with Nikhil, Shanika and her sister Nimna, was one of the best times I had in Australia. I am also thankful for my friendship with Isobel Romero-Shaw, her husband Adrian Turner, Melanie Hampel, and Kymberley Bailey, who have made my life in Australia much more fulfilling. My thanks also extend to all my friends living in other countries but with whom I spent a lot of time on calls. Especially Maria Senchermés, whom I feel closer to than ever despite not having seen her in over three years, and Soheb Mandhai who feels like an old friend even though we have never met in person. Finally, I would also like to thank my family, who always supported my endeavours and allowed me to explore my passions without putting pressure on me.

Contents

Copyright notice	i
Abstract	ii
Declaration	iii
Thesis including published works declaration	iv
Publications during enrolment	v
Acknowledgements	vii
List of Figures	xii
List of Tables	xiv
1 Introduction	1
1.1 Bayesian Inference	2
1.1.1 Bayes' Theorem	2
1.1.2 Computational Implementations	4
1.2 Gravitational-wave Memory	4
1.2.1 Linear Memory	4
1.2.2 Nonlinear Memory	5
1.2.2.1 Memory Phenomenology	6
1.2.2.2 Importance of higher-order modes	7
1.2.3 Calculating and measuring Memory in Practice	9
1.2.4 Prospects of measuring Memory	10
1.2.4.1 Ground-based Detectors	10
1.2.4.2 Space-based Instruments	12
1.2.4.3 Pulsar Timing	12
1.3 Quasi-periodic Oscillations	12
1.3.1 Underlying physical processes	13
1.3.2 Analysis methods	13
2 Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy	15
2.1 Introduction	15
2.2 BILBY Design Philosophy	17
2.3 Code Overview	18

2.3.1	Installation	18
2.3.2	Packages	18
2.3.2.1	The <code>core</code> package	19
2.3.2.2	The <code>gw</code> package	19
2.3.3	The <code>hyper</code> package	20
2.4	Compact Binary Coalescence	20
2.4.1	GW150914: the onset of gravitational wave astronomy	20
2.4.2	Binary black hole merger injection	22
2.4.3	Measuring tidal effects in binary neutron star coalescences	26
2.4.4	Implementing New Waveforms	28
2.4.5	Adding detectors to the network	29
2.5	Alternative signal models	30
2.5.1	Supernovae	31
2.5.2	Neutron star post-merger remnant	31
2.6	Population Inference: hyperparameterizations	34
2.7	Analysis of arbitrary data: an example	34
2.8	Conclusion	36
2.9	Acknowledgements	37
3	Further Information about Bilby	39
3.1	Open Source Projects in the Scientific Context	39
3.2	Extensions to the BILBY Package	40
3.2.1	Bilby pipe	40
3.2.2	PBILBY	40
3.2.3	PESUMMARY	41
3.2.4	Bilby-MCMC	41
3.3	Improvements in the BILBY Package	41
3.3.1	Constrained and conditional Priors	41
3.3.2	Boundary Conditions	44
3.4	Evidence Review	44
4	Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog	47
4.1	Introduction	47
4.2	Methods	49
4.2.1	Signal models	49
4.2.2	Bayesian methods	49
4.2.3	Reweighting study	50
4.2.4	Analyzing real events	51
4.3	GWTC-1 Results	52
4.4	Population study	52
4.5	Conclusion and Outlook	55
4.6	Acknowledgements	55
5	Memory Remains Undetected: Updates from the Second LIGO/Virgo Gravitational-Wave Transient Catalog	56
5.1	Introduction	56
5.2	Methods	57

5.3	Results	59
5.3.1	GW190521	61
5.3.2	GW190521_074359 and similar events	61
5.3.3	Comparison with GWTC-1 analysis	62
5.4	Conclusion and Outlook	63
5.5	Acknowledgements	63
6	Pitfalls of periodograms: The non-stationarity bias in the analysis of quasi-periodic oscillations	65
6.1	Introduction	65
6.2	Methods	66
6.2.1	Whittle likelihood	67
6.2.2	Models of power spectra	67
6.2.3	Model selection and parameter estimation	68
6.3	Pitfalls of Periodograms	70
6.3.1	QPOs in stationary noise	72
6.3.2	QPOs in non-stationary red noise	74
6.4	Simulated data	76
6.4.1	Setup	76
6.4.2	Non-stationary QPO in white noise	76
6.4.3	Non-stationary QPO in a simple transient	77
6.4.4	Non-stationary QPO in non-stationary red noise	77
6.4.5	Tests of stationarity	82
6.5	Solar flare data	82
6.6	Discussion and outlook	85
6.7	Acknowledgements	86
6.8	Appendix: Priors and simulated data parameter tables	86
6.9	Appendix: Deterministic processes impact white noise	86
7	Detecting Quasi-periodic Oscillations with Gaussian Processes	94
7.1	Methods	95
7.1.1	Gaussian Processes Overview	95
7.1.2	Kernel Functions	96
7.1.3	Mean Functions	98
7.1.4	Towards non-stationary Models	99
7.1.5	Assumptions and Limits of GPs	100
7.1.6	Bayesian Inference with Gaussian Processes	101
7.2	Simulated Data	103
7.2.1	Percentile-Percentile Analysis	103
7.2.2	Model Selection	104
7.2.3	Non-stationarity Bias	105
7.3	Real data	107
7.3.1	Gamma-ray Bursts	107
7.3.2	Magnetar bursts	109
7.3.3	Giant Magnetar Flare	110
7.3.4	Solar Flares	113
7.4	Discussion and Outlook	114

8	Summary and future directions	116
8.1	Gravitational-wave Memory	116
8.2	Gaussian Processes and Quasi-periodic Oscillations	117
A	Software implementation for gravitational-wave memory	119
A.1	The GWMEMORY package	119
A.2	The MEMESTR package	120
	Bibliography	123

List of Figures

1.1	IMR _{PHENOMXHM} waveform plus mode	8
1.2	$\ln BF$ and memory matched-filter signal-to-noise ratio	11
2.1	GW150914 posterior source-mass distributions	23
2.2	GW150914 posterior inclination and distance distributions	24
2.3	Sky localisation uncertainty for GW150914	25
2.4	Injecting and recovering a binary black hole gravitational-wave signal	26
2.5	Injecting and recovering a binary neutron star gravitational-wave signal.	28
2.6	Sky location uncertainty with Gingin.	30
2.7	Parameter estimation reconstruction of a supernova signal.	32
2.8	Proxy post-merger gravitational-wave signal from a neutron star.	33
2.9	Population modelling with BILBY hierarchical Bayesian inference module.	35
2.10	The ‘B’ from the BILBY logo.	36
2.11	All letters from the BILBY logo, generated using the BILBY package	36
3.1	Evidence review sampler comparison.	45
3.2	Evidence review percentile-percentile test	46
4.1	Memory signal-to-noise ratio vs. evidence.	51
4.2	GWTC-1 Memory $\ln BF$ summary	53
4.3	Cumulative memory Bayes factors obtained for a set of 2000 injections.	54
5.1	GWTC-2 Memory $\ln BF$ summary.	60
5.2	Wide tails of memory $\ln BF$ values.	61
5.3	GW190521_074359 inclination/distance vs. log weights	64
6.1	Effects of extending a time series with zeros.	71
6.2	Non-stationary QPO extended with white noise results.	78
6.3	Non-stationary QPO in deterministic flare data.	79
6.4	Non-stationary QPO in deterministic flare results.	80
6.5	Non-stationary QPO and red noise in stationary white noise data.	81
6.6	Non-stationary QPO and red noise in stationary white noise results.	89
6.7	χ^2 -tests demonstration.	90
6.8	Solar flare x-ray lightcurve as observed with GOES in the 1 Å to 8 Å band.	91
6.9	GOES lightcurve periodogram maximum likelihood fit.	91
6.10	Non-stationary QPO in white photon counting noise data.	92
6.11	Non-stationary QPO in white photon counting noise results.	93
7.1	Percentile-percentile plot to show that our sampling methods are not biased.	104

7.2	Histogram of $\ln BF_{\text{QPO}}$ obtained from data sets produced with identical kernel	106
7.3	Maximum likelihood fit of the data (black) from Sec. 6.4.4 for $x = 2$ with a non-stationary GP model.	107
7.4	$\ln BF_{\text{QPO}}$ depending on the extension factor x	107
7.5	2D $t_{\text{start}}/t_{\text{end}}$ posterior	108
7.6	$\ln BF$ of the non-stationary GP model relative to the stationary GP model depending on the extension factor.	108
7.7	GRB090709A maximum likelihood fit using two skewed exponentials.	109
7.8	Model evidences for GRB090709A.	110
7.9	Period posterior for GRB090709A using two skewed exponentials.	110
7.10	Magnetar burst 080823478 maximum likelihood fit using two skewed exponentials.	111
7.11	Model evidences for the 080823478 burst.	111
7.12	Period posterior for the 080823478 burst.	111
7.13	Giant flare maximum likelihood fit using two skewed Gaussians.	113
7.14	Maximum likelihood fit of the solar flare from May 12, 2013, observed by GOES.	114

List of Tables

2.1	Default binary black hole priors.	21
2.2	Default binary neutron star priors.	27
6.1	Results from analyzing the segments of the solar flare detected by GOES.	84
6.2	Parameters and priors used throughout all studies in this paper.	87
6.3	Values used to create the simulated data in Sec. 6.4.	87
7.1	Priors for PP-test	105
7.2	Injection parameters for model selection study.	106

Chapter 1

Introduction

Statistical inference in astrophysics is a fast-evolving field with a unique set of challenges. Broadly speaking, astrophysical observations allows us to probe fundamental physics in a way that would be impossible on Earth. The strong gravity around black holes and the extreme states of matter in neutron stars could not possibly be replicated in a laboratory and thus provide a unique way to understand the universe. However, the obvious trade-off is that these objects are far away and much of the most interesting data we record stems from faint, transient events. These transient events can in general be seen as a not controlled and not reproducible natural experiment, and thus require careful statistical treatment in their interpretation.

At the time of writing only 50 confident gravitational-wave transients from the mergers of compact objects have been reported by the LIGO/Virgo Scientific Collaboration [20, 29], and only one optical counterpart was ever detected for GW170817, a merger of two neutron stars [277]. Still, these observations provide some of the best probes of extreme gravity in many aspects. Thus, we must analyse these events carefully and use optimal techniques to infer if their physical behaviour matches our models. One instance where such careful treatment is especially important is the gravitational-wave memory effect, a effect in general relativity which occurs in its purely nonlinear form in the merger of binaries [79, 279]. Memory manifests as a constant offset in the strain long after the wave has passed. It may be possible to detect memory using the LIGO/Virgo/KAGRA network in the second half of this decade by combining the statistical evidence from thousands of events.

There are other phenomena where observations remain scant and careful analysis is paramount. One aspect of interest in time series analysis of these phenomena is the presence of oscillatory behaviour near a fixed frequency, usually referred to as quasi-periodic oscillations (QPOs). Magnetar hyperflares are extremely energetic outbursts of x-rays from highly magnetised neutron stars. Multiple QPOs have been reported in those flares' tails, which has triggered a flurry of theoretical papers about their origin (e.g. [93, 180–182, 245, 260, 261, 263–265]). These hyperflares have only been confidently linked to magnetars thrice, from SGR 0526-66 in 1979 [52, 197], from SGR 1900+14 [152] in 1998, and from SGR 1806-20 in 2005 [153, 218], though the data from the first observation is generally considered to be of low quality. In gamma-ray bursts (GRBs) the observational evidence for QPOs has long remained weak [74, 92, 160], though there have been recent

reports suggesting that QPOs may be more common [274]. In solar flares, QPOs¹ are observed with great regularity, but identifying their origin remains a challenge, see e.g. Ref. [289] and references therein. Additionally, searching for periodicity in solar-flare lightcurves is challenging due to their non-stationary nature. Understanding QPOs from solar flares may lead to better inferences of stellar flares from other stars [289].

In this thesis, I present my work on developing Bayesian inference techniques for astrophysical applications. This chapter serves to provide context on how the presented publications are connected and some additional background information. I provide a short synopsis of Bayesian inference and its use in astrophysics in Sec. 1.1. I have contributed to the BILBY project, which is presented in Chap. 2 and on which I provide additional information in Chap. 3. One application of Bayesian inference is the measuring of the gravitational-memory effect, which I present in detail in Chaps. 4 and 5. Sec. 1.2 in this chapter provides an overview of how the analysis works and what technical challenges had to be overcome. Finally, we take a look at how Bayesian inference helps us to understand the challenges associated with finding QPOs within time series data in Sec. 1.3, and Chaps. 6 and 7. In Chap. 6 I demonstrate that tests of periodicity based on periodograms are often biased by non-stationarity behaviour which is present in any astrophysical transient. This effect can greatly exaggerate the significance of a QPO. In Chap. 7 I present how one can test time series for periodicity using Gaussian process (GP) likelihoods. Unlike periodogram-based methods, GPs operate solely in the time domain and can trivially overcome most biases that arise due to non-stationarities.

1.1 Bayesian Inference

Bayesian inference can be understood as a statistical paradigm that can be used both for parameter estimation and hypothesis testing (model selection). Unlike classical (frequentist) statistical inference, Bayesian inference relies on the choice of priors to describe prior beliefs about the model. Since Bayesian inference does not rely on repeated sampling from a population to draw inferences, it is a natural choice for astrophysics where we often deal with unique, transient events that provide limited data.

1.1.1 Bayes' Theorem

Bayes' theorem stands at the centre of Bayesian statistics. In its most generic form it describes how conditional probabilities of two events E_1 and E_2 relate to each other

$$p(E_1|E_2) = \frac{p(E_2|E_1)p(E_1)}{p(E_2)}, \quad (1.1)$$

where we denote conditional probability $p(E_1|E_2)$ to mean “probability of E_1 if E_2 already occurred”. The conditional probability is formally defined as

$$p(E_1|E_2) = \frac{p(E_1 \cap E_2)}{p(E_2)} \quad (1.2)$$

¹In solar flares QPOs are commonly referred to as quasi-periodic pulsations (QPP). We will stick with “QPO” even in the solar flare context for consistency.

where $E_1 \cap E_2$ describes E_1 and E_2 both happening. Substituting Eq. 1.2 into Eq. 1.1 and using the commutativity of the \cap -operator thus immediately proves the theorem.

Equation 1.1 can be translated into a form that is useful for inference by identifying $E_1 = \theta$, which is the set of parameters for a model M , and $E_2 = d$, which is the data we wish to analyse. We rewrite Eq. 1.1 and re-label the probability functions for clarity

$$p(\theta|d, M) = \frac{\pi(\theta|M)L(d|\theta, M)}{Z(d|M)}. \quad (1.3)$$

Here $p(\theta|d, M)$ is the **posterior** probability of the parameters given the data, $\pi(\theta|M)$ is the **prior** probability of the parameters, $L(d|\theta, M)$ is the **likelihood** (sometimes also called loss function) which describes the probability of obtaining the data given a set of parameters, and $Z(d|M)$ is the **evidence** or fully marginalised likelihood. In terms of parameter estimation, we are interested in calculating the posterior, for which we need to define adequate priors and find a likelihood function.

The evidence is obtained in practise by re-arranging and integrating 1.3 and using the fact that the posterior is normalised by definition

$$Z(d|M) = \int \pi(\theta|M)L(d|\theta, M)d\theta. \quad (1.4)$$

We see in Eq. 1.4 that we obtain the evidence by integrating the likelihood times the prior over the entire parameter space, which explains the likelihood's alternative name "fully-marginalised" likelihood. The evidence thus provides a probability of the data given a model irrespective of any specific parameters. This implies the evidence's usefulness in model selection questions. We define the Bayes factor

$$BF = \frac{Z(d|M_1)}{Z(d|M_2)} \quad (1.5)$$

which describes the ratio of probabilities of obtaining data given two models $M_{1/2}$. Sometimes there are prior beliefs that may lead us to prefer one model over another, e.g. if we have some independent measurement that already provided a Bayes factor. To capture this, we can introduce prior odds $\Pi(M)$, which describes our prior belief about the likelihood of a model. With this we can obtain the odds ratio

$$\mathcal{O} = \frac{\Pi(M_1)}{\Pi(M_2)}BF, \quad (1.6)$$

which combines the prior odds and the Bayes factor. In many astrophysical scenarios we set the prior odds ratio $\Pi(M_1)/\Pi(M_2) = 1$ and effectively just consider the Bayes factor as the deciding criterion for model selection.

It should be of note that there is no need for an "Occam-factor", i.e. a penalty factor for more complex models in Bayesian inference, that we might be tempted to add in the prior odds. More complex models will typically feature more parameters and thus a larger prior volume V_π , which indicates a lower prior probability for any given set of parameters θ . Equation 1.4 indicates that a larger prior volume thus reduces our evidence and thus naturally penalises more complex models.

1.1.2 Computational Implementations

Performing Bayesian inference for the many-dimensional problems that are common in astrophysics cannot be achieved with trivial algorithms in practice. The most naive approach of calculating posteriors and evidences is to cast a grid of points onto the parameter space. However, this approach has computational complexity $\mathcal{O}(N^d)$ with N being the number of grid points in each dimension d . This approach is feasible up to $d = 2$ but does not scale to higher dimensions. To calculate posterior distributions efficiently, we need to use sophisticated Markov-Chain Monte-Carlo algorithms to sample the posterior space. Most of these algorithms lack the ability to calculate evidences. More recent developments such as nested sampling [255, 256] recast the integral in Eq. 1.4 in terms of the prior mass X such that

$$Z = \int_0^1 L(X) dX. \quad (1.7)$$

This integral can be evaluated by finding nested shells of iso-likelihood during the sampling process and is thus called “nested-sampling” [255, 256]. Conveniently, nested sampling also yields the posterior distribution alongside the evidence and thus in effect provides a unified method for parameter estimation and model selection.

1.2 Gravitational-wave Memory

Gravitational-wave memory is a phenomenon where the gravitational strain h does not revert to zero after a burst of gravitational radiation has passed. Instead, h will settle down to a constant value discretely different from zero. This means that the distance between freely falling test masses will permanently change before and after the burst. Memory is an effect within general relativity and thus in principle present in any source of gravitational radiation. The term memory was likely first mentioned in the literature in Ref. [68], though the effect has been first described much earlier in Ref. [316].

In the following, I will briefly introduce the two kinds of memory that are described in the literature, linear and nonlinear memory, and provide an intuitive understanding as to why they exist.

1.2.1 Linear Memory

Linear gravitational-wave memory has been first mentioned by Ref. [316], where the authors examined the emission of gravitational waves in superdense clusters of stars. In such a cluster, bursts of gravitational waves can be sourced by flybys between stars. The authors first note the memory effect in that they find that the gravitational strain does not revert to zero after the burst. Over the following years the theory of gravitational-wave memory, i.e. the behaviour of h in the zero-frequency limit, was developed further. Eventually Ref. [69] provided a simple formula for calculating the amplitude of linear memory in a detector at distance r for a burst sourced from N freely moving systems indexed by A , with mass M_A , moving with velocities v_A at an angle θ_A relative to the

direction to the detector

$$\Delta h_{jk}^{\text{TT}} = \Delta \sum_{A=1}^N \frac{4M_A}{r\sqrt{1-v_A^2}} \left[\frac{v_A^j v_A^k}{1-v_A \cos \theta_A} \right]^{\text{TT}}. \quad (1.8)$$

Here Δ refers to the difference in summation before and after the burst has passed. The subscripts j and k refer to the x , y , and z coordinates and the TT superscript indicates that we are working in the transverse-traceless gauge, i.e. the wave is propagating in the z direction and $h_{jz} = h_{zj} = 0$. The quantities r , v_A^j , and θ_A are measured as in the frame of the detector. From 1.8 it is clear that linear memory can only be sourced in a group of unbound systems such as the flybys mentioned in Ref. [316], as a group of bound systems could easily be recast as a single system with a fixed centre of mass.

Possible detectable sources of linear memory include supernovae and dynamic three black hole interactions. In supernovae, detectable linear memory could be sourced if there is a highly asymmetric emission of neutrinos or the stellar envelope, though could not be detected in current generation detectors [69, 208]. Dynamic interactions at the centre of galaxies could allow for the ejection of a black hole at high enough velocities for the memory to be detected, though it remains unclear if these occur at sufficient rates to be suitable targets for observation [69]. Thus, Linear memory is now seen as a more unlikely candidate for detectable gravitational waves, though early theory thought it possible to be among the first sources of gravitational waves [69].

1.2.2 Nonlinear Memory

Binary black-hole coalescences are the most common sources of gravitational waves we observe today. Since these start as a fairly closely bound binary and finish as a single, stationary black hole, it is clear that they cannot source memory as prescribed in Eq. 1.8. It was long assumed that any nonlinear effects of gravitational waves could be neglected, though Ref. [79] showed that bound systems nevertheless source a memory contribution that can reach the same order of magnitude as the regular oscillatory wave at the detector. Shortly after, Ref. [311] explicitly computed the memory for compact binary systems in the quadrupole approximation. The argument made in Ref. [79] relies on highly complex mathematical analysis and is less suitable for intuitive understanding of the nonlinear memory effect.

A simple explanation, building upon the theory of the linear memory effect, is given in Ref. [279], which pointed out that the result in Ref. [79] can be obtained by considering radiated gravitons in the form of gravitational waves in Eq. 1.8. This works if we consider the energy, rather than the mass, of the gravitational waves as they are naturally massless. As a further extension to Eq. 1.8, we can also write down the time evolution of the memory as this is a function of the energy E of the gravitational wave (see e.g. Ref. [104])

$$\delta h_{jk}^{\text{TT}}(t_R, \Omega) = \frac{4}{r} \int_{-\infty}^{t_R} dt \int_{S^2} d\Omega' \frac{d^2 E}{dt d\Omega'} \left[\frac{n_j n_k}{1 - \mathbf{n} \cdot \mathbf{N}} \right]^{\text{TT}} \quad (1.9)$$

where t_R is the retarded time at the detector, Ω is the solid angle, $\mathbf{n}(\Omega')$ is a unit vector, $\mathbf{N}(\Omega)$ is the unit line-of-sight vector between the source and the detector, and the energy

flux is given by

$$\frac{d^2 E}{dt d\Omega} = \frac{r^2}{16\pi} \left| \frac{dh(t, \Omega)}{dt} \right|^2. \quad (1.10)$$

Here, we have used units of $G = c = 1$. Thus, given the “oscillatory” gravitational-wave strain $h(t, \Omega)$ we can immediately calculate the associated memory.

Equation 1.9 displays the hereditary nature of the memory effect in that it is dependent on the entire history of h , which leads to the constant offset observed from it. This behaviour emerges since memory essentially describes the zero-frequency behaviour of the gravitational-wave emission. As memory is sourced instantaneously with the emission of oscillatory waves at time t , there is an infinitesimal direct current offset applied to the wave at any instant. This hereditary nature and the term “memory” should not be mistaken to think that knowing the memory at time t_R allows us to know the past evolution of the source, like a memory in the conventional sense, allows us to know the past. Instead, it is rather the opposite: To know the memory at time t_R exactly, we need to know the entire history of the event and knowing $\delta h_{jk}^{\text{TT}}(t_R, \Omega)$ only gives vague information about the event’s history. Memory can be somewhat thought of like hysteresis in a magnetic material, which remains magnetised after the external field has been switched off. However, it is different to hysteresis in the sense that the lasting DC offset can not be measured on its own since spacetime is flat.

There are alternative ways to calculate memory other than Eq. 1.9. For instance, one can also employ lengthy calculations using Post Newtonian calculations (see e.g. Refs. [63, 102]). There is also some recent progress in extracting nonlinear memory from numerical relativity simulations, which has been difficult in the past for technical reasons [104, 205]. These simulations may eventually lead to viable surrogate models that can be used for inference. As Chap. 4 lays out, we use the approach in Eq. 1.8 to search for memory signatures in detected binary black hole mergers for practical reasons. Since real-world detections of gravitational waves have all been of oscillatory waves from binary coalescences [20, 29], we need an oscillatory model to fit the oscillatory part of the signal and we can add the memory part to perform model selection. Additionally, extracting memory from numerical relativity simulations is computationally expensive and there are no viable surrogate models available yet [205]. Calculating memory from Eq. 1.9 on the other hand is computationally cheap and is typically no more expensive than calculating the oscillatory part of the waveform. We show an example waveform in Fig. 1.1. In the remainder of the thesis, we will often refer to nonlinear memory just as “memory” as it is the effect we are investigating.

1.2.2.1 Memory Phenomenology

In this section, we will review some of the major results about the shape and properties of memory sourced from compact binaries. Most of this is based on the work of Marc Favata who thoroughly explored this in Refs. [102–104].

In the simplest case, we consider the memory sourced by the leading quadrupolar order modes of the oscillatory waveform. The standard choice of polarization tensors leads to the entire memory effect being present in the plus mode h_+ for circularised binaries. Ref. [102] lists three important relations for the plus polarised memory h_+^{mem} . First, as we would expect intuitively, memory is inversely proportional to the luminosity distance

d_L

$$h_+^{\text{mem}} \propto 1/d_L. \quad (1.11)$$

This relation should be trivially true due to energy conservation. Next, the memory is sensitively dependent on the inclination angle θ_{jn} between line of sight to the binary and direction of total angular momentum

$$h_+^{\text{mem}} \propto \sin^2 \theta_{\text{JN}} (17 + \cos^2 \theta_{\text{JN}}). \quad (1.12)$$

Concretely, this means that edge-on binaries ($\theta_{\text{JN}} = \pi/2$) give us maximum memory whereas face-on ($\theta_{\text{JN}} = 0, \pi$) yield zero memory. This stands in contrast with oscillatory gravitational radiation which is primarily beamed parallel to the angular momentum. Yet, this should come as no surprise, if we remember the picture in which the memory is the part of the wave which is sourced by the wave itself. The oscillatory wave is beamed primarily perpendicular to the orbital plane, whereas the memory is then beamed perpendicular to the oscillatory wave. Overall, Eq. 1.12 is quite favourable for us as observers since we are looking more edge-on rather than face-on at most binaries. As we will see later, this is additionally favourable for us since higher-order modes, which are important to determine the sign of the memory, are also more easily visible for edge-on events. Finally, memory sourced before the merger time depends on the masses of the system

$$h_+^{\text{mem}} \propto \mu = \frac{m_1 m_2}{m_1 + m_2}, \quad (1.13)$$

where μ is the reduced mass, and m_1 and m_2 are the component masses of the binary. This is only true before the merger time, but this is where most of the detectable signal is sourced, so for purposes of memory detection, it is approximately true for the entire waveform. This means that given a fixed total mass $M = m_1 + m_2$, memory is maximal for equal component masses. Unlike Eq. 1.12, this relation should leave us a bit less hopeful. While many binaries observed to date do have relatively close to equal masses, slightly asymmetric masses are important to measure higher-order modes (perfectly equal mass, non-spinning, binaries would show zero higher-order modes). As mentioned briefly above, measuring higher-order modes is important to determine the sign of the memory. Additionally, this relation indicates that the low mass binaries will source little memory. Thus, recently detected binary neutron star and neutron star-black hole systems are uninteresting for memory searches.

1.2.2.2 Importance of higher-order modes

As hinted at above, it is important to measure higher-order modes to correctly determine the sign of the memory in the detector. This is because the quadrupolar modes $(l, m) = (2, \pm 2)$ of a gravitational wave are degenerate under a simultaneous transformation

$$\phi_c \rightarrow \phi_c \pm \pi/2 \quad (1.14)$$

$$\psi \rightarrow \psi \pm \pi/2, \quad (1.15)$$

where ϕ_c is the phase at coalescence and ψ is the polarisation angle [179]. Meanwhile, the same transformation flips the sign of the memory. We can understand the approach from Ref. [179], which we implement in Chap. 4, as adding up the memory signal from multiple events, and thus increase the memory signal-to-noise ratio with square root of the number of events. However, if we are unable to determine the sign of the memory,

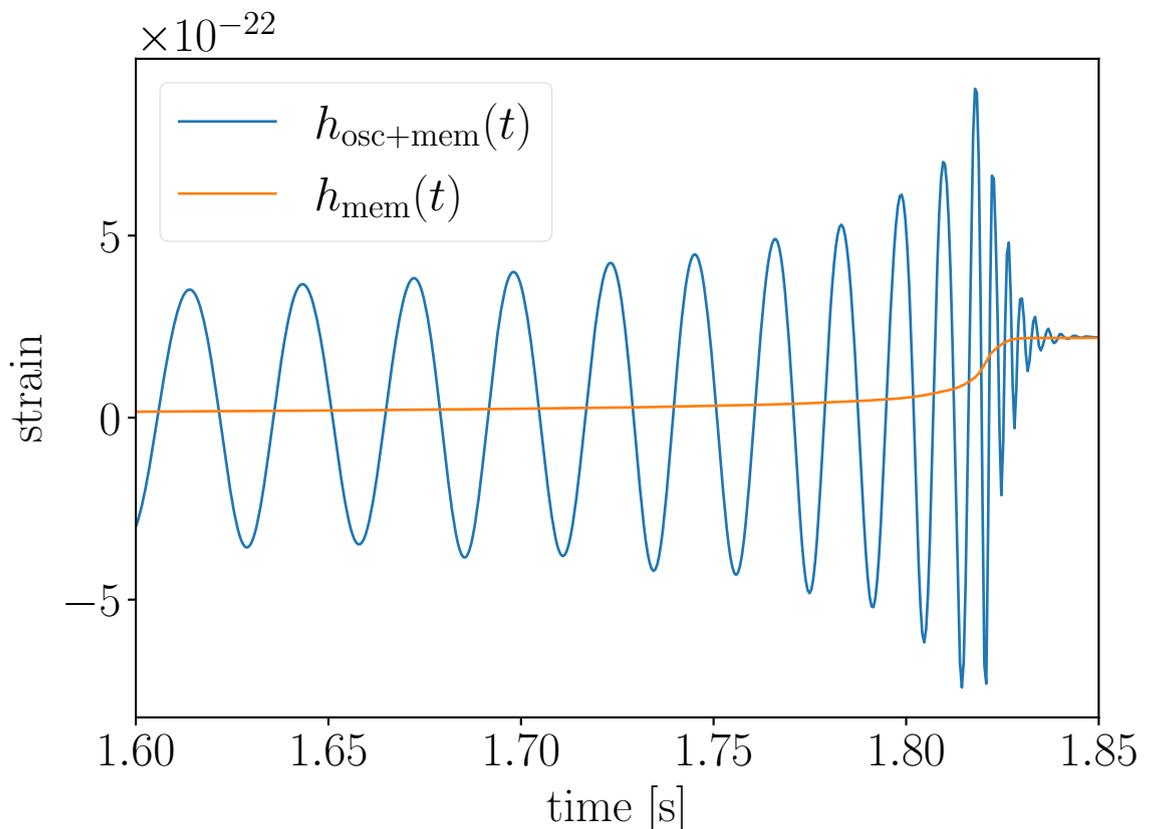


FIGURE 1.1: Plus mode of an IMRPHENOMXHM waveform and memory for a GW150914-like event.

the signal will not add up coherently and the signal-to-noise will increase much slower with the fourth root of the number of events. The reality of this matter turns out to be more complicated as not all gravitational-wave detections are created equal. While including higher-order modes to break the degeneracy in phase and polarisation is helpful to determine the sign of the memory, it will not always be sufficient in practice. None of the events in the first two observing runs shows strong signs of higher-order modes [220]. Even for a strong gravitational-wave signal such as GW150914 [2, 3] the phase and polarisation degeneracy is not broken, likely since it is close to being equal mass (see Fig. 3 in Ref. [220]).

Higher-order modes should be included in the model regardless since they help constrain posterior distributions. As we stated above, memory is sensitive to the inclination angle θ_{JN} , which can be better constrained with higher-order modes [220]. As we show later in the top of Fig. 5.3, even including some of the sub-dominant higher-order modes over just the most dominant ones can shift the posterior distribution of θ_{JN} and affect the inferred Bayes factor for memory. Thus, higher-order modes are also required to reduce the presence of systematic waveform errors by as much as possible.

1.2.3 Calculating and measuring Memory in Practice

There are several practical considerations and approaches when it comes to calculating memory contributions to the waveform. As established in Ref. [103], we could use an analytical “minimal-waveform model”. While this model is much faster than the extraction of memory from the (l, m) modes via Eq. 1.9, it would be hard to match an oscillatory waveform to it in a consistent way. After all, we are measuring memory from binaries where a strong oscillatory signal is already present. To find the memory, we not only have to distinguish memory from noise but also have to distinguish memory from the oscillatory waveform. In the following, we lay out how we can perform model selection that we apply in Chaps. 4 and 5. For more detail about the software packages, we wrote to carry this out in practice see App.A.

We can write the combined oscillatory and memory waveform as

$$h_{\text{osc+mem}} = h_{\text{osc}} + h_{\text{mem}} \quad (1.16)$$

where h_{osc} is the ordinary oscillatory waveform as calculated by waveform approximants, and h_{mem} is the part that we calculate in Eq. 1.9. In order to make a determination about the presence of memory, we need to perform model selection between $h_{\text{osc+mem}}$ and h_{osc} . Bayesian model selection using Bayes factors are well suited for this task as we are comparing discretely different models that have the same parameters [124]. The latter property is especially favourable for Bayesian methods as this means that prior choices will not directly impact the Bayes factor calculation.

Evidences and Bayes factors can be calculated with sophisticated algorithms such as nested sampling. However, as we show in Chaps 4 and 5, the log Bayes factors we obtain for $h_{\text{osc+mem}}$ and h_{osc} are so small that they are swamped by sampling noise using ordinary settings. Instead, we use the fact that $h_{\text{osc+mem}}$ and h_{mem} have the same priors $\pi(\theta)$, and also approximately the same posterior $p(\theta|d)$. Solving Bayes’ theorem for π we get

$$\pi(\theta) = \frac{Z(d)p(\theta|d)}{L(d|\theta)}. \quad (1.17)$$

We perform this step for both models and equate

$$\frac{Z_{\text{osc+mem}}(d)p_{\text{osc+mem}}(\theta|d)}{L_{\text{osc+mem}}(d|\theta)} = \frac{Z_{\text{osc}}(d)p_{\text{osc}}(\theta|d)}{L_{\text{osc}}(d|\theta)}. \quad (1.18)$$

We move $L_{\text{osc+mem}}$ to the right hand side and integrate

$$Z_{\text{osc+mem}}(d) = \frac{Z_{\text{osc}}(d)}{n} \sum_{k=1}^n \frac{L_{\text{osc+mem}}(d|\theta_k)}{L_{\text{osc}}(d|\theta_k)}, \quad (1.19)$$

where n is the number of posterior samples. We use the fact that $p_{\text{osc+mem}}$ is normalised $\int p_{\text{osc+mem}}(\theta|d)d\theta = 1$. The memory Bayes factor BF_{mem} can then be obtained by rearranging

$$BF_{\text{mem}} = \frac{Z_{\text{osc+mem}}(d)}{Z_{\text{osc}}(d)} = \frac{1}{n} \sum_{k=1}^n \frac{L_{\text{osc+mem}}(d|\theta_k)}{L_{\text{osc}}(d|\theta_k)}. \quad (1.20)$$

This method is called ‘‘importance sampling’’ or sometimes ‘‘reweighting’’ in the context of gravitational-wave analyses, and we write the weights as

$$w_k = \frac{L_{\text{osc+mem}}(d|\theta_k)}{L_{\text{osc}}(d|\theta_k)}. \quad (1.21)$$

The great advantage of this method is that we only need to calculate a single posterior using h_{osc} . This benefits us in two major ways. First, we reduce the computational cost by more than half ($h_{\text{osc+mem}}$ is somewhat more expensive to evaluate than h_{osc}). Second, and more crucially, we greatly reduce the effect stochastic sampling noise has on the Bayes factor calculation as we explain in more detail in Chap. 4. Importance sampling works well if the distribution obtained from the oscillatory waveform closely matches the combined waveform. Having a wide range of weights will effectively cause samples with very low weights to be tossed out in the reweighted posterior. We can calculate the number of effective samples remaining

$$n_{\text{eff}} = \frac{(\sum_k w_k)^2}{\sum_k w_k^2}. \quad (1.22)$$

The reweighting framework also sheds additional light on the importance of including higher-order modes into our models. Assume we obtain two posterior samples using L_{osc} , one of which represents the true underlying parameter set θ_{true} , and one that is in the degenerate mode θ_{flip} . In the idealised case of noise being zero $\ln(L_{\text{osc+mem}}(\theta_{\text{true}})) = 0$, since all the residuals vanish. Additionally, $\ln(L_{\text{osc}}(\theta_{\text{true}})) = \ln(L_{\text{osc}}(\theta_{\text{flip}})) = \delta h^2$ where δh^2 represents the weighted least-square difference between the oscillatory model and the combined oscillatory and memory model. In gravitational-wave detectors, δh^2 corresponds to the square of the matched-filter signal-to-noise ratio. Finally, $\ln(L_{\text{osc+mem}}(\theta_{\text{flip}})) = 4\delta h^2$, since model and signal are now $2\delta h$ apart. Putting this together to obtain a Bayes factor, we find

$$BF_{\text{mem}} = \frac{1}{2} \left(\frac{L_{\text{osc+mem}}(\theta_{\text{true}})}{L_{\text{osc}}(\theta_{\text{true}})} + \frac{L_{\text{osc+mem}}(\theta_{\text{flip}})}{L_{\text{osc}}(\theta_{\text{flip}})} \right) \quad (1.23)$$

$$= \frac{1}{2} \left(e^{\delta h^2} + e^{-3\delta h^2} \right) \quad (1.24)$$

$$\approx 1 - \delta h^2, \quad (1.25)$$

where we have used the first order Taylor approximation of the exponential in the last line. This result implies a negative log Bayes factor at least for small δh , though the log Bayes factor turns positive for $\delta h \gtrsim 0.781$. In practice, such small signals will be dominated by noise, so the actual Bayes factors we obtain will not follow this relation exactly. We display the shape of Eq. 1.23 in Fig. 1.2.

1.2.4 Prospects of measuring Memory

1.2.4.1 Ground-based Detectors

As we show in Chap. 4, we require 1830_{-1100}^{+1730} (90% CI) gravitational-wave detections of binary black hole mergers with LIGO/Virgo at design sensitivity to confidently measure memory ($\ln BF_{\text{mem}} \geq 8$). This milestone will likely be reached in the second half of the decade, at the beginning of the LIGO A+/Virgo+ era, when we will observe tens of

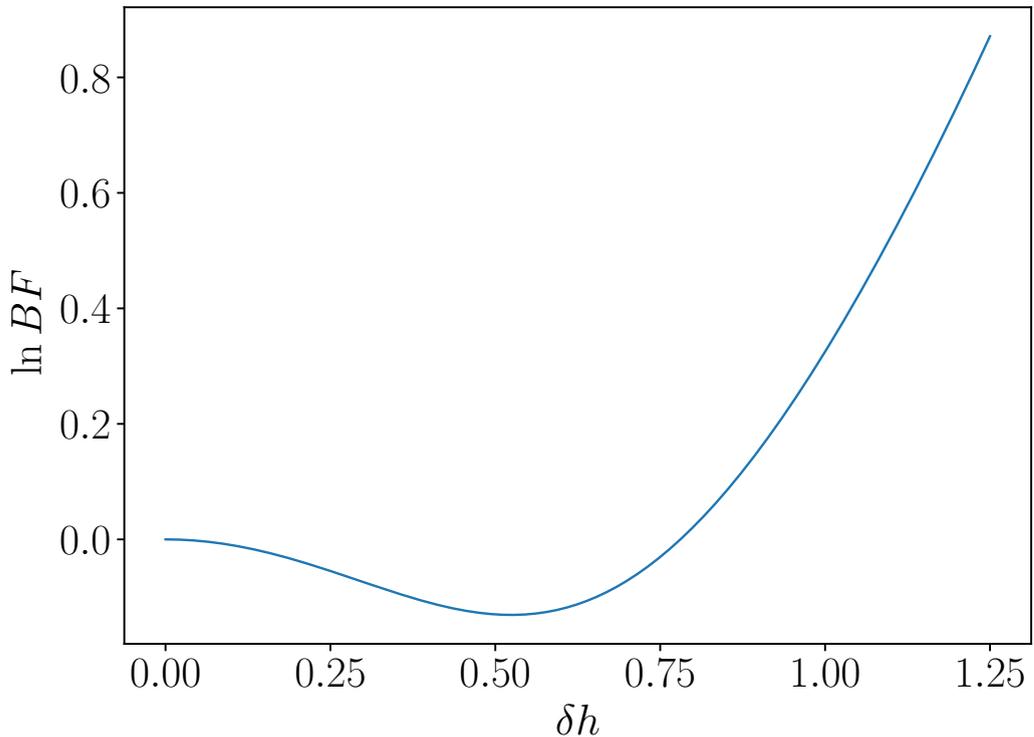


FIGURE 1.2: $\ln BF$ given different memory matched-filter signal-to-noise ratios δh assuming a toy model with two posterior samples, one of which is in the degenerate $\phi_c - \psi$ mode. The $\ln BF$ is negative for small δh , highlighting the need to include higher-order modes to break the degeneracy. Note that the location of $\ln BF = 0$ is only valid for this specific toy model and does not hold in the general case.

events every day. The large uncertainties on the number of required detections do not take into account “unknown unknowns”, such as what specific low-frequency sensitivity of future detector upgrades will yield, the detector duty cycles, and so forth. Reference [64] shows that, in broad agreement with our study, about five years of data collection with LIGO/Virgo at design sensitivity is required to reach a total memory SNR $\rho_{\text{tot}} = 3$, which can be seen as a tentative measurement. While Ref. [64] investigates a wider range of models of waveforms and memory, and uses somewhat different population assumptions, they find that 1488^{+725}_{-879} (90% CI) events are required to reach $\rho_{\text{tot}} = 3$ using the assumptions we made in Chap. 4. It should also be noted that these estimates sensitively depend on inferred merger rates, which will be better understood in the coming years.

Third generation ground-based detectors will most likely be able to measure memory in binary mergers, assuming they reach their sensitivity targets [162]. Einstein Telescope [139, 230] may be able to detect memory with $\rho_{\text{mem}} \gtrsim 8$ from a GW150914-like merger, but is most sensitive to memory from mergers with total mass $M_{\text{tot}} = 1000M_{\odot} - 10000M_{\odot}$ if they exist, where M_{\odot} is the solar mass. Memory from a $M_{\text{tot}} = 5000M_{\odot}$ merger might be detected with $\rho_{\text{mem}} \gtrsim 3$ out to more than 10 Gpc. Cosmic Explorer [9, 237] will boast even greater sensitivity to somewhat smaller masses, with $M_{\text{tot}} = 2000M_{\odot}$ being detectable out to ~ 20 Gpc. Cosmic explorer will also be highly sensitive to memory from events such as in GWTC-1 [162].

1.2.4.2 Space-based Instruments

The Laser Interferometer Space Antenna (LISA) is a proposed space-based gravitational-wave detector, that is scheduled to launch in the early to mid 2030s [37]. Based on the projected sensitivity LISA should be especially suitable for measuring the low-frequency strain sourced by memory. LISA will likely detect between 1 and 10 memory events with $\rho_{\text{mem}} \geq 5$ within its planned 4-year mission [158]. LISA will be most sensitive to events between $M_{\text{tot}} = 10^5 M_{\odot} - 10^7 M_{\odot}$ and thus provide a nice complement to the third generation ground-based detectors.

The Deci-hertz Interferometer Gravitational-wave Observatory (DECIGO) is a proposed space-based detector that is supposed to close the frequency gap between LISA and ground-based detectors [165]. DECIGO will be most sensitive between 0.1 Hz and 10 Hz, which will make memory from GW150914-like mergers easily measurable out to cosmological distances. The same is true for any merger with total mass up to $10^6 M_{\odot}$. Similarly, the Big Bang Observer (BBO), a proposed successor to LISA which may be launched in many decades, will be able to reach similar or higher sensitivity than DECIGO across DECIGO's sensitive band [89, 162].

1.2.4.3 Pulsar Timing

Pulsar timing is an avenue of detecting even lower frequency gravitational waves than space-based detectors and is based on radio astronomy. Pulsar timing arrays work by measuring the time of arrival of radio pulses from galactic millisecond pulsars. Memory would manifest itself as a sudden change in the pulsar's rotational frequency. Pulsar timing remains a complex task as intrinsic noise properties of individual pulsars can mimic a burst. Analyses thus rely on the combination of data from many pulsars (see Ref. [140] for the methods used in pulsar timing).

There are several currently operating and proposed PTAs which can be used for memory measurements from supermassive black hole binaries. The North American Observatory for Gravitational Waves (NANOGrav) [36, 87, 232], the Parkes Pulsar Timing Array (PPTA) [193], and the European Pulsar Timing Array (EPTA) [170] are currently operating, and the International Pulsar Timing Array [221, 299] aims to combine measurements from these groups to increase sensitivity. The Square Kilometer Array (SKA) is a future detector that will likely detect many more pulsars and also improve timing residuals on known pulsars, thus significantly increasing sensitivity. Searches for memory with PTAs have been carried out, though without any success yet (see Refs. [35, 41, 304] for the most comprehensive studies). These searches are most sensitive to memory from binaries between $M_{\text{tot}} = 10^{11} M_{\odot} - 10^{12} M_{\odot}$, which might not exist since supermassive black holes likely cannot grow larger than $\sim 10^{10} M_{\odot}$ [162, 169]. On the other hand, the SKA will likely be able to probe memory from supermassive systems down to $10^9 M_{\odot}$ [162].

1.3 Quasi-periodic Oscillations

QPOs can be thought of as oscillatory behaviour in a time series with changing amplitude or frequency that manifests as a narrow peak in the power spectrum. QPOs are thus

often implicitly understood to arise from stochastic processes. It is thus natural that tests for periodicity have mostly been developed in the frequency domain, where a QPO would stand out as a visual feature. One common method to analyse QPOs is to calculate a periodogram and assess significance by searching for frequency bins with excess power relative to the underlying spectrum. Tests for periodicity vary and include frequentist tests of excess power in individual frequency bins and fitting routines where the QPO is modelled as some narrow feature in the power spectrum on top of the noise continuum.

1.3.1 Underlying physical processes

QPOs exist for many different astrophysical sources and thus do not have a single underlying process. We specifically study QPOs in transient signals, such as magnetar flares, gamma-ray bursts (GRBs), and solar flares. However, QPOs have also been observed and studied in other objects such as X-ray binaries and active galactic nuclei (see Refs. [156, 246] for reviews).

QPOs in magnetar flares have first been established in the SGR 1806_20 Giant flare observed in 2004 [159, 218, 268, 307], which has caused a flurry of theories about their origin. Initial analysis suggested that QPOs are related to torsional oscillation modes [159]. Shortly after, Ref. [181] posed that coupling of crust to the core by strong magnetic fields should cause the oscillations to damp within a few tenths of a second. More realistic magnetic coupling scenarios were investigated shortly after [82, 117, 118, 182, 183, 300]. Additionally, torsional mode identification was explored as a probe to constrain properties of the neutron star such as mass and radius, and its internal composition [83, 204, 245, 261, 262, 264, 306]. However, the exact process by which QPOs arise remains elusive.

There has been speculation about QPOs in long GRBs [195, 318] and a recent claim of several detections [274]. However, the observational literature is found most claims of QPOs to be contentious. Specifically, GRB090709A, which was originally reported to have 8.06 s period, later proved not to be statistically significant [74, 92, 160]. There are several possible sources of QPOs in GRBs. Reference [195] proposes that magnetorotational instabilities in a hyperaccreting disk can cause episodic, intense mass accretion onto a central black hole due to gravitational torque. Alternatively, Ref. [318] conjectures QPOs to arise from a precessing magnetic field.

QPOs in solar flares are regularly observed across the electromagnetic spectrum. The two prominent families of mechanisms proposed to explain these phenomena are magneto-hydrodynamic oscillations present in coronal structures and quasi-periodic regimes of magnetic reconnection [289, 319]. However, the exact origin of these oscillations remains elusive since the models are generally qualitative in nature [319]. There have been efforts in recent years to standardise the analysis of solar flare QPOs, which may eventually shed light on their origin [154, 155].

1.3.2 Analysis methods

In Chap. 6 we show how some of the established and widely used methods are biased if non-stationarities are present in the time series. This is because these methods rely on the Whittle-likelihood function that is only valid for stationary time series (see e.g.

Refs. [75, 309]). However, real astrophysical transients arise due to a combination of deterministic and non-stationary stochastic processes. We demonstrate and quantify this effect using Bayesian techniques using both theoretical and empirical calculations. Additionally, we show how this has affected previous studies of solar flare QPOs.

The results in Chap. 6 motivate the work on Gaussian process (GP) methods presented in Chap. 7. The GP likelihood is a generalisation of the Gaussian likelihood often employed for time series data. This likelihood allows us to model stationary and non-stationary stochastic processes in the time domain via the definition of a covariance matrix, described by a kernel function. Unlike the Whittle likelihood, it also allows us to include known errors for individual data points, which allows for the analysis of heteroscedastic data sets which are common in astrophysics. Additionally, GPs can fit a deterministic mean function simultaneously with the stochastic process.

The main hindrance in the adoption of GPs in time series analysis has likely been the high computational cost of evaluating the likelihood, which is $\mathcal{O}(N^3)$ in time and $\mathcal{O}(N^2)$ in memory for a time series with N elements in the general case. However, recent innovations and software implementations such as CELERITE [113] have made it possible to reduce both the time and memory complexity to $\mathcal{O}(N)$ for the class of stationary, complex exponential kernel functions or combinations thereof. This makes the analysis of light curves using Bayesian inference with GPs feasible. The class of allowed kernel functions in CELERITE is particularly suitable in the analysis of QPOs. Using some minor modifications, CELERITE can also fit some simple non-stationary extensions to stationary models.

In Chap. 7 we develop physically motivated kernel functions for periodic and aperiodic processes and implement these within the CELERITE framework. We then use a set of studies using simulated data to demonstrate that we can accurately recover posterior distributions of Gaussian process parameters using BILBY. Finally, we deploy this method on astrophysical and solar flare data and highlight its advantages over frequency-domain based methods.

Chapter 2

Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy

Published as:

G. Ashton, M. Hübner, P. Lasky, C. Talbot, et al. 2019, *Bilby: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy*, *ApJS*, 241, 27, [arxiv:1811.02042](https://arxiv.org/abs/1811.02042)

Bayesian parameter estimation is fast becoming the language of gravitational-wave astronomy. It is the method by which gravitational-wave data is used to infer the sources' astrophysical properties. We introduce a user-friendly Bayesian inference library for gravitational-wave astronomy, BILBY. This python code provides expert-level parameter estimation infrastructure with straightforward syntax and tools that facilitate use by beginners. It allows users to perform accurate and reliable gravitational-wave parameter estimation on both real, freely-available data from LIGO/Virgo, and simulated data. We provide a suite of examples for the analysis of compact binary mergers and other types of signal model including supernovae and the remnants of binary neutron star mergers. These examples illustrate how to change the signal model, how to implement new likelihood functions, and how to add new detectors. BILBY has additional functionality to do population studies using hierarchical Bayesian modelling. We provide an example in which we infer the shape of the black hole mass distribution from an ensemble of observations of binary black hole mergers.

2.1 Introduction

Bayesian inference underpins gravitational-wave science. Following a detection, Bayesian parameter estimation allows one to estimate the properties of a gravitational-wave source, for example, the masses and spins of the components in a binary merger [e.g., 3–5, 16, 21]. If the detection involves neutron stars, Bayesian parameter estimation is used to study the properties of matter at nuclear densities via the signature of tidal physics imprinted on the gravitational waveform [16, 21, 277]. The posterior probability distributions of source parameters such as inclination angle can be used, in turn, to make inferences about

electromagnetic phenomena such as gamma-ray bursts [e.g., 10]. Such parameter estimation is also used to measure cosmological parameters such as the Hubble constant [11]. By combining data from multiple detections, Bayesian inference is used to understand the population properties of gravitational-wave sources [e.g., 5, 99, 243, 257, 272, 275, 313, and references therein], which is providing insights into stellar astrophysics. By extending the gravitational-wave signal model, Bayesian inference is used to test general relativity and look for evidence of new physics [6, 10, 12, 17, 18]

The field of gravitational-wave astronomy is growing rapidly. We have entered the “open data era,” in which gravitational-wave data has become publicly available [285]. Since Bayesian parameter estimation is central to gravitational-wave science, there is a need for a robust, user-friendly code that can be used by both gravitational-wave novices and experts alike.

The primary tool currently used by the LIGO and Virgo collaborations for parameter estimation of gravitational-wave signals is LALINFERENCE [298]. This pioneering code enabled the major gravitational-wave discoveries achieved during the first two LIGO observing runs [e.g., 3–5, 16, 21]. The code itself is now almost a decade old, and years of development have made it hard for beginners to learn, and difficult for experts to modify and adapt to new challenges. More recently, PYCBC INFERENCE [57] was released; a modern, PYTHON-based toolkit designed for compact binary coalescence parameter estimation. This package provides access to several different samplers and builds on the PYCBC package [213] – an open-source toolkit for gravitational-wave astronomy.

We introduce BILBY, a user-friendly parameter-estimation code for gravitational-wave astronomy. BILBY provides expert-level parameter estimation infrastructure with straightforward syntax and tools that facilitate use by beginners. For example, with minimal user effort, users can download and analyze publicly-available LIGO and Virgo data to obtain posterior distributions for the astrophysical parameters associated with recent detections of binary black holes [2, 5, 7, 12–14] and the binary neutron star merger [277].

One key functional difference between BILBY and LALINFERENCE/PYCBC INFERENCE is its modularity and adaptability. The core library is not specific to gravitational-wave science and has uses outside of the gravitational-wave community. Ongoing projects include astrophysical inference in multimessenger astronomy, pulsar timing, and x-ray observations of accreting neutron stars. The gravitational-wave specific library is also built in a modular way, enabling users to easily define their own waveform models, likelihood functions, etc. This implies BILBY can be used for more than studying compact binary coalescences—see Sec. 2.5. The modularity further ensures the code will be sufficiently extensible to suit the future needs of the gravitational-wave community. Moreover, we believe the wider astrophysics inference community will find the code useful by virtue of having a common interface and ideas that can be easily adapted to a range of inference problems.

