



Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

THE UNIVERSITY OF AUCKLAND

DOCTORAL THESIS

**Bayesian modelling of stellar core
collapse gravitational wave signals and
detector noise**

Matthew Charles EDWARDS

*A thesis submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy in Statistics,
The University of Auckland, 2017.*

August 2017

Declaration of Authorship

I, Matthew Charles EDWARDS, declare that this thesis titled, ‘Bayesian modelling of stellar core collapse gravitational wave signals and detector noise’ and the work presented in it are my own.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“I was born not knowing and have had only a little time to change that here and there.”

Richard Feynman

THE UNIVERSITY OF AUCKLAND

Abstract

Faculty of Science

Department of Statistics

Doctor of Philosophy

Bayesian modelling of stellar core collapse gravitational wave signals and detector noise

by Matthew Charles EDWARDS

A new era of astronomy dawned on September 14, 2015, when the Advanced Laser Interferometer Gravitational-Wave Observatory (Advanced LIGO) detectors observed a gravitational wave signal from a binary black hole merger for the first time. This was followed by two more observations of gravitational waves from black hole binary mergers on December 26, 2015, and January 4, 2017. Bayesian data analysis played a key role in inferring the underlying astrophysics of these events. As more detectors come on-line and new discoveries are made, novel data analysis techniques will be critical to accurately model gravitational wave signals and background noise.

Though stellar core collapse gravitational waves have not been observed yet, parameter estimation routines that can extract important astrophysical parameters encoded in these signals must be designed for their eventual detection. These methods will need to be different from those of binary black hole mergers as stellar core collapse signals are far more complex. A novel method for parameter estimation of stellar core collapse will be discussed here. The signal will first be reconstructed using principal component regression and implemented using Metropolis-within-Gibbs and reversible jump Markov chain Monte Carlo algorithms. Known astrophysical parameters will be fitted to Monte Carlo estimates of the principal component coefficients. Inferences of important physical quantities will then be made by sampling from the posterior predictive distribution and by applying classification and cross-validation methods.

In addition to modelling stellar core collapse signals, the noise spectral density from the ground-based gravitational wave detectors, Advanced LIGO, will be modelled using the methods of

Bayesian nonparametrics. Three different approaches will be presented: the Bernstein polynomial prior; a newly developed B-spline prior; and the recently developed nonparametric correction to a parametric likelihood. These methods will address the limitations of the default parametric noise model used in much of the gravitational wave data analysis literature and in practice.

Acknowledgements

My sincerest thanks go to my wonderful PhD supervisors, Renate Meyer and Nelson Christensen. I was very fortunate to have such talented, patient, and dedicated supervisors, giving me thoughtful guidance and the freedom to explore my research interests. Thanks to my amazing family and my wonderful partner Bex for your unconditional love and support. I couldn't have done this without you. To my fantastic friends (you know who you are), thanks for all the laughs. I am appreciative of my insightful Bayesian time series collaborators, Claudia Kirch and Alexander Meier, and my diligent proof-readers, Rebekah Sherriff and Louise McMillan. I extend my warmest thanks to Blake Seers and Niffe Hermansson from 224, whom I shared much of the past four years with. Your distractions have been welcome. My gratitude goes to the New Zealand eScience Infrastructure (NeSI) for their high performance computing facilities, and the Centre for eResearch at the University of Auckland for their technical support. I am very grateful for the financial support provided by the University of Auckland Doctoral Scholarship. Finally, I would also like to thank the LIGO Scientific Collaboration for letting me be part of this amazing journey.

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	v
Contents	vi
1 Introduction	1
2 Gravitational wave astronomy	5
2.1 Gravitational waves	5
2.2 Gravitational wave sources	6
2.3 Gravitational wave detectors	8
2.3.1 Advanced LIGO	9
2.4 Stellar core collapse	12
2.5 Gravitational wave data analysis	13
2.5.1 Detection significance	14
2.5.2 Parameter estimation	15
2.5.3 Noise modelling	16
3 Statistical methods	18
3.1 The Bayesian paradigm	18
3.1.1 Bayesian inference	19
3.1.2 Summarising the posterior	19
3.1.3 Posterior prediction	20
3.1.4 Model selection	20
3.2 Markov chain Monte Carlo	22
3.2.1 Monte Carlo integration	22
3.2.2 Markov chains	23
3.2.3 Markov chain Monte Carlo simulation	24
3.2.4 Metropolis-Hastings algorithm	26
3.2.5 Gibbs sampler	27

3.2.6	Reversible jump MCMC	28
3.2.6.1	Birth-death moves and nested models	29
3.2.6.2	Automatic scaling: Zeroth order method	30
3.2.7	Parallel tempering	30
3.3	Bayesian nonparametrics	32
3.3.1	Introduction	32
3.3.2	Dirichlet process prior	34
3.3.3	Bernstein polynomial prior	35
3.3.4	B-spline prior	36
3.3.4.1	B-splines and B-spline densities	36
3.3.4.2	B-spline prior	38
3.4	Fourier analysis of time series	41
3.4.1	Stationary time series	41
3.4.2	Fourier transform	42
3.4.3	Properties of the Fourier transform	42
3.4.4	Discrete Fourier transform	43
3.4.5	Sampling a continuous signal	44
3.4.6	Downsampling	46
3.4.7	Linear filters	46
3.5	Spectral density estimation	47
3.5.1	Introduction	47
3.5.2	Periodogram smoothing	50
3.5.3	Time-varying spectra	50
3.5.4	Bayesian nonparametric approaches to spectral density estimation	51
3.5.4.1	Bernstein polynomial prior	51
3.5.4.2	B-spline prior	53
3.5.4.3	Nonparametric correction of a parametric likelihood	55
3.6	Principal component analysis	59
3.7	Classification	60
3.7.1	Supervised machine learning	60
3.7.1.1	Naïve Bayes classifier	60
3.7.1.2	k -nearest neighbour	61
3.7.2	Generalised linear models	61
3.7.2.1	Bayesian ordinal probit regression	62
4	Parameter estimation for rotating stellar core collapse	63
4.1	Introduction	63
4.2	Core collapse waveform catalogue	65
4.3	Signal reconstruction	67
4.3.1	Signal reconstruction model 1: Metropolis-within-Gibbs PCR with known signal arrival time	68
4.3.2	Signal reconstruction model 2: Metropolis-within-Gibbs PCR with unknown signal arrival time	70
4.3.3	Fixed dimension model selection	71
4.3.4	Signal reconstruction model 3: Reversible jump PCR	73

4.4	Ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce	77
4.5	Precollapse differential rotation	81
4.5.1	Supervised learning algorithms	81
4.5.2	Bayesian ordinal probit regression	83
4.6	Nuclear equation of state	84
5	Spectral density estimation for Advanced LIGO noise	88
5.1	Introduction	88
5.2	Advanced LIGO spectral density estimation using the Bernstein polynomial prior	90
5.2.1	PSD estimation	93
5.2.2	Signal reconstruction	93
5.2.3	Nonstationary noise	94
5.2.4	Results	95
5.2.4.1	Estimating the PSD of non-Gaussian coloured noise	96
5.2.4.2	Extracting a rotating core collapse signal in stationary coloured noise	98
5.2.4.3	Comparing input and reconstruction parameters	100
5.2.4.4	Extracting a rotating core collapse signal in time-varying coloured noise	101
5.2.4.5	Detecting a spectral change-point	104
5.2.4.6	Simulated Advanced LIGO noise	105
5.3	Advanced LIGO spectral density estimation using the B-spline prior	108
5.3.1	Simulation study: A comparison of the Bernstein polynomial prior and B-spline prior	110
5.3.2	Application of the B-spline prior to recoloured LIGO S6 data	114
5.3.2.1	Informative prior	115
5.4	Advanced LIGO spectral density estimation using a nonparametric correction to a parametric likelihood	116
6	Conclusions and outlook	121
	Bibliography	125

Dedicated to my parents.

Graduate Centre
 The ClockTower – East Wing
 22 Princes Street, Auckland
 Phone: +64 9 373 7599 ext 81321
 Fax: +64 9 373 7610
 Email: postgraduate@auckland.ac.nz
 www.postgrad.auckland.ac.nz



Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Publication details: Bayesian parameter estimation of core collapse supernovae using gravitational wave simulations, *Inverse Problems*, **30**, 114008 (2014).

Presented in: Chapter 3 (Sections: 3.1.4, 3.7.1) and Chapter 4 (Sections: 4.1, 4.2, 4.3.1, 4.3.2, 4.3.3, 4.4, 4.5.1).

Nature of contribution by PhD candidate	Conducted the study; ran the simulations; wrote the paper.	
Extent of contribution by PhD candidate (%)	90%	

CO-AUTHORS

Name	Nature of Contribution
Renate Meyer	Supervisor. Edited the paper.
Nelson Christensen	Supervisor. Edited the paper.

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and that the candidate wrote all or the majority of the text.

Name	Signature	Date
Renate Meyer		22.2.17
Nelson Christensen		22.2.17

Graduate Centre
 The ClockTower – East Wing
 22 Princes Street, Auckland
 Phone: +64 9 373 7599 ext 81321
 Fax: +64 9 373 7610
 Email: postgraduate@auckland.ac.nz
 www.postgrad.auckland.ac.nz



Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.	
Publication details: Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis, <i>Physical Review D</i> , 92 , 064011 (2015).	
Presented in: Chapter 3 (Sections: 3.3.2, 3.3.3, 3.5.1, 3.5.4.1) and Chapter 5 (Sections: 5.1, 5.2).	
Nature of contribution by PhD candidate	Conducted the study; ran the simulations; wrote the paper.
Extent of contribution by PhD candidate (%)	90%

CO-AUTHORS

Name	Nature of Contribution
Renate Meyer	Supervisor. Edited the paper.
Nelson Christensen	Supervisor. Edited the paper.

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and that the candidate wrote all or the majority of the text.

Name	Signature	Date
Renate Meyer		22.2.17
Nelson Christensen		22.2.17

Graduate Centre
 The ClockTower – East Wing
 22 Princes Street, Auckland
 Phone: +64 9 373 7599 ext 81321
 Fax: +64 9 373 7610
 Email: postgraduate@auckland.ac.nz
 www.postgrad.auckland.ac.nz



Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Publication details: Bayesian nonparametric spectral density estimation using B-spline priors, *Pre-print (submitted to Statistics and Computing)* (2017).

Presented in: Chapter 3 (Sections: 3.2.7, 3.3.4, 3.5.4.2) and Chapter 5 (Sections: 5.3).

Nature of contribution by PhD candidate	Conducted the study; ran the simulations; wrote the paper.	
Extent of contribution by PhD candidate (%)	90%	

CO-AUTHORS

Name	Nature of Contribution
Renate Meyer	Supervisor. Edited the paper.
Nelson Christensen	Supervisor. Edited the paper.

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and that the candidate wrote all or the majority of the text.

Name	Signature	Date
Renate Meyer		22.2.17
Nelson Christensen		22.2.17

Graduate Centre
 The ClockTower - East Wing
 22 Princes Street, Auckland
 Phone: +64 9 373 7599 ext 81321
 Fax: +64 9 373 7610
 Email: postgraduate@auckland.ac.nz
 www.postgrad.auckland.ac.nz



Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.	
Publication details: Section 4.3 of Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis, <i>Pre-print</i> , arXiv:1701.04846 [stat.ME] (2017).	
Presented in: Chapter 5 (Section: 5.4).	
Nature of contribution by PhD candidate	Wrote code; ran simulations; wrote results for LIGO application.
Extent of contribution by PhD candidate (%)	75%

CO-AUTHORS

Name	Nature of Contribution
Claudia Kirch	Author. Edited Section 4.3.
Alexander Meier	Author.
Renate Meyer	Author. Wrote introduction of Section 4.3 (not included in thesis).

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and that the candidate wrote all or the majority of the text.

Name	Signature	Date
Claudia Kirch		21.2.17
Alexander Meier		22.2.17
Renate Meyer		22.2.17

Chapter 1

Introduction

The fields of astronomy, astrophysics, and cosmology forever changed on September 14, 2015, when the two Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors directly observed a gravitational wave signal for the first time (Abbott et al., 2016f). Lasting for 0.2 s, the signal GW150914 came from the inspiral of two stellar mass black holes (about 36 and 29 times the mass of the Sun). These coalesced to form a single rotating black hole (about 62 times the mass of the Sun), matching predictions from Einstein’s general theory of relativity (Einstein, 1916). During this process, roughly three times the mass of the Sun was converted into gravitational wave energy, which travelled at the speed of light for a distance of approximately 1.3 billion light-years to reach Earth. While passing through Earth, the two Advanced LIGO detectors were on-line and sensitive enough to detect the signal.

As well as being the first observation of gravitational waves, GW150914 was the first ever observation of a binary black hole merger. A second binary black hole merger was later observed on December 26, 2015 (Abbott et al., 2016d). The signal GW151226 was much weaker than GW150914, coming from a pair of black holes approximately 14 and 8 times the mass of the Sun. They coalesced to form a spinning black hole of roughly 21 times the mass of the Sun, radiating one solar mass of gravitational wave energy. Subsequently, a third gravitational wave signal (GW170104) was observed by the LIGO detectors on January 4, 2017, coming from a 50 solar mass binary black hole coalescence (Abbott et al., 2017).

Future detections are imminent. Now that gravitational waves have been directly measured for the first time, they have opened an alternative window to view the Universe through, launching a new era in astronomy. It is a new era in the sense that traditional astronomy has generally probed various forms of electromagnetic radiation, but the “dark-side” of the Universe (i.e., opaque to light) can now be explored using gravitational waves. This has the potential to lead to new and exciting insights about the cosmos.

The breakthrough detections of GW150914, GW151226, and GW170104 come after decades of ambitious engineering challenges and the collaborative effort of thousands of scientists, engineers, and data analysts. As gravitational waves are extremely weak after travelling great distances, the physical reality of detecting these cosmic harbingers was debated amongst physicists until 1957 (Abbott et al., 2016f, Saulson, 2011). There should be no more debate. The evidence in favour of the measurement of GW150914 coming from an astrophysical origin is overwhelming, with a false alarm rate of 1 event per 203,000 years.

Bayesian methods have had a brief but fertile history in astronomy. After its first conception by English statistician and minister Thomas Bayes (published posthumously (Bayes, 1763)), Bayes' theorem was independently developed and formalised by French mathematician and astronomer Pierre-Simon Laplace (Laplace, 1812). Under the Bayesian formalism, one's prior beliefs are updated using additional data. In the early 19th century, Laplace applied Bayesian reasoning to compare astronomical observations with his theories of celestial mechanics — the first use of Bayesian probability theory in astronomy!

Fast forward a few hundred years, Bayesian methods are used in a variety of modern astronomical applications, ranging from exoplanet discovery (Brewer and Donovan, 2015) to gravitational lensing (Brewer et al., 2016), from characterising magnetar bursts (Huppenkothen et al., 2015) to inferring star formation history (Walmswell et al., 2013). Further, Bayesian methods have had tremendous success in cosmology since the turn of the millennium, due to the pioneering work of Christensen et al. (2001). This is not so surprising as Bayesian models are incredibly well-suited to astronomy and cosmology (see (Loredo, 1992) for further discussion).

Though Bayesian and frequentist methods can lead to the same conclusions under vague prior information, they are fundamentally different philosophies. The notion of probability, for example, has a different interpretation to the Bayesian than it does to the frequentist. A Bayesian would describe probability as a measure of plausibility of a statement, quantifying uncertainty. On the other hand, a frequentist would argue probability is the limiting relative frequency after repeating an identical trial an infinite number of times. Using these definitions, an observational scientist like the astronomer would rather draw conclusions about a unique observation at hand (using a Bayesian credible interval) than to compare this to thousands of imaginary repetitions of that event (using a frequentist confidence interval) as those repetitions might not be a physical or logical reality (Loredo, 1990).

A rise in popularity of Bayesian and Markov chain Monte Carlo (MCMC) methods has been observed over the past few decades, particularly in astronomy. However, the frequentist approach dominated the statistical literature and teaching in the 19th and 20th centuries. This was partly due to the firm belief held by many that science should not be subjective (and subjectivity can

certainly occur under the Bayesian philosophy through the use of strong prior information), and partly due to the pioneering works on classical hypothesis testing, analysis of variance, and experimental design (among other topics) by the prolific English statistician and biologist, Ronald Fisher (Fisher, 1925, 1935). In recent times, Bayesian inference has gained traction in mainstream statistics and astronomy in part due to the philosophy that it is a generalisation of logic (where rather than dealing with truths, one deals with uncertainty) (Gregory, 2005, Jaynes, 2003), and largely due to improved computing power.

Today, Bayesian analysis, particularly MCMC, is an important tool for gravitational wave astronomy, and played a crucial role in the detections of the binary black hole merger signals, GW150914, GW151226, and GW170104. Based on the pioneering work of Christensen and Meyer (1998), sophisticated MCMC methods have been developed to infer important astrophysical parameters from compact binary coalescence (CBC) signals, such as the mass and spin of black holes before and after coalescence, and the source location in the sky (Abbott et al., 2016h). MCMC routines have also been developed for other gravitational wave sources of astrophysical origin (see for example Clark et al. (2007), Cornish and Littenberg (2015), Coughlin et al. (2014), Logue et al. (2012), Röver et al. (2009)).

Parameter estimation routines and template-based search methods are well-defined for CBC signals. A template-based approach involves “matching” a signal to a template from a large waveform catalogue with a fully explored parameter space. However, this approach is not possible for core collapse supernovae due to a complex interaction between general relativity and particle physics. It becomes computationally infeasible to conduct template-based searches as stellar core collapse signals do not have a closed-form expression, meaning one cannot populate a large library of templates with a fully explored parameter space. Alternative approaches are thus required.

This thesis therefore addresses the parameter estimation problem for rotating stellar core collapse gravitational wave signals. Waveforms are first reconstructed using principal component basis functions. The posterior means of the principal component coefficients are then exploited to extract encoded astrophysical information. This is achieved using posterior predictive sampling, as well as classification and cross-validation methods.

Bayesian nonparametric modelling has received much less attention in the field of gravitational wave data analysis than its parametric counterpart. This may be attributed to the many difficulties inherent within the Bayesian nonparametric framework, such as specifying priors on infinite-dimensional function spaces, implementing computationally convenient sampling algorithms, and ensuring posterior consistency is attained (Hjort et al., 2010). Bayesian nonparametric approaches to gravitational wave data analysis could make a significant impact in the field, particularly in the case where parametric models are misspecified, which often result in misleading inferences.

To this end, Bayesian nonparametric models will play a key role in this thesis, with a focus on estimating the spectral density of noise from the ground-based Advanced LIGO detectors.

The standard approach to modelling noise from terrestrial gravitational wave detectors is to assume it is stationary and Gaussian distributed, estimating the spectral density as an average of periodograms of off-source data. This approach may be considered fundamentally flawed for various reasons. First, high amplitude transient noise artifacts, or glitches, appearing in gravitational wave strain data nullifies the Gaussian assumption (Cornish and Littenberg, 2015). Second, it has been demonstrated that noise from Advanced LIGO is time-varying, voiding the stationarity assumption (Littenberg et al., 2013). Following on from this, it would be prudent to characterise noise on the gravitational wave signal rather than using nearby segments. Bayesian nonparametric priors are used in this thesis to address these limitations. The methods implemented include the Bernstein polynomial prior of Petrone (1999a,b) and Choudhuri et al. (2004), the newly developed B-spline prior, and a nonparametric correction to an autoregressive likelihood.

The work presented in this thesis is based on the published papers by Edwards et al. (2014)¹ and Edwards et al. (2015)², as well as the submitted but unpublished papers by Edwards et al. (2017a) and Kirch et al. (2017), and some passages have been quoted verbatim.

The structure of this thesis is as follows. Chapter 2 gives an introduction to gravitational wave astronomy, while Chapter 3 discusses the statistical tools used in this thesis to conduct gravitational wave data analysis. This includes MCMC methods, Fourier analysis of time series, and the methods of Bayesian nonparametrics. Methods are then applied in Chapter 4 to infer important astrophysical parameters from rotating stellar core collapse gravitational wave signals in additive noise. In Chapter 5, three different Bayesian nonparametric approaches are used to estimate the power spectral density (PSD) of noise from the Advanced LIGO detectors. Finally, Chapter 6 provides concluding remarks and an outlook for future research directions.

¹© IOP Publishing. Reproduced with permission. All rights reserved.

²© APS. Reproduced with permission. All rights reserved.

Chapter 2

Gravitational wave astronomy

2.1 Gravitational waves

In 1916, Albert Einstein published his famous general theory of relativity (Einstein, 1916), a geometrical theory of gravity. This theory uses the notion that gravity arises from a curvature in a four-dimensional manifold called *space-time*. Conversely, space-time is curved in the presence of mass, energy, pressure, and angular momentum. A massive object (like the Sun) curves space-time. The more massive the object, the more space-time curves. A smaller object (like the Earth) will move in a geodesic along curved space-time, orbiting the larger object.

Gravitational waves are a consequence of general relativity. They are ripples in the fabric of space-time, caused by accelerating massive objects. They propagate through the Universe at the speed of light, carrying energy and astrophysical information away from the source. As gravitational waves are extremely weakly interacting, they traverse the Universe, essentially unobscured by intervening matter. As they are neither absorbed, nor scattered, gravitational waves are a rich and pure source of astrophysical information.

Gravitational waves stretch and compress space-time. Their effect is orthogonal to the direction of propagation. Gravitational waves have a dimensionless *strain* h , which is a combination of two polarisations: a plus (+)-polarisation and a cross (\times)-polarisation. First consider the linear polarised case with amplitude h_+ (i.e., a +-polarisation with no \times -polarisation). Let two points be separated by length L along an axis that is orthogonal to the direction of the gravitational wave. The distance separating the two points will be stretched by $\Delta L = \frac{1}{2}h_+L$. The perpendicular axis (still orthogonal to the direction of propagation) with two points separated by L will simultaneously be compressed by $\Delta L = -\frac{1}{2}h_+L$. Each axis will then oscillate back and forth, stretching and

compressing as the gravitational wave propagates through the plane of observation. The \times -polarisation has amplitude h_{\times} and has a similar effect, but rotated by 45 degrees (such that it is still perpendicular to the direction of propagation). These effects can be seen in Figure 2.1.

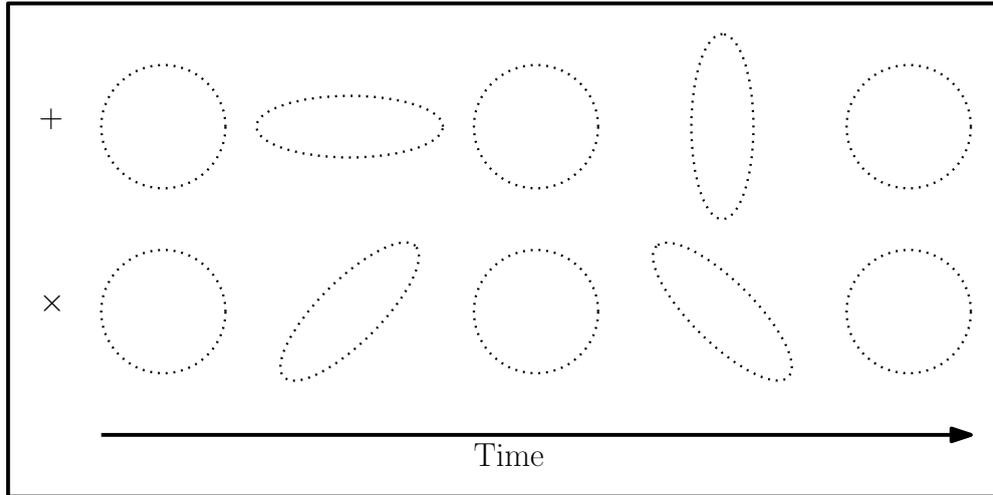


FIGURE 2.1: The effect a passing gravitational wave will have on a circular arrangement of free-falling test masses. The top row illustrates the $+$ -polarisation and the bottom row the \times -polarisation.

Gravitational waves have a very small strain on space (primarily due to source distance, the speed of light, and the gravitational constant) when they reach Earth, and only events with very large mass will produce a signal strong enough for detection. In fact, sources are expected to produce a strain of order $h \approx 10^{-21}$, which corresponds to a relative mirror displacement of approximately 10^{-18} in the ground-based gravitational wave interferometers — many orders of magnitude smaller than the diameter of a proton. This has led to decades of enormous engineering challenges, and it is only very recently (in September 2015) that detectors became sensitive enough to finally directly detect gravitational waves (Abbott et al., 2016d,f, 2017).

Some gravitational wave sources will have electromagnetic counterparts, such as supernovae and the coalescence of a binary neutron star system, which could give new insights into the underlying mechanics of such phenomena. Other sources, such as the oscillations of newly formed black holes, will not emit electromagnetic radiation, giving a direct probe into previously unobservable events. With the recent breakthrough detections of binary black holes (Abbott et al., 2016d,f, 2017), gravitational wave astronomy has opened a new window to view the cosmos through.

2.2 Gravitational wave sources

Gravitational waves are classified into four groups: inspiral, continuous, burst, and stochastic.

Inspirational gravitational waves were the first (and only to date) class to be indirectly (Hulse and Taylor, 1975, Taylor and Weisburg, 1989) and directly (Abbott et al., 2016d,f, 2017) observed. They occur when two massive objects (i.e., two black holes, two neutron stars, or a black hole and a neutron star) inspiral leading to an eventual coalescence. As the two stellar mass objects orbit each other, they emit gravitational waves and lose energy. Their orbital distance will decrease, increasing the orbital frequency and the gravitational wave frequency and amplitude. This type of signal is often referred to as a *chirp*, and is the most well-modelled and understood gravitational wave source (Thorne, 1987).

Continuous gravitational waves are generated by, for example, single massive objects like pulsars (rotating neutron stars) with imperfections, making them slightly non-spherical (Cutler, 2002). These types of signals will typically be long-duration, and will have a near constant amplitude and frequency (quasimonochromatic) (Abbott et al., 2009a). They will also be relatively weak and difficult to integrate over a long time.

Traditional astronomy has typically involved detectors exploring various forms of electromagnetic radiation (e.g., visible light, radio waves, microwaves, X-rays, etc.). Whenever new technologies have been used to study the cosmos, many unexpected discoveries have been made, revolutionising the way humans view the Universe. As gravitational wave astronomy probes the “dark-side” of the Universe (using gravity rather than electromagnetism), new and unexpected astrophysical discoveries could be made. This is where *burst* gravitational waves fit into the picture. Bursts are typically high amplitude, short-duration events. Some burst signals are expected to come from unknown origins, and little is known about their underlying mechanisms. Other burst signals are expected to come from known sources but cannot be modelled well. These include rotating core collapse supernovae followed by protoneutron star formation (Ott et al., 2004), the mechanisms that generate gamma-ray bursts (Meszaros, 2006), cosmic string cusps (Damour and Vilenkin, 2005), pulsar glitches (Andersson and Comer, 2001), and starquakes from magnetar flares (Mereghetti, 2008). Searches for burst signals are performed using excess power searches in the time-frequency domain and coherence in multiple gravitational wave detectors (Klimenko et al., 2008, Sutton et al., 2010).

Stochastic gravitational waves (also called the gravitational wave background) come from a conglomeration of small unresolved gravitational wave signals (from the previously mentioned groups) Abbott et al. (2009c). Analogous to the cosmic microwave background, a stochastic gravitational wave background could be left over from the Big Bang. Because the early Universe was not transparent to light until about 300,000 years after the Big Bang, direct measurements of a stochastic gravitational background will give researchers an opportunity to see further into the past than ever before (seconds after the Big Bang). This will help constrain various cosmological models of the

early Universe. However, as stochastic gravitational waves have the smallest strain on space-time, they will be incredibly difficult to detect.

2.3 Gravitational wave detectors

Direct detection of gravitational waves has proven to be a notoriously difficult and ambitious engineering challenge. As the sources of gravitational waves are far away from Earth, gravitational waves are expected to have a strain amplitude of order $\sim 10^{-21}$ or less. Detector and environmental noise also limit the ability to resolve these small perturbations in space.

Historically, resonant bars (Weber, 1967) and first-generation ground-based interferometers — Initial LIGO (Abbott et al., 2009b), Initial Virgo (Accadia et al., 2011), TAMA 300 (Ando, 2015), GEO 600 (Grote, 2010) — were the first attempts at directly detecting gravitational waves. Though important technologies were developed during these projects, no direct observations of gravitational waves were confirmed during their operation (Abbott et al., 2016a). Currently, GEO 600 is operational, TAMA 300 has been decommissioned, and Initial LIGO and Initial Virgo have been developed into second-generation detectors.

Though gravitational waves were indirectly observed by (Hulse and Taylor, 1975, Taylor and Weisburg, 1989), it was not until September 2015, when the Advanced LIGO detectors began their first operational run (O1) that gravitational waves were directly detected (Abbott et al., 2016d,f, 2017). These discoveries commenced a new era in astronomy. Second-generation ground-based interferometers are now at the forefront of the nascent field of gravitational wave astronomy, with a potential for hundreds of discoveries to be made per year (Abadie et al., 2010, Abbott et al., 2016i). This network includes the two Advanced LIGO detectors in the United States (Aasi et al., 2015), Advanced Virgo in Cascina, Italy (Acernese et al., 2015), and KAGRA in Japan (Somiya, 2012).

Advanced LIGO is currently a network of two ground-based interferometers in Hanford, Washington, and Livingston, Louisiana (Aasi et al., 2015). These instruments have perpendicular L-shaped arms, each 4 km in length. A gravitational wave that passes through the detectors will alternately stretch and compress these arms, generating an interference pattern to be measured by a photo-detector. At their design sensitivity, the Advanced LIGO interferometers are expected to be ten times more sensitive than Initial LIGO, searching 1,000 times the volume of space. The Advanced LIGO interferometers were the only second-generation detectors operating during the discoveries of GW150914, GW151226, and GW170104. A third LIGO detector, called LIGO-India, is expected to be built in India, with the same specifications as the US-based interferometers

(Unnikrishnan, 2013). Additional interferometers in the ground-based network will help improve source localisation and detection confidence (Fairhurst, 2014).

Like the Advanced LIGO detectors, Advanced Virgo is a ground-based Michelson interferometer with 3 km long perpendicular arms, near Pisa, Italy. It is currently undergoing upgrades, and is expected to begin operations in 2017. Likewise, KAGRA is currently under development with plans to operate in 2018. In addition to the state-of-the-art technologies present in the Advanced LIGO and Advanced Virgo detectors, KAGRA will be underground, with plans to use cryogenic mirrors (Somiya, 2012). This will help mitigate thermal noise, which limits the low frequency sensitivity in ground-based detectors.

The European Space Agency (ESA) is planning a space-based mission, called the Laser Interferometer Space Antenna (LISA) (Amaro-Seoane et al., 2016). The goal is to send three satellites with test masses into space, forming an equilateral triangle with vertexes separated by 2.5×10^6 km. The satellites will be strategically placed (in an Earth-trailing heliocentric orbit), and the distances between them will be monitored using quasi-Michelson interferometry to detect distortions in space-time caused by passing gravitational waves. As ground-based detectors are limited in their low frequency sensitivity due to seismic noise, they cannot resolve frequencies much lower than 10 Hz. However, the LISA satellites will be sensitive to low frequency gravitational waves from 0.1 mHz to 1 Hz, and may detect sources such as binary supermassive black holes. Results from the pilot mission, LISA Pathfinder, are promising (Armano et al., 2016), demonstrating great potential for the future success of LISA. The expected LISA launch date is set for 2034.

The Einstein telescope (Sathyaprakash et al., 2012) will be part of the third generation of ground-based gravitational wave detectors. Similar to KAGRA, it will use cryogenic mirrors and will be built underground to limit thermal noise. The Einstein telescope will be a set of triangular interferometers like LISA, but with 10 km long arms.

As pulsars produce regular pulses of radio waves, they can be considered to be extremely accurate clocks. In contrast to interferometric detectors, the International Pulsar Timing Array project (Hobbs et al., 2010) aims to utilise this property of pulsars, comparing the observed pulse arrival times with predicted times, in the hopes of discovering extremely low frequency ($\sim 10^{-9}$ – 10^{-8} Hz) gravitational waves, from sources such as binary coalescing supermassive black holes.

2.3.1 Advanced LIGO

Advanced LIGO consists of two second generation ground-based gravitational wave detectors, both in the USA: one in Hanford, Washington, and the other in Livingston, Louisiana (Aasi et al.,

2015). Each instrument is a Michelson interferometer with two perpendicular 4 km arms. An aerial view of these observatories can be seen in Figure 2.2.



(a) Hanford, Washington.

(b) Livingston, Louisiana.

FIGURE 2.2: The Advanced LIGO interferometers in the United States. Both ground-based detectors have two perpendicular arms, each 4 km in length. Photos courtesy of Caltech/MIT/LIGO Laboratory (LIGO).

The Advanced LIGO detectors use a method called *laser interferometry*. A beam of light is split into two beams and sent in a vacuum down two equal-length perpendicular tunnels. Test masses are suspended on pendulums at the end of each arm to imitate free-fall, with special mirror coatings to reflect light. Since light has a constant speed by Einstein's special relativity, and the tunnels are the same length, when the light beams recombine at the photo detector, they will have a destructive interference pattern, and no light will be detected. However, if a detectable gravitational wave happens to pass through the interferometer, it will change the relative lengths of each arm, causing a constructive interference pattern. This interference pattern is then analysed.

Fabry-Pérot cavities are used to increase the interaction time between the gravitational wave and light by increasing the effective length of the interferometers (from 4 km to about 1120 km long). This improves detector sensitivity, particularly at lower frequencies. A Fabry-Pérot cavity contains two special mirrors that reflect light back and forth (about 280 times), storing photons for about 1 ms to lengthen the interaction time gravitational waves have with the light.

To improve the resolution of the detectors, the method of *power recycling* is used. A power recycling mirror is placed between the laser source and beam splitter. Most of the light from the interferometer arms reflects back to the power recycling mirror rather than the photodetector, which in turn, reflects back down the arms. This process amplifies the power of the laser.

A schematic of the Advanced LIGO detectors can be seen in Figure 2.3.

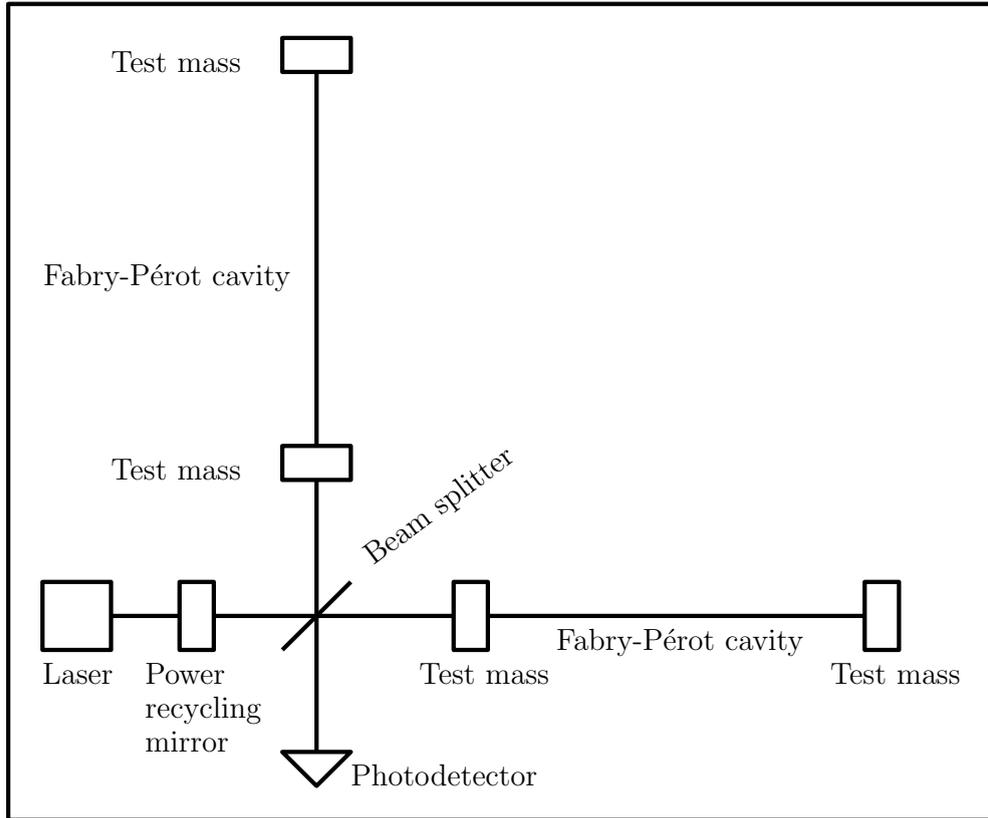


FIGURE 2.3: Schematic of an Advanced LIGO gravitational wave interferometer (not to scale).

The two detectors are separated by a large distance to mitigate local noise events that could be misinterpreted as gravitational wave triggers/signals. A third detector in the Advanced LIGO network is being planned in India (Unnikrishnan, 2013). Additional ground-based gravitational wave detectors in the network will not only help improve source localisation in the sky, but will also improve detection confidence (Fairhurst, 2014).

Gravitational wave observations from ground-based interferometers, such as Advanced LIGO and Advanced Virgo, are subject to many sources of noise. It is important to characterise environmental and instrument noise to make accurate detection statements and astrophysical inferences. At its design sensitivity, Advanced LIGO will be sensitive to gravitational waves in the frequency band from ~ 10 Hz to 8 kHz. The main noise sources for ground-based interferometers include seismic noise, thermal noise, and quantum noise (Aasi et al., 2015). Seismic noise is caused by movements in the ground from phenomena such as earthquakes, ocean waves, wind, etc., and limits the low frequency sensitivity of the detectors. Thermal noise is the predominant noise source in the most sensitive frequency band of Advanced LIGO (around 100 Hz), and arises from the test mass mirror suspensions and the Brownian motion of the mirror coatings. Quantum mechanics limits the accuracy to which the position of the test masses can be measured. Quantum noise comes from two sources: photon shot noise and radiation pressure. Photon shot noise is due to quantum uncertainties in the detected photon arrival rate, and dominates the high frequency

sensitivity of the detectors. Radiation pressure, caused by momentum imparted to the mirrors upon reflection, limits detector sensitivity at low frequencies. The Advanced LIGO noise power spectrum also contains, sharp-featured, high power *spectral lines* at particular frequencies (which may wander). These are caused by the AC electrical supplies and mirror suspensions, among other phenomena.

Each LIGO detector records approximately 200,000 auxiliary channels to monitor environmental noise and instrument behaviour. Electromagnetic noise from sources such as lightning, solar winds, radio frequency communication, and cosmic ray showers, as well as anthropogenic noise from human activities near the detectors are monitored (Abbott et al., 2016b). These noise sources are monitored using instruments such as seismometers, magnetometers, microphones, and accelerometers. The amplitude of noise disturbances in these auxiliary channels are compared to the amplitude of a gravitational wave event to rule out noise triggers. Before analysing gravitational wave strain data, data quality vetoes are created. Time segments are removed where identified instrumental or environmental noise are coupled with the gravitational wave strain channel (Nuttall et al., 2015).

2.4 Stellar core collapse

Main sequence stars convert hydrogen to helium through nuclear fusion in the core, releasing vast amounts of energy. When a star is hot enough, courtesy of gravity, heavier elements will fuse together. As elements fuse together, there is an outward thermal pressure that ebbs and flows in a delicate balance with the inward pull of gravity throughout a star's main sequence.

At the end of a massive star's (greater than about eight solar masses) life, it will run out of nuclear fuel when an iron core is formed — iron only absorbs energy and can therefore not fuse into heavier elements, nor can it release energy by fission. At this point, no more energy is released and the outward thermal pressure can no longer equilibrate the inward pull of gravity, causing the core to collapse on itself.

Gravitation then overcomes electron degeneracy pressure (the pressure due to electrons being compressed into small volumes). Protons and electrons are then forced together into neutrons, over-riding the weak nuclear force (the force that is involved in the decay of neutrons to protons, electrons, and neutrinos). Due to conservation laws, neutrinos are also released during this process, carrying unfathomable amounts of energy away from the core. This causes the core to cool down and compress more.

The inner core then collapses until it reaches nuclear density (the density of an atomic nucleus), at which point the strong nuclear force (the force that holds the nucleus of an atom together against the repulsion of protons) takes over, causing the inner core to “bounce”, creating a shock wave that blasts into the inward falling outer core. If the shock wave stalls, collapse ensues, leading to black hole formation. However, the shock wave could be regenerated by some underlying mechanism, heating the star up, and producing new elements. When the shock wave reaches the surface, the star explodes in a brilliant supernova. Depending on the final mass of the core, either a black hole or neutron star is formed.

The mechanism generating the shock wave revival is currently up for debate among theorists (Logue et al., 2012), though the two front-runners are the *neutrino mechanism* (Bethe and Wilson, 1985) and the *magnetorotational mechanism* (Shibata et al., 2006). The neutrino mechanism relies on the notion that some high-energy neutrinos are trapped behind the stalled shock and this could be the underlying mechanism regenerating the shock wave. The magnetorotational mechanism suggests that the shock wave is regenerated by the strong differential rotation in the outer protoneutron star.

Electromagnetic observations of stellar core collapse can only directly probe the envelope of a star, and not the core. Like neutrinos, gravitational waves are emitted deep in the core of a progenitor and traverse the Universe mostly unobscured by astrophysical objects between the source and a detector on Earth. Gravitational waves from stellar core collapse will provide direct observations about the multi-dimensional core collapse dynamics and supernova mechanisms, which could lead to many new insights and theoretical developments.

2.5 Gravitational wave data analysis

Direct detections of gravitational waves have not only required the development of cutting-edge technologies, but have also relied on sophisticated data analysis techniques.

To begin analysing gravitational wave data $y(t)$, one may assume it is a time series of the following form:

$$y(t) = s(t; \boldsymbol{\theta}) + \epsilon(t), \quad t = 1, \dots, n, \quad (2.1)$$

where $s(t; \boldsymbol{\theta})$ is a (deterministic) gravitational wave signal with parameters $\boldsymbol{\theta}$, and $\epsilon(t)$ is additive background/detector noise.

The field of gravitational wave data analysis can be broadly categorised into three key domains: *detection significance*, *parameter estimation*, and *noise modelling*. Detection significance refers to

the statistical methodologies used to calculate a *false alarm probability* and *false alarm rate*, both of which determine the level of significance of a candidate gravitational wave. Parameter estimation involves using statistical methods (mostly Bayesian) to extract important astrophysical parameters (e.g., mass, spin) encoded in gravitational wave signals. Noise modelling involves estimating the power spectral density (PSD) of background/detector noise. This step has historically received the least amount of attention and is often ignored in favour of a default method (stationary, Gaussian, known PSD), though one could argue that it is just as important as parameter estimation, as a poor understanding of noise could lead to systematic biases in estimates of important astrophysical parameters.

When using multiple detectors to search for gravitational waves, methods can be classed as either *coincident* or *coherent*. Coincident methods involve creating a short-list of gravitational wave triggers/candidates using data that look similar between detectors and are coincident in time (within a certain propagation time between detectors). Coincidence works by assuming that detectors separated by large distances should have uncorrelated noise, meaning the probability of accidental coincidence should be small (Maggiore, 2008). This may be the case for most noise sources, though magnetic fields from Schumann resonances have recently been shown to negate the assumption of uncorrelated noise (Kowalska-Leszczynska et al., 2016). Coherent methods use all data from each detector simultaneously and are extremely useful for rejecting background noise glitches.

2.5.1 Detection significance

Well-modelled gravitational wave sources, such as CBC, generally have a large number of *template* signals with various parameter configurations. In these cases, template bank searches (i.e., *matched filtering*) are used to find an optimal match between templates and the true signal. Matched filtering maximises the output signal-to-noise ratio (SNR) and involves forming a convolution between the data and the template. Matched filters are only optimal in the Gaussian white noise case, but still maximise SNR when noise is non-Gaussian.

Two independent LIGO Scientific Collaboration (LSC) pipelines, called PyCBC (Dal Canton et al., 2014) and GstLAL (Cannon et al., 2012) use template-based search methods to identify coincident candidate events that are within 10 ms of each other (to take into account the time a gravitational wave will take to travel from one interferometer to another). A *detection statistic* (or ranking statistic) is used to rank triggers, and is then compared to an estimated background to calculate the false alarm probability and the level of significance of a trigger.

Detection statistics are calculated using traditional goodness-of-fit tests. PyCBC uses a chi-squared statistic to determine whether data in different frequency bins follow what would be expected from a particular template. This is then used to rescale the SNR. The reweighted SNR is then used to rank triggers. GstLAL, on the other hand, uses the likelihood ratio test.

To calculate the background data set necessary to assess the significance of triggers, a time-slide technique similar to randomisation/permutation tests is conducted. The idea here is that one detector's data are shifted by multiples of 0.1 s, and a new detection statistic is computed. This is replicated many times and the number of replicated detection statistics at least as great as the observed one are counted. Poisson statistics are then used to determine a false alarm probability.

The PyCBC and GstLAL pipelines were both used to determine detection confidence and significance in the discovery of GW150914 (Abbott et al., 2016c).

In contrast to inspiral gravitational waves, it is not possible to use template-based search methods like matched filtering for unmodelled bursts (such as core collapse supernovae) as these signals are not well-modelled and do not have a well-understood phase evolution. Alternative methods are needed for these signals. Excess power searches in the time-frequency domain via wavelets are a popular approach. The difficulty with such methods is distinguishing an unmodelled burst signal from noise transients, often called *glitches*.

Excess power search methods such as the coherent WaveBurst (cWB) pipeline (Klimenko et al., 2008) and X-Pipeline (Sutton et al., 2010) are useful for identifying gravitational wave bursts. The cWB pipeline is a coherent method that uses a wavelet transformation and constrained likelihood approach, looking for excess power in the time-frequency domain. One version of cWB is a low-latency transient search that aims to provide fast alerts for electromagnetic follow-up (Abbott et al., 2016e). It also provides the first estimates of sky location and other parameters. Regarding detection significance, the cWB detection statistic is a function of coherent energy, and false alarm rates are computed using the same time-slide approach as CBC signals.

Methods used to determine detection significance are primarily frequentist in nature. However, BayesWave and omicron-LALInference-Bursts (oLIB) are alternative approaches that use the Bayesian framework. BayesWave (Cornish and Littenberg, 2015) is a follow-up analysis to cWB that models signals and glitches using a Morlet-Gabor continuous wavelet basis and reversible jump MCMC (RJMCMC) for parameter estimation (Green, 1995). A Bayes factor is calculated, comparing a signal model against a glitch model (Abbott et al., 2016g). This is then used as a detection statistic.

Like cWB, oLIB (Lynch et al., 2016) is another low-latency search pipeline that searches for excess power events. This is a coincident method where short-listed triggers are fitted using wavelets,

with key parameters estimated using Bayesian methods. As signals have a coherent phase across detectors and glitches do not, Bayes factors are calculated for a coherent signal against Gaussian noise, and a coherent signal against an incoherent glitch. The ratio of these Bayes factors is used as the detection statistic.

2.5.2 Parameter estimation

The Bayesian statistical framework has proven to be a powerful tool for parameter estimation in astrophysical and cosmological settings (Loredo, 1992). Christensen et al. (2001) were the first to apply Bayesian methods (particularly the Metropolis-Hastings algorithm) to infer cosmological parameters from the cosmic microwave background. Bayesian gravitational wave data analysis was also first pioneered by Christensen and Meyer (1998). Christensen and Meyer (2001) then demonstrated the usefulness of the Gibbs sampler (Geman and Geman, 1984) for estimating five physical parameters from coalescing binary signals. Christensen et al. (2004b) then went on to show how a custom-built Metropolis-Hastings algorithm (Hastings, 1970, Metropolis et al., 1953), a generalisation of the Gibbs sampler, was a superior and more suited routine for eventual implementation into the LSC algorithm library (LAL).

