

Physical modelling of galaxy clusters and Bayesian inference in astrophysics

Kamran Javid

Fitzwilliam College



May 2018

A dissertation submitted for the degree of Doctor of Philosophy

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

SUMMARY

This thesis is concerned with the modelling of galaxy clusters, applying these models to real and simulated data using Bayesian inference, and the development of Bayesian inference algorithms applicable to a wide range of astrophysical problems.

I present a comparison of mass estimates for 54 galaxy cluster candidates from the second Planck catalogue (PSZ2) of Sunyaev–Zel’dovich sources. I compare the mass values obtained with data taken from the Arcminute Microkelvin Imager (AMI) radio interferometer system and from the Planck satellite. The former of these uses a Bayesian analysis pipeline that parameterises a cluster in terms of its physical quantities, and models the dark matter & baryonic components of a cluster using Navarro-Frenk-White (NFW) and generalised-NFW profiles respectively. The mass estimates derived from Planck data are obtained from the results of the Bayesian detection algorithm PowellSnakes (PwS). I also analyse simulated AMI data with input values based on PwS mass estimates.

I then compare three cluster models using AMI data for the 54 cluster sample. The two observational models considered only model the gas content of the cluster. To compare the physical and observational models I consider their posterior parameter estimates, including the calculation of a metric defined between two probability distributions. The models’ fit to the cluster data is evaluated by looking at the Bayesian evidence values.

Improvements to the physical modelling of galaxy clusters are then considered, either by relaxing some of the assumptions underlying the physical model, or by introducing a new profile for the dark matter component of clusters. The resultant models are compared with the physical model introduced previously.

The final part of the cluster analysis work focuses on Bayesian analysis using a joint likelihood function of data from both AMI and the Planck satellite simultaneously. The results of this joint analysis are compared with those obtained from the individual likelihood analyses using simulated data and with real data taken from the 54 cluster sample.

Finally, a new Bayesian inference algorithm based on nested sampling is presented. The algorithm, named the "geometric nested sampler", is an adaption of the Metropolis-Hastings nested sampler and makes use of the geometrical interpretation of sets of parameters to sample from their domains efficiently. The geometric nested sampler is tested on several toy models as well as a model representing the emission of gravitational waves from binary black hole mergers. The results obtained using the geometric nested sampler are compared with those from popular nested sampling algorithms.

ACKNOWLEDGEMENTS

Completing my PhD has been a bit of a roller coaster ride, and a journey that has been full of helpful and supportive people.

First of all I would like to thank my primary supervisor, Richard Saunders. Your support and attention through a number of issues including ones I couldn't have imagined experiencing before starting my PhD was wonderful. Your thoroughness and attention to detail have changed the way I approach problems and I am eternally grateful for it. You have been family-like figure during my time at Cambridge as well as my work colleague, and I know that our friendship will continue beyond Cambridge.

A massive thank you to my secondary supervisor Yvette Perrott, who has done a fabulous job of providing me support, and giving direction to my research. Your understanding of the underlying astrophysics associated with my research, ability to give recommendations and ways to improve my work will be very much missed post-Cambridge.

Another massive thank you to Will Handley for his support while we collaborated on the nested sampling project. You are a fantastic supervisor of research, and I know your career will go very far. You are a pleasure to work with and have helped me confirm that statistics is one of my true passions.

I would also like to thank everyone else who I have collaborated with during my PhD: Clare Rumsey, Pedro Carvalho, Keith Grainge, Mike Hobson, Anthony Lasenby, Farhan Feroz, and others. A special thanks goes to Malak Olamaie who was incredibly helpful as a stand-in supervisor during the early stages of my PhD while Yvette was on maternity leave.

Thanks to Dave Green for his invaluable L^AT_EX support while writing my first year report, research papers, and my thesis. I would also like to thank Dave Titterington and Greg Willatt from the Cavendish Astrophysics IT team, and Stuart Rankin from the Cambridge High Performance Computing team. Your impeccable services and support have kept the 'experimental equipment' of my PhD in tip-top shape, and so I am truly indebted to you.

A special shout out to the University food halls across the city, who have done a fantastic job of the not so easy task of keeping my belly full the past few years. In particular, Hughes Hall did an exquisite job of appreciating my love for food by giving me over-sized, nutritious and tasty portions.

I would also like to thank the students I have supervised at Trinity College during my PhD, you have provided a mental workout like no other. You are undoubtedly some of the smartest people I will ever meet, and it was a pleasure to get an insight into your way of thinking.

To my friends in the astrophysics office: Terry Jin, Maximilian Günther, Marion Neveu, Richard Hall, Zoe Ye and Julia Riley. Thank you for providing fun and support in one of the places I needed it the most!

To my friends in Sheffield, Cambridge, and London, a big thank you for helping me keep sane during the last few years with our fun adventures together. A special thanks to Rob Hull for teaching me how to electronically draw images of three-dimensional shapes neatly for the nested sampling project!

The Mathematics department at Wales High School deserves a lot of credit for getting me interested in mathematics in the first place. Without your passion and enthusiasm for the subject, I doubt I would have gone to university at all. I am eternally grateful and I hope to re-pay you one day by helping to inspire our next generation of students to love the subject.

Saving the best until last, I would like to thank my family. Without your love there's no way I would have finished this PhD. Everything I do, I do it to make you proud of me. You are my drive and I will strive to make you proud for the rest of my days. To my mother, given we're a single-parent, only-child family, I'm sure no one would have expected us getting this far. Putting a smile on your face and making you proud is worth a lifetime of my blood, sweat and tears.

CONTENTS

Contents	ix	
1	Introduction	1
1.1	Galaxy clusters	1
1.2	The Sunyaev–Zel’dovich effect	1
1.3	Planck mission	2
1.4	AMI	4
1.5	Remainder of this thesis	4
1.6	Conventions	5
2	Introductory theory	7
2.1	Interferometry	7
2.2	Measuring the SZ effect with an interferometer	10
2.3	Cluster model selection	13
2.4	A physical model for AMI data	13
2.5	Recognised radio-sources and general noise contributions	19
2.6	Bayesian inference	21
2.7	Parameter prior distributions	23
2.8	The likelihood function	24
3	Physical modelling of clusters detected by Planck	27
3.1	Selection and observation of the cluster sample	27
3.2	AMI data analysis	29
3.3	PSZ2 redshift values	30
3.4	PSZ2 methodology for deriving cluster mass estimates	30
3.5	Obtaining AMI mass estimates	33
3.6	AMI and PSZ2 mass estimates	38
3.7	AMI simulations with PSZ2 mass inputs	40
3.8	Conclusions	51
4	Comparison of physical and observational galaxy cluster modelling using AMI data	53
4.1	Physical model estimates of observational parameters	53
4.2	Observational models	54

4.3	AMI model comparisons	56
4.4	Conclusions	65
5	Physical modelling of galaxy clusters using Einasto dark matter profiles	67
5.1	Einasto physical model	68
5.2	Results	73
5.3	Conclusions	82
6	Enhanced physical modelling I: relaxing the f_{gas} assumption	85
6.1	Total mass equations	85
6.2	Determining cluster profile parameters	86
6.3	Mass profile plots	88
6.4	Conclusions	89
7	Enhanced physical modelling II: Inclusion of non-thermal pressure	93
7.1	Non-thermal gas pressure	94
7.2	Modelling non-thermal gas pressure	94
7.3	Non-thermal pressure profiling	100
7.4	Conclusions	104
8	Joint analysis of AMI and Planck data	107
8.1	Joint likelihood analysis	108
8.2	Joint likelihood analysis hyperparameters	110
8.3	Cluster models	113
8.4	Cluster simulations	115
8.5	Cluster simulation results	116
8.6	Application of joint analysis to real cluster data	127
8.7	Conclusions	129
9	Monte Carlo sampling methods	135
9.1	Inverse transform sampling	135
9.2	Rejection sampling	136
9.3	Markov Chain Monte Carlo sampling	137
9.4	Nested sampling	141
9.5	Plotting posterior samples	146
10	Geometric nested sampling	149
10.1	Nested sampling prior distributions	149
10.2	Satisfying the likelihood constraint	150
10.3	Geometric nested sampling	151
10.4	Applications of geometric nested sampling	163
10.5	Geometric nested sampling implementation	175
10.6	Conclusions	178
A	Appendix A: Results of physical modelling of Planck clusters	181

A.1	Results table	182
B	Appendix B: Y_{tot} derivation and results of physical and observational modelling comparison	185
B.1	GNFW Y_{tot} analytical solution	185
B.2	Results tables	187
C	Appendix C: Einasto model derivations and results of PM I and PM II comparisons	193
C.1	Einasto mass integral	193
C.2	Determining r_{500} iteratively	194
C.3	Simulations results table	196
D	Appendix D: Supplementary statistical derivations and results	197
D.1	Metropolis-Hastings and detailed balance	197
D.2	Evidence integral transformation	199
D.3	Probability density function of the largest of n numbers	200
D.4	Determining $\log(\mathbb{E}[\mathcal{Z}])$ and $\log(\mathbb{E}[\mathcal{Z}^2])$ from $\log(\mathcal{L})$ and $\log(X)$.	201
References		203

INTRODUCTION

1.1 Galaxy clusters

In the local Universe and out to redshifts of around two, clusters of galaxies are observed as massive gravitationally bound structures, often roughly spherical and with very dense central cores (see reviews by e.g. [Rosati, Borgani, & Norman 2002](#), [Voit 2005](#), [Allen, Evrard, & Mantz 2011](#), and [Giodini et al. 2013](#)). It is over eighty years ago that it was first postulated that a galaxy cluster's mass is dominated by dark matter ([Zwicky 1933](#) and [Zwicky 1937](#)). More recently it has been shown that dark matter contributes $\approx 90\%$ of the cluster mass (see e.g. [Vikhlinin et al. 2006](#) and [Komatsu et al. 2011](#)). Stars, gas and dust in galaxies, as well as a hot ionised intra-cluster medium (ICM) make up the rest of the mass in a cluster, with the latter being the most massive baryonic component. The galaxies emit in the optical and infrared wavebands, whilst the ICM emits in X-ray via thermal Bremsstrahlung and also interacts with cosmic microwave background (CMB) photons via inverse Compton scattering. This last effect is what is known as the Sunyaev–Zel'dovich (SZ) effect ([Sunyaev & Zeldovich 1970](#)).