The remainder of this paper is structured to highlight the versatile, yet user-friendly nature of the code. To that end, the paper is example driven. We assume familiarity with the mathematical formalism of Bayesian inference and parameter estimation (priors, likelihoods, evidence, etc.) as well as familiarity with gravitational-wave data analysis (antenna-response functions, power spectral densities, etc.). Readers looking for an introduction to Bayesian inference in general are referred to Ref. [255], while gravitational-wave specific introductions to inference can be found in Refs. [280, 298]. Section 2.2 describes the BILBY design philosophy, and Sec. 2.3 provides an overview

of the code including installation instructions in Sec. 2.3.1. Subsequent sections show worked examples. The initial examples are the sort of simple calculations that we expect will be of interest to most casual readers. Subsequent sections deal with increasingly complex applications that are more likely of interest to specialists.

The worked examples are as follows. Section 2.4 is devoted to compact binary coalescences. In 2.4.1, we carry out parameter estimation with publicly-available data to analyze GW150914, the first ever gravitational-wave event. In 2.4.2, we study a simulated binary black hole signal added to Monte Carlo noise. In 2.4.3, we study the matter effects encoded in the gravitational waveforms of a binary neutron star inspiral. In 2.4.4, we show how it is possible to add more sophisticated gravitational waveform phenomenology, for example, by including memory, eccentricity, and higher order modes. In 2.4.5, we study an extended gravitational-wave network with a hypothetical new detector.

Section 2.5 is devoted to signal models for sources that are not compact binary coalescences. In 2.5.1, we perform model selection for gravitational waves from a core collapse supernova. In 2.5.2, we study the case of a post-merger remnant. Section 2.6 is devoted to *hyper*-parameterization, a technique used to study the population properties of an ensemble of events. Closing remarks are provided in Section 2.8.

2.2 BILBY Design Philosophy

Three goals guide the design choices of BILBY. First, we seek to provide a parameter-estimation code that is sufficiently powerful to serve as a workhorse for expert users. Second, we aim to make the code accessible for novices, lowering the bar to work on gravitational-wave inference. Third, we desire to produce a code that will age gracefully; advances in gravitational-wave astronomy and Bayesian inference can be incorporated straightforwardly without resort to inelegant workarounds or massive rewrites. To this end, we adhere to a design philosophy, which we articulate with four principles.

- **Modularity.** Wherever possible, we seek to modularize the code and follow the abstraction principle [222], reducing the amount of repeated code and easing development. For example, the sampler is a modularized object, so if a problem is initially analyzed using the PYMULTINEST [72] sampler for example, one can easily switch to the EMCEE [114] sampler or even a custom-built gravitational-wave sampler. For example, BILBY accesses samplers through a common interface; as a result it is trivial to easily switch between samplers to compare performance or check convergence issues.
- **Consistency.** We enforce strict style guidelines, including adherence to the PEP8 style guide for PYTHON ¹. As a result, the code is relatively easy-to-follow and intuitive. In order to maintain integrity of the code while responding to the needs of a large and active user base, we employ GITLAB’s merge request feature. Updates require approval by two experts. The PEP8 protocol is enforced using continuous integration.
- **Generality.** Wherever possible, we keep the code as general as possible. For example, the gravitational-wave package is separate from the package that passes

¹<https://www.python.org/dev/peps/pep-0008/>

the likelihood and prior to the sampler. This generality provides flexibility. For example, in Section 2.6, we show how BILBY can be used to carry out population inference, even though the likelihood function is completely different to the one used for gravitational-wave parameter estimation. Moreover, a general design facilitates the transfer of ideas into and out of gravitational-wave astronomy from the greater astro-statistics community.

- **Usability.** We observe that historically, people find it difficult to get started with gravitational-wave inference. In order to lower the bar, we endeavour to make basic things doable with very few lines of code. We provide a large number of tutorials that can serve as a blueprint for a large variety of real-world problems. Finally, we endeavour to follow the advice of the PEP20 style guide for PYTHON²: “There should be one—and preferably only one—obvious way to do it.” In other words, once users are familiar with the basic layout of BILBY, they can intuit where to look if they want to, for example, add a new detector (see Section 2.4.5) or include non-standard polarization modes.

2.3 Code Overview

2.3.1 Installation

BILBY is open-source, MIT licensed, and written in python. The simplest installation method is through PyPI³. The following command installs from the command line:

```
$ pip install bilby
```

This command downloads and installs the package and dependencies. The source-code itself can be obtained from the `git` repository [46], which also houses an issue tracker and merge-request tool for those wishing to contribute to code development. Documentation about code installation, functionality, and user syntax is also provided [47]. Scripts to run all examples presented in this work are provided in the `git` repository.

2.3.2 Packages

BILBY has been designed such that logical blocks of code are separated and, wherever possible, code is abstracted away to allow future re-use by other models. At the top level, BILBY has three packages: `core`, `gw`, and `hyper`. The `core` package contains the key functionalities. It passes the user-defined priors and likelihood function to a sampler, harvests the posterior samples and evidence calculated by the sampler, and returns a `result` object providing a common interface to the output of any sampler along with information about the inputs. The `gw` package contains gravitational-wave specific functionality, including waveform models, gravitational-wave specific priors and likelihoods. The `hyper` package contains functionality for the hierarchical Bayesian inference (see Sec. 2.6). A flowchart showing the dependency of different packages and modules is available on the `git` repository [46].

²<https://www.python.org/dev/peps/pep-0020/>

³<https://pypi.org/project/BILBY/>

2.3.2.1 The core package

The `core` package provides all of the code required for general problems of inference. It provides a unified interface to several different samplers listed below, standard sets of priors including arbitrary user-defined options, and a universal result object that stores all important information from a given simulation.

Prior and likelihood functions are implemented as classes, with a number of standard types implemented in the `core` package: e.g., the `Normal`, `Uniform`, and `LogUniform` priors, and `GaussianLikelihood`, `PoissonLikelihood`, and `ExponentialLikelihood` likelihoods. One can write their own custom prior and likelihood functions by writing a new class that inherits from the parent `Prior` or `Likelihood`, respectively. The user only needs to define how the new prior or likelihood is instantiated and calculated, with all other house-keeping logic being abstracted away from the user.

The prior and likelihood are passed to the function `run_sampler`, which allows the user to quickly change the sampler method between any of the pre-wrapped samplers, and to define specific run-time requirements such as the number of live points, number of walkers, etc. Pre-packaged samplers include Markov Chain Monte Carlo Ensemble samplers `emcee` [114], `ptemcee` [303], `PyMC3` [244], and Nested samplers [255, 256] `MultiNest` [106–108] (through the PYTHON implementation `pyMultiNest` [72]), `Nestle` [53], `Dynesty` [266], and `CPNest` [296]. The `Sampler` class again allows users to specify their own sampler by following the other examples.

Despite the choice of sampler, the output from `BILBY` is universal: an `hdf5` file [276] that contains all output including posterior samples, likelihood calculations, injected parameters, evidence calculations, etc. The `Result` object can be used to load in these output files, and also perform common operations such as generating corner plots, and creating plots of the data and maximum posterior fit.

2.3.2.2 The gw package

The `gw` package provides the core functionality for parameter estimation specific to transient gravitational waves. Building on the `core` package, this provides prior specifications unique to such problems, e.g., a prior that is uniform in co-moving volume distance, as well as the standard likelihood used when studying gravitational-wave transients ([e.g., see 298] and Eq. 2.1), defined as the `GravitationalWaveTransient` class. The `gw` package also provides an implementation of current gravitational-wave detectors in the `detector` module, including their location and orientation, as well as different noise power spectral densities for both current and future instruments. Standard waveform approximants are also included in the `source` module, which are handled through the `LALSIMULATION` package [184].

The `gw` package also contains a set of tools to load, clean and analyse gravitational-wave data. Many of these functions are built on the `GWpy` [191] code base, which are contained within `bilby.gw.detector` and primarily accessed by instantiating a list of `Interferometer` objects. This functionality also allows one to implement their own gravitational-wave detector by instantiating a new `Interferometer` object—we show an explicit example of this in Sec. 2.4.5.

2.3.3 The hyper package

The `hyper` package contains all required functionality to perform hierarchical Bayesian inference of populations. This includes both a `Model` module and a `HyperparameterLikelihood` class. This entire package is discussed in more detail in Sec. 2.6.

2.4 Compact Binary Coalescence

In this section, we show a suite of BILBY examples analyzing binary black hole and binary neutron star signals.

We employ a standard Gaussian noise likelihood \mathcal{L} for strain data d given source parameters θ [287, 288, 297]:

$$\ln \mathcal{L}(d|\theta) = -\frac{1}{2} \sum_k \left\{ \frac{[d_k - \mu_k(\theta)]^2}{\sigma_k^2} + \ln(2\pi\sigma_k^2) \right\}, \quad (2.1)$$

where k is the frequency bin index, σ is the noise amplitude spectral density, and $\mu(\theta)$ is the waveform. The waveform is a function of the source parameters θ , which consist of (at least) eight intrinsic parameters (primary mass m_1 , secondary mass m_2 , primary spin vector \vec{S}_1 , secondary spin vector \vec{S}_2) and seven extrinsic parameters (luminosity distance d_L , inclination angle ι , polarization angle ψ , time of coalescence t_c , phase of coalescence ϕ_c , right ascension and declination ra and dec, respectively). Table 2.1 shows the default priors implemented for binary black hole systems. We show how these priors can be called in Secs. 2.4.1 and 2.4.2. Unless otherwise specified, $\mu(\theta)$ is given using the `IMRPhenomP` approximant [251]. However, the approximant can be easily changed; see Secs. 2.4.2 and 2.4.3. Moreover, it is relatively simple to sample in different parameters than those listed above (e.g., chirp mass and mass ratio instead of m_1 and m_2); examples for doing this are provided in the `git` repository [46].

2.4.1 GW150914: the onset of gravitational wave astronomy

The first direct detection of gravitational waves occurred on the 14th of September, 2015, when the two LIGO detectors [1] in Hanford, Washington and Livingston, Louisiana detected the coalescence of a binary black hole system [2]. The gravitational waves swept through the two detectors with a $6.9_{-0.4}^{+0.5}$ ms time difference which, when combined with polarization information, allowed for a sky-location reconstruction covering an annulus of 590 deg^2 [2]. The initially-published masses of the colliding black holes were given as $36_{-4}^{+5} M_\odot$ and $29_{-4}^{+4} M_\odot$ [3]. Subsequent analyses with more accurate precessing waveforms constrained the masses to be $35_{-3}^{+5} M_\odot$ and $30_{-4}^{+3} M_\odot$ at 90% confidence [4]. The distance to the source is determined to be 440_{-180}^{+160} Mpc [4].

In this example, we use BILBY to reproduce the parameter estimation results for GW150914. The data for published LIGO/Virgo events is made available through the Gravitational Wave Open Science Center [285]. Built-in BILBY functionality downloads and parses this data. We begin with the following two lines.

TABLE 2.1: Default binary black hole priors. The intrinsic variables are the two black hole masses $m_{1,2}$, their dimensionless spin magnitudes $a_{1,2}$, the tilt angle between their spins and the orbital angular momentum $\theta_{1,2}$, and the two spin vectors describing the azimuthal angle separating the spin vectors $\delta\phi$ and the cone of precession about the system’s angular momentum ϕ_{JL} . The extrinsic parameters are the luminosity distance d_L , the right ascension ra and declination dec, the inclination angle between the observers line of sight and the orbital angular momentum ι , the polarisation angle ψ , and the phase at coalescence ϕ_c . The phase, spins, and inclination angles are all defined at some reference frequency. We do not set a default prior for the coalescence time t_c . ‘sin’ and ‘cos’ priors are uniform in cosine and sine, respectively, and ‘comoving’ implies uniform in comoving volume.

variable	unit	prior	minimum	maximum
$m_{1,2}$	M_{\odot}	uniform	5	100
$a_{1,2}$	-	uniform	0	0.8
$\theta_{1,2}$	rad.	sin	0	π
$\delta\phi, \phi_{\text{JL}}$	rad.	uniform	0	2π
d_L	Mpc	comoving	10^2	5×10^3
ra	rad.	uniform	0	2π
dec	rad.	cos	$-\pi/2$	$\pi/2$
ι	rad.	sin	0	π
ψ	rad.	uniform	0	π
ϕ_c	rad.	uniform	0	2π

```
>>> import bilby
>>> interferometers = bilby.gw.detector.get_event_data("GW150914")
```

The first line of code imports the BILBY code-base into the PYTHON environment. The second line returns a set of objects that contain the relevant data segments and associated data products relevant for the analysis for both the LIGO Hanford and Livingston detectors. By default, BILBY downloads and windows the data. A local copy of the data is saved along with diagnostic plots of the gravitational-wave strain amplitude spectral density.

In addition to the data, the two key ingredients for any Bayesian inference calculation are the likelihood and the prior. Default sets of priors can be called from the `gw.prior` module, and we also employ the default Gaussian noise likelihood (Eq. 2.1).

```
prior = bilby.gw.prior.BBHPriorDict(filename="GW150914.prior")
likelihood = bilby.gw.likelihood.get_binary_black_hole_likelihood(
interferometers)
```

The above code calls the GW150914 prior, which differs from the priors described in Tab. 2.1 in two main ways. Firstly, to speed up the running of the code it restricts the mass priors to between 30 and 50 M_{\odot} for the primary mass, and 20 and 40 M_{\odot} for the secondary mass. Moreover, this prior call restricts the time of coalescence to 0.1 seconds before and after the known coalescence time. One can revert to the priors in Tab. 2.1 by replacing the above file call with `filename="binary_black_holes.prior"`, but this would require separately setting a prior for the coalescence time. We show how this can be done in Sec. 2.4.2.

The next step is to call the sampler:

```
>>> result = bilby.core.sampler.run_sampler(likelihood, prior)
```

This line performs parameter inference using the sampler default DYNesty [266], with a default 500 live points. This number can be increased by passing the `nlive=` keyword argument to `run_sampler()`. The sampler returns a list of posterior samples, the Bayesian evidence, and metadata, which is stored in an `hdf5` file. One may plot a corner plot showing the posterior distribution for all parameters in the model using the command

```
>>> result.plot_corner()
```

The above example code produces posterior distributions that by eye, agree reasonably well with the parameter uncertainty associated with the published distributions for GW150914. The shape of the likelihood for the extrinsic parameters presents significant challenges for samplers, due to strong degeneracy's between different sky locations, distances, inclination angles, and polarization angles [see e.g., 100, 234]. For more accurate results, we use the nested sampling package CPNEST [296], which is invoked by changing the `run_sampler` function above to include the additional argument `sampler='cpnest'`. We also change the number of live points by adding `nlive=5000` to the same function, and specify a keyword argument `maxmcmc=5000`, which is the maximum number of steps the sampler takes before accepting a new sample. To resolve the issue with the phase at coalescence, we analytically marginalize over this parameter [100] by adding the optional `phase_marginalization=True` argument to the instantiation of the likelihood. BILBY has built in analytic marginalization procedures for the time of coalescence [100] and distance [253, 254], which can both be invoked using `time_marginalization=True` and `distance_marginalization=True`, respectively. These decrease the run time of the code by minimizing the dimensionality of the parameter space. Posterior distributions can still be determined for these parameters by reconstructing them analytically from the full set of posterior samples [e.g., see 280].

Using BILBY we can plot marginalized distributions by simply passing the `plot_corner` function the optional `parameters=...` argument. In Fig. 2.1 we show the marginalized, two-dimensional posterior distribution for the masses of the two black holes as calculated using the above BILBY code (shown in blue). In orange we show the LIGO posterior distributions from Ref. [5], calculated using the LALINFERENCE software [298], and hosted on the Gravitational Wave Open Science Center [285].

In Fig. 2.2 we show the marginalized posterior distribution of the luminosity distance and inclination angle, where the BILBY posteriors are again shown in blue, and the LALINFERENCE posteriors in orange. Figure 2.3 shows the sky localisation uncertainty for both BILBY and LALINFERENCE.

The above example does not make use of detector calibration uncertainty, which is an important feature in LIGO data analysis. Such calibration uncertainty is built in to BILBY using the cubic spline parameterization [101], with example usage in the BILBY repository.

2.4.2 Binary black hole merger injection

BILBY supports both the analysis of real data as in the previous section, as well as the ability to inject simulated signals into Monte Carlo data. In the following two sections we

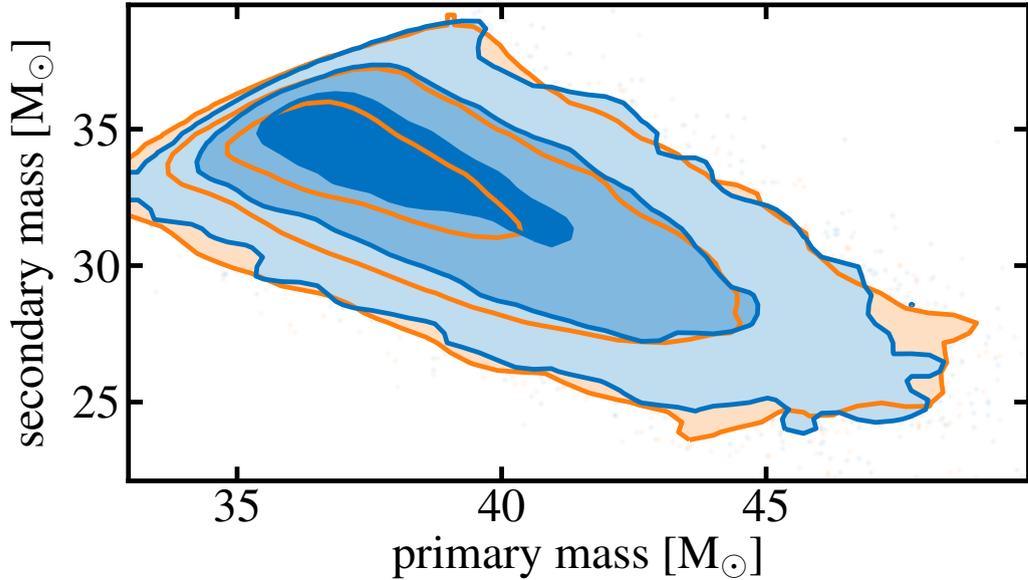


FIGURE 2.1: Marginalised posterior source-mass distributions for the first binary black hole merger detected by LIGO, GW150914. We show the posterior distributions recovered using BILBY (blue), and those using LALINFERENCE (orange), using open data from the Gravitational Wave Open Science Centre [285]. The five lines of BILBY code required for reproducing the posteriors are shown in Section 2.4.1.

inject a binary black hole signal and a binary neutron star signal, respectively, showing how one can easily inject and recover signals and their astrophysical properties.

In this first example⁴, we create a binary black hole signal with parameters similar to GW150914 [3], albeit at a luminosity distance of $d_L = 2$ Gpc (cf. $d_L \approx 400$ Mpc for GW150914). We inject the signal into a network of LIGO-Livingston, LIGO-Hanford [1] and Virgo interferometers [32], each operating at design sensitivity. When doing examples of this nature, it is time intensive to sample over all fifteen parameters in the waveform model. Therefore, to get quick results that can be run on a laptop, we only sample over four parameters in the waveform model: the two black-hole masses $m_{1,2}$, the luminosity distance d_L , and the inclination angle ι . BILBY supports simple functionality to limit or extend the number of parameters included in the likelihood calculation, as shown below.

We begin by setting up a `WaveformGenerator` object using a frequency domain strain model that takes the signal injection parameters and specific waveform arguments such as the waveform approximant as arguments. The `WaveformGenerator` also takes data duration and sampling frequency as input parameters. With the source model defined, we now instantiate an `interferometer` object that takes the strain signal from the `WaveformGenerator` and injects it into a noise realisation of the three interferometers. One could choose to do a zero-noise simulation by simply including the flag `zero_noise=True`.

⁴This example is found in the BILBY git repository at https://git.ligo.org/Monash/bilby/blob/master/examples/injection_examples/basic_tutorial.py.

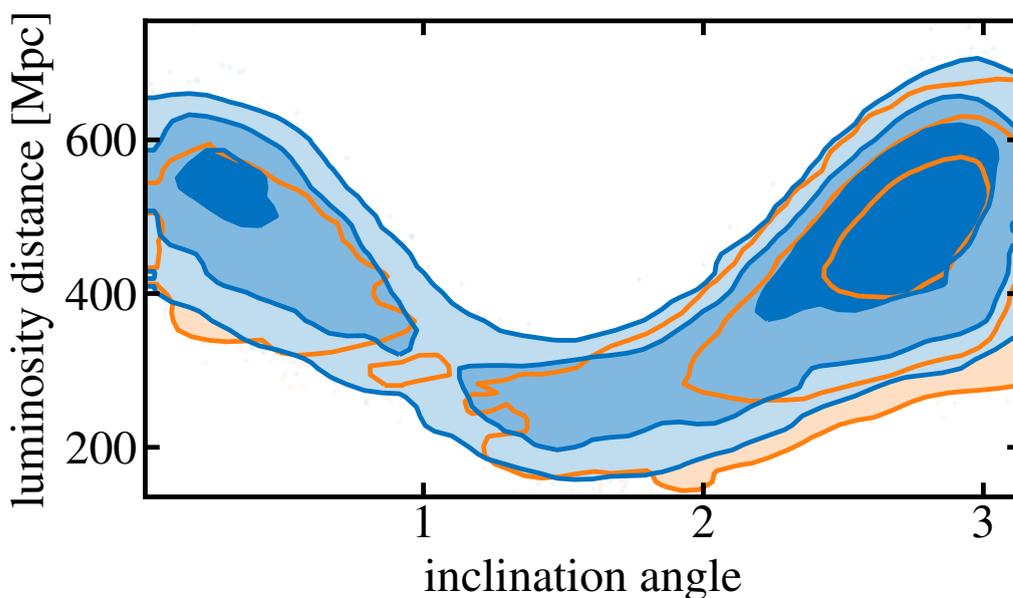


FIGURE 2.2: Marginalized posterior distributions on the binary inclination angle and luminosity distance for the first binary black hole merger detected by LIGO, GW150914. We show the posterior distributions recovered using BILBY (blue), and those using LALINFERENCE (orange), using open data from the Gravitational Wave Open Science Centre [285].

Priors are set up as in the previous open data example, except we call the `binary_black_holes.prior` file instead of the specific prior file for GW150914. Moreover, to hold all but four of the parameters fixed, we set the value of the prior for those other parameters to the injection value. For example, setting

```
>>> prior['a_1']=0
```

sets the prior on the dimensionless spin magnitude of the primary black hole to a delta-function at zero.

In general, we can change the prior for any parameter with one line of code. For example, to change the prior on the primary mass to be uniform between $m_1 = 25 M_\odot$ and $35 M_\odot$, say, one includes

```
>>> prior['mass_1']=bilby.core.prior.Uniform(minimum=25, maximum=35,
↪ unit=r'$M_\odot$')
```

BILBY knows about many different types of priors that can all be called in this way. For this example we are also required to define priors on the coalescence time, which we define to be a uniform prior with minimum and maximum one second either side of the injection time.

The likelihood is again set up similarly to the open-data example of Sec. 2.4.1, although this time we must pass the `interferometer`, `waveform_generator`, and `prior`. Finally,

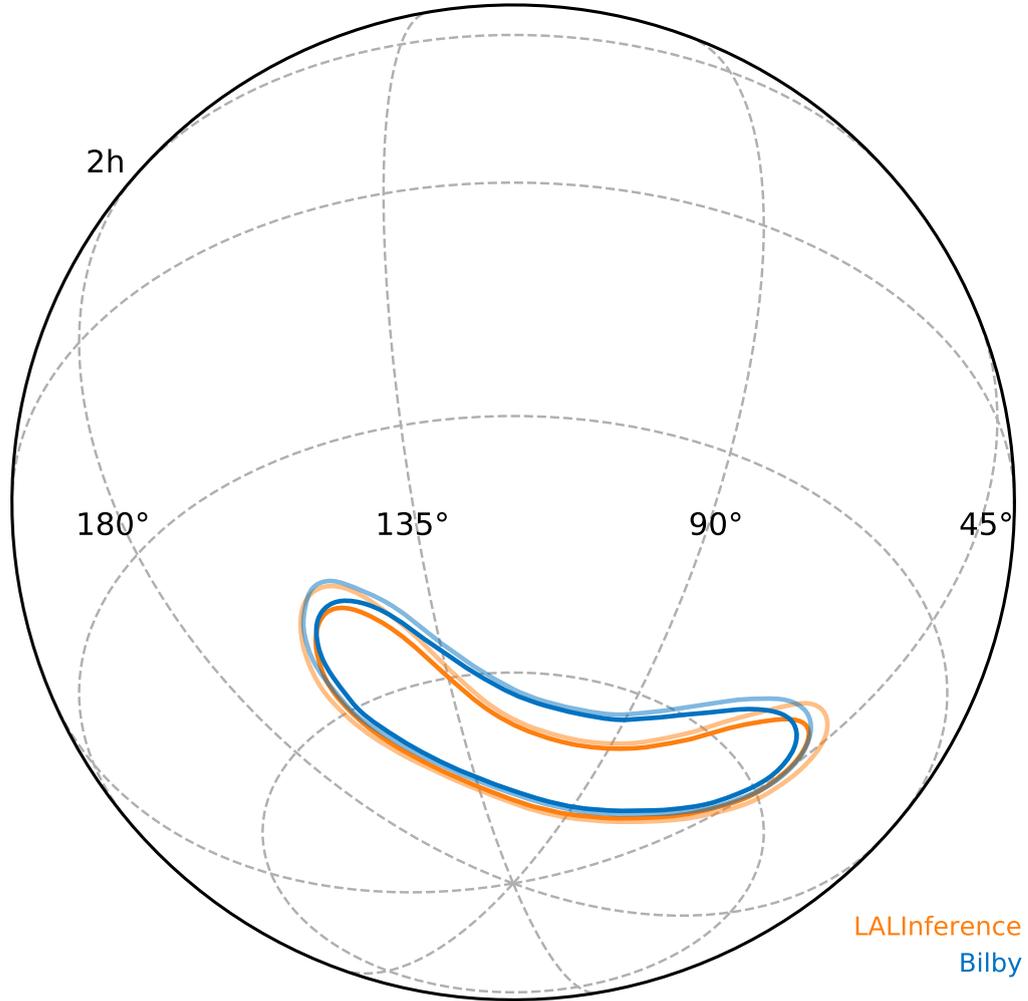


FIGURE 2.3: Sky localisation uncertainty for GW150914. The blue marginalized posterior distributions are those recovered using BILBY, and the orange are those recovered using LALINFERENCE, using open data from the Gravitational Wave Open Science Center [285].

the sampler can be called in the same way as Sec. 2.4.1; for this example we use the `pyMultiNest` nested sampler [72].

Figure 2.4 shows the recovered posterior distributions (blue) and the injected parameter values (orange). For this example, using the `PYMULTINEST` [72] nested sampling package with 6000 live points took approximately 30 minutes on a laptop to sample fully the four-dimensional parameter space. The parameters in Fig. 2.4 are recovered well with the usual degeneracy present between the luminosity distance and inclination angle of the source, d_L and ι , respectively.

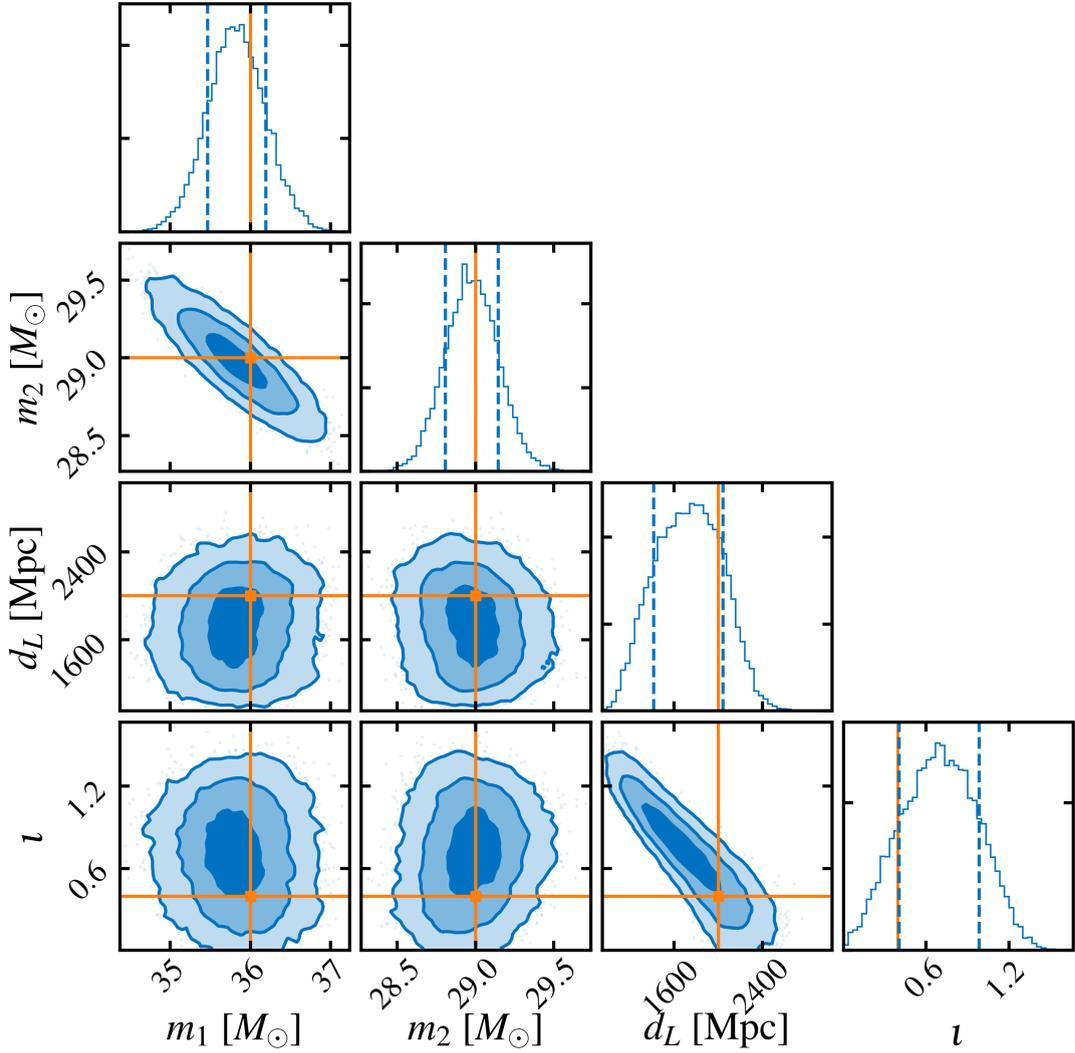


FIGURE 2.4: Injecting and recovering a binary black hole gravitational-wave signal with BILBY. We inject a signal into a three-detector network of LIGO-Livingston, LIGO-Hanford, and Virgo and perform parameter estimation. The posterior distributions are shown in blue and the injected values in orange. To speed up the simulation we only search over the two black hole masses m_1 and m_2 , the luminosity distance d_L , and the inclination angle i .

2.4.3 Measuring tidal effects in binary neutron star coalescences

The first detection of binary neutron star coalescence GW170817 was a landmark event signalling the beginning of multimessenger gravitational-wave astronomy [185, 277]. Gravitational-wave parameter estimation of the inspiral is what ultimately determined that both objects were likely neutron stars, and provides the best-yet constraints on the nuclear equation of state of matter at supranuclear densities [10, 185, 277].

TABLE 2.2: Default binary neutron star priors. $\Lambda_{1,2}$ are the tidal deformability parameters of the primary and secondary neutron star defined in Eq. 2.2. For other variable definitions, see Tab. 2.1. Note our commonly-used waveform approximant does not allow misaligned neutron star spins, implying we do not require priors on those spin parameters.

variable	unit	prior	minimum	maximum
$m_{1,2}$	M_{\odot}	uniform	1	2
$a_{1,2}$	-	uniform	-0.05	0.05
$\Lambda_{1,2}$	-	uniform	0	3000
d_L	Mpc	comoving	10	500
ra	rad.	uniform	0	2π
dec	rad.	cos	$-\pi/2$	$\pi/2$
ι	rad.	sin	0	π
ψ	rad.	uniform	0	π
ϕ_c	rad.	uniform	0	2π

One of the key measurements in determining the equation of state from binary neutron star coalescences is that of the tidal parameters. The dimensionless tidal deformability

$$\Lambda = \frac{2k_2}{3} \left(\frac{c^2 R}{Gm} \right)^5, \quad (2.2)$$

is a fixed parameter for a given equation of state and neutron star mass. Here, k_2 is the second Love number, R and m are the neutron star radius and mass, respectively. The binary neutron star merger GW170817 provided constraints of $\Lambda_{1,4} = 190_{-120}^{+390}$ [16, 91], where the subscript denotes this is the estimate on Λ assuming a $1.4 M_{\odot}$ neutron star, and the uncertainty is the 90% credible interval.

BILBY can be used to study neutron star coalescences in both real and simulated data. We inject a binary neutron star signal using the `TaylorF2` waveform approximant into a three-detector network of the two LIGO detectors and Virgo, all operating at design sensitivity⁵. Our injected signal is an $m_1 = 1.3 M_{\odot}$, $m_2 = 1.5 M_{\odot}$ binary at $d_L = 50$ Mpc with dimensionless spin parameters $a_{1,2} = 0.02$, and tidal deformabilities $\Lambda_{1,2} = 400$. Setting up such a system in BILBY is equivalent to doing the binary black hole injection study of Sec. 2.4.2, except we call the `lal_binary_neutron_star` source function, which requires the additional $\Lambda_{1,2}$ arguments. We also have specific binary neutron star priors; the default set can be called using

```
>>> priors = bilby.gw.prior.BNSPriorDict()
```

The standard set of binary neutron star priors are shown in Tab. 2.2. In this example we use the `Dynesty` sampler [266].

The tidal deformability parameters Λ_1 and Λ_2 are known to be highly correlated. The terms that appear explicitly due to the tidal corrections in the phase evolution are instead $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ [110] (for definitions of these parameters, see Eqs. (14) and (15) of Ref. [175]). We therefore sample in $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$, instead of Λ_1 and Λ_2 . Although we sample in all

⁵This example is found in the BILBY git repository at https://git.ligo.org/Monash/bilby/blob/master/examples/injection_examples/binary_neutron_star_example.py.

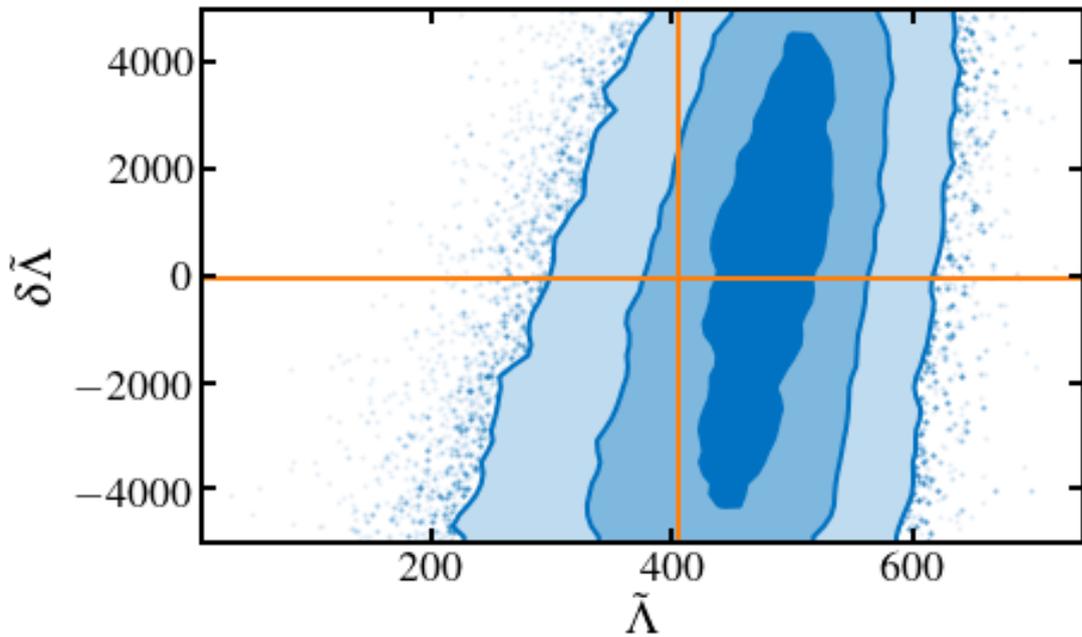


FIGURE 2.5: Injecting and recovering a binary neutron star gravitational-wave signal with BILBY. We inject a signal into the three-detector network, and show here only the marginalized two-dimensional posterior on the two tidal deformability parameters (blue) with the injected values shown in orange.

binary neutron star parameters, we show only the two-dimensional marginalized posterior distribution for $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ in Fig. 2.5. The corresponding injected values of $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ are shown as the orange vertical and horizontal lines, respectively.

2.4.4 Implementing New Waveforms

The preceding subsections have only given a flavour of what can be achieved with BILBY for compact binary coalescences. It is trivial to implement more complex signal models that include, for example, higher order modes, eccentricity, gravitational-wave memory, non-standard polarizations. Examples showing different signal models are included in the `git` repository [46]. BILBY has already been used in one such application: testing how well the orbital eccentricity of binary black hole systems can be measured with Advanced LIGO and Advanced Virgo [189]. An example script reproducing those results can be found in the `git` repository [46].

If a signal model exists in the LAL software [184], then calling that signal model and defining which parameters to include in the sampler is as simple as the above examples. In Sec. 2.5 we also show how to include a user-defined source model. Moreover, one is free to define and sample models in either the time or frequency domain. We include examples for both cases in the `git` repository. The latter case of using a time-domain source model requires doing little more than selecting the argument `time_domain_source_model` in the `WaveformGenerator`, rather than selecting `frequency_domain_source_model`.

Of course, one may also want to set up the injection and the sampler using two different waveform models, for example to inject a numerical relativity signal into Monte Carlo

data and recover it with a waveform approximant (see also Sec. 2.5.1). This is possible by simply instantiating two `WaveformGenerators`, injecting with one and passing the other to the `likelihood`.

2.4.5 Adding detectors to the network

The full network of ground-based gravitational-wave interferometers will soon consist of the two LIGO detectors in the US, Virgo, LIGO-India [161] and the KAGRA detector in Japan [48], all of which are implemented in BILBY. A gravitational-wave interferometer is specified by its geographic coordinates, orientation, and noise power spectral density. By default, BILBY includes descriptions of current detectors including LIGO, Virgo, and KAGRA, as well as proposed future detectors, A+ [202], Cosmic Explorer [9], and the Einstein Telescope [230]. It is also possible to define new detectors, which is useful for developing the science case for proposals and to optimize the design and placement of new detectors. Among other things, this can be used in developing the science-case for interferometer design and placement.

BILBY provides a common interface to define detectors by their geometry, location, and frequency response. By way of example, we place a new four-kilometer-arm interferometer in the Shire of Gingin, located outside of Perth, Australia; the current location of the Australian International Gravitational Observatory (AIGO). We assume a futuristic network configuration of the Australian Observatory together with the two LIGO detectors in Hanford and Livingston, all operating at A+ sensitivity [202]. We generate A+ power spectral densities in the same script used to run BILBY by using the PYGWINC software [239], which creates an array containing the frequency and noise power spectral density⁶ (one could equally use more sophisticated software such as FINESSE [71] to create more detailed interferometer sensitivity curves). We then create a new `Interferometer` object using `bilby.gw.detector.Interferometer()`, which takes numerous arguments including the position and orientation of the detector, minimum and maximum frequencies, and the power or amplitude noise spectral density. The noise spectral density can be passed as an ascii file containing the frequency and spectral noise density. With the new detector defined, one can again calculate a noise realisation and signal injection in a manner similar to what is done in Sec. 2.4.

In this example we inject a GW150914-like binary black hole inspiral signal at a luminosity distance of $d_L=4$ Gpc, and recover the masses, sky location, luminosity distance and inclination angle of the system. In this example we use the `Nestle` sampler [53]. Figure 2.6 shows the two-dimensional marginalized posterior for the sky-location uncertainty when including (blue) and not including (orange) the Australian detector in Gingin. In this instance, the sky localisation uncertainty decreases by approximately a factor four when including the third detector.

While this example includes three detectors, it is straightforward to extend this analysis to an arbitrary detector network. The likelihood evaluation simply loops over the number of detectors passed to it and multiplies the likelihood for each detector to get a combined likelihood for each point in the parameter space.

⁶This example is found in the BILBY git repository at https://git.ligo.org/Monash/bilby/blob/master/examples/injection_examples/Australian_detector.py.

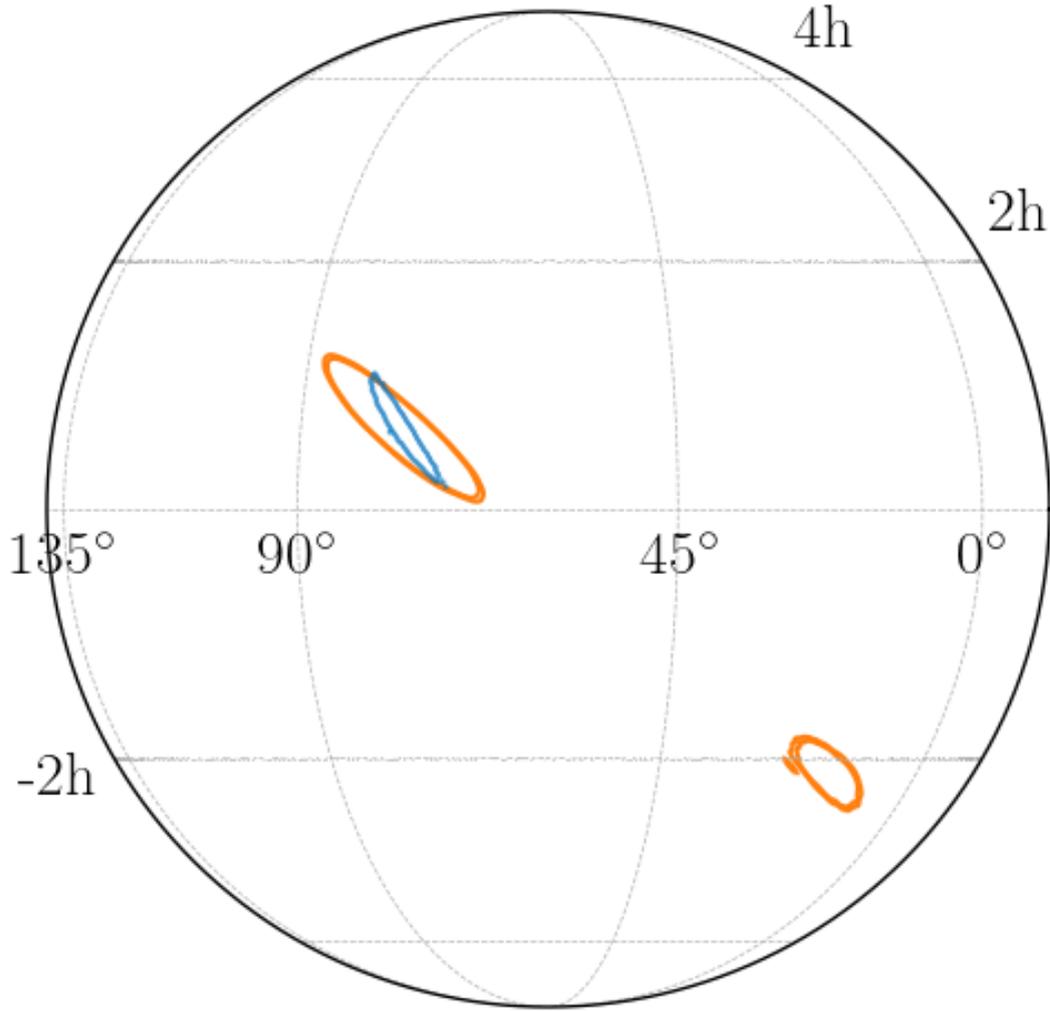


FIGURE 2.6: Sky location uncertainty when including a gravitational-wave detector in Gingin, Australia. Shown are the sky localisations (marginalized two-dimensional posterior distributions) for an injected binary black hole signal using a two-detector network of gravitational-wave interferometers Hanford and Livingston (orange) and a three-detector network that also includes the Australian detector (blue).

2.5 Alternative signal models

Section 2.4 focuses on compact binary coalescences. However, the BILBY `gw` package enables parameter estimation for any type of signal for which a signal model can be defined. In this section, we show two illustrative examples: the injection and recovery of a core-collapse supernovae signal, and a much-simplified model of a hypermassive neutron star following a binary neutron star merger. The former example highlights two key pieces of infrastructure; the ability to inject numerical relativity signals, and to develop one's own source model that is not built into BILBY. The latter example highlights the use of a different likelihood function that only uses the amplitude of the signal, and throws away the phase information.

2.5.1 Supernovae

Gravitational-wave signals from core-collapse supernovae are complicated and not well understood in terms of their specific phase evolution. Numerous techniques have been developed to deal with both detection and parameter estimation. One such method for the latter problem involves principal component analysis [186, 224, 226], where the signal is reconstructed using a weighted sum of orthonormal basis vectors. In this example, we inject a gravitational-wave signal from a numerical relativity simulation [209] and recover the principal components using BILBY⁷.

The injection is performed by defining a new signal class that, in this case, simply reads in an ascii text file containing the gravitational-wave strain time series. The injection is then performed in a way akin to the binary black hole and binary neutron star examples in Sec. 2.4. We inject signal L15 from Ref. [209], which comes from a three-dimensional simulation of a non-rotating core-collapse supernova with a 15 M_{\odot} progenitor star. The signal is injected at a distance of 5 kpc in the direction of the galactic center. The amplitude spectral density of the injected signal is shown in Fig. 2.7 as the orange trace.

The signal is reconstructed using principal component analysis, such that the strain is expressed as

$$\tilde{h}(f) = A \sum_{j=1}^k \beta_j U_j(f), \quad (2.3)$$

where A is an amplitude factor, β_j and U_j are the complex principal component amplitudes and vectors, respectively. Equation (2.3) is implemented into BILBY as another new signal model that takes the β_j coefficients, luminosity distance (which is a proxy for A), and sky location as inputs. Priors for each of the new parameters are established in the same way as the example with the mass in Sec. 2.4.2. In this case, we set $k = 5$ and use uniform priors between -1 and 1 for each of the β_j 's.

Figure 2.7 shows the injected (orange) and recovered (blue) gravitational-wave signal in the frequency domain. The dark blue curve shows the maximum likelihood curve, and the shaded blue region is a superposition of many reconstructed waveforms from the posterior samples.

2.5.2 Neutron star post-merger remnant

There are a number of physical scenarios that can occur following the merger of two neutron stars, including the existence of short- or long-lived neutron star remnants. In the early phases post-merger ($\lesssim 1$ s), these neutron stars are highly dynamic, and can emit significant gravitational radiation potentially observable by Advanced LIGO and Virgo at design sensitivity out to ~ 50 Mpc [e.g., 80, and references therein]. While the ultimate fate of binary merger GW170817 is unknown, no gravitational waves from a post-merger remnant were found [15, 21], which is not surprising given the interferometers were not operating at design sensitivity and the distances involved.

⁷This example is found in the BILBY git repository at https://git.ligo.org/Monash/bilby/blob/master/examples/supernova_example/supernova_example.py.

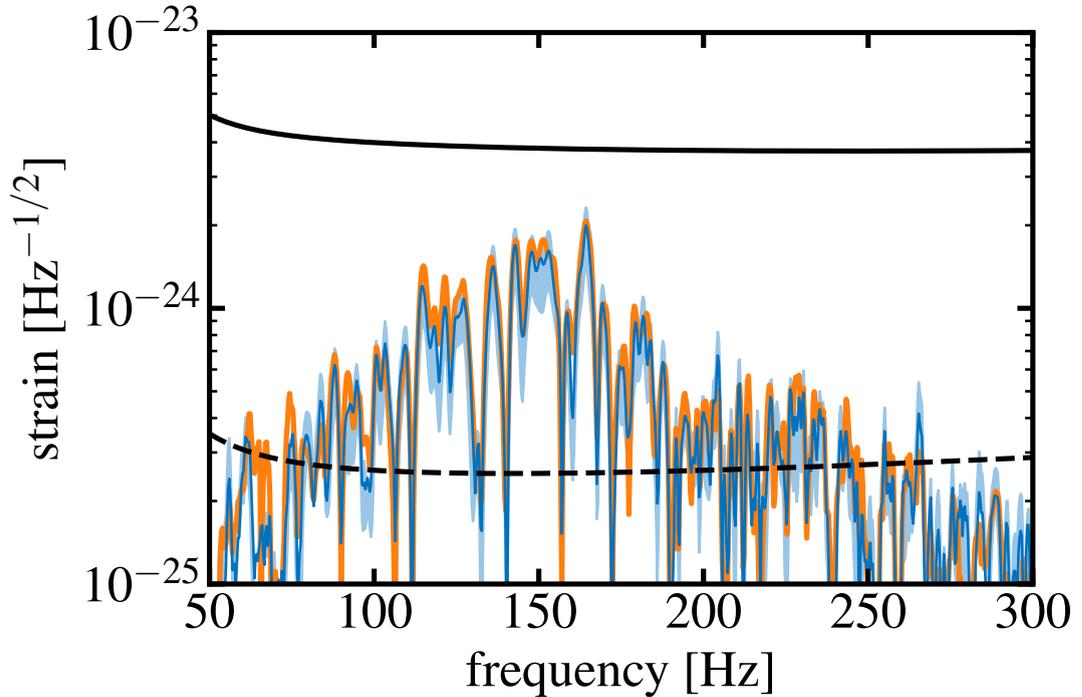


FIGURE 2.7: Parameter estimation reconstruction of a numerical relativity supernova signal. A numerical relativity supernovae signal (orange) is injected into a three-detector network of the two Advanced LIGO detectors and Advanced Virgo, all operating at design sensitivity. The maximum likelihood reconstruction of the signal is shown in dark blue, and the blue light band shows the superposition of many reconstructed waveforms from the posterior samples.

Providing the sensitivity of gravitational-wave interferometers continues to increase, it is possible a gravitational-wave signal from a post-merger remnant could be detected in the relative near future. Such a detection would provide an excellent opportunity to understand the nuclear equation of state of matter at extreme densities, as well as the rich physics of these exotic objects [e.g., 50, 236, 252]. Parameter inference of such short-lived signals is in its infancy [e.g., see 76], largely due to the paucity of reliable waveforms [81, 97]. This is an ongoing challenge due to the expensive nature of numerical relativity simulations and the complex physics that must be included in such simulations.

Simple models that provide approximate gravitational-wave signals fit to a handful of numerical relativity waveforms exist [67, 97, 200], which may eventually be used for full parameter inference. The phase evolution of such numerical relativity simulations is rapid, and very difficult to model [97, 200]. However, it is the frequency content of the signal that carries information about the equation of state and the physics of the remnant [e.g., 271, and references therein]. It is therefore possible that parameter-estimation algorithms may require one to throw away information about the phase, and only keep amplitude spectral content. Such a process requires a different likelihood function than the one that has been used to this point. This therefore provides good motivation for showing how to include a different likelihood function in BILBY code.

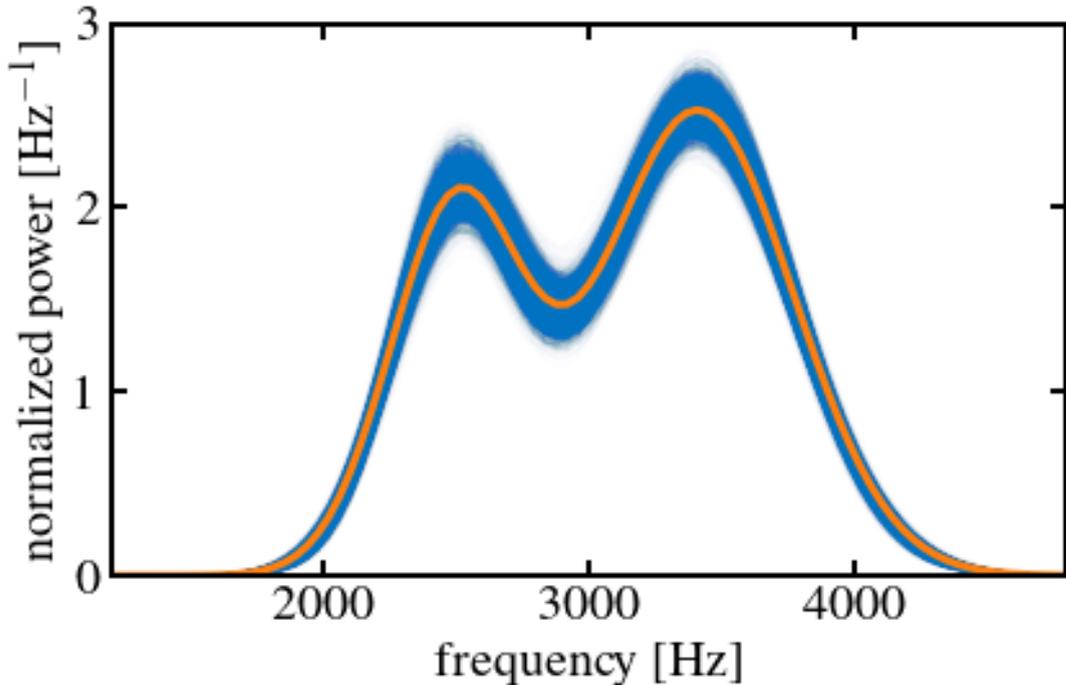


FIGURE 2.8: A proxy post-merger gravitational-wave signal from a short-lived neutron star showing the implementation of a different likelihood function in BILBY. The orange curve is an injected, double-peaked Gaussian signal injected into a constant noise realisation. The blue band shows the waveform reconstructions from the posterior samples using a power-spectrum likelihood function; i.e., one that only uses the amplitude of the signal and ignores the phase.

We implement a power-spectral density (“burst”) likelihood

$$\ln \mathcal{L}(|d| | \theta) = \sum_{i=1}^N \left[\ln I_0 \left(\frac{|\tilde{h}_i(\theta)| |\tilde{d}_i|}{S_n(f_i)} \right) - \frac{|\tilde{h}_i|^2 + |\tilde{d}_i|^2}{2S_n(f_i)} + \ln |\tilde{h}_i(\theta)| - \ln S_n(f_i) \right], \quad (2.4)$$

where I_0 is the zeroth-order modified Bessel function of the first kind. This requires setting up a new Likelihood class, that contains a `log_likelihood` function that reads in the frequency array, noise spectral density and waveform model, and outputs a single likelihood evaluation. Having defined a new likelihood function, one calls the remaining functions in the usual way; the likelihood function is instantiated and passed to the `run_sampler()` command.