Parameter estimation for compact binary inspirals has subsequently become more sophisticated in recent years (see for example Aasi et al. (2013), Raymond et al. (2009), Röver et al. (2006, 2007a,b), van der Sluys et al. (2008), Veitch and Vecchio (2010)). The `LALInference` library (Veitch et al., 2015) utilises many of these earlier developments, and was integral in the parameter estimation studies of the recent discoveries of binary black hole mergers (Abbott et al., 2016h). MCMC routines for inferring the physical parameters of pulsars have also been developed (Christensen et al., 2004a, Clark et al., 2007, Umstätter et al., 2004).

Sophisticated burst parameter estimation methods have been developed in recent times. As mentioned in the previous subsection, `BayesWave` (Cornish and Littenberg, 2015) is a very general approach to modelling unknown burst signals and transient noise glitches using a Morlet-Gabor wavelet basis with reversible jump MCMC. Bayes factors are used to rule out glitches. This is a big improvement over the existing methods for making statements about completely unmodelled bursts, which have generally only focused on sky localisation (Essick et al., 2015).

Specialised Bayesian parameter estimation routines have also been developed for neutron star r -modes (Coughlin et al., 2014) and stellar core collapse. Heng (2009) and Röver et al. (2009) used principal component regression models to reconstruct rotating core collapse signals. Logue et al. (2012) and Powell et al. (2016) used nested sampling to infer an underlying supernova

mechanism. Abdikamalov et al. (2014) used matched filtering and nested sampling to infer total angular momentum for rotating progenitors.

Posterior predictive sampling was used to infer the ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce, as well as differential rotation in the paper by Edwards et al. (2014). The results from that paper are also included in Chapter 4.

2.5.3 Noise modelling

As mentioned earlier in 2.3.1, there are many sources of environmental and detector noise. Noise limits detector sensitivity and must be accurately modelled to ensure astrophysical parameter estimates are accurate, minimising potential systematic biases.

Noise modelling has often been ignored in parameter estimation routines in favour of a default model. This default model assumes that gravitational wave noise is stationary and Gaussian distributed, with a known PSD that is estimated using off-source data (not on a candidate signal) via the Welch method (Littenberg et al., 2013). However, it has been found that these assumptions are too strict for real gravitational wave data, which are often non-Gaussian and nonstationary (Christensen, 2010). Littenberg and Cornish (2015) demonstrated that the noise PSD in noise from LIGO’s sixth science run (S6) (Abadie et al., 2012b) was in fact time-varying. It was also demonstrated by Aasi et al. (2013) that fluctuations in the PSD can moderately bias parameter estimates of binary coalescence gravitational wave signals in LIGO S6 data. The time-varying nature of real LIGO data invalidates the stationarity assumption. It is also common to see high amplitude non-Gaussian transient noise (glitches) in real detector data. This voids the Gaussian noise assumption, which could be particularly problematic for unmodelled burst signals.

There have been attempts reported in the literature to improve the modelling of noise present in gravitational wave data, primarily concentrating on noise with embedded signals from well-modelled gravitational wave sources, such as binary inspirals. Of note, Röver et al. (2011) used a Student-t likelihood under the Bayesian framework, Littenberg et al. (2013) included additional scale parameters and marginalised over uncertainty in the PSD, and Vitale et al. (2014) used iteratively reweighted least squares to analytically marginalise out background noise. Regarding noise models for unmodelled bursts, Littenberg and Cornish (2015) implemented the BayesLine algorithm, using a cubic spline to model the smooth-changing broadband noise spectrum while modelling spectral lines using Lorentzians, and Cornish and Littenberg (2015) implemented the BayesWave algorithm to model non-Gaussian transient glitches.

One of the main focuses of this thesis is to create novel solutions to the noise modelling problem using Bayesian nonparametric methods. Materials from Edwards et al. (2015), Edwards et al.

(2017a), and Kirch et al. (2017) will be presented in Chapters 3 and 5. In Edwards et al. (2015), the nonparametric Bernstein polynomial prior of Petrone (1999a,b) and Choudhuri et al. (2004) was used to model simulated Advanced LIGO noise. A rotating stellar core collapse gravitational wave signal was simultaneously extracted. Data were also broken into locally stationary components to deal with nonstationarities. However, Bernstein polynomials proved to be inadequate for modelling the complex structure of real Advanced LIGO noise due to the presence of sharp and abrupt spectral lines. A generalisation to the Bernstein polynomial prior, called the B-spline prior, was later developed in Edwards et al. (2017a) to improve coverage and to adequately model data from the LIGO S6 run. The local support property of B-splines allowed for modelling of sharp peaks in the log-spectral density. Finally, in Kirch et al. (2017), a nonparametric correction to an autoregressive parametric likelihood was used to generalise on the commonly used Whittle likelihood. In this approach, one can take advantage of a parametric model's efficiency while mitigating misspecifications via a nonparametric adjustment. This approach was also applied to LIGO S6 data.

Chapter 3

Statistical methods

3.1 The Bayesian paradigm

In this thesis, the Bayesian paradigm is the chosen philosophy to conduct gravitational wave data analysis in. Not only is Bayesian probability theory well-suited to astronomy from a philosophical perspective (Loredo, 1992), it can be considered an extended theory of logic (Gregory, 2005, Jaynes, 2003). That is, Bayesian probability theory may be considered a generalisation of logic that deals with levels of uncertainty rather than truths. Updating one's pre-existing state of knowledge with new data thus gives the user a completely logical framework to draw conclusions.

Given this, it is important to review the key Bayesian ideas that form the essence of this thesis. This includes the notions of Bayes' theorem, posterior inference, posterior prediction, and model selection. Of particular importance is the posterior predictive distribution, which is used in Chapter 4 to predict important astrophysical information encoded in rotating stellar core collapse supernova gravitational wave signals, such as the ratio of rotational kinetic energy to gravitational potential energy. Just as important are model selection methods, which are also used in Chapter 4 to determine an appropriate statistical model for reconstructing these signals from a set of principal component basis functions.

As this is only a brief review of a field with broad scope, not all topics can be covered and the reader is referred to the following text books by Gelman et al. (2013), Gilks et al. (1996), Gregory (2005), Jaynes (2003) for more detailed coverage.

3.1.1 Bayesian inference

Bayesian inference involves updating one's prior beliefs about some unknown parameters $\boldsymbol{\theta}$ using observable data \mathbf{y} . This is done by assigning probability distributions to the data, as well as the unknowns.

Bayesian inference requires three pivotal quantities. The *likelihood* function $p(\mathbf{y}|\boldsymbol{\theta})$ is the probability density function (pdf) of the data \mathbf{y} , conditional on the random vector of model parameters $\boldsymbol{\theta}$. The *prior* $p(\boldsymbol{\theta})$ is the pdf of the model parameters that takes into account all of the information known about $\boldsymbol{\theta}$ before the data are observed. The *posterior* $p(\boldsymbol{\theta}|\mathbf{y})$ is the updated pdf of the model parameters after the data are observed. These quantities are related via *Bayes' theorem*

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (3.1)$$

where $p(\mathbf{y}) = \int p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}$ is called the *marginal likelihood*, *evidence*, or *prior predictive distribution*. This is the probability of observing the data regardless of the parameter configuration. This quantity is often not computed when the dimension of the parameter space d is large, and is treated as a *normalising constant*, as it is independent of model parameters $\boldsymbol{\theta}$. The posterior can thus be thought of as being proportional to the product of the prior and the likelihood:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}). \quad (3.2)$$

It is of interest to sample values from the posterior to make inferences about the model parameters after observing the data. These points are referred to as *posterior samples*.

3.1.2 Summarising the posterior

One can summarise the posterior distribution of parameter θ using point estimators and credible intervals. A point estimator (e.g., posterior expectation, posterior median, and posterior mode) of θ can be written as $\hat{\theta}$.

The posterior expectation is given by

$$\hat{\theta} = \int \theta p(\theta|\mathbf{y})d\theta, \quad (3.3)$$

the posterior median is $\hat{\theta}$ such that

$$\int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{y})d\theta = \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{y})d\theta = 0.5, \quad (3.4)$$

and the posterior mode is

$$\hat{\theta} = \arg \max_{\theta} p(\theta|\mathbf{y}). \quad (3.5)$$

An equal-tailed $100(1 - \alpha)\%$ credible interval (c_l, c_h) is such that $(\alpha/2)\%$ of the posterior density is below c_l and $(\alpha/2)\%$ above c_h . Credible intervals are useful for making probabilistic statements such as “there is a $100(1 - \alpha)\%$ probability that θ is between c_l and c_h ”.

3.1.3 Posterior prediction

Once the data \mathbf{y} are observed, predictions of new observable data \mathbf{y}_{new} can be made by sampling from the *posterior predictive distribution*. This is computed by taking the product of the posterior distribution and the likelihood of the new data, and marginalising over the model parameters. That is,

$$p(\mathbf{y}_{\text{new}}|\mathbf{y}) = \int p(\mathbf{y}_{\text{new}}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}. \quad (3.6)$$

Alternatively, one can view this as the expected value of the sampling distribution of the new data, when taken over the posterior distribution. That is,

$$p(\mathbf{y}_{\text{new}}|\mathbf{y}) = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}} [p(\mathbf{y}_{\text{new}}|\boldsymbol{\theta})]. \quad (3.7)$$

The posterior part of posterior predictive distribution comes from conditioning on data \mathbf{y} , and the predictive part comes from the prediction of observable \mathbf{y}_{new} .

3.1.4 Model selection

Bayes factors can be used in the Bayesian analogue of classical hypothesis testing, and are often used in model comparison. Consider the two hypotheses $H_0 : \boldsymbol{\theta} \in \Theta_0$ and $H_1 : \boldsymbol{\theta} \in \Theta_1$, such that $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$. A Bayes factor is the ratio of marginal likelihoods under these competing hypotheses, and can be calculated using the ratio of posterior odds to prior odds, where odds quantify how plausible H_1 is compared to H_0 . The Bayes factor in favour of H_1 over

H_0 is given by:

$$BF_{10} = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} \quad (3.8)$$

$$= \frac{p(H_1|\mathbf{y})}{p(H_0|\mathbf{y})} \bigg/ \frac{p(H_1)}{p(H_0)}. \quad (3.9)$$

A large BF_{10} provides evidence in favour of H_1 , giving the practitioner a useful tool for model comparison and selection.

Alternatively, one can use approximations such as the *deviance information criterion* (DIC) (Spiegelhalter et al., 2002). DIC is a generalisation of the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for hierarchical models, making similar use of the quality of fit statistic, *deviance*. Deviance is defined as $D = -2 \log p(\mathbf{y}|\boldsymbol{\theta})$, where $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood of a statistical model. DIC has the following formulation:

$$\text{DIC} = \bar{D} + p_D \quad (3.10)$$

$$= 2\bar{D} - D(\bar{\boldsymbol{\theta}}), \quad (3.11)$$

where

$$\bar{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}], \quad (3.12)$$

$$\bar{D} = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}[D], \quad (3.13)$$

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}). \quad (3.14)$$

Here \bar{D} is the mean deviance over posterior samples, calculated by evaluating the deviance for each of the stored model parameters $\boldsymbol{\theta}$ that have been sampled from their joint posterior pdf, and taking the average, p_D is the effective number of parameters, and $D(\bar{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior means of the parameters, calculated by finding the posterior mean of each of the model parameters $\bar{\boldsymbol{\theta}}$ and then evaluating the deviance. When comparing competing statistical models, the lowest DIC is preferred. \bar{D} is a relative measure of fit (where a smaller value indicates a better fit), and p_D is a measure of model complexity used to penalise models with too many parameters. Equation (3.10) therefore illustrates how DIC incorporates Occam's razor into the model selection process, allowing one to select a parsimonious model, balancing between fit and complexity. Equation (3.11), on the other hand, provides a simple method for computing DIC.

When conducting Bayesian model comparison, DIC may be preferred to Bayes factors for two reasons. First, Bayes factors require computing the marginal likelihood (or evidence) from equation (3.1), which often involves multi-dimensional integration over a large number of parameters. Numerical techniques such as *nested sampling* (Skilling, 2006) can be used to derive the evidence

of competing models, but these methods require significant computational power and time. On the other hand, DIC is readily computed from posterior samples. The second reason for using DIC over Bayes factors is that improper priors do not violate any conditions of use. Bayes factors, however, are no longer relevant when improper priors are used due to the so-called *marginalisation paradox* (Stone and Dawid, 1972).

A popular alternative to the model selection problem comes in the form of the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm of Green (1995). This algorithm allows for movement between spaces of differing dimension, such as models with different numbers of parameters. A benefit of this approach is that uncertainty in model dimensionality can be accounted for. Model selection is also essentially automatic, as one could either conduct model averaging or choose the posterior mode. However, these algorithms are often fraught with implementation difficulties. RJMCMC will be discussed in more detail in 3.2.6.

The Bayesian nonparametric paradigm arguably provides the best solution to the model selection problem by allowing for a data-driven choice of model complexity. Rather than fitting many competing models as is done when using Bayes factors and DIC, Bayesian nonparametric methods involve fitting one model, and adapting its complexity to the data (Gershman and Blei, 2011). Bayesian nonparametric methods will be discussed in 3.3.

3.2 Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are at the heart of modern Bayesian inference. They are extremely useful in a number of statistical problems, and are used to sample from complicated probability distributions when no other sampling mechanisms are available. These methods have already proven to be invaluable in the context of gravitational wave data analysis, with the pioneering work of Christensen and Meyer (1998) paving the way for conducting parameter estimation in the recent discoveries of binary black hole gravitational waves (Abbott et al., 2016d,f, 2017). A concise review of Markov chains and the key algorithms used in this thesis is provided within this section.

3.2.1 Monte Carlo integration

In Bayesian inference, one is often interested in evaluating integrals involving the posterior distribution. If it is not possible to analytically integrate these integrals, stochastic numerical methods are required. One such method is *Monte Carlo integration* — the process of approximating an

expectation using the sample mean of a function of simulated random variables. Let X be a random variable with probability density function f , and let h be a function of X . Let N be the (large) number of samples drawn from f . The value of h is calculated for each of these samples before averaging. The *Monte Carlo estimate* of h is thus given as:

$$\mathbb{E}_f[h(X)] = \int h(x)f(x)dx \quad (3.15)$$

$$\approx \frac{1}{N} \sum_{i=1}^N h(x_i). \quad (3.16)$$

Analogously, other quantities such as quantiles may be estimated using posterior samples.

Justification of Monte Carlo integration comes from the strong law of large numbers, which states that a sample mean \bar{X} of a random sample of size N from a distribution with true mean μ converges to μ almost surely as N tends to infinity.

If it is not easy to sample from f , alternative methods such as *importance sampling* can be employed, though for brevity, such methods will not be discussed here, but can be found in Gelman et al. (2013).

3.2.2 Markov chains

Markov chains play a fundamental role in MCMC simulations (Gilks et al., 1996). A (discrete-time) Markov chain is a stochastic process that satisfies the *Markov property*. That is, the sequence of random variables $\{X_0, X_1, \dots\}$ is a Markov chain if for state space S ,

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}), \quad (3.17)$$

where X_n is the random variable representing the state of the system at time n . This means that the probability of moving to the next state only depends on the current state. In other words, given the present state, future and past states are independent. Note that state space S is usually continuous (and often \mathbb{R}^d). This is certainly the case for the gravitational wave applications presented in this thesis. However, the following concepts are more straightforward to explain using discrete state spaces (keeping the continuous state space in mind) (Gilks et al., 1996).

The probability of moving from state i to state j is called the *transition probability* and is written as

$$p_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i). \quad (3.18)$$

Let n be the number of states in state space S . A *transition matrix* P is an $n \times n$ matrix whose entries are transition probabilities from each state to every other state ($p_{ij}, i, j = 1, \dots, n$).

The vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n)$ is called a *stationary distribution* if $\mathbb{P}(X_0 = i) = \pi_i$ implies $\mathbb{P}(X_n = i) = \pi_i, \forall n \geq 0$. A stationary distribution can be found by solving the *full balance* equations,

$$\boldsymbol{\pi}P = \boldsymbol{\pi}. \quad (3.19)$$

In addition, the following constraint must hold to ensure a full rank system of equations is solved:

$$\sum_{i \in S} \pi_i = 1. \quad (3.20)$$

Alternatively, it is often simpler to find a stationary distribution by solving the *detailed balance* equations,

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad \forall i, j \in S. \quad (3.21)$$

A Markov chain is *reversible* if there exists a stationary distribution $\boldsymbol{\pi}$ such that detailed balance is satisfied. Detailed balance is not a necessary condition, but is a sufficient condition for a Markov chain to have a stationary distribution, implying all reversible Markov chains have stationary distributions. Reversibility is an important concept for many MCMC samplers, such as the Metropolis-Hastings algorithm (Hastings, 1970, Metropolis et al., 1953), and the Gibbs sampler (Geman and Geman, 1984), as it provides a simple condition to ensure the desired stationary distribution is found.

A Markov chain is *irreducible* if it is possible to move from each state to every other state in a finite number of moves. A Markov chain is *aperiodic* if it does not return to a state at regular intervals.

For MCMC sampling, it is crucial that the target distribution of interest is the stationary distribution of the Markov chain, and that the Markov chain is *ergodic*. Ergodicity means the stationary distribution can be reached regardless of the initial state. For a chain to be ergodic, it must be *irreducible* and *aperiodic*. A Markov chain has a unique stationary distribution if it is irreducible and aperiodic. This is also the *limiting distribution* for the Markov chain.

3.2.3 Markov chain Monte Carlo simulation

It is often possible to directly sample points from named posterior distributions using pseudo-random number generators. Posterior samples are often generated using algorithms such as the

inverse-transform method and the *acceptance-rejection* method. However, when this is not possible, as is the case with many complex statistical problems, MCMC methods are extremely useful and valuable.

MCMC methods work by constructing a Markov chain whose stationary distribution is the target (posterior) distribution of interest. One can generate (dependent) samples from the support of the posterior in the correct proportions using this Markov chain. One can then use these samples to compute Monte Carlo estimates to approximate integrals of interest from the posterior, hence the name Markov chain Monte Carlo.

The simplest and most popular MCMC methods are the Metropolis-Hastings algorithm and a special case called the Gibbs sampler, which will be described in 3.2.4 and 3.2.5 respectively. Other useful MCMC algorithms include reversible jump MCMC (Green, 1995) (which will be described in 3.2.6), as well as auxiliary methods such as the slice sampler (Neal, 2003) and Hamiltonian Monte Carlo (Neal, 2011).

It is important to consider the *mixing* of Markov chains in MCMC routines, ensuring there is a reliable exploration of the parameter space, at an efficient rate. The faster an algorithm mixes, the faster the dependence between successive iterations decays, and the faster it converges. One may need to *tune* the algorithm if mixing is too slow.

After a certain *burn-in* period, a Markov chain is said to have *converged* to its stationary distribution. These burn-in samples are discarded. Though this step could be considered wasteful from a computational perspective, it ensures the stationary distribution is less dependent on starting values. The burn-in period may be long if the chain has poor mixing, getting stuck in small areas of the state space.

Due to ergodicity, an MCMC chain will eventually converge to its target distribution. However, an inappropriate starting value for a chain can affect the speed of convergence, so one must exercise caution when selecting an initial state.

It is common to assess convergence visually using trace plots. Trace plots show the values that a parameter takes during the running of an MCMC algorithm. One can often assess how well the chain is mixing and how long the burn-in period should be based on these plots. It can also be valuable to see if multiple runs of the chain (using a diffuse set of starting points) end up at the same stationary distribution. Empirical diagnostics such as the Geweke time series diagnostic, the Gelman and Rubin multiple sequence diagnostic, the Raftery and Lewis integrated diagnostic, and the Heidelberger and Welch diagnostic are also popular (Cowles and Carlin, 1996).

One may decide to *thin* Markov chains by taking every i^{th} sample (e.g., every tenth sample) to reduce disk space. This will also help reduce autocorrelation in the Markov chain, which may lead

to independent and identically distributed (iid) samples. Due to the ergodic theorem, it is not necessary to have iid samples from the posterior to make inferences, though one would need more dependent samples to achieve the same accuracy as with iid samples.

3.2.4 Metropolis-Hastings algorithm

The *Metropolis algorithm* (Metropolis et al., 1953) is a random walk sampler where a *candidate* for the *target* (posterior) distribution is sampled by centering on the sample from the previous iteration and adding noise from a symmetric *proposal* distribution. The candidate is either accepted or rejected according to an acceptance probability. To sample from the (univariate) posterior distribution $p(\theta|\mathbf{y})$, the Metropolis algorithm is set up as follows:

1. Initialise by choosing a starting point $\theta^{(0)}$ such that $p(\theta^{(0)}|\mathbf{y}) > 0$.
2. For $i = 1, 2, \dots, N$:
 - (a) Propose a candidate $\theta^{(*)}$ from proposal distribution $J(\theta^{(*)}|\theta^{(i-1)})$.
 - (b) Calculate the *acceptance ratio* as the ratio of densities:

$$r = \frac{p(\theta^{(*)}|\mathbf{y})}{p(\theta^{(i-1)}|\mathbf{y})}. \quad (3.22)$$

- (c) Accept or reject by setting

$$\theta^{(i)} = \begin{cases} \theta^{(*)} & \text{with probability } \alpha = \min(1, r), \\ \theta^{(i-1)} & \text{otherwise.} \end{cases} \quad (3.23)$$

The acceptance-rejection step is implemented by simulating a uniform random number $u \sim \text{Uniform}[0, 1]$, and comparing this to the *acceptance probability* α . That is, accept candidate $\theta^{(*)}$ if $u < \alpha$, and reject otherwise. Often for numerical stability and mathematical convenience, this is computed on the log scale, in which case one would accept $\theta^{(*)}$ if $\log u < \log \alpha$, and reject otherwise.

The proposal distribution for the Metropolis algorithm must be symmetric, i.e., it must satisfy the condition $J(\theta^{(a)}|\theta^{(b)}) = J(\theta^{(b)}|\theta^{(a)})$ for all $\theta^{(a)}$ and $\theta^{(b)}$.

This algorithm, after burn-in and thinning, will converge to a stationary distribution that is the target distribution. The efficiency of the algorithm also depends on choosing an adequate starting point $\theta^{(0)}$, and choosing a proposal distribution that closely envelopes the target distribution.

The optimal acceptance rates for the Metropolis algorithm in one dimension and high dimensions are around 0.44 and 0.23 respectively (Gelman et al., 1996). It is possible to construct an *adaptive* algorithm that is automatically tuned by increasing or decreasing the proposal variance, with the aim of bringing the acceptance rate to its optimum value. Roberts and Rosenthal (2009) discuss many additional examples of adaptive MCMC. Note that if an adaptive method is used, then the proposal distribution would have a subscript i such as $J_i(\theta^{(*)}|\theta^{(i-1)})$.

The *Metropolis-Hastings algorithm* (Hastings, 1970) is a generalisation of the Metropolis algorithm, allowing for asymmetric proposal distributions.

The acceptance ratio of the Metropolis-Hastings algorithm changes to

$$r = \frac{p(\theta^{(*)}|\mathbf{y})}{p(\theta^{(i-1)}|\mathbf{y})} \frac{J(\theta^{(i-1)}|\theta^{(*)})}{J(\theta^{(*)}|\theta^{(i-1)})}. \quad (3.24)$$

3.2.5 Gibbs sampler

The *Gibbs sampler* (Geman and Geman, 1984) is another common MCMC algorithm, where the Markov chain is constructed using a sequence of conditional distributions. It involves alternately setting all parameters constant except for one, then drawing a sample from the conditional distribution given these fixed parameters. The Gibbs sampler can be viewed as a special case of the Metropolis-Hastings algorithm.

Suppose one is interested in sampling from the joint posterior $p(\boldsymbol{\theta}|\mathbf{y})$, where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_d\}$ is a d -dimensional parameter vector. Let $\boldsymbol{\theta}_{-j} = \{\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d\}$ be the subset of the parameter vector $\boldsymbol{\theta}$ without θ_j . The aim is to create a Markov chain that alternately cycles through the full conditional distributions $p(\theta_j|\boldsymbol{\theta}_{-j}, \mathbf{y})$ for $j = \{1, 2, \dots, d\}$.

The Gibbs sampling algorithm is then given as follows:

1. Choose starting values $\boldsymbol{\theta}^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)}\}$.
2. At the i^{th} iteration ($i > 0$), a single Gibbs cycle is completed by drawing from the d full conditional distributions given by:

$$\theta_1^{(i)} \sim p(\theta_1|\theta_2^{(i-1)}, \theta_3^{(i-1)}, \dots, \theta_{d-1}^{(i-1)}, \theta_d^{(i-1)}, \mathbf{y}) \quad (3.25)$$

$$\theta_2^{(i)} \sim p(\theta_2|\theta_1^{(i)}, \theta_3^{(i-1)}, \dots, \theta_{d-1}^{(i-1)}, \theta_d^{(i-1)}, \mathbf{y}) \quad (3.26)$$

$$\vdots \quad (3.27)$$

$$\theta_{d-1}^{(i)} \sim p(\theta_{d-1}|\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_{d-2}^{(i)}, \theta_d^{(i-1)}, \mathbf{y}) \quad (3.28)$$

$$\theta_d^{(i)} \sim p(\theta_d|\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_{d-2}^{(i)}, \theta_{d-1}^{(i)}, \mathbf{y}). \quad (3.29)$$

3. Repeat until the chain has converged.

In the basic form of the Gibbs sampler (where all conditional distributions can be directly sampled), all samples are accepted. The *Metropolis-within-Gibbs* sampler is an extension of the Gibbs sampler where it is not possible to directly sample from some of the full conditional distributions, and random walk Metropolis steps are therefore required within a full conditional sampling step.

Another extension of the algorithm involves sampling from conditional distributions of groups of parameters instead of one at a time. This is called a *blocked* Gibbs sampler. Blocked Gibbs samplers can often lead to speed improvements in code by reducing the length of `for` loops, but often at the expense of slower mixing. However, blocked Gibbs samplers may also improve mixing if well constructed (Roberts and Sahu, 1997). If parameters are correlated then separate Gibbs proposals may often be small and mixing can be slow. Joint proposals can take advantage of this correlation and make proposals that move the chain larger distances through the posterior.

3.2.6 Reversible jump MCMC

Often used for model selection, *reversible jump* Markov chain Monte Carlo (RJMCMC) is a generalisation of the Metropolis-Hastings algorithm, allowing for jumps between spaces of differing dimension (Green, 1995). RJMCMC will be used in 4.3.4 to select the number of principal components in a principal component regression model.

Suppose one is interested in the competing models $\{M_1, M_2, \dots, M_k, \dots\}$, where M_i has continuous parameter space Θ_i and $\boldsymbol{\theta}_i$ is a typical element of length n_i from this space. It is of interest to sample from the target density

$$p(M_i, \boldsymbol{\theta}_i | \mathbf{y}) \propto p(M_i)p(\boldsymbol{\theta}_i | M_i)p(\mathbf{y} | M_i, \boldsymbol{\theta}_i), \quad (3.30)$$

where $p(M_i)$ is the prior model probability, $p(\boldsymbol{\theta}_i | M_i)$ is the within-model prior density, and $p(\mathbf{y} | M_i, \boldsymbol{\theta}_i)$ is the within-model likelihood for M_i .

Moving from model M_i to a higher dimensional model M_j (i.e., $n_j > n_i$) first involves generating a random vector \mathbf{V} of length $n_j - n_i$ from some proposal density $\phi(\cdot)$, where the joint density of \mathbf{V} is given by

$$\phi_{n_j - n_i}(\mathbf{v}) = \prod_{i=1}^{n_j - n_i} \phi(v_i). \quad (3.31)$$

The next step involves proposing a move from $\boldsymbol{\theta}_i$ to $\boldsymbol{\theta}_j = f_{i,j}(\boldsymbol{\theta}_i, \mathbf{V})$ using *jump function*

$$f_{i,j} : \Theta_i \times \mathbb{R}^{n_j - n_i} \rightarrow \Theta_j. \quad (3.32)$$

Here, one is mapping the current state of $\boldsymbol{\theta}_i$ along with the random vector \mathbf{V} to a higher dimensional space.

The proposal is accepted with probability

$$\alpha\{(M_i, \boldsymbol{\theta}_i), (M_j, \boldsymbol{\theta}_j)\} = \min\{1, A_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)\}, \quad (3.33)$$

where $A_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$ is the following acceptance ratio

$$A_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \frac{p(M_j, \boldsymbol{\theta}_j | \mathbf{y})}{p(M_i, \boldsymbol{\theta}_i | \mathbf{y})} \times \frac{r_{ji}(\boldsymbol{\theta}_j)}{r_{ij}(\boldsymbol{\theta}_i)} \times \frac{J_{i,j}^f(\boldsymbol{\theta}_i, \mathbf{v})}{\phi_{n_j - n_i}(\mathbf{v})} \quad (3.34)$$

$$\propto \frac{p(M_j)}{p(M_i)} \times \frac{p(\boldsymbol{\theta}_j | M_j)}{p(\boldsymbol{\theta}_i | M_i)} \times \frac{p(\mathbf{y} | M_j, \boldsymbol{\theta}_j)}{p(\mathbf{y} | M_i, \boldsymbol{\theta}_i)} \times \frac{r_{ji}(\boldsymbol{\theta}_j)}{r_{ij}(\boldsymbol{\theta}_i)} \times \frac{J_{i,j}^f(\boldsymbol{\theta}_i, \mathbf{v})}{\phi_{n_j - n_i}(\mathbf{v})}. \quad (3.35)$$

The probability of proposing to jump from M_i to M_j is given by $r_{ij}(\boldsymbol{\theta}_i)$, and $J_{i,j}^f(\boldsymbol{\theta}_i, \mathbf{v}) = \left| \frac{\partial f_{i,j}(\boldsymbol{\theta}_i, \mathbf{v})}{\partial(\boldsymbol{\theta}_i, \mathbf{v})} \right|$ is the Jacobian (which arises from a change in variables).

If a discrete uniform prior is placed on model dimension, the $p(M_i)$ and $p(M_j)$ in 3.35 cancel out to give the following acceptance ratio

$$A_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) \propto \frac{p(\boldsymbol{\theta}_j | M_j)}{p(\boldsymbol{\theta}_i | M_i)} \times \frac{p(\mathbf{y} | M_j, \boldsymbol{\theta}_j)}{p(\mathbf{y} | M_i, \boldsymbol{\theta}_i)} \times \frac{r_{ji}(\boldsymbol{\theta}_j)}{r_{ij}(\boldsymbol{\theta}_i)} \times \frac{J_{i,j}^f(\boldsymbol{\theta}_i, \mathbf{v})}{\phi_{n_j - n_i}(\mathbf{v})}. \quad (3.36)$$

If $n_j < n_i$ then one proposes to move to a lower dimensional space, and the acceptance ratio becomes

$$A_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = A_{j,i}(\boldsymbol{\theta}_j, \boldsymbol{\theta}_i)^{-1}. \quad (3.37)$$

This is essentially the same as Equation 3.35, but inverts the final ratio to the proposal density over the Jacobian. Also note that \mathbf{v} in this case is the vector of random variables proposed to be removed.

3.2.6.1 Birth-death moves and nested models

The *birth-death* move in RJMCMC algorithms was first introduced by Richardson and Green (1997) to jump between models with varying numbers of mixture components in a mixture model. This step involves proposing either the adding (birth) or removing (death) of a parameter from

the current model in the algorithm, and thus model dimension can increase or decrease by one. This is often intermingled with Metropolis-Hastings *stay* steps, where model dimension remains the same and the current model's parameter space is further explored.

Nested models appear often in model comparison problems where all of the parameters of a smaller model occur in a larger one. That is, the smaller model is a subset of the larger model. Formally, let $M_1 \subset M_2 \subset \dots \subset M_d$, where d is number of competing models. Nested models are among the simplest models handled by RJMCMC, and are easily implemented using birth-death moves.

A common scheme is to assign one third probability to each of the jump up, jump down, and stay proposals. That is, $r_{i,i-1}(\boldsymbol{\theta}_i) = r_{i,i}(\boldsymbol{\theta}_i) = r_{i,i+1}(\boldsymbol{\theta}_i) = \frac{1}{3}$ if $i \neq \{1, d\}$. However, one must take care if $i = 1$ or $i = d$. If $i = 1$ then $r_{1,1}(\boldsymbol{\theta}_1) = r_{1,2}(\boldsymbol{\theta}_1) = \frac{1}{2}$, and if $i = d$ then $r_{d,d-1}(\boldsymbol{\theta}_d) = r_{d,d}(\boldsymbol{\theta}_d) = \frac{1}{2}$.

There are many other types of moving schemes such as *split and merge* (Richardson and Green, 1997) but as these are not used in this thesis, a discussion is omitted for brevity.

3.2.6.2 Automatic scaling: Zeroth order method

To ensure efficient trans-dimensional mixing in a RJMCMC algorithm, it may be necessary to scale the proposal density $\phi(\cdot)$. There are many proposed methods for automatically scaling $\phi(\cdot)$. Of the simplest approaches is the so-called *zeroth order* method (Brooks et al., 2003). This is only effective when the posterior and prior are similar.

The idea first involves a *centering* function $c(\cdot)$ that maps from the space Θ_i to space Θ_j . For the simple nested case, a birth move from the space Θ_i to Θ_{i+1} has centering function

$$c(\boldsymbol{\theta}_i) = (\boldsymbol{\theta}_i, 0). \tag{3.38}$$

The scale parameter σ for proposal density $\phi(\cdot)$ is chosen such that the acceptance probability for a jump between $\boldsymbol{\theta}_i$ and $c(\boldsymbol{\theta}_i)$ is equal to 1. That is,

$$A\{\boldsymbol{\theta}_i, c(\boldsymbol{\theta}_i)\} = 1. \tag{3.39}$$

Other practical considerations involved when implementing RJMCMC can be found in Brooks et al. (2003).

3.2.7 Parallel tempering

Many statistical models, such as finite and infinite mixture models, involve multimodal posterior distributions. If there are many isolated modes separated by low posterior density, it is important to use a sampling technique that mixes Markov chains efficiently, rather than relying on the random walk behaviour of the Metropolis sampler. In order to mitigate poor mixing and to accelerate convergence of Markov chains, *parallel tempering* (also called replica exchange) can be implemented (Earl and Deem, 2005, Swendsen and Wang, 1986). There are many other alternatives (such as diffusive nested sampling (Brewer et al., 2011)), though only parallel tempering will be described here as this is the only method considered to accelerate convergence in this thesis.

The idea of parallel tempering is borrowed from physical chemistry, where a system may be replicated multiple times at a series different temperatures. Higher temperature replicas are able to sample larger volumes of the parameter space, whereas lower temperature replicas may become stuck in local modes. The method works by allowing the exchange of information between neighbouring systems. Information from the high temperature replicas can trickle down to the low temperature systems (including the posterior distribution of interest), providing more representative posterior samples.

In the context of MCMC, parallel tempering involves introducing an auxiliary variable called *inverse-temperature*, denoted T_c^{-1} for chains $c = \{1, 2, \dots, C\}$. This variable becomes an exponent in the target distribution for each parallel chain, $p_c(\cdot)$. That is,

$$p_c(\boldsymbol{\theta}|\mathbf{y}) \propto (p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}))^{T_c^{-1}}, \quad (3.40)$$

where $\boldsymbol{\theta}$ are the model parameters, and \mathbf{y} are the data. If $T_c = 1$, this is the (unnormalised) posterior distribution. All other inverse-temperature values produce tempered target distributions. As $T_c \rightarrow \infty$, the target distribution flattens out. Each chain moves on its own in parallel and occasionally swaps states between chains according to the following Metropolis acceptance ratio:

$$r = \left(\frac{p(\boldsymbol{\theta}_j)p(\mathbf{y}|\boldsymbol{\theta}_j)}{p(\boldsymbol{\theta}_i)p(\mathbf{y}|\boldsymbol{\theta}_i)} \right)^{T_i^{-1} - T_j^{-1}}, \quad (3.41)$$

where information is exchanged between chains i and j and $i < j$.

Alternatively, one may choose to only temper the likelihood such that

$$p_c(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})^{T_c^{-1}}. \quad (3.42)$$

Under this formulation, it is easy to see that if $T_c = 1$, this yields the posterior of interest, and as $T_c \rightarrow \infty$, the likelihood component flattens out, giving more preference to the prior. Therefore, T_c^{-1} balances between the prior and the posterior. In this case, the acceptance ratio is given as:

$$r = \left(\frac{p(\mathbf{y}|\boldsymbol{\theta}_j)}{p(\mathbf{y}|\boldsymbol{\theta}_i)} \right)^{T_i^{-1} - T_j^{-1}}, \quad (3.43)$$

which only depends on the likelihood and inverse temperatures, and not the prior. This set-up would certainly make sense under an informative prior.

When implementing parallel tempering, one may encounter many formidable difficulties, such as choosing an appropriate temperature scheme, how many parallel chains should be used, and how often swaps should be proposed. Parallelising existing code is also a challenge, but the method is extremely powerful when used in conjunction with parallel computing facilities, such as large CPU clusters, as little additional computing time is required (compared to a serial version of the code).

3.3 Bayesian nonparametrics

Though Bayesian nonparametric models have received little attention in the field of gravitational wave data analysis, they are promising alternatives to parametric models, particularly in the case where the parametric model is poorly specified. One such example in the field of gravitational wave data analysis is the Gaussian noise model for terrestrial detectors. One problem with this model is that real data often have high amplitude non-Gaussian glitches, voiding the assumptions of the parametric model. Bayesian nonparametric approaches are used in Chapter 5 to address this issue, though it is first important to understand the framework and the key tools such as the Dirichlet process. This will be followed by a description of the Bernstein polynomial prior, the recently developed B-spline prior, and a nonparametric correction to a parametric likelihood, each of which are used to model the spectral density of gravitational wave detector noise.

3.3.1 Introduction

Statistical models can be classified into two groups — parametric and nonparametric. *Parametric* models have a fixed and finite set of parameters, are relatively easy to analyse, and are powerful when their underlying assumptions are correctly specified. However, if the model is misspecified, inferences will be unreliable. *Nonparametric* models have far fewer restrictions, but are less efficient and less powerful than their parametric counterparts. No assumption about the underlying

distribution of the data is made in nonparametric modelling, and the number of parameters is not fixed (and potentially infinite dimensional). Instead, the effective number of parameters increases with more data, providing the model structure. *Semiparametric* models contain both parametric and nonparametric components.

For example, parametric regression (including linear models, nonlinear models, and generalised linear models) uses the following equation:

$$\mathbf{y} = g(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k | \boldsymbol{\beta}) + \boldsymbol{\epsilon}, \quad (3.44)$$

where \mathbf{y} is the response variable, $g(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k | \boldsymbol{\beta})$ is a function of k explanatory variables (that aim to explain the variability in \mathbf{y}) given some model parameters $\boldsymbol{\beta}$. The statistical error, $\boldsymbol{\epsilon}$, is usually assumed to be iid Gaussian random variables, with 0 mean and constant variance σ^2 . Here, the functional form of $g(\cdot)$ is specified in advance, such as in linear regression, where

$$g(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k | \boldsymbol{\beta}) = \beta_0 \mathbf{1} + \beta_1 \mathbf{x}_1 + \dots + \beta_k \mathbf{x}_k. \quad (3.45)$$

Nonparametric regression has a similar set-up, but assumes that the functional form of $g(\cdot)$ is unknown and to be learnt from the data. In a nonparametric setting, function $g(\cdot)$ could be thought of as an uncountably infinite-dimensional parameter.

Bayesian nonparametrics is the interface between Bayesian inference and nonparametric modelling. It is characterised by large parameter spaces and probability measures over these spaces (Hjort et al., 2010). Bayesian inference allows one to incorporate prior knowledge into a statistical framework, and is particularly powerful when these priors accurately represent one's beliefs. On the other hand, nonparametric methods are useful for constructing flexible and robust alternatives to parametric models. Under certain conditions, the combination of these two ideologies provides an arguably superior statistical framework. An additional benefit of Bayesian nonparametric models is their ability to automatically infer model complexity from the data, without explicitly conducting model comparison.

Bayesian nonparametrics is a relatively nascent field in statistics, and faces many challenges. The most obvious one is the mathematical difficulty in specifying well-defined probability distributions on infinite-dimensional function spaces. Constructing priors on these spaces can be arduous, and in the case of noninformative priors, one must ensure large topological support so as not to put too much mass on a small region of the parameter space. Further, creating computationally convenient algorithms to sample from complicated posterior distributions presents its own set of challenges. It is also important to ensure that the nonparametric posterior distribution is consistent (the truth

is uncovered asymptotically), as some procedures do not automatically possess this quality (Hjort et al., 2010).

Bayesian nonparametric priors (and posteriors) are stochastic processes rather than parametric distributions. Ferguson (1973) provided the seminal paper for the field of Bayesian nonparametrics, introducing the Dirichlet process, an infinite-dimensional generalisation of the Dirichlet distribution, now commonly used as a prior in infinite mixture models. This is a popular model (often expressed as the Chinese Restaurant Process) for classification problems where the number of classes is unknown and to be inferred from the data. A formal definition of the Dirichlet distribution and Dirichlet process can be found in 3.3.2.

3.3.2 Dirichlet process prior

The *beta distribution* has the following pdf:

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad (3.46)$$

$$\propto x^{\alpha-1}(1-x)^{\beta-1}, \quad (3.47)$$

where $x \in (0, 1)$ is the support of the distribution, the shape parameters are positive real numbers (i.e., $\alpha > 0$ and $\beta > 0$), and $\Gamma(\cdot)$ is the gamma function defined as the following improper integral

$$\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx. \quad (3.48)$$

The *Dirichlet distribution* is a multivariate generalisation of the beta distribution, with a pdf defined on the K -dimensional simplex

$$\Delta_K = \left\{ (x_1, \dots, x_K) : x_i > 0, \sum_{i=1}^K x_i = 1 \right\}. \quad (3.49)$$

The pdf of the Dirichlet distribution is defined as

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}{\prod_{i=1}^K \Gamma(\alpha_i)} \prod_{i=1}^K x_i^{\alpha_i-1}, \quad (3.50)$$

where $\alpha_i > 0$ for $i = 1, \dots, K$.

The *Dirichlet process* (DP) is an infinite-dimensional generalisation of the Dirichlet distribution (Ferguson, 1973). It is a probability distribution on the space of probability distributions (i.e.,

each draw from a Dirichlet process is a distribution), and is often used in Bayesian inference as a prior for infinite mixture models (aptly called the *Dirichlet process prior*).

For any finite measurable partition $\{A_1, \dots, A_m\}$ of a measurable set S , if $G \sim \text{DP}(M, G_0)$, then $(G(A_1), \dots, G(A_m)) \sim \text{Dirichlet}(MG_0(A_1), \dots, MG_0(A_m))$, where G_0 is the centre/base distribution (i.e., the prior expected value of G is G_0) with Lebesgue density g_0 , and $M > 0$ is the precision/concentration parameter, where larger M implies the Dirichlet process will concentrate more of its mass around the centre. G has the same support as base measure G_0 .

One of the many representations of the Dirichlet process is Sethuraman's *stick-breaking construction* (Hjort et al., 2010, Sethuraman, 1994), which is incredibly useful for implementing MCMC sampling algorithms. The stick-breaking representation has the following hierarchical structure:

$$G(\cdot) = \sum_{i=1}^{\infty} p_i \delta_{Z_i}(\cdot), \quad (3.51)$$

$$p_i = \left(\prod_{j=1}^{i-1} (1 - V_j) \right) V_i, \quad (3.52)$$

$$Z_i \sim G_0, \quad (3.53)$$

$$V_i \sim \text{Beta}(1, M). \quad (3.54)$$

Here, $\delta_a(x)$ is a Dirac measure (indicator function) that is one if $x = a$ and zero otherwise.

To intuit the stick-breaking process, it may be useful to consider a stick of unit length. The weights p_i associated with points Z_i can be thought of as breaking this stick randomly into an infinite number of segments. First, break the stick at location $V_1 \sim \text{Beta}(1, M)$, assigning the mass V_1 to the random point $Z_1 \sim G_0$. Then break the remaining length of the stick $1 - V_1$ by the proportion $V_2 \sim \text{Beta}(1, M)$, assigning the mass $(1 - V_1)V_2$ to the random point $Z_2 \sim G_0$. At the i^{th} step, break the remaining length of the stick $\prod_{j=1}^{i-1} (1 - V_j)$ by the proportion $V_i \sim \text{Beta}(1, M)$, assigning the mass $\left(\prod_{j=1}^{i-1} (1 - V_j) \right) V_i$ to the random point $Z_i \sim G_0$. This process is repeated infinitely many times, though in practice is truncated to a large but finite number L as one cannot sample an infinite number of parameters.

Alternative representations of the Dirichlet process such as the Chinese Restaurant Process and the Polya Urn can be found in Hjort et al. (2010).

3.3.3 Bernstein polynomial prior

Bernstein polynomials are a special case of B-splines where there are no internal knots (B-splines will be defined in 3.3.4). To define Bernstein polynomials, it is first necessary to discuss Bernstein

basis polynomials. There are $k + 1$ Bernstein basis polynomials of degree k , with the following structure:

$$b_{j,k}(x) = \binom{k}{j} x^j (1-x)^{k-j}, \quad j = 0, 1, \dots, k. \quad (3.55)$$

A Bernstein polynomial is then the following linear combination of Bernstein basis polynomials:

$$B_k(x) = \sum_{j=0}^k \beta_j b_{j,k}(x), \quad (3.56)$$

where β_j for $j = 0, 1, \dots, k$ are called the Bernstein coefficients.

The Bernstein polynomial prior of Petrone (1999a,b) and Choudhuri et al. (2004) is based on the Weierstrass approximation theorem that states that any continuous real-valued function on $[0, 1]$ can be uniformly approximated to any desired accuracy by (Bernstein) polynomials. Let $G(\cdot)$ denote a cumulative distribution function (cdf) with continuous density $g(\cdot)$ on $[0, 1]$, then the following mixture

$$\hat{G}_r(\omega) = \sum_{j=1}^r G\left(\frac{j-1}{r}, \frac{j}{r}\right] I\beta(\omega; j, r-j+1) \quad (3.57)$$

$$= \sum_{j=1}^r w_{j,r} I\beta(\omega; j, r-j+1), \quad \omega \in [0, 1], \quad (3.58)$$

converges uniformly to $G(\omega)$, where $G(u, v] = G(v) - G(u)$ and $I\beta(\omega; a, b)$ and $\beta(\omega; a, b)$ denote the cdf and density of the beta distribution with parameters a and b , respectively. This becomes the Bernstein polynomial prior when a Dirichlet process is placed on $G(\cdot)$.

3.3.4 B-spline prior

The *B-spline prior* was developed as a generalisation to the Bernstein polynomial prior to provide a flexible framework to estimate complicated spectral densities present in real gravitational wave detector noise (Edwards et al., 2017a). Bernstein polynomials have poor coverage on the unit interval, meaning they cannot adequately approximate spectral density functions with sharp and abrupt peaks, called spectral lines. Due to the local support property of B-splines, spectral lines may be accurately estimated, thus the development of the B-spline prior.

3.3.4.1 B-splines and B-spline densities

A spline function of order $r + 1$ is a piecewise polynomial of degree $\leq r$ with so-called *knots* where the piecewise polynomials connect. A spline is continuous at the knots (or continuously differentiable to a certain order depending on the multiplicity of the knots). The number of internal knots must be $\geq r$. Any spline function of order $r + 1$ defined on a certain partition can be uniquely represented as a linear combination of basis splines, *B-splines*, of the same order over the same partition (Cai and Meyer, 2011, Powell, 1981). B-splines can be parametrised either recursively (de Boor, 1993), or by using divided differences and truncated power functions (Cai and Meyer, 2011, Powell, 1981). The former convention is defined below.