1.2 The Sunyaev–Zel'dovich effect

The SZ effect is particularly strong in the cluster ICM, where temperatures range between 10^7 – 10^8 K. The nature of the CMB spectrum means that the effect leads to an increase in intensity at

frequencies above 217 GHz and a decrease for frequencies below (Figure 1.1). The measurement of the SZ surface brightness increment / decrement has the crucial characteristic that it is redshift independent (see Section 2.2). The SZ effect has the additional advantage over X-ray analysis, that it only depends on the electron number density linearly (see Section 2.2), whereas X-ray Bremsstrahlung emission is proportional to electron number density squared. This means that SZ can in practice be used to analyse a cluster at higher radius. The Planck telescope (Section 1.3) and the Arcminute Microkelvin Imager radio interferometer system (AMI, see Section 1.4) both observe galaxy clusters by measuring the SZ effect.

1.3 Planck mission

The Planck mission^a is a European Space Agency (ESA) mission, whose primary objective is to investigate the CMB. The Planck telescope was a space telescope which was launched in May 2009 and deactivated in October 2013. The combination of Planck's low-frequency and high-frequency instruments (LFI and HFI) provides nine frequency channels in the range 37 GHz – 857 GHz. The LFI has angular resolutions of 33, 24, and 14 arcminutes at respective frequencies of 30, 44, and 70 GHz. The HFI has angular resolutions of 10, 7.1, and 5.5 arcminutes at 100, 143, and 217 GHz and 5.0 arcminutes at each of 353, 545, and 857 GHz. For more information on the Planck telescope I refer the reader to the Scientific Programme of Planck ([The Planck Collaboration 2006](#)). In addition to all-sky coverage, Planck has its own advantages for SZ work: a very wide range of frequency channels, polarisation capability, and a channel at the 217-GHz null frequency of SZ all help to remove contamination from synchrotron, Bremsstrahlung and dust emissions. Of particular importance for the work described here are the Planck cluster-catalogues (see [Planck Collaboration XXIX 2014](#), [Planck Collaboration XXXII 2015](#) and [Planck Collaboration XXVII 2016](#) for papers relating to catalogues PSZ1, PSZ1.2 and PSZ2 respectively, where 'PSZX' refers to the Xth Planck SZ catalogue). These provide e.g. cluster candidate positions, redshift (z) values (see Section 3.3), integrated Comptonisation parameter (Y) values and mass (M) estimates. PSZ2 is the most recent all-sky Planck cluster catalogue, and is the one which I refer to unless stated otherwise.

^a<http://www.esa.int/Planck/>.

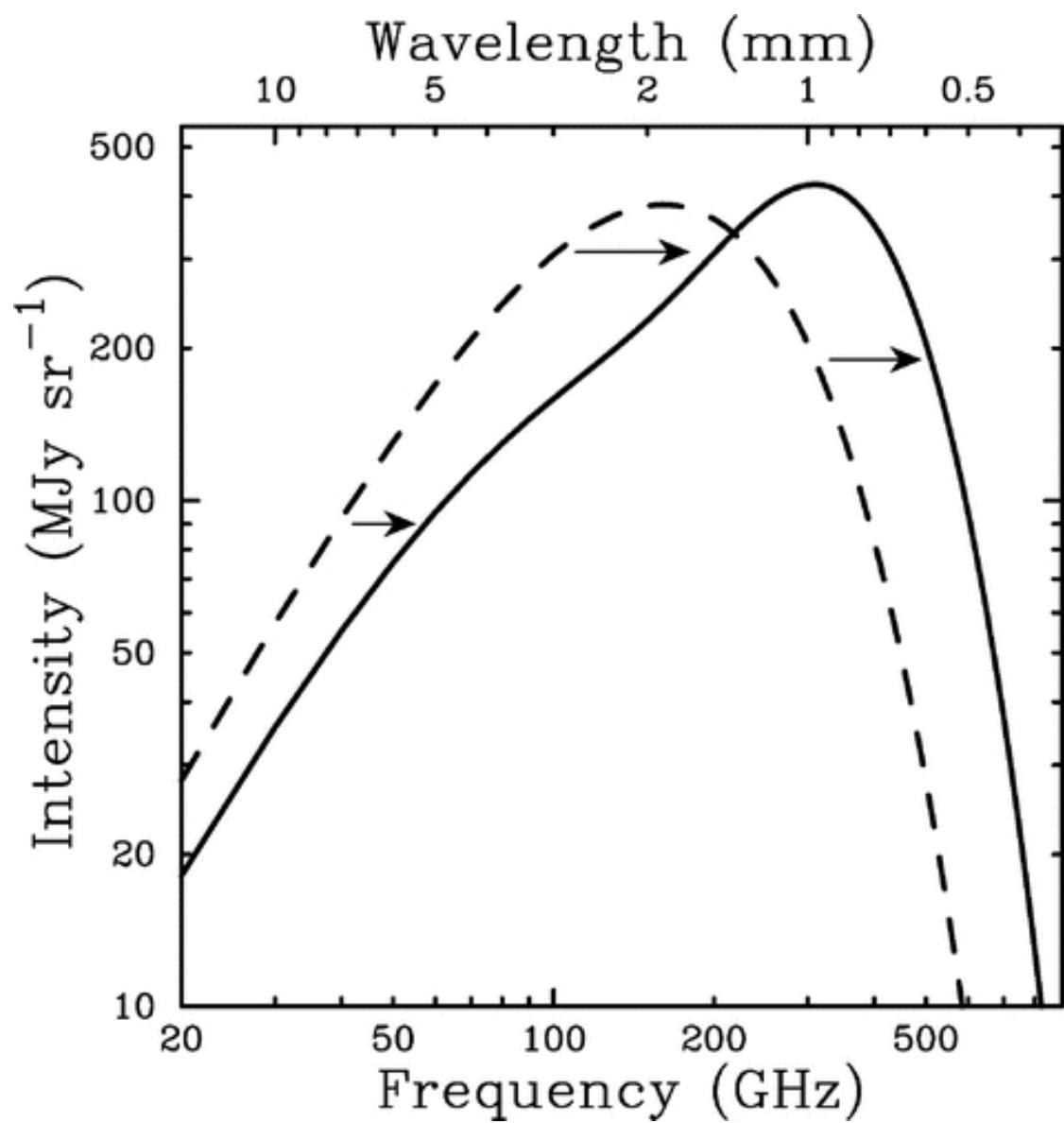


Figure 1.1: Radiation intensity as a function of frequency. Note the dashed line represents the incident radiation, whilst the solid line represents the energy-boosted inverse Compton scattered radiation. Taken from [Carlstrom, Holder, & Reese \(2002\)](#).

	SA	LA
Antenna diameter	3.7 m	12.8 m
Number of antennas	10	8
Baseline lengths (current)	5 – 20 m	18 – 110 m
Primary beam FWHM (at 15.7 GHz)	20.1 arcmin	5.5 arcmin
Typical synthesised beam FWHM	3 arcmin	30 arcsec
Flux sensitivity	30 mJy s ^{1/2}	3 mJy s ^{1/2}

Table 1.1: Summary of AMI characteristics. Both arrays measure the same linear polarisation.