We inject a double-peaked Gaussian, shown in Fig. 2.8 as the solid orange curve. We recover this signal using the same model (with a constant noise spectral density), where we use uniform priors for the amplitudes, widths and frequencies of each of the peaks. Figure 2.8 shows the waveform reconstruction for each of the posterior samples, which can be seen to cover the injected signal.

2.6 Population Inference: hyperparameterizations

Individual detections of binary coalescences can provide stunning insights into various physical and astrophysical questions. Increased detector sensitivities imply significantly more events will be detected, enabling statements to also be made about ensemble properties of populations [e.g., 5, 99, 243, 257, 272, 275, 313, and references therein]. Extracting information from a population of events is performed using hierarchical Bayesian inference where the population is described by a set of hyper-parameters, Λ . BILBY has built-in support for calculating Λ from multiple sets of posterior samples from individual events.

BILBY implements the conventional method whereby the posterior samples θ_i^j for each event j are re-weighted according to the ratio of the population model prior $\pi(\theta|\Lambda)$ and the sampling prior $\pi(\theta)$ to obtain the hyper-parameter likelihood

$$\mathcal{L}(h|\Lambda) = \prod_j^N \frac{\mathcal{Z}_j}{n_j} \sum_i^{n_j} \frac{\pi(\theta_i^j|\Lambda)}{\pi(\theta_i^j)}. \quad (2.5)$$

Here, \mathcal{Z}_j is the Bayesian evidence for the data given the original model and n_j is the number of posterior samples in the j^{th} event.

The BILBY implementation requires the user to define $\pi(\theta|\Lambda)$ and $\pi(\theta)$ which, along with the set of posterior samples θ_i^j , are passed to the `HyperparameterLikelihood` in BILBY’s `hyper` package. The hyperparameter priors are then set up in the usual way, and passed to the standard `run_sampler` function.

As a demonstration ⁸ of this method we reproduce results [272] recovering parameters describing a postulated excess of black holes due to pulsational pair-instability supernovae (PPSN) [138, 312]. The posterior distribution for the hyperparameters determining the abundance and characteristic mass of black holes formed through this mechanism are shown in Fig. 2.9. The hyperparameter λ is the fraction of binaries where the more massive black hole formed through PPSN, μ_{pp} is the typical mass of these black holes and σ_{pp} determines the width of the “PPSN graveyard”.

This model contains seven additional hyperparameters describing the remainder of the distribution of black hole masses that we hold fixed for the purposes of this example. Additional hyperparameters may be added straightforwardly.

2.7 Analysis of arbitrary data: an example

BILBY is more than a tool for gravitational-wave astronomy; it can also be used as a generic and versatile inference package. In the documentation examples, we demonstrate how BILBY can be applied to generic time-domain data from radioactive decay processes. Furthermore, BILBY is currently being used to analyse radio and x-ray data from neutron stars, and to study multi-messenger signals associated with binary neutron star mergers. Here we show an example that calculates posterior distributions for one of the letters in the BILBY logo.

⁸This example is found in the BILBY git repository at https://git.ligo.org/Monash/bilby/blob/master/examples/other_examples/hyper_parameter_example.py.

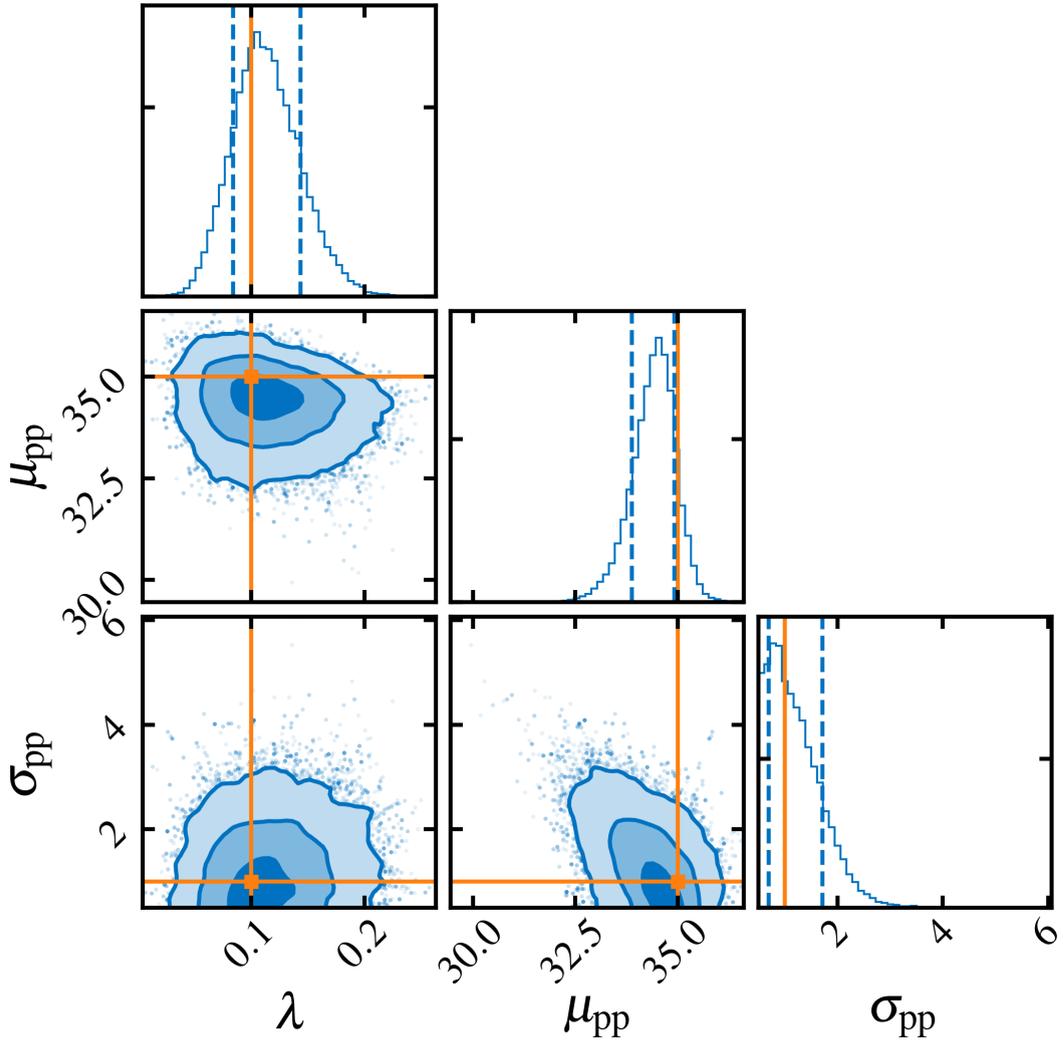


FIGURE 2.9: Population modelling with BILBY hierarchical Bayesian inference module. We show the recovery of parameters describing part of the mass distribution of binary black holes using the model described in Ref. [272]. The population parameters are drawn from values shown in orange, and the posterior distributions for the hyperparameters shown in blue. Here, λ is the fraction of binaries where the more massive black hole formed through pulsational pair-instability supernovae, μ_{pp} and σ_{pp} are the typical mass of these black holes and the width of the “PPSN graveyard”, respectively.

We import an image file containing the letter, map this to an x - y coordinate system and sample in both dimensions with likelihood

$$\ln \mathcal{L} \propto \frac{-1}{xy}, \quad (2.6)$$

assuming uniform priors on both variables. Figure 2.10 shows the posterior distribution for the “B” in the BILBY logo. All letters are shown in Fig 2.11, where the axis labels have been removed. The code for making this plot, and all other posterior distributions in the logo, are available with the git repository [46] in `sample_logo.py`.

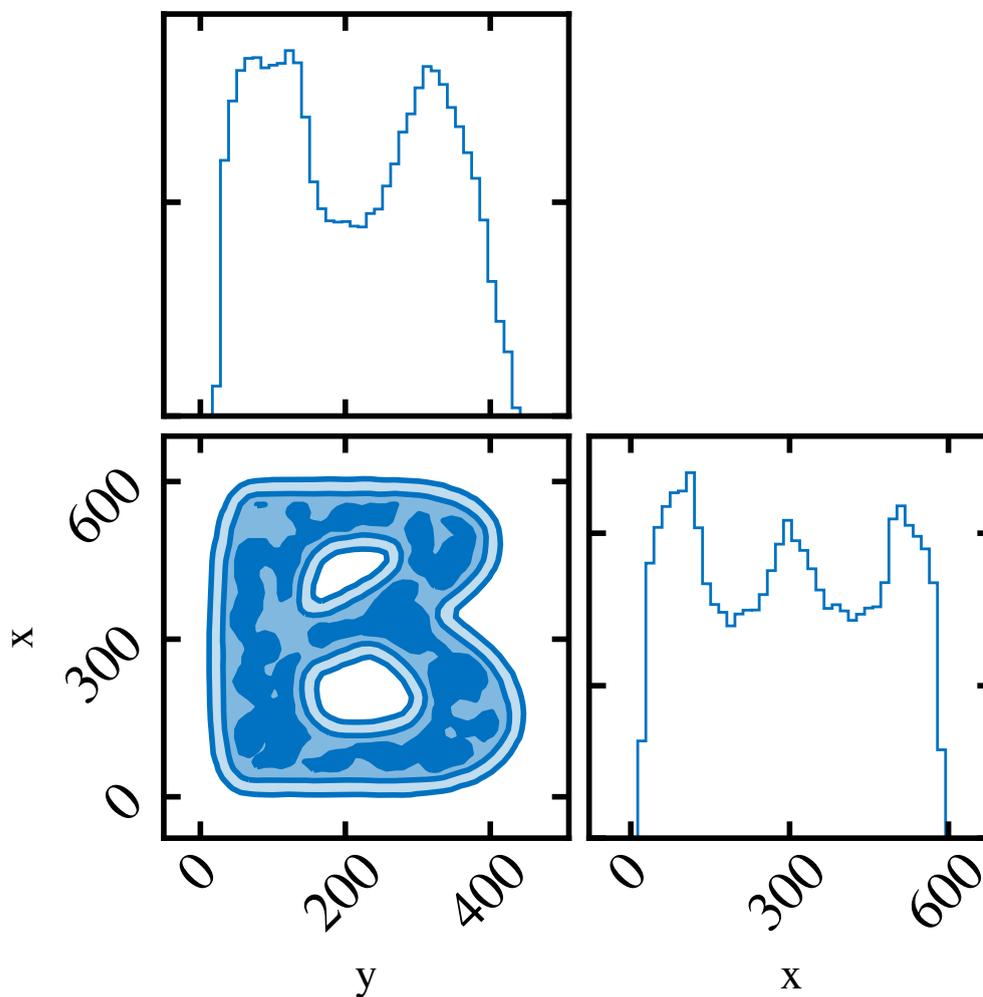


FIGURE 2.10: The ‘B’ from the BILBY logo, generated using the BILBY package; see Sec. 2.7

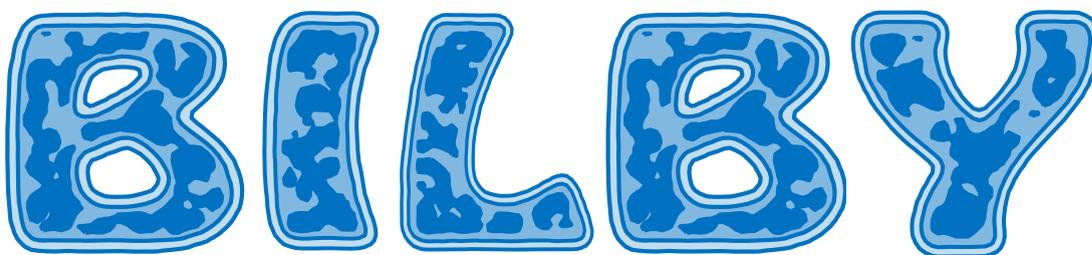


FIGURE 2.11: All letters from the BILBY logo, generated using the BILBY package; see Sec. 2.7

2.8 Conclusion

Gravitational-wave astronomy is fast becoming a data-rich field. With the significantly increased activity in the field, there is a developing need for robust, easy-to-use inference

software that is also modular and adaptable. We present BILBY: the Bayesian inference library for gravitational-wave astronomy. BILBY is open-source software that can be used to perform Bayesian inference. It is easily applied to data from LIGO/Virgo, including open data available from the Gravitational Wave Open Science Center. We access and manipulate LIGO data using `GWPy` [191]. Alternatively, BILBY may be used to study simulated data. BILBY can also be used to perform hierarchical Bayesian inference for population studies.

We present examples highlighting BILBY’s functionality and usability, including examples using open data from the first gravitational-wave detection GW150914. Only five lines of code are required to reconstruct the astrophysical parameters of GW150914. One can redo the analysis using different priors, alternative waveform models, and/or a different sampling method with only modest changes. We show how to inject binary black hole and binary neutron star signals into Monte Carlo noise. We show how to define new gravitational-wave detectors.

We emphasise that BILBY is a front-end system that provides a unified interface to a variety of samplers, which are a primary workhorse of Bayesian inference. While numerous off-the-shelf samplers are implemented (see Sec. 2.3.2.1), to the best of our knowledge there is no universal sampling solution to gravitational-wave parameter estimation problems. BILBY is therefore only as good as the implemented samplers; initial studies show that `CPNest` [296], `Dynesty` [266], and `emcee` [114, 303] sample the extrinsic parameters of binary coalescences more accurately than `Nestle` [53] and `pyMultiNest` [72]. A systematic comparison of all off-the-shelf and boutique samplers is currently underway using BILBY.

BILBY is designed so as to be applicable to arbitrary signal models, not just compact binary coalescences. To this end we show two examples: one of an injected numerical relativity supernova waveform that we reconstruct using principal component analysis, and another using a proxy for a neutron star post-merger waveform. The former example highlights how one can include their own signal models to perform both injections and signal recoveries, while the latter example demonstrates the ability to add a likelihood function that is different from the standard gravitational-wave transient likelihood.

2.9 Acknowledgements

We are grateful to John Veitch and Christopher Berry who provided valuable comments on the manuscript, and also the LIGO/Virgo Parameter Estimation group for insightful discussions. This work is supported through Australian Research Council (ARC) Centre of Excellence CE170100004. PDL is supported through ARC Future Fellowship FT160100112 and ARC Discovery Project DP180103155. SB is partially supported by the Australian-American Fulbright Commission. MDP is funded by the UK Science & Technology Facilities Council (STFC) under grant ST/N005422/1. ET is supported through ARC Future Fellowship FT150100281 and CE170100004. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (<https://www.gw-openscience.org>), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della

Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. The BILBY package makes use of the standard scientific PYTHON stack [163, 198, 216], MATPLOTLIB [144], CORNER [112], and HEALPY [129] for the generation of figures, DEEPDISH [177] for HDF5 [276] file operations, and ASTROPY [229, 238] for common astrophysics-specific operations.

Chapter 3

Further Information about Bilby

3.1 Open Source Projects in the Scientific Context

The BILBY project has been a huge success by any measure since we published the first paper in 2018 [44]¹. It has become one of the standard Bayesian inference tools by the LIGO Scientific Collaboration for the analysis of gravitational waves (see e.g. Refs. [25–27, 29–31, 278]). It has also been used in many specific studies such as measurements of eccentricity [240, 241], physical detector calibration models [219, 302], and studies of gravitational waves from supernovae [225] to name a few. Additionally, there has been adoption outside gravitational-wave physics, for instance in the analysis of gamma-ray bursts (e.g. Refs. [188, 247, 305]), fast radio bursts (e.g. Refs. [77, 98, 172, 231]), pulsar timing (e.g. Refs. [127, 128, 171, 190]), the 2016 Vela pulsar glitch [42, 206], and others [119, 201, 317].

As laid out in Chap. 2, we have put effort into using good coding practices, using a modular, abstract, object-oriented structure, and enforcing consistent code with the PEP8 protocol. Over the development of the project, we ensured backwards compatibility in the application programming interface (API), which ensures that users can safely upgrade to newer versions of BILBY. We additionally created unit tests which are covering large parts of the code base [132]. Unit tests aim to test the code one feature at a time, ensuring that the code can handle all edge cases correctly. Unit testing also encourages overall good coding practises as it forces us to write “testable” code with small functions and methods that perform as little logic at a time as possible. We execute unit tests automatically on several docker images with reference PYTHON setups every time a commit is pushed onto the repository.

Project management also has played a large role in the success of BILBY. The project follows standard methods of software development on its GITLAB page. We track identified issues and changes are submitted via merge requests on separate branches. Code is only merged after it has been reviewed by at least two of the package maintainers. We typically go through rounds of comments on merge requests to ensure contributors add code that complies with the package standards. We also foster a community of users on a dedicated SLACK workspace.

¹The repository is available at <https://git.ligo.org/lscsoft/bilby>

3.2 Extensions to the BILBY Package

The BILBY project has focused on providing a stable and intuitive API. This has allowed for extension packages and projects to be created that are in part or wholly based off of BILBY.

3.2.1 Bilby pipe

With the increasing number of observed gravitational-wave events and the adoption of BILBY as a standard tool, it has become vital to automate the deployment of the analysis as much as possible. As an extension to BILBY, the BILBY_PIPE² package was introduced in Ref. [242]. BILBY_PIPE provides an interface either via the command line or via the creation of submission files such as the one in Listing A.1. Creating configuration files makes it much easier in practice to track the analysis settings and identify issues. BILBY_PIPE interfaces with BILBY and GWPY to retrieve an analysis segment around a specified trigger time and then calls to BILBY to perform the data analysis.

Production gravitational-wave analysis is typically run on supercomputers that use a scheduling system for job submissions. BILBY_PIPE automatically creates the required job submission files for the SLURM and HTCONDOR systems that are used in the OzStar and LIGO compute clusters. BILBY_PIPE is now the standard tool for job submission and has been used in the GWTC-1 re-analysis in Ref. [242], and is actively being used for O3 analysis runs [25–27, 29–31, 278]. Additionally, BILBY_PIPE can also create simulated data and can thus be used for large-scale injection studies.

3.2.2 pBILBY

Inference runs on gravitational-wave events can be exceptionally costly if the waveform model is expensive to evaluate. Standard single-core analysis sometimes does not terminate within an acceptable time for production analysis. Based off the BILBY package, Ref. [258] introduced the “parallel Bilby” pBILBY package. pBILBY uses Message Parsing Interface (MPI) to distribute a BILBY inference run that uses DYNESTY over multiple cores. Effectively, pBILBY achieves a speedup by using a pool of CPUs to draw samples in parallel at each iteration of the nested sampling algorithm. The speedup S of this depends on the number of live points n_{live} and the number of cores n_{core}

$$S = n_{\text{live}} \ln \left(1 + \frac{n_{\text{core}}}{n_{\text{live}}} \right) \approx n_{\text{core}}, \quad (3.1)$$

where the approximation is valid if $n_{\text{live}} \gg n_{\text{core}}$.

Using pBILBY is especially useful for events where we deal with expensive waveform models and segment lengths. This is typically the case for binary neutron star mergers and neutron star black hole mergers, or some lower-mass black hole binary systems. Reference [258] shows that the wall time of runs with some of the most expensive waveform models such as SEOBNRv4PHM [217], can be analysed in $\mathcal{O}(\text{week})$ for typical events by leveraging hundreds of CPUs.

²Code available at https://git.ligo.org/lscsoft/bilby_pipe

3.2.3 PESUMMARY

While BILBY and BILBY_PIPE are central to performing the data analysis, practical collaboration work requires a standardised way of visualising and distributing results. Each posterior from an analysed event still requires manual revision to ensure that analysis parameters and priors were sufficient and that the posterior is converged. Within the LIGO collaboration this is achieved with PESUMMARY³ [142], which performs consistent post-processing operations for BILBY and other LIGO inference software, LALINFERENCE [298], PyCBC [57], and RIFT [176]. PESUMMARY standardises the output into a single “meta-file” that can be shared more easily across the collaboration and also implements several plotting functions. More importantly, PESUMMARY creates summary pages for the runs which can be accessed via a browser window. This means that collaborators can be more easily involved in reviewing results without requiring expert knowledge about the software.

3.2.4 Bilby-MCMC

BILBY-MCMC is a parallel-tempered ensemble Metropolis-Hastings sampler, that implements gravitational-wave specific proposals [43]. It has been released as part of the main BILBY package and can be accessed the same way as the other sampling packages by using the `sampler='bilbymcmc'` keyword argument in the sampling interface. For gravitational-wave specific problems, BILBY-MCMC reaches a higher efficiency in drawing new live points, thereby speeding up the sampling process. Evidence calculation within BILBY-MCMC is performed via thermodynamic integration [178]. One of the main advantages of BILBY-MCMC is that it is embarrassingly simple and asynchronously parallelised, which allows it to greatly reduce the time required to perform analyses. Unlike PBILBY, BILBY-MCMC does not require much communication between computing nodes. This makes it more suitable for use on high-throughput computing clusters compared to PBILBY.

3.3 Improvements in the BILBY Package

As a follow-up to Chap. 2, we published Ref. [242], which introduced new features and validated the use of BILBY for gravitational-wave analyses. Many more features have been added to the code base by the community of developers. In the following, I highlight features that I have contributed in part or wholly to the project.

3.3.1 Constrained and conditional Priors

In many inference problems, we wish to remove parts of the prior space by placing restrictions on relationships between parameters. For example, the usual convention in gravitational-wave analysis when sampling in component masses is to enforce $m_1 \geq m_2$. We can implement this constraint in BILBY by adding a `Constraint` prior and a conversion function to the regular BILBY `PriorDict`. While the specific case of $m_1 \geq m_2$ is coded into the `CBCPriorDict`, we show how to do it manually as a pedagogical

³The repository is located at <https://git.ligo.org/lscsoft/pesummary/>

example in Listing 3.1. To define a constrained prior, we have to convert the component mass parameters to a new parameter in which the constraint is given by a simple prior bound. In our case, we use the asymmetric mass ratio $q = m_2/m_1$, which we can cap to be 1 at maximum to enforce the constraint. After applying constraints, BILBY prior distributions are no longer correctly normalised. To compensate for that we use an updated normalisation based on a Monte-Carlo integral.

While constrained priors are handy in some simple applications, there are limits to their usefulness. When drawing a prior sample, the constraint is evaluated after all other parameters have been drawn and the sample is thrown out if the constraint is not fulfilled. For our example with the component masses, this means we discard 50% of the prior samples initially, though we discard far fewer samples as the live points are converging towards the posterior distribution. However, if the constraints exclude a far larger fraction of the parameter space, sampling can become relatively inefficient. For example, suppose we would like to fit a time series with a number of identical Gaussians

$$f(t; A, \mu_i, \sigma) = \sum_{i=1}^N A \exp\left(-\frac{(\mu_i - t)^2}{2\sigma^2}\right), \quad (3.2)$$

where the A and σ are amplitude and width of each Gaussian, and the μ_i are the times of their peaks. This model will have $N!$ degenerate modes arising because the μ_i are not ordered. Creating constraints by, e.g. enforcing $\delta\mu_i = \mu_{i+1} - \mu_i > 0$, is not scalable to more than a few parameters since the probability of drawing all μ_i in the correct order is $1/N!$. One alternative is to use conditional priors. A conditional prior on a parameter θ_1 conditioned on θ_0 can be written as $\pi(\theta_1|\theta_0)$. This means we draw θ_1 after θ_0 and can thus ensure that the posterior sample always fulfils our side constraints. Coming back to the example in Eq. 3.2, we can define conditional priors for our problem in the following way:

$$\begin{aligned} \pi(\mu_1) &= \text{Unif}(t_{\min}, t_{\max}), \\ \pi(\mu_2|\mu_1) &= \text{Unif}(\mu_1, t_{\max}), \\ \pi(\mu_3|\mu_2) &= \text{Unif}(\mu_2, t_{\max}), \\ &\dots \\ \pi(\mu_N|\mu_{N-1}) &= \text{Unif}(\mu_{N-1}, t_{\max}). \end{aligned} \quad (3.3)$$

By iteratively drawing the μ_i we can now always be sure that our prior sample will be valid and that there are no degeneracies due to the μ_i . There are multiple ways to define conditional priors for this example. In Chap. 7 we show that we can use conditional Beta functions as priors for the μ_i if we assume that all peaks were originally uniformly distributed in time without being ordered in a specific way. Beyond that, BILBY's conditional priors are also used in the analysis of hierarchical models used to infer the population properties of compact objects with LIGO/Virgo [30, 242]. For example, one popular parametrisation of the population distribution of binary black holes is

$$\begin{aligned} p(m_1|m_{\min}, m_{\max}, \alpha) &= \frac{m_1^{-\alpha}}{m_{\max}^{1-\alpha} - m_{\min}^{1-\alpha}}, \\ p(q|m_1, m_{\min}, \beta) &= \frac{m_1^{1+\beta} q^\beta}{m_1^{1+\beta} - m_{\min}^{1+\beta}}, \end{aligned} \quad (3.4)$$

where m_{\min} and m_{\max} are the minimum and maximum allowed masses for the heavier black holes, and α and β are power law indices[22, 109, 242].

Conditional priors are implemented in BILBY using a generic factory class pattern that creates a conditional prior class for each regular prior class, e.g. there is a `ConditionalUniform` class that is created by passing the `Uniform` prior class into the `conditional_prior_factory` in `bilby.core.prior.conditional`. This approach highlights the advantage of the modular, object-oriented approach BILBY is built on. To evaluate conditional priors in practice, we have to use the `ConditionalPriorDict`, which implements the logic of resolving the correct order in which priors need to be drawn. This logic can resolve any allowed chains or forks of conditions. We show how to implement a conditional prior to solve the ordering of m_1 and m_2 in Listing 3.2. To do this, we need to implement a function describing our condition. The first parameter of this function is a dictionary of reference parameters, which refers to parameters of the prior rather than of the model. For example, these reference parameters for a uniform prior are the minimum and maximum we initially define. For a power-law shaped prior it would also include the power-law index as a reference parameter. The remaining arguments are the names of the other priors this conditional prior relies on, which is m_1 in the case of a conditional prior on m_2 . Since conditional priors in BILBY use the underlying logic of regular priors, they are also automatically normalised without the need for Monte-Carlo estimates.

Conditional and constrained priors are both valid options for some examples like component masses, but they should be used with caution as they yield different marginal prior distributions in these parameters. When choosing between these two options some care should be taken to understand how we impact the posterior distribution with these choices. In some instances, it may also be desirable to simply reparametrise the problem to a different set of parameters that are easier to sample in. As we show in Chap. 7, we can also use conditional and constrained priors on different parameters within the same `ConditionalPriorDict`.

```
import bilby

def conversion_function(samples):
    out_samples = samples.copy()
    out_samples['q'] = samples['m_2']/samples['m_1']
    return out_samples

priors = bilby.core.prior.PriorDict()
priors['m_1'] = bilby.core.prior.Uniform(minimum=5, maximum=100)
priors['m_2'] = bilby.core.prior.Uniform(minimum=5, maximum=100)
priors['q'] = bilby.core.prior.Constraint(minimum=0.0, maximum=1.0)
priors.conversion_function = conversion_function

samples = priors.sample(100)
print(samples['m_2']/samples['m_1'])
```

LISTING 3.1: Minimal script to ensure $m_1 \geq m_2$ during sampling using a constrained prior.

```
import bilby

def condition_function_m_2(reference_parameters, m_1):
    new_prior_m_2_prior_bounds = dict(minimum=reference_parameters['minimum'],
                                      maximum=m_1)
    return new_prior_m_2_prior_bounds

priors = bilby.core.prior.ConditionalPriorDict()
```

```

priors['m_1'] = bilby.core.prior.Uniform(minimum=5, maximum=100)
priors['m_2'] = bilby.core.prior.ConditionalUniform(
    condition_func=condition_function_m_2, minimum=5, maximum=100)

samples = priors.sample(100)
print(samples['m_2']/samples['m_1'])

```

LISTING 3.2: Minimal script to ensure $m_1 \geq m_2$ during sampling using a conditional prior.

3.3.2 Boundary Conditions

BILBY now implements three types of boundary conditions, **periodic**, **reflective**, and **None**, for samplers that support these settings. Assume we have a parameter θ with a uniform prior between 0 and 1. The **periodic** boundary condition moves a proposed point $1+\epsilon$ to ϵ , which means that 0 and 1 effectively describe the same point. **reflective** boundaries on the other hand move the same point to $1 - \epsilon$, and the **None** condition rejects the point outright. Not using boundary conditions can bias the posterior samples close to the edges of the prior as we reject a disproportionate amount of points in their vicinity [242]. Some problems can be rewritten to avoid these issues, but many parameters in gravitational-wave analyses such as the mass ratio q , the spin magnitudes $a_{1/2}$, or the sky position parameters, often have significant support close to their boundaries

The boundary conditions are implemented as a settable attribute in the **Prior** classes. **DYNESTY**, which is the reviewed sampler used for LIGO/Virgo analyses, implements all three boundary conditions [266], **PYMULTINEST** implements the **periodic** and **None** conditions [72]. All other samplers do not support boundary conditions at this stage and are effectively treated as **None**.

3.4 Evidence Review

Extensive studies with BILBY to find best practices for analysis settings for the analysis of compact binary coalescences have been carried out in Ref. [242]. One validation study that in part led us to adopt **DYNESTY** as the default sampling back end was the review of evidence calculations, which we have presented in Ref. [242], and provide more detail on here⁴. Evidence estimates the sampling packages produce may be systematically biased without us noticing. We can test if this is the case if we use a likelihood with analytically known evidence for a set problem. In this case, we use a 15-dimensional multivariate normal distribution as a likelihood that we implemented using the `scipy.stats.multivariate_normal` and is available via

`bilby.core.likelihood.AnalyticalMultidimensionalCovariantGaussian`, as this reflects the number of parameters we encounter in a typical binary system. The entries in the covariance matrix are set such that COV_{ij} ranges between 0.15 and 0.25, which means all parameters show some covariance with each other. We use uniform priors in the range of $[-20, 20]$, and since the likelihood is a normalised distribution with respect to the parameters in this case, the evidence is approximately given by the prior volume V_π

$$\ln Z \approx -\ln V_\pi = -\ln 40^{15} \approx -55.3. \quad (3.5)$$

⁴The code for this study is available at https://git.ligo.org/moritz.huebner/evidence_review/

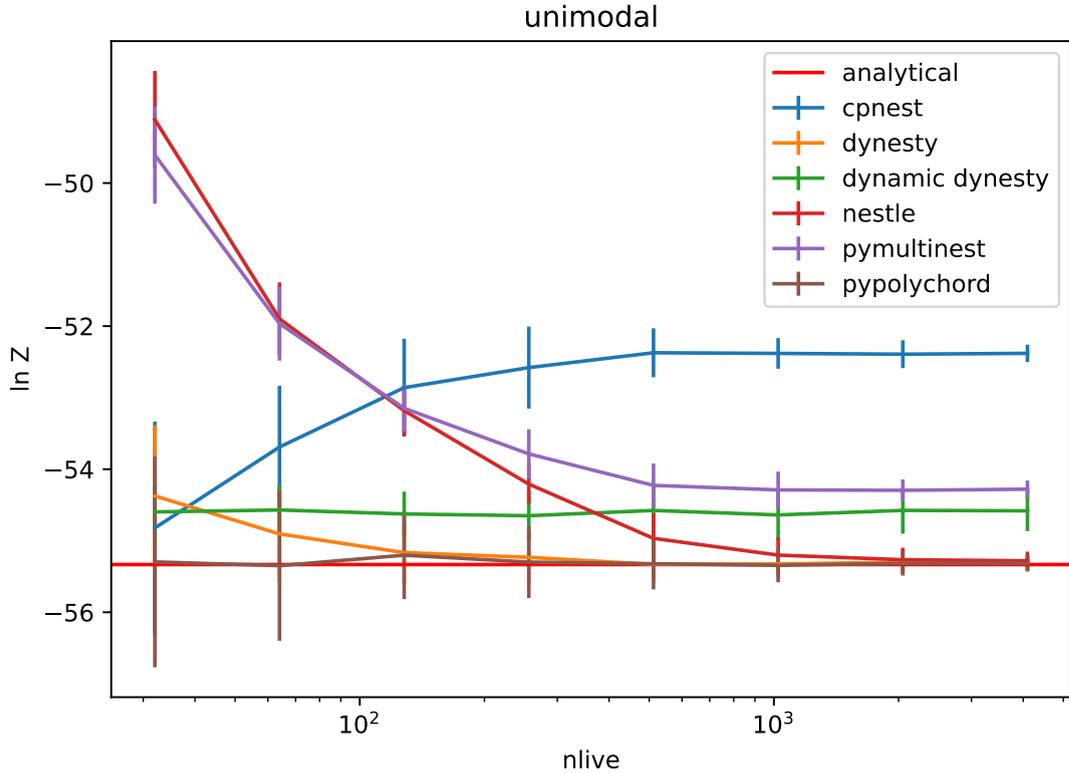


FIGURE 3.1: $\ln Z$ versus the number of live points for the 15-dimensional multivariate Gaussian distribution averaging 100 inference runs. We show that DYNESTY, NESTLE, and PYPOLYCHORD infer the true value without bias given a sufficient number of live points whereas the other samplers show systematic bias.

Nested sampling packages estimate an error on the inferred $\ln Z$. The primary questions thus are if the inferred $\ln Z$ is consistent with the analytically known one, and if the $\ln Z$ error quoted by the nested sampling package is as broad as the one we can estimate by running the same analysis many times.

We show a comparative study using six different back end samplers called with a varying number of live points and the BILBY default settings in Fig. 3.1. We obtained the error bars by taking the standard deviation of the inferred $\ln Z$ over 100 identical runs. We see that DYNESTY [266] and PYPOLYCHORD [133] perform the best in that they obtain the correct evidence unbiased for less than 1024 live points. NESTLE [53] also recovers the evidence without bias for 1024 live points or more. More concerning, the PYMULTINEST [72], CPNEST [296], and the dynamic version of DYNESTY appear to be systematically biased even for high numbers of live points. It may be possible to eliminate this bias by choosing analysis settings more carefully for these samplers, though BILBY mostly remains faithful to the defaults the packages set themselves.

In Fig. 3.2, we investigate DYNESTY’s quoted $\ln Z$ errors more closely. We obtain a percentile-percentile plot by testing what $x\%$ of the time the true evidence is within the x -percentile of the quoted $\ln Z$ credible interval. We find that the quoted $\ln Z$ errors are broadly acceptable if we use more than 1000 live points. The slightly “S”-shaped curve in the percentile-percentile analysis indicates that the evidence errors are not Gaussian

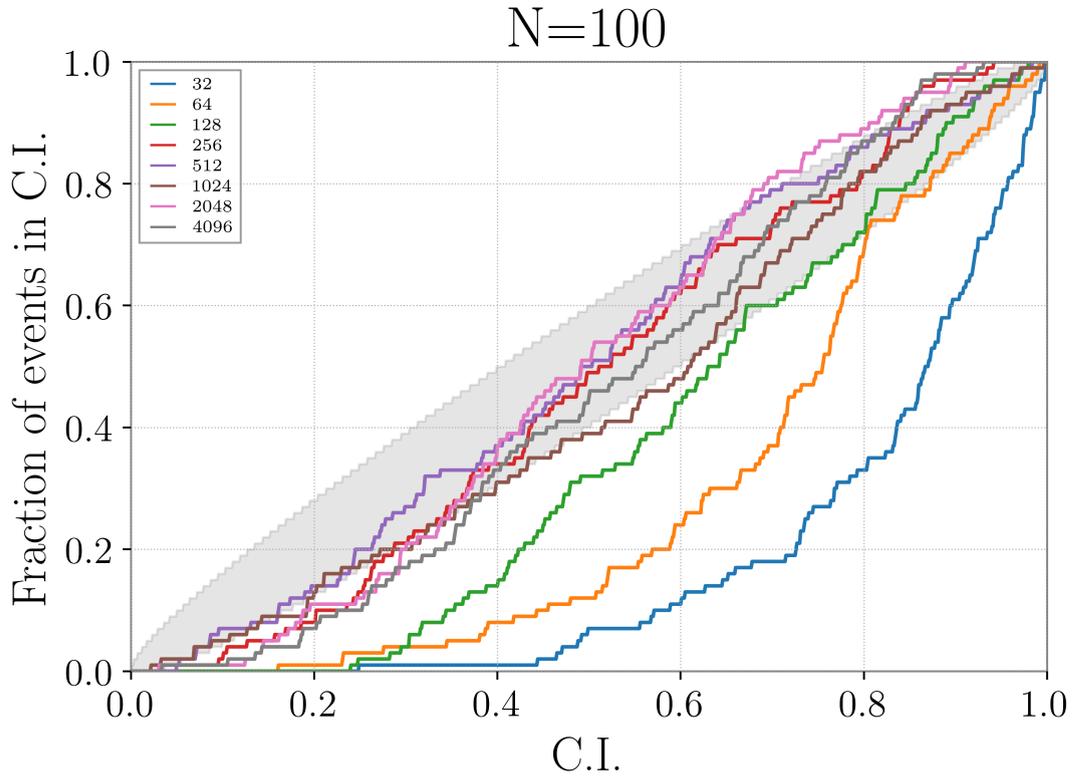


FIGURE 3.2: Illustration of the frequency with which the true evidence is within a given credible interval for a unimodal Gaussian-shaped likelihood when using the DYNESTY sampler. The legend shows how many live points are used to produce the individual curves. For a lower number of live points, systematic errors in the evidence estimation cause significant underestimates of the error. Starting at 1024 live points, the evidence error reasonably reflects the true uncertainty. The grey band shows the 90% confidence interval. This plot was originally shown in [242]

distributed. Instead, the distribution appears to be tighter and produce fewer outliers compared to a Gaussian.

We have more complex posteriors in the analysis of compact binaries. Specifically, the shape of the sky position posterior is often shaped like a ring or a banana, and there are degeneracies in phase and polarisation angle. Thus, we normally combine five runs when analysing compact binaries to ensure good convergence. We tested this by performing the same analysis on a bimodal multivariate Gaussian distribution for which we found very similar overall results.

Chapter 4

Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog

Published as:

M. Hübner, C. Talbot, P. Lasky, E. Thrane 2020, *Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog*, *Phys. Rev. D*, 101, 023011, [arxiv:1911.12496v1](https://arxiv.org/abs/1911.12496v1)

Gravitational-wave memory, a strong-field effect of general relativity, manifests itself as a permanent displacement in spacetime. We develop a Bayesian framework to detect gravitational-wave memory with the Advanced LIGO/Virgo detector network. We apply this algorithm on the ten binary black hole mergers in LIGO/Virgo's first transient gravitational-wave catalog. We find no evidence of memory, which is consistent with expectations. In order to estimate when memory will be detected, we use the best current population estimates to construct a realistic sample of binary black hole observations for LIGO/Virgo at design sensitivity. We show that an ensemble of $\mathcal{O}(2000)$ binary black hole observations can be used to find definitive evidence for gravitational-wave memory. We conclude that memory is likely to be detected in the early A+/Virgo+ era.

4.1 Introduction

Gravitational waves from binary black hole mergers are now observed regularly with LIGO and Virgo [1, 20, 32]. These observations allow us to investigate aspects of general relativity that could not have been studied observationally until now [13, 23, 24, 126]. One such aspect is gravitational-wave memory, a strong-field effect of general relativity that is sourced from the emission of gravitational waves. Memory causes a permanent displacement between freely falling test masses [69, 279, 316].

In general, memory can arise both in the linearized Einstein field equations and in their full non-linear form. Early research focused on the production of linear memory from unbound systems such as supernovae or triple black hole interactions [69]. Non-linear

contributions to memory were originally thought to be negligibly small [79]¹. However, further investigations showed that bound systems such as binary black holes produce significant non-linear memory [79, 279]. Non-linear memory can be interpreted as the component of a gravitational wave that is sourced by the emission of the gravitational wave itself [279]. The amplitude of memory is typically no more than $\mathcal{O}(5\%)$ of the peak oscillatory waveform amplitude in typical binary black hole systems [69]. Detecting gravitational-wave memory from a single merger with current generation detectors is improbable due to the low amplitude of memory [162, 179].

Memory will be detectable from single events with proposed future detectors such as LISA, Cosmic Explorer, and the Einstein Telescope [9, 158, 230]. While detecting memory with LIGO/Virgo [1, 32] directly from a single merger is not possible, it is potentially detectable using an ensemble of mergers [179]. Proposed low-frequency improvements to LIGO could substantially increase the sensitivity to the memory effect [315]. Searches for memory from supermassive black-hole binaries with pulsar timing arrays also have been proposed [86, 192, 290] and carried out (e.g. [35, 41, 304]), although without any detection yet. Future pulsar timing arrays, using data from the Square Kilometer Array [95], may be able to detect memory from supermassive black hole binaries [162].

There are a number of proposed sources of memory besides binaries. These include high-frequency sources outside the LIGO band such as dark matter collapse in stars [173], black hole evaporation [130, 210], or cosmic strings [90]. While such sources are purely conjectural, they would be able to produce memory that is detectable within the LIGO band [199].

Recent work has also shown that there is a redshift enhancement in memory at cosmological distances, which will become relevant for future detectors [55, 56]. Other theoretical work has shown the links between the memory effect, soft gravitons, and asymptotic symmetries in general relativity, which has implications for the black hole information paradox [135, 164, 270]. Measurements of memory with gravitational waves may eventually prove useful studying these phenomena, though, it is not yet clear how.

In this paper, we perform the first search for gravitational-wave memory using the ten binary black hole mergers that LIGO and Virgo observed during their first two observing runs [20]. We find no evidence for memory, consistent with expectations. However, the infrastructure developed here will be used on future observations. We show that, using 1830_{-1100}^{+1730} gravitational-wave observations we will be able to accumulate enough evidence to definitively detect gravitational-wave memory. With the memory signal firmly established, it will then be possible to characterize the properties of memory to see if they are consistent with general relativity.

We structure the remainder of this paper as follows. In Section 4.2, we discuss the methods required to detect memory. In Section 4.3, we apply our algorithm to the first ten binary black hole observations and report the results. In Section 4.4, we use binary black hole population estimates from the first two LIGO/Virgo observing runs to create a realistic sample of future binary black-hole merger observations and calculate the required number to detect memory. Finally, in Section 4.5 we provide an outlook for future developments.

¹Reference [79] shows that nonlinear contributions are not negligible, but states that they were conventionally thought to be so at the time.

4.2 Methods

4.2.1 Signal models

The first major consideration in our analysis is the choice of our signal model. The most precise signal models for binary black hole mergers are numerical-relativity simulations, which solve the Einstein field equations numerically given a set of initial conditions. However, numerical-relativity simulations may take months to carry out even for single mergers. Surrogate models, i.e., models that interpolate between a set of pre-computed waveforms, are hence preferred to create high fidelity waveforms in $\mathcal{O}(1\text{s})$ [58, 293]. Unfortunately, numerical-relativity waveforms and their associated surrogates typically do not include memory since memory is hard to resolve when carrying out numerical-relativity simulations [104].

Recent advances have made it practical to calculate memory directly from the oscillatory part of the waveform [102, 103, 273]. We use the GWMEMORY package [273], which calculates memory from arbitrary oscillatory waveforms, which we then add to the oscillatory component to obtain the full waveform. We compute the memory using IMRPHENOMD [166], a phenomenological model that describes the gravitational wave during the inspiral, merger, and ringdown phase for aligned-spin binary black holes.

One additional consideration was pointed out in [179]. The memory changes sign under a transformation $\phi \rightarrow \phi \pm \pi/4$ and $\psi \rightarrow \psi \pm \pi/4$ ². Here ϕ is the phase at coalescence and ψ is the polarization angle of the waveform. At the same time, this transformation leaves the lower order spin-weighted spherical harmonic modes $(l, m) = (2, \pm 2)$ unaffected, which causes a degeneracy in the (ϕ, ψ) posterior space. If we only use $(l, m) = (2, \pm 2)$ modes, this degeneracy implies the sign of the memory is unknown, which causes the signal to add incoherently (like the fourth root of the number of mergers). Including higher-order modes in the signal model to break this degeneracy is hence advantageous, as they help us to determine the sign of the memory (which causes the signal to grow like the square root of the number of mergers).

4.2.2 Bayesian methods

In order to determine whether a set of gravitational-wave observations contains a memory signature, we perform Bayesian model selection using LIGO/Virgo data. We define our “full” signal model to be the waveform that includes both the oscillatory and memory part of the waveform. We test this model against an “oscillatory only” model (abbreviated “osc”) that only contains the oscillatory part of the waveform.

The Bayes factor describes how much more likely one hypothesis is to have produced the available data compared to another. We define the memory Bayes factor as

$$\text{BF}_{\text{mem}} = \frac{\mathcal{Z}_{\text{full}}}{\mathcal{Z}_{\text{osc}}}, \quad (4.1)$$

where $\mathcal{Z}_{\text{full}}$ and \mathcal{Z}_{osc} are each an evidence (fully-marginalized likelihood) corresponding to our two models. See Ref. [280] for a review of Bayesian statistics in the context

²This statement is incorrect and should actually be $\phi \rightarrow \phi \pm \pi/2$ and $\psi \rightarrow \psi \pm \pi/2$. This chapter is identical to the published version, which also contains the incorrect transformation.

of gravitational-wave astronomy. The total memory Bayes factor $\text{BF}_{\text{mem}}^{\text{tot}}$ can then be accumulated over a series of N gravitational-wave observations,

$$\text{BF}_{\text{mem}}^{\text{tot}} = \prod_{i=1}^N \text{BF}_{\text{mem}}^i. \quad (4.2)$$

Following convention (e.g. [179]), we consider $\ln \text{BF}_{\text{mem}}^{\text{tot}} \geq 8$ a detection.

We calculate both the posterior probability distributions for the model parameters and the evidence using a nested sampling algorithm [106, 256, 266]. In practice, we perform all runs in this paper using the interface to the nested-sampling package DYNESTY [266] within BILBY. Stochastic sampling noise in evidence calculations dominate our results if the difference in evidence between both models is small. We resolve this issue by sampling with the oscillatory-only model and reweighting the posterior samples to the full model to determine the Bayes factor between these two models following the prescription from [220]. A similar analysis has recently been carried out to search for eccentricity in the existing binary catalog [241]. Given a set of n posterior samples θ_k and the observed data d , we calculate the memory Bayes factor BF_{mem} using the oscillatory-only likelihood \mathcal{L}_{osc} and the full likelihood $\mathcal{L}_{\text{full}}$:

$$\text{BF}_{\text{mem}} = \frac{1}{n} \sum_{k=1}^n \frac{\mathcal{L}_{\text{full}}(\theta_k|d)}{\mathcal{L}_{\text{osc}}(\theta_k|d)} \equiv \frac{1}{n} \sum_{k=1}^n w_k. \quad (4.3)$$

We refer to the likelihood ratio w_k as ‘‘weights.’’ This approach is valid if both models have similar posterior distributions, which is true in our case. Since the Bayes factor is now based on the same set of samples for both models, the stochastic sampling noise cancels.

4.2.3 Reweighting study

In order to study the performance of the reweighting technique, we simulate GW150914-like events with different signal strengths in the LIGO/Virgo detector network at design sensitivity with a zero-noise realization using the BILBY software package [45]. We create the oscillatory part of the waveform with IMRPHENOMD and add the memory part of the waveform by using the GWMEMORY package [273].

We use these software injections to compare reweighting to the naive method in which we carry out separate sampling runs with \mathcal{L}_{osc} and $\mathcal{L}_{\text{full}}$. Since this study is purely illustrative, we artificially break the (ϕ, ψ) degeneracy, by restricting the prior space by $\pm\pi/4$ around the injected values for ϕ and ψ . By re-running the sampling algorithm eight times for each distance, we obtain an estimate of the uncertainty in the Bayes factor for both methods. Finally, we also compare the estimates for the Bayes factor with the likelihood ratio at the injected parameter values, as this yields the Bayes factor one would obtain assuming perfect knowledge of the binary parameters. The results are shown in Figure 4.1. The upper panel 4.1 shows that reweighting is generally much better at recovering the Bayes factor whereas separately sampling both models can lead to significant sampling noise. In the lower panel of Figure 4.1 we display the stochastic error of both methods after eight runs. This error ($\Delta \ln \text{BF}$) is defined as the standard error of the sample mean of the eight $\ln \text{BF}$ s we obtained. Notably, the reweighting technique yields a reduction of about a factor 10^2 in stochastic sampling noise. Stochastic sampling

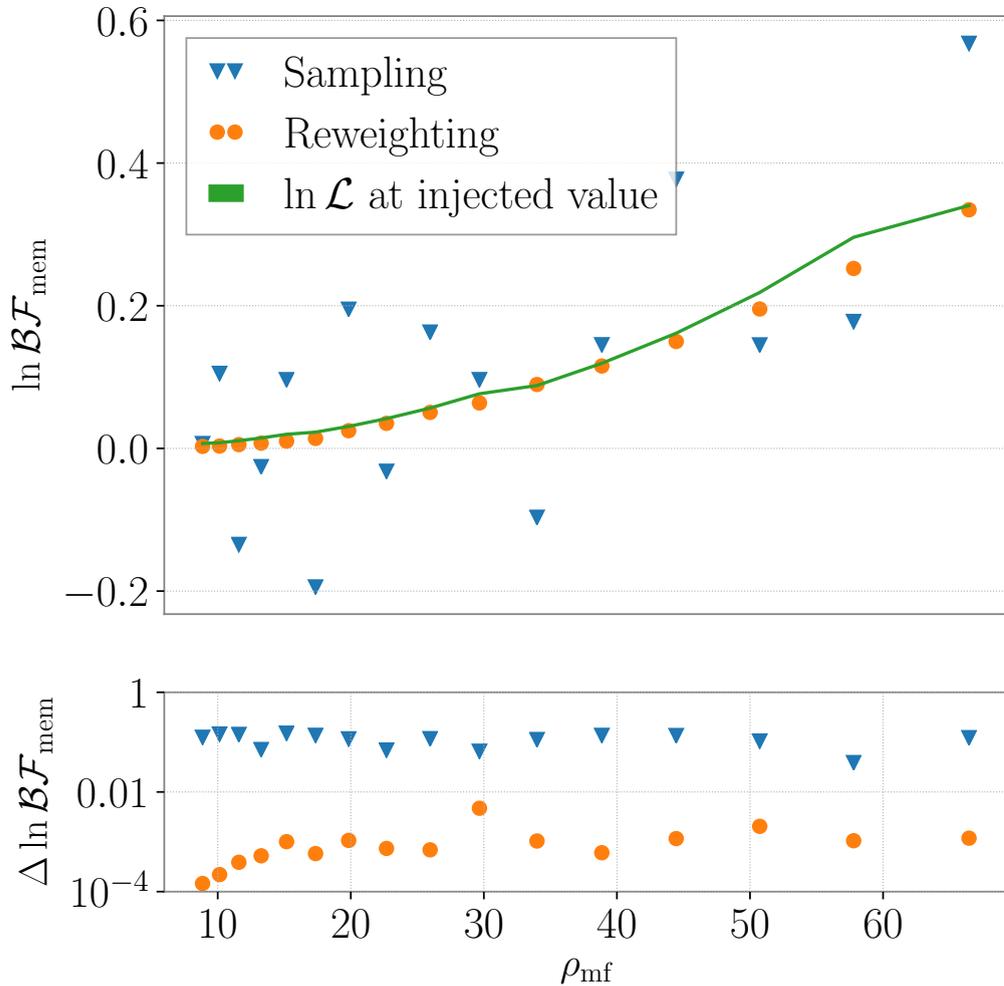


FIGURE 4.1: We compare the precision of calculating the Bayes Factors using different techniques for a set different signal-to-noise ratios ρ_{mf} . In the upper panel, we see that using the naive method of dividing the Bayes Factors from eight separate sampling runs (blue triangles), we see a much wider spread away from the fiducial line of likelihood ratios at the injected value (green curve) compared to the Bayes factors we obtained using reweighting (orange squares). In the lower panel, we see that using the reweighting method yields a statistical error that is about $\mathcal{O}(10^2)$ times smaller.

noise vanishes with computation time t as $\Delta \ln \text{BF} \propto t^{-1/2}$ [78], which implies that the $\sim 10^2$ improvement is equivalent to what would have been achieved by increasing the computation time by a factor of $\sim 10^4$.

4.2.4 Analyzing real events

The analysis of real events mostly follows the prescription in [220]. Initially, we perform inference with the IMRPHENOMD model to obtain a “proposal” posterior distribution. Reweighting these posterior samples first with the NRHYBSUR3DQ8, a surrogate waveform model that includes modes (ℓ, m) up to $(5, 5)$ [293], yields the Bayes factor for

higher-order modes BF_{hom} , since IMRPHENOMD does not contain these modes. Then reweighting with the full NRHYBSUR3DQ8 plus memory model yields the combined higher-order mode plus memory Bayes factor $\text{BF}_{\text{hom+mem}}$. The memory Bayes factor is

$$\text{BF}_{\text{mem}} = \frac{\text{BF}_{\text{hom+mem}}}{\text{BF}_{\text{hom}}} \quad (4.4)$$

A final issue in the analysis is that NRHYBSUR3DQ8 and IMRPHENOMD define the phase ϕ and time at coalescence t_c differently, and there is no analytic way to map posterior samples between those two definitions. Following [220], we map the posterior samples from IMRPHENOMD to NRHYBSUR3DQ8 by maximizing the waveform overlap in terms of ϕ and t_c between both models for each posterior sample. The maximum overlap can be quickly found using common optimization techniques. Furthermore, optimizing over the (ϕ, t_c) plane does not require us to evaluate the expensive NRHYBSUR3DQ8 waveform at every step since these are not intrinsic parameters of the waveform. Instead, we produce the waveform once for each posterior sample and project it into the (ϕ, t_c) space as desired. Results using this method analysing the gravitational-wave transient catalog are presented in Section 4.3.

4.3 GWTC-1 Results

We apply the reweighting technique on posterior samples of the first ten binary black hole mergers from the first two LIGO/Virgo observation runs. The results are summarized in Figure 4.2. The original posterior samples for the proposal run are the same as in [220]. The total $\ln \text{BF}_{\text{mem}}^{\text{tot}} = 3.0 \times 10^{-3}$ provides no significant support for or against the memory hypothesis. However, this small Bayes factor is expected; we explore why in the subsequent section. We see that even the loudest event in the catalog, GW150914 ($\rho_{\text{mf}} \approx 26$), contributes only weak evidence in favour of the memory hypothesis.

4.4 Population study

We construct a simulated population of gravitational-wave events observed by the LIGO/Virgo detector network at design sensitivity so that we can estimate the number of required observations until we reach $\ln \text{BF} \geq 8$. We assume a power-law distribution both in primary mass and in mass ratio as outlined in [22]. The mass distribution parameters are still poorly constrained given the low number of observations in the first two observing runs. From the posterior distributions in [22] we choose parameters that correspond to the points of maximal posterior probability. We choose minimum and maximum black hole masses $m_{\text{min}} = 8 M_{\odot}$ and $m_{\text{max}} = 45 M_{\odot}$ respectively, and use $\alpha = 1.5$ and $\beta = 3$ as spectral indexes for the primary mass and mass ratio distribution, respectively.