Without loss of generality, assume the global domain of interest is the unit interval $[0, 1]$. For a set of k B-splines of degree $\leq r$ for some integer $r \geq 0$, define a nondecreasing knot sequence

$$\boldsymbol{\xi} = \{0 = \xi_0 = \xi_1 = \dots = \xi_r \leq \xi_{r+1} \leq \dots \leq \xi_{k-1} \leq 1 = \xi_k = \xi_{k+1} = \dots = \xi_{k+r}\} \quad (3.59)$$

of $k + r + 1$ knots, comprised of $k - r + 1$ internal knots and $2r$ external knots. The external knots outside or on the boundary of $[0, 1]$ (i.e., $\xi_0 \leq \dots \leq \xi_{r-1} \leq \xi_r = 0$ and $1 = \xi_k \leq \xi_{k+1} \dots \leq \xi_{k+r}$) are required for B-splines to constitute a basis of spline functions on $[0, 1]$. The external knots may be assumed to be exactly on the boundary (and indeed this is an assumption used in the analysis presented in this thesis). The knot sequence $\boldsymbol{\xi}$ yields a partition of the interval $[0, 1]$ into $k - r$ subsets.

For $j = \{1, 2, \dots, k\}$, each individual B-spline of degree r , $B_{j,r}(\cdot; \boldsymbol{\xi})$, depends on $\boldsymbol{\xi}$ only through the $r + 2$ consecutive knots $(\xi_{j-1}, \dots, \xi_{j+r})$. The number of internal knots is equal to the degree of the B-spline $B_{j,r}$ if there are no knot multiplicities. There can be a maximum of $r + 1$ coincident knots for continuity. These knots determine the shape and location of each B-spline.

A B-spline with degree 0 is the following indicator function

$$B_{j,0}(\omega; \boldsymbol{\xi}) = \begin{cases} 1, & \omega \in [\xi_{j-1}, \xi_j), \\ 0, & \text{otherwise.} \end{cases} \quad (3.60)$$

Higher degree B-splines can then be defined recursively using

$$B_{j,r}(\omega; \boldsymbol{\xi}) = \frac{\omega - \xi_{j-1}}{\xi_{j+r-1} - \xi_{j-1}} B_{j,r-1}(\omega; \boldsymbol{\xi}) + \frac{\xi_{j+r} - \omega}{\xi_{j+r} - \xi_j} B_{j+1,r-1}(\omega; \boldsymbol{\xi}), \quad (3.61)$$

where $r > 0$ is the degree.

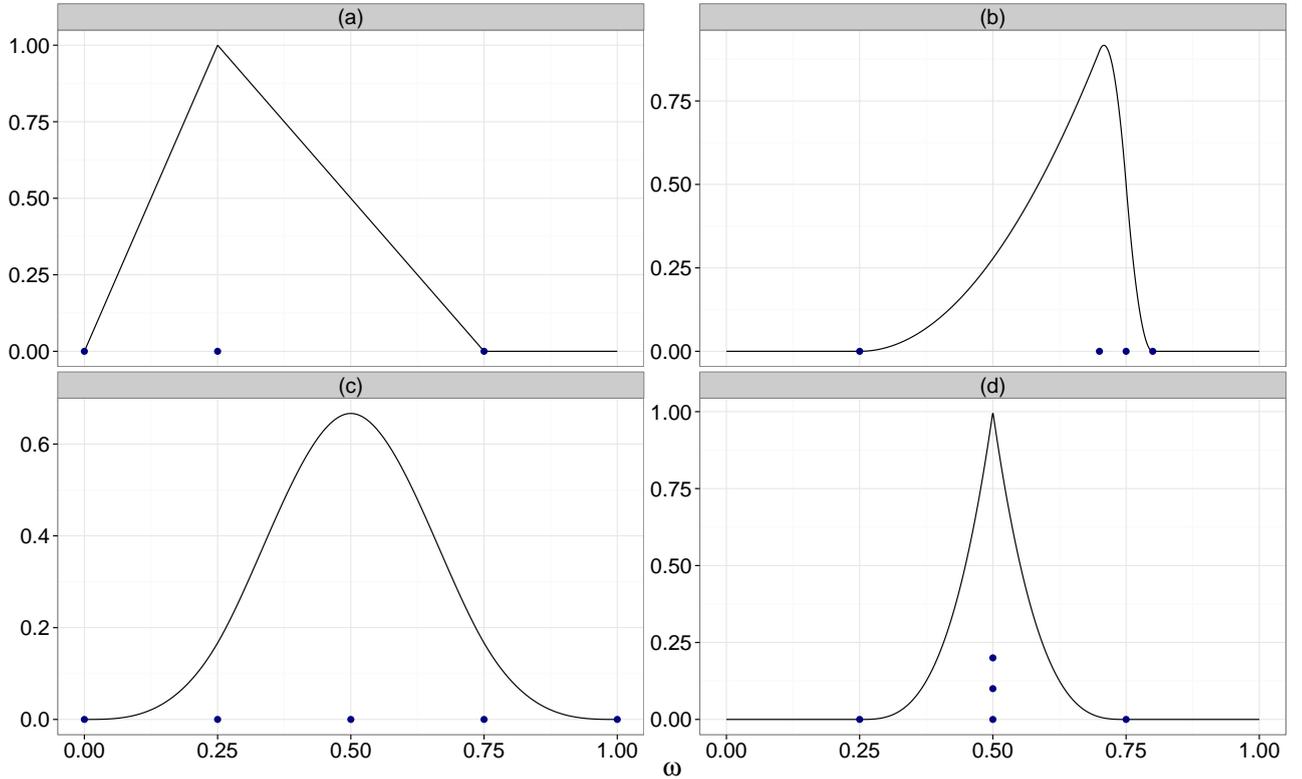


FIGURE 3.1: (a): A linear (triangular) B-spline ($r = 1$) with knots at $\omega = \{0, 0.25, 0.75\}$. (b): A quadratic B-spline ($r = 2$) with knots at $\omega = \{0.25, 0.7, 0.75, 0.8\}$. (c): A cubic B-spline ($r = 3$) with equidistant knots at $\omega = \{0, 0.25, 0.5, 0.75, 1\}$. (d): A cubic B-spline ($r = 3$) with unique knots at $\omega = \{0.25, 0.75\}$ and three coincident knots at $\omega = 0.5$.

Some examples of various B-splines are presented in Figure 3.1, and the following properties are highlighted: local support on the unit interval, where the B-splines are only non-zero between their end knots; the number of knots for a degree r B-spline is $r + 2$; the higher the degree r , the smoother the B-spline (if there are no coincident knots); and coincident knots (and knots that are sufficiently close together) lead to sharp peaks.

B-spline densities are the usual B-spline basis functions, normalised so they each integrate to 1 (Cai and Meyer, 2011). The recursive B-spline parametrisation used in this thesis allows one to easily analytically integrate each B-spline, which is then used as a normalisation constant for the B-spline density, which is defined as

$$b_{j,r}(\omega; \boldsymbol{\xi}) = \frac{B_{j,r}(\omega; \boldsymbol{\xi})}{\int_{\xi_{j-1}}^{\xi_{j+r}} B_{j,r}(\omega; \boldsymbol{\xi}) d\omega}. \quad (3.62)$$

3.3.4.2 B-spline prior

Now that the Bernstein polynomial prior and B-splines have been defined, the B-spline prior can be formalised.

Define $\mathcal{F} = \{F : F \text{ is a cdf on } [0, 1]\}$ and $\mathcal{F}_r = \{F : F \text{ is a mixture of } I\beta(j, r-j+1) \text{ distributions, } j = 1, \dots, r\}$. The loss function $l(\cdot)$ associated with approximating \mathcal{F} by the $r - 1$ dimensional space \mathcal{F}_r is defined as

$$l(\mathcal{F}, \mathcal{F}_r) = \sup_{G \in \mathcal{F}} \inf_{F \in \mathcal{F}_r} \rho(G, F), \quad (3.63)$$

where $\rho(G, F) = \sup_{x \in [0, 1]} |G(x) - F(x)|$. As shown by Perron and Mengersen (2001), this loss can not be made arbitrarily small. Thus the mixture of beta cdfs does not provide adequate coverage of the space of cdfs on $[0, 1]$. However, Perron and Mengersen (2001) showed that if one replaces the beta distributions by B-spline distributions of fixed order (shown for order 2, i.e., triangular distributions) but with variable knots, the loss can be made arbitrarily small by increasing the number of knots. This is the rationale for using a mixture of B-spline distributions with variable knots in the following specification of a *sieve prior* (a prior that is constructed from a sequence of priors living on finite dimensional spaces) (Shen and Wasserman, 2001).

The B-spline prior has the following representation as a mixture of B-spline densities:

$$s_r(\omega; k, \mathbf{w}_k, \boldsymbol{\xi}) = \sum_{j=1}^k w_{j,k} b_{j,r}(\omega; \boldsymbol{\xi}), \quad (3.64)$$

where k is the number of B-splines of fixed degree $\leq r$ in the mixture, $\mathbf{w}_k = (w_{1,k}, \dots, w_{k,k})$ is the weight vector (with constraint $\sum_{j=1}^k w_{j,k} = 1$), and $\boldsymbol{\xi}$ is the knot sequence. Rather than placing a prior on the \mathbf{w}_k 's whose dimension changes with k , one can follow the approach of Choudhuri et al. (2004) and assume that the weights are induced by a cdf G on $[0, 1]$. Similarly, assume that the $k - r$ internal knot differences $\Delta_j = \xi_{j+r} - \xi_{j+r-1} = H\left(\frac{j-1}{k-r}, \frac{j}{k-r}\right)$ for $j = \{1, \dots, k-r\}$ are induced by a cdf H on $[0, 1]$. Or equivalently, $\xi_{j+r} = H\left(\frac{j}{k-r}\right)$ for $j = \{1, \dots, k-r\}$. Note that if B-splines are defined on the unit interval with boundary knots on 0 and 1, then $\sum_j \Delta_j = 1$.

The B-spline prior is parametrised in terms of k , G , and H :

$$s_r(\omega; k, G, H) = \sum_{j=1}^k G\left(\frac{j-1}{k}, \frac{j}{k}\right) b_{j,r}(\omega; H). \quad (3.65)$$

Independent Dirichlet process priors are then placed on G and H and a discrete prior on the number of mixture components k .

The B-spline prior is similar in nature to the Bernstein polynomial prior introduced by Petrone (1999a,b) and applied to spectral density estimation by Choudhuri et al. (2004). The primary difference is that the B-spline prior is a mixture of B-spline densities with *local support* rather than beta densities with *full support* on the unit interval as shown in Figure 3.2. Moreover, placement of knots is fully data-driven.

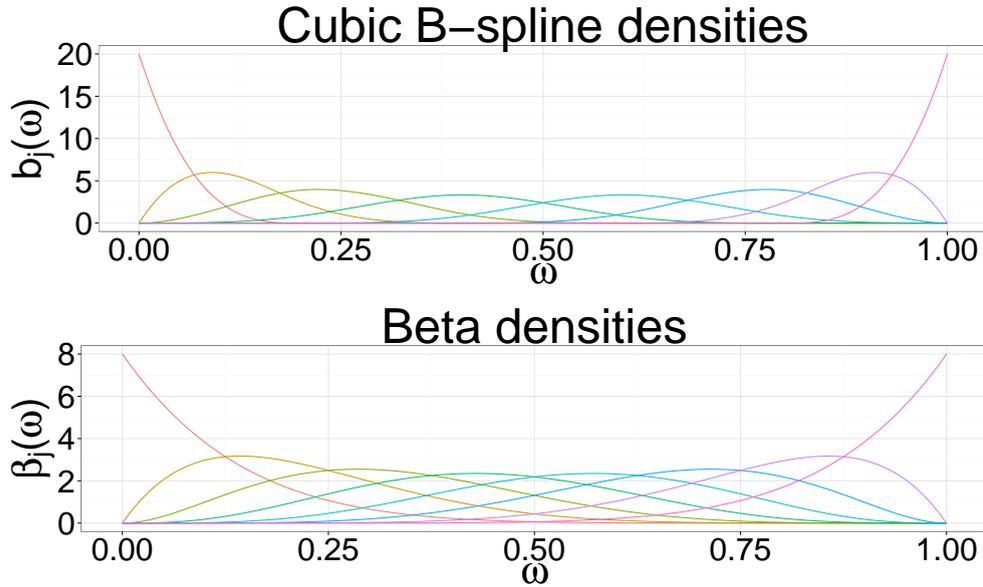


FIGURE 3.2: Top panel: Eight cubic B-spline densities with equidistant knots at $\omega = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Notice the local support. Bottom panel: Eight beta densities with full support on the entire unit interval.

When there are no internal knots, the B-spline basis becomes a Bernstein polynomial basis. Bernstein polynomials are thus a special case of B-splines, and the B-spline prior could be regarded as a generalisation of the Bernstein polynomial prior.

Figure 3.3 demonstrates that it is possible to construct curves (B-spline mixtures) with sharp peaks if knots are sufficiently close together. The top panel shows a set of B-spline density functions and the bottom panel shows a mixture of these with random weights. The local support property of B-splines is the reason the B-spline prior will be instrumental in estimating a spectral density with sharp peaks.

3.4 Fourier analysis of time series

Gravitational wave strain data are a finite duration time series. It is thus important to understand the theory and practice of time series analysis and digital signal processing. Much of gravitational wave data analysis is conducted in the frequency domain, making the Fourier transform a prerequisite. To ensure gravitational wave data are fit-for-purpose prior to conducting parameter estimation, the time series may need to be manipulated (using tools such as windowing (3.4.5), downsampling (3.4.6), and filtering (3.4.7)), and it is necessary to understand how these processes work. This section provides an overview of the important tools needed to analyse a gravitational wave time series.

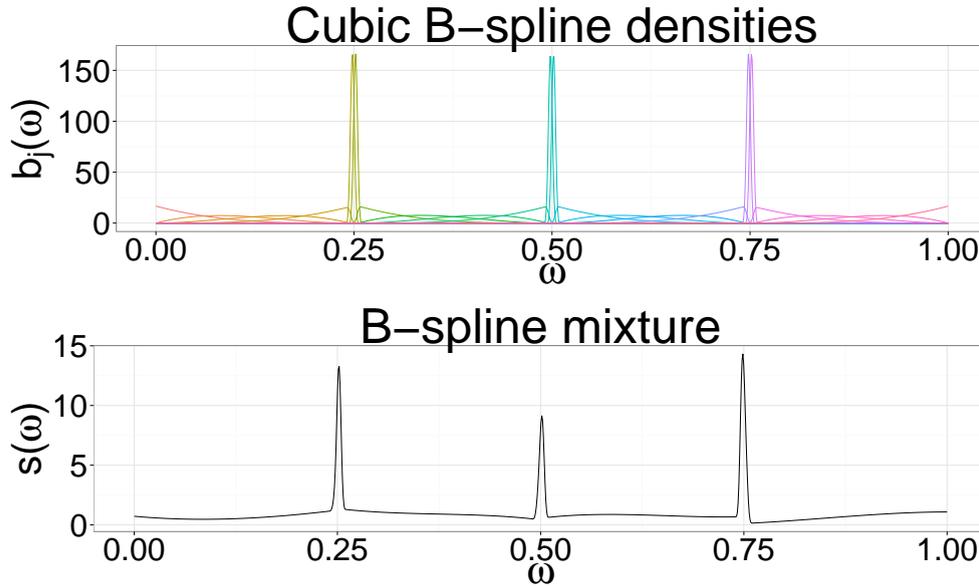


FIGURE 3.3: Top panel: Cubic B-spline densities with many knots close to each of the locations $\omega = \{0.25, 0.5, 0.75\}$. Bottom panel: A random mixture of these B-spline densities. It is possible to construct a B-spline mixture with abrupt, sharp peaks.

3.4.1 Stationary time series

A weakly (or second order) stationary time series $\{Y_t\}$ is a stochastic process that has constant and finite mean and variance over time (i.e., $\mathbb{E}[Y_t] = \mu < \infty$ and $\text{Var}[Y_t] = \sigma^2 < \infty$, for all t), and an autocovariance function $\gamma(\cdot)$ that depends only on the time lag h . That is, for a zero-mean weakly stationary process, the autocovariance function has the form

$$\gamma(h) = \mathbb{E}[Y_t Y_{t+h}], \quad \forall t, \quad (3.66)$$

where $\mathbb{E}[\cdot]$ is the expected value operator, and t represents time.

The autoregressive moving average (ARMA) model is a common approach to modelling stationary time series. This is made up of two components: an autoregressive component and a moving average component. Assuming the time series $\{Y_t\}$ is mean-centred (i.e., the mean has been subtracted from the time series to yield zero mean time series), an ARMA(p, q) model has the following formulation:

$$Y_t - \sum_{i=1}^p \varphi_i Y_{t-i} = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}, \quad (3.67)$$

where p is the number of autoregressive terms, q is the number of moving average terms, and ϵ_t is white noise with zero mean and constant variance (often assumed to be, but not necessarily, independent normally distributed random variables). The summation on the left-hand side of the equation is the autoregressive part, and the summation on the right-hand side is the moving

average part. If $q = 0$, this reduces to an $\text{AR}(p)$ model (autoregressive model of order p). If $p = 0$, the ARMA model reduces to a moving average model of order q , or $\text{MA}(q)$. If $p = q = 0$, then the model reduces to white noise.

A *nonstationary* time series is one that does not meet the conditions of (weak) stationarity. For example, a time series with a trend does not have a constant mean over time and is therefore not stationary. Stationarity of a time series can be tested using classical hypothesis tests such as the Augmented Dickey-Fuller test (Said and Dickey, 1984), Phillips-Perron unit root test (Phillips and Perron, 1988), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992).

3.4.2 Fourier transform

The Fourier transform is a transformation that maps a function of time to a function of frequency by using complex sinusoids. In the context of time series analysis (also called signal processing in the engineering literature), the Fourier transform is immensely useful, as many problems are naturally solved in the *frequency domain* as opposed to the *time domain*.

Formally, let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a real-valued function of time t , then its Fourier transform $\tilde{h} : \mathbb{R} \rightarrow \mathbb{C}$ is a complex-valued function defined as

$$\tilde{h}(\lambda) = \int_{-\infty}^{\infty} h(t) \exp(-i\lambda t) dt, \quad (3.68)$$

where $\lambda = 2\pi\nu$ is angular frequency in radians per unit time, and ν is frequency measured in cycles per unit time. If the unit for time is seconds, the unit for frequency is Hertz, denoted Hz.

The operation that reverses this transformation from the frequency domain to the time domain is called the *inverse Fourier transform*, and is defined as

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{h}(\lambda) \exp(i\lambda t) d\lambda. \quad (3.69)$$

The Fourier transform and its inverse can be thought of as a *Fourier transform pair*, denoted

$$h(t) \rightleftharpoons \tilde{h}(\lambda). \quad (3.70)$$

3.4.3 Properties of the Fourier transform

Since h is real-valued, then \tilde{h} is *Hermitian symmetric*. That is,

$$\tilde{h}(-\lambda) = \overline{\tilde{h}(\lambda)}, \quad (3.71)$$

where $\overline{\tilde{h}(\lambda)}$ is the complex conjugate of $\tilde{h}(\lambda)$. This means attention may be restricted to positive frequencies, as no new information is contained by the redundant negative frequencies.

The Fourier transform is *linear*. That is, let $h(t) \rightleftharpoons \tilde{h}(\lambda)$ and $g(t) \rightleftharpoons \tilde{g}(\lambda)$ be Fourier transform pairs, then

$$ah(t) + bg(t) \rightleftharpoons a\tilde{h}(\lambda) + b\tilde{g}(\lambda), \quad (3.72)$$

for any real numbers a and b .

Through the *convolution theorem*, a convolution in the time domain forms a Fourier transform pair with multiplication in the frequency domain. That is,

$$h(t) * g(t) \rightleftharpoons \tilde{h}(\lambda)\tilde{g}(\lambda), \quad (3.73)$$

where

$$h(t) * g(t) = \int_{-\infty}^{\infty} h(t - \tau)g(\tau)d\tau \quad (3.74)$$

is the *convolution* of h and g . The same is true in reverse. That is, multiplication in the time domain forms a Fourier transform pair with a convolution in the frequency domain.

The Fourier transform has the following *time-shifting* property:

$$h(t - T) \rightleftharpoons \exp(-i\lambda T)\tilde{h}(\lambda), \quad (3.75)$$

for any time lag T . That is, shifting a time series h by lag T can be done in the frequency domain by multiplying the original Fourier transform \tilde{h} by $\exp(-i\lambda T)$.

3.4.4 Discrete Fourier transform

Analogous to the Fourier transform, the *discrete Fourier transform* (DFT) converts functions of a finite sequence of discrete time points into a finite set of Fourier coefficients in the frequency domain. More formally, let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a real-valued time series of finite (even) length n , with sampling rate $\frac{1}{\Delta_t}$, sampling interval Δ_t , and discrete time points $t = 0, \Delta_t, 2\Delta_t, \dots, (n - 1)\Delta_t$.

The DFT $\tilde{h} : \mathbb{R} \rightarrow \mathbb{C}$ is then defined as

$$\tilde{h}(\lambda) = \sum_{j=0}^{n-1} h(j\Delta_t) \exp(-ij\lambda\Delta_t), \quad (3.76)$$

where $\lambda = 0, \Delta_\lambda, 2\Delta_\lambda, \dots, (n-1)\Delta_\lambda$ are discrete angular frequencies, with interval $\Delta_\lambda = \frac{2\pi}{n\Delta_t}$.

The integral in Equation 3.68 is replaced by a sum in Equation 3.76. The DFT can therefore be seen as a discrete approximation to the Fourier transform of a continuous function.

The *inverse DFT* is the transformation back to the time domain, and is given by

$$h(t) = \frac{1}{n} \sum_{j=0}^{n-1} \tilde{h}(j\Delta_\lambda) \exp(ijt\Delta_\lambda). \quad (3.77)$$

The DFT and inverse DFT are efficiently computed using the *fast Fourier transform* (FFT) algorithm (Cooley and Tukey, 1965). Like the Fourier transform, the DFT also has Hermitian symmetry if h is real-valued. If n is even, then there are $\frac{n}{2} + 1$ non-redundant DFT elements. Note also that $\tilde{h}(0)$ and $\tilde{h}(\frac{n}{2}\Delta_\lambda)$ are always real-valued (zero imaginary components). This means there are $\frac{n}{2} + 1$ real parts and $\frac{n}{2} - 1$ imaginary parts, which can be turned into a vector of length n for easy manipulation.

3.4.5 Sampling a continuous signal

Many issues may arise when sampling (digitising) a continuous (analog) time series. The problem of *aliasing* occurs if there is an insufficient sampling rate. If the sampling rate is too slow, information about frequency and amplitude may be missed, leading to incorrect conclusions. Power contained in frequencies above a critical frequency (called the *Nyquist frequency*) will be *folded* onto lower frequency components. The Nyquist frequency is defined to be half of the signal sampling rate. Frequencies above the Nyquist frequency will be indistinguishable from the lower frequency components, hence the name aliasing.

Aliasing can be avoided by sampling a continuous signal at a rate that is at least two times the maximum required frequency. One can then restrict attention to the bandwidth of frequencies below the Nyquist frequency. Another common approach to avoid aliasing is to apply a low-pass (anti-aliasing) filter to the signal to remove power from the frequency components above the Nyquist frequency (see 3.4.7).

Spectral leakage arises due to finite duration discrete sampling of an infinite duration signal. This is a problem as frequencies with smaller power can be lost when the power from frequencies with larger power leak across neighbouring frequencies. As signals are generally time-limited, no information about the signal is known before or after its measurement period. The FFT assumes a signal is periodic and repeats itself after it has been measured. However, if a signal is not perfectly periodic with the number of samples in the signal, n , there will be a discontinuity at the end-points (i.e., the end-points do not match) and additional frequency components will be introduced, leading to additional spectral leakage.

Spectral leakage cannot be completely annihilated, but may be mitigated by either taking more samples, or by choosing an appropriate *window function*. *Windowing* can be used to mitigate the detrimental effects of spectral leakage. A discrete time series of length n can be multiplied by a window function $w(\cdot)$ in the time domain prior to applying the discrete Fourier transform.

The default window (when no window is used) is called the *rectangular window*. It is simply given as:

$$w(t) = 1, \quad t = 0, 1, \dots, n - 1. \quad (3.78)$$

By default, a finite duration signal has a rectangular window, and a window (multiplication) in the time domain implies a convolution in the frequency domain. The Fourier transform of a rectangular window is an aliased sinc function, and as the sampling rate increases to infinity, this approaches the sinc function, which is defined as

$$\text{sinc}(x) = \frac{\sin(x)}{x}, \quad x \neq 0. \quad (3.79)$$

This convolution leads to a smearing of the spectrum across frequencies far away from a spectral peak, and thus spectral leakage.

A rectangular window has the lowest total leakage, but this is spread across a large amount of frequencies, which may blur out details about the spectrum at frequencies further away from the original frequency. Non-rectangular windows have higher amounts of total leakage, but are less problematic as they distribute this to frequencies closer to the original frequency. Two important non-rectangular windows in the context of gravitational wave data analysis are the Hann window and Tukey window.

The *Hann window* is often applied to gravitational wave data when estimating the noise PSD (see 3.5). This is defined as:

$$w(t) = \frac{1}{2} \left(1 - \cos \left(\frac{2\pi t}{n-1} \right) \right), \quad t = 0, 1, \dots, n - 1. \quad (3.80)$$

Applying a Hann window results in low aliasing effects, but decreases the resolution of the time series.

The *Tukey window* (also called the tapered cosine window) is applied to gravitational wave data prior to Fourier transforming. It has a parameter α that determines the fraction of the window that follows a tapered cosine, and the fraction that is rectangular. The Tukey window is defined as:

$$w(t) = \begin{cases} \frac{1}{2} \left(1 - \cos \left(\frac{t\pi}{\frac{\alpha}{2}n} \right) \right), & 0 \leq t \leq \frac{\alpha}{2}n, \\ 1, & \frac{\alpha}{2}n \leq t \leq (1 - \frac{\alpha}{2})n, \\ \frac{1}{2} \left(1 - \cos \left(\frac{(n-t)\pi}{\frac{\alpha}{2}n} \right) \right), & (1 - \frac{\alpha}{2})n \leq t \leq n - 1. \end{cases} \quad (3.81)$$

The cosine parts occur at the first and last $\frac{\alpha}{2}$ fraction of the time series. The rectangular and Hann windows are special cases of the Tukey window where $\alpha = 0$ and 0.5 respectively.

A useful discussion on windowing can be found in Harris (1978).

3.4.6 Downsampling

Downsampling or *decimation* is the process that reduces the sampling rate of a time series. This is often done to reduce the volume of data processed. If done correctly, downsampled data will have no information loss (but may introduce aliasing).

It is common to downsample by integer or rational factors. Downsampling by an integer factor R involves two steps: applying an appropriate low-pass (anti-aliasing) filter to eliminate the high frequency components of the time series, which is necessary to avoid aliasing at lower sampling rates; and downsampling the filtered time series by integer factor R to achieve the desired sampling rate (keeping every R^{th} sample). Downsampling by a rational factor $\frac{R}{S}$ (where $R, S \in \mathbb{Z}$ and $R > S$), involves upsampling by an integer factor S via interpolation, then downsampling by an integer factor R .

A discussion of appropriate anti-aliasing filters can be found in Oppenheim et al. (1999).

3.4.7 Linear filters

A *linear filter* transforms an input time series into an output time series to meet a particular objective, or to show special features of the data. For example, it may be of interest to remove particular frequency components in a given time series. A *low-pass filter* allows low frequencies through but cuts out the higher frequencies. A *high pass filter* does the opposite. A *band-pass*

filter has a chosen band of frequencies allowed to pass through, and a *band-stop filter* only passes the frequencies outside a specified band.

An important low-pass filter used often in gravitational wave data analysis is the *Butterworth filter* (Butterworth, 1930). It is an important anti-aliasing filter, used prior to downsampling data to prevent aliasing at higher frequencies. The useful property of the Butterworth filter is its flat frequency response in the pass-band (i.e., it accurately reproduces the input signal) .

The *matched filter* is also hugely important in gravitational wave astronomy, and played a key role in the recent discoveries of gravitational waves (Abbott et al., 2016d,f, 2017). Matched filtering requires a library of known signals, called *templates* and works by finding the complex conjugate of a time-reversed template, and convolving this with an unknown signal (Turin, 1960). It is also the linear filter that maximises the signal-to-noise ratio (SNR). The SNR measures the ratio of a signal's power to the power of background noise, and is defined as:

$$\varrho = \sqrt{2 \int_0^\infty \frac{|\tilde{s}(\lambda)|^2}{|\tilde{\epsilon}(\lambda)|^2} d\lambda}, \quad (3.82)$$

where $\tilde{s}(\cdot)$ is the Fourier transformed signal, $\tilde{\epsilon}(\cdot)$ is the Fourier transformed noise, and λ is the angular frequency.

Matched filtering is a natural method for black hole binary signals as the waveforms for these sources are well-modelled and relatively easy to produce, meaning it is possible to generate a large list of templates using a grid of values for all of the important astrophysical parameters.

3.5 Spectral density estimation

One fundamental notion in time series analysis is the *spectral density*, which characterises a signal in the frequency domain. Characterising noise from ground-based gravitational wave observatories, such as Advanced LIGO, requires computing an estimate of its underlying spectral density. The standard (Gaussian and stationary) approach used in much of the literature may be considered somewhat rudimentary, particularly because the assumptions of the parametric model are not realistic. In this section, basic smoothing methods will be described. This will be followed by more complicated Bayesian nonparametric techniques for inferring the spectral density of a time series, forming a central part of this thesis.

3.5.1 Introduction

Useful information about a stationary time series is encoded in its spectral density, sometimes called the power spectral density (PSD). This quantity describes the strength (or power) each individual frequency component contributes to the overall time series, and forms a Fourier transform pair with the autocovariance function. More formally, assuming an absolutely summable autocovariance function ($\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$), the spectral density function $f(\cdot)$ of a zero-mean weakly stationary time series is defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-ih\lambda), \quad \lambda \in (-\pi, \pi], \quad (3.83)$$

where λ is the angular frequency and $\gamma(h) = \text{Cov}[Y_t, Y_{t+h}] = \mathbb{E}[Y_t Y_{t+h}]$.

To intuit, a time series that rapidly oscillates is said to have strong high frequency components, and a time series with a slowly varying cycle has strong low frequency components. A time series that has a flat spectral density is said to be a *white noise process*. Analogous to white light, all of the frequencies are present at the same strength.

Spectral density estimation methods can be broadly classified into two groups: parametric and nonparametric. Parametric approaches to spectral density estimation are primarily based on fitting an autoregressive moving average (ARMA) model to a time series (Brockwell and Davis, 1991), though these methods tend to give misleading inferences when the parametric model is poorly specified.

The periodogram $I_n(\cdot)$ is a common nonparametric estimate of the spectral density of a stationary time series that randomly fluctuates around the true PSD. It is a popular object of interest in time series analysis, as it is easily and efficiently computed as the (normalised) squared modulus of Fourier coefficients using the FFT. That is,

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=0}^{n-1} Y_t \exp(-it\lambda) \right|^2, \quad \lambda \in (-\pi, \pi], \quad (3.84)$$

where λ is angular frequency, and Y_t is a stationary time series with discrete time points, $t = 0, 1, \dots, n-1$.

Though the periodogram is an asymptotically unbiased estimator of the spectral density, it is not a consistent estimator (Brockwell and Davis, 1991). Smoothing techniques such as Bartlett's method (Bartlett, 1950), Welch's method (Welch, 1967), and the multitaper method (Thomson, 1982) aim to reduce the variance of the periodogram by dividing a time series into (potentially

overlapping) segments, calculating the periodogram for each segment, and averaging over all of these. Unfortunately, these techniques are sensitive to the choice of smoothing parameter (i.e., the number of segments), resulting in a variance/bias trade-off. Reducing the length of each segment also leads to poorer frequency resolution.

Another common nonparametric approach to spectral estimation involves the use of splines. Smoothing spline techniques are not new to spectral estimation (see e.g., Cogburn and Davis (1974) for an early reference). Wahba (1980) used splines to smooth the log-periodogram, with an automatic data-driven smoothing parameter, avoiding the difficult problem of having to choose this quantity. Kooperberg et al. (1995) used maximum likelihood and polynomial splines to approximate the log-spectral density function.

Bayesian nonparametric approaches to spectrum estimation have gained momentum in recent times. In the context of splines, Gangopadhyay et al. (1999) used a fixed low-order piecewise polynomial to estimate the log-spectral density of a stationary time series. They implemented a RJMCMC algorithm (Green, 1995), placing priors on the number of knots and their locations, with the goal of estimating spectral densities with sharp features. Choudhuri et al. (2004) placed a Bernstein polynomial prior (Petrone, 1999a,b) on the spectral density. The Bernstein polynomial prior is essentially a finite mixture of beta densities with weights induced by a Dirichlet process. The number of mixture components is a smoothing parameter, chosen to have a discrete prior. Zheng et al. (2010) generalised this and constructed a multi-dimensional Bernstein polynomial prior to estimate the spectral density function of a random field. Also extending the work of Choudhuri et al. (2004), Macaro (2010) used informative priors to extract unobserved spectral components in a time series, and Macaro and Prado (2014) generalised this to multiple time series.

Other interesting Bayesian nonparametric approaches include Carter and Kohn (1997) inducing a prior on the log-spectral density using an integrated Wiener process, and Tonellato (2007) placing a Gaussian random field prior on the log-spectral density. Liseo et al. (2001), Rousseau et al. (2012), and Chopin et al. (2013) used Bayesian nonparametric methods to estimate spectral densities from long memory time series, and Rosen et al. (2012) focused on time-varying spectra in nonstationary time series.

The majority of the Bayesian nonparametric methods (for short memory time series) mentioned here make use of Whittle's approximation to the Gaussian likelihood, often called the *Whittle likelihood* (Whittle, 1957). The Whittle likelihood $L_n(\cdot)$ for a mean-centred weakly stationary

time series Y_t of length n with spectral density $f(\cdot)$ has the following formulation:

$$L_n(\mathbf{y}|f) \propto \exp \left(- \sum_{l=1}^{\lfloor \frac{n-1}{2} \rfloor} \left(\log f(\lambda_l) + \frac{I_n(\lambda_l)}{f(\lambda_l)} \right) \right), \quad (3.85)$$

where $\lambda_l = 2\pi l/n$ are the positive Fourier frequencies, $\lfloor (n-1)/2 \rfloor$ is the greatest integer value less than or equal to $(n-1)/2$, and $I_n(\cdot)$ is the periodogram defined in Equation (3.84). If the PSD is assumed known, the $\log f$ term in Equation (3.85) is a constant and can be ignored.

Compare Whittle's pseudo-likelihood to the true likelihood for a Gaussian time series,

$$L_n(\mathbf{y}|f) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} \mathbf{y}' \Sigma^{-1} \mathbf{y} \right), \quad (3.86)$$

where Σ is the $n \times n$ autocovariance matrix with entries $\gamma(i-j)$ given by

$$\gamma(h) = \int_{-\pi}^{\pi} f(\lambda) \exp(ih\lambda) d\lambda. \quad (3.87)$$

The Whittle likelihood has an advantage over the true Gaussian likelihood as it has a direct dependence on the PSD rather than the autocovariance function. The cost of inverting the $n \times n$ autocovariance matrix is $O(n^3)$ whereas the Whittle approximation can be evaluated in $O(n \log n)$ operations. The Whittle likelihood is only exact for Gaussian white noise but works well under certain conditions, even when the data are not Gaussian (Shao and Wu, 2007). That is, the Fourier coefficients are asymptotically independent and Gaussian for a large class of non-Gaussian and non-linear time series. However, Contreras-Cristán et al. (2006) show that there is a loss of efficiency when using the Whittle likelihood compared to the true (Gaussian or non-Gaussian) likelihood.

The Whittle likelihood is ubiquitous in gravitational wave data analysis and a literature review of spectral density estimation in the context of Advanced LIGO data can be seen in Chapter 5.

3.5.2 Periodogram smoothing

Periodogram smoothing is a simple approach to spectrum estimation, with the aim of reducing the variance of the periodogram to uncover the truth. The most common approaches are Bartlett's method (Bartlett, 1950) and Welch's method (Welch, 1967).

Bartlett's method works by splitting a time series of length n into k non-overlapping equal length segments, then computing the periodogram for each segment, and taking the average of these. The

original method averages via the mean, though the median can provide more robust estimates. Welch’s method extends on Bartlett’s method by allowing segments to overlap, and by multiplying each segment by an appropriate window function.

One obvious cost of these approaches is their poorer frequency resolution. However, this is not a problem in the context of gravitational wave data analysis, as the Welch method is applied by considering noise segments close to, but not on, the true gravitational wave signal.

3.5.3 Time-varying spectra

When a time series is nonstationary, it is common to look at time-varying spectra, or time-frequency maps, that show how the power distributed across each frequency bin changes over time. The most common time-frequency map is the *spectrogram*, which makes use of the *short-time Fourier transform*. A time series is broken into (usually overlapping) windowed segments, and the power spectrum at each frequency in each time segment is measured using the squared modulus of the short-time Fourier transform for that segment.

It is important to note that time-frequency maps are subject to the uncertainty principle — there is a trade-off between time and frequency resolutions. Intuitively, there will be poorer frequency resolution for shorter time segments.

Excess power searches of time-frequency maps are used by the gravitational wave community in the search for short duration unmodelled gravitational wave signals, or *bursts*. These methods include X-Pipeline by Sutton et al. (2010) and coherent WaveBurst (cWB) by Klimenko et al. (2008). Time-frequency maps have also made their way to the analysis of long-duration transients (Thrane et al., 2011). Wavelets are generally used as a more sophisticated alternative to the spectrogram, and are an integral part of the BayesWave algorithm (Cornish and Littenberg, 2015), which also characterises burst-type gravitational wave signals.

An alternative approach for constructing time-frequency maps is the AdaptSpec algorithm of Rosen et al. (2012). This estimates the number of segments and segment lengths using reversible jump MCMC. Within each segment, it uses smoothing splines to estimate the spectral density. Some work presented in Chapter 5 will demonstrate how one can break a nonstationary time series into locally stationary segments, and use the nonparametric Bernstein polynomial prior of Petrone (1999a,b) and Choudhuri et al. (2004) to estimate the spectral density of each segment.

3.5.4 Bayesian nonparametric approaches to spectral density estimation

3.5.4.1 Bernstein polynomial prior

The Bernstein polynomial prior, defined in 3.3.3, has been used to estimate the spectral density of stationary time series by Choudhuri et al. (2004), Edwards et al. (2015), Macaro (2010), Macaro and Prado (2014), Zheng et al. (2010).

As the spectral density is not defined on the unit interval (which is the support of the beta probability density function), one must first reparameterise $f(\lambda)$ such that

$$f(\pi\omega) = \tau \times q(\omega), \quad \omega \in [0, 1], \quad (3.88)$$

where $\tau = \int_0^1 f(\pi\omega)d\omega$ is the normalising constant. Note that if working with angular frequency λ , then the scaled frequency is given by $\omega = \frac{\lambda}{\pi}$. If working with frequency ν (in Hertz), then the scaled frequency is $\omega = \frac{\nu}{\nu_0}$ where ν_0 is the Nyquist frequency. Here, Equation 3.88 becomes

$$f(\nu_0\omega) = \tau \times q(\omega), \quad \omega \in [0, 1]. \quad (3.89)$$

To specify a prior on spectral density $f(\cdot)$, a Bernstein polynomial prior is placed on $q(\omega)$, using the following hierarchical scheme of Choudhuri et al. (2004):

- $q(\omega) = \sum_{j=1}^k G\left(\frac{j-1}{k}, \frac{j}{k}\right] \beta(\omega|j, k-j+1)$, where G is a cdf, and $\beta(\omega|a, b)$ is a beta probability density with parameters a and b .
- G is a Dirichlet process distributed random probability measure with base measure G_0 and precision parameter M .
- k has a discrete probability mass function such that $p(k) \propto \exp(-\theta_k k^2)$, $k = 1, 2, \dots$
- τ has an Inverse-Gamma(α_τ, β_τ) distribution.
- G , k , and τ are *a priori* independent.

The stick-breaking representation of Sethuraman (1994) is used to construct the Dirichlet process. This is an infinite-dimensional model that is truncated to a large but finite number of mixture components L for computational purposes. The choice of a large L will provide a more accurate approximation but at the expense of increasing computation time. G is parametrised to

$(Z_0, Z_1, \dots, Z_L, V_1, \dots, V_L)$ such that

$$G(\omega) = \left(\sum_{l=1}^L p_l \delta_{Z_l}(\omega) \right) + \left(1 - \sum_{l=1}^L p_l \right) \delta_{Z_0}(\omega), \quad (3.90)$$

where $p_1 = V_1$, $p_l = \left(\prod_{j=1}^{l-1} (1 - V_j) \right) V_l$ for $l \geq 2$, $V_l \sim \text{Beta}(1, M)$ for $l = 1, \dots, L$, $Z_l \sim G_0$ for $l = 0, 1, \dots, L$, and $\delta_a(x)$ is an indicator function that is one if $x = a$ and zero otherwise. This yields the finite mixture prior of the spectral density

$$f(\pi\omega) = \tau \sum_{j=1}^k w_{j,k} \beta(\omega|j, k - j + 1), \quad (3.91)$$

with weights $w_{j,k} = \sum_{l=0}^L p_l I\{\frac{j-1}{k} < Z_l \leq \frac{j}{k}\}$ and $p_0 = 1 - \sum_{l=1}^L p_l$.

Abbreviating the vector of parameters to $\boldsymbol{\theta} = (\mathbf{v}, \mathbf{z}, k, \tau)$, the joint prior is

$$p(\boldsymbol{\theta}) \propto \left(\prod_{l=1}^L M(1 - v_l)^{M-1} \right) \left(\prod_{l=0}^L g_0(z_l) \right) p(k)p(\tau), \quad (3.92)$$

and is updated using the Whittle likelihood to produce the unnormalised joint pseudo-posterior.

3.5.4.2 B-spline prior

The Bernstein polynomial prior was found to be ill-suited for modelling the PSD of real Advanced LIGO detector noise (Edwards et al., 2015), leading to the development of the B-spline prior.

The B-spline prior was first defined in 3.3.4. Similar to the Bernstein polynomial prior, to place a prior on the spectral density $f(\cdot)$ of a stationary time series defined on the interval $[0, \pi]$, one must first transform to the interval $[0, 1]$:

$$f(\pi\omega) = \tau \times s_r(\omega; k, G, H), \quad \omega \in [0, 1], \quad (3.93)$$

where $\tau = \int_0^1 f(\pi\omega) d\omega$ is the normalising constant, and $s_r(\cdot)$ is the B-spline prior defined in Equation (3.65). Note again that if working with angular frequency λ , then scaled frequency is given by $\omega = \frac{\lambda}{\pi}$, and if working with frequency ν (in Hertz), then scaled frequency is given by $\omega = \frac{\nu}{\nu_0}$ where ν_0 is the Nyquist frequency, in which case Equation 3.93 becomes

$$f(\nu_0\omega) = \tau \times s_r(\omega; k, G, H). \quad (3.94)$$

The prior for $f(\cdot)$ then has the following hierarchical structure:

- G determines the weights (i.e., scale) for each of the k B-spline densities. Let $G \sim \text{DP}(M_G, G_0)$, where $M_G > 0$ is the precision parameter and G_0 is the base probability distribution function with density g_0 .
- H determines the location of knots and hence the shape and location of the B-spline densities. Let $H \sim \text{DP}(M_H, H_0)$, where $M_H > 0$ is the precision parameter and H_0 is the base probability distribution function with density h_0 .
- k is the number of B-splines in the mixture and has discrete probability mass function $p(k) \propto \exp(-\theta_k k^2)$ for $k = 1, 2, \dots, k_{\max}$. Here k_{\max} is the largest possible value k is allowed to take. That is, limit the maximum value of k for computational reasons and do pilot runs to ensure a larger k_{\max} is not required.
- τ is the normalising constant. Let $\tau \sim \text{IG}(\alpha_\tau, \beta_\tau)$.

Assume all of these parameters are *a priori* independent.

As Dirichlet process priors have been placed on G and H , an algorithm to sample from these distributions is required. To sample from a Dirichlet process, Sethuraman's stick-breaking construction (Sethuraman, 1994) is utilised. For computational purposes, the number of mixture distributions for Dirichlet process representations of G and H is truncated to large but finite positive integers (L_G and L_H respectively).

To set up the stick-breaking process, reparameterise G to $(Z_0, Z_1, \dots, Z_{L_G}, V_1, \dots, V_{L_G})$ such that

$$G(\omega) = \left(\sum_{l=1}^{L_G} p_l \delta_{Z_l}(\omega) \right) + \left(1 - \sum_{l=1}^{L_G} p_l \right) \delta_{Z_0}(\omega), \quad (3.95)$$

$$p_1 = V_1, \quad (3.96)$$

$$p_l = \left(\prod_{j=1}^{l-1} (1 - V_j) \right) V_l, \quad l \geq 2, \quad (3.97)$$

$$p_0 = 1 - \sum_{l=1}^{L_G} p_l, \quad (3.98)$$

$$V_l \sim \text{Beta}(1, M_G), \quad l = 1, \dots, L_G, \quad (3.99)$$

$$Z_l \sim G_0, \quad l = 0, 1, \dots, L_G, \quad (3.100)$$

and H to $(X_0, X_1, \dots, X_{L_H}, U_1, \dots, U_{L_H})$ such that

$$H(\omega) = \left(\sum_{l=1}^{L_H} q_l \delta_{X_l}(\omega) \right) + \left(1 - \sum_{l=1}^{L_H} q_l \right) \delta_{X_0}(\omega), \quad (3.101)$$

$$q_1 = U_1, \quad (3.102)$$

$$q_l = \left(\prod_{j=1}^{l-1} (1 - U_j) \right) U_l, \quad l \geq 2, \quad (3.103)$$

$$q_0 = 1 - \sum_{l=1}^{L_H} q_l, \quad (3.104)$$

$$U_l \sim \text{Beta}(1, M_H), \quad l = 1, \dots, L_H, \quad (3.105)$$

$$X_l \sim H_0, \quad l = 0, 1, \dots, L_H, \quad (3.106)$$

where $\delta_a(x) = 1$ if $x = a$ and is 0 otherwise.

Conditional on k , the above hierarchical structure provides a finite mixture prior for the spectral density of a stationary time series

$$f(\pi\omega) = \tau \sum_{j=1}^k w_{j,k} b_{j,r}(\omega; \boldsymbol{\xi}), \quad (3.107)$$

with weights

$$w_{j,k} = \sum_{l=0}^{L_G} p_l I \left\{ \frac{j-1}{k} < Z_l \leq \frac{j}{k} \right\}, \quad (3.108)$$

and knot differences

$$\Delta_j = (\xi_{j+r} - \xi_{j+r-1}) \quad (3.109)$$

$$= \sum_{l=0}^{L_H} q_l I \left\{ \frac{j-1}{k-r} < X_l \leq \frac{j}{k-r} \right\}, \quad (3.110)$$

for $j = \{1, \dots, k-r\}$ and $k > r$. The denominator $k-r$ in Equation (3.110) comes from assuming the exterior knots are the same as the boundary knots. Note that the lower internal boundary knot is assumed to be $\xi_r = 0$, meaning the first knot difference is $\Delta_1 = \xi_{r+1} - \xi_r = \xi_{r+1}$. The subsequent knot placements are determined by taking the cumulative sum of the knot differences. Note also that if the upper boundary knot is 1, then $\sum_j \Delta_j = 1$.