1.4 AMI

AMI is an interferometer system near Cambridge, designed for SZ studies (see e.g. [Zwart et al. 2008](#)). It consists of two arrays: the Small Array (SA), optimised to couple to SZ signal, with an angular resolution of ≈ 3 arcmin and sensitivity to structures up to ≈ 10 arcmin in scale; and the Large Array (LA), with angular resolution of ≈ 30 arcsec, which is largely insensitive to SZ, and is used to characterise and subtract confusing radio-sources (see Section [2.5.1](#)). Both arrays operate at a central frequency of ≈ 15.7 GHz and, at the time the AMI data for this paper were taken, with a bandwidth of ≈ 4.3 GHz, divided into six channels. Both arrays actually operate over the wide frequency range of $\approx 12.0 – 18.0$ GHz for sensitivity, and the correlator splits this range into eight separate channels each approximately 0.72 GHz wide to reduce chromatic aberration over the fields of view to manageable levels. However, due to satellite interference at the lower end of the spectrum, data from the bottom two channels are excluded, giving the effective bandwidth of 4.3 GHz across six channels mentioned above). A summary of AMI’s characteristics is given in Table [1.1](#). More detail on AMI is given in Section [2.1.1](#). Note that AMI has recently received a new digital correlator ([Hickish et al. 2018](#)), but all data used in this thesis were obtained by the system with its analogue correlator.

1.5 Remainder of this thesis

In Chapter [2](#) I give an overview of the theory underlying various topics which are heavily relied upon throughout the thesis: interferometry, measuring the SZ effect, galaxy cluster modelling, and Bayesian inference.

In Chapter [3](#) I apply a cluster model to data from AMI of clusters detected by Planck, and compare the results with those obtained directly from Planck data. I also analyse simulated cluster

data whose inputs are based on the mass estimates obtained from Planck data, to see if AMI simulations & the cluster model are capable of inferring the correct cluster masses.

Chapter 4 presents the results of a cluster model comparison for the sample of 54 clusters considered in the previous Chapter; for the three models I compare the parameter estimates and Bayesian evidence values obtained for each cluster.

A new cluster model is presented in Chapter 5 which uses an Einasto profile to model the dark matter component of a cluster. By looking at cluster parameter profiles, and performing Bayesian analysis on simulated & real data, I compare the new model with the one presented in Chapter 2. Chapters 6 and 7 detail further attempts to enhance galaxy cluster modelling. I first try to relax the mass assumption associated with the models detailed in Chapters 2 and 5, and plot the resulting mass profiles for a range of clusters (Chapter 6). I then try to incorporate non-thermal pressure into the cluster models in Chapter 7, and plot the resultant parameter profiles.

In Chapter 8 I introduce a joint AMI-Planck analysis method, which revolves around evaluating the likelihood functions associated with each instrument simultaneously. I then present the results of this method applied to both simulated and real datasets, and compare with the results obtained from conducting the individual instrument analyses separately.

An overview of Monte Carlo sampling methods is given in Chapter 9. This includes an introduction to nested sampling, the method upon which the algorithm presented in Chapter 10 is based on. I also explain briefly how samples can be used to approximate the distribution from which they originate.

In Chapter 10 I provide the motivation & technical details of the nested sampling algorithm I have created and refer to as the "geometric nested sampler". I apply the algorithm to several toy models & to an astrophysical application (detecting gravitational waves from a black hole binary merger system), and compare its performance with pre-existing nested sampling algorithms.

1.6 Conventions

A ‘concordance’ flat Λ CDM cosmology is assumed: $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_R = 0$, $\Omega_K = 0$, $h = 0.7$, $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\sigma_8 = 0.8$, $w_0 = -1$, and $w_a = 0$. The first four parameters correspond to the (dark + baryonic) matter, the cosmological constant, the radiation, and the curvature densities respectively. h is the dimensionless Hubble parameter, while H_0 is the Hubble parameter now and σ_8 is the power spectrum normalisation on the scale of $8 h^{-1} \text{ Mpc}$ now. w_0 and w_a are the equation of state parameters of the Chevallier-Polarski-Linder parametrisation ([Chevallier & Polarski 2001](#)).

CHAPTER 2

INTRODUCTORY THEORY

2.1 Interferometry

In addition to the advantage of high angular resolution from long baselines, interferometers possess a number of advantages over single-dish telescopes, particularly for CMB work. Among these are their relative insensitivity to atmospheric emission (see e.g. [Watson et al. 2003](#)), the ease with which systematic errors such as ground spill ([Lay & Halverson 2000](#)) can be dealt with; and radio-source contamination (see e.g. [Grainger et al. 2002](#)) can be kept to a minimum. Furthermore, the angular sensitivity of an interferometer can be fine-tuned by adjusting baseline lengths.

To understand how an interferometer works, consider a two-antenna system similar to the one constructed by Ryle and Vonberg ([Ryle & Vonberg 1948](#)). Figure 2.1 shows two antennas on an east-west baseline of length b tracking a visible patch of sky which, initially, meets three conditions: (i) contains only one radio-source; (ii) this source is at the centre of the tracked patch; and (iii) this source is unresolved by the interferometer. At hour angle θ (as defined in Figure 2.1), the voltages V_1 and V_2 measured by each antenna at time t are

$$\begin{aligned} V_1 &= V_0 e^{i(\omega t + kb \sin \theta)}, \\ V_2 &= V_0 e^{i\omega t}, \end{aligned} \tag{2.1}$$

where V_0 is the signal voltage amplitude, ω is the angular frequency of the source radiation

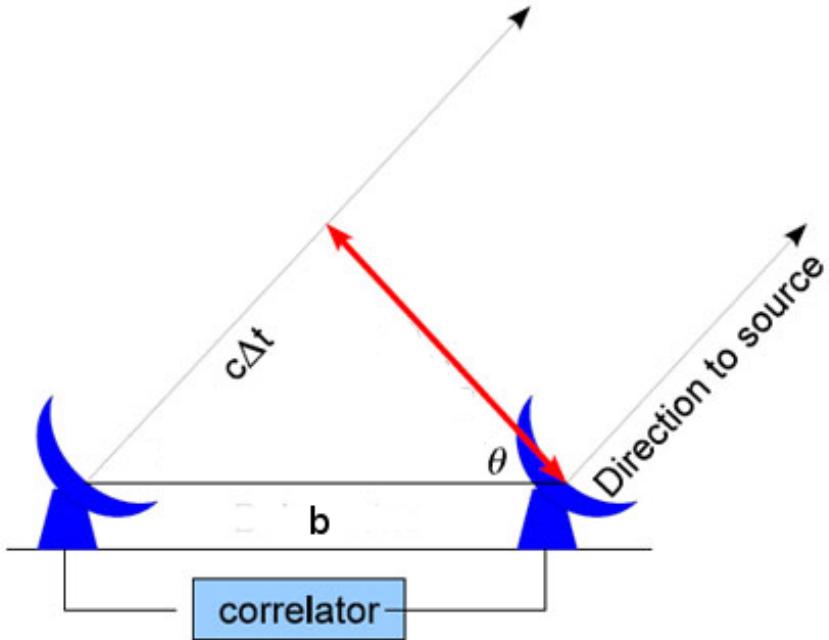


Figure 2.1: Simple east-west single baseline interferometer tracking a patch of sky containing a single radio-source. For a baseline b and a source at angle θ from the vertical axis, the wavefront has to travel an additional distance $b \sin \theta$ to the further antenna. Image taken from Zaroubi (2013).

being observed and k is the corresponding wavenumber. The ωt -dependent parts are removed and the correlator multiplies the remaining components of 2.1 together to give a response, termed visibility, proportional to

$$e^{ikb \sin \theta}, \quad (2.2)$$

in which the constant of proportionality (including V_0^2 , the effects of integration time, dish area and so on) which in practice is evaluated by observation of a bright, unresolved radio-source with well known properties. Unless the observing bandwidth $\Delta\omega$ is very low (and thus the coherence length $2\pi c/\omega$ is very long), the baseline must be ‘phased up’ by inserting an additional path equivalent to $b \sin \theta$ into the interferometer arm which the radiation hits first. This compensates for the extra path $c\Delta t = b \sin \theta$ involved in the other arm.

We now relax condition (ii). If the source is offset from the pointing centre by an angle α , the extra path becomes $b \sin(\theta + \alpha)$. The path compensation is set for the pointing centre so that

multiplying the equivalent expressions of 2.1 now gives a visibility proportional to

$$\begin{aligned} & e^{ikb(\sin(\theta+\alpha)-\sin\theta)} \\ &= e^{ikb\alpha \cos\theta}, \end{aligned} \quad (2.3)$$

using the small angle approximation for α .

We now relax condition (iii). The response to a source, which has a top-hat surface-brightness distribution in α -space of width $\Delta\alpha$ and centred on α , is averaged over $\Delta\alpha$, giving a response proportional to

$$\begin{aligned} & \frac{1}{\Delta\alpha} \int_{-\frac{\Delta\alpha}{2}}^{\frac{\Delta\alpha}{2}} e^{ikb\alpha' \cos\theta} d\alpha' \\ &= \text{sinc}\left(\frac{kb\Delta\alpha \cos\theta}{2}\right) \end{aligned} \quad (2.4)$$

Thus sources with a large angular size on the sky ($\Delta\alpha \gg kb \cos\theta$) are resolved out by the interferometer since $\text{sinc}(x) \rightarrow 0$ as $x \rightarrow \infty$.