We assume an aligned spin prior distribution [176], with a maximal allowed spin magnitude of $a_{\text{max}} = 0.5$. Higher spins are disfavoured observationally [22] and on theoretical grounds [116]. At any rate, we do not expect the spin distribution to greatly affect the memory search because the absolute memory amplitude is mostly driven by the overall signal amplitude, which primarily depends on the masses and the luminosity distance of the source. Spin only has an $\mathcal{O}(10\%)$ effect on the memory of a given binary. The

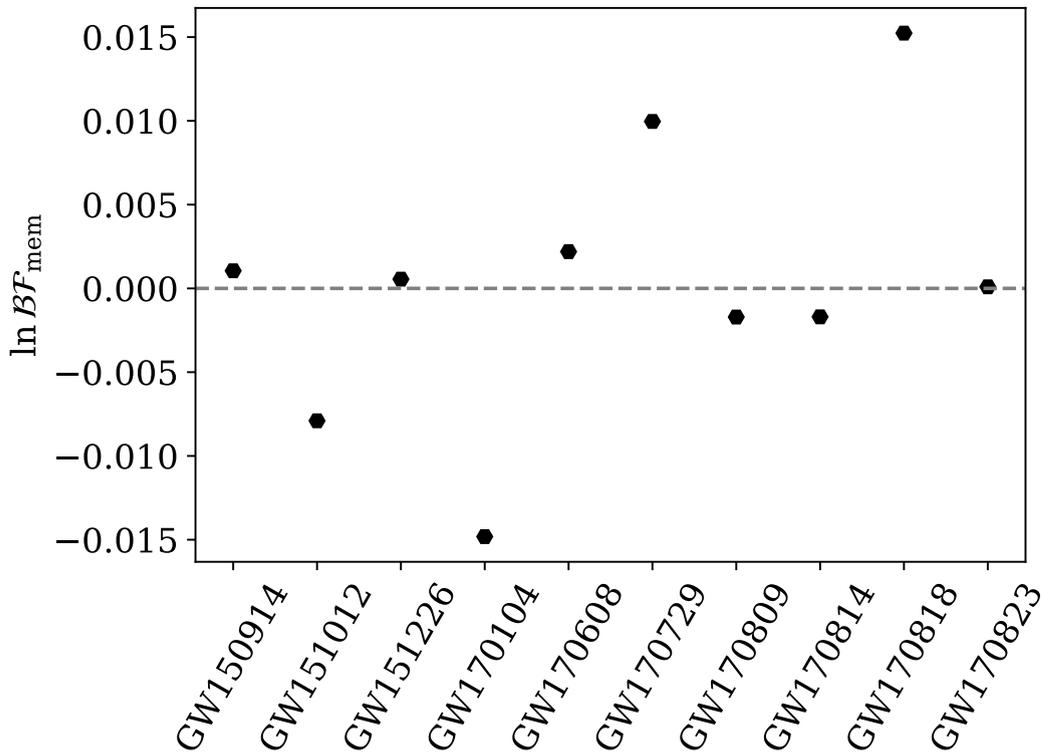


FIGURE 4.2: Memory Bayes factors obtained for the first two LIGO/Virgo observation runs. Overall, there is no significant evidence for or against the memory hypothesis.

remaining extrinsic parameters (inclination, luminosity distance, sky position, time and phase at coalescence, polarisation angle) are chosen using standard priors. We restrict the maximum luminosity distance to 5000 Mpc since more distant events are unlikely to be detected.

We randomly sample parameters from the distributions in intrinsic and extrinsic parameters. However, the LIGO/Virgo detector network will only be able to actually detect a fraction of all occurring binary black hole mergers in the Universe. We therefore only keep events with a matched filter signal-to-noise ratio greater than 12 in the network and/or greater than 8 in any single detector. Otherwise, the event is considered to be undetected.

Following the steps outlined Section 4.2.4, we obtain Bayes factors for each event. In practice, this works reliably up to a matched filter signal-to-noise ratio $\rho_{\text{mf}} \approx 32$, i.e. we recover the injected parameters and obtain an acceptable number of effective samples after reweighting [220]. At higher ρ_{mf} , systematic differences between IMRPHENOMD and NRHYBSUR3DQ8 can cause the inference runs to converge to non-overlapping regions in parameter space. In those cases the reweighting technique using the IMRPHENOMD model becomes invalid if the posterior does not extend over the true value of the injected NRHYBSUR3DQ8 data. We resolve this issue by performing inference with the NRHYBSUR3DQ8 model directly and then reweighting the posterior samples to the NRHYBSUR3DQ8 plus memory model. Since sampling with NRHYBSUR3DQ8 is of far greater computational expense, we do not extend its use to the $\rho_{\text{mf}} < 32$ events, which comprise

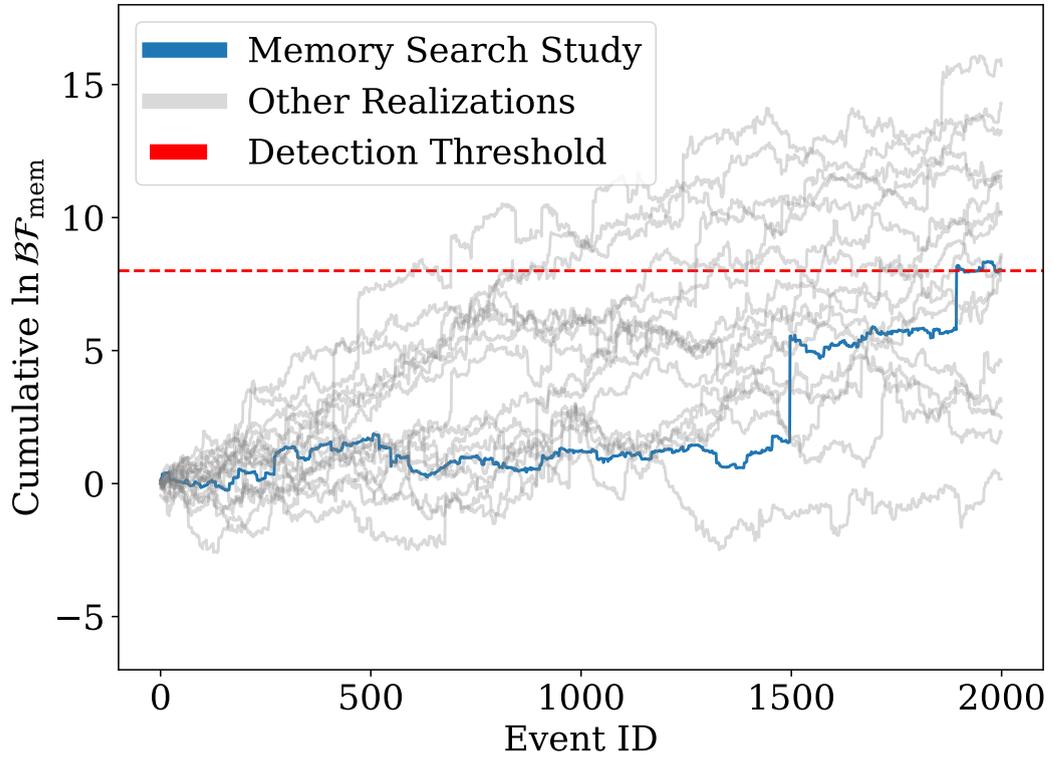


FIGURE 4.3: Cumulative memory Bayes factors obtained for a set of 2000 injections. The blue curve shows the recovered Bayes Factors using nested sampling and reweighting. The dashed red line displays the threshold for detection. The gray lines show a set of realizations using the the likelihood ratio at the injected parameters.

92.5% of all events in our population set. Instead, we use the reweighting technique with IMRPHENOMD proposal distribution for these events.

We perform the analysis on a set of 2000 events and re-run inference until each combined posterior has at least 20 effective samples. By requiring this number of effective samples, we ensure that the samples are reasonably closely converged to the injected value. Otherwise, the weights would wildly diverge and the number of effective samples would hence always be close to unity.

We display the results of our population study in Figure 4.3 (blue curve). The population passes $\ln \text{BF} > 8$ after about 2000 events. We also simulate many more events for which we estimate the Bayes factor by using the likelihood ratio at the injected values (gray curves). Using this much larger population, we estimate the required number of events to reach $\ln \text{BF} \geq 8$ to be 1830^{+1730}_{-1100} at the 90% confidence level. Although this study likely overestimates the Bayes factors since it implicitly assumes that we can always break the (ϕ, ψ) degeneracy, we still consider this to be a good approximation since most support for memory comes from very few events with exceptionally high signal-to-noise ratios.

4.5 Conclusion and Outlook

We have found a combined $\ln \text{BF} = 0.003$ for the existence of memory in the gravitational waves from the ten binary black holes observed by LIGO/Virgo in their first two observing runs. We have shown that we need 1830_{-1100}^{+1730} events to reach $\ln \text{BF} = 8$, which can be considered to be a detection of memory [179]. This is likely to take place in the early days of A+/Virgo+, when observatories will be detecting $\mathcal{O}(10)$ events a day. Adding KAGRA [259] and LIGO-India [284] to the network will further reduce the time until memory is detected. Furthermore, reducing noise at low frequencies has also been shown to substantially decrease the number of detections required [315], reducing the time to detection by a factor of 3. Once memory is observed, it may be possible to use it to probe the nature of black holes and to look for physics beyond general relativity; see, e.g., [314].

We have shown how recent innovations, such as memory waveforms [273], and waveforms with higher-order modes enable us to know the sign of the memory, despite the computational challenges. By introducing likelihood reweighting we reduce the stochastic sampling error by a factor of $\mathcal{O}(10^2)$, which is equivalent in terms of error reduction to an increase in sampling time by $\mathcal{O}(10^4)$. Additionally, we show that by fine-tuning sampling parameters we can obtain confident measures of the Bayes factor within one week of computation time even if we have to use costly waveform models.

4.6 Acknowledgements

We would like to thank Robert Wald for helpful comments. We would also like to thank Ethan Payne for kindly providing his posterior samples and helpful conversations. This work is supported through Australian Research Council (ARC) Centre of Excellence CE170100004. PDL is supported through ARC Future Fellowship FT160100112 and ARC Discovery Project DP180103155. ET is supported through ARC Future Fellowship FT150100281 and CE170100004. This is LIGO Document No. DCC P1900346.

Chapter 5

Memory Remains Undetected: Updates from the Second LIGO/Virgo Gravitational-Wave Transient Catalog

Published as:

M. Hübner, P. Lasky, E. Thrane 2021, *Memory remains undetected: Updates from the second LIGO/Virgo gravitational-wave transient catalog*, [Phys. Rev. D](#), **104**, 023004, [arxiv:2105.02879](#)

The LIGO and Virgo observatories have reported 39 new gravitational-wave detections during the first part of the third observation run, bringing the total to 50. Most of these new detections are consistent with binary black-hole coalescences, making them suitable targets to search for gravitational-wave memory, a non-linear effect of general relativity. We extend a method developed in previous publications to analyse these events to determine a Bayes factor comparing the memory hypothesis to the no-memory hypothesis. Specifically, we calculate Bayes factors using two waveform models with higher-order modes that allow us to analyse events with extreme mass ratios and precessing spins, both of which have not been possible before. Depending on the waveform model we find a combined $\ln \text{BF}_{\text{mem}} = 0.024$ or $\ln \text{BF}_{\text{mem}} = 0.049$ in favour of memory. This result is consistent with recent predictions that indicate $\mathcal{O}(2000)$ binary black-hole detections will be required to confidently establish the presence or absence of memory.

5.1 Introduction

The gravitational-wave memory effect is a non-oscillatory part of any gravitational wave. It can be understood as the part of the gravitational wave that is sourced by previously emitted waves. Memory causes a permanent distortion in spacetime long after the wave has passed [62, 69, 79, 279, 316]. Memory effects are not included in most numerical relativity models and are hence typically not incorporated in gravitational waveforms of compact binary coalescences. This is because memory appears in in the $m = 0$

modes of the waveform which can not currently be resolved with numerical relativity simulations [104, 205, 223].

The slow build-up of memory during compact binary coalescences causes low-frequency contributions to the gravitational-wave signal. Due to the relatively low sensitivity at lower frequencies ($f \lesssim 20$ Hz), the probability of measuring memory in any individual compact binary detection with current generation gravitational-wave observatories is vanishingly small [162, 179]. However, recent advances in modelling and gravitational-wave signal analysis have made it possible to coherently search for the presence of memory in an ensemble of gravitational-wave signals [45, 45, 143, 179, 273]¹. Previous studies have shown that thousands of gravitational-wave detections with LIGO/Virgo [1, 8, 19, 20, 29, 32, 33, 73, 134, 194, 283] operating at their design sensitivity may be required to confidently detect memory, a milestone which is likely to occur during the LIGO A+/Virgo+ era [64, 143].

Since memory arises due to explicitly non-linear effects in general relativity, any measurement of memory can be considered to be a test of the theory in this regime. In this study, we focus on measuring the presence of memory in the gravitational-wave signal itself. Beyond the detection of memory, future studies will then be able to focus on tests of the exact amplitude and shape of the memory part of the wave, which are motivated by modified theories of gravity [122, 141, 248, 314], cosmology [55, 56] and possibly cross-checks with waveform models [167].

5.2 Methods

We follow mostly the description laid out in Ref. [143], but make some adjustments to our waveform models. For a detailed description of our methods see Refs. [143, 179]. Recent advances in waveform modelling have made it more feasible to perform lengthy sampling processes with models that include higher-order modes (see e.g. Refs. [84, 120, 228, 292, 293]). Including higher-order mode effects into the analysis is not just necessary to avoid systematic errors in the waveform models that could affect the analysis, they are also required to break a degeneracy that leaves the sign of the memory ambiguous [179]. This degeneracy arises because memory changes its sign under a simultaneous 90° rotation of the polarisation angle ψ and the phase at coalescence ϕ_c . The same transformation leaves the quadrupolar (leading order) modes unchanged.

For our analysis we use two waveform models. First, the IMRPHENOMXHM waveform model [84, 120, 227], which is an aligned-spin model and includes several of the most dominant higher-order modes. Aligned-spins in this context refer to the black-hole spins being parallel (or antiparallel) to the orbital angular momentum of the binary. IMRPHENOMXHM covers mass ranges of the more extraordinary gravitational-wave observations such as GW190814 [26], which has not been possible for memory analyses until now [143]. Furthermore, IMRPHENOMXHM natively provides the time domain representation of the waveform in the spherical harmonic mode decomposition, which is required to calculate the memory contribution using the GWMEMORY package [273]. Finally, IMRPHENOMXHM is among the most computationally efficient waveform models that contain higher-order modes [120]. Additionally, we perform our analysis with the NRSUR7DQ4 waveform model [292]. While NRSUR7DQ4 only extends to a mass ratio

¹Reference [143] is also Chap. 4 in this thesis

$q = m_2/m_1 = 1/4$, which mainly excludes GW190412, GW190814, and parts of the posterior distributions of a few other events, it does include spin precession effects, which have been shown to exist in the population of binaries [30]. NRSUR7DQ4 has been trained on numerical-relativity simulations to a mass ratio of $q = 1/4$, and compares favourably with alternative waveforms such as SEOBNRv3 in terms of mismatch down to $q = 1/6$, though the mismatch rises quickly past $q < 1/4$ (see Fig. 10 in Ref. [292]). We conservatively restrict the minimum of the prior to $q = 1/4$ when using NRSUR7DQ4. Using both IMRPHENOMXHM and NRSUR7DQ4 provides us with a cross-check on our results and allows us to deploy at least one model on all events ².

Although precession effects are present at the population level, we do not expect these to cause substantially different memory estimates. None of the individual binary black-hole mergers reported so far show significant signs of precession, though, GW190412 and GW190521 exhibit a mild preference for precession [20, 25, 29]. Additionally, we find in injection studies that precessing spins do not meaningfully change the signal-to-noise ratio of the memory part of the waveform.

Recent studies of GW190521 suggest that it may be a highly eccentric ($e \geq 0.2$) binary [28, 123, 240]. Eccentric binaries may become an interesting target to measure memory in future generation detectors [105], however, since they are not yet firmly established, we assume all binaries to be circular for this work.

We measure memory by performing model comparison on each event by calculating a Bayes factor (BF_{mem}) for the presence of memory. For a more in-depth discussion of these methods we refer to our previous work [143, 179]. The memory Bayes factor can be understood to be the fraction of evidences $Z_{\text{osc+mem}}$ and Z_{osc} for the combined oscillatory plus memory waveform and oscillatory-only waveform, respectively

$$\text{BF}_{\text{mem}} = Z_{\text{osc+mem}}/Z_{\text{osc}}. \quad (5.1)$$

In order to calculate the Bayes factors in practice, we perform initial sampling runs with the DYNESTY [266] implementation within BILBY [45, 242] using the IMRPHENOMXHM and NRSUR7DQ4 waveform models without any memory contribution. We use the standard, frequency-domain Gaussian likelihood L for gravitational-wave strain data d given source parameters θ [287, 288, 297],

$$\ln L(d|\theta) = -\frac{1}{2} \sum_i \left\{ \frac{[d_i - \mu_i(\theta)]^2}{\sigma_i^2} + \ln(2\pi\sigma_i^2) \right\}, \quad (5.2)$$

where i refers to the frequency bin index, σ is the noise amplitude spectral density, and μ is the waveform, i.e. IMRPHENOMXHM or NRSUR7DQ4 in our case.

²There are more waveform models such as IMRPHENOMXPHM that both model extreme mass ratios and precession [228], however it currently poses practical issues for memory calculations. The IMRPHENOMXPHM implementation in the LALSUITE package [184] only provides the spherical-harmonic mode decomposition in the frequency domain. Frequency-domain waveforms within LALSUITE, unlike time-domain waveforms, are defined so that the merger is exactly at the beginning/end of the data segment. Calculating memory, which monotonically rises throughout the binary's history, thus creates a discontinuity at the merger time which can not fully be remedied by techniques such as signal windowing. Furthermore, frequency-domain waveforms are cut-off at 20 Hz which leads to an information loss below this threshold when we apply an inverse Fourier transform.

We employ standard priors as laid out in Ref. [45], with the following modifications. We sample using a flat detector-frame total mass prior, and a prior flat in mass ratio which is cut-off to match the limits of our waveforms. For the luminosity distance we use the uniform source-frame prior as described in Ref. [45]. We use isotropic spin priors for the NRSUR7DQ4 runs and the aligned spin prior laid out in Ref. [176] for the IMR-PHENOMXHM runs. We omit the use of calibration envelopes from this analysis since calibration uncertainty does not impact inferences made for binary black hole systems with a signal-to-noise ratio of less than 30 [219, 302].

Non-linear memory is sourced by the energy-flux of the radiated gravitational waves and can thus be expressed as a time-integral of the multipole moments of the oscillatory part of the gravitational wave [104, 279]. Thus we require our waveforms in the time-domain and split in their (l, m) -mode decomposition. In order to calculate the memory contribution in practice we adapt the GWMEMORY PYTHON package [273] to support IMR-PHENOMXHM and NRSUR7DQ4 and make additional modifications to produce waveforms with a consistent length which is required for practical inference tasks.

Next, we combine the oscillatory and memory waveforms in the time-domain by simple addition. We apodize the waveform using a Tukey-window [60] to avoid Gibbs' phenomena before we Fourier transform the combined waveform using standard methods in BILBY [45]. Specifically, we use a shape factor $\alpha = 0.05, 0.025, 0.0125$ as defined in SCIPY [301] for signals with lengths 4 s, 8 s, 16 s. If the apodization is omitted, the Fourier-transformed signal is dominated by an unphysical $1/f$ power law. Using the combined model we calculate the Bayes factor for each event using importance sampling [143, 220, 280]. We obtain the memory Bayes factor by summing the memory weights w_{mem} over all n_{post} posterior samples θ_i

$$\text{BF}_{\text{mem}} = \frac{1}{n_{\text{post}}} \sum_{i=1}^{n_{\text{post}}} w_{\text{mem}}(\theta_i). \quad (5.3)$$

The weights are defined to be the ratios between the likelihoods L of the two hypotheses

$$w_{\text{mem}}(\theta_i) = \frac{L_{\text{mem+osc}}(\theta_i)}{L_{\text{osc}}(\theta_i)}. \quad (5.4)$$

Using importance sampling effectively suppresses stochastic sampling noise in the evidence calculation, making it more suitable than performing inference, e.g. using nested sampling [255], with both hypotheses separately and comparing the resulting evidence values [143].

Since including memory does not add any additional parameters to our problem, we do not need to consider effects due to increased prior volume. We set our prior odds on the presence of memory to be 1, i.e. we give equal weight to memory either being present or not. This means we could also interpret the memory Bayes factor as an odds.

5.3 Results

As discussed in Ref [143], $\ln \text{BF}_{\text{mem}} > 8$ can be considered to be very strong evidence for the presence of gravitational-wave memory. We present the new findings for the unambiguous 36 new binary-black hole observations plus GW190814 additionally to a

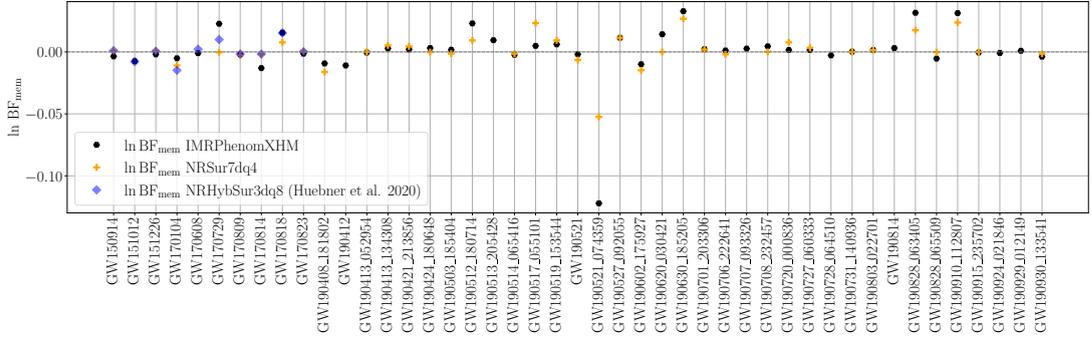


FIGURE 5.1: Memory Bayes factors obtained for the first and second gravitational-wave transient catalog using IMRPHENOMXHM and NRSUR7DQ4 as fiducial waveform models, as well as the results from [143] which used NRHYBSUR3DQ8. Overall, there is no evidence for or against the memory hypothesis.

re-analysis of the first ten binary-black hole observations in Fig. 5.1. We calculate a cumulative $\ln \text{BF}_{\text{mem}} = 0.025$ using IMRPHENOMXHM and $\ln \text{BF}_{\text{mem}} = 0.049$ using NRSUR7DQ4, which indicate that there is no strong evidence favouring or disfavouring the presence of memory in the signals.

We omit the analysis of GW190425 and GW190426_152155, which are most likely to be a binary neutron star and a black hole-neutron star binary system, respectively, for two reasons. First, low mass binaries produce far less memory within the LIGO band than heavier binary black holes, which makes them less useful for memory studies [162]. Second, neither IMRPHENOMXHM nor NRSUR7DQ4 model neutron star physics, and we would thus need to implement and test another waveform model for very marginal benefit. Additionally, GW190426_152155 has a relatively high false alarm rate of 1.4 per year and thus might not be of astrophysical origin. Since they are the other most likely events to not be of astrophysical origin we also exclude GW190719_215514 and GW190909_114149. We are otherwise liberal and include all events that have been reported in the catalogs so far. For the NRSUR7DQ4 runs, we exclude GW151012, GW190412, GW190814, GW190513_205428, GW190707_093326, GW190728_064510, GW190924_021846, and GW190929_012149 since they show substantial posterior support for $q < 1/4$.

We visually verify that our parameter estimates are broadly consistent with what has been reported in [29]. While we occasionally find minor differences, this is likely due to the fact that some of the runs in the catalog [29] were performed using a sampling frequency of 512 Hz, which implies that no physics beyond 256 Hz was included in the analysis. However, most events have some contributions at higher frequencies due to higher-order modes, which cause the posterior to shift. As we have explained earlier, higher-order modes are important to determine the sign of the memory. Thus we use a sampling frequency of 2048 Hz to ensure that these effects are fully represented. While we do not find evidence for or against the presence of memory in any of the observed systems, some of the obtained Bayes factors stand out and some events deserve our attention.

5.3.1 GW190521

GW190521 is the highest mass event that has been reported so far [28]. Even though higher mass systems should in principle create more memory since they radiate off more energy in gravitational waves, their memory is shifted outside of the observable LIGO/Virgo band [162]. This is consistent with our finding that the measured memory Bayes factor indicates GW190521 to be uninformative about memory.

5.3.2 GW190521_074359 and similar events

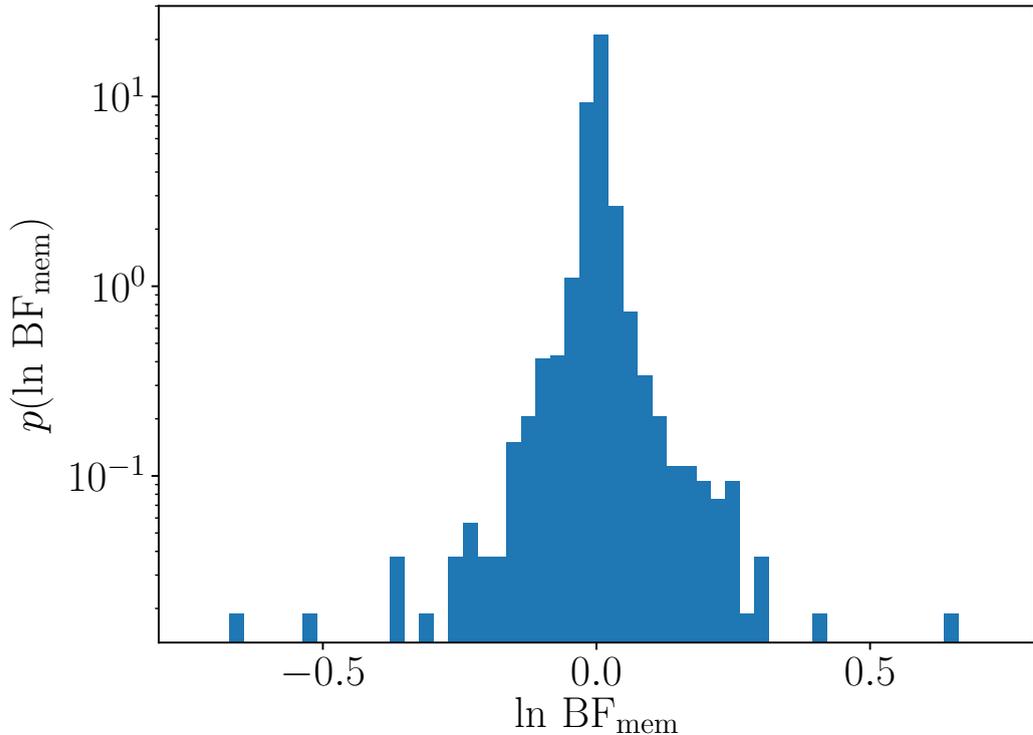


FIGURE 5.2: The distribution of Bayes factors for a simulated population of events from Ref. [143] exhibits wide tails. For visualization, we do not show the two most extreme Bayes factors, which have values of $\ln \text{BF}_{\text{mem}} = 2.4$ and 4.0 .

GW190521_074359 is a near equal mass binary with a total mass of around $75M_{\odot}$ and is also one of the loudest events observed so far ($\text{SNR} \approx 26$) [29]. These properties generally point towards it being a favourable event with which to measure memory. Despite this, we find $\ln \text{BF}_{\text{mem}} = -0.12$ using IMRPHENOMXHM and $\ln \text{BF}_{\text{mem}} = -0.05$ using NRSUR7DQ4, which is the lowest memory Bayes factor but the highest by absolute value for both waveform models. A negative $\ln \text{BF}_{\text{mem}}$ for any individual event is not concerning since they are expected to arise from noise fluctuations. To show this, we re-examine the population study in Ref. [143]. While the population study is not perfectly comparable with the set of the actual measured events as the former is based on a point estimate of the inferred population of GWTC-1 [22], and used different waveform models, it can still provide us with a cross-check to see if the observed distribution of Bayes factors is sensible. We find that out of 2000 simulated events, 28 have $\ln \text{BF}_{\text{mem}} < -0.12$ despite

memory being present, which indicates that our measurement GW190521_074359 is broadly consistent with our expectations. Furthermore, as we show in Fig. 5.2, the distribution of Bayes factors has wide tails, which means that single outlying values are to be expected.

How is it then that the event that looks most likely to contain measurable memory returns the lowest memory Bayes factor in the catalog? Events that are highly unfavourable to measure memory with will return $\ln \text{BF}_{\text{mem}} \approx 0$ as they can only be uninformative. On the other hand, events like GW190521_074359 are more informative about memory, but memory is still weak relative to detector noise. Hence, their signals are also prone to noise fluctuations that may randomly cancel out the memory contributions, which results in a negative log Bayes factor. This is only true for weak memory signals, though. If the signal power of the memory is sufficiently larger than the noise power, it becomes much less likely that we measure a negative log Bayes factor due to noise fluctuations. As Fig. 5.2 demonstrates, the log Bayes factor of GW190521_074359 is still in the regime in which we expect noise fluctuations to be able to change the overall sign of the result.

Other events with high absolute memory log Bayes factors (e.g. GW190630_185205, GW190828_065509, GW190910_112807) follow a similar pattern to GW190521_074359 in that they are relatively high signal-to-noise ratio events, close to equal mass, and have a total mass between $50 - 80M_{\odot}$.

While most of the differences between the Bayes factors from our two waveform models are minor, they do appreciably diverge for GW190521_074359. The difference is unlikely to be due to stochastic sampling noise as this is strongly suppressed in the importance sampling step [143]. In order to understand this difference, we examine the posteriors of both IMRPHENOMXHM and NRSUR7DQ4. In Fig. 5.3, we display the posterior as a contour plot in terms of the obtained memory log weights $\ln w_{\text{mem}}$, and the inclination angle θ_{JN} as well as the luminosity distance d_L . The observed memory strain h_{mem} is highly sensitive to inclination angle θ_{JN} we are viewing the binary at

$$h_{\text{mem}}(\theta_{\text{JN}}) \propto \sin^2 \theta_{\text{JN}} (17 + \cos^2 \theta_{\text{JN}}). \quad (5.5)$$

See e.g. Ref. [103] for a detailed derivation of this relation. Thus, memory is most easily seen edge-on ($\theta_{\text{JN}} = \pi/2$) as opposed to the oscillatory part which is preferably emitted face-on ($\theta_{\text{JN}} = 0, \pi$). In the posteriors, NRSUR7DQ4 has stronger support to be closer to face-on whereas IMRPHENOMXHM shows support for GW190521_074359 being an edge-on binary. This leads to the IMRPHENOMXHM weights obtaining larger absolute log weights. The preference of NRSUR7DQ4 being closer to face-on conversely corresponds to a higher inferred luminosity distance than IMRPHENOMXHM. Overall, these inferred differences in posteriors are expected due to systematic differences in the waveform models, e.g. NRSUR7DQ4 contains precession effects and all modes up to $(\ell, |m|) = (4, 4)$, whereas IMRPHENOMXHM only has aligned spins and modes $(\ell, |m|) = (2, 2), (2, 1), (3, 3), (3, 2), (4, 4)$.

5.3.3 Comparison with GWTC-1 analysis

As part of our analysis we redo the analysis in Ref. [143] for the events of the first two observing runs in which we originally used the hybridized surrogate model NRHYB-SUR3DQ8 [293]. We find that the difference between NRHYBSUR3DQ8 and either IMRPHENOMXHM and NRSUR7DQ4 to be in the same order of magnitude as differences

between IMRPHENOMXHM and NRSUR7DQ4. Again, this is most likely to be due to systematic differences in the waveform models as stochastic sampling noise is suppressed in the importance sampling step. We estimate the effect of stochastic sampling noise by considering the Bayes factor we would obtain from alternative posterior samples. To do this we repeatedly draw half the posterior samples without replacement and calculate the Bayes factor with importance sampling. The distribution of these Bayes factors indicates that stochastic sampling noise at worst causes deviations only on the scale of $\mathcal{O}(\Delta \ln \text{BF}_{\text{mem}}) = 10^{-4}$. We also note that stochastic sampling error scales with the square root of the number of events so even an error of 10^{-4} per event would only scale up to an error of $\Delta(\ln \text{BF}_{\text{mem}}) \approx 0.005$ for the $\mathcal{O}(2000)$ events required to reach $\ln \text{BF}_{\text{mem}} = 8$.

5.4 Conclusion and Outlook

We implement the memory waveforms associated with two waveform models, IMRPHENOMXHM and NRSUR7DQ4, using the memory calculation method laid out in Refs. [104, 273]. We perform Bayesian model comparison to search for the presence of memory in the data. Using the IMRPHENOMXHM (NRSUR7DQ4) model we find a combined $\ln \text{BF}_{\text{mem}} = 0.025$ (0.049) in the first and second gravitational-wave transient catalog. This is consistent with our expectation that $\mathcal{O}(2000)$ events are required to reach $\ln \text{BF}_{\text{mem}} = 8$, which we consider to be very strong evidence [64, 143]. We find that differences in the Bayes factors for each event are likely due to systematic differences in the waveforms and to a lesser extent due to stochastic sampling noise.

We have shown that our approach outlined in our previous paper (Ref. [143]) is scalable up to a large number of events, demonstrating the possibility to coherently search for memory in the future. Given the rapid developments in the waveform community and innovations such as massively parallel Bayesian inference [258], we anticipate that more advanced waveform models can be used for inference in the future. These waveform models may allow us to calculate higher-order and precessing effects at greater mass ratios and thus remove the need for using multiple waveform models to cover all observed events.

5.5 Acknowledgements

We would like to thank Ethan Payne, Cecilio Garcia-Quiros, Colm Talbot, Bernard Whiting, and the referees for helpful discussions. This work is supported through Australian Research Council (ARC) Centre of Excellence CE170100004. PDL is supported through ARC Future Fellowship FT160100112 and ARC Discovery Project DP180103155. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This work was performed on the OzSTAR national facility at Swinburne University of Technology. The OzSTAR program receives funding in part from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government. This is LIGO Document No. DCC P2100125.

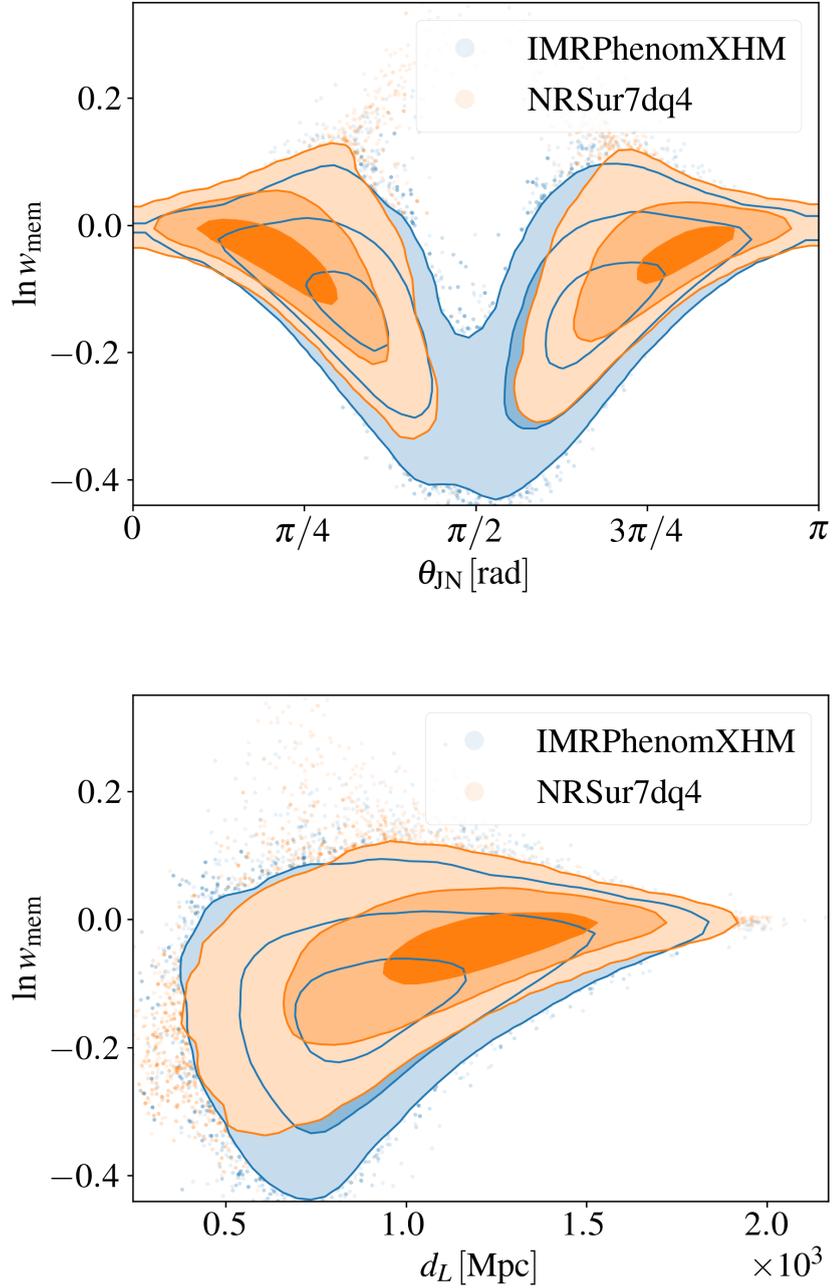


FIGURE 5.3: Two dimensional contour plots of the inferred posterior of GW190521_074359 in terms of inclination and luminosity distance versus the calculated memory weights for IMRPHENOMXHM (blue) and NRSUR7DQ4 (orange). We show inclination (θ_{JN}) in the top and luminosity distance (d_L) in the bottom subfigure. The top subfigure demonstrates why the IMRPHENOMXHM have samples with on average larger weights. NRSUR7DQ4 samples are somewhat further constrained away from $\theta_{\text{JN}} = \pi/2$ which corresponds to an edge-on binary for which observed memory is maximal. The bottom subfigure demonstrates that samples at closer distances have larger absolute log weights on average, corresponding to the fact that closer events are more informative than ones further away.

Chapter 6

Pitfalls of periodograms: The non-stationarity bias in the analysis of quasi-periodic oscillations

Submitted manuscript:

M. Hübner, D. Huppenkothen, P. Lasky, A. Inglis 2021, *Pitfalls of periodograms: The non-stationarity bias in the analysis of quasi-periodic oscillations*, submitted to *Astrophysical Journal Supplement Series*, [arxiv:2108.07418](https://arxiv.org/abs/2108.07418)

Quasi-periodic oscillations (QPOs) are an important key to understand the dynamic behavior of astrophysical objects during transient events like gamma-ray bursts, solar flares, and magnetar flares. Searches for QPOs often use the periodogram of the time series and perform spectral density estimation using a Whittle likelihood function. However, the Whittle likelihood is only valid if the time series is stationary since the frequency bins are otherwise not statistically independent. We show that if time series are non-stationary, the significance of QPOs can be highly overestimated and estimates of the central frequencies and QPO widths can be overconstrained. The effect occurs if the QPO is only present for a fraction of the time series and the noise level is varying throughout the time series. This can occur for example if background noise from before or after the transient is included in the time series or if the low frequency noise profile varies strongly over the time series. We confirm the presence of this bias in previously reported results from solar flare data and show that significance can be highly overstated. Finally, we provide some suggestions that help identify if an analysis is affected by this bias.

6.1 Introduction

Quasi-periodic oscillations (QPOs)¹ are a common astrophysical phenomenon that are regularly observed across a variety of sources. While there is ample discussion about how they emerge in their respective sources, it is worth re-examining existing techniques for their detection and characterization. Observations of QPOs are scant for some objects

¹In solar physics the term “quasi-periodic pulsation” (QPP) is preferred. We apply the term QPO generically throughout for simplicity.

such as in magnetar flares [147–149, 159, 203, 267, 269, 307], or contentious in others like gamma-ray bursts (GRBs) [74, 92, 160, 207, 274], and false detections may lead to unjustified theoretical inferences. Moreover, even if there are ample detections of QPOs like in solar flares [211, 289, 319] or X-ray binaries [156], making accurate inferences about their parameters and significance is essential for understanding the mechanisms that produce them.

There is an array of methods being applied to time series data to find QPOs. The most common tests involve periodograms and are either based on outlier analyses or Bayesian tests [294, 295]. Other methods include wavelet analyses [115, 174, 282] and those based on Gaussian processes [88, 113, 233, 317].

In this paper, we explore the effects that the non-stationary nature of a transient lightcurve, and the potential non-stationarity of a QPO within it, have on detection significance and characterization when methods are used that assume the underlying processes to be stationary. We show that we are likely to overestimate the significance of the quasi-periodic signal under these conditions. This bias arises because neighboring bins in periodograms are only statistically independent for stationary time series. We thus refer to this effect as the “non-stationarity bias”, which to the best of our knowledge has not been reported in the astrophysical literature to date. We show that merely adding photon counting noise before or after a transient leads one to greatly overestimate the significance of a QPO and overconstrain the QPO parameters. More critically, this effect also occurs if only the transient is selected for analysis but the QPO is only present for a fraction of it.

It is our intuition that astrophysicists would rather be conservative and include more of the time series that does not include the QPO, rather than “cherry-picking” the part of the lightcurve that appears most likely to contain it. The erroneous reasoning behind this might be the belief that adding more noise to the signal should not increase the significance of the QPO.

We structure this paper as follows. In Sec. 6.2 we lay out the Bayesian methods and models that we use throughout. Thereafter, in Sec. 6.3, we explain how the non-stationarity bias arises both on a conceptual level and with detailed mathematical arguments. We show in Sec. 6.4 based on simulated data that the non-stationarity bias exists empirically. In Sec. 6.5 we show how the non-stationarity bias affects the analysis of solar flare lightcurves. We conclude in Sec. 6.6 and provide some outlook on how alternative methods can potentially better handle non-stationary data sets.

6.2 Methods

In the following, we recapitulate the Fourier-based methods commonly used to analyze QPOs which are relevant to this study. A general comprehensive overview on the topic of spectral density estimation for time series analysis is given in many popular textbooks such as chapter 7 in Chatfield [75], as well as in Barret & Vaughan [54], van der Klis [286] in the context of astrophysical lightcurves.

6.2.1 Whittle likelihood

A periodogram is an estimate of the power spectral density (PSD) of the signal based on a given time series $x(t)$. We calculate the periodogram $I(f)$ as the absolute square of the discrete Fourier transform of the signal

$$I(f) = \left| \sum_{i=1}^N x(t_i) \exp(-i2\pi f t_i) \right|^2. \quad (6.1)$$

Periodograms are established as a standard method in the search for QPOs in solar flares [137, 154, 155] and QPOs in astrophysical transients elsewhere [e.g. 147–149, 151, 203]. Since individual frequency bins $I(f_j) = I_j$ are calculated by taking the sum of the square of the normal-distributed real and imaginary parts of the Fourier series, it follows that they are χ_2^2 -distributed around the PSD $S(f_j) = S_j$ [75, 309]. This specific case of a χ_2^2 -distribution is identical to an exponential distribution around S_j

$$p(I_j|S_j) = \frac{1}{S_j} \exp(-I_j/S_j), \quad (6.2)$$

where $p(I_j|S_j)$ is the conditional probability to observe the power I_j given an underlying PSD S_j . This relation is generically true for any individual frequency bin in any periodogram. Assuming that all bins are statistically independent from another, we obtain the Whittle likelihood function by taking the product over all $N/2$ frequency bins of a periodogram corresponding to a time series with N points

$$L(I|S) = \prod_{j=1}^{N/2} \frac{1}{S_j} \exp(-I_j/S_j). \quad (6.3)$$

We emphasize here that this is only true in the stationary limit, and this will lead to biased estimates of the PSD in general, as we show in Sec. 6.3. Nevertheless, Eq. 6.3 is the standard likelihood that is used for spectral density estimation.

6.2.2 Models of power spectra

Many astrophysical transients show excess power at low frequencies, and it is often assumed that this can be modeled using a red-noise process (see e.g. Broomhall et al. 70, Huppenkothen et al. 148, Inglis et al. 154, 155, Ingram & Motta 156, Miller et al. 203). One basic noise model that is commonly used is a combination of a red-noise power law with amplitude A and spectral index α , and a white noise amplitude C [154, 155], i.e.

$$\begin{aligned} S^{\text{RW}}(f) &= S^{\text{R}}(f) + S^{\text{W}} \\ &= A f^{-\alpha} + C, \end{aligned} \quad (6.4)$$

where we use the R and W superscripts as short-hand for “red noise” and “white noise” respectively. This spectral shape usually emerges as a combination from the overall structure of the transient, and additional variability on smaller time scales. The shape is thus not due to a stationary process but rather due to a combination of unknown

deterministic and non-stationary stochastic processes adding up to mimic a red noise spectrum [148].

A common way to model a QPO is to add a Gaussian or Lorentzian enhancement to S^{RW} , e.g.

$$\begin{aligned} S^{\text{RWQ}}(f) &= S^{\text{RW}}(f) + S^{\text{Q}}(f) \\ &= S^{\text{RW}}(f) + \frac{B}{\pi\sigma} \frac{\sigma^2}{(f - f_0)^2 + \sigma^2}, \end{aligned} \quad (6.5)$$

where B is the QPO amplitude, f_0 is its mean frequency, and σ is the half-width half-maximum scale parameter. Explicitly modeling the red noise component is important as the QPO is likely to overlap with the red noise dominated part of the PSD. Not considering the red noise dominated part of the PSD may thus lead to false positives.

There are extensions to the noise model, such as the broken or bent power law (B superscript), that can fit more structured red noise and are often a better fit to the data

$$S^{\text{BW}}(f) = Af^{-\alpha_1} \left(1 + \left(\frac{f}{\delta} \right)^{\alpha_2 - \alpha_1} \right) + C, \quad (6.6)$$

where $\alpha_{1,2}$ are the power law indices present before and after the break frequency δ where the power law changes, and we also enforce $\alpha_2 < \alpha_1$ to avoid degeneracies. There are also various different formulations of bent or broken power laws with various degrees of smoothness.

6.2.3 Model selection and parameter estimation

There are several ways to assert the significance of a QPO when using spectral density estimation. A widely used frequentist way to detect QPOs is to use outlier statistics. As a first step, one fits the PSD using a model that does not contain a QPO, e.g. Eq. 6.4 or Eq. 6.6. Thereafter, every frequency bin is checked if its amplitude exceeds a set probability threshold based on the χ^2_2 -distribution in Eq. 6.2. One also has to account for the number of trials, i.e. the number of frequency bins tested by applying a Bonferroni-correction [66]. Additionally, it is possible to rebin the periodogram into a smaller number of bins. For QPOs with widths larger than the width of a single frequency bin, rebinning may be useful because it increases the significance of the QPO relative to the noise continuum.

Alternative to the outlier statistics, one can cast this as a model Bayesian selection problem where we find the preferred model to fit our data, in our case Eqs. 6.4, 6.5, and 6.6. There are multiple ways to quantify model preference in Bayesian statistics. One approach that has been used in solar physics [154, 155, 157] is to calculate the Bayesian Information Criterion (*BIC*)

$$BIC = k \ln(n) - 2 \ln(L_{\text{max}}) \quad (6.7)$$

where k is the number of free parameters (e.g. three for S^{RW} , six for S^{RWQ} , and five for S^{BW}), $n = N/2$ is the number of data points, and L_{max} is the maximum likelihood value. This method is relatively cheap computationally as the calculation of the maximum likelihood can be obtained with relatively few likelihood evaluations using a maximization

algorithm. A lower BIC indicates a better fit to the data, thus the difference in BIC for S^{RW} and S^{RWQ} is a measure of statistical significance of the QPO,

$$\begin{aligned} \Delta BIC &= BIC(S^{RWQ}) - BIC(S^{RW}) \\ &= (k_{RWQ} - k_{RW}) \ln(N/2) - 2 \ln \frac{L_{\max}(I|S^{RWQ})}{L_{\max}(I|S^{RW})}, \end{aligned} \quad (6.8)$$

where $\Delta BIC < 0$ would indicate that S^{RWQ} is preferred and vice versa. We can also perform model selection via calculation of Bayes factors BF . To understand Bayes factors, we start with Bayes' theorem

$$p(\theta|d, S) = \frac{\pi(\theta|S)L(d|\theta, S)}{Z(d|S)}, \quad (6.9)$$

where θ are the model parameters, d are the data, i.e. the periodogram in our case, p is the posterior probability of the parameters, π is the prior probability of the parameters, L is the likelihood of the data given the parameters, and Z is the evidence, or fully marginalized likelihood. All these probabilities are conditioned on a model we want to evaluate, e.g. a PSD S . The Bayes factor comparing two models is the ratio of their evidences. For example, the Bayes factor comparing S^{RWQ} and S^{RW} is

$$BF = \frac{Z(d|S^{RWQ})}{Z(d|S^{RW})}. \quad (6.10)$$

It thereby measures the relative odds of the underlying data to have been produced by either model, though it does not measure if the model itself is a good fit to the data, similar to the BIC . The evidence is calculated by rearranging and integrating Bayes' theorem

$$Z(d|S) = \int \pi(\theta)L(d|\theta, S)d\theta. \quad (6.11)$$

Evaluating this integral is much more computationally challenging than calculating the BIC , but can be achieved thanks to improvements in algorithms such as nested sampling [256], and accessible software implementations such as BILBY [45, 242, 266]. The Bayes factor obtained via evidence calculation is seen as the superior standard for model selection in Bayesian statistics since it involves prior beliefs about the distribution of parameters [308]. The penalty factor $k \ln(n)$ for the BIC also increases with the size of the data set and can thus favor overly simplistic models [124, 308]. Since the BIC relies on a point estimate of the likelihood, it is also a less reliable measure if the likelihood has multiple modes. Additionally, nested sampling provides posterior distributions on the model parameters which are interesting in their own right.

Bayes factors and BIC s work best if the underlying models are truly discrete, but S^{RW} is a special case of both S^{RWQ} and S^{BW} . There are reasonable criticisms on the use of Bayes factors or the BIC for cases like this as S^{RW} is a subset of both S^{RWQ} and S^{BW} , and thus not a truly discrete model [124]. In this case, the Bayes factor depends on prior choices of the parameters specific to S^{RWQ} and S^{BW} . Alternatively, we could use tests that measure if the posterior of the QPO amplitude B is inconsistent with being zero and provides a better fit to the data. Nevertheless, we use Bayes factors as prior choices since they will yield a constant difference in our Bayes factors and will not distort the overall trend when we investigate the non-stationarity bias. We also consider the effects on the ΔBIC s where it is instructive.

For all the Bayes factor calculations and posterior samples in this publication we use DYNesty via the BILBY interface [45, 242, 266].

6.3 Pitfalls of Periodograms

Assume a discrete time series of length T with N data points containing a QPO and noise. The underlying stochastic process is characterized in the frequency domain as a PSD $S(f)$, which can be estimated by calculating a periodogram $I(f)$. We recall that the value in any specific of the $N/2$ frequency bins in the periodogram is proportional to the absolute square of the Fourier amplitude.

In the top panel of Fig. 6.1 we extend a stationary time series by appending zeros on either end until we reach a length xT with xN data points, where we call x the *extension factor*. As we display in the bottom panel, this increases the number of frequency bins in the periodogram to $xN/2$, while decreasing the periodogram to $I(f)/x$ if we do not apply a normalization. The non-stationarity bias occurs here because an increased number of frequency bins boosts the calculated significance of a QPO in terms of the ΔBIC or $\ln BF$ between, e.g. S^{RWQ} and S^{RW} .

The example displayed in Fig. 6.1 is not realistic in practice but serves to illustrate the effect. A more realistic scenario occurs if the stationary time series is instead extended with low amplitude white noise and the QPO is within the red noise dominated part of the PSD. This raises the white noise level relative to red noise and the QPO. Yet it still increases the significance of the QPO, which is determined by its amplitude relative to the dominant red noise level. Such a scenario might occur if we overselect data surrounding a transient

Another scenario is that the overall time series is non-stationary, and instead, the QPO and the overall noise processes vary, or switch on and off, during a selected time window. For example, the QPOs reported in the 2004 SGR1806-20 hyperflare are directly associated with specific rotational phases of the magnetar [147, 159, 203, 267, 268, 307]. Parts of the time series might only contribute low levels of noise and thereby increase the number of bins without suppressing the QPO in the periodogram. These parts, therefore, increase the significance of the QPO without having delivered any actual information about it.

We now turn towards a more mathematical explanation of the non-stationarity bias. Let us consider how we can calculate the SNR of a QPO when red and white noise are present². Given a single frequency bin f_j , this is

$$\rho(f_j) = \frac{S^{\text{Q}}(f_j)}{S^{\text{RW}}(f_j)}, \quad (6.12)$$

where we denote the SNR as ρ . Assuming all frequency bins in a periodogram are statistically independent, the SNR adds in quadrature

$$\rho_{\text{tot}} = \sqrt{\sum_{j=1}^{N/2} \rho(f_j)^2}, \quad (6.13)$$

²The argument here holds for arbitrary noise.