Abbreviating the vector of parameters to $\boldsymbol{\theta} = (\mathbf{v}, \mathbf{z}, \mathbf{u}, \mathbf{x}, k, \tau)$, the joint prior is

$$\begin{aligned} p(\boldsymbol{\theta}) &\propto \left(\prod_{l=1}^{L_G} M_G(1 - v_l)^{M_G-1} \right) \left(\prod_{l=0}^{L_G} g_0(z_l) \right) \\ &\times \left(\prod_{l=1}^{L_H} M_H(1 - u_l)^{M_H-1} \right) \left(\prod_{l=0}^{L_H} h_0(x_l) \right) \\ &\times p(k)p(\tau). \end{aligned}$$

To produce the unnormalised joint pseudo-posterior, this joint prior is updated using the Whittle likelihood defined in Equation (3.85).

The R package `bsplinePsd` was developed to conduct this analysis in, and is available on the Comprehensive R Archive Network (CRAN) (Edwards et al., 2017b). Implementation details as they apply to Advanced LIGO data can be seen in Chapter 5.

3.5.4.3 Nonparametric correction of a parametric likelihood

This subsection presents an alternative approach to modelling the spectral density of a time series by generalising the Whittle likelihood. The key notion here is that one can take advantage of the efficiency of parametric models while mitigating potential misspecifications using a nonparametric correction. This is joint work from the paper by Kirch et al. (2017) and for the sake of completeness, a brief review of this paper is given here.

Consider a parametric working model (with mean 0) for time series $\mathbf{Y}_n = (Y_1, \dots, Y_n)'$ (e.g., ARMA), and assume this is a reasonable approximation to the true time series. Denote $f_{\text{param}}(\cdot)$ as the spectral density of this parametric working model. If the model is misspecified, then this spectral density will be incorrect and must be corrected to obtain the correct second-order dependence structure. A *correction matrix* is then defined as

$$\begin{aligned} C_n &= C_n(f, f_{\text{param}}) \\ &= \begin{cases} \text{diag} \left(\frac{f(\lambda_0)}{f_{\text{param}}(\lambda_0)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \dots, \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}, \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}, \frac{f(\lambda_{n/2})}{f_{\text{param}}(\lambda_{n/2})} \right) & n \text{ even,} \\ \text{diag} \left(\frac{f(\lambda_0)}{f_{\text{param}}(\lambda_0)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \frac{f(\lambda_1)}{f_{\text{param}}(\lambda_1)}, \dots, \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)}, \frac{f(\lambda_N)}{f_{\text{param}}(\lambda_N)} \right) & n \text{ odd,} \end{cases} \end{aligned}$$

where $N = \lfloor (n-1)/2 \rfloor$.

Let F_n be the $n \times n$ orthonormal matrix that when multiplied by a time series \mathbf{Y}_n , produces a vector of discrete Fourier coefficients, $F_n \mathbf{Y}_n$. First, start with \mathbf{Y}_n in the time domain and Fourier transform this to the frequency domain to yield $F_n \mathbf{Y}_n$. Then multiply this by square root of the

correction matrix C_n to produce $C_n^{1/2} F_n \mathbf{Y}_n$, and inverse Fourier transform back to the time domain to get $F_n' C_n^{1/2} F_n \mathbf{Y}_n$. One can then produce the following nonparametrically corrected likelihood function under the parametric working model

$$p_{\text{param}}^C(\mathbf{Y}_n|f) \propto \det(C_n)^{-1/2} p_{\text{param}}(F_n' C_n^{-\frac{1}{2}} F_n \mathbf{Y}_n), \quad (3.111)$$

where p_{param} denotes the parametric likelihood.

As an example, one can restrict attention to an AR(p) parametric working model such that $Y_i = \sum_{l=1}^p a_l Y_{i-l} + \epsilon_i$ where $\{\epsilon_i\}$ are iid $N(0, 1)$ random variables with density $\varphi(\cdot)$. Note that one can set $\sigma^2 = 1$ without loss of generality. This yields the following parametric likelihood of the working model for order $p \geq 0$ and coefficients $\mathbf{a} = (a_1, \dots, a_p)$:

$$p_{\text{param}}(\mathbf{Y}_n|\mathbf{a}) \propto p_{\text{param}}(\mathbf{Y}_p|\mathbf{a}) \prod_{i=p+1}^n \varphi\left(Y_i - \sum_{l=1}^p a_l Y_{i-l}\right), \quad (3.112)$$

with spectral density

$$f_{\text{param}}(\lambda; \mathbf{a}) = \frac{1}{2\pi} \left| 1 - \sum_{l=1}^p a_l e^{-il\lambda} \right|^{-2}. \quad (3.113)$$

The time series is assumed to be stationary and causal (completely determined by the past) *a priori*.

One can then use the nonparametric Bernstein polynomial prior of Choudhuri et al. (2004) on the spectral density to update the nonparametrically corrected likelihood (rather than the Whittle likelihood) to produce a pseudo-posterior. However, rather than specifying a prior on the spectral density $f(\cdot)$, one can do so on a pre-whitened version of it, thus incorporating the spectral density of the parametric working model. That is, consider an *eta-damped correction function*:

$$c_\eta(\lambda) = c_\eta(\lambda; \mathbf{a}) = f(\lambda)/f_{\text{param}}(\lambda; \mathbf{a})^\eta, \quad (3.114)$$

where $\eta \in [0, 1]$ specifies the level of confidence in the parametric model. If η is close to 1, the model is well-specified and $c_\eta(\cdot)$ will be smoother than the original spectral density as $f_{\text{param}}(\cdot)$ is already a reasonable fit to the data. Estimating $c_\eta(\cdot)$ should therefore take less effort than $f(\cdot)$. If η is close to 0, any erroneous spectral peaks introduced by a misspecified parametric model will be damped.

The Bernstein polynomial prior of Choudhuri et al. (2004) is placed on $c_\eta(\cdot)$ rather than $f(\cdot)$. The only difference in the set-up is to choose $p(k) \propto \exp(-\theta_k k \log(k))$ as the discrete prior for k .

The prior on $c_\eta(\cdot)$ induces a prior on $f(\cdot)$ via multiplication with $f_{\text{param}}(\cdot; \mathbf{a})^\eta$. The pseudo-posterior can then be defined as:

$$p_{\text{post}}^C(V_1, \dots, V_L, Z_0, Z_1, \dots, Z_L, k, \tau | \mathbf{Y}_n, \mathbf{a}, \eta) \\ \propto p(V_1, \dots, V_L, Z_0, Z_1, \dots, Z_L, k, \tau) \det(C_n)^{-1/2} p_{\text{param}}(F_n' C_n^{-1/2} F_n \mathbf{Y}_n | \mathbf{a}),$$

where $C_n = C_n(c_\eta(\lambda; \mathbf{a}) f_{\text{param}}(\lambda; \mathbf{a})^{\eta-1})$.

One can include estimation of the parametric working model parameters into the Bayesian setting. That is, for fixed p , include parameters \mathbf{a} , as well as spectral shape parameter η . One must ensure stationarity, causality, and identifiability of the parametric model by placing a prior on the partial autocorrelations $\boldsymbol{\rho} = (\rho_1, \dots, \rho_p)$ with $\rho_l \in [-1, 1]$ for $l = 1, \dots, p$. For reparametrisation details, see Kirch et al. (2017).

The prior specification for the spectral density then becomes:

$$f(\lambda) = c_\eta(\lambda) f_{\text{param}}(\lambda; \boldsymbol{\rho})^\eta, \quad (3.115)$$

where a Uniform[0, 1] prior is placed on η , and Uniform[-1, 1] priors are placed on $\boldsymbol{\rho}$.

To visualise why a prior is placed on the eta-damped correction function, consider two simulations: AR(1) data with a well-specified AR(1) parametric working model; and AR(4) data with a misspecified AR(1) parametric working model. The nonparametric correction algorithm is run for 20,000 iterations (with a burn-in of 10,000 and thinning factor of 2) assuming these parametric working models are to be nonparametrically corrected. Figures 3.4 and 3.5 show the relationship between the estimated median PSD $f(\cdot)$, the estimated parametric PSD $f_{\text{param}}(\cdot)^\eta$ using the median values for ρ and η , and the eta-damped correction function $c_\eta(\cdot)$.

For the well-specified case, $\hat{\eta} = 0.9672$, and it can be seen in Figure 3.4 that $c_\eta(\cdot)$ is smooth and small in scale relative to $f(\cdot)$ and $f_{\text{param}}(\cdot)^\eta$. This is much easier to approximate using Bernstein polynomials than $f(\cdot)$ itself. Also notice how close $f(\cdot)$ is to $f_{\text{param}}(\cdot)^\eta$. Compare this to the misspecified case in Figure 3.4 where $\hat{\eta} = 0.0744$, $c_\eta(\cdot)$ is less smooth and close to $f(\cdot)$. This is the case where estimating $c_\eta(\cdot)$ using Bernstein polynomials will be just as challenging as estimating $f(\cdot)$, as both functions are similar.

The R package `beyondWhittle` was developed to conduct this analysis in, and is available on CRAN (Meier et al., 2017). Earlier versions of the code were also used for the methods of 3.5.4.1.

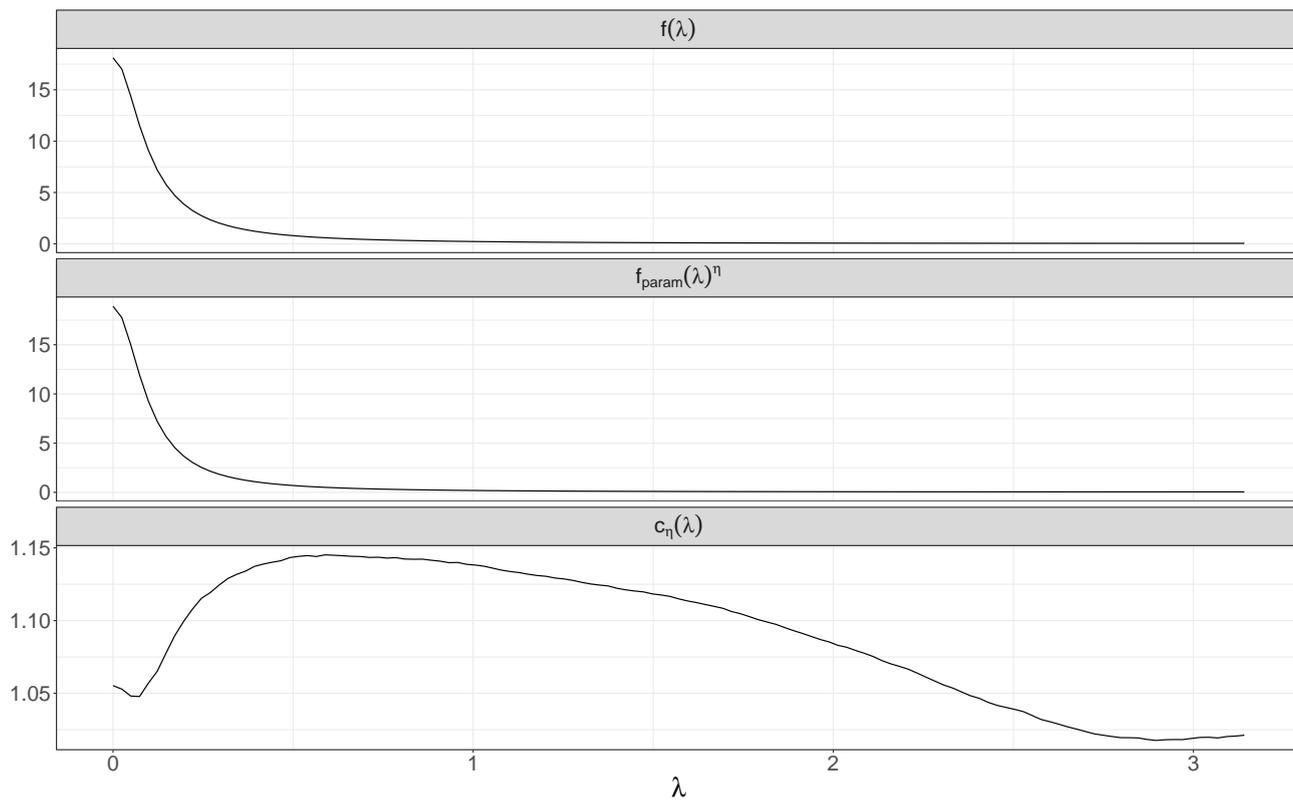


FIGURE 3.4: Nonparametric correction with an AR(1) working model applied to AR(1) data with $a_1 = 0.9$. This parametric working model is well-specified.

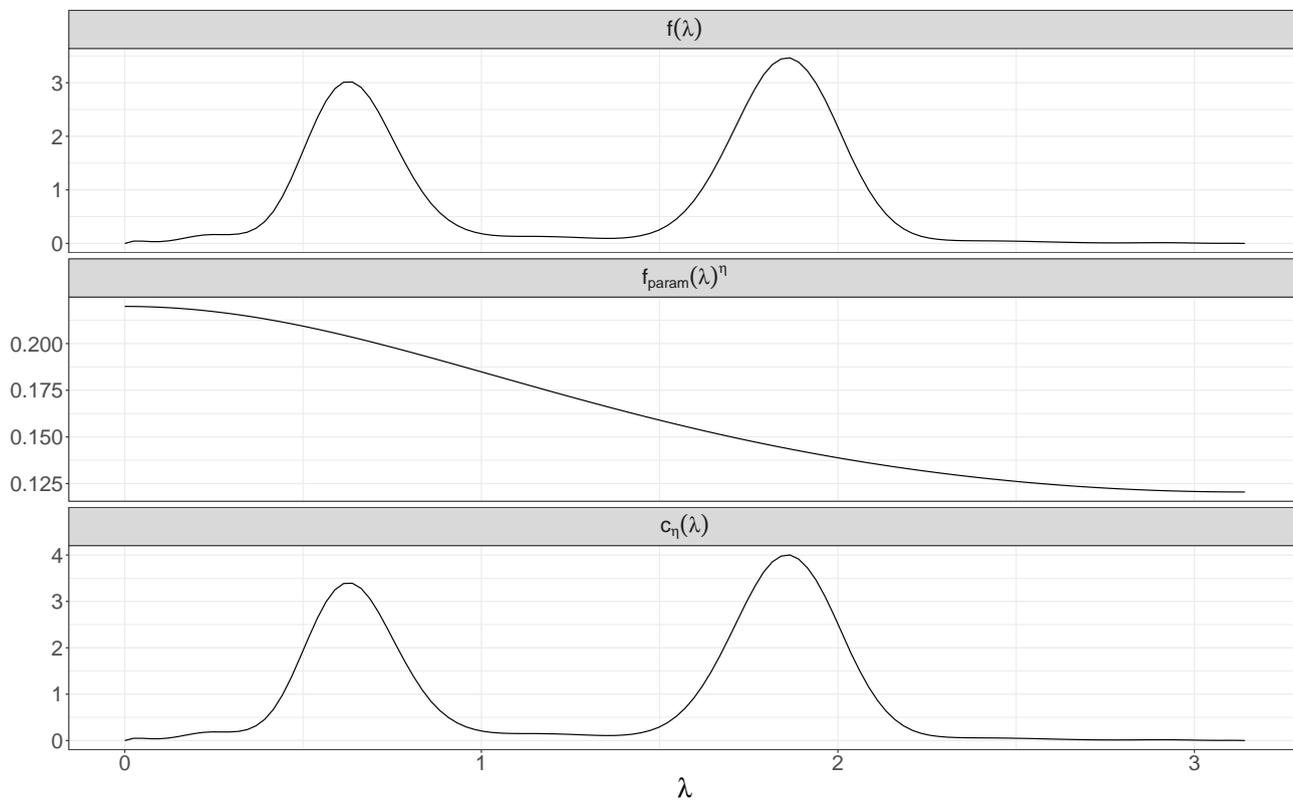


FIGURE 3.5: Nonparametric correction with an AR(1) working model applied to AR(4) data with $a_1 = 0.9, a_2 = -0.9, a_3 = 0.9,$ and $a_4 = -0.9$. This parametric working model is misspecified.

3.6 Principal component analysis

In Chapter 4, principal component analysis and principal component regression are used to construct signals from rotating stellar core collapse that are embedded in noise. A brief review of these methods is given here.

Principal component analysis (PCA) is a multivariate technique used in dimension reduction (ordination) problems, and transforms a set of potentially correlated variables into a set of linearly uncorrelated principal components (PCs). PCA works by considering a data cloud with an orthogonal set of axes, and finding a set of eigenvectors from the covariance matrix of the data by centering and rotating these axes (Rencher, 2002). This produces a new coordinate system that the data are projected onto. Each vector that undergoes PCA can be represented as a linear combination of the orthogonal basis vectors, where the projection of the data onto the first basis vector has the largest variance (or eigenvalue), the projection onto the second basis vector has the second highest variance, and so on.

Suppose X is an $m \times n$ data matrix with m mean-centred row vectors of length n . The goal of PCA is to transform (rotate) data matrix X to a matrix of uncorrelated variables Y (PC scores) by finding an orthonormal matrix P such that $Y = PX$, where Y has a diagonalised covariance matrix $\Sigma = \frac{1}{n-1}YY'$. As XX' is a symmetric matrix, this is achieved using an eigendecomposition (also called spectral decomposition). That is, XX' may be decomposed into a matrix of its orthonormal eigenvectors E , and a diagonal matrix D (whose diagonal elements are the eigenvalues of XX') using $XX' = EDE'$. It follows that if one sets $P = E'$, then $\Sigma = \frac{1}{n-1}D$. The principal components are the eigenvectors of XX' (or rows of P) and the variances are the diagonal elements of Σ .

Principal component regression (PCR) is linear regression with the PC basis vectors as explanatory variables. Dimension reduction is achieved by considering only the first d basis vectors (with the highest eigenvalues), the so-called d PCs, that preserves as much of the information in the original data as possible. The choice of d is non-trivial, though is often based heuristically on Scree plots. However, traditional model comparison techniques may be used.

In 4.3.3, the choice of d is made by considering a constrained optimised version of DIC. In 4.3.4, a simple birth-death reversible jump MCMC algorithm is implemented to automatically select d .

3.7 Classification

Classification is a common statistical task. Given a set of groups, is it possible to predict the class of an object by probing its key features? In the context of gravitational wave data analysis,

classification methods could be used to, for example, classify various types of signals or glitches.

In this thesis, the classification methods to follow will be used to infer the precollapse differential rotation profile of a rotating stellar core collapse event. This variable is categorical, and in particular, ordinal, making classification methods well-suited for this challenge.

3.7.1 Supervised machine learning

3.7.1.1 Naïve Bayes classifier

The *Naïve Bayes classifier* (Ripley, 1996) is a common supervised learning algorithm and discriminant method used to group objects into a discrete set of classes based on a set of features. The algorithm requires a *training set* of objects with known groupings and observed features. Once the algorithm has learnt from the training set, each object in a *test set* (containing a set of observed features and potentially unknown classes) is assigned to the group that it has the highest probability of belonging to.

The “Bayes” component of the method refers to Bayes’ theorem

$$p(c|\mathbf{u}) \propto p(c)p(\mathbf{u}|c) \quad (3.116)$$

where $c \in C$ is the class that an object could belong to, and \mathbf{u} are the features exploited to classify the object. That is, given some observed features \mathbf{u} , what is the posterior probability of an object belonging to class c ?

The “naïve” component refers to the assumption of conditional independence of the model features $\mathbf{u} = (u_1, u_2, \dots, u_d)$, where d is the number of different features. This assumption implies that the joint pdf $p(\mathbf{u}|c)$ can be factorised as the product of marginal distributions

$$p(\mathbf{u}|c) = \prod_{i=1}^d p(u_i|c), \quad (3.117)$$

and so equation (3.116) becomes

$$p(c|\mathbf{u}) = p(c) \prod_{i=1}^d p(u_i|c). \quad (3.118)$$

Given class c , each feature (u_1, u_2, \dots, u_d) is assumed to be independently normally distributed. The model parameters are approximated using the relative frequencies from the training set. The class prior probabilities $p(c)$ are specified as the number of objects in class c in the training set

divided by the total number of objects. Objects are grouped into the class that yields the highest posterior probability. This is known as the maximum *a posteriori* (MAP) decision rule.

3.7.1.2 *k*-nearest neighbour

An alternative supervised machine learning algorithm to the naïve Bayes classifier is the *k*-nearest neighbour algorithm (Ripley, 1996), which uses a measure of “closeness” between objects rather than a probabilistic framework. If $k = 1$, an object in the test set is assigned to the class of its single nearest neighbour in the training set, where ties in distance are settled at random.

The definition of closeness in this context depends on the choice of metric, though Euclidean distance is often assumed in the literature (Ripley, 1996). For any object with features $\mathbf{u} = (u_1, u_2, \dots, u_d)$ in the test set, the *k*-nearest neighbour algorithm finds the object with features $\mathbf{v} = (v_1, v_2, \dots, v_d)$ in the training set that minimises the Euclidean distance

$$\text{distance}(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^d (u_i - v_i)^2}, \quad (3.119)$$

and then assigns \mathbf{u} to the class of \mathbf{v} .

3.7.2 Generalised linear models

Generalised linear models (GLMs) are a generalisation of general linear models, allowing for categorical response variables and error distributions that are not normally distributed (but belong to the exponential family of distributions) (Agresti, 2002). GLMs are also useful for binary, nominal, and ordinal classification problems. The key component of GLMs is the *link function* $g(\cdot)$ that links the expected value of the response variable Y to the linear predictor $X\boldsymbol{\alpha}$. That is, $E[Y] = g^{-1}(X\boldsymbol{\alpha})$.

3.7.2.1 Bayesian ordinal probit regression

If one has an ordered categorical response variable, it is possible to conduct a Bayesian analysis using a *Bayesian ordinal probit regression* model. Though ordinal logit models are generally more popular under the frequentist framework, ordinal probit models are easier to implement in the Bayesian context.

Using the framework presented by Albert and Chib (1993), let Y_1, \dots, Y_n be categorical response variables, with ordered classes $c = 1, \dots, C$. Let the cumulative probabilities for each observation $i = 1, \dots, n$ and each class $c = 1, \dots, C$ be denoted $\eta_{ic} = \mathbb{P}(Y_i \geq c) = \Phi(\gamma_c - \mathbf{x}'_i \boldsymbol{\alpha})$. That is, assume there is a continuous latent variable $Z_i \sim N(\mathbf{x}'_i \boldsymbol{\alpha}, 1)$, where $Y_i = c$ if $\gamma_{c-1} < Z_i \leq \gamma_c$. Here, $(\gamma_0, \gamma_1, \dots, \gamma_C)$ are the cut-points for each category, where it is assumed that $\gamma_0 = -\infty, \gamma_1 = 0$ and $\gamma_C = \infty$.

Assuming diffuse priors on $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, the joint posterior is:

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, Z|\mathbf{y}) \propto \prod_{i=1}^n \left(\sqrt{\frac{1}{2\pi}} \exp \left\{ -\frac{(Z_i - \mathbf{x}'_i \boldsymbol{\alpha})^2}{2} \right\} \times \left(\sum_{c=1}^C I_{\{Y_i=c\}} I_{\{\gamma_{c-1} < Z_i \leq \gamma_c\}} \right) \right). \quad (3.120)$$

Sampling from the posterior is done via a Gibbs sampler. However, implementation details may differ depending on how one samples the cut-points (Cowles, 1996).

Chapter 4

Parameter estimation for rotating stellar core collapse

4.1 Introduction

Due to the analytical intractability and complex multi-dimensional nature of rotating stellar core collapse events, a significant amount of computational time must go into numerically simulating the gravitational waveforms. This involves simulating all of the physics relating to stellar core collapse, which can be considered a *mélange* of general relativity, particle physics, and nuclear physics. Unlike binary inspiral events, one cannot simply use template-based search methods for supernova burst events as it is computationally impossible to cover the entire signal parameter space. It is therefore important to derive alternative parameter estimation techniques.

The expected rate of core collapse supernovae in the Milky Way is around three per century (Adams et al., 2013). It is therefore important to ensure that appropriate data analysis routines are in place so as not to miss an opportunity to detect and make timely statements about these rare events. Some of the following approaches have aimed to estimate key parameters of stellar core collapse events using their gravitational wave signals.

Summerscales et al. (2008) utilised the maximum entropy framework to deconvolve noisy data from multiple (coherent) detectors, with the goal of extracting a core collapse supernova gravitational wave signal. Inference on amplitude and phase parameters was conducted using cross correlation between the recovered waveform and the set of competing waveforms from the Ott et al. (2004) catalogue. A match was defined as the model with the maximum cross correlation to the recovered waveform.

Heng (2009) first proposed a PCA approach to simplify the problem by reducing a given supernova waveform catalogue space down to a small number of basis vectors. Röver et al. (2009) extended this approach and created a novel Metropolis-within-Gibbs sampler to reconstruct test signals from the Dimmelmeier et al. (2008) catalogue in noisy data using a PCR model with random effects and unknown signal arrival time. The authors then attempted to exploit the structure of the posterior PC coefficients with a simple χ^2 measure of distance to determine which catalogue waveform best matched the injected test signal. Although the Bayesian reconstruction method showed much promise, extraction of the underlying physical parameters had limited success.

Logue et al. (2012) used the nested sampling algorithm of Skilling (2006) to compute Bayesian evidence for PCR models under three competing supernova mechanisms — neutrino, magnetorotational, and acoustic mechanisms. Each supernova mechanism has a noticeably distinct gravitational waveform morphology, and the method was successful at correctly inferring a vast majority of injected signals. The authors found that for signals embedded in simulated Advanced LIGO noise, the magnetorotational mechanism could be distinguished to a distance of up to 10 kpc, and the neutrino and acoustic mechanisms up to 2 kpc. This work was later extended by Powell et al. (2016), using real noise from a network of detectors and Bayes factors to select an appropriate number of PCs.

Abdikamalov et al. (2014) generated a new rotating core collapse waveform catalogue and applied matched filtering to infer total angular momentum to within $\pm 20\%$ for rapidly rotating cores. For slowly rotating cores, this changed to $\pm 35\%$. Along with matched filtering, they employed the Bayesian model selection method presented by Logue et al. (2012) to illustrate that under certain assumptions of the rotation law, the second generation of gravitational wave detectors (Advanced LIGO, Advanced Virgo, and KAGRA), could also extract information about the degree of precollapse differential rotation. The two methods worked particularly well for rapidly rotating cores.

In contrast to the Bayesian approach, Engels et al. (2014) used multivariate regression and classical hypothesis testing to analyse important astrophysical parameters from rotating core collapse signals. Rather than reconstructing the waveforms using a linear combination of PCs, the authors used least squares to find an encoded relationship between the PC basis functions and the astrophysical parameters. They could identify the most important astrophysical parameters in the presence of simulated detector noise.

The majority of this chapter describes the research presented by Edwards et al. (2014). The aim of this research was to show that it is possible to extract astrophysically meaningful information about a rotating progenitor undergoing core collapse using its statistically reconstructed gravitational wave signal. Bayesian PCR models were used to reconstruct a gravitational wave signal embedded

in simulated Advanced LIGO noise. Known astrophysical parameters were then regressed on the posterior means of the PC coefficients using a Bayesian linear regression model. The ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce was estimated by sampling from the posterior predictive distribution, and precollapse differential rotation was classified using two supervised machine learning algorithms.

In addition to the results of Edwards et al. (2014), some new (unpublished) results are also presented. An alternative signal reconstruction method based on birth-death reversible jump MCMC is implemented to remove the arbitrariness of choosing an appropriate number of principal components. A Bayesian ordinal probit model is used to classify precollapse differential rotation. Further parameter estimation results based on model comparison and cross-validation are also presented, with the aim of inferring the nuclear equation of state (EOS).

4.2 Core collapse waveform catalogue

A description of the rotating core collapse and bounce gravitational wave simulations by Abdikamalov et al. (2014) is given in this section. This data catalogue is used to show in the sections to follow that it is possible to extract astrophysically meaningful information about rotating stellar core collapse using Bayesian methods and supervised machine learning algorithms.

The waveforms come from two-dimensional numerical axisymmetric general-relativistic hydrodynamic rotating core collapse and bounce supernova simulations. Based on findings that gravitational wave signals are essentially independent of the progenitor zero age main sequence (ZAMS) mass by Ott et al. (2012), a single presupernova progenitor model (the 12 solar mass at ZAMS solar-metallicity progenitor model from Woosley and Heger (2007)) was adopted. The cylindrical rotation law from Ott et al. (2004) was also assumed.

The gravitational wave catalogue is partitioned into a training set and a test set. The training set contains $l = 92$ signals with five levels of precollapse differential rotation A (where higher values of A represent weaker differential rotation), a grid of values for initial central angular velocity Ω_c , as well as a grid of values for the ratio of rotational kinetic energy to gravitational energy of the inner core at bounce $\beta_{ic,b}$ (since $\beta_{ic,b}$ is a function of Ω_c for a fixed progenitor structure). Each signal in the training set was generated using the microphysical finite-temperature Lattimer-Swesty (LS) EOS (Lattimer and Swesty, 1991), parametrised deleptonisation scheme from Dimmelmeier et al. (2008), and neutrino leakage scheme from Ott et al. (2012). As well as varying A , Ω_c , and $\beta_{ic,b}$, the test set contains $m = 47$ signals with differing EOS and deleptonisation parametrisations $Y_e(\rho)$. Specifically, some test signals were generated using the Shen EOS (Shen et al., 1998), or

an increase/decrease in $Y_e(\rho)$ parametrisation by $\sim 5\%$. The values of Ω_c and $\beta_{ic,b}$ in the test set are in the same parameter space as those in the training set, but with an alternative grid. The object of this analysis is to infer the physical parameters ($\beta_{ic,b}$, A , and EOS) of the signals in the test set using information gleaned about signals in the training set.

The signals were initially sampled at 100 kHz and subsequently downsampled by a rational factor to 16384 Hz (the sampling rate of the Advanced LIGO detectors). As mentioned in Section 3.4.6, downsampling by a rational factor essentially involved two steps: upsampling by an integer factor via interpolation and then applying a low-pass filter to eliminate the high frequency components (this step is necessary to avoid aliasing at lower sampling rates); and downsampling by an integer factor to achieve the desired sampling rate. The resampled data was zero-buffered to ensure each signal was the same length, $n = 16384$, which corresponded to 1 s of data at the Advanced LIGO sampling rate. Each signal was then aligned so that the first negative peak (not necessarily the global minimum), corresponding to the time of core bounce, occurred halfway through the time series.

In this analysis, the source of a gravitational wave emission is assumed to be optimally oriented (perpendicular) to a single interferometer. Each signal is linearly polarised with zero cross-polarisation.

A general waveform morphology is illustrated in Figure 4.1. During core collapse, there is a slow increase in gravitational wave strain until the first local maximum is reached (before 0.5 s). This is followed by core bounce, where the strain rapidly decreases towards a local minimum (at 0.5 s). This corresponds to the time when the inner core expands at bounce. After this, there is a period of ring-down oscillations. For slowly rotating progenitors (as seen in the top panel of Figure 4.1), the gravitational wave strain is essentially the same during collapse and bounce and only differs during the stochastic ring-down. For the rapidly rotating progenitors (presented in the bottom panel of Figure 4.1), larger precollapse differential rotation results in: a smaller local maximum during core collapse; a more negative local minimum during core bounce; and a larger first ring-down peak. Because of these patterns, Abdikamalov et al. (2014) concluded that inferences about precollapse differential rotation could in principle be made for rapidly rotating cores.

The data analysed are rotating core collapse gravitational wave signals injected in additive Gaussian noise, coloured by the Advanced LIGO design sensitivity spectral density, $f(\cdot)$. According to LALAdvLIGOPsd (Sathyaprakash, 2007), the Advanced LIGO noise PSD is

$$f(\nu) = f_0 \left\{ \left(\frac{\nu}{\nu_0} \right)^{-4.14} - 5 \left(\frac{\nu_0}{\nu} \right)^2 + 111 \left(\frac{1 - \left(\frac{\nu}{\nu_0} \right)^2 + 0.5 \left(\frac{\nu}{\nu_0} \right)^4}{1 + 0.5 \left(\frac{\nu}{\nu_0} \right)^2} \right) \right\}, \quad (4.1)$$

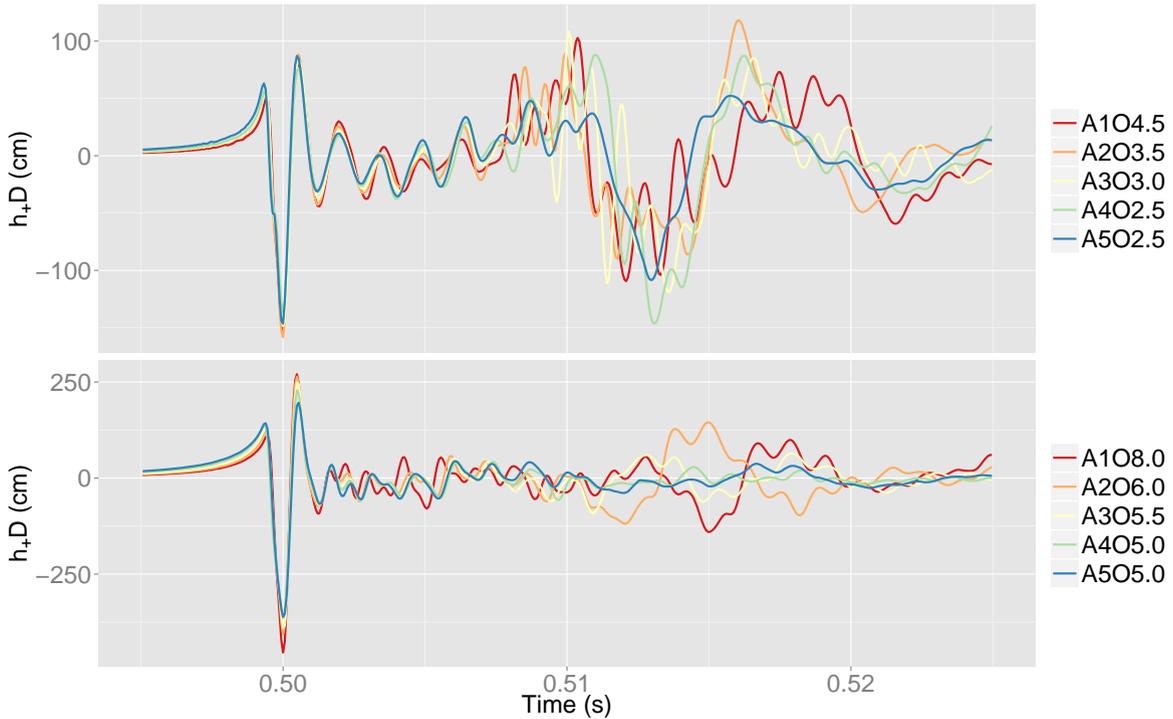


FIGURE 4.1: A snapshot of the Abdikamalov et al. (2014) catalogue. The top panel shows the gravitational wave strain (scaled by source distance) for five models with different levels of precollapse differential rotation (from strongest differential rotation $A1$ to weakest $A5$), each with $\beta_{ic,b} \sim 0.03$ (i.e., slowly rotating progenitors). The bottom panel is the same, but for rapidly rotating progenitors with $\beta_{ic,b} \sim 0.09$.

where ν is the frequency in Hertz and $\nu_0 = 215$ Hz (not to be confused with the Nyquist frequency of the same symbol). The returned value is scaled up by $f_0 = 10^{49}$. Noise is generated in the frequency-domain by sampling Gaussian random variables with zero mean and variances given in 4.3. These complex-valued random variables are inverse Fourier transformed and then the real parts are taken to be the time-domain Gaussian random variables.

The data are then Tukey windowed to mitigate spectral leakage. Rather than fixing source distance to 10 kpc (as done by Abdikamalov et al. (2014)), this analysis assumes a fixed SNR of $\varrho = 20$. As done by Röver et al. (2009), $f(\cdot)$ is estimated *a priori* by averaging 1000 empirical periodograms from identically simulated Advanced LIGO noise using Bartlett's method. This corresponds to a realistic scenario where the noise spectrum must be estimated as well. Although supernovae from the Milky Way will not produce SNRs as small as $\varrho = 20$, this value was chosen to illustrate that the methods are robust at lower SNRs.

4.3 Signal reconstruction

PCA and PCR approaches to signal reconstruction and parameter estimation of supernova gravitational wave signals have become popular over recent years (Abdikamalov et al., 2014, Heng, 2009, Logue et al., 2012, Powell et al., 2016, Röver et al., 2009). Signal reconstruction models 1 and 2 in the following subsections are based on the work of Röver et al. (2009) and used in Edwards et al. (2014) to infer important astrophysical parameters. Model 3 is a novel alternative to the problem, based on trans-dimensional reversible jump MCMC and model averaging.

4.3.1 Signal reconstruction model 1: Metropolis-within-Gibbs PCR with known signal arrival time

A PCA is applied to the training set to reduce dimensionality. Each training waveform is represented as a linear combination of orthonormal basis vectors, where the projection of the data onto the first basis vector has maximum variance, the projection onto the second basis vector has second highest variance, and so on. By considering only projections on the first $d < l$ basis vectors, the so-called d PCs, a parsimonious representation of the catalogue signals in d dimensions is achieved that preserves as much of the information of the original training set as possible. The first four PCs from the Abdikamalov et al. (2014) catalogue are illustrated in 4.2. Notice particularly how the first PC contains the obvious collapse (first positive peak) and bounce (large negative peak) morphology, and that the higher PCs (i.e., the ones with the lower eigenvalues) show more of the post-bounce ringdown morphology, and are on a smaller scale.

Once PCA is conducted, the first d PCs are treated as the explanatory variables of a linear model. The data analysed are a time series vector \mathbf{y} of length n and decomposes into additive signal and noise components. Let $\tilde{\mathbf{y}}$ be the Fourier transformed data vector of length n and let \tilde{X} be the $n \times d$ design matrix, whose columns are the Fourier transformed mean-centred PC vectors from the base catalogue. The frequency domain linear model is

$$\tilde{\mathbf{y}} = \tilde{X}\boldsymbol{\alpha} + \tilde{\boldsymbol{\epsilon}}, \quad (4.2)$$

where $\boldsymbol{\alpha}$ is the vector of PCR coefficients and $\tilde{\boldsymbol{\epsilon}}$ is the Fourier transformed coloured zero-mean Gaussian noise vector whose variance terms are

$$\sigma_{\nu_j}^2 = \frac{n}{4\Delta_t} f(\nu_j). \quad (4.3)$$

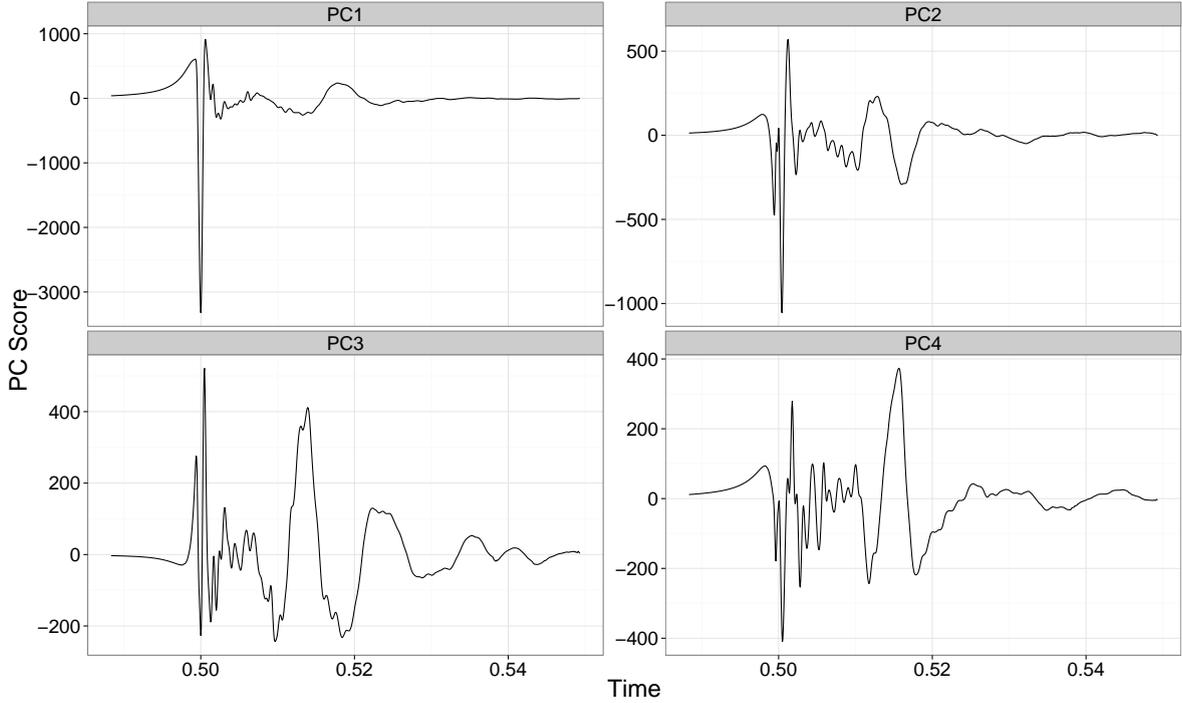


FIGURE 4.2: The first four principal components from the Abdikamalov et al. (2014) catalogue.

Here, ν_j refers to (nonredundant positive) frequency in Hertz, Δ_t is the sampling interval, and $f(\cdot)$ is the *a priori* known (i.e., pre-estimated) noise spectral density. Due to Hermitian symmetry, the frequency domain data vector $\tilde{\mathbf{y}}$ contains only the non-redundant real and imaginary components and is therefore the same length as the time domain vector \mathbf{y} . Conversion between time and frequency domains is conducted using a FFT.

The likelihood for the Bayesian PCR model with known signal arrival time is

$$p(\tilde{\mathbf{y}}|\boldsymbol{\alpha}) \propto \exp \left(-2 \sum_{j=1}^n \frac{\frac{\Delta_t}{n} \left(\tilde{y}_j - (\tilde{X}\boldsymbol{\alpha})_j \right)^2}{f(\nu_j)} \right). \quad (4.4)$$

Assuming Uniform $(-\infty, \infty)$ (flat) priors on $\boldsymbol{\alpha}$, the posterior distribution for the PC coefficients is

$$\boldsymbol{\alpha}|\tilde{\mathbf{y}} \sim \text{N}(\boldsymbol{\mu}, \Sigma), \quad (4.5)$$

where

$$\Sigma = (\tilde{X}' D^{-1} \tilde{X})^{-1}, \quad (4.6)$$

$$\boldsymbol{\mu} = \Sigma \tilde{X}' D^{-1} \tilde{\mathbf{y}}, \quad (4.7)$$

and $D = \text{diag}(\sigma_{\nu_j}^2)$ is the diagonal covariance matrix of individual variances for the noise component. This multivariate normal distribution can be sampled directly with no MCMC required.

Noninformative priors were chosen for this model as there is very little information about the constraints of the PCs. Most information learnt in the algorithm therefore comes from the likelihood rather than the prior. It was important to keep the data and prior knowledge separate and distinct, and to avoid using information from the waveform catalogue for both purposes. That is, an empirical Bayes approach was not deemed appropriate for the given scenario. As the only data available for analysis were the generated gravitational waves, complete prior ignorance on all reconstruction model parameters was assumed.

For comparability with the signal reconstruction model in the next subsection, 100,000 values are directly sampled from the posterior, 10,000 are removed as “burn-in”, and the remaining samples are thinned by a factor of 5.

4.3.2 Signal reconstruction model 2: Metropolis-within-Gibbs PCR with unknown signal arrival time

The Bayesian PCR model presented in the previous section assumed a known signal arrival time. The precise arrival time of a gravitational wave signal to an interferometer will generally not be known in practice, and must therefore be included as an additional unknown parameter in the statistical model.

Let T be a cyclical time shift representing the unknown signal arrival time, and let \tilde{X}_T be the Fourier transformed design matrix \tilde{X} shifted by lag T , such that the Fourier transformed PCs are aligned with the Fourier transformed data vector $\tilde{\mathbf{y}}$. This transformation can be done directly in the frequency domain as a phase shift by multiplying the columns of \tilde{X} by $\exp(-2\pi i\nu T)$, where ν is frequency in Hertz (not angular frequency).

This signal reconstruction model is based on the model presented by Röver et al. (2009), although the primary goal in this analysis is inferring the physical parameters of a supernova progenitor, and not signal reconstruction.

Using the same reasoning described in the previous section, assume flat priors on $\boldsymbol{\alpha}$ and T . The likelihood for the Bayesian PCR model with unknown signal arrival time is

$$p(\tilde{\mathbf{y}}|\boldsymbol{\alpha}, T) \propto \exp\left(-2\sum_{j=1}^n \frac{\frac{\Delta t}{n} \left(\tilde{y}_j - \left(\tilde{X}_T \boldsymbol{\alpha}\right)_j\right)^2}{f(\nu_j)}\right). \quad (4.8)$$

For a given time shift T , the conditional posterior distribution for the PC coefficients $\boldsymbol{\alpha}|T$ is

$$\boldsymbol{\alpha}|T, \tilde{\mathbf{y}} \sim \text{N}(\boldsymbol{\mu}_T, \Sigma_T), \quad (4.9)$$

where

$$\Sigma_T = (\tilde{X}'_T D^{-1} \tilde{X}_T)^{-1}, \quad (4.10)$$

$$\boldsymbol{\mu}_T = \Sigma_T \tilde{X}'_T D^{-1} \tilde{\mathbf{y}}. \quad (4.11)$$

To estimate $\boldsymbol{\alpha}$ and T , a Markov chain is constructed, whose stationary distribution is the posterior distribution of interest using Metropolis-within-Gibbs sampler. This is essentially a Gibbs sampler that alternates between the full set of conditional posterior distributions $p(\boldsymbol{\alpha}|T, \tilde{\mathbf{y}})$ and $p(T|\boldsymbol{\alpha}, \tilde{\mathbf{y}})$. The former can be sampled directly using equation (4.9), and the latter requires a random walk Metropolis step.

After initialisation, step $i + 1$ in the Metropolis-within-Gibbs algorithm is:

1. Directly sample the conditional posterior of $\boldsymbol{\alpha}^{(i+1)}|T^{(i)}$ using equation (4.9);
2. Propose $T^{(*)}$ from $t_{\nu_{\text{df}}}(T^{(i)}, \zeta^2)$ and accept $T^{(i+1)} = T^{(*)}$ with the Metropolis acceptance probability

$$r = \min \left(1, \frac{p(T^{(*)}|\boldsymbol{\alpha}, \tilde{\mathbf{y}})}{p(T^{(i)}|\boldsymbol{\alpha}, \tilde{\mathbf{y}})} \right). \quad (4.12)$$

Otherwise reject and set $T^{(i+1)} = T^{(i)}$.

A Student- t distribution was chosen as the proposal distribution for the algorithm. It has a similar (symmetrical) shape to the normal distribution but has heavier tails and an additional degrees-of-freedom parameter, ν_{df} . The heavier tails of the Student- t distribution results in bolder proposals than the normal distribution. This mitigates getting stuck in local modes and aims to improve mixing. The degrees-of-freedom parameter was set to $\nu_{\text{df}} = 3$, which is the smallest integer that yields a distribution with finite variance. Note that smaller ν_{df} implies heavier tails in the Student- t distribution, which gives higher probability density towards the tails. This is what allows for bolder proposals, which could improve mixing. The proposal for $T^{(i+1)}$ is centred on $T^{(i)}$, and has scale parameter ζ^2 that is initially and arbitrarily set to 0.05, and subsequently automatically tuned during the algorithm to ensure good mixing and acceptance rates using an approach that aims to achieve an acceptance probability of 0.44.