One can similarly examine the effect of the observing bandwidth. Repeating the above analysis for k gives

$$\begin{aligned} & \frac{1}{\Delta k} \int_{-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} e^{ik'b\alpha \cos\theta} dk' \\ &= \text{sinc}\left(\frac{\alpha b \Delta k \cos\theta}{2}\right). \end{aligned} \quad (2.5)$$

So a large enough bandwidth also causes the signal to fall, this time due to chromatic aberration. This explains the need for independent frequency channels which are a feature of AMI.

Finally, replacing condition (i) by a surface-brightness distribution $I(\theta, \alpha)$, and incorporating the primary beam function $A(\alpha)$, gives a visibility proportional to

$$\iiint A(\alpha') I(\theta', \alpha') e^{ik'b\alpha' \cos\theta'} d\theta' d\alpha' dk'. \quad (2.6)$$

2.1.1 AMI interferometry

The compensation for the path-length differences between each antenna and the cluster having been done in cables, the analogue correlator multiples the signal from each antenna at time t by the signal at times $t - 7\Delta t, t - 6\Delta t, \dots, t, \dots, t + 7\Delta t, t + 8\Delta t$; Fourier transforming these lag products gives the amplitude and phase values of each of eight frequency channels. A problem with the analogue correlator is that each timelag $\Delta t = \Delta l/v_{\text{group}}$ is not the same because each Δl , nominally 25 mm, varies by some 5–10% because the circuit boards providing the Δl s have non-uniform relative permittivities and thicknesses.

Thompson, Moran, & Swenson (2011) discuss cross-correlator performance in terms of cross-correlation correlation coefficient ρ ,

$$\frac{\int (a_i - \langle a_i \rangle)(a_j - \langle a_j \rangle)dt}{\sqrt{\int (a_i - \langle a_i \rangle)^2 dt \int (a_j - \langle a_j \rangle)^2 dt}}, \quad (2.7)$$

where a_i is the instantaneous voltage from antenna i , and $\langle \rangle$ denotes average over the few-second integration time τ , and the integrals are over τ . However, radio astronomy cross-correlators do not measure the denominator of equation 2.7; what they do measure, for each lag, is effectively the numerator. The signal power is described as $A \exp(i\phi)$ where A is "amplitude" and ϕ is "phase". The noise power is that from the front-end amplifiers, the atmosphere, and the CMB. The signal in the cross-correlation increases coherently over time, so the signal energy increases as time, while the noise increases incoherently so the noise energy increases as time $^{1/2}$. τ is chosen such that over it, signal energy \ll noise energy. For the measurements over τ to be meaningful, you want each receiver chain, from each front-end amplifier to correlator input, to produce a power that is *stable* over the whole (typically 6–8 hour) observation run. This is achieved with automatic gain controls designed to keep the power going into each correlator input constant. (Note that neither the gains of the receiver chains nor the output powers have to be the same – astronomical calibration deals with this).

However, ensuring the power at a correlator input is maintained at a constant level will bias measurements if, for example, the weather changes: cloud, rain, and raindrops on the receiver cover all emit at GHz-frequencies, thus raising (compared with fine weather) the noise power and so lowering the signal. This effect is (ideally) removed by the noise injection system (at Cambridge misleadingly called the ‘rain gauge’) which works as follows. Low-level noise (of power $\approx 1\%$ of the power due to front-end amplifier, CMB and atmosphere), of constant mean power and known signature $P(t)$ is injected into the waveguide that feeds the astronomical signal into the front-end amplifier. At each correlator input, the noise power due to $P(t)$ is extracted by synchronous detection and compared with the total noise power so that the noise power due to front-end amplifier, CMB and atmosphere, which determines the system temperature, is measured.

2.2 Measuring the SZ effect with an interferometer

For a small field size, an interferometer samples from the two-dimensional complex visibility plane \mathbf{u} , also known as the u - v plane, where u and v are orthogonal projected baselines in units of

observing wavelength. For a given frequency ν the quantity measured by an interferometer corresponds (see equation 2.6) to the Fourier components of the sky brightness distribution $\tilde{I}_\nu(\mathbf{u})$. $\tilde{I}_\nu(\mathbf{u})$ is given by the weighted Fourier transform of the surface brightness $I_\nu(\mathbf{x})$,

$$\tilde{I}_\nu(\mathbf{u}) = \int_{-\infty}^{\infty} A_\nu(\mathbf{x}) I_\nu(\mathbf{x}) e^{2\pi i \mathbf{u} \cdot \mathbf{x}} d^2 \mathbf{x}, \quad (2.8)$$

where \mathbf{x} is the position in the sky relative to the phase centre and $A_\nu(\mathbf{x})$ is the primary beam of the (identical) antennas for a given frequency; note that I and A here are parameterised in terms of spatial coordinates rather than angular. The positions at which $\tilde{I}_\nu(\mathbf{u})$ are sampled from is therefore determined by the physical orientation of the antennas.

The change in CMB surface brightness due to the thermal SZ effect in a galaxy cluster is given by (see e.g. [Birkinshaw 1999](#))

$$\delta I_{\text{cl},\nu} = T_{\text{CMB}} y f_\nu \frac{\partial B_\nu}{\partial T} \Big|_{T=T_{\text{CMB}}} \quad (2.9)$$

where the last factor is the derivative of the blackbody spectrum with respect to temperature evaluated at the temperature of the CMB, which at present is $T_{\text{CMB}} = 2.728$ K ([Fixsen et al. 1996](#)). The surface brightness per unit frequency of blackbody radiation is given (see e.g. [Kogure & Leung 2007](#)) by

$$B_\nu(T) = \frac{2h_p \nu^3}{c^2} \frac{1}{e^{h_p \nu / k_B T} - 1}, \quad (2.10)$$

where h_p is the Planck constant and k_B is the Boltzmann constant. Hence the derivative is given by

$$\frac{\partial B_\nu}{\partial T} \Big|_{T=T_{\text{CMB}}} = \frac{2h_p^2 \nu^4}{c^2 k_B T_{\text{CMB}}^2} \frac{e^{h_p \nu / k_B T}}{(e^{h_p \nu / k_B T} - 1)^2}. \quad (2.11)$$

The function f_ν expresses the spectral dependence of the SZ signal and is derived from the Kompaneets equation ([Kompaneets 1957](#)). Relativistic treatments of f_ν have been considered in e.g. [Rephaeli \(1995\)](#), [Itoh, Kohyama, & Nozawa \(1998\)](#), [Challinor & Lasenby \(1998\)](#), [Nozawa, Itoh, & Kohyama \(1998\)](#), and [Pointecouteau, Giard, & Barret \(1998\)](#), by incorporating relativistic terms into the Kompaneets equation. Relativistic effects may be important in clusters where the ICM temperatures are high. Indeed [Arnaud et al. \(1994\)](#) and [Markevitch et al. \(1996\)](#) have shown that electrons in the ICM can reach energies above 10 keV. Challinor & Lasenby show that these effects lead to a small decrease in the SZ effect. However, Rephaeli argues that the non-relativistic treatment of Compton scattering adopted in [Zeldovich & Sunyaev \(1969\)](#) remains valid at frequencies well below the CMB peak value. For the observing frequencies of AMI (≈ 15 GHz), it can be assumed that this condition holds. Furthermore Rephaeli claims that

for the unmodified Kompaneets equation to be valid, the optical depth of the cluster τ , must be sufficiently large to justify using a diffusion approximation for the scattering process. It is clear that at AMI observing frequencies $h_p \nu \ll m_e c^2$ where m_e is the mass of an electron; and so the photons can be assumed to scatter in the Thomson limit. In this limit the scattering rate is $\propto \sigma_T n_e$ where σ_T is the Thomson scattering cross-section and n_e is the electron number density in the ICM. Thus the optical depth is given by

$$\tau = \int n_e(r) \sigma_T dl, \quad (2.12)$$

where r is the radius from the galaxy cluster centre and the integral is along the line of sight. The non-relativistic form for f_ν is given by

$$f_\nu = X \coth(X/2) - 4, \quad (2.13)$$

where

$$X = \frac{h_p \nu}{k_B T_{\text{CMB}}}. \quad (2.14)$$

Referring back to equation 2.9, y is the Comptonisation parameter which is the number of collisions multiplied by the mean fractional change in energy of the photons per collision, integrated along the line of sight. On average the electrons in the ICM transfer an energy $k_B T_e(r)/m_e c^2$ to the scattered CMB photons, where $T_e(r)$ is the temperature of an electron in the ICM. In the Thomson scattering regime described above this leads to

$$y = \frac{\sigma_T k_B}{m_e c^2} \int T_e(r) n_e(r) dl. \quad (2.15)$$

If the electron gas is assumed to be ideal, then in terms of the gas pressure $P_e(r)$, the Comptonisation parameter is given by

$$y = \frac{\sigma_T}{m_e c^2} \int P_e(r) dl. \quad (2.16)$$

Combining equations 2.11, 2.13, & 2.16 one obtains the following expression for $\delta I_{\nu, \text{cl}}$ in the non-relativistic limit

$$\delta I_{\text{cl}, \nu} = \frac{2\sigma_T (k_B T_{\text{CMB}})^3 X^4 e^X}{h_p^2 c^4 (e^X - 1)^2} [X \coth(X/2) - 4] \int P_e(r) dl. \quad (2.17)$$

Thus for a given cluster $\delta I_{\text{cl}, \nu}$ is independent of z . Since the Fourier transform is a linear operator $\delta I_{\nu, \text{cl}}$ can be substituted directly into equation 2.8 to calculate $\tilde{\delta I}_{\nu, \text{cl}}$.