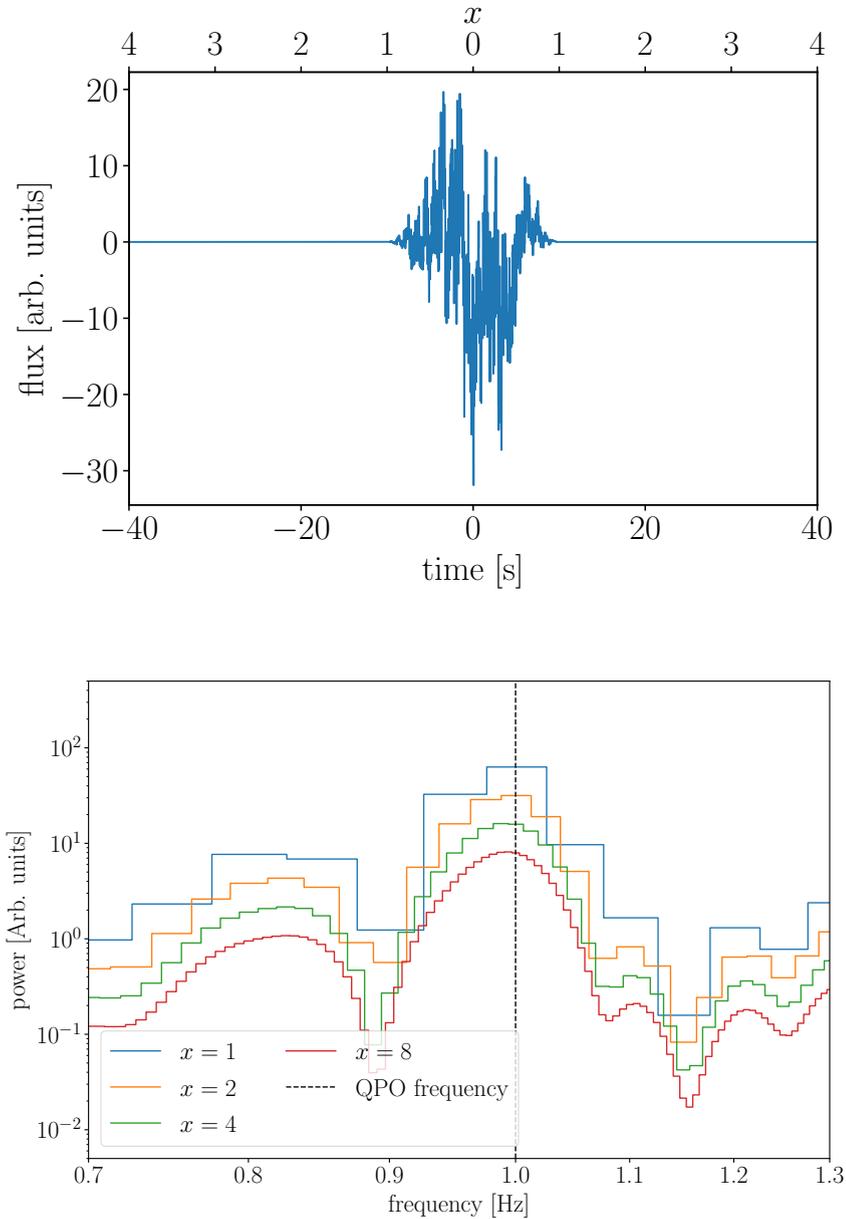


FIGURE 6.1: Effects of extending a time series with zeros. In the top panel, we show a stationary time series between -10 s and 10 s that we multiply with a Hann window to ensure a smooth turn on from zero. We extended this time series with zeros on either end up to a total duration of 200 s. The upper x-axis shows how the extension factor x maps onto the time axis. Taking only the stationary data from -10 s to 10 s corresponds to $x = 1$, i.e. no extension has been applied. The extension factor in this case implies that we select data from $-10x$ s to $10x$ s. In the bottom panel, we show the periodogram given different extension factors, with the blue curve representing the stationary data ($x = 1$). The other colors (orange, green, red) show the effect of extending the time series with zeros on either end. In these curves, neighboring bins are not statistically independent, which manifests itself in the emerging smooth structure of the periodogram. The bins are also decreasing inversely proportional to x since we do not apply a normalization.

where we have summed over the $N/2$ bins in the periodogram and ρ_{tot} denotes the total SNR of the QPO.

6.3.1 QPOs in stationary noise

As the most simple scenario, we consider a stationary time series with N elements made up of a QPO and some arbitrary noise. We denote the SNR of this time series as ρ_{tot}^N . Suppose we look at an extended version of this time series with xN elements, which we create with the same QPO and noise process. We intuitively expect the SNR to increase proportionally to \sqrt{x} and indeed, we calculate

$$\begin{aligned}\rho_{\text{tot}}^{xN} &= \sqrt{\sum_{j=1}^{xN/2} \rho^x(f_j)^2} \\ &= \sqrt{x \sum_{j=1}^{N/2} \rho(f_j)^2} \\ &= \sqrt{x} \rho_{\text{tot}}^N,\end{aligned}\tag{6.14}$$

where we write the SNR at a given frequency in the extended time series as $\rho^x(f_j)$, which is the same as in the original time series. We also use the N and xN superscripts for the SNRs to indicate whether we are looking at the original or extended time series. Note that we have implicitly changed how the f_j are indexed between the steps. Additionally, we assume in the first step that the PSD is only slowly varying between frequency bins and we can therefore approximate the newly added frequency bins by their closest neighbors from the original PSD.

Alternatively, let us consider what happens when we consider a non-stationary extended time series in which the noise is present over xN elements but the QPO is only present for N elements. In that case, S^{RW} is the same as in the original time series whereas S^{Q} gets reduced to S^{Q}/x as we are effectively ‘‘diluting’’ the QPO, similarly to how the periodograms decrease in amplitude with increasing x in the lower panel of Fig. 6.1. We calculate Eq. 6.13 for the extended time series again

$$\begin{aligned}\rho_{\text{tot}}^{xN} &= \sqrt{\sum_{j=1}^{xN/2} \rho^x(f_j)^2} \\ &= \sqrt{x \sum_{j=1}^{N/2} \rho^x(f_j)^2} \\ &= \sqrt{x \sum_{j=1}^{N/2} \rho(f_j)^2 / x^2} \\ &= \frac{1}{\sqrt{x}} \rho_{\text{tot}}^N,\end{aligned}\tag{6.15}$$

where we used the first same step as in Eq. 6.14 and used the fact that the QPO is suppressed by the extension in the second step. These first two scenarios are fairly intuitive, analyzing more of the same stationary process should naturally boost the significance, whereas only adding more noise will dilute the QPO and reduce the significance.

Let us now consider what happens if we extend the time series with zeros on either end, which is illustrated in Fig. 6.1. This scenario is not truly physical, as we always expect some background noise. However, it is an instructive example to understand the effects of non-stationarities. Extending with zeros implies both $S^{\text{RWQ}} \rightarrow S^{\text{RWQ}}/x$ and $S^{\text{RW}} \rightarrow S^{\text{RW}}/x$, so $\rho^x(f) = \rho(f)$ and thus following the same steps as before

$$\rho_{\text{tot}}^{xN} = \sqrt{x}\rho_{\text{tot}}^N, \quad (6.16)$$

which is identical to Eq. 6.14. This is a curious result as extending a time series by zeros clearly should not increase the SNR nor the significance of a signal. The resolution to this seeming paradox is that it is invalid to assume that SNR adds in quadrature as we did in Eq. 6.14. By extending the time series with zeros, we added bins to the periodogram that are not statistically independent.

A further consequence is that the product in the Whittle likelihood becomes invalid. Due to the extended segment length both the PSD and the periodogram need to be divided by a factor of x . Therefore, we can rewrite the likelihood as

$$\begin{aligned} L^{xN}(I|S(f|\theta)) &= \prod_{j=1}^{xN/2} \frac{x}{S_j(\theta)} \exp(-I_j/S_j(\theta)) \\ &\approx x^{xN/2} \left(\prod_{j=1}^{N/2} \frac{1}{S_j(\theta)} \exp(-I_j/S_j(\theta)) \right)^x, \end{aligned} \quad (6.17)$$

where θ is the set of parameters in the PSD and we use the N and xN superscripts on the likelihoods to indicate if we are looking at the original or extended time series. Again, we assume that the PSD is slowly varying with frequency and thus statistically non-independent bins are reasonably approximated by the closest frequency bins obtained from the stationary time series. The factor $x^{xN/2}$ that we have introduced does not matter since it is a constant. Hence we can write

$$\ln L^{xN}(I|S(\theta)) \propto x \ln L^N(I|S(\theta)). \quad (6.18)$$

This means $\ln L^{xN}(I|S(\theta)) \approx x \ln L^N(I|S(\theta))$, i.e. we have steepened the log likelihood function by a factor of x

$$\frac{\partial \ln L^{xN}(I|S(\theta))}{\partial \theta} \approx x \frac{\partial \ln L^N(I|S(\theta))}{\partial \theta}. \quad (6.19)$$

This steepening means that inferred posterior distributions are generally tighter which leads to overconstrained parameter estimates. We calculate how this changes the Bayesian information criterion

$$\begin{aligned} \Delta BIC^{xN} &= (k_{\text{RWQ}} - k_{\text{RW}}) \ln(xN/2) - 2x \ln \frac{L_{\text{max}}^N(I|S^{\text{Q}})}{L_{\text{max}}^N(I|S^{\text{RW}})} \\ &= x \Delta BIC + (k_{\text{RWQ}} - k_{\text{RW}}) (\ln x - (x-1) \ln(N/2)), \end{aligned} \quad (6.20)$$

where $k_{\text{RWQ}} = 6$ and $k_{\text{RW}} = 3$ are the number of free parameters for the respective models, and xN is used as a superscript. Thus, ΔBIC is approximately proportional to x or N since the impact of the $\ln x$ and constant terms are minimal. The impact this has on the Bayes factor is not straight-forward to calculate in the general case. Since likelihood ratios and ΔBIC s can be understood as a related quantity of the $\ln BF$, we expect a similar approximately linear behavior, which we find empirically in Sec. 6.4.

6.3.2 QPOs in non-stationary red noise

Assume a stationary time series with red noise, white noise, and a QPO that sits in the red-noise dominated part of the PSD. If we extend this time series with zeros, the argument in Sec. 6.3.1 holds and the SNR calculated with Eq. 6.13 grows proportional to \sqrt{x} . Extending with white noise instead of zeros will also lead to the SNR growing with \sqrt{x} as long as the QPO remains in the red noise dominated part of the PSD. To show this we define f_{break} as the breaking frequency where white noise and red noise are of equal magnitude

$$f_{\text{break}} = \left(\frac{C}{A}\right)^{-1/\alpha}, \quad (6.21)$$

which means that for $f \ll f_{\text{break}}$ red noise dominates whereas white noise dominates for $f \gg f_{\text{break}}$. If our QPO has $f_0 < f_{\text{break}}$ it mainly competes with red noise in terms of SNR, thus a rising level of white noise will not meaningfully reduce the SNR until $f_0 \approx f_{\text{break}}$. We can write this as

$$\rho_{\text{tot}} \approx \rho_{\text{tot}}(f < f_{\text{break}}), \quad (6.22)$$

where $\rho_{\text{tot}}(f < f_{\text{break}})$ is the total SNR just based on frequencies less than the breaking frequency. If we extend the time series with white noise or zeros by a factor of x , we suppress the red noise term by $1/x$ and shift the breaking frequency

$$f_{\text{break}}^x = \left(\frac{Cx}{A}\right)^{-1/\alpha} = x^{-1/\alpha} f_{\text{break}}, \quad (6.23)$$

where the superscript x indicates that we are looking at the extended time series. We can equate Eq. 6.23 to f_0 and re-arrange to calculate the extension factor x_{break} for which the QPO will be in the white-noise dominated part of the periodogram

$$x_{\text{break}} = \left(\frac{f_{\text{break}}}{f_0}\right)^\alpha = \frac{A}{C} f_0^{-\alpha}. \quad (6.24)$$

Therefore, we are prone to artificially increase the SNR by extending our time series with white noise up to an extension factor of x_{break} . Conversely, if our QPO mainly competes with white noise $f_0 > f_{\text{break}}$, we use the argument in Sec. 6.3.1 to show that we suppress

the SNR with increasing x

$$\begin{aligned}
 \rho_{\text{tot},x} &= \sqrt{\sum_{j=1}^{xN/2} \left(\frac{S^{\text{Q}}(f_j)/x}{S^{\text{W}}} \right)^2} \\
 &\approx \frac{1}{\sqrt{x}} \sqrt{\sum_{j=1}^{N/2} \left(\frac{S^{\text{Q}}(f_j)}{S^{\text{W}}} \right)^2} \\
 &\approx \frac{\rho_{\text{tot}}}{\sqrt{x}},
 \end{aligned} \tag{6.25}$$

where we have used the fact that the PSD is slowly varying with frequency in the first approximation, and that the frequencies where the QPO exceeds the noise continuum are all white-noise dominated in the second approximation.

The most obvious case where this could become a practical problem is when we have to decide where to start and end the segment. Not being aware of this bias may lead us to overselect the data by $\mathcal{O}(10\%)$ and infer a somewhat higher SNR and significance, which by itself is not a big issue. But the issue becomes much worse in practice if the QPO itself is non-stationary and only appears for a part of the transient, which implies that selecting the entire transient may already be overextending the segment in which the QPO is present.

In principle, extending the time series with red noise instead of white noise should not cause the same issue as we would expect the same suppression by $1/\sqrt{x}$ that we have for white noise in Sec. 6.3.1. However, in real transients, the low-frequency noise continuum arises not due to a stationary noise process but due to a combination of deterministic and non-stationary stochastic processes. Different parts of the time series might add up roughly to a single power law or a broken power law, but smaller segments within the transient may have vastly different shapes. For example, the rising and falling edge of the transient may create different power laws, we may have segments that are fairly flat and thus mostly add white noise, and so on. Additionally, deterministic aspects of the time series, if not properly subtracted, are not appropriately modeled with a Whittle likelihood as we show in Sec. 6.4.3. The deterministic parts of the lightcurve can not be safely removed by filtering or smoothing methods as they are prone to create an artificial structure that looks like oscillatory behavior [49]. Clearly, it is not possible to write down a rigorous treatment of this rather large class of possible constellations in which non-stationarity bias could occur.

We note here that using the outlier analysis is not prone to false detections due to the non-stationarity bias since we are testing neighboring bins individually instead of combining SNR from neighboring bins. The opposite is the case, since we add more bins by having extended the time series, it is harder to exceed the Bonferroni-corrected significance level. However, in practice periodograms are sometimes rebinned to add up neighboring bins [e.g. 149]. The motivation behind this is to combine the SNR from neighboring bins, making it easier to detect QPOs. Thus, rebinned outlier analyses are prone to the same non-stationarity bias that affects the Bayesian methods as well.

6.4 Simulated data

To demonstrate the impact of non-stationarities empirically, we analyze simulated data using a variety of setups. For all studies on simulated data, we assume white noise levels to be constant along with the time series. We discuss the impact of the Poissonian nature of the photon counting process in Appendix 6.9.

6.4.1 Setup

To produce the time domain data, we use the algorithm by Timmer & König [281], as it is implemented in the BILBY software package. Concretely, this works by creating a white noise frequency series by randomly drawing both amplitudes and phases from a normal distribution, coloring the noise by multiplying it with the square root of the PSD and then applying an inverse Fourier transform to obtain a stationary time series. In general, we create a continuous, non-stationary time series by using a combination of addition, concatenation, and convolution with window functions. For example, we start by creating a stationary time series using the QPO model and red noise $S^{\text{RQ}} = S^{\text{R}} + S^{\text{Q}}$. We apply a Hann window to ensure a smooth turn on from zero. This helps us to avoid Fourier artifacts such as side-lobes in the periodogram which can appear if discontinuities are present. Next, we create a much longer time series containing just white noise using the Timmer & König [281] method, and add the S^{RQ} time series in the center. We are also interested in what happens when we extend the time series with zeros instead of white noise. To do this we mask the parts containing just white noise and set those values to zero. In all cases, we center the part of the time series in which the QPO is present. This is to prevent the Hann window, which is applied when calculating the periodogram, from diminishing the amplitude of the red noise and QPO components. Note that this is separate from the Hann window we use to ensure the smooth turn-on from zero.

For all simulated data in this section, we assume a constant white noise level throughout the time series and place the QPO in the center of the time series between -10 s and 10 s. We use a sampling frequency of 40 Hz and extend the time series up to 200 s ($x = 10$) or 400 s ($x = 20$) depending on the scenario. The specific parameters for each simulation are listed in the tables in App. 6.8. We discuss the effect of non-stationary white noise in App. 6.9. All parameters and priors are listed in Tabs. 6.2 and 6.3.

6.4.2 Non-stationary QPO in white noise

We start by considering a combination of a QPO and white noise where the setup is effectively identical to what we describe in 6.4.1, but we use S^{Q} instead of S^{RQ} to omit red noise. We consider both scenarios discussed in Sec. 6.3.1 where we either extend the time series with zeros or with white noise. In terms of the parameter inference, we deviate from the standard practice of using $S^{\text{RWQ}}(f)$ as there is no red noise component. Instead we compare S^{WQ} against the white noise hypothesis $S^{\text{W}} = C$.

First, we consider what happens when we extend the time series with white noise. In principle, the results should follow our discussion in Sec. 6.3.1 for QPOs in the white-noise-dominated part of the periodogram, i.e. the SNR should decrease with $1/\sqrt{x}$ and

the $\ln BF$ should decrease roughly inversely to x . The results, which we display in the top panel of Fig. 6.2, are in agreement with our expectations.

We now consider the case where we extend the time series with zeros, similar to what we did to produce Fig. 6.1. As we discussed in Sec. 6.3, we expect the SNR to increase with \sqrt{x} and the $\ln BF$ roughly linearly. As we show in the bottom panel of Fig. 6.2, this is almost perfectly the case for the $\ln BF$ and is also qualitatively true for the SNR.

6.4.3 Non-stationary QPO in a simple transient

We investigate the effect of analyzing a non-stationary QPO that appears on top of a deterministic transient flare shape in the presence of white noise. The time series is displayed in the top panel of Fig. 6.3. The flare shape mimics a red-noise continuum that overlaps with the QPO in the periodogram (bottom panel of Fig. 6.3), though it is not well described by a Whittle likelihood in the periodogram since it is not a stochastic process. We see that there is conspicuously little fluctuation at low frequencies in Fig. 6.3. For the flare shape we use a continuous exponential rise and fall described by

$$f(t) = A_{\text{flare}} \begin{cases} \exp\left(\frac{t-t_0}{\tau_r}\right) & \text{if } t < t_0 \\ \exp\left(-\frac{t-t_0}{\tau_f}\right) & \text{if } t \geq t_0 \end{cases} \quad (6.26)$$

where A_{flare} is the amplitude, t_0 is the peak time, and τ_r and τ_f are the rise and fall timescales, respectively. This model is similar to the FRED model which is popular to fit time series of gamma-ray bursts [214]. The advantage of using an exponential shape for this simulation is that it produces a power law in the periodogram which matches our red noise model. Since the flare is deterministic, low frequencies in the periodogram are not well described by a Whittle likelihood. To create this time series, we repeat the steps in Sec. 6.4.2, though we start with different parameters listed in Tab. 6.3, and add the deterministic flare shape in the last step.

We perform the same test as in the previous section of extending the time series starting from the 20s segment in which the QPO is present (inset in the top panel of Fig. 6.3). It is less intuitive why non-stationarity bias may occur and how it will manifest in this scenario. Extending beyond the peak of the flare might reduce the apparent significance as it will provide the strongest contribution to the low-frequency noise continuum. At the same time, we extend towards the tail of the flare, which contributes little more than white noise. As we can see in Fig. 6.4, the SNR and $\ln BF$ do indeed increase sharply until $x = 3$, which corresponds to the point where we extend the selected window to the peak of the transient. Past this point low-frequency contributions from the rising edge of the transient obscure the QPO in the periodogram and quickly lead to the QPO becoming undetectable for $x > 7$. We note that using a different set of parameters to create the data can easily create situations where the $\ln BF$ continues to increase past $x = 3$.

6.4.4 Non-stationary QPO in non-stationary red noise

In this scenario, we create a time series using a combination of red noise, white noise, and a QPO using the S^{RWQ} model. We extend this time series with either white noise

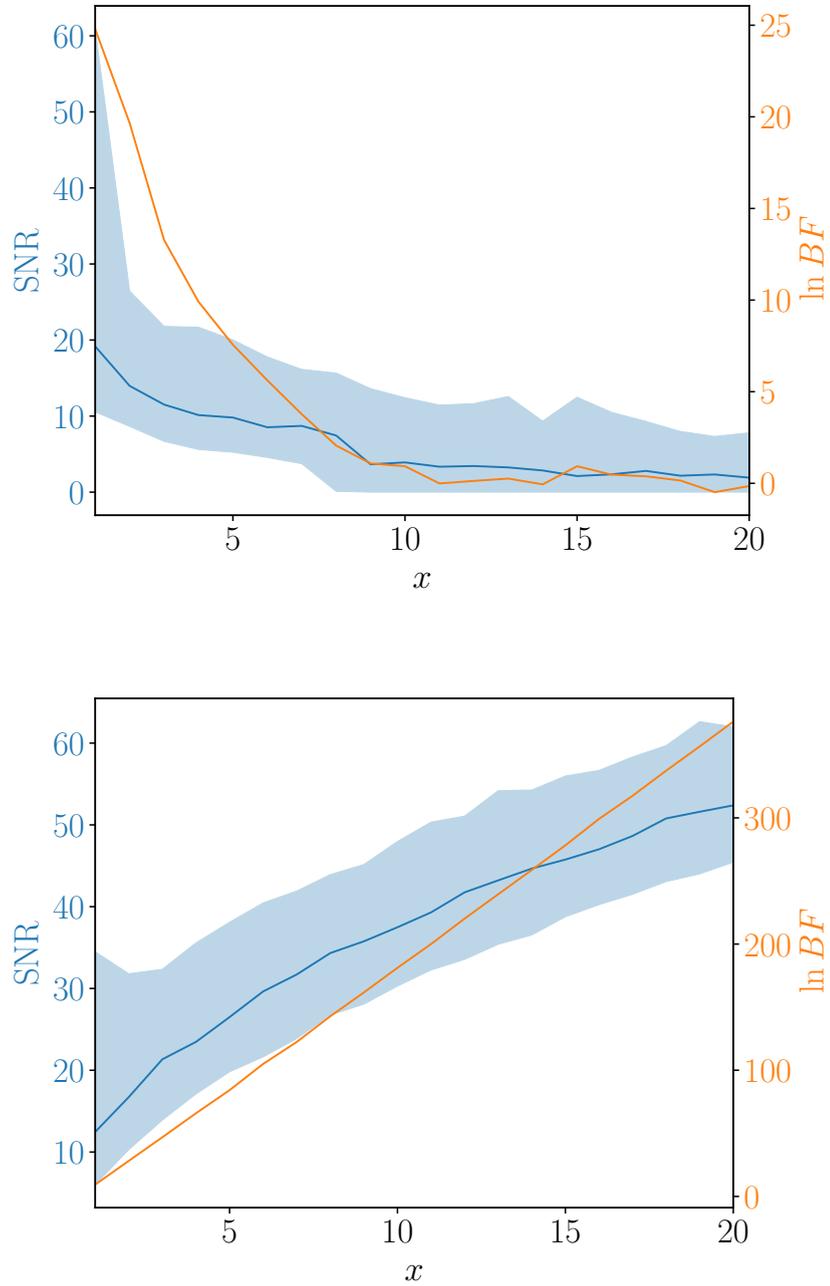


FIGURE 6.2: SNR (blue) and $\ln BF$ (orange) vs the extension factor for a simulated QPO plus white noise signal that is extended with white noise (top) or zeros (bottom). The solid blue line displays the SNR of the maximum likelihood point and the shaded blue region displays the SNR's 90% credible interval. In the top panel both SNR and $\ln BF$ decrease with increasing x which corresponds to the QPO vanishing in noise. In the bottom panel the $\ln BF$ increases almost perfectly linearly due to the increased number of frequency bins that make up the QPO, whereas the SNR increases roughly with the square root.

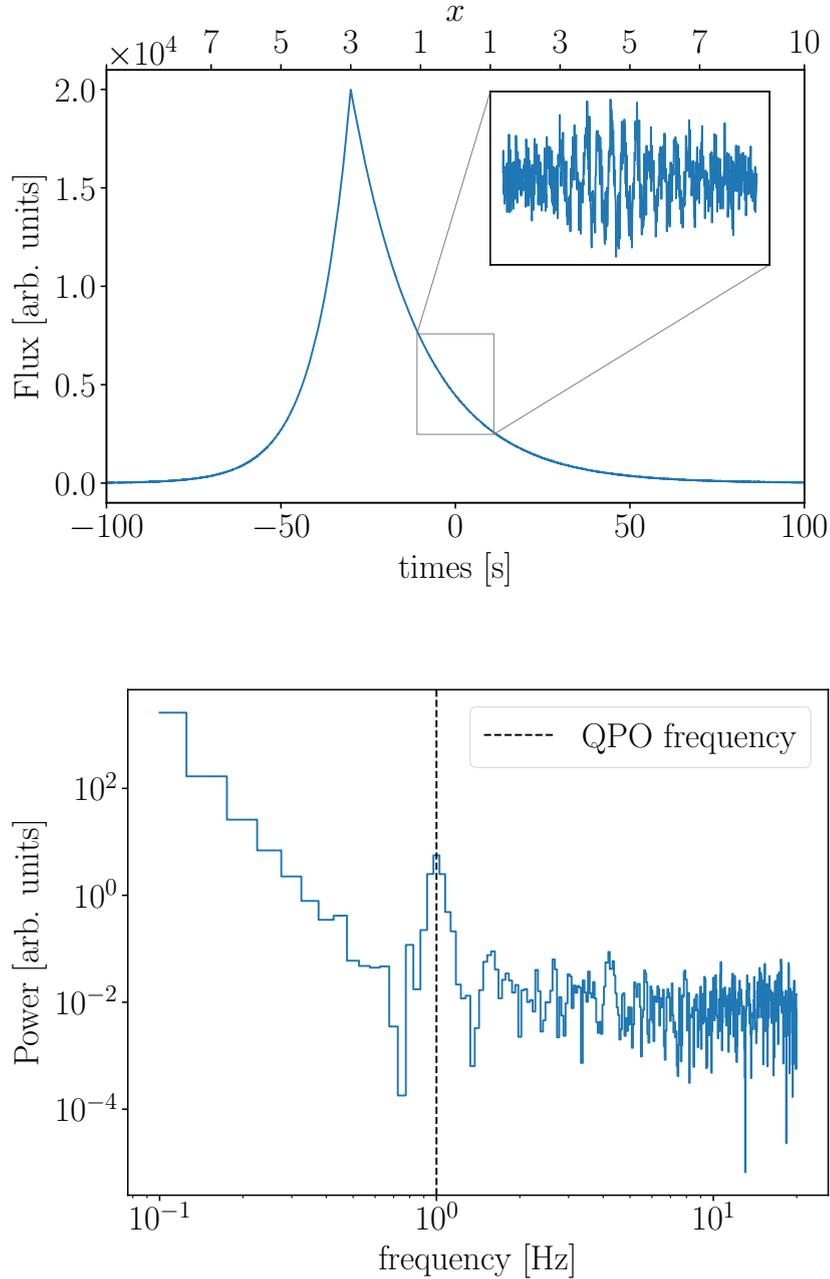


FIGURE 6.3: Simulated data of a non-stationary QPO in a deterministic transient flare shape in the presence of white noise. The time series (top) shows the dominant transient and the QPO that is present for 20s on the tail of the flare (inset). We remove the deterministic trend of the flare in the inset to make the QPO easily visible. The periodogram (bottom) shows a visible QPO at 1 Hz and displays a low-frequency continuum that arises due to the transient flare shape. The periodogram corresponds only to the segment of the lightcurve that contains the QPO. Note that there is much less variability than we would expect from the χ^2_2 -distribution for the bins at low frequencies since the noise continuum is not due to a stationary stochastic process but due to a deterministic process.

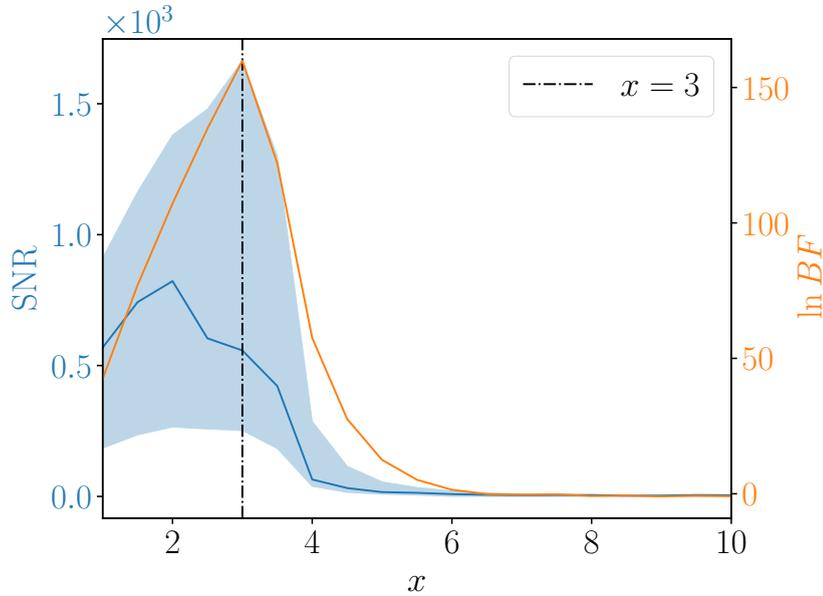


FIGURE 6.4: SNR (blue) and $\ln BF$ (orange) vs the extension factor for the QPO shown in Fig. 6.3. The solid blue line displays the SNR of the maximum likelihood point and the shaded blue region displays the SNR’s 90% credible interval. Both the $\ln BF$ and the upper edge of the SNR’s credible interval peak at $x = 3$ (black dashed vertical line), which corresponds to the point where the selected data segment exceeds past the peak of the flare. The strong differences in $\ln BF$ and SNR across extension factors shows that differences data selection can greatly influence the results.

or zeros. Fig. 6.5 displays the time series for $|t| < 30$ s (top panel) and the periodogram as obtained for $|t| < 10$ s.

In the case of extending with zeros, we expect the same relation as in Sec. 6.4.2, which we recover in the bottom panel of Fig. 6.6. On the other hand, when we extend with white noise the SNR first rises quickly until the extension factor reaches around $x_{\text{break}} = 2$ (see Eq.6.24) and then turns towards a slow descent proportional to $1/\sqrt{x}$. We note that the behavior in the case of $x < x_{\text{break}}$ for white noise is not well captured, but greater extension factors would be less realistic and the simulated data set aims to qualitatively show all effects.

We use models S^{RW} and S^{RWQ} as described earlier to calculate posterior distributions, Bayes factors, and ΔBIC s. As we show in the bottom panel of Fig. 6.6, the $\ln BF$ increases almost perfectly linearly when we extend with zeros. In the top panel, we see that if we extend with white noise the $\ln BF$ also increases with x even past x_{break} , though turns around eventually. The $\ln BF$ may be increasing past x_{break} if the QPO feature in the periodogram is better fit by the model for higher x compared to the feature’s shape at x_{break} .

We have shown that overall we can see a non-stationarity bias if white noise remains stationary, but red noise vanishes for part of the lightcurve. This is in principle different from the effect in Sec. 6.4.3, as the red noise is due to a stochastic rather than a deterministic process.

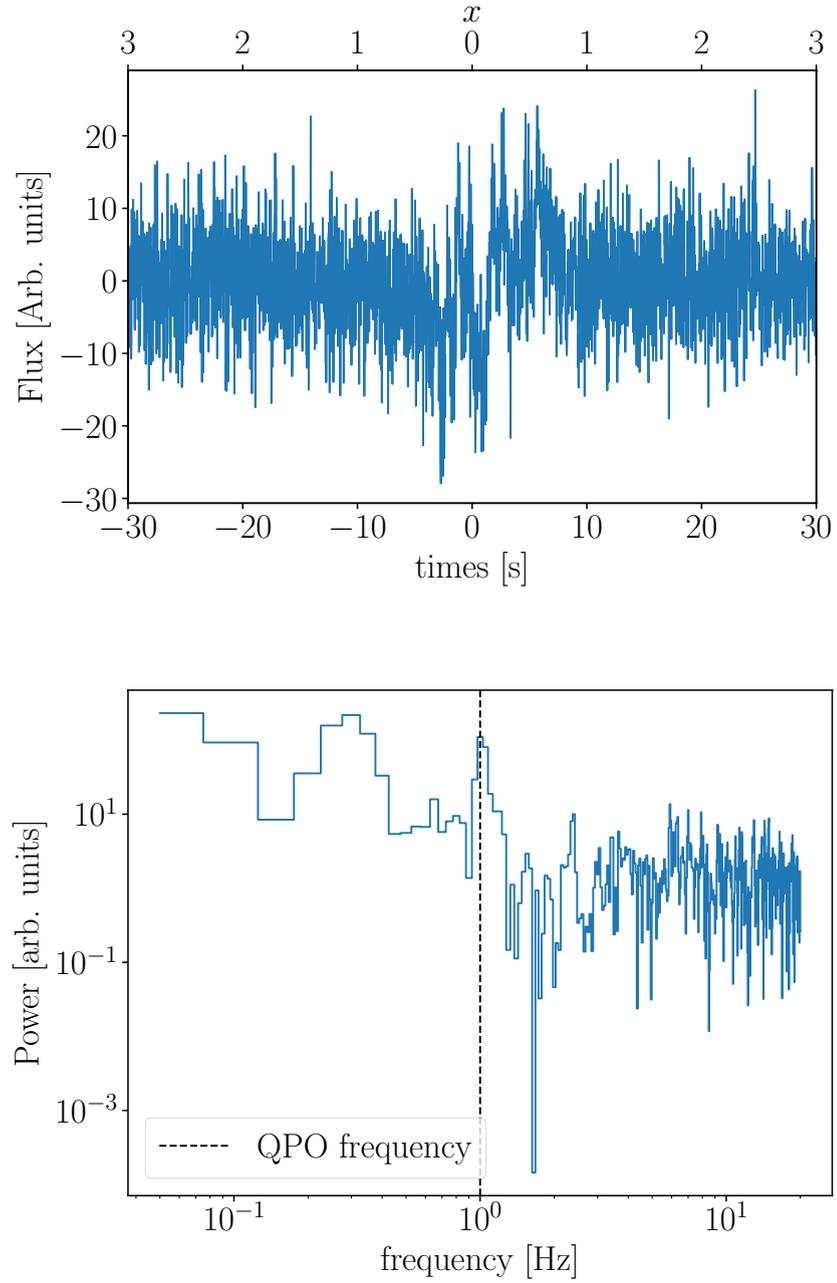


FIGURE 6.5: Simulated data of a non-stationary QPO and red noise in surrounding stationary white noise. In the time series (top) we show how QPO and red noise are transitioning smoothly into the surrounding part of the time series that is pure white noise ($|t| > 10$ s) due to the Hann window that we have applied. In the periodogram (bottom) for $|t| \leq 10$ s the QPO is clearly visible by eye.

6.4.5 Tests of stationarity

One noticeable property of the bias we have found is that it passes regular tests employed to detect fitting validity. Nita et al. [212] demonstrates a χ^2 -like test for exponentially distributed data

$$\chi_\nu^2 = \frac{1}{\nu} \sum_{j=1}^{N/2} \left(1 - \frac{I_j}{\hat{S}_j} \right)^2, \quad (6.27)$$

where ν are the number of degrees freedom and \hat{S} is the maximum likelihood PSD. Analogously to the regular χ^2 -test, $\chi_\nu^2 \approx 1$ indicates an appropriate fit to the data, whereas $\chi_\nu^2 > 1$ indicates underfitting, and $\chi_\nu^2 < 1$ overfitting. Poor fits should in general occur with non-stationary behavior in the red noise or QPO dominated part of the PSD. That is because neighboring bins will be covariant and a fit may either align more closely or further away than should be possible for independent bins, though $\chi^2 \approx 1$ may still randomly occur. However, in practice, non-stationarities will affect the low-frequency part of the PSD while high frequencies are dominated by white noise. Since frequency bins are linearly spaced, the vast majority of all bins are almost always fitted well in terms of the χ^2 -statistic if the periodogram reaches a white noise floor at high frequencies. As we show in Fig. 6.7, the χ^2 -value for the entire PSD is very close to the χ^2 -value for all the frequencies above the QPO frequency plus twice its width. For comparison, setting $f_0 = 1$ Hz with a 40 Hz sampling frequency, as it is the case in the bottom panel of Fig. 6.7, about 95% of all frequency bins are greater than f_0 . Thus even if we overfit locally around the QPO, it is unlikely that this greatly impacts the overall χ^2 -value.

We propose a local χ^2 -test, where we calculate the χ^2 -value for just the frequencies within 2σ around the QPO frequency. We show this test in Fig. 6.7 for data sets used in Secs. 6.4.2 (top) and 6.4.4 (bottom). We note in the top panel that we obtain a reasonable $\chi^2 \approx 0.92$ for the local test if we just look at the stationary time series at $x = 1$, and obtain $\chi^2 < 0.6$ for large x . This method has the downside that it will not work if there are only very few bins available as is the case for $x \lesssim 6$ in the bottom panel of Fig. 6.7. We need at least seven data points to compensate for the six degrees of freedom in the S^{RWQ} model. If more bins are available, this test can reliably flag whether the QPO has been overfitted or underfitted. This test is required but not sufficient to show that non-stationarities are not impacting the analysis. In general, we may find $\chi^2 \approx 1$ even for QPOs that are strongly affected by non-stationarity bias. For example, a time series may in principle contain two QPOs at similar frequencies and non-stationarity bias. In that instance, the overfitting due to covariance of neighboring bins may be compensated by underfitting of the underlying multiple QPO shapes and yield $\chi^2 \approx 1$ for the local test.

6.5 Solar flare data

QPOs are a regularly reported phenomenon that have been observed in solar flares over a wide range of wavelengths for decades, from radio waves to EUV as well as soft and hard X-rays (see Nakariakov & Melnikov [211], Van Doorselaere et al. [289] for recent reviews). In this domain, they are usually referred to as quasi-periodic pulsations (QPPs). Due to their ubiquity, progress has been undertaken to automate their analysis using

Whittle likelihoods, e.g. with the *Automated Flare Inference of Oscillations* (AFINO) method [137, 154, 155]. An extensive catalog of AFINO analyses is available online³. AFINO defines models similar to S^{RW} , S^{RWQ} , and S^{BW} , although it uses a Gaussian instead of a Lorentzian pulse as the QPO component, and a slightly different broken power-law model. AFINO uses a computationally efficient SCIPY fitting routine to determine the maximum likelihood values and calculate the ΔBIC with Eq. 6.8, though it does not calculate Bayes factors or posterior distributions. The QPO model is “strongly favored” if $\Delta BIC < -10$, relative both to the red noise and broken power-law model. This ensures a small number of false-positive values, as has been demonstrated on simulated data in Broomhall et al. [70]. Additionally, AFINO performs the χ^2 -test for exponentially distributed data using Eq. 6.27. AFINO flags the results if the implied p-value from the χ^2_ν test is below or above a set threshold.

For a comparative study on the impact of non-stationarity bias, it is instructive to consider some of the most significant QPOs. This is because these QPOs are visually identifiable from the lightcurve, which guides us in their analysis. We look at a solar flare observed by the X-Ray Sensor (XRS) onboard the GOES-15 satellite on May 12, 2013, at 20:17 UT for which AFINO reports the highest significance for a QPO with $\Delta BIC = -451.6$ relative to red noise and $\Delta BIC = -278.2$ relative to the broken power law. This flare was of GOES magnitude M1.9 and originated from NOAA active region 11748, which was located on the East limb of the Sun at the time. AFINO finds a high quality fit for the QPO model with $\chi^2 = 1.06$ but flags the fits with the broken power law and red noise due to their high χ^2 values, likely due to the very pronounced QPO present at $P = 12.6$ s. We show the 46 minutes long GOES x-ray lightcurve we analyze in Fig. 6.8. The QPO is visible by eye on the tail end of the distribution (inset) and persists for about 200 s, much shorter than the 2760 s data segment. This particular QPO may not be of solar origin, since it is not observed in other flare-observing instruments. However, a similar signature is coincidentally seen in the GOES magnetometer data, suggesting a possible artifact. Regardless of the origin, the QPO is clearly present and confined to a small portion of the data, providing an ideal test case.

To demonstrate the presence of an artificial amplification of the SNR, we split the lightcurve into three segments. The first segment covers the time before the QPO (0 s to 1680 s) (after 20:17 UT), the second selects just the QPO (1680 s to 1880 s), and the third segment selects everything after the QPO (1880 s to 2760 s). We perform Bayesian inference separately on each of these segments and the combined segment independently with the model S^{RW} and S^{RWQ} . We use the same priors as in the studies on simulated data which we list in Tab. 6.2. If the lightcurve were stationary, it would be valid to combine the Bayes factors and f_0 posteriors as they represent independent draws from the same distribution. However, as we show in Tab. 6.1, we are unable to detect the reported QPO in the first and third segments. There is some weak evidence towards a QPO in the first segment, though not at the reported 12.6 s period. The third segment shows some weak support for a QPO with $\ln BF = 1.4$, but the maximum likelihood fit indicates that this is rather due to a broad feature in the periodogram that is better fitted with a broken power law. As we show in Tab. 6.1, both the first and third segments are better fitted with S^{BW} than with S^{RWQ} , which indicates that it is likely no QPO present. In the second segment we find the QPO independently with a very high significance $\ln BF = 27.5$, and it is clearly visible in the fitted periodogram in Fig. 6.9. Finally, we analyze the combined segments together and find the QPO with

³<https://aringlis.github.io/AFINO/>

Segment	$\ln BF$ S^{RWQ}	$\ln BF$ S^{BW}	ΔBIC S^{RWQ}	ΔBIC S^{BW}
0 s to 1680 s	3.0 ± 0.3	6.8 ± 0.3	-11.3	-5.0
1680 s to 1880 s	27.5 ± 0.2	-0.7 ± 0.2	-66.8	7.8
1880 s to 2760 s	1.4 ± 0.3	7.7 ± 0.2	-0.7	-6.8
0 s to 2760 s	229.4 ± 0.3	93.0 ± 0.3	-465.6	-175.8

TABLE 6.1: Results from analyzing the selected segments of the solar flare detected by GOES. All errors are given based on $1-\sigma$ confidence or credible intervals. The segments are given in seconds after 20:17 UT. All $\ln BF$ and ΔBIC values are calculated relative to S^{RW} . We reiterate that either a positive $\ln BF$ or a negative ΔBIC indicate that S^{RWQ} or S^{BW} are preferred over S^{RW} . We find broadly that the ΔBIC values in the last row are in agreement with what AFINO reported. Deviations are likely due to differences, in our model and slightly different data selection. While in all instances Bayes factors and ΔBIC s give the same indication about the preferred model, the significance differs to some extent. This may be in part due to our wide prior choices.

$\ln BF = 229.4$, a significance much higher than in segment two. We note that we find the QPO at a slightly different period of $12.34\text{s} \pm 0.06\text{s}$ compared to the value reported of AFINO. This may be due to slightly different QPO modeling choices between AFINO and our method.

We perform the local χ^2 -test we introduced in Sec. 6.4.5 for the entire lightcurve and find $\chi^2 = 0.29$ for the frequency bins surrounding the QPO. This indicates that the QPO has been overfitted and is non-stationary.

It is evident that the solar flare segment we have analyzed contains a QPO and an overestimation of its significance is not as critical as it would be for a marginal detection. Trying to find an instance where a marginal detection was turned into a very confident detection due to non-stationarity bias would be much harder. We would not be able to determine the location of the QPO in the lightcurve by eye. Instead, we would have to take a systematic approach and split the lightcurve into several smaller segments and determine if the QPO exists in them individually. Of course, the significance in the individual segments would always be lower, so it would be hard to determine whether this is due to the reduced non-stationarity bias or because we split a persistent QPO into multiple segments. As mentioned previously, we do not expect AFINO to report many false detections, as has been established in Broomhall et al. [70]. This is due to its conservative detection threshold $\Delta BIC < -10$. Thus, if AFINO is coupled with a more careful data selection method it could be possible to find more QPOs.

The non-stationarity bias in solar flare data arises most likely as a combination of the effects we describe in Sec. 6.3, i.e. a non-stationary QPO, non-stationary noise processes, and a deterministic overall flare shape all contribute to some extent. We can not easily discern what the impact of each of these effects is.

Non-automated analyses of QPOs in solar flares are also likely less affected than the stretch of GOES data we analyze in this section. Typically, parts of the lightcurve that are of interest are selected manually for detailed analyses [e.g. 136], though this by no means provides a guarantee that such results are without bias. As we have established previously, the $\ln BF$ or the ΔBIC can grow linearly with x , thus even a mild overselection of the lightcurve can lead to erroneous detections and inferences.

6.6 Discussion and outlook

In this paper, we show that analyses based on Whittle likelihoods likely overstates the significance of QPOs in lightcurves if the QPO or noise features in time series are non-stationary. Specifically, this is the case if data is included in the time series that contains only white noise when the QPO is in the red-noise-dominated part of the PSD. The effect can also occur if the QPO is in the white-noise-dominated part of the PSD, but the white noise is non-stationary (e.g. if the noise arises due to photon counting) and we include weak white noise from parts of the lightcurve that have low photon counting rates. Selecting an appropriate time segment for the analysis of transients is thus important but remains difficult to do generically and without bias. A scientist’s natural intuition may be to select data conservatively, i.e. select a longer time series while being aware that the QPO may only be present for a part of it. The erroneous reasoning may be that by selecting a longer segment they are not “cherry-picking” the segment with the most pronounced QPO features. As we demonstrate in this paper, this choice may increase the significance of the QPO artificially and lead to false detections due to the non-stationarity bias.

There are some possible mitigation strategies that one may apply when the use of periodograms is still preferred. However, this means that it is a lot harder to devise a generic process with which to analyze time series. Firstly, if we suspect that the QPO is indeed only present for part of the transient, and the white noise level is relatively low compared to the low-frequency continuum, as is indeed the case for x-ray lightcurves of solar flares, it is reasonable to seek out the part of the lightcurve that looks by eye most likely to contain a QPO, and then set the limits on the start and end of the segment manually. However, this method requires significant human supervision and thus will not scale to large data sets. It is open to other forms of selection bias, too. Secondly, for such transients, we also suggest splitting the lightcurve up into several parts to see if one can detect the QPO in all segments or only in some of them. For the final analysis, one should only use segments if there is some evidence for the QPO. Next, for shorter transients where we suspect the QPO to be present for most of the time, we should conservatively trim the lightcurve after the onset and before the end of the segment. This way we are less likely to make the lightcurve non-stationary. Finally, we have found that regular χ^2 -like tests are unsuitable to detect non-stationary bias since only a few bins around the QPO are affected. The overall χ^2 -value is mostly determined by the far greater share of frequency bins that lie in the white-noise-dominated part of the PSD. We outline that alternatively calculating χ^2 based solely on the frequencies surrounding the QPO can detect overfitting, which is a hallmark of statistically non-independent frequency bins.

Aside from periodograms, other methods that can be used for the search for QPOs. Wavelet transforms are a popular way to analyze the time series by convolving a wavelet function with a time series of interest [282]. Wavelet analyses are not restricted to stationary data sets and may help us to find the specific times when a QPO occurs. However, the statistics of detecting QPOs with wavelets in the presence of red noise remains contentious since detrending methods are likely to lead to false detections [49]. Alternatively, we can avoid frequency-domain methods altogether and model the time series as a Gaussian process with some mean function. The main drawback of Gaussian processes is that they have computational complexity $\mathcal{O}(N^3)$ and thus are only suitable for short time series in the general case [233], though progress has been made

to reduce complexity to $\mathcal{O}(N)$ for stationary, complex exponential kernels or combinations thereof [113]. We note that we expect the same non-stationarity bias if we apply a stationary Gaussian process kernel to a non-stationary time series since the Whittle likelihood explicitly derives from a Gaussian process likelihood. We have also found this result empirically in some preliminary analyses (Huebner et al. 2021, in prep).

One advantage of Gaussian processes is that they allow us to fit the overall shape of the transient and the stochastic process simultaneously, instead of lumping the shape of the transient into the red noise, which can help us to prevent the bias demonstrated in Sec. 6.4.3. It thereby allows us to avoid nonparametric methods that are likely to create false periodicities [49]. Additionally, the Gaussian process likelihood can explicitly take in the known variance for flux values within a lightcurve if they are available. Thus, they also provide a natural resolution to the bias shown in App. 6.9. Finally, Gaussian processes allow us to model relatively simple, non-stationary extensions to the fast, stationary models that are popular now (Huebner et al. 2021, in prep). Further development of Gaussian process methods may eventually allow us to create more complex models of QPOs and their possible intermittency within transients.

6.7 Acknowledgements

This work is supported through ARC Centre of Excellence CE170100004. P.D.L. is supported through the Australian Research Council (ARC) Future Fellowships FT160100112, ARC Discovery Project DP180103155. D.H. is supported by the Women In Science Excel (WISE) programme of the Netherlands Organisation for Scientific Research (NWO). This work was performed on the OzSTAR national facility at Swinburne University of Technology. The OzSTAR program receives funding in part from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government.

6.8 Appendix: Priors and simulated data parameter tables

We list the priors and parameters used for all the studies in Secs. 6.4 and 6.5 in Tabs. 6.2 and 6.3.

6.9 Appendix: Deterministic processes impact white noise

White noise observed in astrophysical lightcurves using photon counting does not arise due to intrinsic properties of the source, but rather due to the Poissonian noise nature of photon counting. Concretely, given a rate λ , the distribution of the number of observed photons k in a unit of time is Poissonian

$$\text{Pois}(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \quad (6.28)$$

Photon counting noise thus scales proportionally to the standard deviation of the Poisson distribution $\sqrt{\lambda}$. This relation implies that photon-counting noise throughout a transient

Parameter	Description	Prior	Minimum	Maximum
A	Red noise amplitude	LogUniform	$\exp(-30)$	$\exp(30)$
$\alpha, \alpha_1, \alpha_2$	red noise / BPL spectral index	Uniform	0	10
B	QPO amplitude	LogUniform	$\exp(-60)$	$\exp(60)$
f_0	QPO central freq.	LogUniform	$2\Delta f$	f_{\max}
σ	QPO HWHM	Conditional-LogUniform	$\Delta f/\pi$	$0.25f_0$
C	White noise amplitude	LogUniform	$\exp(-30)$	$\exp(30)$
δ	Breaking freq. for BPL	LogUniform	Δf	f_{\max}

TABLE 6.2: Parameters and priors used throughout all studies in this paper. Priors are referred to by their implementation in BILBY. We select wide priors in A , B , and C for simplicity. Although the selection of priors this wide may have a slight impact on the Bayes factor calculation, it will not qualitatively change the scaling of Bayes factor with x . We also set a wide prior on α , or α_1 and α_2 , and enforce $\alpha_2 < \alpha_1$ using a **Constraint** prior in BILBY to avoid degeneracies. The prior on f_0 is motivated by the available frequencies in the periodogram. We set the minimum f_0 at twice the difference between neighboring frequencies Δf to ensure better convergence, and refer to the highest frequency in the periodogram as f_{\max} . We set σ , the width of the QPO at half width half maximum (HWHM), to be conditional uniform in log between $\Delta f/\pi$, i.e. on the scale of a single frequency bin, and $0.25f_0$, to prevent it from converging towards wider features in the periodogram. For the broken power law analysis of solar flare data, we use priors for $\alpha_{1,2}$ identical to α , and the listed δ prior

Parameter	Sec. 6.4.2 z	Sec. 6.4.2 wn	Sec. 6.4.3	Sec. 6.4.4	App. 6.9
A	-	-	-	4	-
α	-	-	-	2	-
B	3	15	80	8	100
f_0	5	5	1	1	5
σ	0.1 Hz	0.1 Hz	0.01 Hz	0.02 Hz	0.1 Hz
C	2	2	2	2	-
A_{flare}	-	-	20000	-	1000 s^{-1}
t_0	-	-	70 s	-	200 s
τ_r	-	-	10 s	-	-
τ_f	-	-	20 s	-	-
σ_{flare}	-	-	-	-	20 s
c_0	-	-	-	-	10 s^{-1}
Segment length	400 s	400 s	200 s	400 s	400 s
Sampling frequency	40 Hz	40 Hz	40 Hz	40 Hz	40 Hz

TABLE 6.3: Values used to create the simulated data in Sec. 6.4. The column heads refer to the subsection in which this set of parameters was used. We use two different sets of parameters in Sec. 6.4.2 depending on whether we extend using zeros (z) or more white noise (wn).

does not remain constant. We demonstrate a limit of this in Sec. 6.3.1 when we consider what happens if we extend a time series with zeros. This limit corresponds to the case in which a detector sees no photons at all. In realistic detectors, we will not reach this limit as there are at least some background photons, though we can get close enough to it for it to lead to wrong inferences about QPO significance.

The above effect points to a deficiency in periodograms more generally. For the calculation of a periodogram we use the values $x(t_i)$ of a discrete time series, but we do

not include their variances $\Delta x(t_i)$ if they are available. In effect, this implies a loss of information as the white noise is already encoded in the variance associated with the photon counts. Instead, we infer it after the fact via power spectral density estimation independently and lose knowledge about non-stationarity in the white noise.

To demonstrate this effect, we construct a transient similar to Sec. 6.4.3, though in a way that its low-frequency contributions do not overlap with the QPO frequency, as to demonstrate that this effect arises due to non-stationarity in the white noise. Concretely, we construct the transient by generating a Gaussian profile plus a constant

$$f(t; A_{\text{flare}}, t_0, \sigma_{\text{flare}}) = A_{\text{flare}} \exp\left(-\frac{(t - t_0)^2}{2\sigma_{\text{flare}}^2}\right) + c_0, \quad (6.29)$$

where A_{flare} is the amplitude, t_0 is the time of the peak, σ_{flare} is the width, and the constant c_0 represents a possible background count rate. We add a 20s non-stationary QPO in the same manner as we describe in Sec. 6.4.3. We create photon counts, which we show in Fig. 6.10, by simulating the Poisson process using the implementation in `scipy.stats.poisson.rvs`. The specific parameters of this time series are listed in Tab. 6.3. By construction, the QPO is set on the top of the Gaussian profile, which corresponds to the highest level of photon counting noise. The Gaussian profile, unlike the exponential from Sec. 6.4.3, contributes powers at lower frequencies than the exponential, roughly up to $1/\sigma_{\text{flare}} \approx 0.05$ Hz. Additionally, we set our QPO frequency at 5 Hz and cut off frequencies below 0.5 Hz so that we can avoid using a red noise component in our modeling in the same way we did in Sec. 6.4.2.

On the tails of the flare, we expect the same effect as we observe when we extend with zeros in our other simulations, though Fig. 6.11 clearly shows that we artificially increase the significance already with very small extension factors. This is possible since the flare quickly falls off on either side and the added photon counting noise does not compensate for the increased number of frequency bins that contain the QPO. Overall, this effect would be hard to account for just based on better methods for setting data cuts because varying levels of white noise are expected in many transient lightcurves.