Simulations using this model run for 100,000 iterations, with a burn-in period of 10,000 and a thinning factor of 5.

4.3.3 Fixed dimension model selection

An important statistical task is to select a prudent number of model dimensions whilst incorporating Occam’s razor into the decision making process. More specifically, one needs to balance model fit against complexity to ensure over-fitting is avoided. Over-fitting becomes a significant issue when making predictions outside of a training data set. In the context of PCR, the decision is usually made based on the amount of variation the first d PCs contribute to the data set (i.e., analysing Scree plots and using some kind of “elbow criterion”). This approach is arbitrary and deals specifically with dimension reduction, but not Occam’s razor. The choice of the number of PCs has also been somewhat arbitrary in most of the supernova gravitational wave parameter estimation literature and this number has usually been $d = 10$ (see for example Röver et al. (2009) and Abdikamalov et al. (2014)). An alternative approach, involving DIC and constrained optimisation is proposed.

The optimal choice of d proposed here is based on careful analysis of the DIC for competing models and constrained optimisation. Since PCs are ordered by the total amount of variation they make up in the data set, PCA provides a convenient ordering system for nested modelling. Let M_d represent the set of possible PCR models, where $d = 1, 2, \dots, l$ is the number of explanatory variables in the model. The models are nested such that M_1 has one explanatory variable (PC_1), M_2 has two explanatory variables (PC_1 and PC_2), and so on.

For each of the $l = 92$ signals in the training set (each added to the same realisation of simulated Advanced LIGO noise for consistency), all of the models $\{M_1, M_2, \dots, M_{92}\}$ are fitted and then compared using DIC. The model with the lowest DIC is the best fit to the data. However, models with an absolute difference in DIC of $\lesssim 5$ are generally taken to be indistinguishable from one another (Spiegelhalter et al., 2002) and so to prevent over-fitting, a constrained optimisation routine is conducted, where the selected model has the smallest d such that the difference in DIC between M_d and the model with the minimum DIC is less than 5. More specifically, let M_{\min} be the model with the minimum DIC, then find d such that

$$\operatorname{argmin}_d \left\{ \text{DIC}(M_d) - \text{DIC}(M_{\min}) < 5 \right\}. \quad (4.13)$$

This routine is employed for each of the $l = 92$ training set signals, and the distribution of M_d ’s over all signals is analysed. The median of this distribution seems a prudent choice for a general-purpose number of PCs due to a skewed distribution. The median is also robust against outliers. It is important to note here that under this framework, it is not possible to choose a different value for d for each signal as this would lead to a sparse design matrix (which may be singular or cause numerical errors) when sampling from the posterior predictive distribution. Therefore, one

should conduct the proposed constrained optimisation model selection method on all of the $l = 92$ training set signals and take the median of the distribution of d 's as the general-purpose d .

The change in DIC as model dimensionality increases is analysed. Figure 4.3 illustrates DIC as a function of model dimensionality for signal *A1O2.5* from the Abdikamalov et al. (2014) catalogue. This is the typical shape of the DIC curve for all signals in the training set and a good visual aid of Occam's razor in action. There tends to be a sharp decrease in DIC as the model dimension increases at the beginning, where model fit is improving. DIC flattens out and then reaches a minimum, where there is the best balance between fit against complexity. After this, there is a slow rise in DIC as the model dimension increases and becomes too complex.

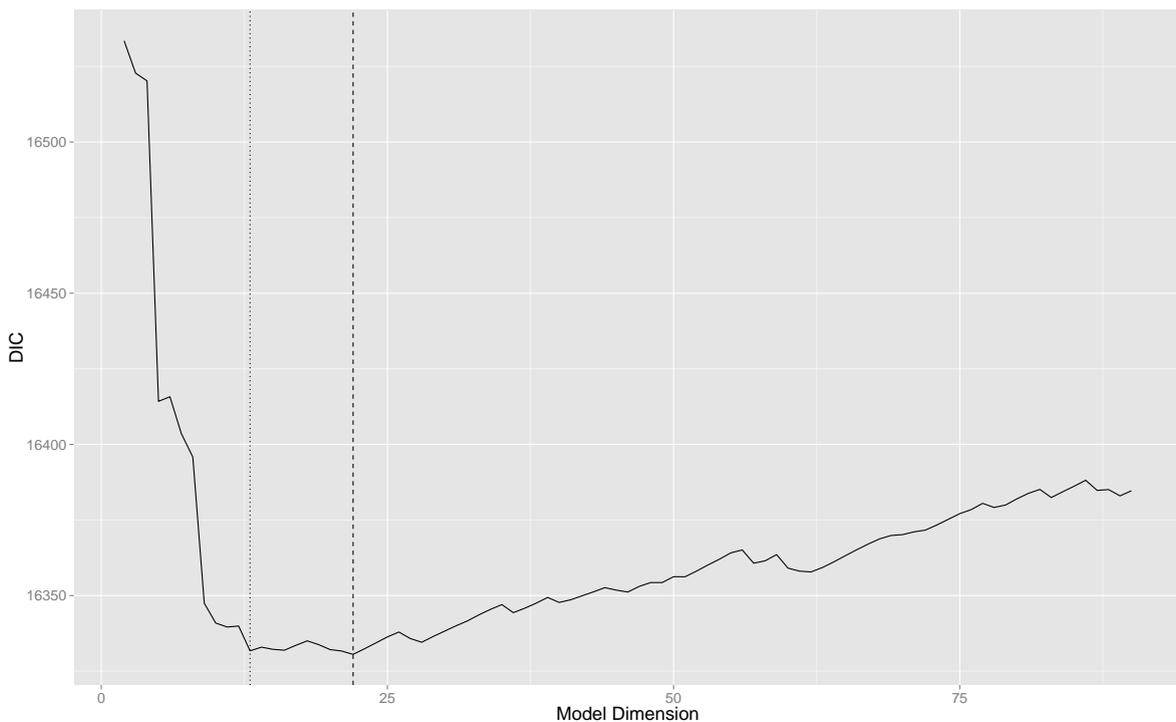


FIGURE 4.3: DIC as a function of model dimensionality for model *A1O2.5* from the Abdikamalov et al. (2014) catalogue. The dashed vertical line to the right represents the model with the minimum DIC ($M_{\min} = M_{22}$). The dotted vertical line to the left represents the model dimension after constrained optimisation ($M_d = M_{13}$).

The flat basin around the global minimum in Figure 4.3 is of particular interest. Since models with an absolute difference in DIC of less than 5 are essentially indistinguishable, it is sensible to select the model with the smallest number of dimensions in this region to prevent over-fitting. For signal *A1O2.5*, there is a significant decrease in model dimensionality from $M_{\min} = M_{22}$ to $M_d = M_{13}$. The choice of d for this particular signal (*A1O2.5*) is $d = 13$.

The histogram in Figure 4.4 shows the distribution of d for all $l = 92$ signals in the training set. It is positively skewed, with a median (and mode) of 14 PCs and mean of 17 PCs. The number of PCs is chosen to be $d = 14$, and is based on the median of this distribution. This is the number

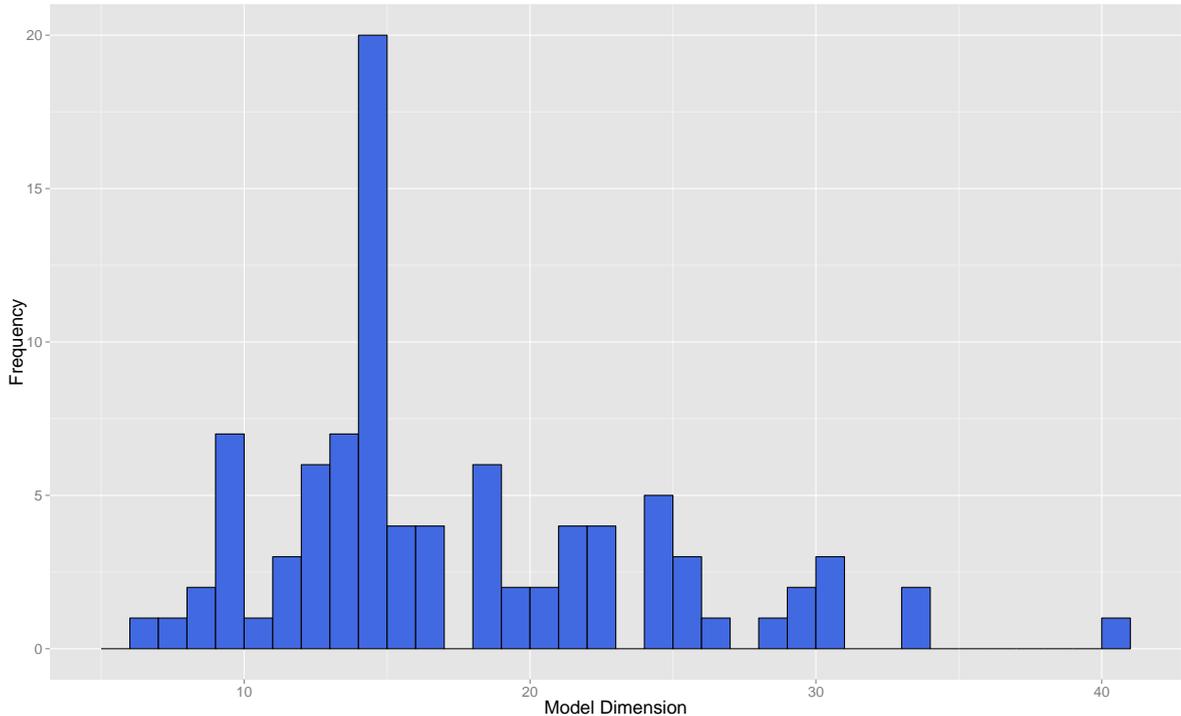


FIGURE 4.4: Distribution of model dimensionality for all $l = 92$ signals in the training set under the constrained optimisation routine.

of explanatory variables that are used in signal reconstruction models 1 and 2. Heuristically, this should mitigate the risks of both over-fitting and under-fitting as best as possible. A critique of the method is that 14 PCs will not be the optimal number of PCs for each signal. This cannot be avoided here due to the need for a fixed number of PCs to construct an explanatory matrix for posterior predictive sampling in later sections.

4.3.4 Signal reconstruction model 3: Reversible jump PCR

Rather than using a fixed number of PCs to reconstruct a rotating core collapse gravitational wave signal, the reversible jump MCMC algorithm of Green (1995) can be implemented to allow for trans-dimensional jumps. Model selection is also automatic under this framework via model averaging.

As PCs are organised according to the amount of variation they contribute to the catalogue (their eigenvalues), there is a natural ordering and nested structure. A *nested model* therefore suits this scenario well.

The algorithm is implemented with a *birth-death* regime. The algorithm decides whether to stay at the current model M_i , to move to model M_{i+1} by adding the PC with the next highest eigenvalue (birth), or to move to model M_{i-1} by removing the PC with the smallest eigenvalue (death). If the

algorithm stays at M_i , an ordinary Metropolis-Hastings move is made. As mentioned in 3.2.6.1, a common scheme is to assign one third probability to each of the jump up, jump down, and stay proposals, and this is adopted here. That is, let $r_{i,i-1}(\boldsymbol{\theta}_i) = r_{i,i}(\boldsymbol{\theta}_i) = r_{i,i+1}(\boldsymbol{\theta}_i) = \frac{1}{3}$ if $i \neq \{1, d\}$. Note that if $i = 1$ then $r_{1,1}(\boldsymbol{\theta}_1) = r_{1,2}(\boldsymbol{\theta}_1) = \frac{1}{2}$, and if $i = d$ then $r_{d,d-1}(\boldsymbol{\theta}_d) = r_{d,d}(\boldsymbol{\theta}_d) = \frac{1}{2}$.

For simplicity, assume the gravitational wave signal is embedded in Gaussian white noise. The time domain model set-up is

$$\mathbf{y} = X\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (4.14)$$

where X is the PC design matrix, $\boldsymbol{\alpha}$ are the PC coefficients, and $\boldsymbol{\epsilon}$ is additive Gaussian white noise such that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_\epsilon^2 I)$.

Assume conjugate priors for all of the model parameters. The model parameters have the following *a priori* distributions:

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma_\alpha^2 I), \quad (4.15)$$

$$\sigma_\epsilon^2 \sim \text{IG}(a_0, b_0). \quad (4.16)$$

Setting $\sigma_\alpha^2 = 1000$ and $a_0 = b_0 = 0.001$ provides a noninformative prior set-up. However, using the zeroth order method described in 3.2.6, one can adjust σ_α^2 to tune the jumping proposal variance, σ^2 .

The acceptance ratio moving from $\boldsymbol{\theta}_i$ to $c(\boldsymbol{\theta}_i) = (\boldsymbol{\theta}_i, 0)$ is

$$A\{\boldsymbol{\theta}_i, (\boldsymbol{\theta}_i, 0)\} = \frac{1}{(2\pi\sigma_\alpha^2)^{1/2}} \frac{r_{i+1,i}}{r_{i,i+1}} \frac{\sigma}{(2\pi)^{-1/2}}. \quad (4.17)$$

One can then set this equal to 1 to solve for the proposal variance σ^2 . This is given as

$$\sigma^2 = \sigma_\alpha^2 \left(\frac{r_{i,i+1}}{r_{i+1,i}} \right)^2. \quad (4.18)$$

Thus, one can automatically tune the proposal using information from the prior.

An example, demonstrating signal reconstruction of a representative signal from the Dimmelmeier et al. (2008) catalogue is now presented. This signal is embedded in Gaussian white noise with zero mean and standard deviation 0.25. The RJMCMC algorithm runs for 100,000 iterations with a burn-in of 50,000 and a thinning factor of 5. The trace plot of model dimension is shown in Figure 4.5. For this particular signal, the number of PCs converges in the region of $d \approx 20$.

To compute the model-averaged reconstructed signal, multiply the posterior samples associated with each of these sampled models by the PC design matrix and take the pointwise mean for

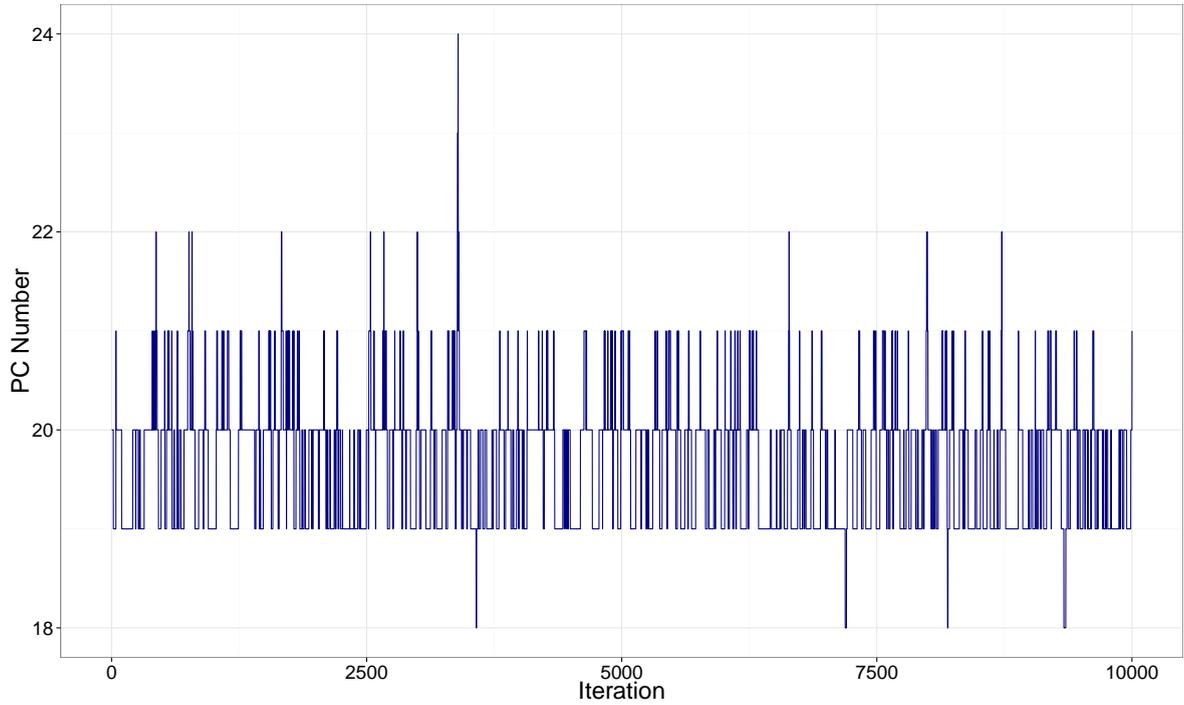


FIGURE 4.5: RJMCMC trace plot for model dimensionality for a representative signal from the Dimmelmeyer et al. (2008) catalogue.

each time point. The 90% credible region is computed in the same way, using the 5th and 95th percentiles. The reconstructed signal can be seen in Figure 4.6.

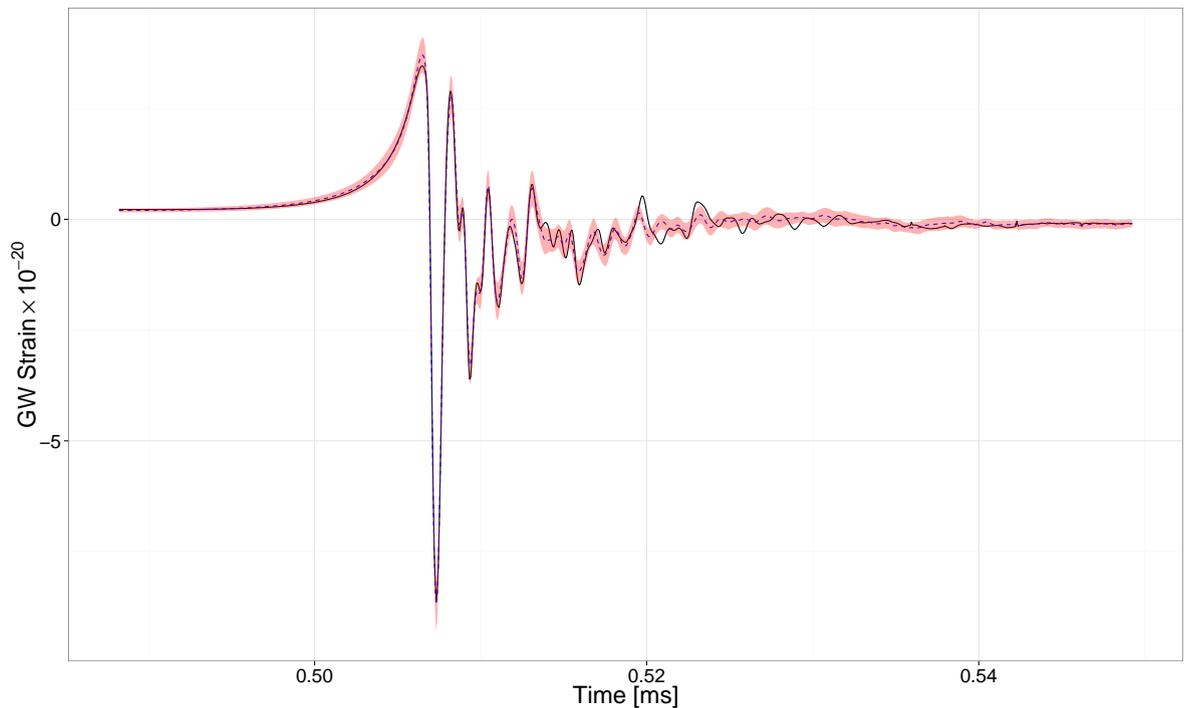


FIGURE 4.6: Reconstructed signal from the Dimmelmeyer et al. (2008) waveform catalogue. The true signal (solid black) and model-averaged estimate (dashed blue) are overlaid with the model-averaged 90% credible region (shaded pink).

The reconstructed signal in Figure 4.6 is very close to the true signal and the 90% credible region mostly contains the truth. Note that different signals will settle in a different regions of PC space depending on the need for more or less PCs to explain the signal. This will also depend largely on the prior, particularly the variance term σ_α^2 , which is used to automatically tune the proposal variance σ^2 . Larger values of σ_α^2 will favour a smaller number of PCs.

4.4 Ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce

In this section, the ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce (denoted $\beta_{ic,b}$) from stellar core collapse gravitational wave signals is estimated.

As mentioned in 4.2, the Abdikamalov et al. (2014) rotating core collapse gravitational wave catalogue is split into a *training set* and *test set*. Patterns in the data are learned using the $l = 92$ signals in the training set, and these patterns are used to make inferences about the $m = 47$ signals in the test set. For each signal in the training and test sets, Bayesian PCR signal reconstruction models (from 4.3.1 and 4.3.2) are fitted, using the first d PCs as explanatory variables (where the choice of $d = 14$ is explained in 4.3.3).

An $l \times (d + 1)$ design matrix P is then constructed, with rows equal to the *posterior means* of the d PC coefficients (plus an intercept term), for each of the l signals in the training set. That is, $P_{ij} = \hat{\alpha}_{ij}$ where $\hat{\alpha}_i$ is the PC coefficient parameter estimate for training signal i .

The primary goal is to exploit the posterior PC coefficient space to make inferences on the physical parameters of rotating core collapse stellar events in the test set. This is accomplished by fitting a linear regression model with the *known* physical parameters from the training set as the response variable on the design matrix P using

$$\boldsymbol{\beta} = P\boldsymbol{\gamma} + \boldsymbol{\epsilon}_\beta, \quad (4.19)$$

where $\boldsymbol{\beta}$ is the $l \times 1$ vector of a single known continuous physical parameter from the training set, $\boldsymbol{\gamma}$ is the vector of regression coefficients, and $\boldsymbol{\epsilon}_\beta$ is an error term, assumed to come from an iid normal distribution with zero mean and variance σ_β^2 . Estimates using the *posterior predictive distribution* are the primary interest in this analysis, and not the model parameters (PC coefficients) themselves.

Assuming the convenient noninformative prior distribution that is uniform on $(\boldsymbol{\gamma}, \log \sigma_\beta)$, the posterior predictive distribution for a normal linear model is a multivariate Student- t distribution

and can be sampled from directly with no MCMC (Gelman et al., 2013). The formula is

$$\check{\beta}|\beta \sim t_{l-d-1} \left(\check{P}\hat{\gamma}, s^2 \left(I + \check{P}V_{\beta}\check{P}' \right) \right), \quad (4.20)$$

where $\check{\beta}$ is the vector of outcomes to infer (i.e., the physical parameters from signals in the test set), \check{P} is the $m \times (d+1)$ matrix whose rows are the posterior means of the signals in the test set (and an intercept term) from the Bayesian PCR step, I is the $m \times m$ identity matrix, and

$$V_{\beta} = (P'P)^{-1}, \quad (4.21)$$

$$\hat{\gamma} = V_{\beta}P'\beta, \quad (4.22)$$

$$s^2 = \frac{1}{l-d-1}(\beta - P\hat{\gamma})'(\beta - P\hat{\gamma}). \quad (4.23)$$

Each of the $l = 92$ training set signals and $m = 47$ test set signals are added to simulated (design sensitivity) Advanced LIGO noise with SNR $\varrho = 20$. These are then fitted using the two Bayesian PCR signal reconstruction models (from 4.3.1 and 4.3.2). Using the constrained optimisation approach of 4.3.3, the first $d = 14$ PC basis vectors are chosen to reconstruct each signal.

The known physical values of $\beta_{ic,b}$ are then regressed on the posterior means of the sampled PC coefficients from the training set signals. Samples are then directly taken from the posterior predictive distribution of the test set signals to infer the unknown $\beta_{ic,b}$ for these new data.

Figures 4.7–4.10 show the estimates of $\beta_{ic,b}$. The true value from the test set (red triangle) is compared with the estimated value (blue circle) and uncertainty is measured using 90% credible intervals (black lines). Figures 4.7 and 4.8 assume a *known* signal arrival time (see 4.3.1). The signal arrival time T is *unknown* for Figures 4.9 and 4.10 (see 4.3.2). The change in background gradient for Figures 4.7 and 4.9 represents the varying precollapse differential rotation model A for signals with LS EOS and standard $Y_e(\rho)$ parametrisation. For Figures 4.8 and 4.10, the background shade represents gravitational wave signals (from a precollapse differential rotation model $A1$) with a Shen EOS, or increase/decrease in $Y_e(\rho)$ of $\sim 5\%$. $\beta_{ic,b}$ is scaled up by a factor of 100 in these plots.

Accurate estimates of $\beta_{ic,b}$ are made for the vast majority of test signals in Figure 4.7. The posterior predictive means are close to the true values of $\beta_{ic,b}$ and the 90% credible intervals generally contain the truth. However, $\beta_{ic,b}$ from signal 27 ($A5O3.25$ from the Abdikamalov et al. (2014) catalogue) has a particularly poor estimate. This signal comes from a slowly rotating core with uniform rotation, and the poor estimate here is likely to be due to the strong stochastic components in the gravitational wave signal from prompt postbounce convection (Abdikamalov et al., 2014). The true values of $\beta_{ic,b}$ are on the boundary of the 90% credible intervals for signals 3 ($A1O10.25$),

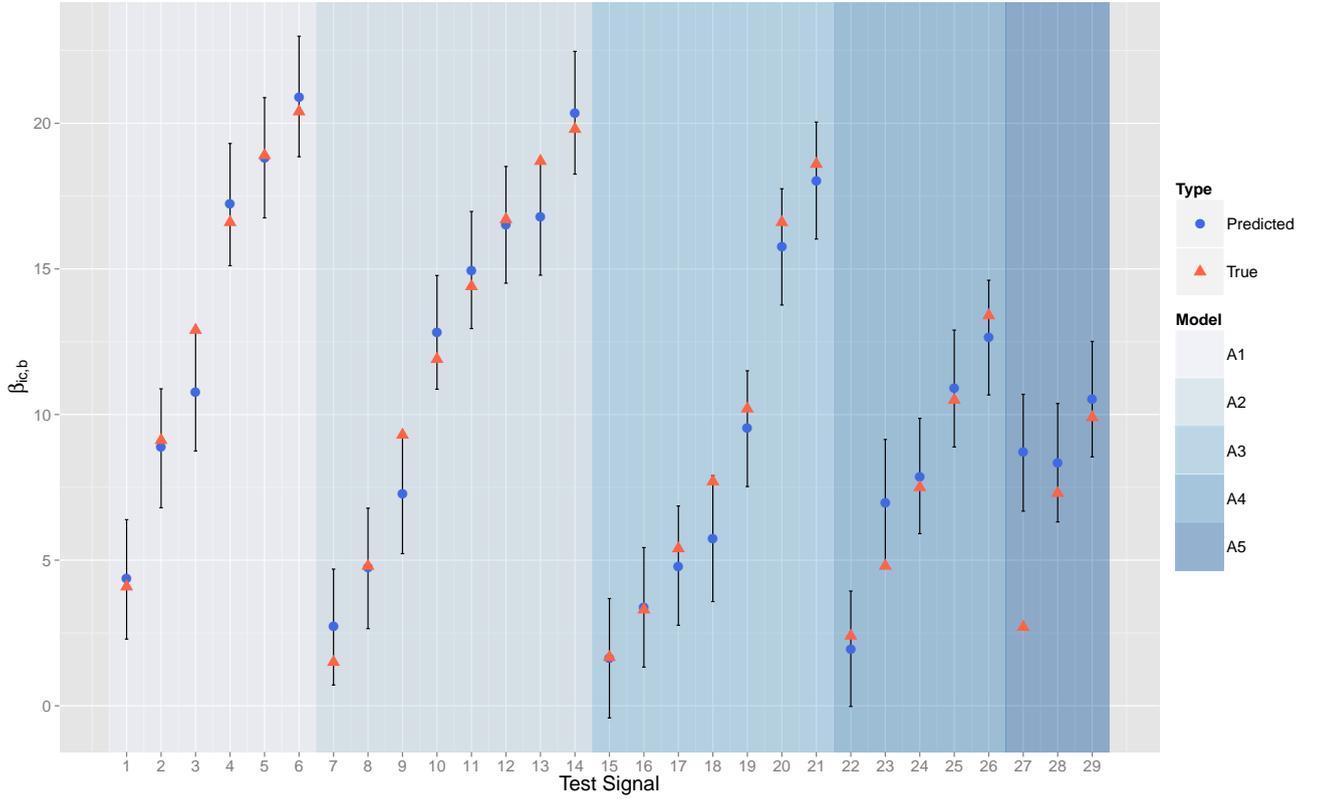


FIGURE 4.7: 90% credible intervals of $\beta_{ic,b}$ for the 29 test signals with the LS EOS and standard $Y_e(\rho)$ parametrisation. T is *known*.

9 ($A2O6.25$), 18 ($A3O5.25$), and 23 ($A4O3.25$), but there is no distinguishable pattern between these signals. The credible intervals are relatively small, at approximately four units ($\times 10^{-2}$) long. This means that it is particularly easy to distinguish $\beta_{ic,b}$ between gravitational wave signals.

The credible interval length widens by a factor of ~ 1.5 when changing from a known to unknown signal arrival time. Incorporating an unknown time shift increases the uncertainty of the PC coefficients since the MCMC algorithm draws $\alpha|T$. That is, the additional uncertainty from conditioning on T propagates through to α . However, inferences are still accurate in most cases. It can be seen in Figure 4.9 that signal 27 ($A5O3.25$) has a poor estimate again. Signal 23 ($A4O3.25$) also has an unfavourable estimate, with credible interval on the negative side of the number line. This is a physically impossible range for a strictly positive variable, and is a consequence of the priors only constraining the linear model parameters (γ, σ_β^2) . More specifically, it is not possible to place priors on the response variable of physical parameters β to constrain the estimated physical parameters $\tilde{\beta}$. A similarity between the two signals with unfavourable estimates (signals 23 and 27) are that they both come from a slowly rotating core with weak differential rotation.

The methods worked moderately well when making inferences about gravitational wave signals with varying EOS and deleptonisation parametrisations, although some signals with moderate rotation in Figure 4.8 are underestimated. Three of these signals come from an increase of $Y_e(\rho)$

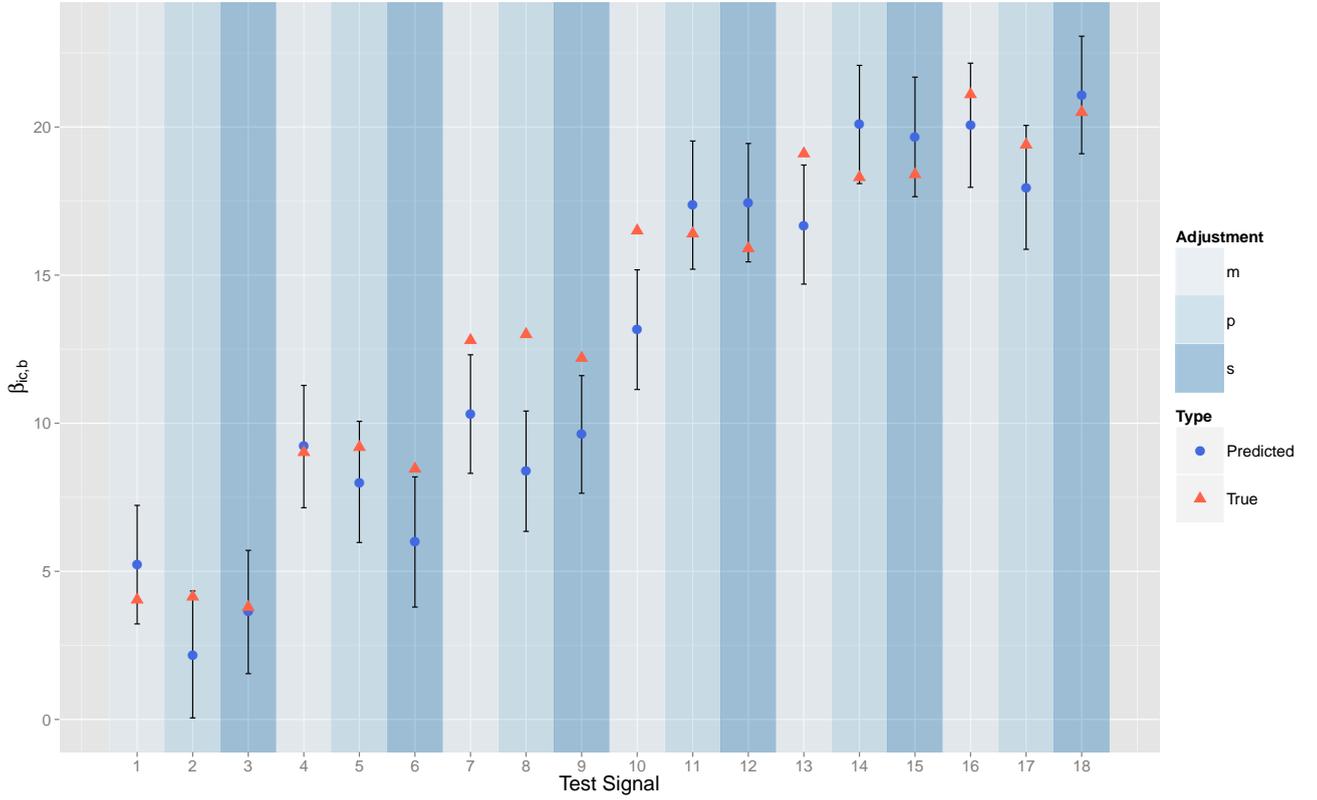


FIGURE 4.8: 90% credible intervals of $\beta_{ic,b}$ for the 18 test signals with varying EOS and $Y_e(\rho)$ parametrisation. Note that m refers to an increase in $Y_e(\rho)$ of 5%, p refers to a decrease in $Y_e(\rho)$ of 5%, and s refers to the Shen EOS. T is *known*.

parametrisation, one from a decrease of $Y_e(\rho)$ parametrisation, and two from the Shen EOS. When incorporating an unknown time shift in Figure 4.10, the uncertainty of T increases and covers the true parameters. The increase in the width of credible interval makes it more difficult to distinguish $\beta_{ic,b}$ between signals.

One could conclude that the methods employed in this section are slightly sensitive to uncertainties in $Y_e(\rho)$ and EOS. Dimmelmeier et al. (2008) found that a gravitational wave signal has relatively weak dependence on the nuclear EOS. It will also be shown in 4.6 that it was only possible to correctly identify between the LS and Shen EOS for 50% of the signals in the Dimmelmeier et al. (2008) waveform catalogue using cross-validation and model comparison techniques.

The results presented here assumed a SNR of $\varrho = 20$. To test robustness, simulations with SNRs of $\varrho = 50$ and $\varrho = 100$ were trialled (but are not presented here for brevity). These are more realistic levels for detecting rotating stellar core collapse events in the Milky Way. Estimates and credible intervals of $\beta_{ic,b}$ were approximately the same, regardless of the SNR. This can be attributed to using only the posterior means of the PC coefficients in constructing design matrix P , and not the full posterior distributions. Choosing to use the posterior means therefore artificially removed

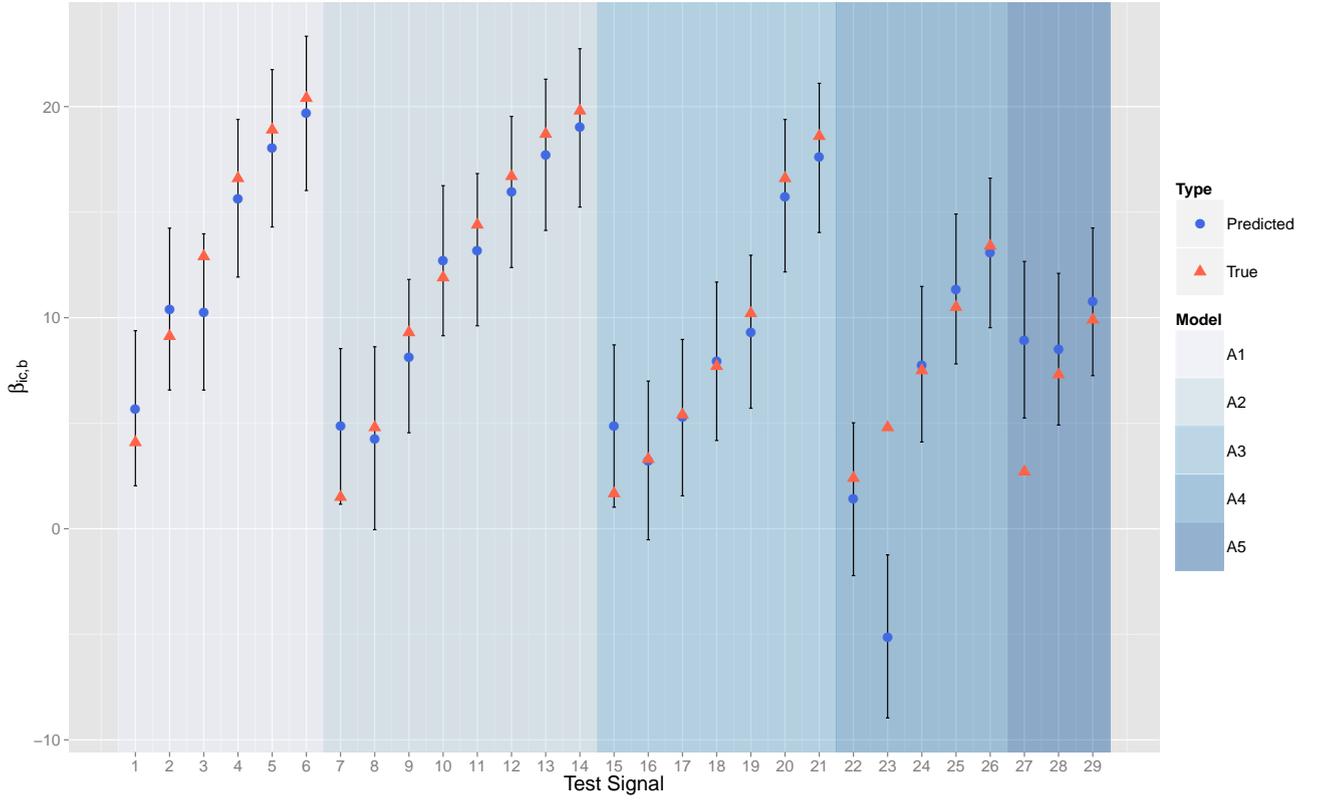


FIGURE 4.9: 90% credible intervals of $\beta_{ic,b}$ for the 29 test signals with the LS EOS and standard $Y_e(\rho)$ parametrisation. T is *unknown*.

some of the uncertainty (due to Advanced LIGO noise and signal reconstruction) when estimating $\beta_{ic,b}$.

4.5 Precollapse differential rotation

The aim of this section is to infer precollapse differential rotation from rotating stellar core collapse gravitational wave signals embedded in simulated Advanced LIGO noise. Precollapse differential rotation is treated as a categorical variable with five different levels in this analysis. Define the set of classes $C = \{A1, A2, A3, A4, A5\}$, where $A1$ has the strongest precollapse differential rotation and $A5$ has the weakest.

4.5.1 Supervised learning algorithms

In Section 4.4, gravitational wave signals were reconstructed using the PCR models of 4.3.1 and 4.3.2. Design matrix P and matrix \check{P} were then constructed using the posterior means of the PC coefficients of the reconstructed signals from the training set and test set respectively. These

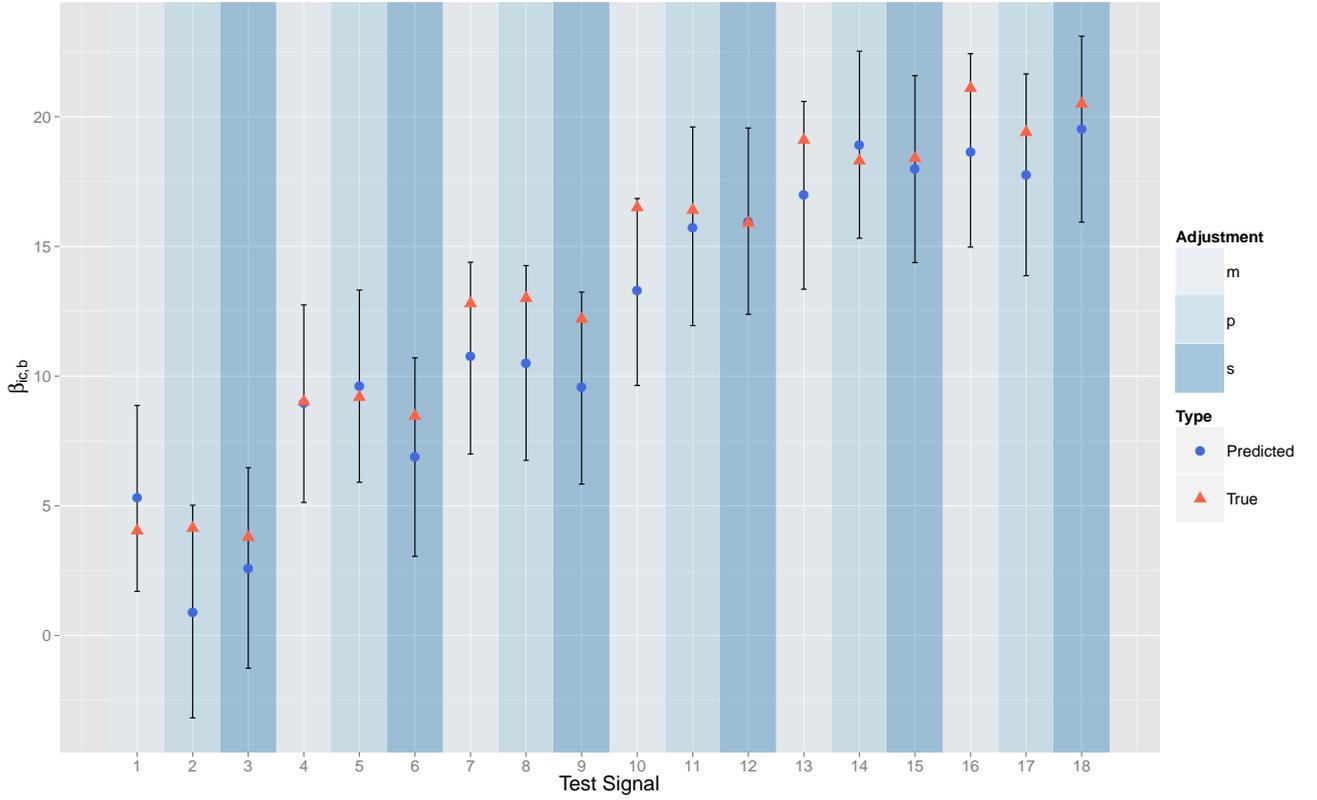


FIGURE 4.10: 90% credible intervals of $\beta_{ic,b}$ for the 18 test signals with varying EOS and $Y_e(\rho)$ parametrisation. Note that m refers to an increase in $Y_e(\rho)$ of 5%, p refers to a decrease in $Y_e(\rho)$ of 5%, and s refers to the Shen EOS. T is *unknown*.

feature matrices are inputs to the naïve Bayes classifier and k -nearest neighbour supervised learning algorithms (described in 3.7.1.1 and 3.7.1.2), used to extract precollapse differential rotation from each of the signals in the test set. The goal of this analysis is to let both algorithms learn from the training set to discriminate gravitational wave signals in the test set. Let $k = 1$ so that the single nearest neighbour is selected in the k -nearest neighbours algorithm. For naïve Bayes, assume the prior probabilities for each class are proportional to the number of signals in that class.

Table 4.1 shows the percentage of signals in the test set that have a correctly identified level of A using the naïve Bayes and k -nearest neighbour algorithms. The methods are compared using $\bar{\alpha}$ and $\bar{\alpha}$ from data with known and unknown signal arrival times.

The results between models with known and unknown signal arrival times are similar. The standard gravitational waves from class $A1$ are discriminated well by both algorithms. The decrease (and to some degree, the increase) in $Y_e(\rho)$ parametrisation did not affect the algorithms' abilities to discriminate. Both algorithms performed particularly poorly for the Shen EOS test signals, which may suggest that A is slightly sensitive to the EOS (at least for strong pre-collapse differential rotation).

TABLE 4.1: Percentage of signals in the test set with correctly identified precollapse differential rotation A using naïve Bayes (NBC) and k -nearest neighbours (k -NN).

Differential Rotation, A	Known T (%)		Unknown T (%)	
	NBC	k -NN	NBC	k -NN
$A1$				
– Standard	83	83	83	83
– $\uparrow Y_e(\rho)$	67	50	67	50
– $\downarrow Y_e(\rho)$	67	83	83	100
– Shen EOS	33	17	0	17
$A2$	50	75	50	50
$A3$	43	57	29	57
$A4$	0	80	20	80
$A5$	33	33	0	33

The k -nearest neighbours algorithm generally performs better than naïve Bayes for gravitational wave signals with weak to moderate differential rotation ($A3, A4, A5$). This could be attributed to the choice in prior classes for the naïve Bayes method. Since models with stronger differential rotation are more populated in the training set, they have a higher prior probability than those with weaker differential rotation.

4.5.2 Bayesian ordinal probit regression

Precollapse differential rotation is technically a continuous variable, but treated as a categorical variable in these analyses due to a limited grid of values available. Rather than treating the classes as nominal, it may make sense to use an ordinal model as $A1$ exhibits strong differential rotation, weakening to $A5$.

Using the `MCMCoprobit` function from the `MCMCpack` R package (Martin et al., 2011), a Bayesian ordinal probit regression model is fitted, with PC explanatory matrix P constructed using the algorithm of 4.3.1. The data augmentation algorithm of Cowles (1996) is used to sample the cut-points $(\gamma_2, \dots, \gamma_{C-1})$. The algorithm is run for 500,000 iterations, with burn-in of 250,000 and thinning factor of 10 to yield 25,000 final samples. Improper uniform (diffuse) priors are placed on all parameters. Note that the signal arrival time T is assumed to be known here.

Once the ordinal probit model is fitted to the training set, it is of interest to make inferences about what class c each signal in the test set belongs to. To do this, one can sample from the *posterior*

predictive distribution. Using Monte Carlo integration, this can be approximated as:

$$p(\mathbf{y}_{\text{new}} = c | \mathbf{y}) \approx \frac{1}{N} \sum_{j=1}^N I(\gamma_{c-1}^{(j)} < Z_i^{(j)} \leq \gamma_c^{(j)}), \quad (4.24)$$

where $Z_i^{(j)} \sim \mathcal{N}(\mathbf{x}_i' \boldsymbol{\alpha}^{(j)}, 1)$ are the continuous latent variables for signals $i = 1, \dots, 47$ at iteration j . That is, for each iteration ($j = 1, \dots, N$), multiply the posterior samples of $\boldsymbol{\alpha}$ by explanatory variables \mathbf{x}_i to find the mean of Z_i , then randomly sample one point the normal distribution with this mean, and variance 1. Next, count the number of Z_i 's in each class c according to the sampled cut-points $(\gamma_2, \dots, \gamma_{C-1})$, noting that $\gamma_0 = -\infty$, $\gamma_C = \infty$, and γ_1 is normalised to equal 0. The posterior predictive probability of being in class c is then the proportion of samples allocated into that group. For this analysis, the highest class probability then decides which differential rotation class each test signal most likely belongs to. The class probabilities for each signal in the test set can be viewed in Table 4.2.

Overall, there is 65% success rate for estimating differential rotation class using the Bayesian ordinal probit model. Of the 16 incorrectly classified signals, 10 were selected into an adjacent class. This is particularly true for signals with differential rotation class *A2* and *A4*. This is promising and demonstrates that even though the class selection for these signals is incorrect, most of these estimates are not far away from their true differential rotation, keeping in mind that differential rotation is actually a continuous parameter (but treated as categorical for these analyses due to a limited grid of values).