[Bartlett & Silk \(1994\)](#) noted that the total Comptonisation parameter Y , which is the integral of y over the solid angle $d\Omega$ subtended by the galaxy cluster is proportional to the volume

integral of the gas pressure. Y can be written in terms of spherical coordinates as

$$Y_{\text{sph,phys}}(r) = \frac{\sigma_T}{m_e c^2} \int_0^r P_e(r') 4\pi r'^2 dr'. \quad (2.18)$$

Note that $Y_{\text{sph,phys}}(r)$ has dimensions [length²]. $Y_{\text{sph}}(r) \equiv Y_{\text{sph,phys}}(r)/D_A^2$ (where D_A is the angular diameter distance to the cluster), which has dimensions [angle²] and is the quantity referred to in this thesis unless stated otherwise. Thus $Y_{\text{sph,phys}}$ measured out to large r is, with caveats, the total thermal energy of the cluster.

2.3 Cluster model selection

To determine $\delta I_{\text{cl},\nu}$, one must select a model which calculates the electron temperature (equation 2.15) or pressure (equation 2.16) profile of a cluster. The AMI consortium has implemented a number of cluster models over the years. [Marshall, Hobson, & Slosar \(2003\)](#) considered the Navarro-Frenk-White (NFW) profile ([Navarro, Frenk, & White 1995](#)) as a cluster mass model; their model assumes spherical symmetry and hydrostatic equilibrium, and is used in a joint analysis between SZ and gravitational lensing data (see e.g. [Schramm & Kayser 1994](#) for how lensing can be used to investigate cluster properties). Marshall also used the Beta model ([Cavaliere & Fusco-Femiano 1976, 1978](#)) to model the cluster gas profile; the Beta model is another spherically symmetric model, but is purely empirical. [Feroz et al. \(2009\)](#) (from here on FF09) built on this work, but concentrated on modelling multi-frequency SZ data with the Beta model, but using the hydrostatic equilibrium assumption to derive an expression for the cluster mass. Most recently [Olamaie, Hobson, & Grainge \(2012\)](#) (MO12) presented a new, physical model to describe the baryonic matter as well the dark matter component in order to give a more thorough treatment of the make-up of galaxy clusters; I refer to this as a physical model.

2.4 A physical model for AMI data

2.4.1 Model assumptions

The model presented here is largely based on the one introduced in MO12 but includes the adaptions mentioned in Sections 2.4.4 and 3.2. For any model it is important to know the underlying assumptions which allow it to be valid. The four main assumptions in the physical model are as follows.

- The cluster is spherically symmetric. This means that the cluster can be parameterised in terms of the scalar radius r (rather than its vector equivalent \mathbf{r}) from the centre of the cluster.
- The cluster is in hydrostatic equilibrium up to radius r_{200} (defined below). This means at any radius up to r_{200} the outward pushing pressure force created by the pressure differential at that point must be equal to the gravitational binding force due to the mass enclosed within that radius (see e.g. Bahcall & Sarazin 1977, and equation 2.30 below).
- The gas mass fraction $f_{\text{gas}}(r)$ is much less than unity up to radius r_{200} , so that the total mass is $M(r_{200}) \approx M_{\text{dm}}(r_{200})$. Consequently the total mass out to r_{200} is given by the integral of the dark matter density along the radius of the cluster (see equation 2.23 below).
- The cluster gas is assumed to be an ideal gas, so that the electron temperature can be trivially represented in terms of its pressure.

2.4.2 Dark matter profile

The model uses an NFW profile (Navarro, Frenk, & White 1995) the dark matter density as a function of cluster radius r ,

$$\rho_{\text{dm}}(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1 + \frac{r}{r_s}\right)^2}, \quad (2.19)$$

where ρ_s is an overall density normalisation coefficient and r_s is a characteristic radius defined by $r_s \equiv r_{200}/c_{200}$ and is the radius at which the logarithmic slope of the profile $d \ln \rho(r)/d \ln r$ is -2 . r_{200} is the radius at which the average cluster density is $200 \times \rho_{\text{crit}}(z)$. $\rho_{\text{crit}}(z)$ is the *critical* density of the Universe at the cluster z which is given by $\rho_{\text{crit}}(z) = 3H(z)^2/8\pi G$ where $H(z)$ is the Hubble parameter (at the cluster redshift) and G is Newton's constant. c_{200} is the concentration parameter at this radius. Following Olamaie, Hobson, & Grainge (2013), we can calculate c_{200} for an NFW dark matter density profile taken from the expression in Corless, King, & Clowe (2009)

$$c_{200} = \frac{5.26}{1+z} \left(\frac{M(r_{200})}{10^{14} h^{-1} M_{\text{Sun}}} \right)^{-0.1}, \quad (2.20)$$

here, M_{Sun} denotes units of solar mass. The $1/(1+z)$ factor comes from Wechsler et al. (2001) and is obtained from N-body simulated dark matter halos between $z = 0$ and $z = 7$. The remainder of the relation was derived in Neto et al. (2007) by fitting a power-law for c_{200} to N-body simulated cluster data. Note that the sample used in Neto et al. (2007) was assumed to contain clusters that are relaxed. In equation 2.20 $M(r_{200})$ is the mass enclosed at radius r_{200} . Thus for given values of z and $M(r_{200})$, c_{200} can be calculated.

2.4.3 Electron pressure profile

Following [Nagai, Kravtsov, & Vikhlinin \(2007\)](#), the generalised-NFW (GNFW) model is used to parameterise the electron pressure as a function of radius from the cluster centre

$$P_e(r) = \frac{P_{ei}}{\left(\frac{r}{r_p}\right)^c \left(1 + \left(\frac{r}{r_p}\right)^a\right)^{(b-c)/a}}, \quad (2.21)$$

where P_{ei} is an overall pressure normalisation factor and r_p is another characteristic radius, defined by $r_p \equiv r_{500}/c_{500}$. The parameters a , b and c describe the slope of the pressure profile at $r/r_p \approx 1$, $r/r_p \gg 1$ and $r/r_p \ll 1$ respectively. For values $r/r_p \ll 1$ the logarithmic slope ($d \ln P_e(r)/d \ln r$) converges to $-c$. For values $r/r_p \gg 1$ the logarithmic slope converges to $-b$. The value of a dictates how quickly (in terms of r) the slope switches between these two values, and in the limit that a tends to zero, the logarithmic slope is $-(b+c)/2$ for all r . Note that [Nagai, Kravtsov, & Vikhlinin \(2007\)](#) choose to parameterise the pressure profile with the GNFW model because it closely matches the observed profiles of the Chandra X-ray clusters and results of numerical simulations in their outskirts. In addition to this, the gas pressure distribution is primarily determined by the gravitationally dominant dark matter component (which is fitted with the NFW profile), they argue that it makes sense to parameterise the pressure profile using the generalised NFW model.

Consistent with many of the Planck follow-up papers (see e.g. [Planck Collaboration XI 2011](#)) and with MO12 the slope parameters are taken to be $a = 1.0620$, $b = 5.4807$ and $c = 0.3292$. These ‘universal’ values are from [Arnaud et al. \(2010\)](#) and are the GNFW slope parameters derived for the standard self-similar case using scaling relations from a REXCESS sub-sample (of 20 well-studied low- z clusters observed with XMM-Newton), as described in appendix B of the paper ([Böhringer et al. 2007](#)). I also use the Arnaud et al. value for the concentration parameter $c_{500} \equiv r_{500}/r_p$ of 1.156. I note however that in [Perrott et al. \(2015\)](#) (from here on YP15) using simulations it was shown that the disagreement between Planck and AMI parameter estimates may indicate pressure profiles deviating from the ‘universal’ profile.

2.4.4 Model calculations

The three cluster model input parameters required to calculate the electron pressure given by equation 2.21 in the physical model are $M(r_{200})$, z , and $f_{\text{gas}}(r_{200})$. $f_{\text{gas}}(r_{200})$ is the fraction of the total mass attributed to the gas mass up to radius r_{200} . Note that in general the total mass out

to r_A is given by

$$M(r_A) = \frac{4\pi}{3} A \rho_{\text{crit}}(z) r_A^3. \quad (2.22)$$

Hence r_{200} can be calculated from $M(r_{200})$, and the mass can be determined at other (known) radii (e.g. r_{500}).