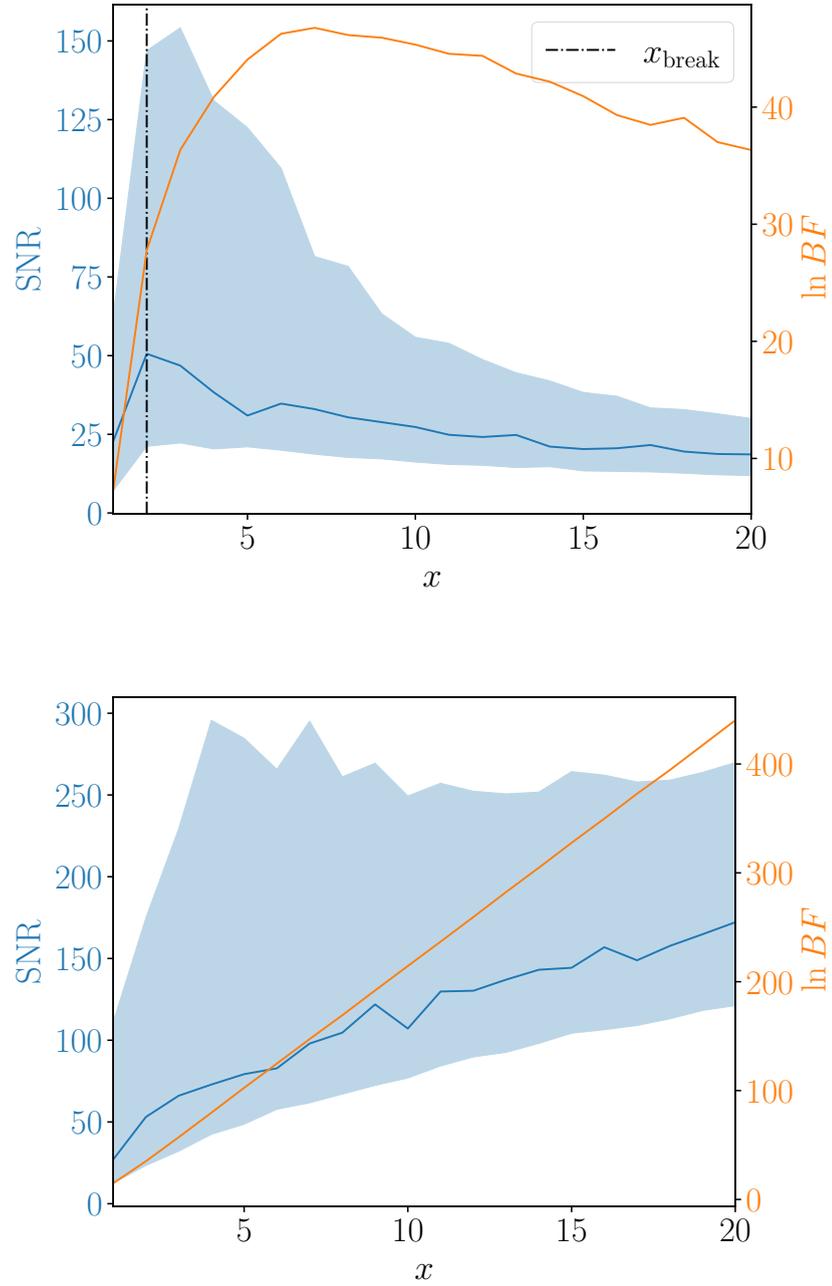


FIGURE 6.6: SNR (blue) and $\ln BF$ (orange) vs the extension factor for simulated data of a non-stationary QPO and red noise with white noise as described in Sec. 6.4.4 for extension with white noise (top) and zeros (bottom). The solid blue line displays the SNR of the maximum likelihood point and the shaded blue region displays the SNR's 90% credible interval.

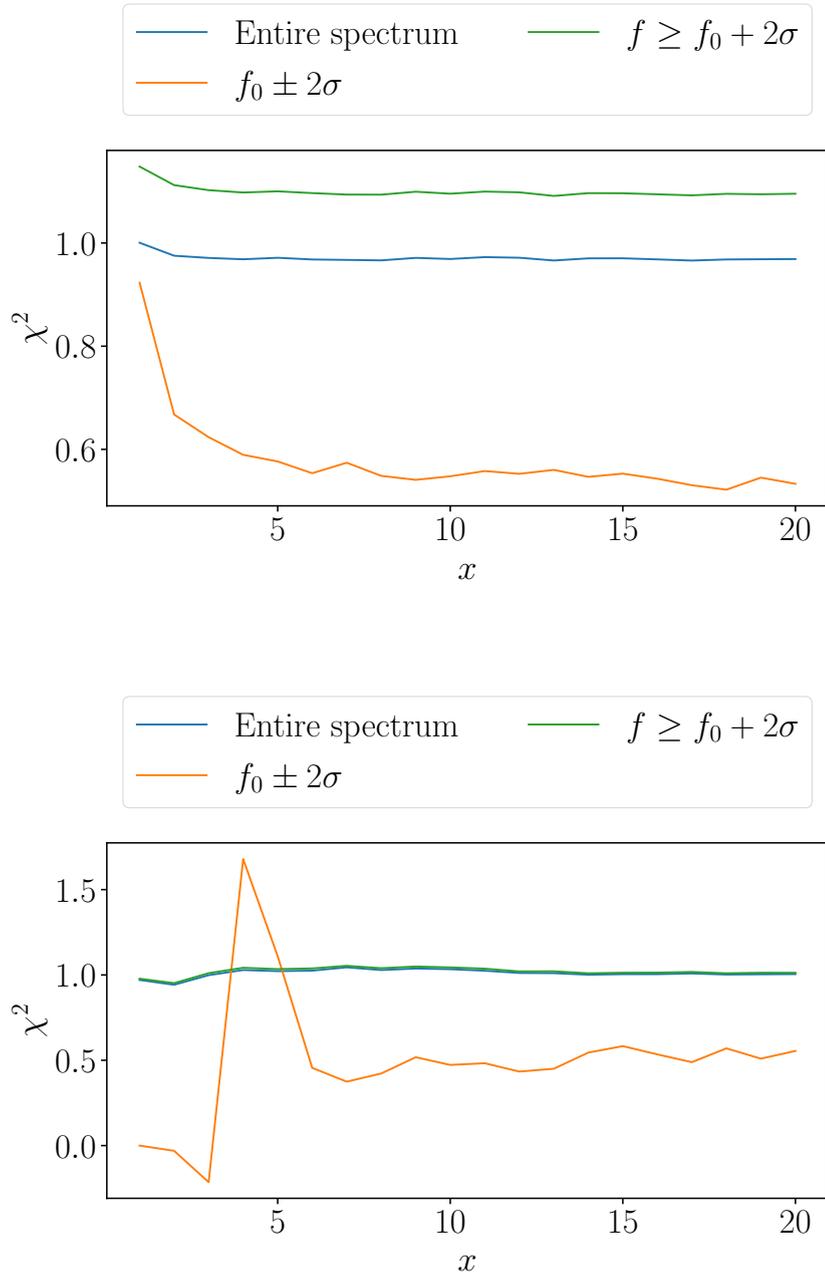


FIGURE 6.7: Nita et al. [212] χ^2 -tests in the case of a QPO and white noise extended with zeros as described in Sec. 6.4.2 (top), and the case of QPO, red noise and white noise extended with white noise as described in Sec. 6.4.4 (bottom). We see in the top panel that the overall fit (blue) and the fit for frequencies above the QPO frequency (green) have $\chi^2 \approx 1$, indicating a good fit. The local χ^2 -test around the QPO frequency f_0 indicates a good fit only for $x = 1$ and falls much below that for $x > 1$, indicating overfitting. This implies that the local χ^2 -test can detect non-stationarity bias in this instance. In the bottom panel, we display some of the limitations of the local χ^2 -test when we deal with narrow QPOs. For $x \lesssim 6$ the local χ^2 -test (orange) around f_0 swings widely as we have too few frequency bins within the range. Above the $x \approx 6$ enough frequency bins accumulate within the range and we detect the overfitting. We also note that the χ^2 -test for the entire PSD (blue) and frequencies above the QPO frequency (green) indicate an overall good fit despite us overfitting the QPO

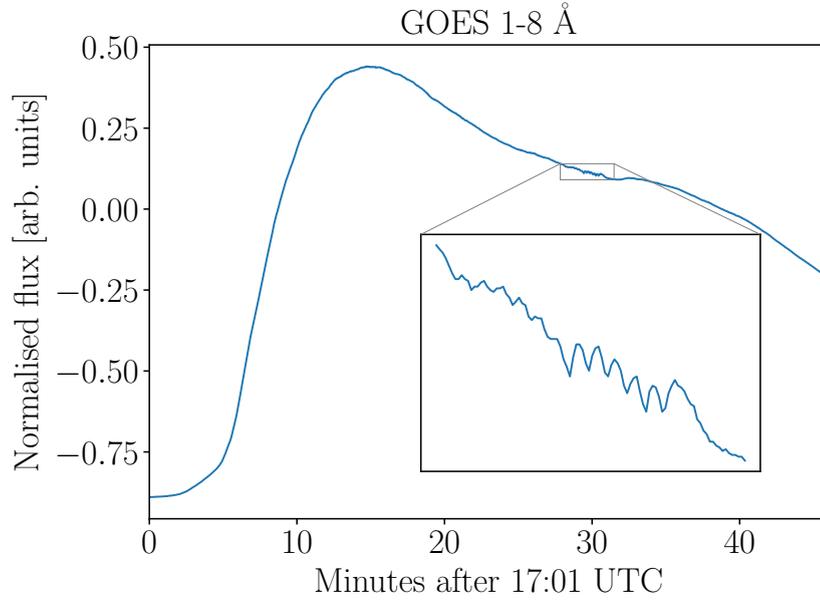


FIGURE 6.8: Solar flare x-ray lightcurve as observed with GOES in the 1 \AA to 8 \AA band. The figure shows the time selected as done with AFINO. The inset zooms in on the tail of the lightcurve where the QPO is clearly visible.

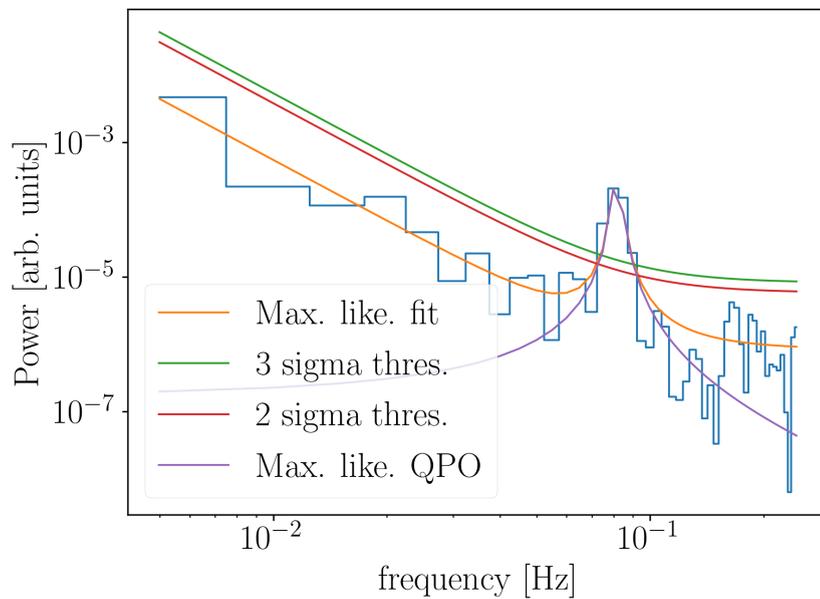


FIGURE 6.9: Maximum likelihood fit of the periodogram just for the second segment of the GOES lightcurve. The QPO is clearly visible and highly significant. We draw the 2σ and 3σ detection limits based on the Bonferroni-corrected frequentist statistics. We note that the QPO is located in the red noise dominated part of the PSD.

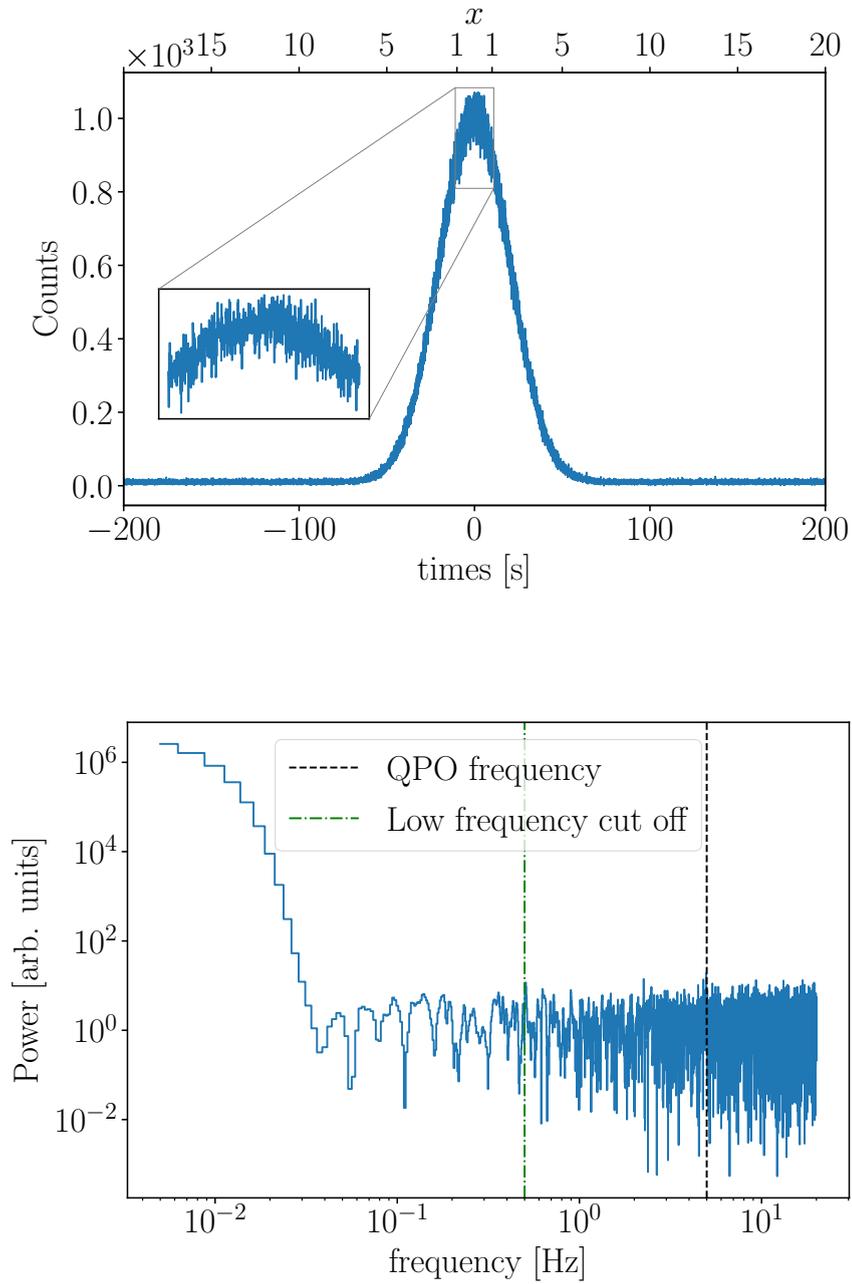


FIGURE 6.10: Simulated data of a non-stationary QPO in a deterministic transient flare shape in the presence of white photon counting noise. The time series (top) shows the dominant transient and the QPO that is present for 20s on the top of the flare (inset). The periodogram for the entire 400s time series (bottom) has a relatively wide ($\sigma = 0.1$) QPO at 5 Hz (barely visible). For our analysis, we cut off all frequencies below 0.5 Hz to avoid any effects arising from the power in low frequencies due to the flare shape.

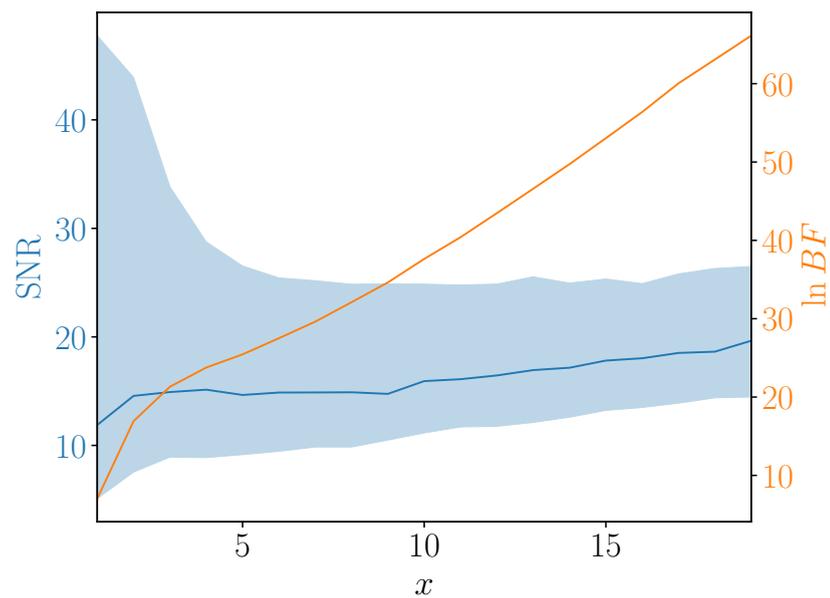


FIGURE 6.11: We calculate the Bayes factors for the presence of a QPO in a transient flare for different extension factors with $x = 1$ corresponding the 20 s inset from the top of Fig. 6.10. While we already start with very strong evidence of a QPO at $x = 1$, the Bayes factors continue to increase as we extend the time series to include more of the transient.

Chapter 7

Detecting Quasi-periodic Oscillations with Gaussian Processes

We are following up on the discussion from the preceding chapter about how non-stationarities can bias the search for quasi-periodic oscillations (QPOs) when we use periodogram-based analysis methods. An alternative way to look at time-series data is to understand them as a Gaussian process (GP), i.e. a combination of a deterministic and stochastic process in which all data points are part of a multivariate normal distribution (see Ref. [233] for a comprehensive introduction). A GP is characterised by a kernel function $k(t_0, t_1)$, which describes the stochastic aspect via the covariance between two different points in time t_i , and the mean function $\mu(t)$, which describes the deterministic aspect. GPs are entirely modelled in the time domain and thus obviate the need for Fourier transforms. Using GPs also has several challenges associated with it, especially motivating the choice of kernel function and computing the GP likelihood in an acceptable amount of time.

In this chapter, we discuss how we can use a GP approach to search for QPOs and how this can explicitly address the issues we encountered in the previous chapter. Time-domain modelling allows us to distinguish stochastic from deterministic aspects of the light curve as these form different parts of the model which can be inferred concurrently. Although creating and evaluating general non-stationary QPO models is complicated, we show how we can make some relatively easy tweaks to stationary models to allow the exploration of simple, non-stationary structures.

As we show, there are also some trade-offs to using GPs. The statistics only works for Gaussian data and is unsuitable if we are analysing, e.g., Poissonian data with very low count rates, which are common in some x-ray observations. Next, GP likelihoods are computationally expensive to evaluate in the general case ($\mathcal{O}(N^3)$ for N data points), though fast evaluation algorithms exist for some classes of kernel functions. We utilise an $\mathcal{O}(N)$ complexity algorithm that is available for the class of complex exponential kernel functions and combinations thereof [113], which is suitable for exploring periodicity ¹.

In this chapter, we show that using Bayesian inference, properties of a QPO such as its frequency can be inferred, and that model selection between QPOs and alternative

¹Strictly speaking, CELERITE has $\mathcal{O}(NJ^2)$ complexity, where J is the number of complex exponential terms we combine to create a kernel function, which is no more than two in this study

models such as red noise can be performed reliably. We layout in detail what GPs are and motivate a kernel function that corresponds to QPOs in Sec. 7.1. In Sec. 7.2, We demonstrate on simulated data that both parameter estimation and model selection work reliably. This method is easily transferable to many astrophysical data sets. We show this by applying the method to some selected astrophysical time-series data, such as GRBs, magnetar flares, and the solar flare we investigated in the preceding chapter, in Sec. 7.3. We discuss the results in Sec. 7.4.

7.1 Methods

In the following, we briefly introduce the core definitions of GPs, specifically in the one-dimensional case relevant to this study. For a more in-depth introduction, the reader can refer to Ref. [233].

7.1.1 Gaussian Processes Overview

A GP is defined by a covariance matrix $K(\alpha)$ and a mean function $\mu_\theta(t)$, with α and θ being their respective parameters². The mean function describes non-stochastic behaviour in our data and could thus also be called the ‘‘trend function’’ in the context of astrophysical light curves which are a function of time.

We define the coordinate vector \mathbf{t} and data vector \mathbf{y} with N entries each. The GP log-likelihood is defined as

$$\ln L(\theta, \alpha) = -\frac{1}{2} \mathbf{r}_\theta^T K_\alpha^{-1} \mathbf{r}_\theta - \frac{1}{2} \ln \det K_\alpha - \frac{N}{2} \ln(2\pi), \quad (7.1)$$

where

$$\mathbf{r}_\theta = \mathbf{y} - \mu_\theta(\mathbf{t}), \quad (7.2)$$

is the residual vector and the covariance matrix elements are given by the kernel function $[K_\alpha]_{nm} = \sigma_n^2 \delta_{nm} + k_\alpha(t_n, t_m)$ where σ_n^2 are the variances due to white noise and δ_{nm} is the Kronecker delta. In the simplest case of $k_\alpha(t_n, t_m) = 0$ the likelihood function becomes a simple Gaussian likelihood which is typically employed to fit Gaussian data. The fact that we can take the σ_n into account, unlike for periodograms, means that we automatically correctly treat heteroscedastic data.

It is important to underline that σ_n , i.e. the white noise estimate on any data point if known, is intrinsically part of the GP likelihood, which is not the case for the Whittle likelihood. If we analyse a time series using the Whittle likelihood we have to infer the white noise level from \mathbf{y} , and will only be able to assume the white noise level to be constant, which is not the case in most astrophysical light curves. This points directly to an advantage of GP modelling.

The σ_n in astrophysical light curves often arise due to photon counting which is described by a Poisson distribution

$$\text{Pois}(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (7.3)$$

²In the context of GP regression, the parameters α are often referred to as ‘‘hyperparameters’’, which should not be confused with the hyperparameters used in hierarchical inference.

for k photon counts given a photon rate λ . Both mean and variance of the distribution are λ . Since the distribution is Gaussian for high counting rates, it is valid to apply a GP likelihood by approximating $\sigma_n \approx \sqrt{\lambda(t_n)} \approx k_n$, where k_n is the number of photon counts in the n th bin.

One of the main drawbacks of GPs is that in general, the likelihood takes $\mathcal{O}(N^3)$ steps to evaluate [233]. There are numerous sophisticated approaches to reduce this complexity to a more manageable level [111, 121, 310], including GPU acceleration [94]. Most suitable for us is the CELERITE software package, which solves performance issues by restricting itself to one-dimensional problems and the class of complex exponential kernel functions and combinations thereof which reduces the complexity to $\mathcal{O}(NJ^2)$ [113] with N being the number of data points and J being the number of exponential terms that form the kernel. CELERITE achieves this by exploiting the semi-separable structure of the covariance matrices, which allows for the use of a fast solver for the Cholesky factorisation [113]. We investigate in the following section why the class of exponential kernel functions is sufficient for our purposes.

7.1.2 Kernel Functions

As explained in the previous section, the kernel and the mean function are the two elements that model a GP. There exists exhaustive literature on popular kernel functions, but it is not always clear which kernel function should be used and how to interpret their hyperparameters. In fact, for many applications of GPs, this is not even the goal. Instead one may be interested in GP's predictive power to interpolate and extrapolate data, which can be tested and tuned on training sets. However, we are interested in being able to interpret the kernel function and associate it with an underlying physical process. By choosing to use CELERITE, we need to use a kernel that is in the class of complex exponential kernels.

We start by defining a kernel describing a periodic oscillation

$$k_{\text{po}}(\tau) = a \cos(2\pi f\tau), \quad (7.4)$$

where a is the amplitude of the oscillation and f is its frequency. This kernel corresponds to a perfect harmonic oscillation; its representation in the power spectrum would be a delta peak at f . In order to make this kernel quasi-periodic, we add another factor to account for variations in amplitude over time

$$k_{\text{qpo}}(\tau) = \exp(-c\tau)k_{\text{po}}(\tau). \quad (7.5)$$

This kernel corresponds to a QPO with a parameter c that describes on which time scale $1/c$ oscillations are excited and damped. In the frequency domain, this kernel corresponds to a Lorentzian function with peak frequency f and full-width at half maximum c , which directly corresponds to the model that we used in the preceding chapter.

In practice, many astrophysical systems will display both red noise and QPO at the same time. In the context of CELERITE, we can model red noise as a simple exponential

$$k_{\text{rn}}(\tau) = a \exp(-c\tau). \quad (7.6)$$

Physically, this kernel corresponds to a damped random walk (Ornstein-Uhlenbeck process), and its power spectrum is a f^{-2} power law.

In order to obtain a kernel that both describes the QPO and the red noise process, we simply add both kernels

$$k_{\text{qpo+rn}}(\tau) = k_{\text{qpo}}(\tau) + k_{\text{rn}}(\tau) \quad (7.7)$$

$$= a_{\text{qpo}} \exp(-c_{\text{qpo}}\tau) \cos(2\pi f\tau) + a_{\text{rn}} \exp(-c_{\text{rn}}\tau) \quad (7.8)$$

denoting QPO and red noise component with the subscripts ‘‘qpo’’ and ‘‘rn’’ respectively. This kernel is somewhat different to a QPO kernel proposed in [113]

$$k(\tau) = \frac{a}{2+b} \exp(-c\tau) [\cos(2\pi f\tau) + (1+b)], \quad (7.9)$$

which also both models QPO and red noise features, but makes them share the c parameter. Since we do not find that $c_{\text{qpo}} = c_{\text{rn}}$ in general, we use Eq. 7.7 as our model for the general case and Eq. 7.5 for the case that we only encounter a QPO but no red noise.

There are alternative ways to describe periodic behaviour outside the domain of CELERITE compatible kernel functions. One example has been used in Ref. [39] in the astrophysical context, who adapted it from Ref. [233] who used it to model seasonal changes in atmospheric CO2 levels

$$k(\tau) = a \exp\left(-\frac{\tau^2}{2\ell^2} - \Gamma \sin^2(\pi f\tau)\right). \quad (7.10)$$

This kernel combines a squared exponential with an oscillatory term such that a parameter Γ controls the amount of covariance between two points that are roughly one period away from each other. For high values of Γ , only points exactly integer multiples of a period away from each other have high covariance. This way arbitrary curves with a repeating shape can be described using this kernel, although a QPO kernel as in Eq. 7.7 has shown to be sufficient to correctly infer the frequency [113]. Furthermore, while Eq. 7.10 is useful to fit arbitrary curves, it may be hard to interpret the underlying physics from the inferred parameters and comes with the high computational cost of general GP kernels. Alternatively, Eq. 7.10 could also be approximated by a Fourier series since the CELERITE kernel family forms a Fourier basis. However, this approach would also incur far higher computational costs since the CELERITE likelihood complexity scales quadratically with the number of terms required to build the kernel and the Fourier coefficients to be inferred would drastically increase the parameter space.

There are also alternative ways to define aperiodic kernels. Within the support of CELERITE models, Ref. [113] proposed using a critically damped stochastic harmonic oscillator

$$k(\tau) = S_0 \omega_0 e^{-\frac{1}{\sqrt{2}}\omega_0\tau} \cos\left(\frac{\omega_0\tau}{\sqrt{2} - \frac{\pi}{4}}\right), \quad (7.11)$$

which is commonly used to model background granulation noise in asteroseismic and helioseismic oscillations [113]. Outside of the CELERITE support, the squared exponential, Matern-3/2 and rational square kernel functions are employed in a great number of contexts. Recently some of these kernels have also been applied in the search for oscillations in blazar light curves [88], but they are difficult to physically motivate and the $\mathcal{O}(N^3)$ complexity restricted their use to maximum likelihood estimates rather than Bayesian

evidence calculation. We stick to the red noise kernel for our analyses as we consider this to be a basic framework that may be able to describe many physical systems to a first degree.

7.1.3 Mean Functions

Choosing a mean model is similarly difficult in general if there is no universally physically motivated model for the underlying background trend. The simplest mean model is to just take the (weighted) average of all recorded data points. This is tantamount to assuming that all variability in the light curve is due to stochastic processes. If there is a substantial background trend present this assumption quickly breaks down and leads to incorrect inferences about the red noise content in the light curve. It is thus not a suitable way to approach the problem in most cases.

Another approach is to use filter methods such as the boxcar filter or the Savitzky-Golay filter [249] and subtract the filtered light curve from the original light curve. As we have mentioned in the preceding chapter, using filtering methods is prone to cause artificial periodicities [49]. Some problems feature a relatively simple background trend that could be modelled as a linear trend or another low-order polynomial instead of a filter. Other problems may have existing physical or phenomenological models for their trend, e.g. there are popular models of gamma-ray burst (GRB) light curves such as the “fast-rise exponential decay” (FRED) model.

For general problems, one can employ methods such as shapelet fitting, in which the problem is expressed as a sum of orthonormal basis functions. Alternatively, for burst-like light curves, it is often possible to identify several base flare shapes (e.g. (skewed) Gaussians, fast-rise exponential decays [150], etc.) that can be used to model the overall shape of the burst. While one has to be careful with the definition of priors in this inherently degenerate problem, we show that this may be a worthwhile approach for many problems.

In the following we introduce some flare mean models that we use throughout this Chapter. As the most basic models we used skewed Gaussians and skewed exponentials, i.e.

$$\mu_{\text{GAUSS}}(t; A, t_0, \sigma_1, \sigma_2) = \begin{cases} A \exp\left(-\frac{(t-t_0)^2}{2\sigma_1^2}\right) & \text{if } t \leq t_0 \\ A \exp\left(-\frac{(t-t_0)^2}{2\sigma_2^2}\right) & \text{if } t > t_0 \end{cases} \quad (7.12)$$

and

$$\mu_{\text{EXP}}(t; A, t_0, \sigma_1, \sigma_2) = \begin{cases} A \exp\left(-\frac{t-t_0}{\sigma_1}\right) & \text{if } t \leq t_0 \\ A \exp\left(\frac{t-t_0}{\sigma_2}\right) & \text{if } t > t_0. \end{cases} \quad (7.13)$$

There are a number of other approaches to define FRED models. We focus on the ones defined in Refs. [214, 215], which can assume a wide variety of shapes. The regular FRED model is defined as

$$\mu_{\text{FRED}}(t; A, t_0, \psi, \Delta) = A \exp\left(-\psi \left(\frac{t+\Delta}{t_0} + \frac{t_0}{t+\Delta}\right)\right) \exp(2\psi) \quad (7.14)$$

for $t < 0$. The maximum of this model coincides with $t = t_0$, and we can interpret Δ as an “offset” parameter that allows us to fit the full range of flare shapes with arbitrary offsets from $t = 0$. Finally, ψ serves as a symmetry parameter, with the curve becoming increasingly symmetrical for large values of ψ . The additional factor $\exp(2\psi)$ is optional in the model definition and serves as a normalisation factor such that $\mu(t = t_0; A, t_0, \psi, 0) = A$, which eases the definition of the amplitude prior.

We can extend the regular FRED model by adding extra exponents to the terms in the exponential. This allows the model to assume some more “peaky” structures.

$$\begin{aligned} \mu_{\text{FRED-x}}(t; A, t_0, \psi, \Delta, \mu, \nu) = \\ A \exp\left(-\psi^\mu \left(\frac{t + \Delta}{t_0}\right)^\mu - \psi^\nu \left(\frac{t_0}{t + \Delta}\right)^\nu\right) \exp(2\psi) \end{aligned} \quad (7.15)$$

As we show later, any of the models we have introduced here perform reasonably well in terms of fitting general flare shapes. Performing Bayesian inference with these models thus naturally yields comparable evidences. For general problems, it is prudent to run the analysis with multiple mean models and select the one that yields the highest evidence.

7.1.4 Towards non-stationary Models

We have already laid out how GPs resolve the issues we found with Whittle-likelihood based analyses in the case of heteroscedastic data and in case a deterministic trend seeps into frequency bins that contain the QPO. The other issue we face is if the QPO and noise process are non-stationary. Specifically, we are considering the following scenario: a time series that is split into three parts, the first and last contain only white noise whereas the middle contains both a QPO and red noise. Additionally, we allow for a continuous mean function to be present everywhere. We show how we can implement this basic, non-stationary process and how we can use it to infer when QPOs start and end.

Stationarity in the context of GPs is defined as

$$k(t_0, t_1) = k(\|t_1 - t_0\|) = k(\tau), \quad (7.16)$$

i.e. the matrix looks the same if we move forwards or backwards in time. As a simple example, we consider a system that has some regular x-ray emission characterised by white noise, then flares for some time, and then goes back to its regular emission. If we do not select the data we want to analyse carefully, we are likely to incur some biases due to the non-stationarity bias, which occurs due to the white noise at the beginning and the end of the segment. As Whittle likelihoods are based on the same assumptions as stationary GP likelihoods, the bias is very similar in both cases.

By breaking the time series into three segments, we can model a very basic, non-stationary process without having to resort to a more general likelihood solver. Instead, we split the likelihood into two parts, one for the two disjoint, white noise segments on either end and one for the stochastic process in the centre. The covariance matrix can then be decomposed into two submatrices correspondingly. We obtain one “inner” covariant submatrix by deleting all rows and columns that only contain the diagonal σ_n and one “outer” non-covariant submatrix by deleting all the other rows and columns

from the original covariance matrix. Since the covariant submatrix forms a block, we can calculate the inverse and the determinant independent from the remaining entries. The same operations are trivial for the non-covariant blocks. We can re-write Eq. 7.1 to describe this explicitly

$$\begin{aligned} \ln L(\theta, \alpha) = & -\frac{1}{2} \sum_{n_{\text{out}}} \frac{r_{n_{\text{out}}}^2}{\sigma_n^2} - \sum_{n_{\text{out}}} \ln \sigma_{n_{\text{out}}} - \frac{N_{\text{out}}}{2} \ln(2\pi) \\ & - \frac{1}{2} r_{\text{in},\theta}^{\mathbf{T}} K_{\text{in},\alpha}^{-1} r_{\text{in},\theta} - \frac{1}{2} \ln \det K_{\text{in},\alpha} - \frac{N_{\text{in}}}{2} \ln(2\pi) \end{aligned} \quad (7.17)$$

where the likelihood coming from the outer non-covariant submatrix is in the first line and the inner covariant submatrix is in the second line. Here n_{out} are the indices of the outer N_{out} non-covariant row/columns, $K_{\text{in},\alpha}$ is the inner covariant submatrix, and r_{in} are the residuals of the N_{in} inner rows/columns. We introduce two additional parameters to our model t_{start} , and t_{end} , that describe the transition point between those segments.

7.1.5 Assumptions and Limits of GPs

While GPs show great flexibility, their use implies certain assumptions about the systems we investigate. Firstly, we are assuming that QPOs are an additive process, i.e. the QPO does not depend on the value of the mean function at any point. This is likely not a perfect model. At the very least we expect no QPOs long before and after a transient. This implies that there is some connection between the mean function and the GP, which is not modelled using our framework.

The QPO model we propose is also inherently stationary as this is one of the requirements of the CELERITE package, and we can only model a narrowly limited set of non-stationary time series as we have shown in Sec. 7.1.4. QPOs that drift in frequency over long time scales can not be properly modelled and would require us to use a different framework of GP solvers. Once frequency shifts become too large, our QPO kernel is certain to fail. We can detect shifting frequencies by splitting the light curve into multiple segments that we analyse independently and compare the QPO frequency posteriors. For a more rigorous analysis, we would have to evaluate a more expensive non-stationary QPO model.

Astrophysical x-ray data is typically recorded by counting the number of photons that arrived at a detector in a specific time interval, a process that is by its nature Poissonian. Real data is often more complicated and may feature aspects of Gaussian and Poissonian properties. Meanwhile, GP modelling inherently assumes that the underlying data are Gaussian. While Poisson counting data are Gaussian to a good approximation if there are sufficient counts per bin, data with lower count rates are not correctly modelled by GPs. There are some approaches to deal with photon-counting data with low count rates. The easiest way is to apply a variance stabilising transform which makes the data approximately Gaussian with $\sigma = 1$ [40, 51]. However, these transformations are still far off for bins with just zero or one photon. Other methods like a sigmoidal Gaussian-Cox process treat Poisson data effectively as Gaussian by converting the data with a sigmoid function [34]. In any case, using these methods makes it harder to interpret the hyperparameters because they also will be transformed. Alternatively, we can also use the GPs ability to treat data that is not equally spaced. If we encounter Poissonian data that in parts has bins with few counts, we combine neighbouring bins with low counts.

By doing that we need to divide the combined bin by the number of bins we combined to account for the increased value.

We also assume that dead-time, i.e. the property of detectors that they are unable to detect a second photon for a short period after they counted the first photon, does not play a significant role in our data. Dead-time effects are particularly a problem for strongly flaring sources. Modelling dead-time in practice is difficult and requires sophisticated techniques such as simulation-based inference [145]. For this study, we are assuming that our data are not significantly influenced by dead-time effects.

7.1.6 Bayesian Inference with Gaussian Processes

Bayesian inference with GPs works much the same as the general framework that we have laid out in Sec. 1.1. Similar to Chap. 6, we are interested in distinguishing a combination of red noise and QPOs from pure red noise, given some mean model that is identical in both cases. Bayes factor calculations for model selection has normally not been considered viable (Refs. [88, 317] are notable exceptions). Alternative methods such as the Akaike or Bayesian information criterion can be computed much faster based on a maximum likelihood estimate. However, these metrics are less reliable if the posterior distribution is not Gaussian or features multiple modes.

The main motivation for using approximate methods is that these are computationally cheap to obtain. However, when we deal with CELERITE models, performing the necessary number of likelihood evaluations for full posterior and evidence calculations is relatively trivial. We find that nested sampling can be performed in $\mathcal{O}(\text{minutes})$ using CELERITE for a light curve with a few hundred bins when using the QPO GP model and a skewed Gaussian mean model. Furthermore, even beyond the class of CELERITE models, full Bayesian inference may be possible on many data sets with the use of efficient GP evaluation algorithms [94, 111, 121, 310] or the use of massively parallel inference [258].

Priors for GP parameters must be chosen carefully. Generally, we employ priors that are uniform in logarithm for amplitude a , inverse decay time c , and frequency f , since these prior spaces span several orders of magnitude. These priors can also be adjusted depending on the application. We must also consider the ranges of the priors for two reasons. First, the prior range impacts the evidence calculation; making the prior range overly large would suppress the evidence and disfavour the model. Second, there are parts of the parameter space that are not reasonable to sample in, e.g. amplitude parameters that are much too small for the given data points.

The prior range of the amplitude parameters a for red noise and QPO can be set fairly widely, though if there is indeed no QPO/red noise present in the data, we only obtain an upper limit below which we infer the prior distribution. To avoid this, we set the lower limit to be the same as the smallest σ_n , meaning that we assume that the QPO/red noise process has to be at least as strong as the white noise. This helps us to distinguish $k_{\text{qpo+rn}}$ and k_{rn} because $k_{\text{qpo+rn}}$ can no longer look like k_{rn} by having very low a_{qpo} values. Setting a generic upper limit on a is somewhat more difficult, but also has a smaller impact on the results. When we compare evidence for red noise against evidence for QPO plus red noise, a wider prior on the amplitude will disfavour the latter evidence more as it will contain a greater prior volume V_π . In case the prior is uniform, then the

prior can be extracted from the evidence integral

$$Z = \frac{1}{V_{\pi,\theta}} \int \mathcal{L}(d|\theta) d\theta \quad (7.18)$$

For example, in the case that we set the prior to be $0 < \ln a < 20$, which is much larger than any data set we expect to encounter, we have a QPO amplitude prior volume $\ln V_{\pi,\ln a} = \ln 20 \approx 3.0$. CELERITE is not numerically stable much above 30 because it encounters issues with the limited size of floats. However, such a wide prior is may not be reasonable since the possible amplitudes do not span this many magnitudes. Suppose we find that $0 < \ln a_{\text{qpo}} < 5$ is a more reasonable prior, then the associated prior volume is $\ln V_{\pi,\ln a} = \ln 5 \approx 1.6$. Overall, this will only make a difference of about 1.4 in $\ln Z$, which will not greatly affect the interpretation of our results.

The frequency parameter f should not be less than $1/T$, with T being the length of the data segment, and not be greater than the Nyquist frequency, which is half the sampling rate. This is assuming the data points are equally spaced.

The decay parameter c can in principle take on infinitely low values (e.g. for an undamped harmonic oscillation) which is rarely the case in real data sets of interest. We find that if we have a data set with a persistent, high-quality QPO, the c_{qpo} posterior will be consistent with zero and have some upper limit. Below some threshold then all values of c_{qpo} are effectively equivalent, and the $\ln c_{\text{qpo}}$ posterior will be flat in this regime. It is thus sensible to define a lower limit based on the segment length. We set the minimum of c to be $1/(10T)$, which means that a QPO or red noise kernel has to at least fall by $\approx 10\%$ throughout the data segment.

The upper limit on c requires more care and a distinction for the QPO and red noise case. In the red noise case, the upper limit should be not greater than the sampling frequency $f_s = 1/\Delta t$ of the data because otherwise, the kernel will decay to less than $1/e$ within Δt . At this point, the red noise model is not functionally different from a white noise model. Additionally, the kernel is functionally the same for even higher values of c_{rn} which manifests itself as a plateau in the posterior distribution of $\ln c_{\text{rn}}$. A similar argument can be applied for c_{qpo} , however, we do not consider the sampling frequency of the data, but the frequency of the QPO. In the limit of low QPO frequency, f_{qpo} and short decay time $1/c_{\text{qpo}}$ the QPO kernel is identical to a red noise kernel. Hence we employ the limit that the QPO kernel may only decay to $1/e$ within one period, which implies an upper limit of f_{qpo} on the prior. We implement this relation in practice using the constraint prior features in BILBY (see Chap. 3).

Finally, we also consider the priors of our analytic mean flare shapes. Again, these depend in general on the underlying data we want to analyse. For the constant offset, we apply we choose a prior uniform in logarithm that reaches from the minimum to the maximum value in the data set. The amplitude priors are uniform in logarithm between 10% and 120% of the difference between minimum and maximum of the data. The prior on the width parameter of our flares depends on the model. In general, we aim for the width parameter to have a uniform-in-logarithm prior that spans roughly from the time difference between two data points to the entire time range of the data. We also have to parametrise the peak time of the flare. This prior requires especially careful treatment because the mean model is degenerate under a permutation in the order of the flares. Not breaking this degeneracy makes the parameter space much harder to sample since we end up with n degenerate permutations for n flare components. For this, we use a

result from order statistics for uniform distributions. Assuming n flares are uniformly distributed throughout the interval. We label these flares in order from 0 to $n-1$ indexed by k . The $t_{0,k}$ prior is distributed according to a conditional beta distribution with $\alpha = 1$ and $\beta = n - k$ between the $t_{0,k-1}$ flare and the end of the interval [125]. Putting this together, we obtain

$$\pi(t_{0,k}|t_{0,k-1}) = (n - k) \left(1 - \frac{t_{0,k} - t_{0,k-1}}{t_{\max} - t_{0,k-1}} \right)^{n-k-1}, \quad (7.19)$$

where t_{\max} is the time of the last element of the light curve. For the special case of $k = 0$ the prior is not conditional and we have instead

$$\pi(t_{0,k=0}) = n \left(1 - \frac{t_{0,k=0} - t_{\min}}{t_{\max} - t_{\min}} \right)^{n-1}, \quad (7.20)$$

where t_{\min} is the first element of the light curve.

The priors on the parameters of non-stationary extension of the CELERITE model, t_{start} , and t_{end} , are chosen in the same way as in Eq. 7.19 and Eq. 7.20, but with $n = 2$.

7.2 Simulated Data

We begin validating our model on simulated data before we move on to real data. This ensures that we can be confident that our parameter estimates are robust and unbiased.

In the following we look at different configurations of the models we described earlier. We examine both the QPO kernel and our red noise kernel with different parameters. We also investigate what happens when we infer parameters with the wrong model, e.g. if we create red noise data and perform inference with $k_{\text{qpo+rn}}$.

Given a set of parameters, GP data is generated in two steps. First, we obtain the deterministic part of the data by evaluating the mean function. Second, we produce the stochastic process that arises due to the covariance between the data points. We create the stochastic process data by drawing a sample from the multivariate normal distribution that is associated with this covariance matrix. To do this in practise we use the `GP.sample()` method in CELERITE. Finally, we add the deterministic and stochastic parts to create the overall time series.

7.2.1 Percentile-Percentile Analysis

If the inference process is unbiased then the true value of the parameters has to be in the x -percentile of the posterior distribution $x\%$ of the time. This kind of test is called a percentile-percentile (PP) analysis and can be used to tune the analysis settings based on simulated data before the analysis is deployed on real data [85]. PP-tests fail if the analysis settings are not sufficient to sample the posterior without bias, e.g. if we use too few live points. This occurs if parameters are hard to sample. PP-tests should pass if we use the same model and priors for data creation and inference. Concretely, the PP-test finds a p-value for each parameter using a Kolmogorov–Smirnov test [196] to check if the fraction of events in a particular credible interval is drawn from a uniform distribution.

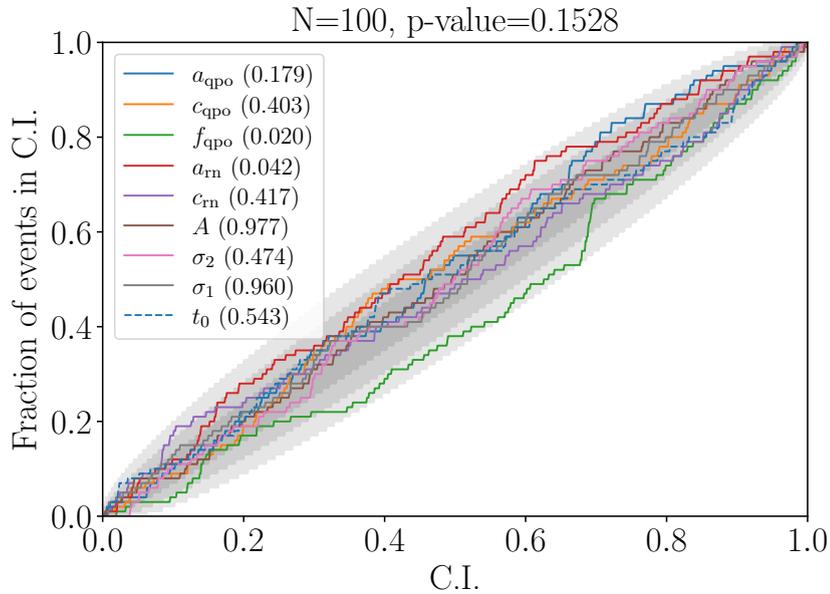


FIGURE 7.1: Percentile-percentile plot to show that our sampling methods are not biased. PP-plots are obtained using the respective function in BILBY [242]. We show the confidence interval (CI) on the horizontal axis, and what fraction of events have the true value within the CI on the vertical axis. The grey bands indicate the 1σ , 2σ , and 3σ -levels. The overall p-value of 0.1528 indicates that the observed deviations are consistent with randomness, though the individual relatively low values for f_{qpo} and a_{rn} indicate that these parameters may be harder to sample and may benefit from more finely tuned settings.

The PP-test combines these individual p-values to a combined p-value, e.g. using the `scipy.stats.combine_pvalues` function in `scipy` package. Conventionally, $p < 0.05$ indicates that the fraction of events is not uniformly distributed to the 2σ -level.

We perform the PP-test using the $k_{\text{qpo+rn}}$ model and a single skewed Gaussian as an arbitrary mean model. We create 100 simulated 1 s long data sets sampled with 256 time bins t_i that are randomly uniform distributed in time and using parameters randomly drawn from the priors in Tab. 7.1. We choose a random distribution of t_i instead of an equidistant one since some astrophysical light curves have gaps or changing sampling frequencies. Using a random distribution thus validates the method in the general case.

As we display in Fig. 7.1, using DYNESTY’s random walk sampling with 1500 live points is sufficient to sample the parameter space without significant biases in the posterior. The overall p-value of 0.1528 indicates that the observed deviations are consistent with randomness, though the individual relatively low values for f_{qpo} and a_{rn} indicate that these parameters may be harder to sample and may benefit from more finely tuned settings.

7.2.2 Model Selection

We want to have an understanding of how much the obtained Bayes factors fluctuate if we produce multiple noise realisations with the same set of parameters. GP simulations are

Parameter	Prior class	Minimum	Maximum
a_{rn}	LogUniform	$\exp(-1)$	$\exp(1)$
c_{rn}	LogUniform	$\exp(-1)$	$\exp(1)$
a_{qpo}	LogUniform	$\exp(-1)$	$\exp(1)$
c_{qpo}	LogUniform	$\exp(-1)$	$\exp(1)$
f_{qpo}	LogUniform	1	64
A	LogUniform	10	100
t_0	Uniform	0	1
σ_1	LogUniform	0.1	1
σ_2	LogUniform	0.1	1

TABLE 7.1: Priors for PP-test. The prior is used to randomly draw parameters for simulated data sets and during the inference process. There is also a prior constraint such that $c_{\text{qpo}} < f_{\text{qpo}}$, which we introduced in Sec. 7.1.6.

different from normal analytical models because noise realisations of the same parameters can indeed be vastly different.

For this study, we set the mean to be constant at zero. We create two different sets of simulated data, using k_{rn} and $k_{\text{qpo+rn}}$. Next, we use both kernel functions to calculate a $\ln BF_{\text{QPO}}$ for both sets of data. We expect that the $\ln BF$ should generally be positive if the data was produced using $k_{\text{qpo+rn}}$, and negative otherwise. We create the data on a 1 s interval sampled equidistantly at 256 Hz using the parameters listed in Tab. 7.2.

In Fig. 7.2 we show the result of performing this analysis for 1000 simulated data sets produced from either kernel. The data sets containing a QPO are about equally split between having a positive or negative $\ln BF_{\text{QPO}}$, but some data sets yield $\ln BF_{\text{QPO}} > 20$, indicating very high significance. This split is because QPO features are sometimes not pronounced enough to be recovered and k_{rn} is favoured due to its smaller prior volume. On the other hand, 92% of the data sets containing only red noise yield a $\ln BF_{\text{QPO}} < 0$, though the distribution is much narrower. This narrower distribution is likely because k_{rn} is a limiting case of $k_{\text{qpo+rn}}$. A data set created using k_{rn} thus is always fit well with $k_{\text{qpo+rn}}$, and the preference for the k_{rn} kernel model is achieved from the smaller prior volume. On the other hand, if there is significant oscillatory behaviour, k_{rn} can not provide a good fit.

If we chose a longer segment or a higher sampling frequency, the $\ln BF_{\text{QPO}}$ should increase at least linearly with the number of data points. For example, we consider one 2 s segment drawn from a GP. If we calculate a $\ln BF_{\text{QPO}}$ for the first and second half individually, it should approximately add up to the $\ln BF_{\text{QPO}}$ we obtain for the entire segment. In reality, there is some covariance between the two consecutive 1 s segments which further enhances our understanding of the data. Especially for the case of QPOs, having more oscillation periods is advantageous. Thus, all other parameters being equal, higher frequency QPOs should be more easily identifiable.

7.2.3 Non-stationarity Bias

Since the Whittle likelihood is directly derived from the assumption that the time series is a stationary GP, it is intuitive that the non-stationarity bias should also exist in GP likelihoods. GPs account for heteroscedasticity and deterministic trends, but the

Parameter	k_{rn} values	$k_{\text{qpo+rn}}$ values
a_{rn}	exp(1)	exp(1)
c_{rn}	exp(1)	exp(1)
a_{qpo}	-	exp(-2)
c_{qpo}	-	exp(1)
f_{qpo}	-	20

TABLE 7.2: Injection parameters for model selection study.

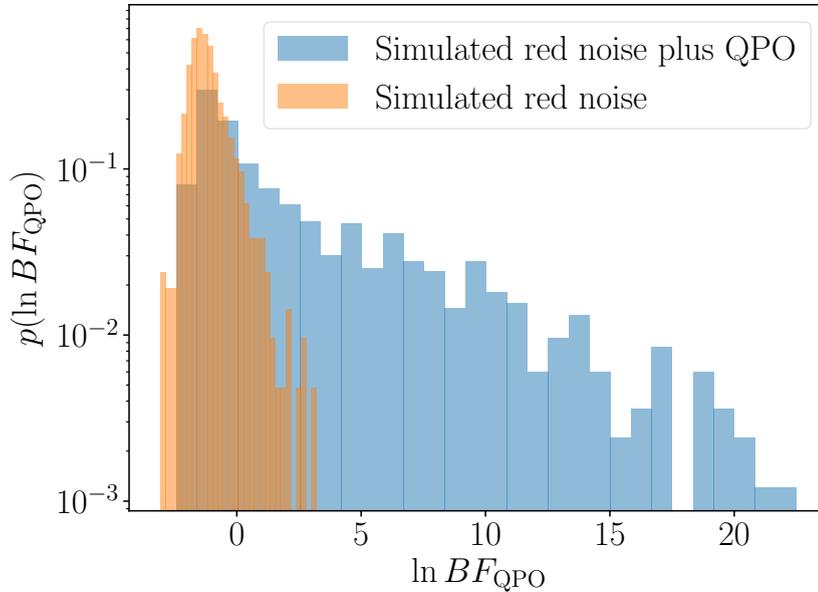


FIGURE 7.2: Normalised histogram of the $\ln BF_{\text{QPO}}$ obtained from data sets produced with identical $k_{\text{qpo+rn}}$ (blue) and k_{rn} (orange). Only 52.9% of the $k_{\text{qpo+rn}}$ data sets are correctly identified as such (The logarithmic scale is somewhat deceiving). This shows that there is substantial spread in terms of the possible Bayes Factors.

stationary kernels we are using can not account for non-stationary behaviour in the noise as we explored in Sec. 6.4.4 with periodograms. We laid out in Sec. 7.1.4 how we can create a simple, non-stationary GP model using CELERITE.

We apply the k_{rn} and $k_{\text{qpo+rn}}$ kernel on the data from Sec. 7.1.4 using both the stationary and the non-stationary GP model. Fig. 7.3 shows the maximum likelihood fit of the $x = 2$ time series using the non-stationary GP model. The $t_{\text{start/end}}$ parameters delineate where the transition between GP and white noise occurs. Because of this, the non-stationary GP yields a near-constant $\ln BF_{\text{QPO}}$ regardless of the extension factor, as we show in Fig. 7.4. On the other hand, the $\ln BF_{\text{QPO}}$ that we obtain for stationary GP models or the periodogram rises first due to non-stationarity bias and then falls off again as the QPO is overwhelmed by increasing white noise. The bias is stronger for the periodogram, though it is not clear why. Possibly, windowing choices made during data creation are having an impact on the inferred significance with the GP models.