The aggregated results for each signal type can be seen in Table 4.3. Compare this to the results of Table 4.1. For a known signal arrival time, the Bayesian ordinal probit model outperforms both of the supervised learning algorithms for strong and weak precollapse differential rotation (*A1* and *A5* respectively), yielding better estimates. This is particularly noticeable for *A1* signals that have an increased deleptonisation parametrisation, as well as a Shen EOS. For moderate precollapse differential rotation models *A2*, *A3* and *A4*, the Bayesian ordinal probit model performs better than naïve Bayes, but not k -nearest neighbours.

4.6 Nuclear equation of state

The Dimmelman et al. (2008) waveform catalogue has a large population of 128 gravitational wave signals from rotating stellar core collapse. The catalogue uses two nuclear equations of state, where half of the signals are generated using the LS EOS and half with the Shen EOS. This is a much more ripe data set for attempting to infer a binary nuclear equation of state than

TABLE 4.2: Differential rotation class probabilities under the Bayesian ordered probit model.

Signal	A1	A2	A3	A4	A5	Estimate	Truth
A1O05.25	0.29	0.35	0.23	0.09	0.03	A2	A1
A1O05.25m	0.50	0.33	0.13	0.04	0.01	A1	A1
A1O05.25p	0.11	0.24	0.28	0.20	0.17	A3	A1
A1O05.25s	0.08	0.22	0.29	0.22	0.18	A3	A1
A1O08.25	0.91	0.08	0.01	0.00	0.00	A1	A1
A1O08.25m	0.51	0.31	0.13	0.04	0.01	A1	A1
A1O08.25p	0.68	0.23	0.07	0.02	0.00	A1	A1
A1O08.25s	0.65	0.24	0.08	0.02	0.00	A1	A1
A1O10.25	0.74	0.19	0.05	0.01	0.00	A1	A1
A1O10.25m	0.38	0.36	0.18	0.06	0.02	A1	A1
A1O10.25p	0.86	0.11	0.02	0.00	0.00	A1	A1
A1O10.25s	0.05	0.19	0.30	0.26	0.21	A3	A1
A1O12.25	0.92	0.06	0.01	0.00	0.00	A1	A1
A1O12.25m	0.88	0.10	0.02	0.00	0.00	A1	A1
A1O12.25p	0.69	0.21	0.07	0.02	0.01	A1	A1
A1O12.25s	0.57	0.29	0.11	0.03	0.00	A1	A1
A1O13.75	0.90	0.08	0.01	0.00	0.00	A1	A1
A1O13.75m	0.71	0.22	0.06	0.01	0.00	A1	A1
A1O13.75p	0.69	0.23	0.06	0.01	0.00	A1	A1
A1O13.75s	0.55	0.30	0.11	0.03	0.01	A1	A1
A1O15.25	0.69	0.23	0.07	0.01	0.00	A1	A1
A1O15.25m	0.85	0.13	0.02	0.00	0.00	A1	A1
A1O15.25p	0.43	0.32	0.17	0.06	0.02	A1	A1
A1O15.25s	0.78	0.17	0.04	0.01	0.00	A1	A1
A2O02.25	0.33	0.36	0.21	0.08	0.03	A2	A2
A2O04.25	0.58	0.28	0.11	0.03	0.01	A1	A2
A2O06.25	0.19	0.32	0.28	0.14	0.07	A2	A2
A2O07.25	0.42	0.35	0.17	0.05	0.01	A1	A2
A2O08.25	0.34	0.36	0.21	0.07	0.02	A2	A2
A2O09.25	0.33	0.36	0.21	0.08	0.02	A2	A2
A2O10.25	0.29	0.36	0.23	0.09	0.03	A2	A2
A2O11.25	0.57	0.28	0.10	0.03	0.01	A1	A2
A3O02.25	0.08	0.23	0.30	0.23	0.16	A3	A3
A3O03.25	0.06	0.18	0.28	0.25	0.23	A3	A3
A3O04.25	0.15	0.29	0.30	0.17	0.09	A3	A3
A3O05.25	0.54	0.29	0.12	0.04	0.01	A1	A3
A3O06.25	0.03	0.14	0.27	0.29	0.28	A4	A3
A3O08.25	0.03	0.15	0.28	0.27	0.27	A3	A3
A3O09.25	0.16	0.31	0.29	0.16	0.08	A2	A3
A4O02.25	0.11	0.28	0.31	0.19	0.11	A3	A4
A4O03.25	0.24	0.32	0.25	0.13	0.07	A2	A4
A4O04.25	0.01	0.05	0.15	0.25	0.55	A5	A4
A4O05.25	0.04	0.16	0.27	0.26	0.26	A3	A4
A4O06.25	0.02	0.09	0.22	0.28	0.39	A5	A4
A5O03.25	0.28	0.34	0.23	0.10	0.04	A2	A5
A5O04.25	0.01	0.07	0.18	0.27	0.46	A5	A5
A5O05.25	0.01	0.05	0.15	0.26	0.53	A5	A5

TABLE 4.3: Percentage of signals in the test set with correctly identified precollapse differential rotation A using a Bayesian ordinal probit (BOP) model.

Differential Rotation, A	BOP (%)
$A1$	
– Standard	83
– $\uparrow Y_e(\rho)$	100
– $\downarrow Y_e(\rho)$	83
– Shen EOS	67
$A2$	63
$A3$	57
$A4$	0
$A5$	67

the Abdikamalov et al. (2014) waveform catalogue. Note that as both catalogues assume the magnetorotational supernova mechanism, the waveform morphologies are similar.

The Dimmelmeier et al. (2008) catalogue is first split in two according to which EOS each signal belongs to. One subset contains only signals with the LS EOS and the other subset contains only signals with the Shen EOS.

A leave-one-out cross-validation routine is then implemented, where one signal is removed (from one of the sub-catalogues), and a PCA is conducted on both sub-catalogues, creating a separate set of LS PCs and Shen PCs. The removed signal is then embedded in simulated Advanced LIGO noise with an SNR of $\varrho = 20$ and then fitted twice using the signal reconstruction model of 4.3.2 with an arbitrary $d = 10$ PCs — one using the LS basis functions, and one using the Shen. The DIC is then computed and compared to infer the best EOS sub-catalogue for each signal (used as a proxy for EOS classification). The lowest DIC shows preference to one sub-catalogue. However, using the same logic as in 4.3.3, a difference in DIC of less than 5 between models is considered to be inconclusive/unidentified.

Figure 4.11 is a histogram showing the proportions of each classification of each waveform as a function of the absolute difference in DIC. Overall, 50% of the 128 waveforms were correctly identified, 21% incorrectly identified, and 29% were inconclusive. The incorrectly identified nuclear EOS decrease proportional to the correctly identified signals as the absolute difference in DIC increases, though there are still some incorrect classifications with large differences in DIC, showing the method is not robust.

Other attempts at inferring nuclear EOS have been unsuccessful. A Bayesian logistic regression model was trialled, but yielded poorer classification than the cross-validation approach. This is not surprising as Dimmelmeier et al. (2008) and Abdikamalov et al. (2014) suggested that nuclear

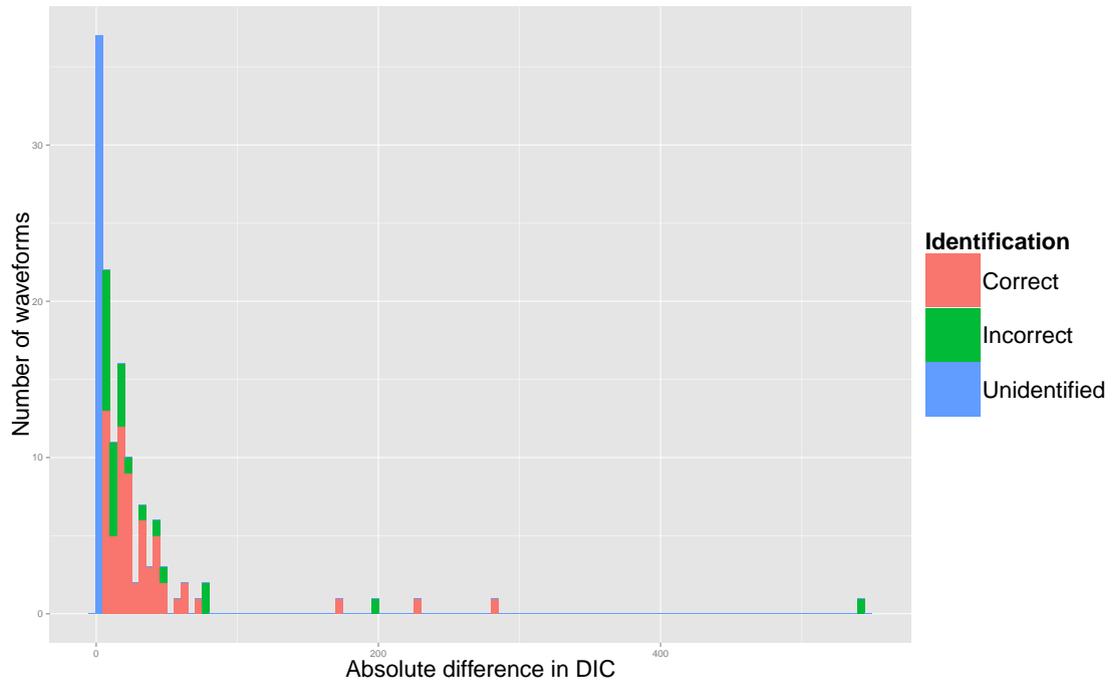


FIGURE 4.11: Histogram of nuclear EOS classification, with bin size 5.

EOS does not significantly influence the gravitational wave morphology for rotating stellar core collapse.

Chapter 5

Spectral density estimation for Advanced LIGO noise

5.1 Introduction

As with any measured signal, gravitational wave observations are subject to noise. Accurate astrophysical inferences will rely on an honest characterisation of background and detector noise sources. At its design sensitivity, Advanced LIGO will be sensitive to gravitational waves in the frequency band from ~ 10 Hz to 8 kHz. The main noise sources for ground-based interferometers include seismic noise, thermal noise, and quantum noise (Aasi et al., 2015). Seismic noise limits the low frequency sensitivity of the detectors, and is caused by phenomena such as earthquakes, ocean waves, and wind. Thermal noise is the predominant noise source in the most sensitive frequency band of Advanced LIGO (around 100 Hz), and arises from the test mass mirror suspensions and the Brownian motion of the mirror coatings. Quantum noise comes from two sources: photon shot noise and radiation pressure. Photon shot noise is due to quantum uncertainties in the detected photon arrival rate, and dominates the high frequency sensitivity of the detectors, whereas radiation pressure limits low frequency sensitivity due to momentum imparted to the mirrors upon reflection.

Standard assumptions about the noise model in the gravitational wave data analysis community rely on detector noise being stationary and Gaussian distributed, with a known PSD that is usually estimated using off-source data (not on a candidate signal) (Littenberg et al., 2013). Real gravitational wave data often depart from these assumptions (Christensen, 2010). Aasi et al. (2013) demonstrated that fluctuations in the PSD can moderately bias parameter estimates of

compact binary coalescence gravitational wave signals embedded in LIGO data from the sixth science run (S6).

High amplitude non-Gaussian transients (or “glitches”) in real detector data invalidate the Gaussian noise assumption, and misspecifications of the parametric noise model could result in misleading inferences and predictions. A more sophisticated approach would be to make no assumptions about the underlying noise distribution by using nonparametric techniques. Unlike parametric statistical models, which have a fixed and finite set of parameters (e.g., the Gaussian distribution has two parameters: μ and σ^2 representing the mean and variance respectively), nonparametric models have a potentially infinite set of parameters, allowing for much greater flexibility.

The theory of spectral density estimation requires a time series to be a stationary process. If data are not stationary (which is often the case for real LIGO data), it is important to adjust for this by introducing a time-varying PSD. Littenberg and Cornish (2015) demonstrated that the noise PSD in real S6 LIGO data is in fact time-varying. Variation in detector sensitivity was also shown by Abadie et al. (2012b). Other gravitational wave literature that discusses nonstationary noise includes Sintes and Schutz (1998) and Finn and Mukherjee (2001). It would be an oversimplification to assume the Advanced LIGO PSD is constant over time, and to use off-source data in characterising this. On-source estimation of the PSD would therefore be preferable to mitigate the issues relating to the time-varying nature of the PSD.

There have been attempts reported in the literature to improve the modelling of noise present in gravitational wave data, primarily concentrating on noise with embedded signals from well-modelled gravitational wave sources, such as binary inspirals (Littenberg and Cornish, 2010, Littenberg et al., 2013, Röver, 2011, Röver et al., 2011, Vitale et al., 2014), and more recently from gravitational wave *bursts* (Cornish and Littenberg, 2015, Littenberg and Cornish, 2015).

Under the Bayesian framework, Röver et al. (2011) used a Student- t likelihood as a generalisation to the commonly used Whittle (approximate Gaussian) likelihood (Whittle, 1957). The benefit of the Student- t set-up is two-fold: uncertainty in the noise spectrum can be accounted for via marginalisation of nuisance parameters; and outliers can be accommodated due to the heavy-tail nature of the Student- t probability density. A drawback of this method is that the choice of hyperparameters can unduly influence posterior inferences.

Using the maximum likelihood approach, Röver (2011) later demonstrated that the Student- t likelihood could be used as a generalisation to the matched-filtering detection method commonly used in the analysis of gravitational wave signals from well-modelled sources. This approach would not be appropriate for gravitational wave bursts, since matched-filtering requires accurate signal models with well-defined parameter spaces.

Littenberg and Cornish (2010) used Bayesian model selection to determine the best noise likelihood function in non-Gaussian noise. The authors considered Gaussian noise with a time-varying mean, noise from a weighted sum of two Gaussian distributions (non-Gaussian tails), and a combination of Gaussian noise and glitches (modelled as a linear combination of wavelets).

Littenberg et al. (2013) demonstrated how one can incorporate additional scale parameters in the Gaussian likelihood, and marginalise over the uncertainty in the PSD to reduce systematic biases in parameter estimates of compact binary mergers in S5 LIGO data. This method requires an initial estimate of the PSD. On a related note, Vitale et al. (2014) highlighted a Bayesian method, similar to iteratively reweighted least squares, that analytically marginalises out background noise and requires no *a priori* knowledge of the PSD. They applied this to simulated data from LISA Pathfinder.

More recently, Littenberg and Cornish (2015) introduced the BayesLine algorithm in conjunction with BayesWave (Cornish and Littenberg, 2015) to estimate the underlying PSD of gravitational wave detector noise. BayesLine is used to model the Gaussian noise component. The authors use a cubic spline to model the smooth changing broadband noise and Lorentzians (Cauchy densities) to model wandering spectral lines (due to AC supply, violin modes, etc.). BayesWave, on the other hand, models the non-Gaussian instrument “glitches” and burst sources with a continuous wavelet basis. Both methods make use of the trans-dimensional RJMCMC algorithm of Green (1995).

In this chapter, the Bayesian nonparametric spectral density estimation results from Edwards et al. (2015), Edwards et al. (2017a), and Kirch et al. (2017) are presented as they apply to Advanced LIGO detector and background noise. In the first study (Edwards et al., 2015), a general framework for improving gravitational wave detector noise spectrum modelling is considered. A Bernstein polynomial prior is used to estimate the PSD of simulated Advanced LIGO noise. A blocked Gibbs sampler is employed to simultaneously estimate the noise PSD, and reconstruct a rotating core collapse supernova gravitational wave signal embedded in this noise. Nonstationary data are modelled by breaking the series into smaller, locally stationary components. The second study (Edwards et al., 2017a) introduces the B-spline prior for spectral density estimation, and compares this to the Bernstein polynomial prior. It was found that due to its slow convergence properties, the Bernstein polynomial prior was not adequate in modelling real detector noise, which has a complicated noise structure, with many *spectral lines* (sharp and abrupt peaks in power at particular frequencies). The B-spline prior aims to ameliorate this. The third study (Kirch et al., 2017) then presents results using a nonparametric correction to a parametric likelihood. This approach can be considered a generalisation to the Whittle likelihood, which is ubiquitous in Bayesian nonparametric time series analysis, as well as gravitational wave data analysis.

5.2 Advanced LIGO spectral density estimation using the Bernstein polynomial prior

In this section, methods and results from Edwards et al. (2015) are presented. A framework for modelling Advanced LIGO background and detector noise is provided, improving on the default noise model used in the gravitational wave data analysis literature and in practice. Recall that the default noise model assumes noise is stationary and Gaussian, with a known PSD. The framework presented here addresses each of these issues. That is, the models here can handle nonstationary and non-Gaussian noise, and do not assume a PSD is known before parameter estimation procedures are conducted. Further, the approach is used to characterise noise that has been added to stellar core collapse gravitational wave signals.

The Bayesian nonparametric spectral density estimation method and Metropolis-within-Gibbs MCMC sampler of Choudhuri et al. (2004) is implemented. Pseudo-posterior samples are obtained by placing a nonparametric Bernstein polynomial prior (finite mixture of beta probability densities) (Petrone, 1999a,b) on the spectral density and updating this using the Whittle likelihood. Note that these are pseudo-posterior samples as the Whittle likelihood is only an approximation to the true Gaussian likelihood, but is much faster to evaluate. Choudhuri et al. (2004) proved that this method yields a consistent estimator for the true spectral density of a (short memory) stationary time series, an attractive feature absent in the periodogram. Posterior consistency in this context essentially means that the posterior probability of an arbitrary neighbourhood around the true PSD goes to 1 as the length of the time series increases to infinity. Thus, as the sample size increases, the posterior distribution of the PSD will eventually concentrate in a neighbourhood of the true PSD (Hjort et al., 2010). This is an important asymptotic robustness quality of the posterior distribution in that the choice of prior parameters should not influence the posterior distribution too much. Because of the high dimensional parameter spaces in Bayesian nonparametrics, many posterior distributions do not automatically possess this quality (Hjort et al., 2010).

Unlike the methods of Röver et al. (2011), Littenberg et al. (2013), and Vitale et al. (2014), noise is not treated as a nuisance parameter to be analytically integrated out. Although the signal parameters are of primary interest, it is also important to quantify one’s uncertainty in the underlying PSD of the noise (in terms of posterior probabilities and credible intervals), providing reliable astrophysical statements.

Assume the data are the sum of a gravitational wave signal and additive noise (from all noise sources), such that

$$\mathbf{y} = \mathbf{s}(\boldsymbol{\alpha}) + \boldsymbol{\epsilon}(\boldsymbol{\theta}), \tag{5.1}$$

where \mathbf{y} is the (coincident) time domain gravitational wave data vector, \mathbf{s} is a gravitational wave signal parameterised by $\boldsymbol{\alpha}$, and $\boldsymbol{\epsilon}$ is noise parameterised by $\boldsymbol{\theta}$. Notation with a tilde on top, such as $\tilde{\mathbf{y}}$, refers to the frequency domain equivalent of the same quantity, obtained by the DFT. Noise in this set-up is treated as the conglomeration of detector noise (such as thermal noise and photon shot noise), background noise (such as seismic noise), and residual errors due to parametric statistical modelling of gravitational wave signals. An important caveat is to ensure the parametric signal model is a good fit to the signal so that residual statistical errors do not dominate the noise. Estimation of spectral lines (sharp and abrupt spectral peaks) are left out of scope of this analysis due to slow convergence of Bernstein polynomials to complicated functions. This will however be addressed in 5.3 and 5.4.

Under this framework, a gravitational wave signal could essentially come from any source as long as a reasonable parametric model exists. To simplify the problem and to demonstrate the novelty of the method, attention is restricted to gravitational waves from rotating core collapse supernovae. Using the gravitational waveform catalogue of Abdikamalov et al. (2014), a PCA is conducted, and a PCR model of the most important principal components is then fitted on an arbitrary rotating core collapse gravitational wave signal (Edwards et al., 2014, Heng, 2009, Röver et al., 2009). This is the same signal reconstruction method presented in 4.3.1.

The (parametric) signal component is easily embedded as an additional Gibbs step in the Metropolis-within-Gibbs MCMC sampler of Choudhuri et al. (2004). A *blocked* Gibbs sampler is thus used to sequentially sample the signal parameters $\boldsymbol{\alpha}$ given the noise parameters $\boldsymbol{\theta}$, and vice versa. As the model now contains a parametric signal component as well as a nonparametric noise component, it can be considered a *semiparametric* model.

To accommodate for nonstationary noise, an idea presented by Rosen et al. (2012) is adapted. A nonstationary time series can be broken down into smaller *locally stationary* segments. For each segment, the PSD is estimated using the method of Choudhuri et al. (2004), and the resulting time-varying spectrum is analysed.

This work will complement existing methods in the gravitational wave literature, with the following benefits:

- A Bayesian framework, allowing one to update prior knowledge based on observed data, as well as quantifying uncertainty in terms of probabilistic statements;
- Posterior consistency of the PSD, i.e., the posterior distribution will concentrate around the true PSD as the sample size increases;
- High amplitude non-Gaussian transients in the noise can be handled due to the nonparametric nature of the method;

- Nonstationarities can be taken into account by splitting the data into smaller locally stationary segments;
- Estimation of noise and signal parameters are done simultaneously using Gibbs sampling;
- Uncertainty in astrophysically meaningful parameter estimates are honest, with less systematic biases;
- Noninformative priors can be chosen, and the PSD does not need to be known *a priori*;
- Useful for any signal with a parametric statistical model (including rotating core collapse supernova gravitational waves).

5.2.1 PSD estimation

Recall the Bernstein polynomial prior for spectral density estimation described in 3.3.3 and 3.5.4.1. The spectral density of a stationary time series is modelled with parameters $\boldsymbol{\theta} = (\mathbf{v}, \mathbf{z}, k, \tau)$. This method is implemented using the Metropolis-within-Gibbs MCMC sampler proposed by Choudhuri et al. (2004). The stick-breaking truncation parameter is chosen to be $L = \max\{20, n^{1/3}\}$. Parameters k and τ are readily sampled from their full conditional posteriors, while \mathbf{V} and \mathbf{Z} require the Metropolis algorithm with uniform proposals. The only variation on this implementation is the sampling of the smoothness parameter k . A Metropolis step is faster than sampling from the full conditional here. The original implementation of Choudhuri et al. (2004) contains a for loop that evaluates the log posterior k_{\max} times, where k_{\max} is chosen (during pilot runs) to be large enough to cater for the roughness of the PSD. For most well-behaved cases, $k_{\max} = 50$ will suffice, but the Advanced LIGO PSD requires many more mixture components (by up to two or more orders of magnitude) due its steepness at low frequencies. This is a significant computational burden, and a well-designed Metropolis step can therefore outperform the original implementation.

5.2.2 Signal reconstruction

To reconstruct a rotating core collapse gravitational wave signal that is embedded in noise, the method described by Heng (2009), Röver et al. (2009), and Edwards et al. (2014), and presented in 4.3.1 is employed. That is, let

$$\tilde{\mathbf{y}} = \tilde{X}\boldsymbol{\alpha} + \tilde{\boldsymbol{\epsilon}}, \quad (5.2)$$

where $\tilde{\mathbf{y}}$ is the length n frequency domain gravitational wave data vector (signal plus noise), \tilde{X} is the $n \times d$ matrix of the d frequency domain principal component basis vectors, $\boldsymbol{\alpha}$ is the vector

of signal reconstruction parameters (PC coefficients), and $\tilde{\boldsymbol{\epsilon}}$ is the frequency domain noise vector with a *known* PSD. Flat priors on $\boldsymbol{\alpha}$ are assumed.

An additional Gibbs step is included in the MCMC sampler described in 3.3.3, 3.5.4.1, and 5.2.1 to simultaneously reconstruct a rotating core collapse gravitational wave signal, whilst also estimating the noise power spectrum. Omitting the conditioning on the data for clarity, the algorithm sequentially samples the full set of conditional posterior densities $p(\boldsymbol{\theta}|\boldsymbol{\alpha})$ and $p(\boldsymbol{\alpha}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\mathbf{v}, \mathbf{z}, k, \tau)$ are the noise parameters defined in 3.3.3 and 3.5.4.1, and $\boldsymbol{\alpha}$ are the signal reconstruction parameters. That is, sample in a cycle from the full conditional posterior distribution of the signal parameters given the PSD parameters, and the full conditionals of the PSD parameters given the signal parameters. This set-up is a *blocked* Gibbs sampler.

To sample the signal parameters, fix the most recent MCMC sample of the PSD parameters. The conditional posterior of $\boldsymbol{\alpha}$ is

$$\boldsymbol{\alpha}|\boldsymbol{\theta} \sim \text{N}(\boldsymbol{\mu}, \Sigma) \quad (5.3)$$

where $\Sigma = (\tilde{X}'D^{-1}\tilde{X})^{-1}$ and $\boldsymbol{\mu} = \Sigma\tilde{X}'D^{-1}\tilde{\mathbf{y}}$. Here $D = 2\pi \times \text{diag}(f(\boldsymbol{\lambda}))$ is the noise covariance matrix, and $f(\boldsymbol{\lambda})$ is the most recent estimate of the PSD. More explicitly, at iteration $i + 1$ in the blocked Gibbs sampling algorithm:

1. Create time domain noise vector: $\boldsymbol{\epsilon}^{(i+1)} = \mathbf{y} - X\boldsymbol{\alpha}^{(i)}$. Due to the linearity of the Fourier transform, $\boldsymbol{\alpha}$ will be the same regardless of whether the time domain or frequency domain is used.
2. Sample the PSD parameters $\boldsymbol{\theta}^{(i+1)}|\boldsymbol{\alpha}^{(i)}$ using the method described in 3.3.3 and 3.5.4.1.
3. Sample the signal parameters $\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\theta}^{(i+1)}$ using Equation (5.3) (since the PSD in iteration $i + 1$ is now known).

5.2.3 Nonstationary noise

Stationary noise has a constant and finite mean and variance over time, and an autocovariance function that depends only on the time lag. Nonstationary noise does not meet these requirements, and has a time-varying spectrum. Stationarity of a time series can be tested using classical hypothesis tests such as the Augmented Dickey-Fuller test (Said and Dickey, 1984), Phillips-Perron unit root test (Phillips and Perron, 1988), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992).

To accommodate nonstationary noise into the current framework, an idea presented by Rosen et al. (2012) is adapted, assuming a time series can be broken down into locally stationary segments. In

their paper, they treated the number of stationary components of a nonstationary time series as unknown, and used RJMCMC (Green, 1995) to estimate the segment breaks.

In a similar fashion, this analysis assumes a nonstationary time series (or gravitational wave data stream) is split into J mutually exclusive, equal-length segments. There are two requirements for the length of these segments: the segment length is large enough for the Whittle approximation to be valid; and the segments are locally stationary according to heuristics or formal stationarity hypothesis tests. This approach moulds well into the current MCMC framework. For each segment, the PSD is estimated using the nonparametric method introduced in 3.3.3 and 3.5.4.1. A benefit of this approach is that change-points in the PSD can be detected without using RJMCMC. Note also that J is fixed in this analysis.

The conditional posterior density for all noise model parameters $\boldsymbol{\theta}$ is the following product

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \tilde{\mathbf{y}}) = \prod_{j=1}^J p_j(\boldsymbol{\theta}_j|\boldsymbol{\alpha}, \tilde{\mathbf{y}}_j), \quad (5.4)$$

where $p_j(\boldsymbol{\theta}_j|\boldsymbol{\alpha}, \tilde{\mathbf{y}}_j)$ is the conditional posterior density of the model parameters $\boldsymbol{\theta}_j$ in the j^{th} segment given the signal parameters $\boldsymbol{\alpha}$ and the j^{th} segment of data $\tilde{\mathbf{y}}_j$.

Note that under this set-up, the PC coefficients $\boldsymbol{\alpha}$ do not depend on segments $j = 1, 2, \dots, J$, as only one set of PC coefficients (not J sets) is required to reconstruct a rotating core collapse gravitational wave signal.

To sample $\boldsymbol{\alpha}|\boldsymbol{\theta}$, the same approach presented in 5.2.2 is used. The only difference is in the construction of the noise covariance matrix. This is constructed as

$$D = 2\pi \times \text{diag}(f_1(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}), \dots, f_J(\boldsymbol{\lambda})), \quad (5.5)$$

where $f_j(\boldsymbol{\lambda})$ is the PSD of the j^{th} noise segment.

5.2.4 Results

For all of the following examples, the stick-breaking truncation parameter is set to $L = \max\{20, n^{1/3}\}$. The same noninformative prior set-up of Choudhuri et al. (2004) is used. That is,

$$\begin{aligned} G_0 &= \text{Uniform}[0, 1], \\ M &= 1, \\ \alpha_\tau &= 0.001, \\ \beta_\tau &= 0.001, \\ \theta_k &= 0.01. \end{aligned}$$

For most examples, $k_{\max} = 50$. However, $k_{\max} = 400$ for the example with simulated Advanced LIGO noise to cater for the steep drop in the PSD at low frequencies.

For the examples with a signal embedded in noise, a $\text{Uniform}(-\infty, \infty)$ prior is placed on the signal reconstruction parameters $\boldsymbol{\alpha}$, and let $d = 25$ PCs. The signals are scaled to an SNR of $\varrho = 50$.

The value of $\varrho = 50$ is physically motivated, as one would expect to see an SNR of approximately 50 to 170 for rotating core collapse supernova gravitational waves at a distance of 10 kpc. The analysis will demonstrate robustness for the lower end of this range.

The units for frequency in most examples are radians per second (rad/s). In the example using simulated Advanced LIGO noise, this is rescaled to kilohertz (kHz). PSD units are the inverse of the frequency units, and the PSD figures are scaled logarithmically. Gravitational wave strain amplitude is unitless.

For all examples, the MCMC sampler is run for 150,000 iterations, with a burn-in period of 50,000 and thinning factor of 10. This results in 10,000 samples retained.

5.2.4.1 Estimating the PSD of non-Gaussian coloured noise

To demonstrate how the noise-only model of 3.3.3, 3.5.4.1 and 5.2.1 is capable of dealing with non-Gaussian transients in the data (or *glitches* as they are sometimes called in the gravitational wave data analysis literature), an illustrative toy example is provided, using coloured noise generated from a first order autoregressive process, abbreviated to AR(1).

A mean-centred AR(1) process $\{Y_t\}$ is defined as

$$Y_t = \rho Y_{t-1} + \epsilon_t, \quad t = 1, 2, \dots, n, \quad (5.6)$$

where ρ is the first order autocorrelation, and $\{\epsilon_t\}$ is a white noise process (not necessarily Gaussian), with zero mean and constant variance σ_ϵ^2 . With this formulation, it is easy to see how the current observation at time t depends on the previous observation at time $t - 1$ through ρ , as well as some white noise ϵ_t , often referred to as *innovations* or the *innovation process* in time series literature.

The AR(1) model is a useful example here since it has a well-defined theoretical spectral density that one can compare against the MC estimates of the (stationary) noise-only PSD estimation algorithm. Assuming $|\rho| < 1$, the AR(1) process is stationary and has spectral density

$$f(\lambda) = \frac{\sigma_\epsilon^2}{1 + \rho^2 - 2\rho \cos \lambda}, \quad \lambda \in (-\pi, \pi]. \quad (5.7)$$

The AR(1) process has a PSD that is not flat, and the noise in the toy example is coloured (non-white), with non-zero autocorrelation — typical of what would be expected with real Advanced LIGO noise. As the AR(1) process has a coloured spectrum, and white noise has a flat spectrum, the term *innovations* will be used to refer to the white noise component of the model to avoid confusion.

For this example, rather than using Gaussian innovations, which is the most common innovation process used in autoregressive models, Student- t innovations with $\nu_{\text{df}} = 3$ degrees of freedom are used. The choice of $\nu_{\text{df}} = 3$ degrees of freedom is the smallest integer that results in a Student- t model with finite variance (a requirement for the innovation process $\{\epsilon_t\}$ of an AR(1) model). This model has wider tails than that of the Gaussian model (and in fact the widest tails possible whilst maintaining the finite variance requirement), meaning extreme values in the tails of the distribution are expected to occur more often. This will be used as a proxy for glitches.

For this example, a length $n = 2^{12}$ AR(1) process with $\rho = -0.9$ and Student- t innovations with $\nu_{\text{df}} = 3$ degrees of freedom is generated. Let this (stationary) time series have sampling interval $\Delta_t = 1/2^{14}$ (the same as Advanced LIGO). The data set-up can be seen in Figure 5.1.

The effect of using $\nu_{\text{df}} = 3$ degrees of freedom can be seen in Figure 5.1. There are transient, high amplitude, non-Gaussian events. These are a result of the wide-tailed nature of the Student- t density. It would be very unlikely to see these high amplitude events if the innovation process was Gaussian.

The (stationary) noise-only algorithm of 3.3.3, 3.5.4.1 and 5.2.1 is used to demonstrate that a PSD from coloured non-Gaussian noise can be accurately characterised. This can be seen in Figure 5.2.

The estimated pointwise posterior median log PSD in Figure 5.2 is very close to the true log PSD, and the 90% credible region generally contains the true log PSD. This demonstrates that

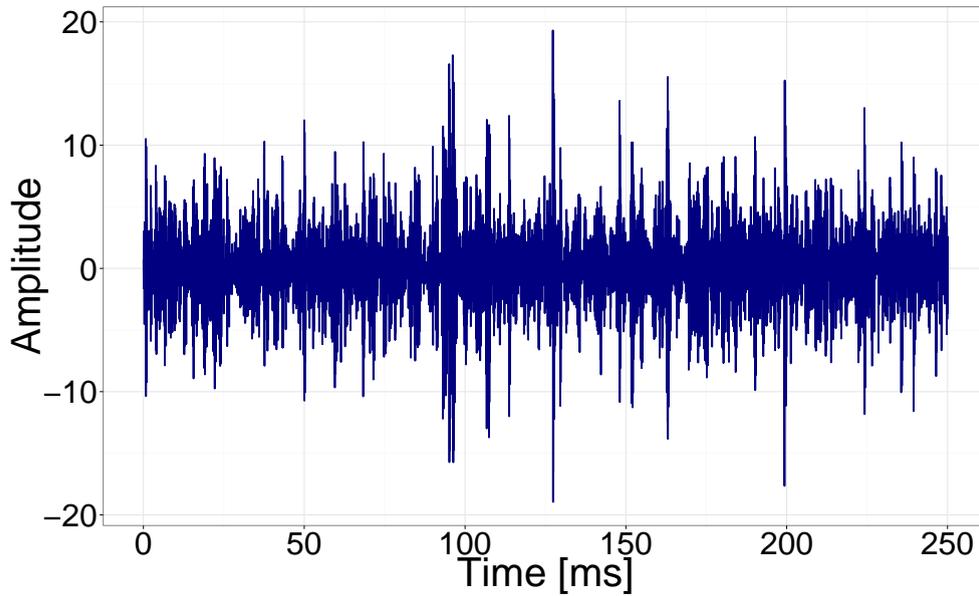


FIGURE 5.1: Simulated stationary AR(1) process with first-order autocorrelation $\rho = -0.9$, and Student- t innovations ($\nu_{df} = 3$ degrees of freedom).

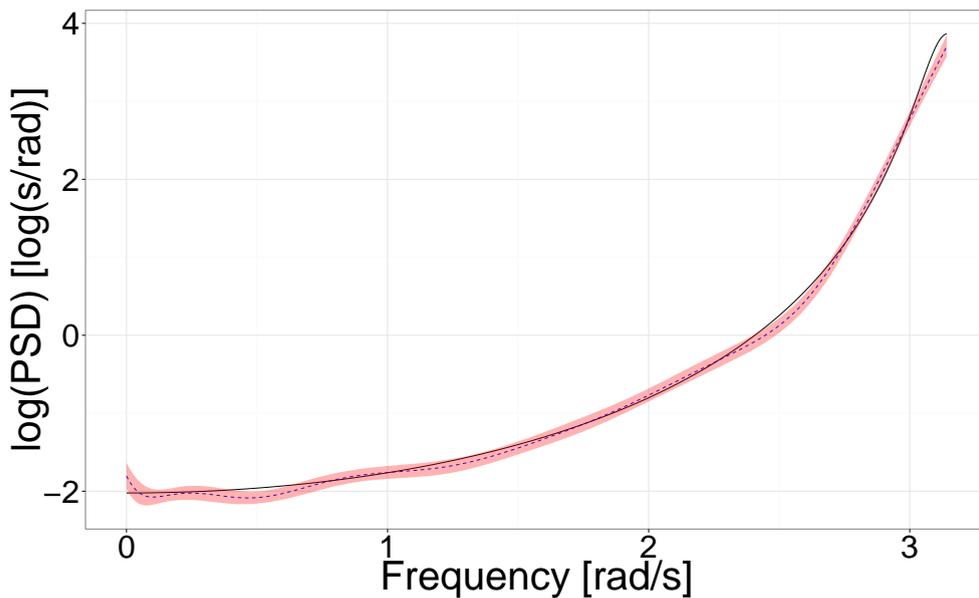


FIGURE 5.2: Estimated log PSD of the AR(1) time series in Figure 5.1. 90% credible region (shaded pink) and posterior median log PSD (dashed blue), superimposed with the true log PSD (solid black).

even if there are non-Gaussian transients in the data (which is certainly the case for real LIGO data), this PSD estimation method performs well. This, however, is not surprising as the Whittle likelihood gives a good approximation to Gaussian and some non-Gaussian (and non-linear) time series (Shao and Wu, 2007).

5.2.4.2 Extracting a rotating core collapse signal in stationary coloured noise

In this example, the goal is to extract a rotating core collapse gravitational wave signal from noisy data using the blocked Gibbs sampler described in 5.2.2. The *A1O12.25* rotating core collapse gravitational wave signal from the Abdikamalov et al. (2014) test set (i.e., a signal not part of the training set used to create the PC basis functions) is embedded in AR(1) noise with $\rho = 0.9$. For clarity, let this process have a Gaussian white noise innovation process with $\sigma_\epsilon^2 = 1$. Let the time series be length $n = 2^{12}$, which corresponds to 1/4 s of data at the Advanced LIGO sampling rate. The signal is scaled to have a SNR of $\varrho = 50$. The reconstructed signal can be seen in Figure 5.3.

The rotating core collapse gravitational wave signal in Figure 5.3 is reconstructed particularly well during the collapse and bounce phases (the first few peaks/troughs). The post-bounce ring-down oscillations are usually poorly estimated due to stochastic dynamics (Abdikamalov et al., 2014, Edwards et al., 2014), but are acceptable for this particular example.

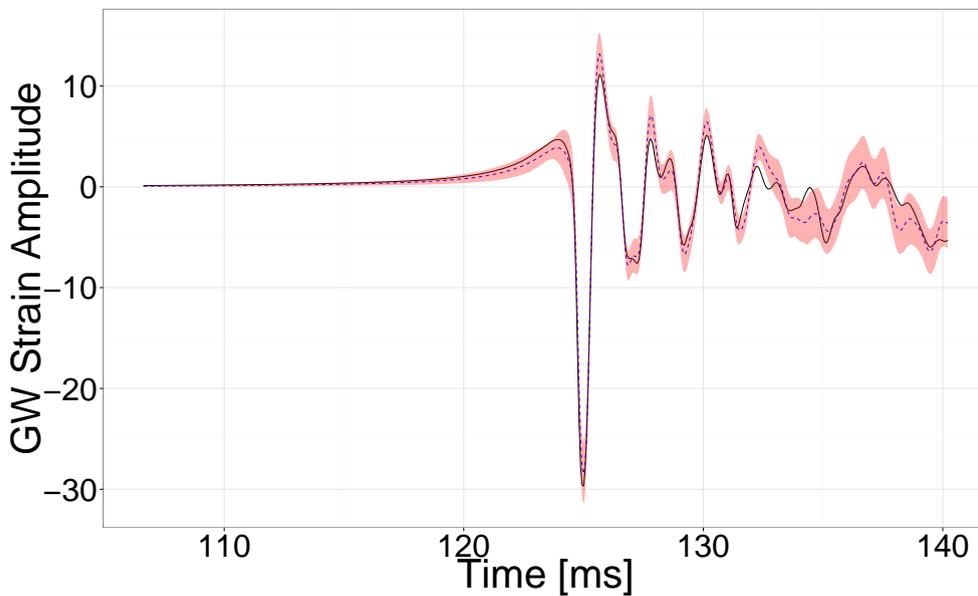


FIGURE 5.3: Reconstructed rotating core collapse gravitational wave signal. 90% credible region (shaded pink) and posterior median signal (dashed blue), superimposed with true *A1O12.25* gravitational wave signal from the Abdikamalov et al. (2014) test set (solid black).

In this example, the signal parameters were simultaneously estimated with the noise PSD using the blocked Gibbs sampler described in 5.2.2. The performance of the estimated noise PSD with and without a signal present is now compared to illustrate that it is possible to simultaneously reconstruct a signal, while not compromising the PSD estimation step. That is, compare the noise PSD estimates between the algorithms presented in 5.2.1 (noise-only model) and 5.2.2 (signal-plus-noise model), using the same noise series for both models.

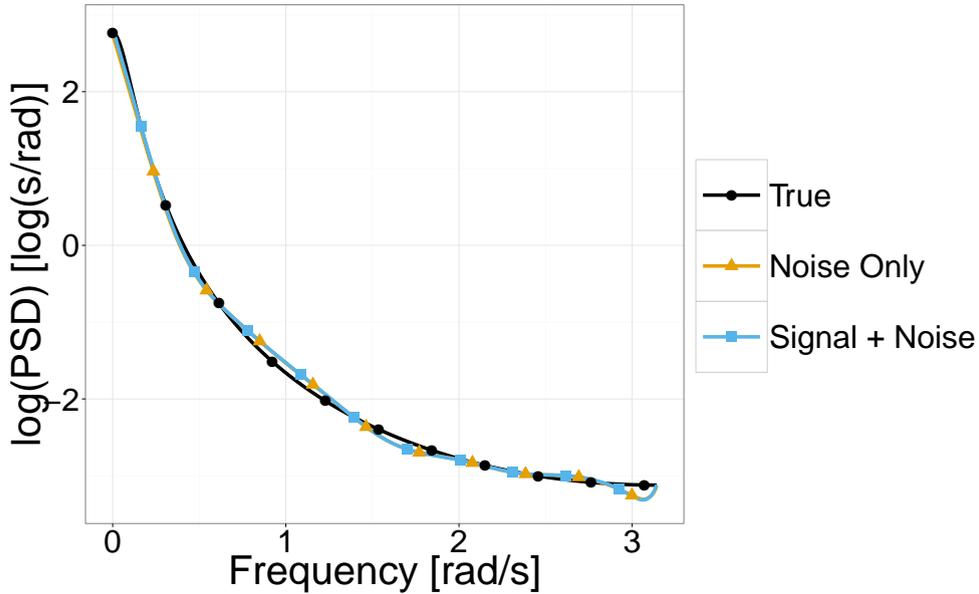


FIGURE 5.4: Comparison of the noise PSD estimates for the noise-only and signal-plus-noise models. Plotted are the pointwise posterior median log noise PSDs with and without a gravitational wave signal. The true log PSD of the AR(1) noise series is overlaid.

Both models (noise-only and signal-plus-noise) in Figure 5.4 perform comparably and look near identical when estimating the PSD of coloured Gaussian noise. The posterior median log PSDs are approximately equal, and are very close to the true log PSD of an AR(1) process with $\rho = 0.9$. This is an important robustness check, and demonstrates that there is a successful decoupling of the signal from the noise.

5.2.4.3 Comparing input and reconstruction parameters

As there is no closed-form expression linking the astrophysical parameters of a rotating core collapse stellar event to its gravitational wave signal, one can only approximate the gravitational wave signal using statistical methods. This is done using PCR (refer to 4.3.1 and 4.3.2), but there are no true input parameters to compare the estimated signal reconstruction parameters against. However, if one were to create a fictitious signal as a known linear combination of PCs, it would be possible to demonstrate the algorithm’s performance in estimating the signal reconstruction parameters.

Consider the following fictitious rotating core collapse gravitational wave signal

$$\mathbf{y} = \sum_{i=1}^d \alpha_i \mathbf{x}_i, \quad (5.8)$$

where \mathbf{y} is the length n signal, $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d)$ are the d PC basis vectors of length n , and $(\alpha_1, \alpha_2, \dots, \alpha_d)$ are the “true” weights, or PC coefficients. To randomise the weights, randomly sample each from the standard normal distribution.

In this example, the fictitious gravitational wave signal of length $n = 2^{12}$ is embedded in AR(1) noise with $\rho = 0.9$ and Gaussian innovations with $\sigma_\epsilon = 1$, setting $d = 10$. The signal is rescaled to have SNR $\rho = 50$, and after the algorithm has run, the estimated PC coefficients are rescaled back to the original scale for comparison. Results are shown in 5.5.

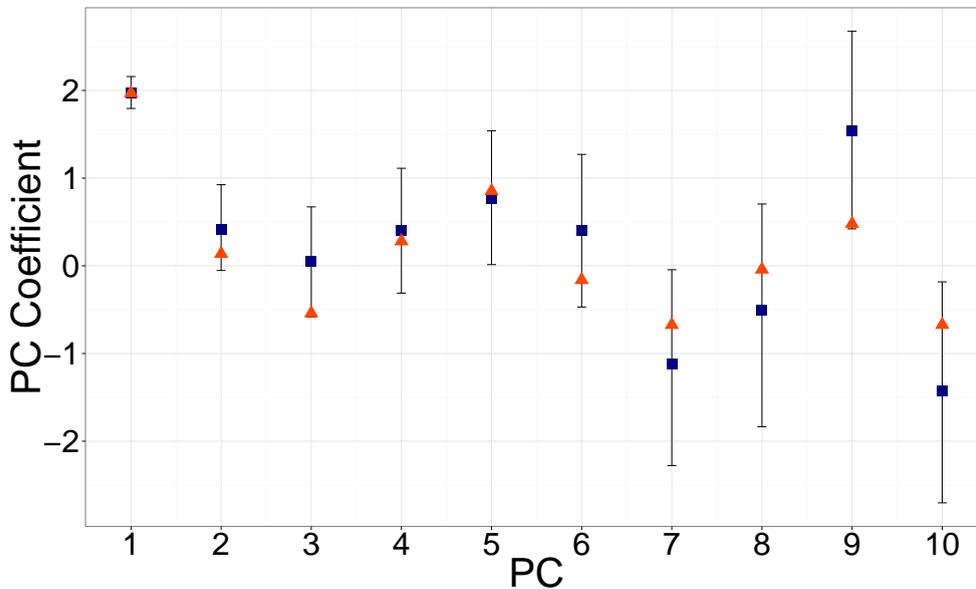


FIGURE 5.5: Posterior median PC coefficients (blue square) and “true” PC coefficients (orange triangle) for the 10 PCs of a fictitious gravitational wave signal embedded in AR(1) noise. The error bands are the 95% credible intervals.

It can be seen in Figure 5.5 that the “true” PC coefficients are generally contained within the 95% credible intervals, demonstrating that the algorithm can estimate a signal’s input parameters well in the presence of stationary coloured noise. Notice also that the credible intervals widen as the principal component number increases. This is due to higher numbered PCs explaining lower amounts of variation in the waveform catalogue, resulting in lower amplitude basis functions, increasing the uncertainty of these PCs embedded in noise.