2.4.4.1 Total enclosed mass

Another analytical solution for $M(r)$ can be derived using the third assumption stated above. Using equation 2.19, $M(r)$ is given by

$$\begin{aligned} M(r) &= \int_0^r 4\pi \rho_{\text{dm}}(r') r'^2 dr' \\ &= \int_0^r 4\pi \frac{\rho_s r'^2}{\left(\frac{r'}{r_s}\right) \left(1 + \frac{r'}{r_s}\right)^2} dr' \\ &= 4\pi \rho_s r_s^3 \left[\ln\left(1 + \frac{r}{r_s}\right) - \left(1 + \frac{r_s}{r}\right)^{-1} \right]. \end{aligned} \quad (2.23)$$

Hence an expression for ρ_s can be obtained by equating 2.22 and 2.23, setting $r = r_{200}$ and solving for ρ_s

$$\rho_s = \frac{200}{3} \left(\frac{r_{200}}{r_s}\right)^3 \frac{\rho_{\text{crit}}(z)}{\left[\ln\left(1 + \frac{r_{200}}{r_s}\right) - \left(1 + \frac{r_s}{r_{200}}\right)^{-1}\right]}. \quad (2.24)$$

One can then obtain an expression r_{500} as follows. Equating 2.22 and 2.23 at r_{500} and substituting in the expression for ρ_s gives

$$\left(\frac{r_s}{r_{500}}\right)^3 \left[\ln\left(1 + \frac{r_{500}}{r_s}\right) - \left(1 + \frac{r_s}{r_{500}}\right)^{-1} \right] = \frac{5}{2} \left(\frac{r_s}{r_{200}}\right)^3 \left[\ln\left(1 + \frac{r_{200}}{r_s}\right) - \left(1 + \frac{r_s}{r_{200}}\right)^{-1} \right]. \quad (2.25)$$

Following Hu & Kravtsov (2003), there is an analytic mapping from r_{200} to r_{500} . Consider the equation

$$g(r_s/r_{500}) = \frac{5}{2} g(r_s/r_{200}), \quad (2.26)$$

where

$$g(x) = x^3 [\ln(1 + x^{-1}) - (1 + x)^{-1}]. \quad (2.27)$$

Equation 2.26 requires that $g(r_s/r_{500})$ be inverted so that

$$\frac{r_s}{r_{500}} = x \left(g_{500} = \frac{5}{2} f(r_s/r_{200}) \right), \quad (2.28)$$

where

$$x(g_{500}) = \left[a_1 g_{500}^{2p} + \frac{9}{16} \right]^{-1/2} + 2g_{500}. \quad (2.29)$$

Here $p = a_2 + a_3 \ln g_{500} + a_4 (\ln g_{500})^2$, and the four fitting parameters correspond to $a_1 = 0.5116$, $a_2 = -0.4283$, $a_3 = -3.13 \times 10^{-3}$ and $a_4 = -3.52 \times 10^{-5}$. This gives a fit to better than 0.3% accuracy for $0 < c_{200} < 20$ and is exact in the limit that $c_{200} \rightarrow 0$. Once r_{500} has been calculated r_p can be calculated from $r_p = r_{500}/c_{500}$.

2.4.4.2 Hydrostatic equilibrium

This requires

$$\frac{dP_g(r)}{dr} = -\rho_g(r) \frac{GM_{\text{tot}}(r)}{r^2}, \quad (2.30)$$

where $\rho_g(r)$ is the gas density and $M(r)$ is the total mass within radius r of the cluster. The gas pressure $P_g(r)$ can be related to the electron pressure as

$$\mu_g P_g(r) = \mu_e P_e(r), \quad (2.31)$$

where μ_e is the mean gas mass per electron and μ_g is the mean mass per gas particle. [Mason & Myers \(2000\)](#) state that for a plasma with the cosmic helium mass fraction $C_{\text{He}} = 0.24$ and the solar abundance values in [Anders & Grevesse \(1989\)](#), then $\mu_e = 1.146$ and $\mu_g = 0.592$ in units of proton mass.

2.4.4.3 Gas density, mass, and temperature

Substituting equations 2.23 and 2.31 into 2.30 and solving for $\rho_g(r)$ gives

$$\begin{aligned} \rho_g(r) = & \frac{\mu_e P_{\text{ei}}}{\mu_g} \frac{1}{4\pi G \rho_s r_s^3} \\ & \times \frac{r}{\ln \left(1 + \frac{r}{r_s} \right) - \left(1 + \frac{r_s}{r} \right)^{-1}} \\ & \times \left(\frac{r}{r_p} \right)^{-c} \left[1 + \left(\frac{r}{r_p} \right)^a \right]^{-(1+(b-c)/a)} \left[b \left(\frac{r}{r_p} \right)^a + c \right]. \end{aligned} \quad (2.32)$$

From this the gas mass $M_g(r)$ can be calculated

$$M_g(r) = \int_0^r 4\pi \rho_g(r') r'^2 dr'. \quad (2.33)$$

Note however that this integral must be solved numerically. Nevertheless, we can determine P_{ei} since we know $M(r_{200})$, $f_{\text{gas}}(r_{200})$ and r_{200} ($M_{\text{g}}(r_{200}) = f_{\text{gas}}(r_{200})M(r_{200})$). Evaluating equations 2.32 and 2.33 at r_{200} and solving for P_{ei} gives the following expression

$$P_{\text{ei}} = \frac{\mu_{\text{g}}}{\mu_{\text{e}}} G \rho_{\text{s}} r_{\text{s}}^3 M_{\text{g}}(r_{200}) \times \frac{1}{\int_0^{r_{200}} \frac{r'^3}{\ln\left(1 + \frac{r'}{r_{\text{s}}}\right) - \left(1 + \frac{r_{\text{s}}}{r'}\right)^{-1}} \left(\frac{r'}{r_{\text{p}}}\right)^{-c} \left[1 + \left(\frac{r'}{r_{\text{p}}}\right)^a\right]^{-(1+(b-c)/a)} \left[b\left(\frac{r'}{r_{\text{p}}}\right)^a + c\right] \text{d}r'}. \quad (2.34)$$

The radial profile of the electron number density is given by $n_{\text{e}}(r) = \rho_{\text{g}}(r)/\mu_{\text{e}}$. Assuming an ideal gas equation of state, the electron temperature $T_{\text{e}}(r)$ is therefore given by

$$T_{\text{e}}(r) = \left(\frac{4\pi\mu_{\text{g}} G \rho_{\text{s}} r_{\text{s}}^3}{k_{\text{B}}} \right) \times \frac{\ln\left(1 + \frac{r}{r_{\text{s}}}\right) - \left(1 + \frac{r_{\text{s}}}{r}\right)^{-1}}{r} \times \left[1 + \left(\frac{r}{r_{\text{p}}}\right)^a\right] \left[b\left(\frac{r}{r_{\text{p}}}\right)^a + c\right]^{-1}, \quad (2.35)$$

which is also equal to the gas temperature $T_{\text{g}}(r)$.

The gas mass can be determined numerically from equation 2.33 as

$$M_{\text{g}}(r) = \frac{\mu_{\text{e}} P_{\text{ei}}}{\mu_{\text{g}}} \frac{1}{G \rho_{\text{s}} r_{\text{s}}^3} \times \int_0^r \frac{r'^3}{\ln\left(1 + \frac{r'}{r_{\text{s}}}\right) - \left(1 + \frac{r_{\text{s}}}{r'}\right)^{-1}} \times \left(\frac{r'}{r_{\text{p}}}\right)^{-c} \left[1 + \left(\frac{r'}{r_{\text{p}}}\right)^a\right]^{-(1+(b-c)/a)} \left[b\left(\frac{r'}{r_{\text{p}}}\right)^a + c\right] \text{d}r'. \quad (2.36)$$

2.4.4.4 Determining $\tilde{\delta}I_{\text{cl},\nu}$

Once r_{p} and P_{ei} have been calculated, the pressure profile can be used in equation 2.16 to calculate the Comptonisation parameter which in turn can be used to calculate $\delta I_{\text{cl},\nu}$ using equation 2.9. $\delta I_{\text{cl},\nu}$ can be Fourier transformed to get the quantity comparable to what an interferometer measures, so that the physical model can be used to analyse data obtained with AMI.

2.5 Recognised radio-sources and general noise contributions

In addition to the SZ decrement, visibilities measured by AMI also contain contributions from radio-sources, primordial CMB anisotropies, and instrumental noise. As defined in [Hobson & Maisinger \(2002\)](#), each visibility measured by an interferometer consists of two components

$$V_\nu(\mathbf{u}_i) = \tilde{I}_\nu(\mathbf{u}_i) + N_\nu(\mathbf{u}_i), \quad (2.37)$$

where $\tilde{I}_\nu(\mathbf{u}_i)$ contains both the contribution from the cluster SZ effect and from the *identified* radio-sources, and $N_\nu(\mathbf{u}_i)$ contains the contributions from *unidentified* radio-sources, primordial CMB and instrumental noise.

2.5.1 Recognised radio-sources

The LA has been (see e.g. [Franzen et al. 2011](#)) and is being used to measure the 15.7-GHz source count. The LA is used to measure radio-sources (without contamination from the SZ effect since the cluster is resolved out), whilst the SA simultaneously measures the combined SZ and source signals.