As we show in Fig. 7.5 for $x = 2$, the two dimensional $t_{\text{start/end}}$ posterior is well constrained. We also note that these parameters have a peaky posterior distribution, making them possibly harder to sample. The posterior we infer for t_{start} is not consistent with

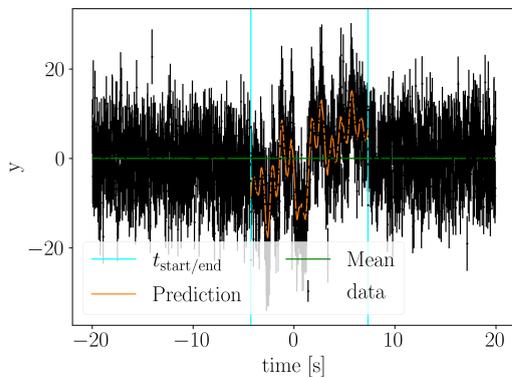


FIGURE 7.3: Maximum likelihood fit of the data (black) from Sec. 6.4.4 for $x = 2$ with a non-stationary GP model. We use a constant zero mean model (green). The teal lines indicate the maximum likelihood $t_{\text{start/end}}$ of the red noise and QPO GP. We underestimate the real duration of the GP which is 20 s. This is because we used a Hann window in the data creation which suppresses the amplitude closer to ± 10 s.

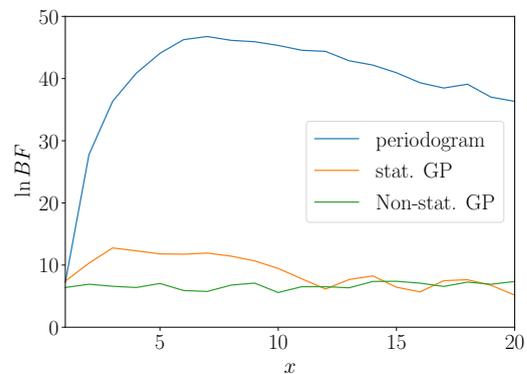


FIGURE 7.4: $\ln BF_{\text{QPO}}$ depending on the extension factor x using a periodogram (blue), a stationary GP model (orange), and the non-stationary extension to the GP model (green). While the periodogram and the stationary GP model are affected by the non-stationarity bias, the non-stationary GP model yields a near constant $\ln BF$.

the true start time of -10 s in Sec. 6.4.4. This is because the Hann window we applied when creating the data strongly suppresses red noise and QPO close to $t = \pm 10$ s.

Figure 7.6 shows the $\ln BF$ between the stationary and non-stationary $k_{\text{qpo+rn}}$ model. As we expect, the non-stationary model is preferred for any x and preferred more strongly for larger x .

7.3 Real data

7.3.1 Gamma-ray Bursts

There is ample speculation about the possibility for QPOs in long GRBs [195, 318], and there have been some recent claims about possible detections [274]. One of the most tantalising events is GRB090709A, a long GRB for which Ref. [74] found a marginal 2σ 8.06 s period QPO in the prompt emission. Independent analyses have since found similar significance levels between 2 and 3.5σ [96, 131, 160, 318]. GRB090709A had its strongest emission for about 100 s and a visible afterglow for several hundreds of seconds afterwards. Ref. [74] primarily relied on the *Swift* light curve for their analysis, though they also considered the *Suzaku* light curve which did not increase the significance. Further analyses of the *Swift* and *XMM-Newton* light curves by Ref. [92], who interpreted GRB090709A as a distant, standard, long GRB, also showed no periodicity above the 3σ -level. *Konus* and *SPI-ACS* also recorded light curves of GRB090709A, though it is not firmly established if the combined data of all instruments would yield higher significance [160]. References [74, 92] detrend the light curve before calculating the

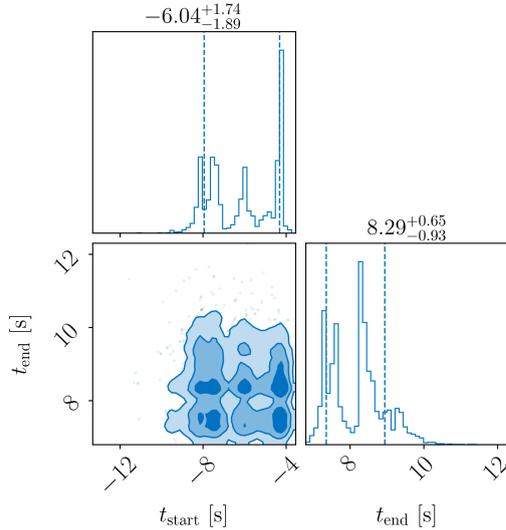


FIGURE 7.5: 2D $t_{\text{start}}/t_{\text{end}}$ posterior for $x = 2$. These parameters are harder to sample as they tend to be multi-modal.

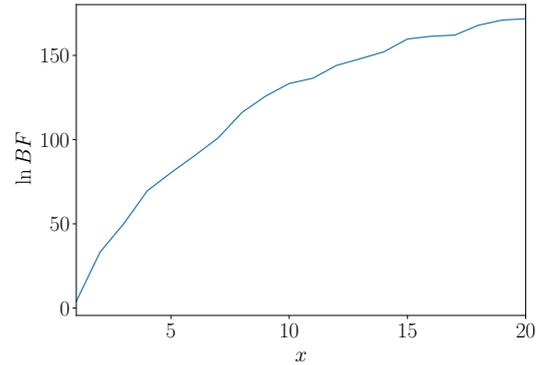


FIGURE 7.6: $\ln BF$ of the non-stationary GP model relative to the stationary GP model depending on the extension factor. The non-stationary model is preferred since it correctly models the change from red noise, white noise, and QPO to just white noise

periodogram to enhance the periodicity. However, Ref. [49] suggests that detrending is very likely to cause false detections of QPOs.

For our re-analysis of this event, we focus on the *Swift* light curve. We select 107s of the overall light curve as suggested by Ref. [74]. We carry out inference with all the models we defined earlier in Sec. 7.1.3, i.e. the skewed exponential and Gaussian models, as well as the FRED and FRED-x model, and allow a constant positive offset with all these models. Furthermore, we analyse the data with 1-3 flare components to see if there is a substantial improvement when using multiple components. As with our framework described in 7.1.2, we perform inference both with k_{rn} and $k_{\text{qpo+rn}}$ for Bayesian model selection.

Fig. 7.7 shows the considered data and a maximum likelihood fit of the mean function as well as the prediction curve generated by CELERITE. We note that there is a large spread in possible parameters in the mean model. Some posterior samples produce mean model curves that are almost consistently above the observed data points. These samples also have a high a_{rn} and low c_{rn} , i.e. an almost constant red noise kernel that does not decay much over the 107s light curve. These red noise kernels allow for a constant offset in the mean model of the magnitude a_{rn} without decreasing the likelihood significantly.

We display the obtained evidences in Fig. 7.8. We find generally that adding more components to our mean models does not meaningfully improve their evidences. This hints that already a single component model is a reasonable fit for the data. Furthermore, while there is some difference between the evidences between the individual mean models, the difference is relatively minor, but the FRED and FRED-x models yield somewhat lower evidences. Overall, we find that the $k_{\text{qpo+rn}}$ hypothesis is disfavoured or about equally favoured consistently in all runs with a $\ln BF_{\text{QPO}}$ between ≈ -3.5 and ≈ 0 .

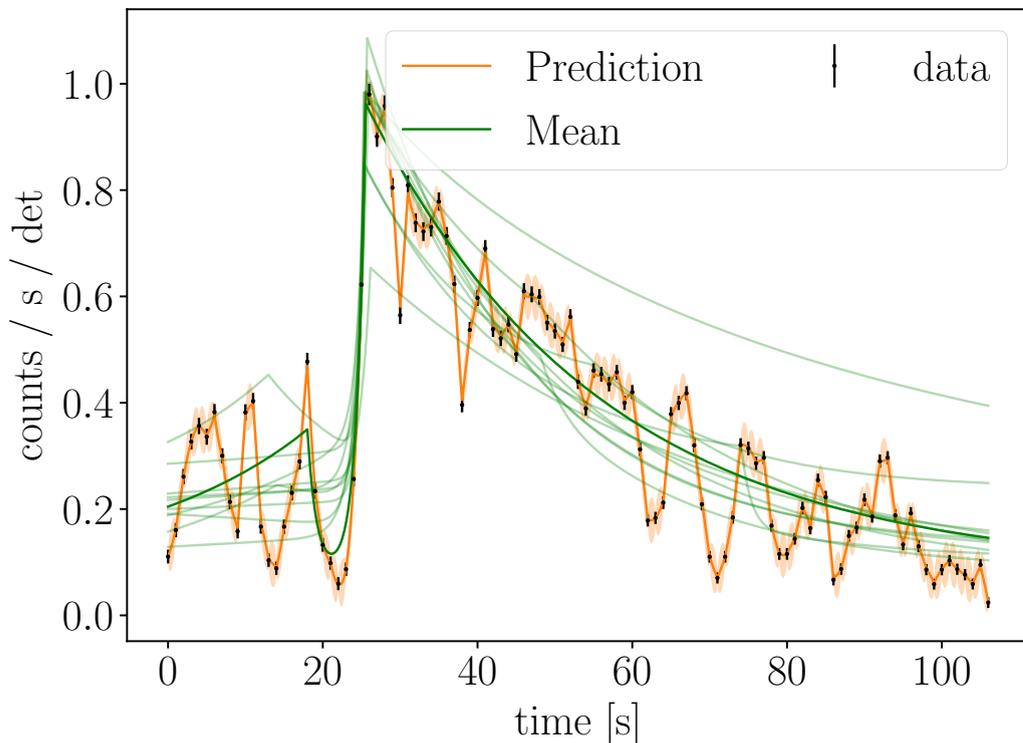


FIGURE 7.7: GRB090709A maximum likelihood fit using two skewed exponentials and the $k_{\text{qpo}+\text{rn}}$ kernel. We show the mean function from the maximum likelihood sample (dark green) and ten other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the $1\text{-}\sigma$ confidence band. Some mean model samples have a strong offset from the data points. This is due to samples with a high a_{rn} and a low c_{rn} .

On the other hand, we find that we obtain a well constrained $P = 1/f = 8.34^{+0.46}_{-0.38}$ (68%CI) posterior (see Fig. 7.9) that is consistent with the previously reported $P = 8.1$ s period [74, 92].

In the future, the combination of data from multiple observatories should also be considered, similar to the approach in Ref. [160]. The challenge is that instruments have different observing bands, different sensitivities, different responses, and are far apart from each other that the light curves need to be barycentred. We can not naively add the significances from different instruments via the addition of $\ln BF_{\text{QPO}}$ since red noise and QPO, unlike white noise, are covariant between different instruments. Data from different instruments should thus not be seen as statistically independent.

7.3.2 Magnetar bursts

Magnetar bursts show a strong bursting behaviour, ranging from series of low-energetic recurrent bursts to rare giant flares of which few have been observed. Some mechanisms may trigger a QPO in a magnetar burst, such as torsional Alfvén oscillations [182, 265], but the observational evidence remains scant. So far there have only been definitive

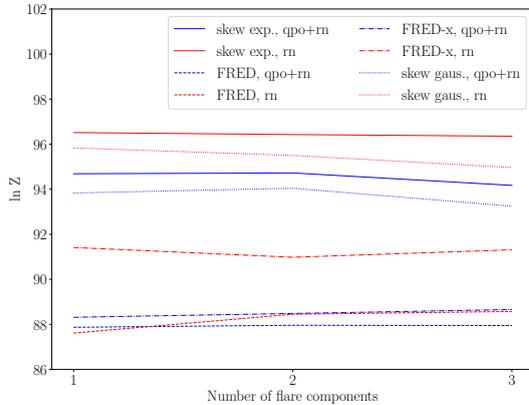


FIGURE 7.8: Model evidences for GRB090709A given different numbers of mean model flare components. We find that k_{rn} models are almost always favoured somewhat. All mean models yield similar evidences.

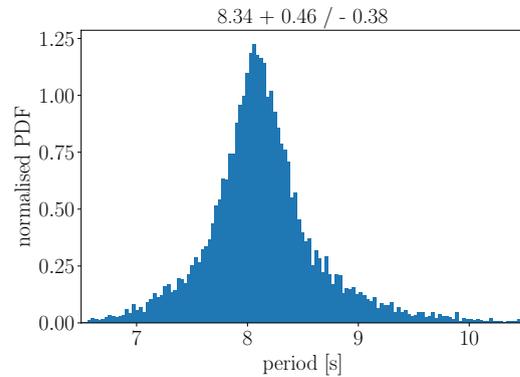


FIGURE 7.9: Period posterior for GRB090709A using two skewed exponentials. The period is well constrained $P = 8.34^{+0.46}_{-0.38}$ (68% CI) and consistent with the results from Refs. [74, 92], but has wide tails with low probability.

detections in the SGR1806-20 and SGR1900+14 Giant flares [159, 267, 268, 307] and in some recurring smaller bursts [146, 148, 149].

We demonstrate our method on a single magnetar burst 080823478 from SGR0501+4516, a magnetar that has been first discovered in 2008 by the *Swift* satellite [235]. As we can see in Fig. 7.10 the burst has a shape that, by eye, appears to be periodic, with 3 roughly equally spaced peaks after the main peak. Following the same steps we have used for the GRB, we find that k_{rn} is slightly preferred over k_{qpo+rn} as we show in Fig. 7.11. Moreover, the period posterior in Fig. 7.12 shows that we are also unable to consistently constrain f_{qpo} . This is in agreement with the conclusions of Ref. [148], which used this burst as a template for studies of red noise in magnetar bursts and also found no periodicity.

Alternatively, we also tested a constant mean model, meaning we model all variability as arising due to the GP. In that case, k_{qpo+rn} is strongly preferred ($\ln BF_{QPO} = 6.8$), but has a much lower evidence ($\ln Z = -382.1$) than any of the runs with mean models in Fig. 7.11. This highlights the benefit of using mean models as they help us to disentangle whether variability arises from the overall shape of the burst or due to a stochastic process.

7.3.3 Giant Magnetar Flare

The SGR1806-20 giant flare was recorded on December 27, 2004, by multiple x-ray observatories including the Rossi X-ray Timing Explorer (RXTE) [153, 218]. The flare, which lasted about 380 s, has been extensively studied for the presence of QPOs [147, 159, 203, 267, 268, 307]. Notably, QPOs were associated with specific phases in the 7.56 s rotational period [268], which indicate that QPOs are associated with a specific region on the magnetar surface that is turning in and out of view. Reference [268] specifically identified the most significant QPOs at 18 Hz, 26 Hz, 29 Hz, 93 Hz, 150 Hz, 625 Hz, and 1837 Hz. These early reported results were based on averaged periodograms that were obtained by selecting the same time interval in subsequent rotational phases

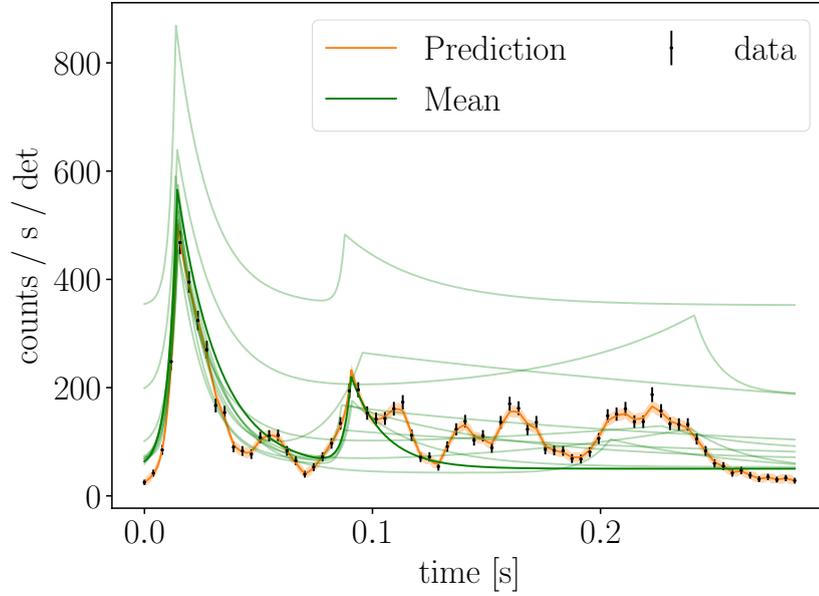


FIGURE 7.10: Magnetar burst 080823478 maximum likelihood fit using two skewed exponentials and the $k_{\text{qpo}+\text{rn}}$ kernel. We show the mean function from the maximum likelihood sample (dark green) and ten other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the $1\text{-}\sigma$ confidence band. Similarly to the fit of GRB090709 in Fig. 7.7, we see some mean fits that have a constant offset relative to the data due to their high a_{rn} and low c_{rn} .

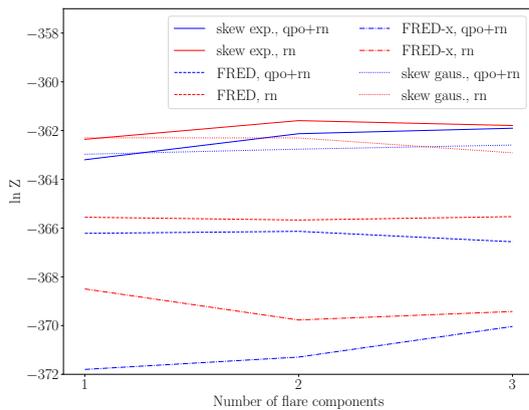


FIGURE 7.11: Model evidences for the 080823478 burst given different numbers of mean model components. We find that k_{rn} models are almost always favoured somewhat. All mean models yield similar evidences.

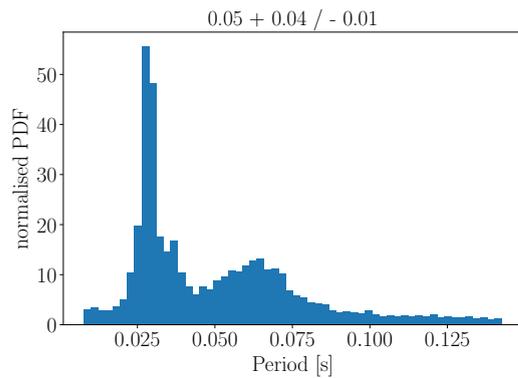


FIGURE 7.12: Period posterior for the 080823478 burst. The posterior is not constrained and has support across the entire prior space. This result is not representative of all runs we performed. Sometimes the period is more narrowly constrained, but $k_{\text{qpo}+\text{rn}}$ is still disfavoured.

of the magnetar. Naturally, these results are therefore prone to be affected by non-stationarity bias.

Ref. [203] took a more systematic approach in analysing the giant flare by using a 1 s sliding window that they moved across the flare light curve in 0.945 s steps. This way they were able to locate when specific QPOs were occurring, though the time resolution is limited to ≈ 1 s. They also used a periodogram-based analysis method, though unlike the method we described in Chap. 6, they used a different likelihood function that normalises the Poisson noise power and fixes the red noise power-law index to 2.

Given our findings of the non-stationarity bias in Chap. 6, identifying when QPOs start and end in the flare light curve is an important key to assessing their significance. We can use our non-stationary model from Sec. 7.1.4 to make quantitative statements for individual QPOs.

For this study, we focus on a specific 23 Hz QPO that was located in the RXTE light curve by Ref. [203] to occur about 122.060 s after the beginning of the light curve at December 27, 2004, 21:30:31.378 UTC. To see whether we can localise when the QPO occurs, we use the non-stationary model and select a 2 s segment starting at 121.060 s. Ref. [203] also reported a 92 Hz QPO to occur at the same segment. Thus, we choose a relatively coarse binning of the time-tagged events of 64 Hz so that we can focus on the lower frequency QPO. As we display in Fig. 7.13, we can constrain the time when the QPO occurs within the light curve. The inferred QPO mean frequency from the non-stationary model is $f_{\text{qpo}} = 22.60_{-0.74}^{+0.60}$ Hz (68% CI). We find that the QPO lasts for $1.20_{-0.18}^{+0.08}$ s (68% CI) and is primarily located on the top and tail side of the peak. We obtain $\ln BF_{\text{QPO}} = 4.9$ using the non-stationary GP model, and $\ln BF_{\text{QPO}} = 3.8$ using the stationary model, and that the non-stationary $k_{\text{qpo+rn}}$ is preferred over the stationary one with $\ln BF = 4.7$.

Next, we analyse this light curve using periodograms as we described in Chap. 6, using the same 2 s segment as before and also a 1 s segment starting at 122.060 s, which is the segment Ref. [203] considered. We bin the light curve at 4096 Hz as higher sampling frequencies help to constrain the white noise level in periodogram-based analyses. To avoid the posterior converging to a higher frequency QPO, we set the prior limit $f_{\text{qpo}} < 32$ Hz. Visual inspection of the periodogram shows that the QPO central frequency is located in the white-noise dominated part of the spectrum. Thus, we expect the non-stationarity bias to primarily occur due to the lower white noise levels on either end of the segment, as we explored in Sec. 6.9. We find $\ln BF_{\text{QPO}} = 6.0$ and $\ln BF_{\text{QPO}} = 9.6$ for the 1 s and 2 s segments, respectively. This result indicates that these $\ln BF_{\text{QPO}}$ are affected by non-stationarity bias and are likely exaggerated.

Using these observations we can also explain why the stationary GP has a lower $\ln BF_{\text{QPO}}$ than the non-stationary one. If the non-stationarity bias is mainly due to varying levels of white noise, then it is already accounted for by construction in the GP likelihood. What is remaining is the overall added white noise on the tails suppressing the significance of the QPO, yielding an overall lower $\ln BF_{\text{QPO}}$.

Expanding on this method by analysing more segments may allow us to explore the temporal structure of the QPOs in more detail. The specific segment we have considered contains one of the most significant QPOs in the entire flare. More marginal QPOs are harder to find and characterise. Eventually, hierarchical models may allow us to better understand the nature of QPOs in giant flares by taking a broader view of the entire

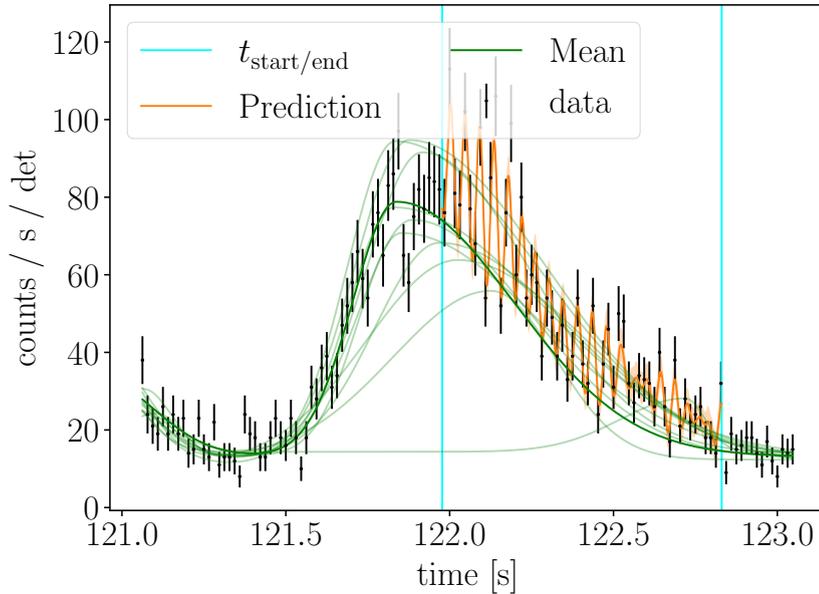


FIGURE 7.13: Giant flare maximum likelihood fit from our selected 2 s segment using two skewed Gaussians using the non-stationary $k_{\text{qpo}+\text{rn}}$ model. We show the mean function from the maximum likelihood sample (dark green) and ten other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the $1\text{-}\sigma$ confidence band.

400 s of data. Concretely, we may be able to constrain where QPOs in the pulse period start and end, and use information from multiple segments to find all frequency modes.

7.3.4 Solar Flares

Analysing solar flare data with GPs is much more difficult compared to the examples we have treated above for several reasons. Firstly, there is practically no white noise in the GOES data and the light curves have a much more smooth and structured overall shape. While this seems advantageous at first glance, it makes finding a mean model that fits the light curve reasonably much harder. Additionally, unlike the solar flare we analysed in Sec. 6.5, the QPO will be hard to spot by eye in most cases. Hence, it is near impossible to determine t_{start} and t_{end} either manually or using our non-stationary GP model.

We revisit the GOES solar flare from Sec. 6.5 and zoom in on the segment containing the QPO. The mean models we employed previously are unsuitable since we only fit a small segment of the flare. It is thus more sensible to use a linear mean function to fit the data. The GOES data also does not contain information about the white noise present. Thus, we add a `jitter_term` in CELERITE, which is a kernel term that is constant across the diagonal of the covariance matrix. We show the fit in Fig. 7.14 and obtain a $\ln BF_{\text{QPO}} = 23.5$ and a period $P = 12.35 \pm 0.18 \text{ s}$ (68% CI), which is similar to the $\ln BF_{\text{QPO}} = 27.5$ and $P = 12.38 \pm 0.22 \text{ s}$ (68% CI) that we obtained in Chap. 6 using periodograms.

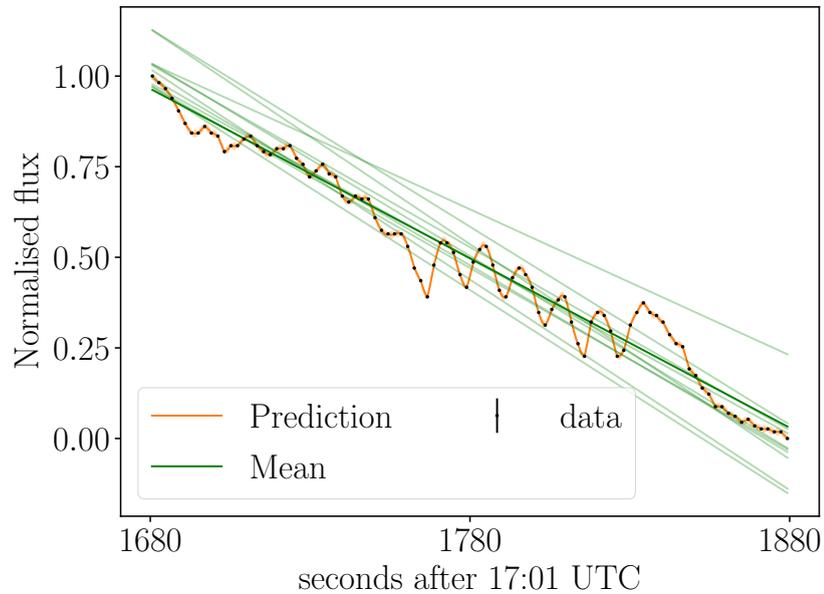


FIGURE 7.14: Maximum likelihood fit of the solar flare from May 12, 2013, observed by GOES using a linear mean model and the $k_{\text{qpo}+\text{rn}}$ kernel. We only select the 200 s segment that contains this QPO. We show the mean function from the maximum likelihood sample (dark green) and ten other samples from the posterior (light green). The orange curve is the prediction based on the maximum likelihood sample and the 1- σ confidence band.

A drawback of the GP model at this point is the limitation to the class of exponential mean models. Using an exponential kernel implies that we are assuming an $\alpha = 2$ power-law model while we found $\alpha = 3.10_{-0.48}^{+0.52}$ (90% CI) using the periodogram. The low-frequency noise in this case is also most likely to arise from a red noise process rather than the linear mean function. This may also explain the different $\ln BF_{\text{QPO}}$ that we obtained from the GP and periodogram methods. This issue can eventually be corrected by implementing kernel functions outside the CELERITE support, specifically with the Matérn class of covariance functions, which can emulate a wide class of power-law processes [233].

7.4 Discussion and Outlook

We have introduced a new method to search for and analyse QPOs in light curves using a combination of GP modelling and Bayesian inference with evidence calculation. Using this method we can directly model all aspects of the x-ray flare in the time domain. We use a set of phenomenological mean models to describe the overall flare shape and different GP models to model either red noise or a combination of red noise and QPOs. Using studies on simulated data we have shown that we can accurately estimate the true parameters, distinguish QPOs from red noise, model simple non-stationary GPs, and avoid some of the biases that non-stationary behaviour causes in Whittle-likelihood based methods. We have demonstrated that this framework can be easily applied with few modifications on many different astrophysical x-ray transients. Overall, the findings we obtain from real data are in agreement with results previously reported in the literature.

We find that the application of GP methods in the analysis of x-ray time series is thus very promising and many improvements and extensions can be implemented soon.

CELERITE limits us to a narrow class of kernel functions, which can be extended by using a more general GP framework with fast solvers. For example, the HODLR solver, which is implemented in the GEORGE package, operates in $\mathcal{O}(N \log^2 N)$ and thus can be run in acceptable time for the real data sets we have analysed [38]. Specifically, this would allow us to implement kernels such as the Matérn class of covariance functions or the squared exponential covariance function, which allows us to model a larger class of noise processes.

Chapter 8

Summary and future directions

I have explored how we can apply Bayesian inference techniques to two specific challenges in astrophysics: The measuring of gravitational-wave memory and identifying QPOs within x-ray light curves. We introduced the BILBY package in Chaps. 2 and 3, which is an object-oriented implementation of Bayesian inference in PYTHON and interfaces with several dedicated sampling packages such as DYNESTY [266]. The wide adoption of BILBY highlights the benefit of building versatile software for science applications.

8.1 Gravitational-wave Memory

As we have shown in Chaps. 4 and 5, LIGO/Virgo are not yet sensitive enough for memory to be detected. Using the events from GWTC-2, we have reached a $\ln BF = 0.024$ or $\ln BF = 0.049$ for the presence of memory depending on the waveform model. This should be considered extremely marginal evidence. We have designed the method to be scalable to a large number of events which will allow us to confidently measure memory in future LIGO/Virgo observing runs. Specifically, we are using BILBY and BILBY_PIPE to automate large parts of the analysis.

We have forecast that memory will be detectable by combining evidence from $\mathcal{O}(2000)$ events, a milestone likely reached early in the second part of the decade. This result is in broad agreement with results found by a similar study [64]. Thanks to its scalability, it is likely that the method we developed will be used to confidently measure memory for the first time. Since waveform models are certainly going to improve in the coming years, it will be necessary to implement these waveforms within the GWMEMORY and MEMESTR packages, which I describe in App. A. There is also the possibility that there will be surrogate models available which contain memory [205]. However, calculating memory from the time-domain spherical harmonic decomposition is likely to remain a widely used method since it allows us to calculate memory for generic waveforms. Whether or not memory will be measured using these surrogate models in the future will depend on their fidelity and whether these surrogate models end up being the preferred mode of analysis for future events. To perform model selection, it will also be necessary to have an option to switch memory contributions on and off.

Measuring memory from compact binary coalescences is a test of general relativity in the nonlinear regime. Moreover, the inclination angle dependence of the memory amplitude

can also be used as a test of general relativity [314]. Beyond that, there are some avenues to test theories using memory. Memory is significantly different in theories with more than four non-compactified spacetime dimensions [122, 141, 248]. In massive graviton theories, the memory amplitude depends on the mass of the graviton and is discretely different from general relativity [168]. Memory is also linked to soft gravitons and asymptotic symmetries in general relativity, which has implications for the black hole information paradox [135, 164, 270]. Measurements of memory with gravitational waves may eventually prove useful studying these phenomena, though, it is not yet clear how.

8.2 Gaussian Processes and Quasi-periodic Oscillations

In Chap. 6 we used Bayesian inference to investigate how non-stationarities in lightcurves impact searches for QPOs with periodogram-based methods. These methods commonly rely on the assumption that frequency bins within periodograms are statistically independent, which is only true for stationary time series. We have found that, counter to what would be intuitive, we may highly overestimate the significance of the QPO even if most of the time series does not contain a QPO, an effect we call the “non-stationarity bias”. Furthermore, we demonstrated that this bias is present at least in some analyses of solar flares. Since transient light curves are always non-stationary to some degree, the non-stationarity bias is likely present in many analyses. We also presented a method based on a modified χ^2 -statistic [212], with which we can identify non-stationarities just from the periodogram after we finish the inference step. It may be valuable to investigate how the non-stationarity bias affects other methods. For example, the Lomb-Scargle periodogram [187, 250, 291] is a commonly used technique to analyse light curves with unequally sampled data.

In Chap. 7 we explored how we can use Gaussian process modelling in conjunction with Bayesian inference as an alternative to periodogram-based methods. Gaussian processes immediately solve non-stationarity biases arising from deterministic trends and heteroscedasticity. Given N data points in a time series, Gaussian processes take in general $\mathcal{O}(N^3)$ steps in time and $\mathcal{O}(N^2)$ space in memory, which is generally considered to be too large to scale. However, using the CELERITE package [113], we can evaluate stationary Gaussian processes in $\mathcal{O}(N)$, by restricting ourselves to the class of complex exponential kernel functions or combinations thereof. We have shown that we can make a slight modification to the CELERITE model which allows us to evaluate some very simple non-stationary models. Using simulated data, we established that this method can reliably distinguish periodic and aperiodic time series. We showed that this framework can be applied to a large class of astrophysical transients, such as GRBs and magnetar flares, as well as solar flares. Gaussian processes can also natively deal with data that are not equally spaced, which is of great benefit for some data. Hence, we believe that Gaussian processes can become an established method for QPO detection and parametric inference in the future.

It will be valuable to apply this method to transients that have previously been characterised using different methods and to gain a better understanding of whether non-stationarities have biased past results. Additionally, it may be valuable to perform comparative studies with Gaussian processes and wavelet transforms. Specifically, creating

simulated data using Gaussian processes could help improve assessing how one should assess significance from wavelet transforms.

One major challenge is to develop and deploy models for non-stationary signals, e.g. a QPO that is continuously drifting in frequency. Aside from CELERITE, there are a number of fast solvers for GPs such as the $\mathcal{O}(N \log^2 N)$ HODLR solver [38], which is implemented in the GEORGE package¹. Additionally, many astrophysical lightcurves are Poisson distributed, which is only approximately Gaussian for sufficient count rates. These light curves may be better modelled using such methods as the Gaussian-Cox process, for which tractable likelihoods exist [34, 111].

¹The GEORGE repository is available at <https://github.com/dfm/george>

Appendix A

Software implementation for gravitational-wave memory

A.1 The GWMEMORY package

Ref. [273] presented the GWMEMORY package which implements the memory extraction by way of integrating the oscillatory $h_{\ell m}$ modes. To keep this process efficient, GWMEMORY splits the evaluation of the double integral in Eq. 1.9 into space and time which are described by the tensors. The spherical integral is pre-computed generically and summarised in a tensor $\Gamma_{\ell m}^{\ell_1 m_1 \ell_2 m_2}$, which links oscillatory input modes $(\ell_1, m_1, \ell_2, m_2)$ to memory output modes (ℓ, m) . During run-time, only the time integral is evaluated and written in terms of a tensor $H_{\ell m}^{\ell_1 m_1 \ell_2 m_2}(t_0, t_R)$ where t_0 is the start time of the oscillatory waveform. In terms of this formalism, we calculate the memory at time t_R as

$$\delta h_{\ell m}(t_R, t_0, \Omega) = \Gamma_{\ell m}^{\ell_1 m_1 \ell_2 m_2}(\Omega) H_{\ell m}^{\ell_1 m_1 \ell_2 m_2}(t_0, t_R). \quad (\text{A.1})$$

Pre-computing $\Gamma_{\ell m}^{\ell_1 m_1 \ell_2 m_2}$ turns out to be of crucial importance as this reduces the evaluation of a single memory waveform from the order of minutes down to a small fraction of a second.

For purposes of measuring memory, we implemented some changes in GWMEMORY that exist on a fork of the project¹. Primarily, this involved interfacing between GWMEMORY and more waveform models. The original implementation featured IMRPHENOMD [166], NRSUR7DQ2 [59], SEOBNRv4 [65], and the minimal waveform model [103]. In Chap. 4 we use the implementation of IMRPHENOMD and the newly implemented NRHYBSUR3DQ8 [293] to calculate memory for the first gravitational-wave transient catalogue. We expand and improve on this in Chap. 5 by implementing IMRPHENOMXHM [227] and NRSUR7DQ4 [292]. The reason for the addition of more waveform models is due to the evolving state of the art of waveform modelling and practical considerations for which waveforms are suitable to perform lengthy sampling processes. NRSUR7DQ2 only has mass ratios down to $q = 1/2$ and is relatively slow, SEOBNRv4 is also relatively expensive to evaluate. IMRPHENOMD on the other hand is faster but does not contain higher-order modes. Thus, for the study in Chap. 4, we implemented NRHYBSUR3DQ8,

¹<https://github.com/MoritzThomasHuebner/gwmemory>

which only models aligned spins but allows for mass ratios down to $q = 1/8$. We ultimately used a combination of IMRPHENOMD and NRHYBSUR3DQ8 as we were able to re-purpose the posterior samples obtained in Ref. [220]. During this, we also discovered and fixed some minor bugs in the original code that caused some higher-order memory modes to receive an incorrect sign. This was specifically due to an error in the spherical-harmonic mode calculation for which we now use the implementation in LAL [184]. Additionally, we modified the call to the spherical harmonic function to be with $\pi - \phi_c$ instead of ϕ to match the LAL convention. Finally, the waveforms in GWMEMORY are calculated in the time domain, whereas the analysis of gravitational waves is usually performed in the frequency domain. Thus, the time domain waveform we obtain must be suitable to be used in a fast Fourier transformation as it is implemented in standard inference software such as BILBY. This means that all waveforms have to be of equal length and contain 2^n elements. NRSUR7DQ4 and similar waveforms are defined to be equal length in terms of geometric time, which depends on the overall mass of the system, and thus do not provide us with the waveform in the desired format. We thus implemented a function that matches the waveform to a given time series given three cases:

1. The waveform is too long: we crop off data points from the inspiral until we fit the time series
2. The waveform is too short: we add zeros to the end of the waveform to match the time series
3. The waveform is the correct length: we return the waveform

A.2 The MEMESTR package

While GWMEMORY provides suitable time-domain waveforms, some steps are still required to perform actual inference. For this, I created the MEMESTR package² (MEMory ESTimator), which takes waveforms provided by GWMEMORY and performs model selection to calculate a Bayes factor for the presence of memory in the signal. Creating waveforms suitable for inference with MEMESTR turns out to be a somewhat tricky undertaking. Firstly, memory waveforms by their nature have a constant offset from zero at the end of the segment. Thus, we need to window the output by GWMEMORY with a Tukey window to avoid Fourier artefacts which would otherwise dominate our signal in the time domain [60, 61]. Next, the LAL convention for waveforms is such that the waveform segment ends with the merger as defined by the maximum point of $|h_+ - ih_\times|$, and we need to shift the correctly time-ordered waveform GWMEMORY provides to make that so. When we calculate combined memory and oscillatory waveforms we take special care at this point as the maximum of $|h_+ - ih_\times|$ needs to be calculated solely based on the oscillatory waveform. If we included the memory here, we would in effect shift the maximum slightly and thus create a discrepancy between the oscillatory, and the combined oscillatory plus memory waveform. Finally, we perform a fast Fourier transform to obtain the frequency domain. We used reference waveforms as provided directly from LAL to perform regression tests to assure that the oscillatory waveforms are identical up to machine precision.

²<https://github.com/MoritzThomasHuebner/memestr>

MEMESTR provides waveforms that are suitable for Bayesian inference with BILBY for IMRPHENOMD, IMRPHENOMXHM, the minimal-waveform model, NRHYBSUR3DQ8, NRSUR7DQ4, and TEOBRESUMS. The waveforms are implemented as a functional PYTHON interface in the `waveforms` subpackage within MEMESTR. We generally provide six functions per waveform model. Those are the (ordinary) oscillatory model, the memory part of the waveform, and the combined oscillatory plus memory model, both in the time and frequency domain. The minimal-waveform model does not have an associated oscillatory model, so we only provide the memory. For IMRPHENOMXHM, we provide a fast interface that does not calculate the modes when only evaluating the oscillatory polarisations. Additionally, we provide a function for IMRPHENOMXHM that allows us to input individual modes.

Scaling is another challenge in measuring memory in many events. We analyse almost 50 events in Chap. 5, which is a challenging task in its own right. One consideration was to use the publicly available posterior samples to calculate Bayes factors, however, we were unable to perfectly reconstruct the analysis settings used to obtain these samples. As memory is such a small effect, even slight differences in the oscillatory waveform can create an overwhelming systematic bias. Moreover, even slight differences in settings, such as reference frequencies, for instance, can completely change some extrinsic parameter estimates such as the phase. Additionally, some of the published posterior samples, as of the time of writing for Chap. 5, used settings that we found insufficient to fully explore the higher-order mode content of these events, and some events were not analysed with NRSUR7DQ4 (for no obvious reason). Instead, we opted for performing Bayesian inference ourselves, running most events with both the IMRPHENOMXHM and NRSUR7DQ4 waveform models. To achieve this scaling, we leverage the `BILBY_PIPE` package, which allows us to define and submit jobs by writing an `ini` file [242]. We also chose established settings for standard sampling runs, which are meant to ensure safe convergence and reasonable evaluation times [242]. We show an example listing for the analysis of GW150914 with NRSUR7DQ4 in Listing A.1. Though there is a long list of settings that we need to change from the defaults, it turns out that there are only seven settings that need to be adjusted for each event.

1. The trigger time
2. The event name
3. The detectors (e.g. `detectors = [H1, L1]`)
4. The duration of the data segment
5. The dictionary with the data channels
6. The dictionary with the PSD files
7. The reference frequency for GW190521 for the NRSUR7DQ4 analysis is 11 Hz instead of 20 Hz

Hence, we decided to automate the writing of `ini` files into a script named `write_ini_files.py`, which is in the `scripts/03a` folder in the MEMESTR project. The main information about all events is stored in `events.py` within the MEMESTR package and contains a comprehensive, and easy-to-edit list of all events with the necessary information to build set up the `ini` files. We obtained information about trigger time and signal

duration from the `ini` files provided in the public data release³. Nevertheless, managing a large number of runs remains a challenging task and all posteriors need to be reviewed to ensure proper convergence.

It is unlikely that we will ever scale up this approach to thousands of events, as this would hardly be manageable by a single person or a small team. Realistically, there are two options for future studies. First, we may want to restrict ourselves to the most promising events in terms of signal-to-noise ratio. As we discuss in Chaps. 4 and 5, only very few events will ever show a $\ln BF$ that is meaningfully different from zero. While we predict $\mathcal{O}(2000)$ events to be required, it will likely be possible to find memory by analysing the loudest 10 – 100 events within that ensemble. Alternatively, future catalogues may perhaps provide posterior samples with more consistent settings, use the same waveforms across most of the events, and provide better ways to reconstruct the waveform, and it might thus be feasible to use these posterior samples.

```

trigger-time = 1126259462.391
outdir = GW150914_prec_2000
detectors = [H1, L1]
duration = 4
channel-dict = {H1:GWOSC, L1:GWOSC}
psd-dict = {H1:GWTC2_PSDs/GW150914_LIGO_Hanford_psd.txt, L1:GWTC2_PSDs/
    ↪ GW150914_LIGO_Livingston_psd.txt}

label = run
accounting = ligo.dev.o3.cbc.pe.lalinference
scheduler = slurm
scheduler-args = mem=4G
scheduler-module = python/3.7.4
scheduler-analysis-time = 7-00:00:00
submit=True

coherence-test = False

sampling-frequency=2048
frequency-domain-source-model = memestr.waveforms.nrsur7dq4.fd_nr_sur_7dq4

calibration-model=None

minimum-frequency=0
reference-frequency=20
deltaT = 0.2
time-marginalization=False
distance-marginalization=False
phase-marginalization=False

prior-file = precessing_spin.prior

sampler = dynesty
sampler-kwargs = {nlive: 2000, sample: rwalk, walks=100, n_check_point=2000, nact
    ↪ =10, resume=True}

n-parallel = 5

create-plots=True
local-generation = True
transfer-files = False
periodic-restart-time = 1209600

```

LISTING A.1: `ini` file example for GW150914

³The catalogue is online available at <https://git.ligo.org/publications/03/o3a-cbc-catalog/-/wikis/home>

Bibliography

- [1] Aasi, J., Abbott, B. P., Abbott, R., et al. 2015, Classical and Quantum Gravity, 32, 74001, doi: [10.1088/0264-9381/32/7/074001](https://doi.org/10.1088/0264-9381/32/7/074001)
- [2] Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Physical Review Letters, 116, 061102, doi: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102)
- [3] —. 2016, Physical Review Letters, 116, 241102, doi: [10.1103/PhysRevLett.116.241102](https://doi.org/10.1103/PhysRevLett.116.241102)
- [4] —. 2016, Physical Review X, 6, 041014, doi: [10.1103/PhysRevX.6.041014](https://doi.org/10.1103/PhysRevX.6.041014)
- [5] —. 2016, Physical Review X, 6, 041015, doi: [10.1103/PhysRevX.6.041015](https://doi.org/10.1103/PhysRevX.6.041015)
- [6] —. 2016, Physical Review Letters, 116, 221101, doi: [10.1103/PhysRevLett.116.221101](https://doi.org/10.1103/PhysRevLett.116.221101)
- [7] —. 2016, Physical Review Letters, 116, 241103, doi: [10.1103/PhysRevLett.116.241103](https://doi.org/10.1103/PhysRevLett.116.241103)
- [8] —. 2016, Physical Review Letters, 116, 131103, doi: [10.1103/PhysRevLett.116.131103](https://doi.org/10.1103/PhysRevLett.116.131103)
- [9] —. 2017, Classical and Quantum Gravity, 34, 044001, doi: [10.1088/1361-6382/aa51f4](https://doi.org/10.1088/1361-6382/aa51f4)
- [10] —. 2017, The Astrophysical Journal, 848, L13, doi: [10.3847/2041-8213/aa920c](https://doi.org/10.3847/2041-8213/aa920c)
- [11] —. 2017, Nature, 551, 85, doi: [10.1038/nature24471](https://doi.org/10.1038/nature24471)
- [12] —. 2017, Physical Review Letters, 119, 141101, doi: [10.1103/PhysRevLett.119.141101](https://doi.org/10.1103/PhysRevLett.119.141101)
- [13] —. 2017, Physical Review Letters, 118, 221101, doi: [10.1103/PhysRevLett.118.221101](https://doi.org/10.1103/PhysRevLett.118.221101)
- [14] —. 2017, The Astrophysical Journal, 851, L35, doi: [10.3847/2041-8213/aa9f0c](https://doi.org/10.3847/2041-8213/aa9f0c)
- [15] —. 2017, The Astrophysical Journal, 851, L16, doi: [10.3847/2041-8213/aa9a35](https://doi.org/10.3847/2041-8213/aa9a35)
- [16] —. 2018, Physical Review Letters, 121, 161101, doi: [10.1103/PhysRevLett.121.161101](https://doi.org/10.1103/PhysRevLett.121.161101)
- [17] —. 2018, Physical Review Letters, 120, 031104, doi: [10.1103/PhysRevLett.120.031104](https://doi.org/10.1103/PhysRevLett.120.031104)

- [18] —. 2018, Physical Review Letters, 120, 201102, doi: [10.1103/PhysRevLett.120.201102](https://doi.org/10.1103/PhysRevLett.120.201102)
- [19] —. 2018, Living Reviews in Relativity, 21, 3, doi: [10.1007/s41114-018-0012-9](https://doi.org/10.1007/s41114-018-0012-9)
- [20] —. 2019, Physical Review X, 9, 031040, doi: [10.1103/PhysRevX.9.031040](https://doi.org/10.1103/PhysRevX.9.031040)
- [21] —. 2019, Physical Review X, 9, 011001, doi: [10.1103/PhysRevX.9.011001](https://doi.org/10.1103/PhysRevX.9.011001)
- [22] —. 2019, The Astrophysical Journal, 882, L24, doi: [10.3847/2041-8213/ab3800](https://doi.org/10.3847/2041-8213/ab3800)
- [23] —. 2019, Physical Review Letters, 123, 011102, doi: [10.1103/PhysRevLett.123.011102](https://doi.org/10.1103/PhysRevLett.123.011102)
- [24] —. 2019, Physical Review D, 100, 104036, doi: [10.1103/PhysRevD.100.104036](https://doi.org/10.1103/PhysRevD.100.104036)
- [25] Abbott, R., Abbott, T. D., Abraham, S., et al. 2020, Physical Review D, 102, 043015, doi: [10.1103/PhysRevD.102.043015](https://doi.org/10.1103/PhysRevD.102.043015)
- [26] —. 2020, The Astrophysical Journal, 896, L44, doi: [10.3847/2041-8213/ab960f](https://doi.org/10.3847/2041-8213/ab960f)
- [27] —. 2020, The Astrophysical Journal, 900, L13, doi: [10.3847/2041-8213/aba493](https://doi.org/10.3847/2041-8213/aba493)
- [28] —. 2020, Physical Review Letters, 125, 101102, doi: [10.1103/PhysRevLett.125.101102](https://doi.org/10.1103/PhysRevLett.125.101102)
- [29] —. 2021, Physical Review X, 11, 021053, doi: [10.1103/PhysRevX.11.021053](https://doi.org/10.1103/PhysRevX.11.021053)
- [30] —. 2021, The Astrophysical Journal Letters, 913, L7, doi: [10.3847/2041-8213/abe949](https://doi.org/10.3847/2041-8213/abe949)
- [31] —. 2021, The Astrophysical Journal Letters, 915, L5, doi: [10.3847/2041-8213/ac082e](https://doi.org/10.3847/2041-8213/ac082e)
- [32] Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, Classical and Quantum Gravity, 32, 024001, doi: [10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001)
- [33] Acernese, F., Agathos, M., Aiello, L., et al. 2019, Physical Review Letters, 123, 231108, doi: [10.1103/PhysRevLett.123.231108](https://doi.org/10.1103/PhysRevLett.123.231108)
- [34] Adams, R. P., Murray, I., & MacKay, D. J. 2009, Proceedings of the 26th International Conference On Machine Learning, ICML 2009, 9
- [35] Aggarwal, K., Arzoumanian, Z., Baker, P. T., et al. 2020, The Astrophysical Journal, 889, 38, doi: [10.3847/1538-4357/ab6083](https://doi.org/10.3847/1538-4357/ab6083)
- [36] Alam, M. F., Arzoumanian, Z., Baker, P. T., et al. 2020, The Astrophysical Journal Supplement Series, 252, 4, doi: [10.3847/1538-4365/abc6a0](https://doi.org/10.3847/1538-4365/abc6a0)
- [37] Amaro-Seoane, P., Audley, H., Babak, S., et al. 2017. <https://arxiv.org/abs/1702.00786>
- [38] Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2016, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252, doi: [10.1109/TPAMI.2015.2448083](https://doi.org/10.1109/TPAMI.2015.2448083)

- [39] Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, *Monthly Notices of the Royal Astronomical Society*, 474, 2094, doi: [10.1093/mnras/stx2109](https://doi.org/10.1093/mnras/stx2109)
- [40] Anscombe, F. J. 1948, *Biometrika*, 35, 246, doi: [10.2307/2332343](https://doi.org/10.2307/2332343)
- [41] Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2015, *The Astrophysical Journal*, 810, 150, doi: [10.1088/0004-637X/810/2/150](https://doi.org/10.1088/0004-637X/810/2/150)
- [42] Ashton, G., Lasky, P. D., Graber, V., & Palfreyman, J. 2019, *Nature Astronomy*, 3, 1143, doi: [10.1038/s41550-019-0844-6](https://doi.org/10.1038/s41550-019-0844-6)
- [43] Ashton, G., & Talbot, C. 2021, *Monthly Notices of the Royal Astronomical Society*, 16, 1, doi: [10.1093/mnras/stab2236](https://doi.org/10.1093/mnras/stab2236)
- [44] Ashton, G., Burns, E., Canton, T. D., et al. 2018, *The Astrophysical Journal*, 860, 6, doi: [10.3847/1538-4357/aabfd2](https://doi.org/10.3847/1538-4357/aabfd2)
- [45] Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, *The Astrophysical Journal Supplement Series*, 241, 27, doi: [10.3847/1538-4365/ab06fc](https://doi.org/10.3847/1538-4365/ab06fc)
- [46] Ashton, G., et al. 2021, <https://git.ligo.org/lscsoft/bilby/>
- [47] —. 2021, <https://lscsoft.docs.ligo.org/bilby/>
- [48] Aso, Y., Michimura, Y., Somiya, K., et al. 2013, *Physical Review D*, 88, 043007, doi: [10.1103/PhysRevD.88.043007](https://doi.org/10.1103/PhysRevD.88.043007)
- [49] Auchère, F., Froment, C., Bocchialini, K., Buchlin, E., & Solomon, J. 2016, *The Astrophysical Journal*, 825, 110, doi: [10.3847/0004-637X/825/2/110](https://doi.org/10.3847/0004-637X/825/2/110)
- [50] Baiotti, L., Giacomazzo, B., & Rezzolla, L. 2008, *Physical Review D*, 78, 084033, doi: [10.1103/PhysRevD.78.084033](https://doi.org/10.1103/PhysRevD.78.084033)
- [51] Bar-Lev, S. K., & Enis, P. 1990, *Statistics and Probability Letters*, 10, 95, doi: [10.1016/0167-7152\(90\)90002-0](https://doi.org/10.1016/0167-7152(90)90002-0)
- [52] Barat, C., Hayles, R., Hurley, K., et al. 1983, *Astronomy and Astrophysics*, 126, 400. <http://www.nature.com/articles/16199>
- [53] Barbary, K. 2021, *Nestle*, doi: <https://github.com/kbarbary/nestle>
- [54] Barret, D., & Vaughan, S. 2012, *Astrophysical Journal*, 746, 131, doi: [10.1088/0004-637X/746/2/131](https://doi.org/10.1088/0004-637X/746/2/131)
- [55] Bieri, L., Garfinkle, D., & Yau, S. T. 2016, *Physical Review D*, 94, 1, doi: [10.1103/PhysRevD.94.064040](https://doi.org/10.1103/PhysRevD.94.064040)
- [56] Bieri, L., Garfinkle, D., & Yunes, N. 2017, *Classical and Quantum Gravity*, 34, 1, doi: [10.1088/1361-6382/aa8b52](https://doi.org/10.1088/1361-6382/aa8b52)
- [57] Biwer, C. M., Capano, C. D., De, S., et al. 2019, *Publications of the Astronomical Society of the Pacific*, 131, 024503, doi: [10.1088/1538-3873/aaef0b](https://doi.org/10.1088/1538-3873/aaef0b)
- [58] Blackman, J., Field, S. E., Galley, C. R., et al. 2015, *Physical Review Letters*, 115, 121102, doi: [10.1103/PhysRevLett.115.121102](https://doi.org/10.1103/PhysRevLett.115.121102)