5.2.4.4 Extracting a rotating core collapse signal in time-varying coloured noise

Nonstationary noise has a time-varying spectrum. To illustrate how the methods presented in this section can handle nonstationarities (or change-points in the spectral structure), a noise series with $J = 2$ locally stationary components of equal length $n_1 = n_2 = 2^{12}$ is simulated. The first segment of the noise series is generated from an AR(1) process with $\rho = 0.5$. The second noise

segment comes from an AR(1) process with $\rho = -0.75$. Both segments use a Gaussian innovation process with variance $\sigma_\epsilon^2 = 1$ for clarity. Part of the *A1O8.25* waveform from the Abdikamalov et al. (2014) catalogue is embedded in the generated noise. This waveform comes from the test set, not included in the construction of PC basis functions. The data set-up can be seen in Figure 5.6.

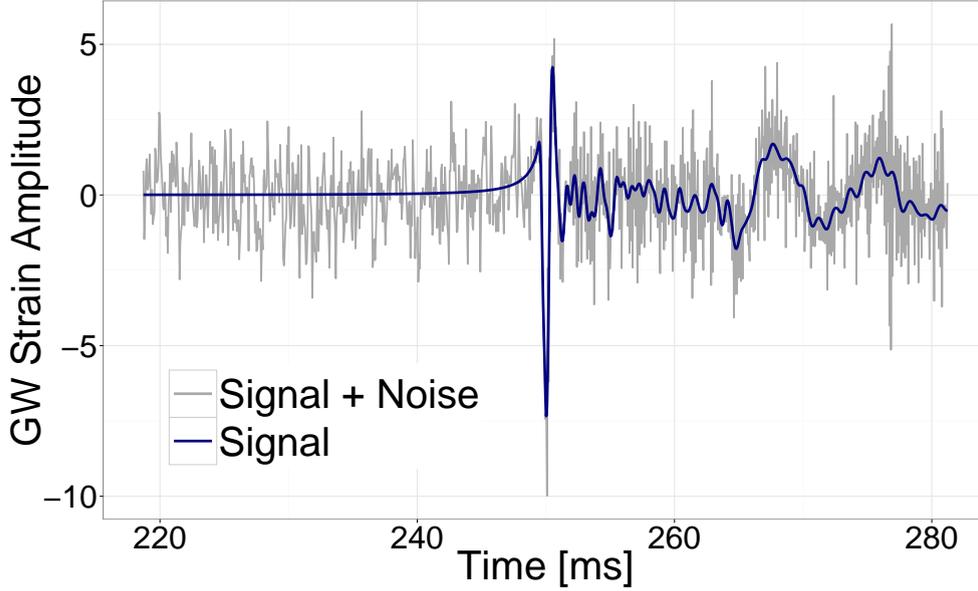


FIGURE 5.6: A snapshot of the signal superimposed on signal plus noise. The noise series has length $n = n_1 + n_2 = 2^{13}$ and is segmented into two equal parts. The first half of the noise is generated from an AR(1) with $\rho = 0.5$, and the second half is generated from an AR(1) with $\rho = -0.75$. Both segments use a Gaussian innovation process with variance $\sigma_\epsilon^2 = 1$. The *A1O8.25* rotating core collapse gravitational wave signal from the Abdikamalov et al. (2014) test set is embedded in this noise with a SNR of $\varrho = 50$.

The aim here is to simultaneously estimate both noise PSDs, as well as reconstructing the embedded gravitational wave signal using the method described in 5.2.3. Assume the change-point between the two noise series is known, though it will be demonstrated in the next section that the method can locate unknown change-points (the accuracy of which depends on the time resolution).

The difference between the first half of the noise series is compared with the second half. Each segment has a different dependence structure, and is therefore coloured differently in the frequency-domain. This results in a different time-domain morphology. Estimates of the noise PSDs can be seen in Figures 5.7 and 5.8.

Figures 5.7 and 5.8 show the estimated log PSDs for the two noise segments. The pointwise posterior median log PSDs are close to the true log PSDs, and the 90% credible regions for both segments mostly contain the true log PSDs, but veer slightly off towards the low frequencies. Due to posterior consistency of the PSD, these estimates will only get better as the sample size increases. Slight imperfections in the PSD estimates may not be such a problem if the embedded

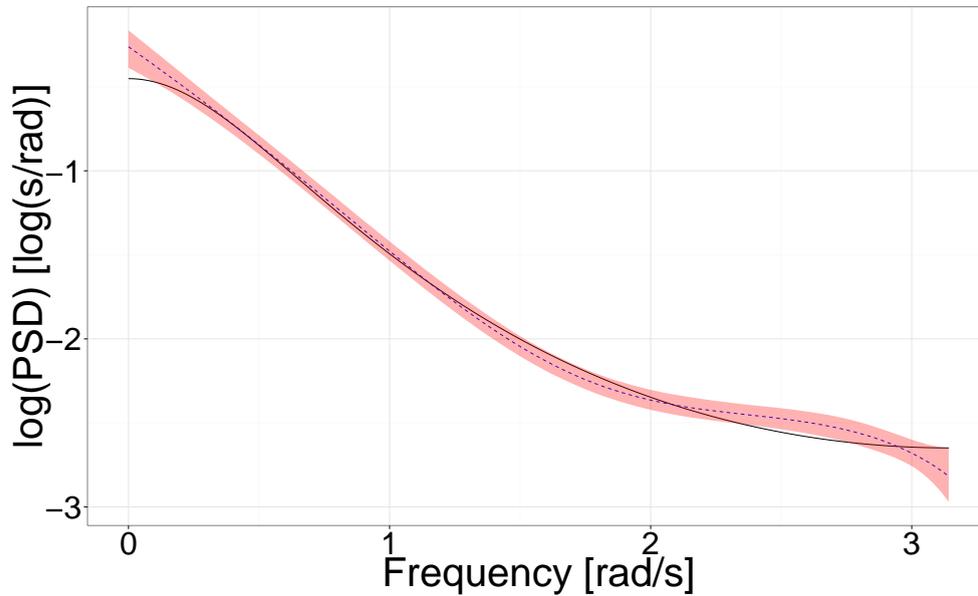


FIGURE 5.7: Spectral density estimate of the first noise segment ($\rho = 0.5$) from Figure 5.6. 90% credible region (shaded pink), posterior median log PSD (dashed blue), and theoretical log PSD (solid black).

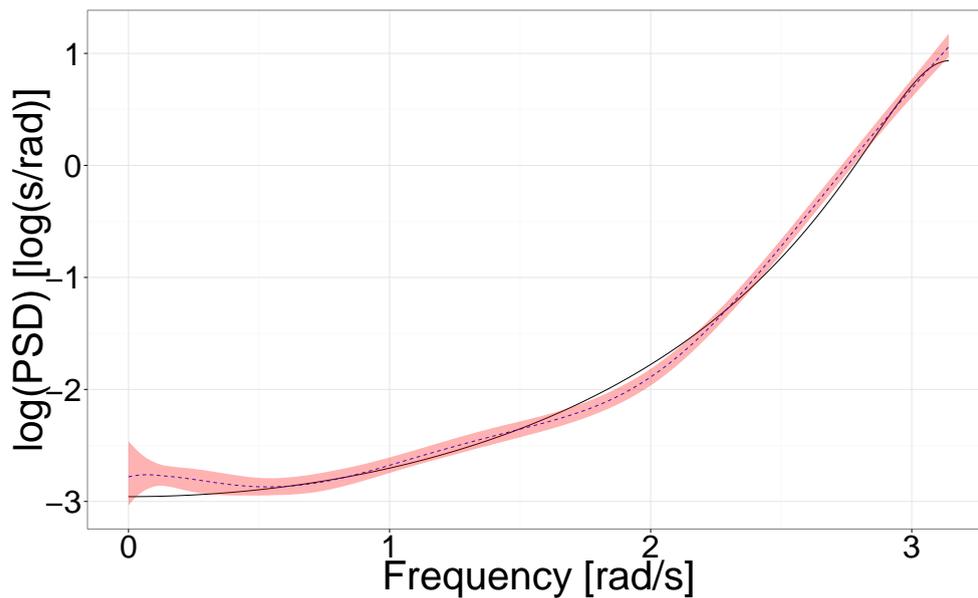


FIGURE 5.8: Spectral density estimate of the second noise segment ($\rho = -0.75$) from Figure 5.6. 90% credible region (shaded pink), posterior median log PSD (dashed blue), and theoretical log PSD (solid black).

gravitational wave signal is extracted well, which happens to be the case in this example. The extracted signal can be seen in Figure 5.9.

The 90% credible region for the reconstructed gravitational wave signal in Figure 5.9 generally contains the true signal, and has performed particularly well during collapse and bounce. Again, the post-bounce ring-down oscillations usually have the poorest reconstruction through the time

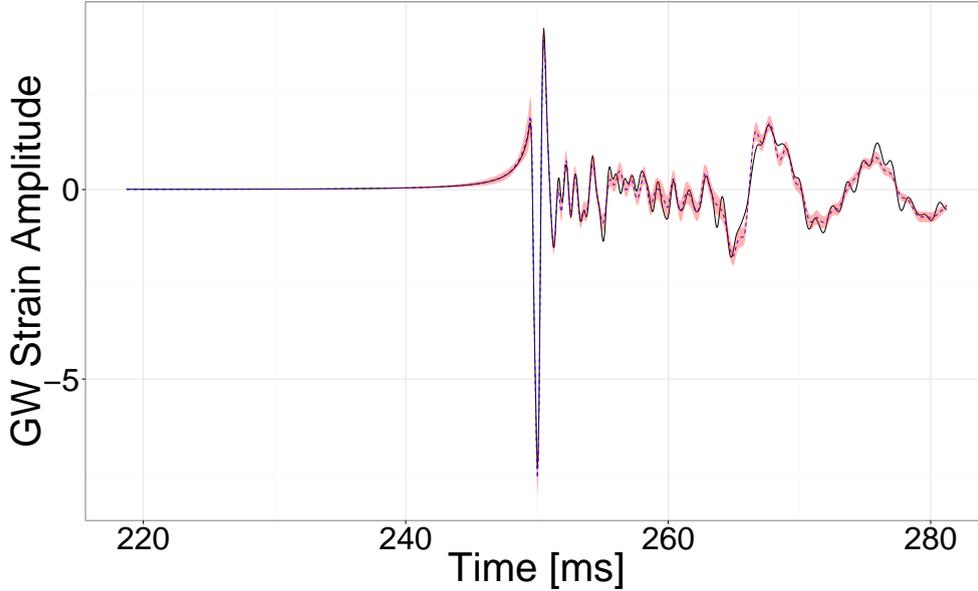


FIGURE 5.9: Reconstructed rotating core collapse gravitational wave. 90% credible region (shaded pink) and posterior median signal (dashed blue), superimposed with true *A1O8.25* gravitational wave signal from the Abdikamalov et al. (2014) test catalogue (solid black). The first half of the signal was embedded in AR(1) noise with $\rho = 0.5$, and the second half had AR(1) noise with $\rho = -0.75$. Both noise segments had Gaussian innovations with $\sigma_\epsilon^2 = 1$.

series, but has performed remarkably well in this example, regardless of the slight imperfections of the PSD estimates.

5.2.4.5 Detecting a spectral change-point

Consider a change-point problem similar to that of the previous section, where a time series exhibits a change in its spectral structure somewhere in the series. A valuable consequence of the algorithm presented in 5.2.3 is its ability to detect change-points regardless of whether the change-point occurs within a segment or on the boundary. For the following examples, let $n = 2^{12}$ and break this into $J = 32$ equal length segments ($n_j = 2^7$). For clarity, assume the time series does not contain an embedded gravitational wave signal.

First consider the case where the change-point occurs on the boundary of two noise series. Let $n_1^* = n_2^* = 2^{11}$ be the lengths of each noise series (noting that each series contains 16 segments of length $n_j = 2^7$), and let the first half of the time series be generated from an AR(1) with $\rho = 0.5$, and the second half from an AR(1) with $\rho = -0.75$. Both AR(1) processes have additive Gaussian innovations with $\sigma_\epsilon^2 = 1$. In this example, the change-point occurs exactly halfway through the series. Figure 5.10 shows a time-frequency map of the estimated log PSDs for each segment.

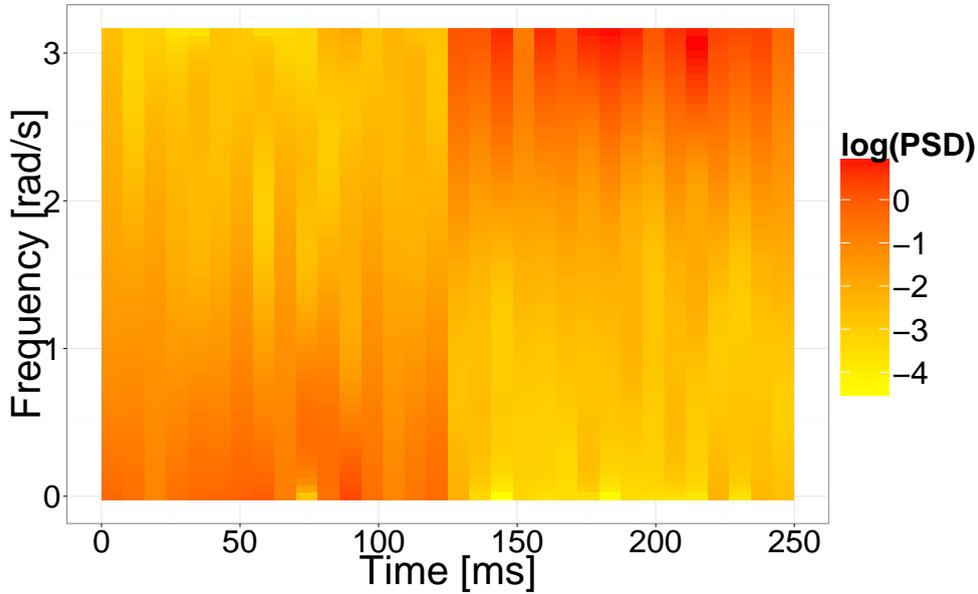


FIGURE 5.10: Time-frequency map showing the estimated posterior median log PSDs for 32 segments of 1/4 s of AR(1) noise. The change-point in spectral structure occurs exactly halfway through the series.

It is obvious that a change-point occurs halfway through Figure 5.10, as there is a sheer change in the spectral structure at this point between segments 16 and 17. The first half of the time-frequency map exhibits stronger low frequency behaviour, whereas the second half has more power in the higher frequencies.

Now consider the case where the change-point occurs during a segment rather than on the boundary. Here, let each segment have the same set-up as before, but instead set $n_1^* = 2^{11} - 2^6$ and $n_2^* = 2^{11} + 2^6$ such that a change-point occurs halfway through segment 16. A time-frequency map of the estimated log PSDs can be seen in Figure 5.11.

Figure 5.11 demonstrates that there is a noticeable change-point roughly halfway through the series. There is a smoother transition from one PSD structure to the other than in the previous example since the true change-point occurs in the middle of a segment rather than on the boundary.

These examples demonstrate that it is possible to detect potentially unknown change-points in a time series. This could be important for understanding nonstationarities in the data that could be due to glitches (or even transient burst signals). These would present themselves as excess power events over a very short duration and over all frequencies. It is important to note that if more segments are used, the time duration within each segment becomes smaller, and the accuracy in detecting the change-point increases. That is, the time at which the change-point occurs becomes more resolved if the segment durations are smaller. However, one must also ensure that the segment durations are long enough for the Whittle approximation to be valid.

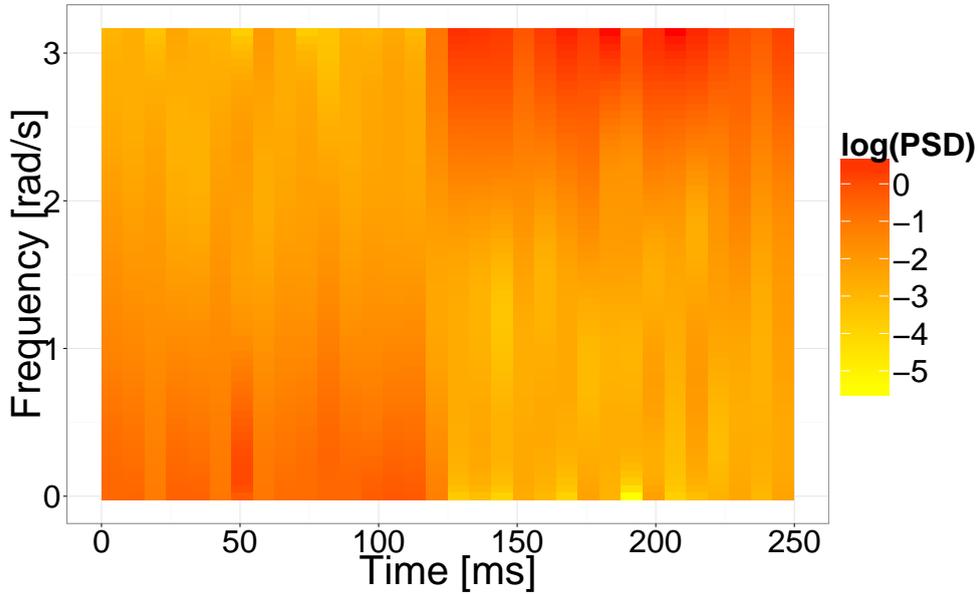


FIGURE 5.11: Time-frequency map showing the estimated posterior median log PSDs for 32 segments of $1/4$ s of AR(1) noise. The change-point in spectral structure occurs in the middle of segment 16 just before the halfway point.

5.2.4.6 Simulated Advanced LIGO noise

In this example, *A1O10.25* rotating core collapse gravitational wave signal from the Abdikamalov et al. (2014) catalogue is added to simulated Advanced LIGO noise, scaled to an SNR of $\varrho = 50$. A one detector set-up is assumed, with a linearly polarised gravitational wave signal (zero cross polarisation). The Advanced LIGO sampling rate is 2^{14} Hz, with a Nyquist frequency of 2^{13} Hz. Let $n = 2^{12}$, which corresponds to 0.25 s of data.

The simulated noise is Gaussian, coloured by the Advanced LIGO design sensitivity PSD. Generating this noise blindly results in a perfect matching of the end-points and their derivatives, due to the simplified frequency domain model. This is not realistic, since real data will often not have matching end-points. In order to make the noise generation more realistic, a longer frequency-domain series (ten times longer) is internally generated, inverse discrete Fourier transformed, and a fraction of it is returned with a random starting point. This is referred to as *padding* the data.

Figure 5.12 shows the estimated log PSD and the 90% credible region, overlaid with the log periodogram. The method performs remarkably well, particularly at higher frequencies. Even though frequencies below $\sim 10 - 20$ Hz will not be resolved at the Advanced LIGO design sensitivity (due to seismic noise), it is still interesting to see how this method performs at lower frequencies. Here, the low frequency estimates are slightly off, but not by much. This is likely to be due to two factors: $1/4$ s of simulated Advanced LIGO noise is actually a nonstationary series (i.e., the mean is not constant over time), and this was not adjusted for (simulated Advanced LIGO data are not

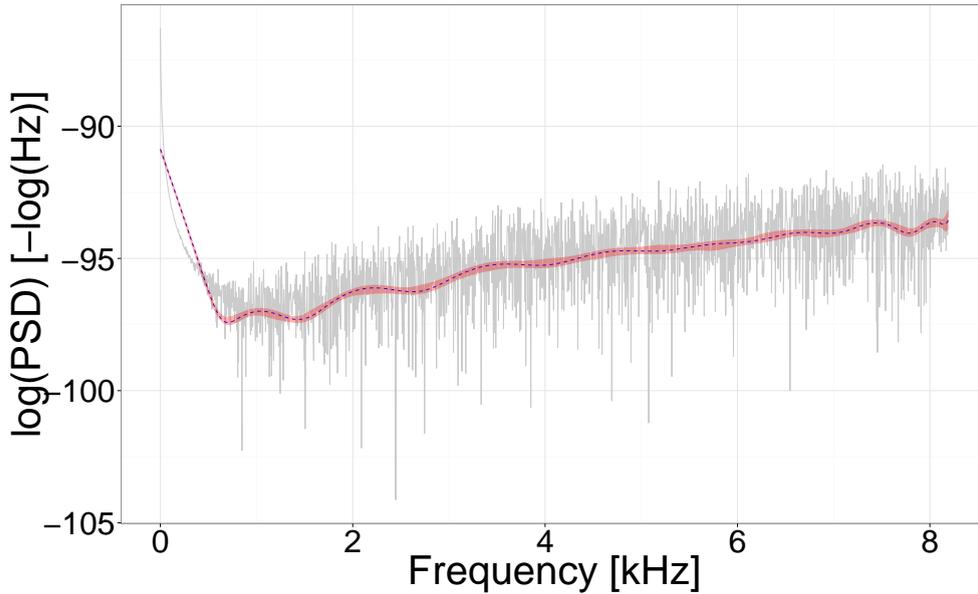


FIGURE 5.12: Estimated log PSD for simulated Advanced LIGO noise. 90% credible region (shaded pink) and posterior median (dashed blue) overlaid with log periodogram (solid grey).

stationary for more than 1/16 s based on the Augmented Dickey-Fuller test, Phillips-Perron unit root test, and KPSS test); and the Bernstein polynomial basis functions are notoriously slow to converge to a true function (Powell, 1981, Shen and Ghosal, 2014). These factors considered, the method still provides a reasonable approximation.

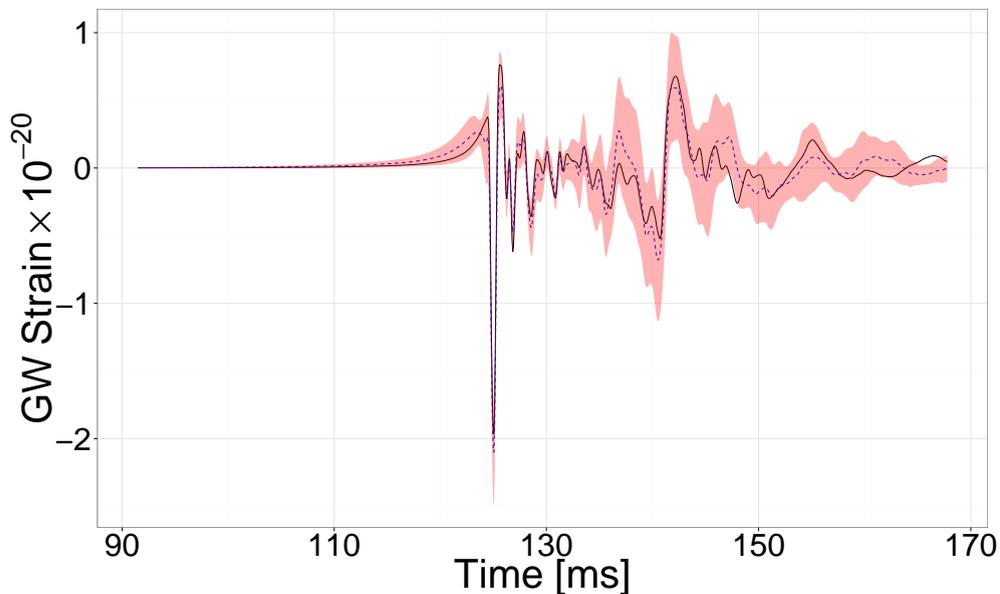


FIGURE 5.13: Reconstructed rotating core collapse gravitational wave signal. 90% credible interval (shaded pink) and posterior mean (dashed blue) overlaid with true A1O10.25 signal (solid black) from Abdikamalov et al. (2014) test catalogue. The signal is scaled to a SNR of $\rho = 50$.

The resultant reconstructed gravitational wave signal can be seen in Figure 5.13. The estimated signal here is very close to the true signal during the the collapse and bounce phases, as well as

during the ring-down oscillations. The 90% credible region contains most of the true gravitational wave signal.

For this example, $d = 25$ PCs were used to reconstruct a rotating core collapse gravitational wave signal, but this could be too many or too few basis functions. Model selection methods similar to Edwards et al. (2014) (and presented in 4.3.3 and 4.3.4) were not investigated here.

Nonstationarities in detector noise are then accommodated by breaking the series into smaller and locally stationary components. The resulting time-varying spectrum is then analysed. This can be seen in Figure 5.14. Rather than choosing $J = 32$ as in 5.2.4.5, nonstationarities in the Advanced LIGO noise become more apparent if the noise series is sliced into fewer segments, each with longer duration. Instead, consider splitting the data into $J = 8$ equal length segments ($n_j = 2^9$). Here, the Whittle approximation is valid, and the segments visually look locally stationary.

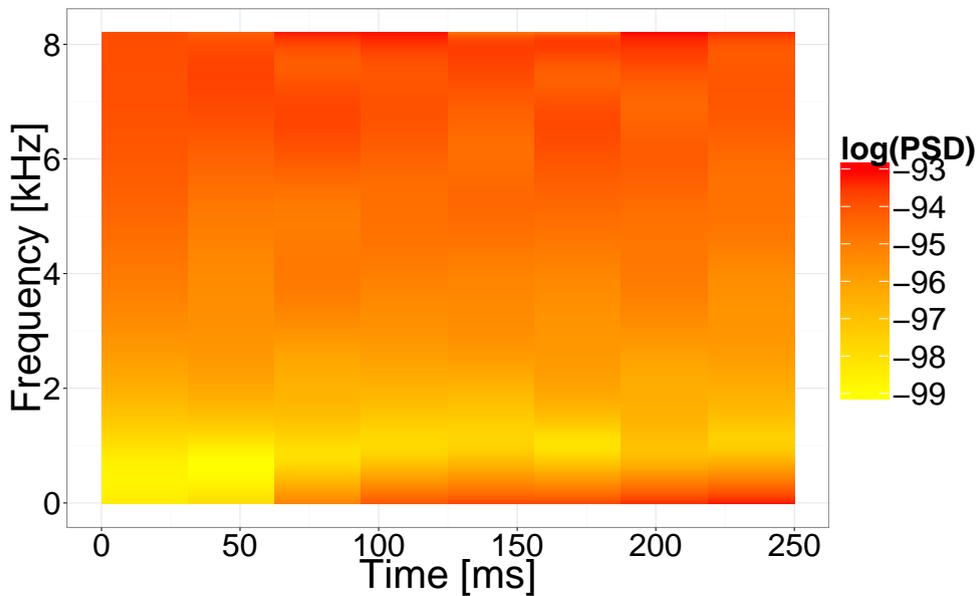


FIGURE 5.14: Time-frequency map of the estimated time-varying noise spectrum for 8 segments of 1/4 s simulated Advanced LIGO noise. The posterior median log PSDs for each noise segment are used.

Figure 5.14 illustrates that the Advanced LIGO PSD is changing slightly over time. Notice that lower frequencies are gaining more power over time here. Assuming each segment is locally stationary (which should be the case since the duration of each segment is less than 1/16 s), it is important to accommodate for the changing nature of the PSD since the Choudhuri et al. (2004) PSD estimation technique is based on the theory of stationary processes. If no adjustments for nonstationarities are made (like the one proposed here), estimates of astrophysically meaningful parameters could become unfavourable or biased.

5.3 Advanced LIGO spectral density estimation using the B-spline prior

In this section, B-spline prior PSD estimation results from Edwards et al. (2017a) are presented. Methodology from that paper has been provided in 3.3.4 and 3.5.4.2. A noise-only model is discussed, as a proof of concept relating to an embedded gravitational wave signal (via a blocked Gibbs sampler) was already demonstrated in 5.2.

The B-spline prior is a new nonparametric prior, developed as a more flexible class of priors than the Bernstein polynomial prior in the context of (but not limited to) Bayesian time series analysis. Bernstein polynomials are not a flexible enough class of functions to handle sharp and abrupt spectral lines in real gravitational wave detector noise as beta cdfs do not provide adequate coverage of the space of cdfs on the unit interval. However, when using B-spline densities instead, the loss function in 3.3.4.2 can be made arbitrarily small by increasing the number of knots.

For the subsections to follow, MCMC implementation details are given here. A Metropolis-within-Gibbs algorithm is used to sample points from the pseudo-posterior, using the same modular symmetric proposal distributions for B-spline weight parameters \mathbf{V} and \mathbf{Z} as described by Choudhuri et al. (2004). That is, say for V_l , propose a candidate from a uniform distribution with $[V_l - \epsilon_l, V_l + \epsilon_l]$, modulo the circular unit interval. If the candidate is greater than 1, take the decimal part only, and if the candidate is less than 0, add 1 to put it back into $[0, 1]$. This is done for all of the \mathbf{V} and \mathbf{Z} parameters. Choudhuri et al. (2004) found that $\epsilon_l = l/(l + 2\sqrt{n})$ worked well for most cases, and this is also adopted here. The same approach is used analogously for the B-spline knot location parameters \mathbf{U} and \mathbf{X} . Parameter τ has a conjugate inverse-gamma prior and may be sampled directly. The number of mixture components k could be sampled directly from its discrete full conditional (as done by Choudhuri et al. (2004)), though this can be computationally expensive for large k_{\max} , so a Metropolis proposal centred on the previous value of k is used instead. Let this proposal have a 75% chance of jumping according to a discrete uniform on $[-1, 1]$, and a 25% chance of boldly jumping according to a discretised Cauchy random variable, to mitigate getting stuck in local modes.

Let the B-spline prior have the following noninformative prior specification:

$$\begin{aligned}G_0 &= \text{Uniform}[0, 1], \\H_0 &= \text{Uniform}[0, 1], \\M_G &= 1, \\M_H &= 1, \\\alpha_\tau &= 0.001, \\\beta\tau &= 0.001, \\\theta_k &= 0.01, \\k_{\max} &= 500.\end{aligned}$$

For comparability, whenever the Bernstein polynomial prior algorithm is used in this analysis, let it have exactly the same prior set-up, but obviously without knot location parameter M_H and distribution H_0 .

For both algorithms, $k_{\max} = 500$. This may seem unnecessarily large for the B-spline prior algorithm as these simple cases converge to a low k . However, it is large enough to ensure the Bernstein polynomial algorithm converges to an appropriate k , without being truncated at k_{\max} .

For the simulation study in 5.3.1 and the Advanced LIGO application in 5.3.2, let the stick-breaking truncation parameters be $L_G = L_H = 20$ (based on Choudhuri et al. (2004)) and increase this to $L_G = L_H = 60$ for the Advanced LIGO application in 5.3.2.1.

The (cubic) B-spline prior algorithm was implemented as the `gibbs_bspline` function in the R package `bsplinePsd` (Edwards et al., 2017b). The Bernstein polynomial prior algorithm was implemented as the `gibbs_NP` function in the R package `beyondWhittle` (Kirch et al., 2017, Meier et al., 2017). Both packages are available on CRAN.

5.3.1 Simulation study: A comparison of the Bernstein polynomial prior and B-spline prior

In this section, a simulation study is conducted to compare the accuracy of spectral density estimates using the Bernstein polynomial prior and the B-spline prior. It will be shown that the B-spline prior outperforms the Bernstein polynomial prior in terms of integrated absolute error (IAE) and uniform coverage probability.

Consider two autoregressive time series of differing order, $p = 1$ and $p = 4$. For the first scenario, an AR(1) with first order autocorrelation $\rho_1 = 0.9$ (a relatively simple spectral density) is generated.

In the second scenario, an AR(4) with parameters $\rho_1 = 0.9, \rho_2 = -0.9, \rho_3 = 0.9,$ and $\rho_4 = -0.9$ is generated, such that the spectral density has two large peaks (a more complicated spectrum). Let each time series have lengths $n = \{128, 256, 512\}$ with unit variance Gaussian innovations. Under these conditions, 1,000 different realisations of AR(1) and AR(4) data are simulated and their spectral densities are modelled by running the Bernstein polynomial prior algorithm of Choudhuri et al. (2004) and the B-spline prior algorithm of Edwards et al. (2017a) on each of these. The MCMC algorithms (without parallel tempering) run for 400,000 iterations, with a burn-in period of 200,000 and thinning factor of 10.

An AR(p) model has theoretical spectral density,

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \sum_{j=1}^p \rho_j \exp(-i\lambda)|^2}, \quad (5.9)$$

where σ^2 is the variance of the white noise innovations and (ρ_1, \dots, ρ_p) are the model parameters. Monte Carlo estimates can be compared to the true spectral density to measure relative performance of the nonparametric priors. One measure of closeness and accuracy is the *integrated absolute error* (IAE), also known as the L_1 -error. This is defined as:

$$\text{IAE} = \|\hat{f} - f\|_1 = \int_0^\pi |\hat{f}(\lambda) - f(\lambda)| d\lambda, \quad (5.10)$$

where $\hat{f}(\cdot)$ is the Monte Carlo estimate (posterior median) of the spectral density $f(\cdot)$. IAE is calculated for each replication and the median IAE over all 1,000 replications is compared. The results are presented in Table 5.1.

TABLE 5.1: Median L_1 -error for the estimated spectral densities using B-spline prior and Bernstein polynomial prior on simulated AR(1) and AR(4) data.

AR(1)	$n = 128$	$n = 256$	$n = 512$
B-spline	0.901	0.756	0.592
Bernstein	0.830	0.706	0.518
AR(4)	$n = 128$	$n = 256$	$n = 512$
B-spline	3.242	2.371	1.886
Bernstein	3.427	2.920	2.656

Table 5.1 compares the median IAE of the estimated spectral densities under the two different nonparametric priors. For the AR(1) cases, the median IAE is only marginally higher for the B-spline prior than the Bernstein polynomial prior. As the AR(1) has a simple spectral structure, this is a case where the global support of the Bernstein polynomials makes sense. However, when estimating the more complicated AR(4) spectral density, the B-spline prior yields more accurate estimates than the Bernstein polynomial prior in terms of IAE. It can also be seen that for both

priors, as n increases, the median IAE decreases. This is a promising sign towards posterior consistency.

For each simulation, two different credible regions are calculated: The usual equal-tailed pointwise credible region, and the uniform (or simultaneous) credible band, popular in the bootstrap literature (Lenhoff et al., 1999, Neumann and Kreiss, 1998, Neumann and Polzehl, 1998, Sun and Loader, 1994). Uniform credible bands are very useful as they allow the calculation of coverage levels for entire curves (spectral densities in this case) rather than pointwise intervals. To compute a $100(1 - \alpha)\%$ uniform credible band, the following formula is used:

$$\hat{f}(\lambda) \pm \zeta_\alpha \times \text{mad}(\hat{f}_i(\lambda)), \quad \lambda \in [0, \pi], \quad (5.11)$$

where $\hat{f}(\lambda)$ is the pointwise posterior median spectral density, $\text{mad}(\hat{f}_i(\lambda))$ is the median absolute deviation of the posterior samples $\hat{f}_i(\lambda)$ kept after burn-in and thinning (which is used as the estimate of dispersion of the sampling distribution of $\hat{f}(\lambda)$), and ζ_α is the quantile chosen such that

$$\mathbb{P} \left\{ \max \left\{ \frac{|\hat{f}_i(\lambda) - \hat{f}(\lambda)|}{\text{mad}(\hat{f}_i(\lambda))} \right\} \leq \zeta_\alpha \right\} = 1 - \alpha. \quad (5.12)$$

The maximum in 5.12 is the largest value for each sampled PSD over all λ (i.e., one number per sampled PSD).

Based on these uniform credible bands, uniform coverage probabilities over all 1,000 replications of the simulation can be computed. That is, calculate the proportion of times that the true spectral density is entirely encapsulated within the uniform credible band. Computed coverage probabilities are shown in Table 5.2.

TABLE 5.2: Coverage probabilities based on 90% uniform credible bands.

AR(1)	$n = 128$	$n = 256$	$n = 512$
B-spline	1.000	1.000	0.998
Bernstein	1.000	0.987	0.499
AR(4)	$n = 128$	$n = 256$	$n = 512$
B-spline	0.936	0.979	0.907
Bernstein	0.000	0.000	0.000

It can be seen in Table 5.2 that the B-spline prior has higher coverage than the Bernstein polynomial prior in all examples (apart from the AR(1) with $n = 128$, where it is the same). The B-spline prior produces excellent coverage probabilities for the AR(1) cases. The Bernstein polynomial prior also performs well in this regard, apart from the $n = 512$ case, where half are not fully covered. An example from one of the 1,000 replications of the AR(1) with $n = 512$ is given in Figure 5.15. Here, the uniform credible band fully contains the true PSD for the B-spline prior

but not for the Bernstein polynomial prior (the true PSD falls outside the uniform credible band at the highest frequencies). The pointwise credible region and posterior median log-PSD for both priors are also very accurate. This is not surprising as the AR(1) has a relatively simple spectral structure.

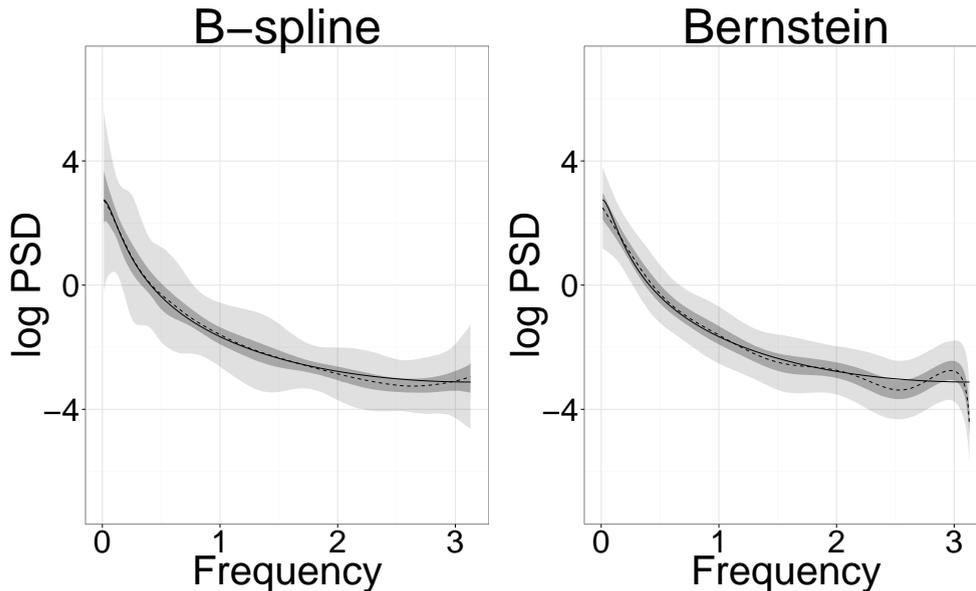


FIGURE 5.15: Estimated log-spectral densities for an AR(1) time series using the B-spline prior (left) and Bernstein polynomial prior (right). The solid line is the true log-PSD; the dashed line is the posterior median log-PSD; the dark shaded region is the pointwise 90% credible region; and the light shaded region is the uniform 90% credible band.

Coverage of the AR(4) spectral density under the B-spline prior is above 90% for each sample size. However, the Bernstein polynomial prior has extremely poor coverage in the AR(4) case, where none of the 1,000 replications are fully covered by the uniform credible band for each sample size. An example of this performance (for $n = 512$) can be seen in Figure 5.16. The Bernstein polynomial prior (under the noninformative prior set-up) tends to over-smooth the second large peak of the PSD, and introduces additional incorrect peaks and troughs throughout the rest of estimate. These false peaks and troughs are present due to the Bernstein polynomial prior algorithm converging to large k in an attempt to approximate the two large peaks of the AR(4) PSD. The B-spline prior gives a much more accurate Monte Carlo estimate. The posterior median log-PSD is close to the true AR(4) PSD, the 90% pointwise credible region mostly contains the true PSD, and the 90% uniform credible band fully contains it.

Of course, the Bernstein polynomial prior could perform better on spectral densities with sharp features if significant prior knowledge was known in advance. This can however be a formidable task, and is not very generalizable to other time series. A benefit of using the B-spline prior is its ability to estimate a variety of different spectral densities using the default noninformative priors used in this paper. Another example of this is demonstrated in the next section.

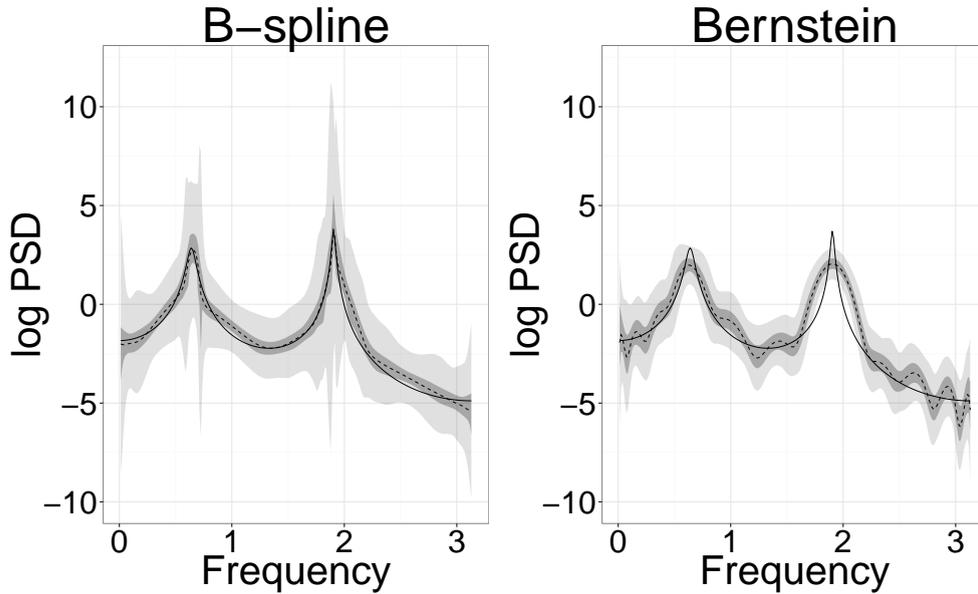


FIGURE 5.16: Estimated log-spectral densities for an AR(4) time series using the B-spline prior (left) and Bernstein polynomial prior (right). The solid line is the true log-PSD; the dashed line is the posterior median log-PSD; the dark shaded region is the pointwise 90% credible region; and the light shaded region is the uniform 90% credible band.

One slight drawback of the B-spline prior algorithm is its computational complexity relative to the Bernstein polynomial prior. B-spline densities must be evaluated many times per MCMC iteration (when sampling k , \mathbf{U} , and \mathbf{X}) due to the variable knot placements, whereas beta densities can be pre-computed and stored in memory, saving much computation time.

Table 5.3 shows the median run-time (over each 1,000 replication) for each of the six AR simulations. It can be seen that the B-spline prior algorithm is approximately 2–3 times slower than the Bernstein polynomial prior algorithm for these examples. Due to the noted advantages (accuracy and coverage) that the B-spline prior has over the Bernstein polynomial prior, particularly for PSDs with complicated structures, the increased computation time is an acceptable trade-off, though for simple spectral densities, the Bernstein polynomial prior should suffice.

TABLE 5.3: Median absolute run-times (hours) and their associated relative run-times.

AR(1)	$n = 128$	$n = 256$	$n = 512$
B-spline	2.967	3.186	3.659
Bernstein	1.423	1.572	1.844
B-spline/Bernstein	2.086	2.026	1.985
AR(4)	$n = 128$	$n = 256$	$n = 512$
B-spline	4.044	4.422	5.174
Bernstein	1.443	1.694	2.281
B-spline/Bernstein	2.802	2.610	2.268

5.3.2 Application of the B-spline prior to recoloured LIGO S6 data

Data collected by the Advanced LIGO observatories are dominated by instrument noise — primarily seismic, thermal, and photon shot noise. There are also high power, narrow band, spectral noise lines caused by the AC electrical supplies and mirror suspensions, among other phenomena. Due to the undesirable properties of the Bernstein polynomial prior (poor coverage and slow convergence), it was not flexible enough to estimate the spectral lines in the spectral density of real LIGO noise. This, coupled with the fact that B-splines have local support, provided the rationale for implementing the B-spline prior instead.

In the following example, consider a 1 s stretch of real LIGO data collected during the sixth science run (S6), recoloured to match the target noise sensitivity of Advanced LIGO (Christensen, 2010). LIGO has a sampling rate of 16384 Hz. To reduce the volume of data processed, a low-pass Butterworth filter (of order 20 and attenuation 0.25) is applied, then the data are downsampled to 4096 Hz. Prior to downsampling, the data are differenced once to become stationary, mean-centred, and Hann windowed to mitigate spectral leakage. Though a 1 s stretch may seem small in the context of gravitational wave data analysis, this time scale is important for on-source characterisation of noise during short-duration transient events (bursts), such as core collapse supernovae (Abadie et al., 2012a). This is particularly true since LIGO noise has a time-varying spectrum, and systematic biases could occur if off-source noise was used to estimate the power spectrum of on-source noise.

A parallel tempered Metropolis-within-Gibbs MCMC algorithm is implemented, running 16 parallel chains (each at different temperatures). These are each run for 400,000 iterations, with a burn-in of 200,000 and thinning factor of 5. Swaps (of all parameters blocked together) are proposed between adjacent chains on every tenth iteration, allowing for the trickling down of information from high to low energy states. For each chain c , the following inverse-temperature scheme gave reasonable results:

$$T_c^{-1} = T_{\min}^{-\Delta_c}, \quad (5.13)$$

where $T_{\min} = 0.005$ is the minimum inverse-temperature allowed, $\Delta_c = \frac{c-1}{C-1}$, and $C = 16$ is the number of chains.

As demonstrated in the previous section (e.g., Figure 5.16), the Bernstein polynomial approach would have struggled to estimate the abrupt changes of power present in real detector data. It can be seen in Figure 5.17 though, that the B-spline prior approach estimates the log-spectral density very well. The estimated log-PSD follows the general broad-band shape of the log-periodogram well, and the primary sharp changes in power are also accurately estimated. The method, however, seems to be less sensitive to the smaller spikes.

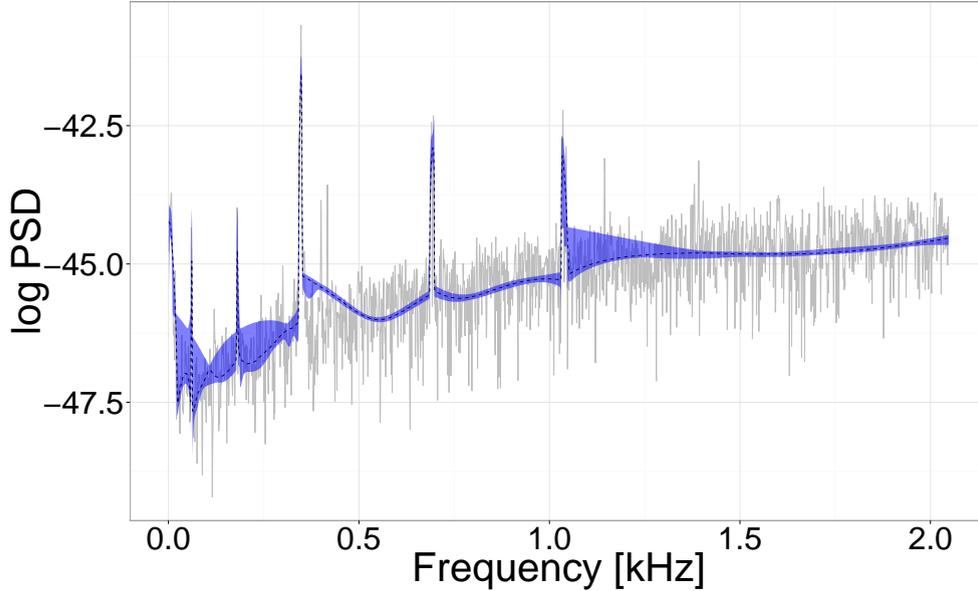


FIGURE 5.17: Estimated \log_{10} PSD for a 1 s segment of recoloured LIGO S6 data. The posterior median log-PSD (dashed black) along with the pointwise credible region (shaded blue) are overlaid with the log-periodogram (grey). The log transform is base 10 here.