The visibility of each recognised radio-source, assuming for illustration that it is unresolved by the LA, is

$$\tilde{I}_{\text{rs},\nu}(\mathbf{u}) = \int A_\nu(\mathbf{x}) S_\nu(\mathbf{x}) \delta(\mathbf{x}_{\text{rs}}) e^{(2\pi i \mathbf{u} \cdot \mathbf{x})} d^2\mathbf{x} = S_\nu(\mathbf{x}_{\text{rs}}) A_\nu(\mathbf{x}_{\text{rs}}) e^{i\phi}, \quad (2.38)$$

where $S_\nu(\mathbf{x})$ is the source flux density at point \mathbf{x} relative to the phase centre, $\phi = 2\pi \mathbf{u} \cdot \mathbf{x}_{\text{rs}}$. The variation in source flux density across the AMI observing band is taken account of via the spectral index α , where

$$S_\nu = S_0 \left(\frac{\nu}{\nu_0} \right)^{-\alpha}, \quad (2.39)$$

where ν_0 is some reference frequency and S_0 is the corresponding source flux density.

2.5.2 General Noise Contributions

2.5.2.1 Instrumental noise

The main source of instrumental noise is Johnson noise. This refers to the thermal agitation of the charge carriers in any circuit ([Nyquist 1928](#)), and in the context of interferometry, the front-end receivers of the antennas. The antennas are cooled to mitigate this effect, but the remaining

contribution is non-negligible. For a given bandwidth $\Delta\nu$, the root mean square of the Johnson noise voltage from a single antenna is given by (see e.g. [Thompson, Moran, & Swenson 2011](#))

$$\sigma_{\text{Johnson}} = \sqrt{4k_B T_{\text{sys}} R \Delta\nu}, \quad (2.40)$$

where T_{sys} is the system temperature and R is the antenna impedance. Note that when limited to a finite bandwidth, Johnson noise is approximately Gaussian (see e.g. [Barry et al. 2004](#)).

2.5.2.2 Primordial CMB

Anisotropies in the temperature of the CMB were predicted as early as [Silk \(1967\)](#) among others, and [Smoot et al. \(1992\)](#) provided the first clear statistical evidence of their existence and [Hancock et al. \(1994\)](#) provided the first direct evidence of individual spatial structures in the CMB. These anisotropies can be separated into two categories: primordial and late time anisotropies. An example of the latter type is the SZ effect. Primordial anisotropies refer to fluctuations in the CMB that have been present since the surface of last scattering (which occurred at $z \approx 1100$ or $t \approx 4 \times 10^5$ years over a period of $\Delta z \approx 60$). On angular scales visible from the ground the acoustic peaks and troughs are the most significant features in the CMB power spectrum. When the Universe was radiation dominated, non-baryonic dark matter began to collapse under gravity to form potential wells^a, but baryonic matter could not clump due to pressure opposition from Thompson scattering of photons by electrons given that there were 10^9 photons per baryon. During recombination the acoustic oscillations imprint the CMB, after recombination the atoms fall into the non-baryonic dark matter potential wells. Acoustic peaks and troughs relate to the waves oscillating in the baryon-photon plasma before recombination occurred. Each successive peak refers to the number of times the wave compressed before the radiation-matter decoupling, and is visible at decreasing angular scale. In this work, the power spectrum for CMB primordial anisotropies is determined via maximum-likelihood methods as written in [Hobson & Maisinger \(2002\)](#) using the results from [Hinshaw et al. \(2013\)](#).

2.5.2.3 Background unrecognised radio-sources

Although the LA is used to identify radio point sources with flux densities $\geq S_{\text{lim}}$ (where S_{lim} is a limiting flux density that is usually taken as $4 \times \sigma$ and σ is the resultant RMS noise in the summed LA data on the particular sky patch), a large enough number of sources with flux densities $< S_{\text{lim}}$ can be a significant contaminant. This type of noise is often referred to as source

^aThis only applies to matter that was in causal contact.

confusion. [Scheuer \(1957\)](#) showed that if such sources obey a power-law number-flux density relation ($n_\nu(S) = dN_\nu(> S)/dS \propto kS^\gamma$ where k & γ are dimensionless constants), then for a random distribution of unresolved radio-sources in the sky, the source confusion noise is given by

$$\sigma_{\text{conf}}^2 = \int_0^{S_{\text{lim}}} S^2 n_\nu(S) dS. \quad (2.41)$$

γ and k were determined empirically in [Davies et al. \(2011\)](#) from the 10C survey to be $\gamma = -1.80$ and $k = 376$ when $n_\nu(S)$ is quoted in units of $\text{Jy}^{-1} \text{sr}^{-1}$, so that when S_{lim} is taken to be $300 \mu\text{Jy}$ (for a standard length AMI cluster observation) $\sigma_{\text{conf}}^2 = 0.185 \text{ Jy}^2 \text{sr}^{-1}$.

2.6 Bayesian inference

2.6.1 Parameter estimation

Given a model \mathcal{M} and a data vector \mathcal{D} , one can obtain model parameters (also known as input parameters or sampling parameters) $\boldsymbol{\Theta}$ conditioned on \mathcal{M} and \mathcal{D} using Bayes' theorem:

$$P(\boldsymbol{\Theta}|\mathcal{D}, \mathcal{M}) = \frac{P(\mathcal{D}|\boldsymbol{\Theta}, \mathcal{M}) P(\boldsymbol{\Theta}|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})}, \quad (2.42)$$

where $P(\boldsymbol{\Theta}|\mathcal{D}, \mathcal{M}) \equiv \mathcal{P}(\boldsymbol{\Theta})$ is the posterior distribution of the input parameter set, $P(\mathcal{D}|\boldsymbol{\Theta}, \mathcal{M}) \equiv \mathcal{L}(\boldsymbol{\Theta})$ is the likelihood function for the data, $P(\boldsymbol{\Theta}|\mathcal{M}) \equiv \pi(\boldsymbol{\Theta})$ is the prior probability distribution for the model parameter set, and $P(\mathcal{D}|\mathcal{M}) \equiv \mathcal{Z}(\mathcal{D})$ is the Bayesian evidence of the data. The evidence can be defined as the factor required to normalise the posterior over the sampling parameter space:

$$\mathcal{Z}(\mathcal{D}) = \int \mathcal{L}(\boldsymbol{\Theta}) \pi(\boldsymbol{\Theta}) d\boldsymbol{\Theta}, \quad (2.43)$$

where the integral is carried out over the N -dimensional parameter space. For the models using AMI data considered here, the input parameters can be split into two subsets, (which are assumed to be independent of one another): cluster parameters $\boldsymbol{\Theta}_{\text{cl}}$ and radio-source or ‘nuisance’ parameters $\boldsymbol{\Theta}_{\text{rs}}$.

2.6.2 Model comparison

While it is the posterior distribution which gives the model parameter estimates from the prior information and data, it is $\mathcal{Z}(\mathcal{D})$ which is crucial to performing model selection. The nested sampling algorithm, [MULTINEST \(Feroz, Hobson, & Bridges 2009\)](#) is a Monte Carlo algorithm

which calculates $\mathcal{Z}(\mathcal{D})$ by making use of a transformation of the N -dimensional evidence integral into a one-dimensional integral that is much easier to evaluate. The algorithm also produces samples from $\mathcal{P}(\boldsymbol{\Theta})$ as a by-product, meaning that it is suitable for both the parameter estimation and model comparison aspects of this work. Nested sampling will be discussed in more detail in Section 9.4. Comparing models in a Bayesian way can be done by considering the following. The probability of a model \mathcal{M} conditioned on \mathcal{D} can be calculated using Bayes' theorem

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}) P(\mathcal{M})}{P(\mathcal{D})}. \quad (2.44)$$

Hence for two models, \mathcal{M}_1 and \mathcal{M}_2 , the ratio of the models conditioned on the same dataset is given by

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M}_1) P(\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2) P(\mathcal{M}_2)}, \quad (2.45)$$

where $P(\mathcal{M}_2)/P(\mathcal{M}_1)$ is the a-priori probability ratio of the models. We set this to one, i.e. we place no bias towards a particular model before performing the analysis. Hence the ratio of the probabilities of the models given the data is equal to the ratio of the evidence values obtained from the respective models (we have defined $\mathcal{Z}_i(\mathcal{D}) \equiv P(\mathcal{D}|\mathcal{M}_i)$). The evidence is simply the average of the likelihood function over the sampling parameter space, weighted by the prior distribution. This means that the evidence is larger for a model with larger areas in its parameter space having higher likelihood values. Moreover, a larger parameter space, either in the form of higher dimensionality or a larger domain, results in a lower evidence value, all other things being equal. Hence the evidence penalises more complex models over basic (lower dimensionality / smaller input parameter space domains) ones which give an equally good fit to the data. Thus the evidence automatically implements Occam's razor: when you have two competing theories that make exactly the same predictions, the simpler one is the better. [Jeffreys \(1961\)](#) provides a scale for interpreting the ratio of evidences as a means of performing model comparison (Table 2.1). A value of $\ln(\mathcal{Z}_1/\mathcal{Z}_2)$ above 5.0 (less than -5.0) presents "strong evidence" in favour of model 1 (model 2). Values $2.5 \leq \ln(\mathcal{Z}_1/\mathcal{Z}_2) < 5.0$ ($-5.0 < \ln(\mathcal{Z}_1/\mathcal{Z}_2) \leq -2.5$) present "moderate evidence" in favour of model 1 (model 2). Values $1 \leq \ln(\mathcal{Z}_1/\mathcal{Z}_2) < 2.5$ ($-2.5 < \ln(\mathcal{Z}_1/\mathcal{Z}_2) \leq -1$) present "weak evidence" in favour of model 1 (model 2). Finally, values $-1.0 < \ln(\mathcal{Z}_1/\mathcal{Z}_2) < 1.0$ require "more information to come to a conclusion" over model preference.