- [59] Blackman, J., Field, S. E., Scheel, M. A., et al. 2017, *Physical Review D*, 96, 28, doi: [10.1103/PhysRevD.96.024058](https://doi.org/10.1103/PhysRevD.96.024058)
- [60] Blackman, R. B., & Tukey, J. W. 1958, *The Bell System Technical Journal*, 37, 185, doi: [10.1002/j.1538-7305.1958.tb03874.x](https://doi.org/10.1002/j.1538-7305.1958.tb03874.x)
- [61] —. 1958, *Bell System Technical Journal*, 37, 485, doi: [10.1002/j.1538-7305.1958.tb01530.x](https://doi.org/10.1002/j.1538-7305.1958.tb01530.x)
- [62] Blanchet, L., & Damour, T. 1992, *Physical Review D*, 46, 4304, doi: [10.1103/PhysRevD.46.4304](https://doi.org/10.1103/PhysRevD.46.4304)
- [63] Blanchet, L., Faye, G., Iyer, B. R., & Sinha, S. 2008, *Classical and Quantum Gravity*, 25, 165003, doi: [10.1088/0264-9381/25/16/165003](https://doi.org/10.1088/0264-9381/25/16/165003)
- [64] Boersma, O. M., Nichols, D. A., & Schmidt, P. 2020, *Physical Review D*, 101, 83026, doi: [10.1103/PhysRevD.101.083026](https://doi.org/10.1103/PhysRevD.101.083026)
- [65] Bohé, A., Shao, L., Taracchini, A., et al. 2017, *Physical Review D*, 95, 044028, doi: [10.1103/PhysRevD.95.044028](https://doi.org/10.1103/PhysRevD.95.044028)
- [66] Bonferroni, C. 1936, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3
- [67] Bose, S., Chakravarti, K., Rezzolla, L., Sathyaprakash, B. S., & Takami, K. 2018, *Physical Review Letters*, 120, 031102, doi: [10.1103/PhysRevLett.120.031102](https://doi.org/10.1103/PhysRevLett.120.031102)
- [68] Braginskii, V. B., & Grishchuk, L. P. 1985, *Journal of Experimental and Theoretical Physics*, 62, 427 . http://jetp.ac.ru/cgi-bin/dn/e_062_03_0427.pdf%0Ahttp://www.jetp.ac.ru/cgi-bin/e/index/e/62/3/p427?a=list
- [69] Braginsky, V. B., & Thorne, K. S. 1987, *Nature*, 327, 123, doi: [10.1038/327123a0](https://doi.org/10.1038/327123a0)
- [70] Broomhall, A. M., Davenport, J. R., Hayes, L. A., et al. 2019, arXiv, doi: [10.3847/1538-4365/ab40b3](https://doi.org/10.3847/1538-4365/ab40b3)
- [71] Brown, D., & Freise, A. 2014, *Finesse*, doi: [10.5281/zenodo.821363](https://doi.org/10.5281/zenodo.821363)
- [72] Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *Astronomy and Astrophysics*, 564, A125, doi: [10.1051/0004-6361/201322971](https://doi.org/10.1051/0004-6361/201322971)
- [73] Buikema, A., Cahillane, C., Mansell, G. L., et al. 2020, *Physical Review D*, 102, 062003, doi: [10.1103/PhysRevD.102.062003](https://doi.org/10.1103/PhysRevD.102.062003)
- [74] Cenko, S. B., Butler, N. R., Ofek, E. O., et al. 2010, *Astronomical Journal*, 140, 224, doi: [10.1088/0004-6256/140/1/224](https://doi.org/10.1088/0004-6256/140/1/224)
- [75] Chatfield, C. 1975, *The Analysis of Time Series: Theory and Practice*. (Chapman & Hall ; distributed by Halsted Press)
- [76] Chatzioannou, K., Clark, J. A., Bauswein, A., et al. 2017, *Physical Review D*, 96, 124035, doi: [10.1103/PhysRevD.96.124035](https://doi.org/10.1103/PhysRevD.96.124035)
- [77] Cho, H., Macquart, J.-P., Shannon, R. M., et al. 2020, *The Astrophysical Journal*, 891, L38, doi: [10.3847/2041-8213/ab7824](https://doi.org/10.3847/2041-8213/ab7824)
- [78] Chopin, N., & Robert, C. P. 2010, *Biometrika*, 97, 741, doi: [10.1093/biomet/asq021](https://doi.org/10.1093/biomet/asq021)

- [79] Christodoulou, D. 1991, *Physical Review Letters*, 67, 1486, doi: [10.1103/PhysRevLett.67.1486](https://doi.org/10.1103/PhysRevLett.67.1486)
- [80] Clark, J., Bauswein, A., Cadonati, L., et al. 2014, *Physical Review D*, 90, 062004, doi: [10.1103/PhysRevD.90.062004](https://doi.org/10.1103/PhysRevD.90.062004)
- [81] Clark, J. A., Bauswein, A., Stergioulas, N., & Shoemaker, D. 2016, *Classical and Quantum Gravity*, 33, 085003, doi: [10.1088/0264-9381/33/8/085003](https://doi.org/10.1088/0264-9381/33/8/085003)
- [82] Colaiuda, A., Beyer, H., & Kokkotas, K. D. 2009, *MNRAS*, 396, 1441, doi: [10.1111/j.1365-2966.2009.14878.x](https://doi.org/10.1111/j.1365-2966.2009.14878.x)
- [83] Colaiuda, A., & Kokkotas, K. D. 2011, *MNRAS*, 414, 3014, doi: [10.1111/j.1365-2966.2011.18602.x](https://doi.org/10.1111/j.1365-2966.2011.18602.x)
- [84] Colleoni, M., Mateu-Lucena, M., Estellés, H., et al. 2021, *Physical Review D*, 103, 024029, doi: [10.1103/PhysRevD.103.024029](https://doi.org/10.1103/PhysRevD.103.024029)
- [85] Cook, S. R., Gelman, A., & Rubin, D. B. 2006, *Journal of Computational and Graphical Statistics*, 15, 675, doi: [10.1198/106186006X136976](https://doi.org/10.1198/106186006X136976)
- [86] Cordes, J. M., & Jenet, F. A. 2012, *The Astrophysical Journal*, 752, 54, doi: [10.1088/0004-637X/752/1/54](https://doi.org/10.1088/0004-637X/752/1/54)
- [87] Cordes, J. M., McLaughlin, M. A., & Collaboration, T. N. 2019, *Bulletin of the AAS*, 51. <https://arxiv.org/abs/1903.08653>
- [88] Covino, S., Landoni, M., Sandrinelli, A., & Treves, A. 2020, *The Astrophysical Journal*, 895, 122, doi: [10.3847/1538-4357/ab8bd4](https://doi.org/10.3847/1538-4357/ab8bd4)
- [89] Crowder, J., & Cornish, N. J. 2005, *Physical Review D*, 72, 083005, doi: [10.1103/PhysRevD.72.083005](https://doi.org/10.1103/PhysRevD.72.083005)
- [90] Damour, T., & Vilenkin, A. 2000, *Physical Review Letters*, 85, 3761, doi: [10.1103/PhysRevLett.85.3761](https://doi.org/10.1103/PhysRevLett.85.3761)
- [91] De, S., Finstad, D., Lattimer, J. M., et al. 2018, *Physical Review Letters*, 121, 091102, doi: [10.1103/PhysRevLett.121.091102](https://doi.org/10.1103/PhysRevLett.121.091102)
- [92] De Luca, A., Esposito, P., Israel, G. L., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 402, 1870, doi: [10.1111/j.1365-2966.2009.16012.x](https://doi.org/10.1111/j.1365-2966.2009.16012.x)
- [93] De Souza, G. H., & Chirenti, C. 2019, *Physical Review D*, 100, 43017, doi: [10.1103/PhysRevD.100.043017](https://doi.org/10.1103/PhysRevD.100.043017)
- [94] Delbridge, I. A., Bindel, D. S., & Wilson, A. G. 2019, arXiv. <https://arxiv.org/abs/1912.12834>
- [95] Dewdney, P., Hall, P., Schilizzi, R., & Lazio, T. 2009, *Proceedings of the IEEE*, 97, 1482, doi: [10.1109/JPROC.2009.2021005](https://doi.org/10.1109/JPROC.2009.2021005)
- [96] Dichiarà, S., Guidorzi, C., Frontera, F., & Amati, L. 2013, *Astrophysical Journal*, 777, doi: [10.1088/0004-637X/777/2/132](https://doi.org/10.1088/0004-637X/777/2/132)
- [97] Easter, P. J., Lasky, P. D., Casey, A. R., Rezzolla, L., & Takami, K. 2019, *Physical Review D*, 100, 043005, doi: [10.1103/PhysRevD.100.043005](https://doi.org/10.1103/PhysRevD.100.043005)

- [98] Farah, W., Flynn, C., Bailes, M., et al. 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 2989, doi: [10.1093/mnras/stz1748](https://doi.org/10.1093/mnras/stz1748)
- [99] Farr, B., Holz, D. E., & Farr, W. M. 2018, *The Astrophysical Journal*, 854, L9, doi: [10.3847/2041-8213/aaa64](https://doi.org/10.3847/2041-8213/aaa64)
- [100] Farr, B., Ochsner, E., Farr, W. M., & O'Shaughnessy, R. 2014, *Physical Review D*, 90, 024018, doi: [10.1103/PhysRevD.90.024018](https://doi.org/10.1103/PhysRevD.90.024018)
- [101] Farr, W. M., Farr, B., & Littenberg, T. 2014
- [102] Favata, M. 2009, *Physical Review D*, 80, 024002, doi: [10.1103/PhysRevD.80.024002](https://doi.org/10.1103/PhysRevD.80.024002)
- [103] —. 2009, *The Astrophysical Journal*, 696, L159, doi: [10.1088/0004-637X/696/2/L159](https://doi.org/10.1088/0004-637X/696/2/L159)
- [104] —. 2010, *Classical and Quantum Gravity*, 27, 084036, doi: [10.1088/0264-9381/27/8/084036](https://doi.org/10.1088/0264-9381/27/8/084036)
- [105] —. 2011, *Physical Review D*, 84, 124013, doi: [10.1103/PhysRevD.84.124013](https://doi.org/10.1103/PhysRevD.84.124013)
- [106] Feroz, F., & Hobson, M. P. 2007, *Monthly Notices of the Royal Astronomical Society*, 384, 449, doi: [10.1111/j.1365-2966.2007.12353.x](https://doi.org/10.1111/j.1365-2966.2007.12353.x)
- [107] Feroz, F., Hobson, M. P., & Bridges, M. 2009, *Monthly Notices of the Royal Astronomical Society*, 398, 1601, doi: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x)
- [108] Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, *The Open Journal of Astrophysics*, 2, 1, doi: [10.21105/astro.1306.2144](https://doi.org/10.21105/astro.1306.2144)
- [109] Fishbach, M., & Holz, D. E. 2017, *The Astrophysical Journal*, 851, L25, doi: [10.3847/2041-8213/aa9bf6](https://doi.org/10.3847/2041-8213/aa9bf6)
- [110] Flanagan, É. É., & Hinderer, T. 2008, *Physical Review D*, 77, 021502, doi: [10.1103/PhysRevD.77.021502](https://doi.org/10.1103/PhysRevD.77.021502)
- [111] Flaxman, S., Wilson, A. G., Neill, D. B., Nickisch, H., & Smola, A. J. 2015, 32nd International Conference on Machine Learning, ICML 2015, 1, 607
- [112] Foreman-Mackey, D. 2016, *The Journal of Open Source Software*, 1, 24, doi: [10.21105/joss.00024](https://doi.org/10.21105/joss.00024)
- [113] Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *The Astronomical Journal*, 154, 220, doi: [10.3847/1538-3881/aa9332](https://doi.org/10.3847/1538-3881/aa9332)
- [114] Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, 125, 306, doi: [10.1086/670067](https://doi.org/10.1086/670067)
- [115] Foster, G. 1996, *The Astronomical Journal*, 112, 1709, doi: [10.1017/CB09781107415324.004](https://doi.org/10.1017/CB09781107415324.004)
- [116] Fuller, J., & Ma, L. 2019, *The Astrophysical Journal*, 881, L1, doi: [10.3847/2041-8213/ab339b](https://doi.org/10.3847/2041-8213/ab339b)
- [117] Gabler, M., Cerdá-Durán, P., Font, J. A., Müller, E., & Stergioulas, N. 2011, *Monthly Notices of the Royal Astronomical Society: Letters*, 410, L37, doi: [10.1111/j.1745-3933.2010.00974.x](https://doi.org/10.1111/j.1745-3933.2010.00974.x)

- [118] —. 2013, *Monthly Notices of the Royal Astronomical Society*, 430, 1811, doi: [10.1093/mnras/sts721](https://doi.org/10.1093/mnras/sts721)
- [119] Galaudage, S., Adamcewicz, C., Zhu, X.-J., Stevenson, S., & Thrane, E. 2021, *The Astrophysical Journal Letters*, 909, L19, doi: [10.3847/2041-8213/abe7f6](https://doi.org/10.3847/2041-8213/abe7f6)
- [120] Garcíá-Quirós, C., Colleoni, M., Husa, S., et al. 2020, *Physical Review D*, 102, 1, doi: [10.1103/PhysRevD.102.064002](https://doi.org/10.1103/PhysRevD.102.064002)
- [121] Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. 2018, *Advances in Neural Information Processing Systems*, 2018-Decem, 7576. <https://arxiv.org/abs/1809.11165>
- [122] Garfinkle, D., Hollands, S., Ishibashi, A., Tolish, A., & Wald, R. M. 2017, *Classical and Quantum Gravity*, 34, 145015, doi: [10.1088/1361-6382/aa777b](https://doi.org/10.1088/1361-6382/aa777b)
- [123] Gayathri, V., Healy, J., Lange, J., et al. 2020. <https://arxiv.org/abs/2009.05461>
- [124] Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, 677. [http://files/25/Gelmanm.fl.-BayesianDataAnalysisThirdedition\(witherrors.pdf](http://files/25/Gelmanm.fl.-BayesianDataAnalysisThirdedition(witherrors.pdf)
- [125] Gentle, J. E. 2009, *Computational Statistics, Statistics and Computing* (New York, NY: Springer New York), doi: [10.1007/978-0-387-98144-4](https://doi.org/10.1007/978-0-387-98144-4)
- [126] Ghosh, A., Ghosh, A., Johnson-McDaniel, N. K., et al. 2016, *Physical Review D*, 94, 021101, doi: [10.1103/PhysRevD.94.021101](https://doi.org/10.1103/PhysRevD.94.021101)
- [127] Goncharov, B., Zhu, X. J., & Thrane, E. 2020, *Monthly Notices of the Royal Astronomical Society*, 497, 3264, doi: [10.1093/mnras/staa2081](https://doi.org/10.1093/mnras/staa2081)
- [128] Goncharov, B., Reardon, D. J., Shannon, R. M., et al. 2021, *Monthly Notices of the Royal Astronomical Society*, 502, 478, doi: [10.1093/mnras/staa3411](https://doi.org/10.1093/mnras/staa3411)
- [129] Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, *The Astrophysical Journal*, 622, 759, doi: [10.1086/427976](https://doi.org/10.1086/427976)
- [130] Greene, J. E. 2012, *Nature Communications*, 3, 1304, doi: [10.1038/ncomms2314](https://doi.org/10.1038/ncomms2314)
- [131] Guidorzi, C., Dichiara, S., & Amati, L. 2016, *Astronomy and Astrophysics*, 589, 1, doi: [10.1051/0004-6361/201527642](https://doi.org/10.1051/0004-6361/201527642)
- [132] Hamil, P. 2009, *Unit Test Frameworks: Tools for High-Quality Software Development* (O'Reilly Media)
- [133] Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015, *Monthly Notices of the Royal Astronomical Society*, 453, 4385, doi: [10.1093/mnras/stv1911](https://doi.org/10.1093/mnras/stv1911)
- [134] Harry, G. M. 2010, *Classical and Quantum Gravity*, 27, 084006, doi: [10.1088/0264-9381/27/8/084006](https://doi.org/10.1088/0264-9381/27/8/084006)
- [135] Hawking, S. W., Perry, M. J., & Strominger, A. 2016, *Physical Review Letters*, 116, 1, doi: [10.1103/PhysRevLett.116.231301](https://doi.org/10.1103/PhysRevLett.116.231301)
- [136] Hayes, L. A., Gallagher, P. T., Dennis, B. R., et al. 2019, arXiv, doi: [10.3847/1538-4357/ab0ca3](https://doi.org/10.3847/1538-4357/ab0ca3)

- [137] Hayes, L. A., Inglis, A. R., Christe, S., Dennis, B., & Gallagher, P. T. 2020, *The Astrophysical Journal*, 895, 50, doi: [10.3847/1538-4357/ab8d40](https://doi.org/10.3847/1538-4357/ab8d40)
- [138] Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, *The Astrophysical Journal*, 591, 288, doi: [10.1086/375341](https://doi.org/10.1086/375341)
- [139] Hild, S., Abernathy, M., Acernese, F., et al. 2011, *Classical and Quantum Gravity*, 28, doi: [10.1088/0264-9381/28/9/094013](https://doi.org/10.1088/0264-9381/28/9/094013)
- [140] Hobbs, G., & Dai, S. 2017, *National Science Review*, 4, 707, doi: [10.1093/nsr/nwx126](https://doi.org/10.1093/nsr/nwx126)
- [141] Hollands, S., Ishibashi, A., & Wald, R. M. 2017, *Classical and Quantum Gravity*, 34, 155005, doi: [10.1088/1361-6382/aa777a](https://doi.org/10.1088/1361-6382/aa777a)
- [142] Hoy, C., & Raymond, V. 2021, *SoftwareX*, 15, 100765, doi: [10.1016/j.softx.2021.100765](https://doi.org/10.1016/j.softx.2021.100765)
- [143] Hübner, M., Talbot, C., Lasky, P. D., & Thrane, E. 2020, *Physical Review D*, 101, 1, doi: [10.1103/PhysRevD.101.023011](https://doi.org/10.1103/PhysRevD.101.023011)
- [144] Hunter, J. D. 2007, *Computing in Science and Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- [145] Huppenkothen, D., & Bachetti, M. 2021, 19, 1. <https://arxiv.org/abs/2104.03278>
- [146] Huppenkothen, D., Heil, L. M., Watts, A. L., & Göğüş, E. 2014, *Astrophysical Journal*, 795, doi: [10.1088/0004-637X/795/2/114](https://doi.org/10.1088/0004-637X/795/2/114)
- [147] Huppenkothen, D., Watts, A. L., & Levin, Y. 2014, *Astrophysical Journal*, 793, doi: [10.1088/0004-637X/793/2/129](https://doi.org/10.1088/0004-637X/793/2/129)
- [148] Huppenkothen, D., Watts, A. L., Uttley, P., et al. 2012, *Astrophysical Journal*, 768, doi: [10.1088/0004-637X/768/1/87](https://doi.org/10.1088/0004-637X/768/1/87)
- [149] Huppenkothen, D., D'Angelo, C., Watts, A. L., et al. 2014, *Astrophysical Journal*, 787, doi: [10.1088/0004-637X/787/2/128](https://doi.org/10.1088/0004-637X/787/2/128)
- [150] Huppenkothen, D., Brewer, B. J., Hogg, D. W., et al. 2015, *Astrophysical Journal*, 810, doi: [10.1088/0004-637X/810/1/66](https://doi.org/10.1088/0004-637X/810/1/66)
- [151] Huppenkothen, D., Younes, G., Ingram, A., et al. 2017, *The Astrophysical Journal*, 834, 90, doi: [10.3847/1538-4357/834/1/90](https://doi.org/10.3847/1538-4357/834/1/90)
- [152] Hurley, K., Cline, T., Mazets, E., et al. 1999, *Nature*, 397, 41, doi: [10.1038/16199](https://doi.org/10.1038/16199)
- [153] Hurley, K., Boggs, S. E., Smith, D. M., et al. 2005, *Nature*, 434, 1098, doi: [10.1038/nature03519](https://doi.org/10.1038/nature03519)
- [154] Inglis, A. R., Ireland, J., Dennis, B. R., Hayes, L., & Gallagher, P. 2016, *The Astrophysical Journal*, 833, 284, doi: [10.3847/1538-4357/833/2/284](https://doi.org/10.3847/1538-4357/833/2/284)
- [155] Inglis, A. R., Ireland, J., & Dominique, M. 2015, *The Astrophysical Journal*, 798, 108, doi: [10.1088/0004-637X/798/2/108](https://doi.org/10.1088/0004-637X/798/2/108)

- [156] Ingram, A. R., & Motta, S. E. 2019, *New Astronomy Reviews*, 85, 101524, doi: [10.1016/j.newar.2020.101524](https://doi.org/10.1016/j.newar.2020.101524)
- [157] Ireland, J., McAteer, R. T., & Inglis, A. R. 2015, *Astrophysical Journal*, 798, doi: [10.1088/0004-637X/798/1/1](https://doi.org/10.1088/0004-637X/798/1/1)
- [158] Islo, K., Simon, J., Burke-Spolaor, S., & Siemens, X. 2019, 1. <https://arxiv.org/abs/1906.11936>
- [159] Israel, G. L., Belloni, T., Stella, L., et al. 2005, *The Astrophysical Journal*, 628, L53, doi: [10.1086/432615](https://doi.org/10.1086/432615)
- [160] Iwakiri, W., Ohno, M., Kamae, T., et al. 2010, *AIP Conference Proceedings*, 1279, 89, doi: [10.1063/1.3509358](https://doi.org/10.1063/1.3509358)
- [161] Iyer, B. R., Souradeep, T., Unnikrishnan, C. S., et al. 2011. <https://dcc.ligo.org/LIGO-M1100296/public>
- [162] Johnson, A. D., Kapadia, S. J., Osborne, A., Hixon, A., & Kennefick, D. 2019, *Physical Review D*, 99, 44045, doi: [10.1103/PhysRevD.99.044045](https://doi.org/10.1103/PhysRevD.99.044045)
- [163] Jones, E., et al. 2001, SciPy: Open source scientific tools for Python. <http://www.scipy.org/>
- [164] Kapec, D., Raclariu, A.-M., & Strominger, A. 2017, *Classical and Quantum Gravity*, 34, 165007, doi: [10.1088/1361-6382/aa7f12](https://doi.org/10.1088/1361-6382/aa7f12)
- [165] Kawamura, S., Ando, M., Nakamura, T., et al. 2008, *Journal of Physics: Conference Series*, 120, doi: [10.1088/1742-6596/120/3/032004](https://doi.org/10.1088/1742-6596/120/3/032004)
- [166] Khan, S., Husa, S., Hannam, M., et al. 2016, *Physical Review D*, 93, 1, doi: [10.1103/PhysRevD.93.044007](https://doi.org/10.1103/PhysRevD.93.044007)
- [167] Khera, N., Krishnan, B., Ashtekar, A., & De Lorenzo, T. 2021, *Physical Review D*, 103, 044012, doi: [10.1103/PhysRevD.103.044012](https://doi.org/10.1103/PhysRevD.103.044012)
- [168] Kilicarslan, E., & Tekin, B. 2019, *The European Physical Journal C*, 79, 114, doi: [10.1140/epjc/s10052-019-6636-4](https://doi.org/10.1140/epjc/s10052-019-6636-4)
- [169] King, A. 2016, *Monthly Notices of the Royal Astronomical Society: Letters*, 456, L109, doi: [10.1093/mnrasl/slv186](https://doi.org/10.1093/mnrasl/slv186)
- [170] Kramer, M., & Champion, D. J. 2013, *Classical and Quantum Gravity*, 30, 224009, doi: [10.1088/0264-9381/30/22/224009](https://doi.org/10.1088/0264-9381/30/22/224009)
- [171] Kumamoto, H., Dai, S., Johnston, S., et al. 2021, *Monthly Notices of the Royal Astronomical Society*, 501, 4490, doi: [10.1093/mnras/staa3910](https://doi.org/10.1093/mnras/staa3910)
- [172] Kumar, P., Shannon, R. M., Osłowski, S., et al. 2019, *The Astrophysical Journal*, 887, L30, doi: [10.3847/2041-8213/ab5b08](https://doi.org/10.3847/2041-8213/ab5b08)
- [173] Kurita, Y., & Nakano, H. 2016, *Physical Review D*, 93, 023508, doi: [10.1103/PhysRevD.93.023508](https://doi.org/10.1103/PhysRevD.93.023508)
- [174] Lachowicz, P., & Done, C. 2010, *Astronomy and Astrophysics*, 515, A65, doi: [10.1051/0004-6361/200913144](https://doi.org/10.1051/0004-6361/200913144)

- [175] Lackey, B. D., & Wade, L. 2015, *Physical Review D*, 91, 043002, doi: [10.1103/PhysRevD.91.043002](https://doi.org/10.1103/PhysRevD.91.043002)
- [176] Lange, J., O’Shaughnessy, R., & Rizzo, M. 2018, 1. <https://arxiv.org/abs/1805.10457>
- [177] Larsson, G., et al. 2021, <https://github.com/uchicago-cs/deepdish>
- [178] Lartillot, N., & Philippe, H. 2006, *Systematic Biology*, 55, 195, doi: [10.1080/10635150500433722](https://doi.org/10.1080/10635150500433722)
- [179] Lasky, P. D., Thrane, E., Levin, Y., Blackman, J., & Chen, Y. 2016, *Physical Review Letters*, 117, doi: [10.1103/PhysRevLett.117.061102](https://doi.org/10.1103/PhysRevLett.117.061102)
- [180] Lee, U. 2007, *Monthly Notices of the Royal Astronomical Society*, 374, 1015, doi: [10.1111/j.1365-2966.2006.11214.x](https://doi.org/10.1111/j.1365-2966.2006.11214.x)
- [181] Levin, Y. 2006, *Monthly Notices of the Royal Astronomical Society: Letters*, 368, doi: [10.1111/j.1745-3933.2006.00155.x](https://doi.org/10.1111/j.1745-3933.2006.00155.x)
- [182] —. 2007, *Monthly Notices of the Royal Astronomical Society*, 377, 159, doi: [10.1111/j.1365-2966.2007.11582.x](https://doi.org/10.1111/j.1365-2966.2007.11582.x)
- [183] Levin, Y., & van Hoven, M. 2011, *MNRAS*, 418, 659, doi: [10.1111/j.1365-2966.2011.19515.x](https://doi.org/10.1111/j.1365-2966.2011.19515.x)
- [184] LIGO Scientific Collaboration. 2018, LIGO Algorithm Library - LALSuite, free software (GPL), doi: [10.7935/GT1W-FZ16](https://doi.org/10.7935/GT1W-FZ16)
- [185] LIGO Scientific Collaboration, Virgo Collaboration, GBM, F., et al. 2017, *Astrophysical Journal Letters*, 848, L12, doi: [10.3847/2041-8213/aa91c9](https://doi.org/10.3847/2041-8213/aa91c9)
- [186] Logue, J., Ott, C. D., Heng, I. S., Kalmus, P., & Scargill, J. H. C. 2012, *Physical Review D*, 86, 044023, doi: [10.1103/PhysRevD.86.044023](https://doi.org/10.1103/PhysRevD.86.044023)
- [187] Lomb, N. R. 1976, *Astrophysics and Space Science*, 39, 447, doi: [10.1007/BF00648343](https://doi.org/10.1007/BF00648343)
- [188] Lower, M. E., Shannon, R. M., Johnston, S., & Bailes, M. 2020, *The Astrophysical Journal*, 896, L37, doi: [10.3847/2041-8213/ab9898](https://doi.org/10.3847/2041-8213/ab9898)
- [189] Lower, M. E., Thrane, E., Lasky, P. D., & Smith, R. 2018, *Physical Review D*, 98, 083028, doi: [10.1103/PhysRevD.98.083028](https://doi.org/10.1103/PhysRevD.98.083028)
- [190] Lower, M. E., Bailes, M., Shannon, R. M., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 494, 228, doi: [10.1093/mnras/staa615](https://doi.org/10.1093/mnras/staa615)
- [191] Macleod, D., Coughlin, S., Urban, A. L., Massinger, T., & al., E. 2018, *gwpy/gwpy*: 0.12.0, doi: [10.5281/zenodo.1346349](https://doi.org/10.5281/zenodo.1346349)
- [192] Madison, D. R., Cordes, J. M., & Chatterjee, S. 2014, *Astrophysical Journal*, 788, doi: [10.1088/0004-637X/788/2/141](https://doi.org/10.1088/0004-637X/788/2/141)
- [193] Manchester, R. N., Hobbs, G., Bailes, M., et al. 2013, *Publications of the Astronomical Society of Australia*, 30, e017, doi: [10.1017/pasa.2012.017](https://doi.org/10.1017/pasa.2012.017)

- [194] Martynov, D. V., Hall, E. D., Abbott, B. P., et al. 2016, *Physical Review D*, 93, 1, doi: [10.1103/PhysRevD.93.112004](https://doi.org/10.1103/PhysRevD.93.112004)
- [195] Masada, Y., Kawanaka, N., Sano, T., & Shibata, K. 2007, *The Astrophysical Journal*, 663, 437, doi: [10.1086/518088](https://doi.org/10.1086/518088)
- [196] Massey, F. J. 1951, *Journal of the American Statistical Association*, 46, 68, doi: [10.2307/2280095](https://doi.org/10.2307/2280095)
- [197] Mazets, E. P., Golenetskii, S. V., Il'inskii, V. N., Aptekar', R. L., & Guryan, Y. A. 1979, *Nature*, 282, 587, doi: [10.1038/282587a0](https://doi.org/10.1038/282587a0)
- [198] McKinney, W. 2010, *Proceedings of the 9th Python in Science Conference*, 51-56
- [199] McNeill, L. O., Thrane, E., & Lasky, P. D. 2017, *Physical Review Letters*, 118, 181103, doi: [10.1103/PhysRevLett.118.181103](https://doi.org/10.1103/PhysRevLett.118.181103)
- [200] Messenger, C., Takami, K., Gossan, S., Rezzolla, L., & Sathyaprakash, B. S. 2014, *Physical Review X*, 4, 041004, doi: [10.1103/PhysRevX.4.041004](https://doi.org/10.1103/PhysRevX.4.041004)
- [201] Meyers, P. M., O'Neill, N. J., Melatos, A., & Evans, R. J. 2021, *Monthly Notices of the Royal Astronomical Society*, 3363, 3349, doi: [10.1093/mnras/stab1952](https://doi.org/10.1093/mnras/stab1952)
- [202] Miller, J., Barsotti, L., Vitale, S., et al. 2015, *Physical Review D*, 91, 062005, doi: [10.1103/PhysRevD.91.062005](https://doi.org/10.1103/PhysRevD.91.062005)
- [203] Miller, M. C., Chirenti, C., & Strohmayer, T. E. 2019, *The Astrophysical Journal*, 871, 95, doi: [10.3847/1538-4357/aaf5ce](https://doi.org/10.3847/1538-4357/aaf5ce)
- [204] Miller, M. C., & Lamb, F. K. 2016, *European Physical Journal A*, 52, 63, doi: [10.1140/epja/i2016-16063-8](https://doi.org/10.1140/epja/i2016-16063-8)
- [205] Mitman, K., Moxon, J., Scheel, M. A., et al. 2020, *Physical Review D*, 102, 104007, doi: [10.1103/PhysRevD.102.104007](https://doi.org/10.1103/PhysRevD.102.104007)
- [206] Montoli, A., Antonelli, M., Magistrelli, F., & Pizzochero, P. M. 2020, *Astronomy and Astrophysics*, 642, 1, doi: [10.1051/0004-6361/202038340](https://doi.org/10.1051/0004-6361/202038340)
- [207] Morris, D., Battista, F., Dhuga, K., & MacLachlan, G. 2010, *AIP Conference Proceedings*, 1279, 394, doi: [10.1063/1.3509323](https://doi.org/10.1063/1.3509323)
- [208] Mukhopadhyay, M., Cardona, C., & Lunardini, C. 2021. <https://arxiv.org/abs/2105.05862>
- [209] Müller, E., Janka, H.-T., & Wongwathanarat, A. 2012, *Astronomy and Astrophysics*, 537, A63, doi: [10.1051/0004-6361/201117611](https://doi.org/10.1051/0004-6361/201117611)
- [210] Nakamura, T., Sasaki, M., Tanaka, T., & Thorne, K. S. 1997, *The Astrophysical Journal*, 487, L139, doi: [10.1086/310886](https://doi.org/10.1086/310886)
- [211] Nakariakov, V. M., & Melnikov, V. F. 2009, *Space Science Reviews*, 149, 119, doi: [10.1007/s11214-009-9536-3](https://doi.org/10.1007/s11214-009-9536-3)
- [212] Nita, G. M., Fleishman, G. D., Gary, D. E., Marin, W., & Boone, K. 2014, *Astrophysical Journal*, 789, doi: [10.1088/0004-637X/789/2/152](https://doi.org/10.1088/0004-637X/789/2/152)

- [213] Nitz, A., Harry, I., Brown, D., Biwer, C. M., & al., E. 2018, gwastro/pycbc: 1.12.3 Release, doi: [10.5281/zenodo.1410598](https://doi.org/10.5281/zenodo.1410598)
- [214] Norris, J. P., Bonnell, J. T., Kazanas, D., et al. 2005, *The Astrophysical Journal*, 627, 324, doi: [10.1086/430294](https://doi.org/10.1086/430294)
- [215] Norris, J. P., Nemiroff, R. J., Bonnell, J. T., et al. 1996, *The Astrophysical Journal*, 459, 393, doi: [10.1086/176902](https://doi.org/10.1086/176902)
- [216] Oliphant, T. E. 2015, *Guide to NumPy*, 2nd edn. (North Charleston, SC, USA: CreateSpace Independent Publishing Platform)
- [217] Ossokine, S., Buonanno, A., Marsat, S., et al. 2020, *Physical Review D*, 102, 044055, doi: [10.1103/PhysRevD.102.044055](https://doi.org/10.1103/PhysRevD.102.044055)
- [218] Palmer, D. M., Barthelmy, S., Gehrels, N., et al. 2005, *Nature*, 434, 1107, doi: [10.1038/nature03525](https://doi.org/10.1038/nature03525)
- [219] Payne, E., Talbot, C., Lasky, P. D., Thrane, E., & Kissel, J. S. 2020, *Physical Review D*, 102, 122004, doi: [10.1103/PhysRevD.102.122004](https://doi.org/10.1103/PhysRevD.102.122004)
- [220] Payne, E., Talbot, C., & Thrane, E. 2019, *Physical Review D*, 100, 123017, doi: [10.1103/PhysRevD.100.123017](https://doi.org/10.1103/PhysRevD.100.123017)
- [221] Perera, B. B. P., DeCesar, M. E., Demorest, P. B., et al. 2019, *Monthly Notices of the Royal Astronomical Society*, 490, 4666, doi: [10.1093/mnras/stz2857](https://doi.org/10.1093/mnras/stz2857)
- [222] Pierce, B. 2002, *Types and Programming Languages* (MIT Press)
- [223] Pollney, D., & Reisswig, C. 2011, *The Astrophysical Journal*, 732, L13, doi: [10.1088/2041-8205/732/1/L13](https://doi.org/10.1088/2041-8205/732/1/L13)
- [224] Powell, J., Gossan, S. E., Logue, J., & Heng, I. S. 2016, *Physical Review D*, 94, 123012, doi: [10.1103/PhysRevD.94.123012](https://doi.org/10.1103/PhysRevD.94.123012)
- [225] Powell, J., & Müller, B. 2019, *Monthly Notices of the Royal Astronomical Society*, 487, 1178, doi: [10.1093/mnras/stz1304](https://doi.org/10.1093/mnras/stz1304)
- [226] Powell, J., Szczepanczyk, M., & Heng, I. S. 2017, *Physical Review D*, 96, 123013, doi: [10.1103/PhysRevD.96.123013](https://doi.org/10.1103/PhysRevD.96.123013)
- [227] Pratten, G., Husa, S., García-Quirós, C., et al. 2020, *Physical Review D*, 102, 1, doi: [10.1103/PhysRevD.102.064001](https://doi.org/10.1103/PhysRevD.102.064001)
- [228] Pratten, G., García-Quirós, C., Colleoni, M., et al. 2021, *Physical Review D*, 103, 104056, doi: [10.1103/PhysRevD.103.104056](https://doi.org/10.1103/PhysRevD.103.104056)
- [229] Price-Whelan, A. M., Sipócz, B. M., Günther, H. M., et al. 2018, *The Astronomical Journal*, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- [230] Punturo, M., Abernathy, M., Acernese, F., et al. 2010, *Classical and Quantum Gravity*, 27, doi: [10.1088/0264-9381/27/19/194002](https://doi.org/10.1088/0264-9381/27/19/194002)
- [231] Qiu, H., Shannon, R. M., Farah, W., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 497, 1382, doi: [10.1093/mnras/staa1916](https://doi.org/10.1093/mnras/staa1916)

- [232] Ransom, S., Brazier, A., Chatterjee, S., et al. 2019, Bulletin of the AAS, 51. <https://arxiv.org/abs/1908.05356>
- [233] Rasmussen, C., & Williams, C. 2006, Gaussian Processes for Machine Learning, Vol. 7 (MIT Press)
- [234] Raymond, V., & Farr, W. M. 2014. <https://arxiv.org/abs/1402.0053>
- [235] Rea, N., Israel, G. L., Turolla, R., et al. 2009, Monthly Notices of the Royal Astronomical Society, 396, 2419, doi: [10.1111/j.1365-2966.2009.14920.x](https://doi.org/10.1111/j.1365-2966.2009.14920.x)
- [236] Read, J. S., Baiotti, L., Creighton, J. D. E., et al. 2013, Physical Review D, 88, 044042, doi: [10.1103/PhysRevD.88.044042](https://doi.org/10.1103/PhysRevD.88.044042)
- [237] Reitze, D., Adhikari, R. X., Ballmer, S., et al. 2019. <https://arxiv.org/abs/1907.04833>
- [238] Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, Astronomy and Astrophysics, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- [239] Rollins, J. G., & al., E. 2017, pygwinc. <https://git.ligo.org/gwinc/pygwinc>
- [240] Romero-Shaw, I., Lasky, P. D., Thrane, E., & Bustillo, J. C. 2020, The Astrophysical Journal, 903, L5, doi: [10.3847/2041-8213/abbe26](https://doi.org/10.3847/2041-8213/abbe26)
- [241] Romero-Shaw, I. M., Lasky, P. D., & Thrane, E. 2019, Monthly Notices of the Royal Astronomical Society, 490, 5210, doi: [10.1093/mnras/stz2996](https://doi.org/10.1093/mnras/stz2996)
- [242] Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, Monthly Notices of the Royal Astronomical Society, 499, 3295, doi: [10.1093/mnras/staa2850](https://doi.org/10.1093/mnras/staa2850)
- [243] Roulet, J., & Zaldarriaga, M. 2019, Monthly Notices of the Royal Astronomical Society, 484, 4216, doi: [10.1093/mnras/stz226](https://doi.org/10.1093/mnras/stz226)
- [244] Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55, doi: [10.7717/peerj-cs.55](https://doi.org/10.7717/peerj-cs.55)
- [245] Samuelsson, L., & Andersson, N. 2007, Monthly Notices of the Royal Astronomical Society, 374, 256, doi: [10.1111/j.1365-2966.2006.11147.x](https://doi.org/10.1111/j.1365-2966.2006.11147.x)
- [246] Sandrinelli, A., Covino, S., Treves, A., et al. 2018, Astronomy and Astrophysics, 615, doi: [10.1051/0004-6361/201732550](https://doi.org/10.1051/0004-6361/201732550)
- [247] Sarin, N., Lasky, P. D., & Ashton, G. 2020, Physical Review D, 101, 063021, doi: [10.1103/PhysRevD.101.063021](https://doi.org/10.1103/PhysRevD.101.063021)
- [248] Satishchandran, G., & Wald, R. M. 2018, Physical Review D, 97, 024036, doi: [10.1103/PhysRevD.97.024036](https://doi.org/10.1103/PhysRevD.97.024036)
- [249] Savitzky, A., & Golay, M. J. E. 1964, Analytical Chemistry, 36, 1627, doi: [10.1021/ac60214a047](https://doi.org/10.1021/ac60214a047)
- [250] Scargle, J. D. 1982, The Astrophysical Journal, 263, 835, doi: [10.1086/160554](https://doi.org/10.1086/160554)
- [251] Schmidt, P., Hannam, M., & Husa, S. 2012, Physical Review D, 86, 104063, doi: [10.1103/PhysRevD.86.104063](https://doi.org/10.1103/PhysRevD.86.104063)

- [252] Shibata, M., & Taniguchi, K. 2006, *Physical Review D*, 73, 064027, doi: [10.1103/PhysRevD.73.064027](https://doi.org/10.1103/PhysRevD.73.064027)
- [253] Singer, L. P., & Price, L. R. 2016, *Physical Review D*, 93, 024013, doi: [10.1103/PhysRevD.93.024013](https://doi.org/10.1103/PhysRevD.93.024013)
- [254] Singer, L. P., Chen, H.-Y., Holz, D. E., et al. 2016, *The Astrophysical Journal*, 829, L15, doi: [10.3847/2041-8205/829/1/L15](https://doi.org/10.3847/2041-8205/829/1/L15)
- [255] Skilling, J. 2004, in *AIP Conference Proceedings*, Vol. 735 (AIP), 395–405, doi: [10.1063/1.1835238](https://doi.org/10.1063/1.1835238)
- [256] Skilling, J. 2006, *Bayesian Analysis*, 1, 833, doi: [10.1214/06-BA127](https://doi.org/10.1214/06-BA127)
- [257] Smith, R., & Thrane, E. 2018, *Physical Review X*, 8, 021019, doi: [10.1103/PhysRevX.8.021019](https://doi.org/10.1103/PhysRevX.8.021019)
- [258] Smith, R. J., Ashton, G., Vajpeyi, A., & Talbot, C. 2020, *Monthly Notices of the Royal Astronomical Society*, 498, 4492, doi: [10.1093/mnras/staa2483](https://doi.org/10.1093/mnras/staa2483)
- [259] Somiya, K. 2012, *Classical and Quantum Gravity*, 29, doi: [10.1088/0264-9381/29/12/124007](https://doi.org/10.1088/0264-9381/29/12/124007)
- [260] Sotani, H., Colaiuda, A., & Kokkotas, K. D. 2008, *Monthly Notices of the Royal Astronomical Society*, 385, 2161, doi: [10.1111/j.1365-2966.2008.12977.x](https://doi.org/10.1111/j.1365-2966.2008.12977.x)
- [261] Sotani, H., Iida, K., & Oyamatsu, K. 2016, *New Astronomy*, 43, 80, doi: [10.1016/j.newast.2015.08.003](https://doi.org/10.1016/j.newast.2015.08.003)
- [262] Sotani, H., Iida, K., & Oyamatsu, K. 2017, *MNRAS*, 464, 3101, doi: [10.1093/mnras/stw2575](https://doi.org/10.1093/mnras/stw2575)
- [263] Sotani, H., & Kokkotas, K. D. 2009, *Monthly Notices of the Royal Astronomical Society*, 395, 1163, doi: [10.1111/j.1365-2966.2009.14631.x](https://doi.org/10.1111/j.1365-2966.2009.14631.x)
- [264] Sotani, H., Kokkotas, K. D., & Stergioulas, N. 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 261, doi: [10.1111/j.1365-2966.2006.11304.x](https://doi.org/10.1111/j.1365-2966.2006.11304.x)
- [265] —. 2008, *Monthly Notices of the Royal Astronomical Society: Letters*, 385, L5, doi: [10.1111/j.1745-3933.2007.00420.x](https://doi.org/10.1111/j.1745-3933.2007.00420.x)
- [266] Speagle, J. S. 2019, *arXiv*, 3158, 3132, doi: [10.1093/mnras/staa278](https://doi.org/10.1093/mnras/staa278)
- [267] Strohmayer, T. E., & Watts, A. L. 2005, *The Astrophysical Journal*, 632, L111, doi: [10.1086/497911](https://doi.org/10.1086/497911)
- [268] —. 2006, *The Astrophysical Journal*, 653, 593, doi: [10.1086/508703](https://doi.org/10.1086/508703)
- [269] —. 2006, *The Astrophysical Journal*, 653, 593, doi: [10.1086/508703](https://doi.org/10.1086/508703)
- [270] Strominger, A., & Zhiboedov, A. 2016, *Journal of High Energy Physics*, 2016, 86, doi: [10.1007/JHEP01\(2016\)086](https://doi.org/10.1007/JHEP01(2016)086)
- [271] Takami, K., Rezzolla, L., & Baiotti, L. 2015, *Physical Review D*, 91, 064001, doi: [10.1103/PhysRevD.91.064001](https://doi.org/10.1103/PhysRevD.91.064001)
- [272] Talbot, C., & Thrane, E. 2018, *The Astrophysical Journal*, 856, 173, doi: [10.3847/1538-4357/aab34c](https://doi.org/10.3847/1538-4357/aab34c)

- [273] Talbot, C., Thrane, E., Lasky, P. D., & Lin, F. 2018, *Physical Review D*, 98, 064031, doi: [10.1103/PhysRevD.98.064031](https://doi.org/10.1103/PhysRevD.98.064031)
- [274] Tarnopolski, M., & Marchenko, V. 2021, *The Astrophysical Journal*, 911, 20, doi: [10.3847/1538-4357/abe5b1](https://doi.org/10.3847/1538-4357/abe5b1)
- [275] Taylor, S. R., & Gerosa, D. 2018, *Physical Review D*, 98, 083017, doi: [10.1103/PhysRevD.98.083017](https://doi.org/10.1103/PhysRevD.98.083017)
- [276] The HDF Group. 1997-NNNN, Hierarchical Data Format, version 5
- [277] The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, B., et al. 2017, *Physical Review Letters*, 119, 161101, doi: [10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101)
- [278] The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, R., et al. 2021. <https://arxiv.org/abs/2108.01045>
- [279] Thorne, K. S. 1992, *Physical Review D*, 45, 520, doi: [10.1103/PhysRevD.45.520](https://doi.org/10.1103/PhysRevD.45.520)
- [280] Thrane, E., & Talbot, C. 2019, *Publications of the Astronomical Society of Australia*, 36, e010, doi: [10.1017/pasa.2019.2](https://doi.org/10.1017/pasa.2019.2)
- [281] Timmer, J., & König, M. 1995, *Astronomy and Astrophysics*, 300, 707
- [282] Torrence, C., & Compo, G. P. 1998, *Bulletin of the American Meteorological Society*, 79, 61, doi: [10.1175/1520-0477\(1998\)079<0061:APGTWA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
- [283] Tse, M., Yu, H., Kijbunchoo, N., et al. 2019, *Physical Review Letters*, 123, 231107, doi: [10.1103/PhysRevLett.123.231107](https://doi.org/10.1103/PhysRevLett.123.231107)
- [284] Unnikrishnan, C. S. 2013, *International Journal of Modern Physics D*, 22, 1341010, doi: [10.1142/S0218271813410101](https://doi.org/10.1142/S0218271813410101)
- [285] Vallisneri, M. 2008, *Physical Review D*, 77, 042001, doi: [10.1103/PhysRevD.77.042001](https://doi.org/10.1103/PhysRevD.77.042001)
- [286] van der Klis, M. 1989, *Annual Review of Astronomy and Astrophysics*, 27, 517, doi: [10.1146/annurev.aa.27.090189.002505](https://doi.org/10.1146/annurev.aa.27.090189.002505)
- [287] van der Sluys, M., Raymond, V., Mandel, I., et al. 2008, *Classical and Quantum Gravity*, 25, 184011, doi: [10.1088/0264-9381/25/18/184011](https://doi.org/10.1088/0264-9381/25/18/184011)
- [288] van der Sluys, M. V., Röver, C., Stroeer, A., et al. 2008, *The Astrophysical Journal*, 688, L61, doi: [10.1086/595279](https://doi.org/10.1086/595279)
- [289] Van Doorselaere, T., Kupriyanova, E. G., & Yuan, D. 2016, *Solar Physics*, 291, 3143, doi: [10.1007/s11207-016-0977-z](https://doi.org/10.1007/s11207-016-0977-z)
- [290] Van Haasteren, R., & Levin, Y. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 2372, doi: [10.1111/j.1365-2966.2009.15885.x](https://doi.org/10.1111/j.1365-2966.2009.15885.x)
- [291] VanderPlas, J. T. 2018, *The Astrophysical Journal Supplement Series*, 236, 16, doi: [10.3847/1538-4365/aab766](https://doi.org/10.3847/1538-4365/aab766)
- [292] Varma, V., Field, S. E., Scheel, M. A., et al. 2019, *arXiv*, 033015, 1, doi: [10.1103/physrevresearch.1.033015](https://doi.org/10.1103/physrevresearch.1.033015)

- [293] —. 2019, *Physical Review D*, 99, 064045, doi: [10.1103/PhysRevD.99.064045](https://doi.org/10.1103/PhysRevD.99.064045)
- [294] Vaughan, S. 2005, *Astronomy and Astrophysics*, 431, 391, doi: [10.1051/0004-6361:20041453](https://doi.org/10.1051/0004-6361:20041453)
- [295] —. 2009, *Monthly Notices of the Royal Astronomical Society*, 402, 307, doi: [10.1111/j.1365-2966.2009.15868.x](https://doi.org/10.1111/j.1365-2966.2009.15868.x)
- [296] Veitch, J., & al., E. 2017, *cpnest*, doi: [10.5281/zenodo.835874](https://doi.org/10.5281/zenodo.835874)
- [297] Veitch, J., & Vecchio, A. 2008, *Physical Review D*, 78, 022001, doi: [10.1103/PhysRevD.78.022001](https://doi.org/10.1103/PhysRevD.78.022001)
- [298] Veitch, J., Raymond, V., Farr, B., et al. 2015, *Physical Review D*, 91, 042003, doi: [10.1103/PhysRevD.91.042003](https://doi.org/10.1103/PhysRevD.91.042003)
- [299] Verbiest, J. P. W., Lentati, L., Hobbs, G., et al. 2016, *Monthly Notices of the Royal Astronomical Society*, 458, 1267, doi: [10.1093/mnras/stw347](https://doi.org/10.1093/mnras/stw347)
- [300] Vietri, M., Stella, L., & Israel, G. L. 2007, *ApJ*, 661, 1089, doi: [10.1086/517506](https://doi.org/10.1086/517506)
- [301] Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- [302] Vitale, S., Haster, C.-J., Sun, L., et al. 2021, *Physical Review D*, 103, 063016, doi: [10.1103/PhysRevD.103.063016](https://doi.org/10.1103/PhysRevD.103.063016)
- [303] Vousden, W. D., Farr, W. M., & Mandel, I. 2016, *Monthly Notices of the Royal Astronomical Society*, 455, 1919, doi: [10.1093/mnras/stv2422](https://doi.org/10.1093/mnras/stv2422)
- [304] Wang, J. B., Hobbs, G., Coles, W., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 1657, doi: [10.1093/mnras/stu2137](https://doi.org/10.1093/mnras/stu2137)
- [305] Wang, Y., Jiang, L.-Y., Li, C.-K., et al. 2021, 1. <https://arxiv.org/abs/2107.10796>
- [306] Watts, A. L., & Reddy, S. 2007, *MNRAS*, 379, L63, doi: [10.1111/j.1745-3933.2007.00336.x](https://doi.org/10.1111/j.1745-3933.2007.00336.x)
- [307] Watts, A. L., & Strohmayer, T. E. 2006, *The Astrophysical Journal*, 637, L117, doi: [10.1086/500735](https://doi.org/10.1086/500735)
- [308] Weakliem, D. L. 1999, *Sociological Methods and Research*, 27, 359, doi: [10.1177/0049124199027003002](https://doi.org/10.1177/0049124199027003002)
- [309] Whittle, P. 1951, *Hypothesis testing in time series analysis* (Almqvist & Wiksell)
- [310] Wilson, A. G., & Nickisch, H. 2015, 32nd International Conference on Machine Learning, ICML 2015, 3, 1775. <https://arxiv.org/abs/1503.01057>
- [311] Wiseman, A. G., & Will, C. M. 1991, *Phys. Rev. D*, 44, R2945, doi: [10.1103/PhysRevD.44.R2945](https://doi.org/10.1103/PhysRevD.44.R2945)
- [312] Woosley, S. E., & Heger, A. 2015, *The Deaths of Very Massive Stars*, ed. J. S. Vink (Springer International Publishing), 199–225, doi: [10.1007/978-3-319-09596-7_7](https://doi.org/10.1007/978-3-319-09596-7_7)
- [313] Wysocki, D., Gerosa, D., O’Shaughnessy, R., et al. 2018, *Physical Review D*, 97, 043014, doi: [10.1103/PhysRevD.97.043014](https://doi.org/10.1103/PhysRevD.97.043014)

-
- [314] Yang, H., & Martynov, D. 2018, Physical Review Letters, 121, 071102, doi: [10.1103/PhysRevLett.121.071102](https://doi.org/10.1103/PhysRevLett.121.071102)
- [315] Yu, H., Martynov, D., Vitale, S., et al. 2018, Physical Review Letters, 120, 141102, doi: [10.1103/PhysRevLett.120.141102](https://doi.org/10.1103/PhysRevLett.120.141102)
- [316] Zel'dovich, Ya., B., & Polnarev, A. G. 1974, Soviet Astronomy, 18, 17, doi: [10.15713/ins.mmj.3](https://doi.org/10.15713/ins.mmj.3)
- [317] Zhu, X. J., & Thrane, E. 2020, arXiv, doi: [10.3847/1538-4357/abac5a](https://doi.org/10.3847/1538-4357/abac5a)
- [318] Ziaepour, H., & Gardner, B. 2011, Journal of Cosmology and Astroparticle Physics, 2011, doi: [10.1088/1475-7516/2011/12/001](https://doi.org/10.1088/1475-7516/2011/12/001)
- [319] Zimovets, I. V., McLaughlin, J. A., Srivastava, A. K., et al. 2021, Space Science Reviews, 217, 66, doi: [10.1007/s11214-021-00840-9](https://doi.org/10.1007/s11214-021-00840-9)