5.3.2.1 Informative prior

Consider the informative prior set-up setting $M_H = n = 4096$. This is the case where there is a stronger prior belief that knot placements are uniform. Also consider increasing the stick-breaking truncation parameters to $L_G = L_H = 60$. This improves the accuracy of the stick-breaking construction, but at the expense of additional computational overheads. The parallel tempered MCMC algorithm is run for 200,000 iterations, with a burn-in of 80,000 and thinning factor of 10. A plot of the estimated log PSD of LIGO S6 noise can be seen in Figure 5.18.

The largest peaks (apart from the one at 1 kHz) are all estimated well under this set-up. This includes the twin peaks around 400 Hz. The broad-band features are also estimated well. To improve the fit even further, the MCMC code will require a significant improvement in efficiency. In its current state, it is infeasible to increase the stick-breaking truncation parameters any further due to computation speed. Further improvements to the code could also come in exploring other tempering schemes.

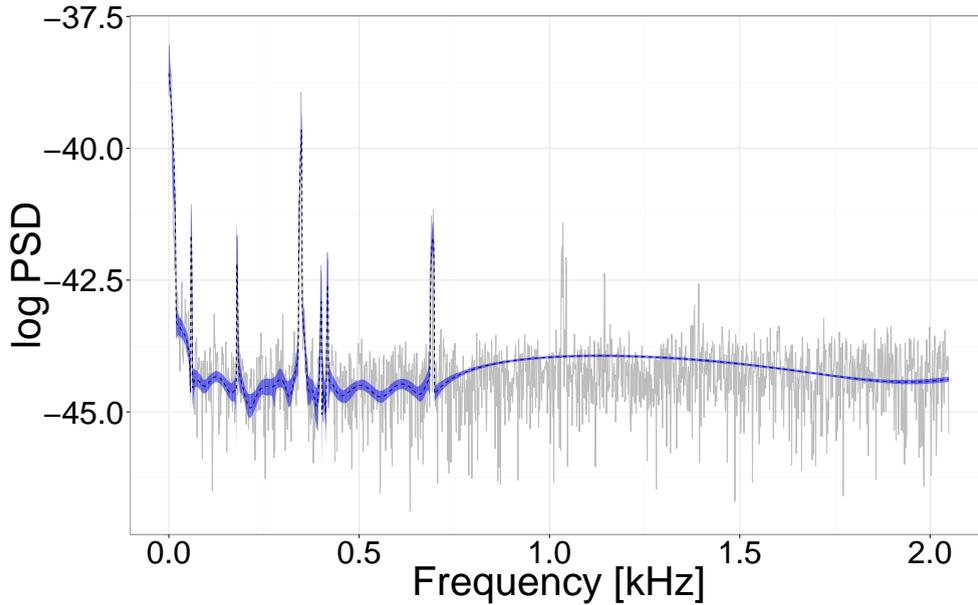


FIGURE 5.18: Estimated \log_{10} PSD for a 1 s segment of recoloured LIGO S6 data, using an informative $M_H = 4096$. The posterior median log-PSD (dashed black) along with the pointwise credible region (shaded blue) are overlaid with the log-periodogram (grey). The log transform is base 10 here.

5.4 Advanced LIGO spectral density estimation using a nonparametric correction to a parametric likelihood

In this section, the nonparametric correction (NPC) procedure presented in 3.5.4.3 is applied to real LIGO noise from the S6 run. Further methodological details (e.g., posterior consistency) and numerical results can be found in the joint work of Kirch et al. (2017).

Consider 1 s of real LIGO data collected during the S6 run, recoloured to match the target noise sensitivity of Advanced LIGO (Christensen, 2010). The data are differenced and then multiplied by a Hann window to mitigate spectral leakage. A low-pass Butterworth filter (of order 20 and attenuation 0.25) is then applied before downsampling from a LIGO sampling rate of 16384 Hz to 4096 Hz, reducing the volume of data.

Apart from changing $k_{\max} = 1000$, the same prior set-up as in 5.2.4 is used. Further, a Uniform[0, 1] prior is placed on spectral shape confidence parameter η , and Uniform[-1, 1] priors are placed on the partial autocorrelation parameters ρ . The results in this section make use of the `gibbs_NPC`, `gibbs_NP`, and `gibbs_AR` functions in the R package `beyondWhittle` (Meier et al., 2017).

With regards to selecting an appropriate $AR(p)$ working model, one could analyse DIC for increasing orders. However, scree-like plots of negative maximum log likelihood for increasing orders often show a clear bend (elbow) in the curve. Similar to scree plots in the context of PCA, the elbow could be considered a reasonable truncation point (“elbow criterion”). Though the model selected

under this criterion will not fully explain the data, adding more parameters generally does not help the nonparametric correction procedure.

First, a pure nonparametric model is run, corresponding to a nonparametrically corrected likelihood with an AR(0) working model (i.e., the Whittle likelihood) to estimate the spectral density. This is then compared to a nonparametrically corrected model with an order of $p = 14$, where a clear elbow can be seen in the negative log-likelihood plot (see Figure 5.19). These simulations run for 100,000 MCMC iterations, with a burn-in of 50,000, and thinning factor of 5. Results are illustrated in Figure 5.20.

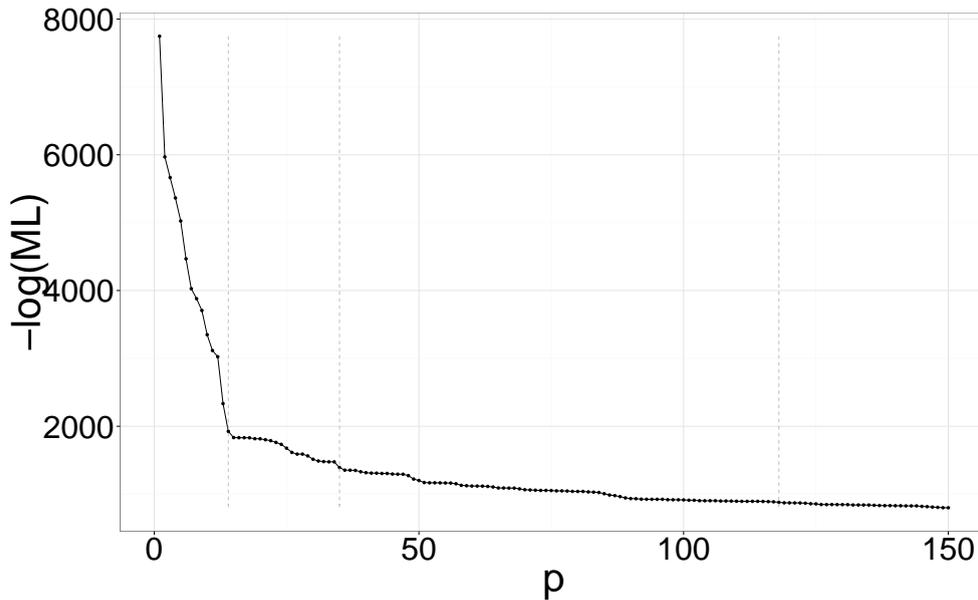


FIGURE 5.19: Negative maximum log likelihood for different AR(p) models applied to Advanced LIGO S6 data.

Even though k converged to $k \approx 900$ mixture components, it is clear that the Bernstein polynomial prior together with the Whittle likelihood is not flexible enough to estimate the sharp peaks of the LIGO spectral density. The parametric AR(14) model (estimated using the Bayesian autoregressive sampler described in Kirch et al. (2017)) captures the four main peaks but not their sharpness. Additionally, it does not capture the structure well in the frequency bands 0 to 450 Hz as well as larger than 1100 Hz. When compared to the AR(0) model, the nonparametrically corrected model based on $p = 14$ estimates the sharp peaks much better. Furthermore, it sharpens all four peaks of the AR(14)-model (with a slight exception around 400 Hz, where seemingly two very sharp peaks overlap, a feature that is not captured by the AR(14) model at all). In the frequency bands 0 to 450 Hz as well as larger than 1100 Hz, where the parametric model fails altogether, the correction yields similar results to the nonparametric Whittle procedure. Similarly to the nonparametric Whittle procedure k tends towards $k_{\max} = 800$ indicating that the Bernstein polynomial prior together with an AR(14)-model is not yet flexible enough for this data set.

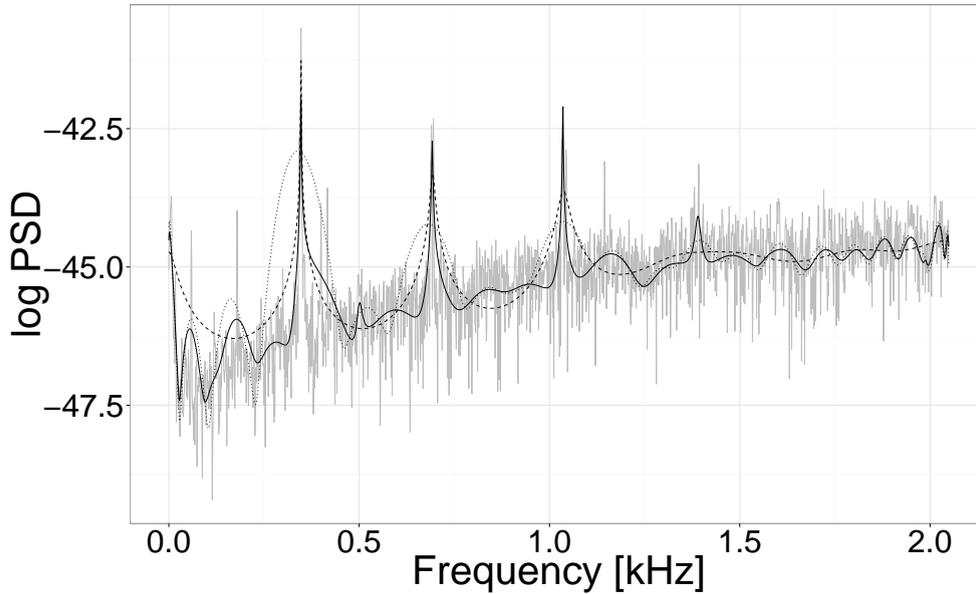


FIGURE 5.20: Estimated log spectral density for a 1 s segment of LIGO S6 data. The posterior median log spectral density estimate of NPC under an AR(14) working model (solid black), AR(14) (dashed black), and NPC under an AR(0) working model (dotted black) are overlaid with the log periodogram (grey).

Looking closer at Figure 5.19, the negative log-likelihood between the models AR(14) and AR(35) decreases by 533, which is not as sharp as the elbow at $p = 14$, but still large — keeping in mind that the LIGO data is a very complex data set. There is a moderately sized jump before $p = 35$, while afterwards the descent slows down. In fact the BIC chooses an order of $p = 118$, where the log-likelihood reaches the level of 879, showing that the difference between $p = 14$ and $p = 35$ (of 533) is comparable to the one between $p = 35$ and $p = 128$ (of 512). This indicates that there is indeed another change of gradient around $p = 35$. When looking at penalised likelihoods, this is also the point where different penalisations start to obviously diverge. The results for $p = 35$ can be found in Figure 5.21.

The parametric AR(35) model already provides a reasonable fit to the periodogram, picking up the major peaks (with the exception around 100 Hz), but under- and over-estimates some of the peaks. In particular there are still major problems in the frequency bands 0 to 300 Hz and 400 to 700 Hz. The NPC procedure with $p = 35$ keeps the peaks that have been captured well by the parametric model but corrects problems most prominently in the above mentioned frequency bands. It is worth mentioning that the correction works in several ways: Sharpening existing peaks (e.g. at 0 Hz), adding new peaks (e.g. at 100 Hz) as well as smoothing out some erroneous peaks (e.g. at 600 Hz). Overall, the resulting estimate seems to capture the structure quite well in all frequency bands. This impression is complemented by the results of the NPC method with AR(35) working model together with the pointwise and uniform credible bounds obtained from the procedure in Figure 5.22.

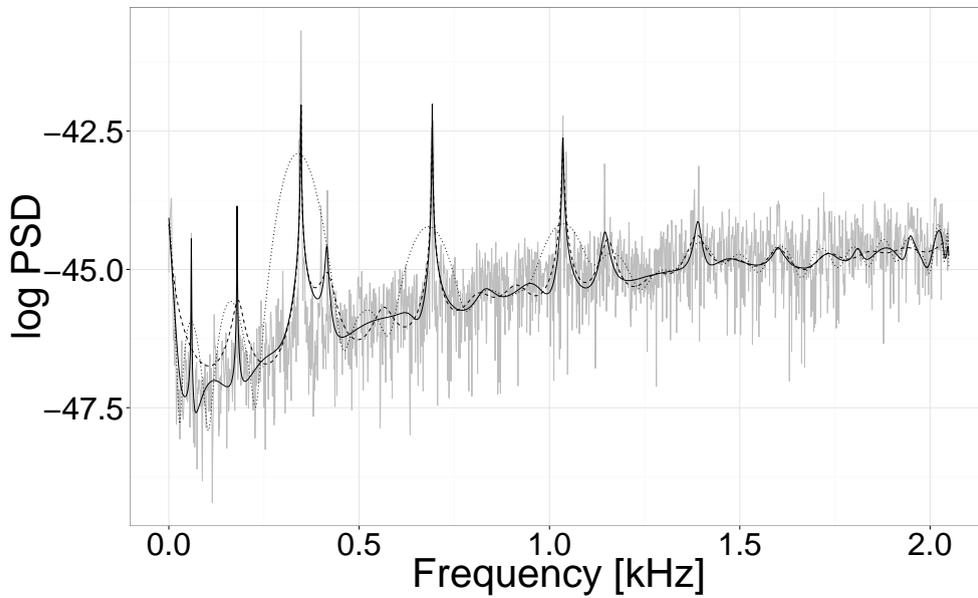


FIGURE 5.21: Estimated log spectral density for a 1 s segment of LIGO S6 data. The posterior median log spectral density estimate of NPC under an AR(35) working model (solid black), AR(35) (dashed black), and NPC under an AR(0) working model (dotted black) are overlaid with the log periodogram (grey).

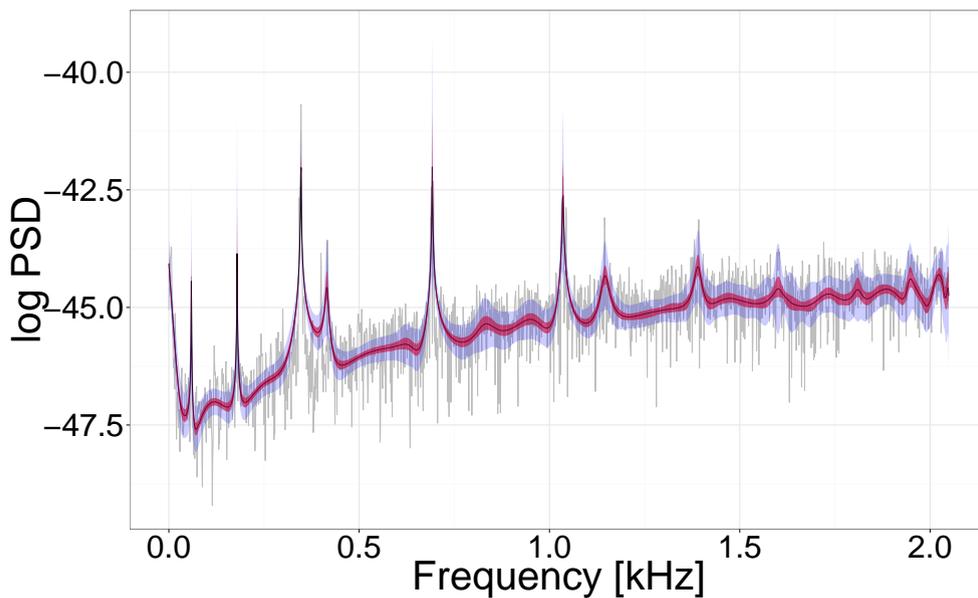


FIGURE 5.22: Estimated log spectral density for a 1 s segment of LIGO S6 data. The posterior median log spectral density estimate of NPC under an AR(35) working model (solid black), pointwise credible region (shaded red), and uniform credible region (shaded violet) are overlaid with the log periodogram (grey).

Chapter 6

Conclusions and outlook

Presented in this thesis were Bayesian approaches to gravitational wave data analysis, with a particular focus on parameter estimation of rotating stellar core collapse gravitational wave signals, and realistic modelling of the power spectral density of LIGO detector noise. A framework for inferring the astrophysical parameters of a rotating stellar core collapse gravitational wave signal was presented, using posterior predictive sampling and supervised machine learning algorithms. Bayesian nonparametric routines were then employed to estimate the PSD of Advanced LIGO noise.

It was shown that with an SNR of $\varrho = 20$ and optimal orientation of detector to source, the ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce can be inferred with reasonable levels of uncertainty using posterior predictive sampling. This level of uncertainty increased when incorporating an unknown signal arrival time into the signal reconstruction model, but this came as no surprise as the uncertainty of the signal arrival time propagated through to the PC coefficients. The posterior predictive sampling approach presented here may be applied generally to any supernova waveform catalogue and to any continuous astrophysical parameter of interest. Indeed, it was initially applied to the Dimmelmeier et al. (2008) waveform catalogue as a proof of concept and then easily transferred to the Abdikamalov et al. (2014) catalogue. Although not presented here, predictions and conclusions relating to the initial central angular velocity were comparable to those of the ratio of rotational kinetic energy to gravitational potential energy of the inner core at bounce. This was expected as the ratio of rotational kinetic energy to gravitational energy of the inner core at bounce is a function of central angular velocity.

Choosing to only use the posterior means of the PC coefficients in the construction of the linear model design matrix removed some of the variability due to LIGO noise and signal reconstruction uncertainty. The uncertainty from the Bayesian PCR modelling step therefore does not propagate

through to the posterior predictive sampling step. An alternative approach for future study would be to incorporate this uncertainty through an errors-in-variables model, which is commonly used when there are measurement errors in the explanatory variables of a regression model. A hierarchical approach would allow for propagation of uncertainty. However, a benefit of the methods presented here were that predictions were essentially independent of SNR (at least for $\varrho \geq 20$).

To extract the level of precollapse differential rotation, two supervised learning algorithms and posterior predictive sampling from a Bayesian ordinal probit model were used. All results were comparable. The k -nearest neighbours algorithm generally performed better than naïve Bayes under the given assumptions, and the Bayesian ordinal probit model outperformed both machine learning algorithms for strong and weak precollapse differential rotation, including signals with a differing nuclear equation of state and deleptonisation parametrisation. It could be useful to investigate how the choice of prior for the naïve Bayes classifier affects classification, and how different metrics such as the Mahalanobis distance (which takes correlations of the data into account) influences classification for k -nearest neighbours.

With regards to PC model selection, two approaches were considered. A constrained optimisation of the DIC was used to select a fixed number of PCs for the Bayesian PCR models. A RJMCMC approach was also implemented, treating the number of PCs as variable. Under the reversible jump regime, one can take advantage of automatic model selection via model averaging. For parameter estimation purposes, the former approach was used as one requires the same number of PCs for each signal when creating the linear model design matrix for posterior predictive sampling. This is not possible under the current reversible jump implementation, as all signals are modelled separately, and different signals will converge to different models. A future extension is to conduct parameter estimation when signals are modelled together in a multivariate setting. Further, alternative signal reconstruction models, such as the Bayesian LASSO and Gaussian process regression, could be considered.

Extracting nuclear equation of state was somewhat unsuccessful. This is most likely due to the fact that the two equations of state analysed were not overly influential in the structure of the gravitational wave signal (Abdikamalov et al., 2014, Dimmelmeier et al., 2008). However, the cross-validation approach gave better-than-expected results, with 50% correctly identified.

The rotating stellar core collapse parameter estimation routines only considered optimal orientation of a gravitational wave source to a single interferometer. Similar to Röver et al. (2007a,b) but in the context of rotating stellar core collapse signals, it would be useful to generalise to a network of detectors, using coherent methods. One could then locate the source of the gravitational wave in the sky. Source localisation could then allow for robust glitch rejection via electromagnetic follow-up.

Parameter estimation routines have dominated the gravitational wave data analysis literature, while noise PSD modelling has often been neglected. It could be argued that PSD estimation is just as important as parameter estimation as one would want to ensure no systematic biases occur due to a poorly specified parametric noise model. The standard approach for estimating the noise PSD involves taking an average of off-source data using the Welch method, then assuming stationary and Gaussian noise with a known PSD. The assumptions of the standard noise model are too restrictive for Advanced LIGO data. Gravitational wave data are subject to high amplitude non-Gaussian transient noise artifacts, meaning the Gaussian assumption is no longer valid. If the noise model is incorrectly specified using a parametric (Gaussian) model, misleading inferences could be made. The stationarity assumption is also not valid as the noise PSD is time-varying. Using off-source data to estimate the PSD may be problematic since the PSD will naturally drift over time, and could have a different structure to the PSD of noise on the gravitational wave source. Bayesian nonparametric methods were employed to address the limitations of the standard noise model.

To address the Gaussian assumption, a nonparametric Bernstein polynomial prior was placed on the LIGO noise PSD. An MCMC framework was then developed to allow for on-source characterisation of the LIGO noise PSD, while simultaneously reconstructing a rotating stellar core collapse gravitational wave signal. Although attention was restricted to gravitational waves from rotating stellar core collapse, the approach is perfectly valid for any parametric gravitational wave signal embedded in noise. To account for nonstationarities in LIGO noise, local rather than global stationarity was considered.

Since the “theoretical” PSD of Advanced LIGO at its design sensitivity has a very steep decrease at low frequencies (until it reaches a minimum at roughly 230 Hz), it is difficult for the Bernstein polynomial prior MCMC algorithm to perfectly characterise the shape at low frequencies without increasing computation significantly. This is due to the well-known slow convergence and poor coverage properties of Bernstein basis functions. That is, many Bernstein polynomials (of order $k = 1000$ or more) are required to accurately characterise the PSD of simulated Advanced LIGO noise, and this is computationally infeasible. Compare this to more well-behaved noise sources, such as those from autoregressive processes, which require $k < 50$. Real Advanced LIGO noise has an even more complicated spectral structure, with sharp and abrupt spectral lines, which this algorithm cannot handle. This led to the development of a more flexible prior, the B-spline prior, which can be considered a generalisation to the Bernstein polynomial prior.

It was shown under default noninformative priors that the B-spline prior outperforms the Bernstein polynomial prior in terms of integrated absolute error and uniform coverage probabilities, yielding superior Monte Carlo estimates, particularly for spectral densities with sharp and abrupt changes in power, such as the LIGO PSD from the S6 run. This was not surprising as B-splines have

local support and better approximation properties than Bernstein polynomials. However, a future statistical challenge is to prove posterior consistency to ensure the posterior estimate converges in the neighbourhood of the true PSD. An additional future direction is to apply the B-spline prior to a more broader statistical context, such as density estimation.

An alternative approach to model the LIGO S6 noise PSD was to use a nonparametric correction to a parametric likelihood. An autoregressive working model was assumed, which was then corrected nonparametrically. Under this methodology, erroneous peaks were damped down, new peaks were introduced, and correct peaks were sharpened. Future directions of this work could be to replace the Bernstein polynomial prior with the B-spline prior. In addition, if there exists a “theoretical” LIGO PSD that includes known spectral lines, one could use this information in the correction matrix. In this way, one can include prior information without using an informative prior via a nonparametric correction to a parametric likelihood.

The path forward is clear — to implement these Bayesian nonparametric PSD estimation methods in the LSC Algorithm Library (LAL). As the gravitational wave data analysis community converges on an adequate noise model, the Bayesian nonparametric framework presented here shows much promise. These methods would be well-suited to many Bayesian parameter estimation routines from a variety of different sources, including stellar core collapse. This will first require a significant overhaul of the code to make the speed improvements necessary for efficient and timely computing.

The methods presented in this thesis are valuable tools going forward. With the recent observations of gravitational waves from binary black hole mergers, it is an exciting time for the fields of astronomy and cosmology. New detections are imminent and could lead to new insights about the cosmos. Robust parameter estimation and noise modelling routines, such as the methods presented in this thesis, are now more important than ever.

Bibliography

- Aasi, J. et al. (2013). Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network. *Physical Review D*, 88:062001.
- Aasi, J. et al. (2015). Advanced LIGO. *Classical and Quantum Gravity*, 32:074001.
- Abadie, J. et al. (2010). Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. *Classical and Quantum Gravity*, 27:173001.
- Abadie, J. et al. (2012a). All-sky search for gravitational-wave bursts in the second joint LIGO-Virgo run. *Physical Review D*, 85:122007.
- Abadie, J. et al. (2012b). Sensitivity achieved by the LIGO and Virgo gravitational wave detectors during LIGO's sixth and Virgo's second and third science runs. *Pre-print*, arXiv:1203.2674 [gr-qc].
- Abbott, B. P. et al. (2009a). Einstein@Home search for periodic gravitational waves in early S5 LIGO data. *Physical Review D*, 80:042003.
- Abbott, B. P. et al. (2009b). LIGO: The Laser Interferometer Gravitational-Wave Observatory. *Reports on Progress in Physics*, 72:076901.
- Abbott, B. P. et al. (2009c). An upper limit on the stochastic gravitational-wave background of cosmological origin. *Nature*, 460:990–994.
- Abbott, B. P. et al. (2016a). All-sky search for long-duration gravitational wave transients with initial LIGO. *Physical Review D*, 93:042005.
- Abbott, B. P. et al. (2016b). Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Classical and Quantum Gravity*, 33:134001.
- Abbott, B. P. et al. (2016c). GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Physical Review D*, 93:122003.
- Abbott, B. P. et al. (2016d). GW151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Physical Review Letters*, 116:241103.

- Abbott, B. P. et al. (2016e). Localization and broadband follow-up of the gravitational-wave transient GW150914. *The Astrophysical Journal Letters*, 826:L13.
- Abbott, B. P. et al. (2016f). Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116:061102.
- Abbott, B. P. et al. (2016g). Observing gravitational-wave transient GW150914 with minimal assumptions. *Physical Review D*, 93:122004.
- Abbott, B. P. et al. (2016h). Properties of the binary black hole merger GW150914. *Physical Review Letters*, 116:241102.
- Abbott, B. P. et al. (2016i). Upper limits on the rates of binary neutron star and neutron-star-black-hole mergers from Advanced LIGO’s first observing run. *The Astrophysical Journal Letters*, 832:L21.
- Abbott, B. P. et al. (2017). GW170104: Observation of a 50-solar-mass binary black hole coalescence at redshift 0.2. *Physical Review Letters*, 118:221101.
- Abdikamalov, E., Gossan, S., DeMaio, A. M., and Ott, C. D. (2014). Measuring the angular momentum distribution in core-collapse supernova progenitors with gravitational waves. *Physical Review D*, 90:044001.
- Accadia, T. et al. (2011). Status of the Virgo project. *Classical and Quantum Gravity*, 28:114002.
- Acernese, F. et al. (2015). Advanced Virgo: A second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32:024001.
- Adams, S. M., Kochanek, C. S., Beacom, J. F., Vagins, M. R., and Stanek, K. Z. (2013). Observing the next galactic supernova. *The Astrophysical Journal*, 778:164.
- Agresti, A. (2002). *Categorical Data Analysis*. Wiley, Hoboken, New Jersey, 2nd edition.
- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.
- Amaro-Seoane, P. et al. (2016). Doing science with eLISA: Astrophysics and cosmology in the millihertz regime. *Pre-print*, arXiv:1201.361v1 [astro-ph.CO].
- Andersson, N. and Comer, G. L. (2001). Probing neutron-star superfluidity with gravitational-wave data. *Physical Review Letters*, 87:241101.
- Ando, M. (2015). Current status of the TAMA300 gravitational-wave detector. *Classical and Quantum Gravity*, 22:S881–889.

- Armano, M. et al. (2016). Sub-femto- g free fall for space-based gravitational wave observatories: LISA Pathfinder results. *Physical Review Letters*, 116:231101.
- Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, 37:1–16.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418.
- Bethe, H. A. and Wilson, J. R. (1985). Revival of a stalled supernova shock by neutrino heating. *The Astrophysical Journal*, 295:14–23.
- Brewer, B. J. and Donovan, C. P. (2015). Fast Bayesian inference for exoplanet discovery in radial velocity data. *Monthly Notices of the Royal Astronomical Society*, 448:3206–3214.
- Brewer, B. J., Huijser, D., and Lewis, G. F. (2016). Trans-dimensional Bayesian inference for gravitational lens substructures. *Monthly Notices of the Royal Astronomical Society*, 455:1819–1829.
- Brewer, B. J., Pártay, L. B., and Csányi, G. (2011). Diffusive nested sampling. *Statistics and Computing*, 21:649–656.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York, 2nd edition.
- Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65:3–55.
- Butterworth, S. (1930). On the theory of filter amplifiers. *Wireless Engineer*, 7:536–541.
- Cai, B. and Meyer, R. (2011). Bayesian semiparametric modeling of survival data based on mixtures of B-spline distributions. *Computational Statistics and Data Analysis*, 55:1260–1272.
- Cannon, K. et al. (2012). Toward early-warning detection of gravitational waves from compact binary coalescence. *The Astrophysical Journal*, 748:136.
- Carter, C. K. and Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Society: Series B (Statistical Methodology)*, 59:255–268.
- Chopin, N., Rousseau, J., and Liseo, B. (2013). Computational aspects of Bayesian spectral density estimation. *Journal of Computational and Graphical Statistics*, 22:533–557.
- Choudhuri, N., Ghosal, S., and Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059.

- Christensen, N. (2010). LIGO S6 detector characterization studies. *Classical and Quantum Gravity*, 27:194010.
- Christensen, N., Dupuis, R. J., Woan, G., and Meyer, R. (2004a). Metropolis-Hastings algorithm for extracting periodic gravitational wave signals from laser interferometric detector data. *Physical Review D*, 70:022001.
- Christensen, N. and Meyer, R. (1998). Markov chain Monte Carlo methods for Bayesian gravitational radiation data analysis. *Physical Review D*, 58:082001.
- Christensen, N. and Meyer, R. (2001). Using Markov chain Monte Carlo methods for estimating parameters with gravitational radiation data. *Physical Review D*, 64:022001.
- Christensen, N., Meyer, R., Knox, L., and Luey, B. (2001). Bayesian methods for cosmological parameter estimation from cosmic microwave background measurements. *Classical and Quantum Gravity*, 18:2677–2688.
- Christensen, N., Meyer, R., and Libson, A. (2004b). A Metropolis-Hastings routine for estimating parameters from compact binary inspiral events with laser interferometric gravitational radiation data. *Classical and Quantum Gravity*, 21:317–330.
- Clark, J., Heng, I. S., Pitkin, M., and Woan, G. (2007). Evidence-based search method for gravitational waves from neutron star ring-downs. *Physical Review D*, 76:043003.
- Cogburn, R. and Davis, H. R. (1974). Periodic splines and spectral estimation. *The Annals of Statistics*, 2:1108–1126.
- Contreras-Cristán, A., Gutiérrez-Peña, E., and Walker, S. G. (2006). A note on Whittle’s likelihood. *Communications in Statistics — Simulation and Computation*, 35:857–875.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301.
- Cornish, N. J. and Littenberg, T. B. (2015). BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, 32:135012.
- Coughlin, M., Christensen, N., Gair, J., Kandhasamy, S., and Thrane, E. (2014). Method for estimation of gravitational-wave transient model parameters in frequency-time maps. *Classical and Quantum Gravity*, 31:165012.
- Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, 6:101–111.

- Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904.
- Cutler, C. (2002). Gravitational waves from neutron stars with large toroidal B-fields. *Physical Review D*, 66:084025.
- Dal Canton, T. et al. (2014). Implementing a search for aligned-spin neutron star - black hole systems with advanced ground based gravitational wave detectors. *Physical Review D*, 90:082004.
- Damour, T. and Vilenkin, A. (2005). Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. *Physical Review D*, 71:063510.
- de Boor, C. (1993). *B(asic)-spline basics*, in: L. Piegl (Ed.), *Fundamental developments of computer-aided geometric modeling*. Academic Press, Washington, DC.
- Dimmelmeier, H., Ott, C. D., Marek, A., and Janka, H.-T. (2008). The gravitational wave burst signal from core collapse of rotating stars. *Physical Review D*, 78:064056.
- Earl, D. J. and Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7:3910–3916.
- Edwards, M. C., Meyer, R., and Christensen, N. (2014). Bayesian parameter estimation of core collapse supernovae using gravitational wave signals. *Inverse Problems*, 30:114008.
- Edwards, M. C., Meyer, R., and Christensen, N. (2015). Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis. *Physical Review D*, 92:064011.
- Edwards, M. C., Meyer, R., and Christensen, N. (2017a). Bayesian nonparametric spectral density estimation using B-spline priors. *Pre-print*, arXiv:1707.04878 [stat.CO].
- Edwards, M. C., Meyer, R., and Christensen, N. (2017b). bsplinePsd: Bayesian nonparametric spectral density estimation using B-spline priors. R package.
- Einstein, A. (1916). Approximative integration of the field equations of gravitation. *Sitzungsberichte Preußischen Akademie der Wissenschaften*, 1916 (Part 1):688–696.
- Engels, W. J., Frey, R., and Ott, C. D. (2014). Multivariate regression analysis of gravitational waves from rotating core collapse. *Physical Review D*, 90:124026.
- Essick, R., Vitale, S., Katsavoundis, E., Vedovato, G., and Klimenko, S. (2015). Localization of short duration gravitational-wave transients with the early Advanced LIGO and Virgo detectors. *The Astrophysical Journal*, 800:81.

- Fairhurst, S. (2014). Improved source localization with LIGO-India. *Journal of Physics: Conference Series*, 484:012007.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.
- Finn, L. S. and Mukherjee, S. (2001). Data conditioning for gravitational wave detectors: A Kalman filter for regressing suspension violin modes. *Physical Review D*, 63:062004.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.
- Gangopadhyay, A. K., Mallick, B. K., and Denison, D. G. T. (1999). Estimation of the spectral density of a stationary time series via an asymptotic representation of the periodogram. *Journal of Statistical Planning and Inference*, 75:281–290.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall / CRC, Boca Raton, FL, 3rd edition.
- Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian statistics*, chapter 5, pages 599–607. Oxford University Press.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.
- Gershman, S. J. and Blei, D. M. (2011). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Chapman and Hall, London.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.
- Gregory, P. (2005). *Bayesian logical data analysis for the physical sciences*. Cambridge University Press, Cambridge.
- Grote, H. (2010). The GEO 600 status. *Classical and Quantum Gravity*, 27:084003.
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66:51–83.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.
- Heng, I. S. (2009). Rotating stellar core-collapse waveform decomposition: A principal component analysis approach. *Classical and Quantum Gravity*, 26:105005.
- Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press, Cambridge, 1st edition.
- Hobbs, G. et al. (2010). The International Pulsar Timing Array project: Using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27:084013.
- Hulse, R. A. and Taylor, J. (1975). Discovery of a pulsar in a binary system. *The Astrophysical Journal*, 195:L51–L53.
- Huppenkothen, D., Brewer, B. J., Hogg, D. W., Murray, I., Frean, M., Elenbaas, C., Watts, A. L., Levin, Y., van der Horst, A. J., and Kouveliotou, C. (2015). Dissecting magnetar variability with Bayesian hierarchical models. *The Astrophysical Journal*, 810:66.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press, Cambridge.
- Kirch, C., Edwards, M. C., Meier, A., and Meyer, R. (2017). Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis. *Pre-print*, arXiv:1701.04846v1 [stat.ME].
- Klimenko, S., Yakushin, I., Mercer, A., and Mitselmakher, G. (2008). Coherent method for detection of gravitational wave bursts. *Classical and Quantum Gravity*, 25:114029.
- Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995). Rate of convergence for logspline spectral density estimation. *Journal of Time Series Analysis*, 16:389–401.
- Kowalska-Leszczynska, I. et al. (2016). Globally coherent short duration magnetic field transients and their effect on ground based gravitational-wave detectors. *Pre-print*, arXiv:1612.01102v1 [astro-ph.IM].
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.
- Laplace, P. S. (1812). *Theorie Analytiques des Probabilités*. Courcier, Paris.
- Lattimer, J. M. and Swesty, F. D. (1991). A generalized equation of state for hot, dense matter. *Nuclear Physics A*, 535:331–376.

- Lenhoff, M. W., Santner, T. J., Otis, J. C., Peterson, M. G., Williams, B. J., and Backus, S. I. (1999). Bootstrap prediction and confidence bands: A superior statistical method for the analysis of gait data. *Gait and Posture*, 9:10–17.
- LIGO. Accessed: 8 February 2017.
- Liseo, B., Marinucci, D., and Petrella, L. (2001). Bayesian semiparametric inference on long-range dependence. *Biometrika*, 88:1089–1104.
- Littenberg, T. B. and Cornish, N. J. (2010). Separating gravitational wave signals from instrument artifacts. *Physical Review D*, 82:103007.
- Littenberg, T. B. and Cornish, N. J. (2015). Bayesian inference for spectral estimation of gravitational wave detector noise. *Physical Review D*, 91:084034.
- Littenberg, T. B., Coughlin, M., Farr, B., and M., F. W. (2013). Fortifying the characterization of binary mergers in LIGO data. *Physical Review D*, 88:084044.
- Logue, J., Ott, C. D., Heng, I. S., Kalmus, P., and Scargill, J. H. C. (2012). Inferring core-collapse supernova physics with gravitational waves. *Physical Review D*, 86:044023.
- Loredo, T. J. (1990). From Laplace to Supernova SN 1987A: Bayesian inference in astrophysics. In Fougère, P. F., editor, *Maximum Entropy and Bayesian Methods*, pages 81–142. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Loredo, T. J. (1992). The promise of Bayesian inference for astrophysics. In Fiegelson, E. D. and Babu, G. J., editors, *Statistical Challenges in Modern Astronomy*, chapter 12, pages 275–297. Springer.
- Lynch, R., Vitale, S., Essick, R., and Katsavounidis, E. (2016). An information-theoretic approach to the gravitational-wave burst detection problem. *Pre-print*, arXiv:1511.05955v2 [gr-qc].
- Macaro, C. (2010). Bayesian non-parametric signal extraction for Gaussian time series. *Journal of Econometrics*, 157:381–395.
- Macaro, C. and Prado, R. (2014). Spectral decompositions of multiple time series: A Bayesian non-parametric approach. *Psychometrika*, 79:105–129.
- Maggiore, M. (2008). *Gravitational Waves: Volume 1 Theory and Experiments*. Oxford University Press, 1st edition.
- Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42(9):1–21.

- Meier, A., Kirch, C., Edwards, M. C., and Meyer, R. (2017). beyondWhittle: Bayesian spectral inference for stationary time series. R package.
- Mereghetti, S. (2008). The strongest cosmic magnets: Soft gamma-ray repeaters and anomalous X-ray pulsars. *The Astronomy and Astrophysics Review*, 15:225–287.
- Meszáros, P. (2006). Gamma-ray bursts. *Reports on Progress in Physics*, 69:2259–2322.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M., Teller, A. H., and E., T. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31:705–676.
- Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov chain Monte Carlo*, chapter 5, pages 113–162. Chapman and Hall / CRC Press.
- Neumann, M. H. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric regression. *The Annals of Statistics*, 26:1570–1613.
- Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9:307–333.
- Nuttall, L. K. et al. (2015). Improving the data quality of Advanced LIGO based on early engineering run results. *Classical and Quantum Gravity*, 32:245005.
- Oppenheim, A. V., Schaffer, R. W., and Buck, J. R. (1999). *Discrete-time Signal Processing*. Prentice Hall.
- Ott, C. D., Abdikamalov, E., O’Conner, E., Reisswig, C., Haas, R., Kalmus, P., Drasco, S., Burrows, A., and Schnetter, E. (2012). Correlated gravitational wave and neutrino signals from general-relativistic rapidly rotating iron core collapse. *Physical Review Letters*, 86:024026.
- Ott, C. D., Burrows, A., Livne, E., and Walder, R. (2004). Gravitational waves from axisymmetric rotating stellar core collapse. *The Astrophysical Journal*, 600:834–867.
- Perron, F. and Mengersen, K. (2001). Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics*, 57:518–528.
- Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics*, 27:105–126.
- Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26:373–393.

- Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75:335–346.
- Powell, J., Gossan, S. E., Logue, J., and Heng, I. S. (2016). Inferring the core-collapse supernova explosion mechanism with gravitational waves. *Physical Review D*, 94:123012.
- Powell, M. J. D. (1981). *Approximation theory and methods*. Cambridge University Press, Cambridge.
- Raymond, V., van der Sluys, M. V., Mandel, I., Kalogera, V., Röver, C., and Christensen, N. (2009). Degeneracies in sky localization determination from a spinning coalescing binary through gravitational wave observations: A Markov-chain Monte Carlo analysis for two detectors. *Classical and Quantum Gravity*, 26:114007.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. John Wiley and Sons Inc., 2nd edition.
- Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59:731–792.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, 1st edition.
- Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18:349–367.
- Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Methodological)*, 59:291–317.
- Rosen, O., Wood, S., and Roy, A. (2012). AdaptSpec: Adaptive spectral density estimation for nonstationary time series. *Journal of the American Statistical Association*, 107:1575–1589.
- Rousseau, J., Chopin, N., and Liseo, B. (2012). Bayesian nonparametric estimation of the spectral density of a long or intermediate memory Gaussian time series. *The Annals of Statistics*, 40:964–995.
- Röver, C. (2011). Student-t based filter for robust signal detection. *Physical Review D*, 84:122004.
- Röver, C., Bizouard, M.-A., Christensen, N., Dimmelmeier, H., Heng, I. S., and Meyer, R. (2009). Bayesian reconstruction of gravitational wave burst signals from simulations of rotating stellar core collapse and bounce. *Physical Review D*, 80:102004.

- Röver, C., Meyer, R., and Christensen, N. (2006). Bayesian inference on compact binary inspiral gravitational radiation signals in interferometric data. *Classical and Quantum Gravity*, 23:4895–4906.
- Röver, C., Meyer, R., and Christensen, N. (2007a). Coherent Bayesian analysis of inspiral signals. *Classical and Quantum Gravity*, 24:S607–S615.
- Röver, C., Meyer, R., and Christensen, N. (2007b). Coherent Bayesian inference of binary inspirals using a network of interferometric gravitational wave detectors. *Physical Review D*, 75:062004.
- Röver, C., Meyer, R., and Christensen, N. (2011). Modelling coloured residual noise in gravitational-wave signal processing. *Classical and Quantum Gravity*, 28:015010.
- Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71:599–607.
- Sathyaprakash, B. et al. (2012). Scientific objectives of Einstein Telescope. *Classical and Quantum Gravity*, 29:124013.
- Sathyaprakash, B. S. (2007). LALAdvLIGOPsd.c. Accessed: 8 February 2017.
- Saulson, P. R. (2011). Josh Goldberg and the physical reality of gravitational waves. *General Relativity and Gravitation*, 43:3289–3299.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.
- Shao, X. and Wu, B. W. (2007). Asymptotic spectral theory for nonlinear time series. *Annals of Statistics*, 35:1773–1801.
- Shen, H., Toki, H., Oyamatsu, K., and Sumiyoshi, K. (1998). Relativistic equation of state of nuclear matter for supernova and neutron star. *Nuclear Physics A*, 637:435–450.
- Shen, W. and Ghosal, S. (2014). Adaptive Bayesian procedures using random series priors. *Preprint*, arXiv:1403.0625 [math.ST].
- Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *The Annals of Statistics*, 29:687–714.
- Shibata, M., Liu, Y. T., Shapiro, S. L., and Stephens, B. C. (2006). Magnetorotational collapse of massive stellar cores to neutron stars: Simulations in full general relativity. *Physical Review D*, 74:104026.
- Sintes, A. M. and Schutz, B. F. (1998). Coherent line removal: Filtering out harmonic related line interference from experimental data, with application to gravitational wave detectors. *Physical Review D*, 58:122003.

- Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1:833–860.
- Somiya, K. (2012). Detector configuration of KAGRA — the Japanese cryogenic gravitational-wave detector. *Classical and Quantum Gravity*, 29:124007.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Society: Series B (Statistical Methodology)*, 64:583–639.
- Stone, M. and Dawid, A. P. (1972). Un-Bayesian implications of improper Bayes inference in routine statistical problems. *Biometrika*, 59:369–375.
- Summerscales, T. Z., Burrows, A., Finn, L. S., and Ott, C. D. (2008). Maximum entropy for gravitational wave data analysis: Recovering the physical parameters from core-collapse supernovae. *The Astrophysical Journal*, 678:1142–1157.
- Sun, J. and Loader, C. R. (1994). Confidence bands for linear regression and smoothing. *The Annals of Statistics*, 22:1328–1345.
- Sutton, P. J., Jones, G., Chatterji, S., Kalmus, P., Leonor, I., Poprocki, S., Rollins, J., Searle, A., Stein, L., Tinto, M., and Was, M. (2010). X-Pipeline: An analysis package for autonomous gravitational-wave burst searches. *New Journal of Physics*, 12:053034.
- Swendsen, R. H. and Wang, J. S. (1986). Replica Monte Carlo simulation of spin glasses. *Physical Review Letters*, 57:2607–2609.
- Taylor, J. and Weisburg, J. (1989). Further experimental tests of relativistic gravity using the binary pulsar PSR 1913 + 16. *The Astrophysical Journal*, 345:434–450.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70:1055–1096.
- Thorne, K. S. (1987). Gravitational radiation. In Hawking, S. W. and Israel, W., editors, *300 Years of Gravitation*, chapter 9, pages 330–358. Cambridge University Press.
- Thrane, E., Kandhasamy, S., Ott, C. D., Anderson, W. G., Christensen, N. L., Coughlin, M. W., Dorsher, S., Giampanis, S., Mandic, V., Mytidis, A., Prestegard, T., Raffai, P., and Whiting, B. (2011). Long gravitational-wave transients and associated detection strategies for a network of terrestrial interferometers. *Physical Review D*, 83:083004.
- Tonellato, S. F. (2007). Random field priors for spectral density functions. *Journal of Statistical Planning and Inference*, 137:3164–3176.

- Turin, G. L. (1960). An introduction to matched filters. *IRE Transactions on Information Theory*, 6:311–329.
- Umstätter, R., Meyer, R., Dupuis, R. J., Veitch, J., Woan, G., and Christensen, N. (2004). Estimating the parameters of gravitational waves from neutron stars using an adaptive MCMC method. *Classical and Quantum Gravity*, 21:S1655–S1665.
- Unnikrishnan, C. S. (2013). IndIGO and LIGO-India: Scope and plans for gravitational wave research and precision metrology in India. *International Journal of Modern Physics D*, 22:1341010.
- van der Sluys, M. V., Röver, C., Stroerer, A., Raymond, V., Mandel, I., Christensen, N., Kalogera, V., Meyer, R., and Vecchio, A. (2008). Gravitational-wave astronomy with inspiral signals of spinning compact-object binaries. *The Astrophysical Journal*, 688:L61–L64.
- Veitch, J. et al. (2015). Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Physical Review D*, 91:042003.
- Veitch, J. and Vecchio, A. (2010). Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. *Physical Review D*, 81:062003.
- Vitale, S. et al. (2014). Data series subtraction with unknown and unmodeled background noise. *Physical Review D*, 90:042003.
- Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association*, 75:122–132.
- Walmswell, J. J., Eldridge, J. J., Brewer, B. J., and Tout, C. A. (2013). A transdimensional Bayesian method to infer the star formation history of resolved stellar populations. *Monthly Notices of the Royal Astronomical Society*, 435:2171–2186.
- Weber, J. (1967). Gravitational radiation. *Physical Review Letters*, 18:498–501.
- Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73.
- Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 19:38–63.
- Woosley, S. E. and Heger, A. (2007). Nucleosynthesis and remnants in massive stars of solar metallicity. *Physical Reports*, 442:269–283.

Zheng, Y., Zhu, J., and Roy, A. (2010). Nonparametric Bayesian inference for the spectral density function of a random field. *Biometrika*, 97:238–245.