$\ln(\mathcal{Z}_1/\mathcal{Z}_2)$	Interpretation	Probability of favoured model
≤ 1.0	better data are needed	≤ 0.75
≤ 2.5	weak evidence in favour of \mathcal{M}_1	0.923
≤ 5.0	moderate evidence in favour of \mathcal{M}_1	0.993
> 5.0	strong evidence in favour of \mathcal{M}_1	> 0.993

Table 2.1: Jeffreys scale for assessing model preferability based on the $\ln \equiv \log_e$ of the evidence ratio of two models.

2.7 Parameter prior distributions

Prior distributions incorporate the prior knowledge we have on the sampling parameters used in Bayesian inference. The prior parameter space for AMI cluster analysis consists both of parameters associated with the cluster $\pi(\boldsymbol{\Theta}_{cl})$ and those associated with each identified radio-source $\pi(\boldsymbol{\Theta}_{rs})$. If one assumes that the cluster parameters are separable from those associated with each recognised radio-source, then the total prior distribution is given by

$$\pi(\boldsymbol{\Theta}_t) = \pi(\boldsymbol{\Theta}_{cl}) \prod_i \pi(\boldsymbol{\Theta}_{rs,i}), \quad (2.46)$$

where i labels each recognised radio-source. The prior distributions assigned to the cluster parameters will be discussed in the Sections where the Bayesian analyses carried out are introduced (i.e. Sections 3.2, 4.2, 5.1.0.7, and 8.3.1).

2.7.1 Radio-source prior distributions

Following FF09, each source can be parameterised by four variables: its position on the sky (x_{rs} , y_{rs}), its measured flux density at some reference frequency ν_0 , $S_{rs,0}$, and its spectral index α_{rs} . Assuming these are independent, then for source i

$$\pi(\boldsymbol{\Theta}_{rs,i}) = \pi(x_{rs,i})\pi(y_{rs,i})\pi(S_{rs,0,i})\pi(\alpha_{rs,i}). \quad (2.47)$$

Delta functions are applied to the prior distributions on x_{rs} and y_{rs} , due to the LA's ability to measure spatial positions to high accuracy: $\pi(x_{rs}) = \delta(x_{rs,LA})$, $\pi(y_{rs}) = \delta(y_{rs,LA})$. Delta priors were also set on $S_{rs,0}$ & α_{rs} (centred on the values measured by the LA), if the measured $S_{rs,0}$ was less than four times the instrumental noise associated with the observation, and the source was more than 5 arcminutes away from the SA pointing centre: $\pi(S_{rs,0}) = \delta(S_{rs,0,LA})$, $\pi(\alpha_{rs}) = \delta(\alpha_{rs,LA})$. Otherwise, a Gaussian prior was set on $S_{rs,0}$ centred at the LA measured value with a standard deviation equal to 40% of the measured value ($\sigma_{rs,0} = 0.4 \times S_{rs,0,LA}$):

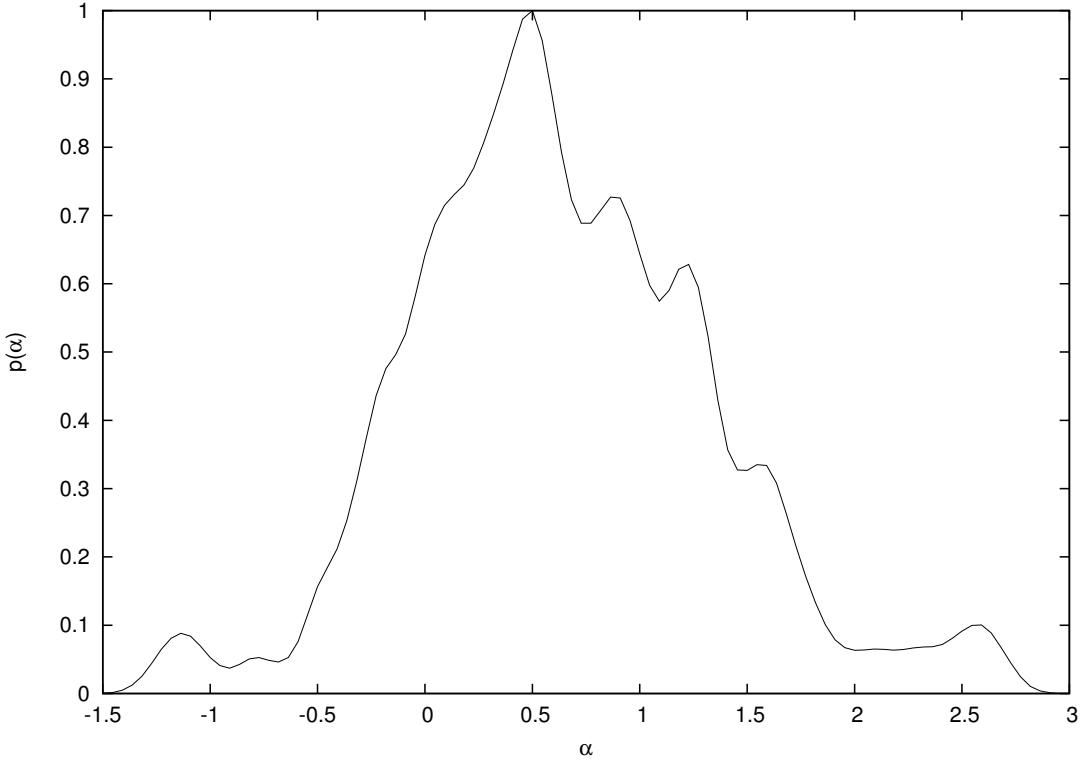


Figure 2.2: Spectral index distribution adapted from [Waldrum et al. \(2007\)](#) from the 9C survey of radio-sources.

$\pi(S_{\text{rs},0}) = \mathcal{N}(S_{\text{rs},0}, \text{LA}, \sigma_{\text{rs},0})$. The spectral index α_{rs} was modelled using the empirical distribution determined in [Waldrum et al. \(2007\)](#): $\pi(\alpha_{\text{rs}}) = \mathcal{W}(\alpha_{\text{rs}})$ and is shown in Figure 2.2.

2.8 The likelihood function

The likelihood function gives the probability of observing data given a set of parameter values. In the case of AMI observations, the data are visibilities observed by AMI and the parameters are those described in the previous Section. Following [Hobson & Maisinger \(2002\)](#) and FF09, it is convenient first to place the $N_{\text{vis},\nu}$ observed complex visibilities $V_\nu(\mathbf{u}_i)$ into a data vector d_ν for each frequency channel (six channels in the case of the analogue correlator AMI data),

ordered such that

$$\mathbf{d}_{\nu,i} = \begin{cases} \text{Re}[V_{\nu}(\mathbf{u}_i)] & (i \leq N_{\text{vis},\nu}) \\ \text{Im}[V_{\nu}(\mathbf{u}_{i-N_{\text{vis},\nu}})] & (N_{\text{vis},\nu} + 1 \leq i \leq 2N_{\text{vis},\nu}). \end{cases} \quad (2.48)$$

Similarly, one can define the noise vectors \mathbf{n}_{ν} containing only the contributions to the noise components $N_{\nu}(\mathbf{u}_i)$. Section 2.5.2 explains the three contributors to $N_{\nu}(\mathbf{u}_i)$. We take the likelihood to be Gaussian

$$\mathcal{L}(\boldsymbol{\Theta}) = \frac{1}{Z_N} e^{-\frac{1}{2}\chi^2}. \quad (2.49)$$

Here χ^2 is a measure of the goodness-of-fit of the model to the data (which is simply the concatenation of data vectors \mathbf{d}_{ν} for all ν) \mathbf{d} and the predicted data $\mathbf{d}^p(\boldsymbol{\Theta})$:

$$\chi^2 = \sum_{\nu,\nu'} (\mathbf{d}_{\nu} - \mathbf{d}_{\nu}^p(\boldsymbol{\Theta}))^T \mathbf{C}_{\nu,\nu'}^{-1} (\mathbf{d}_{\nu'} - \mathbf{d}_{\nu'}^p(\boldsymbol{\Theta})). \quad (2.50)$$

$\mathbf{d}_{\nu}^p(\boldsymbol{\Theta})$ is assumed to consist of the signal measured from the cluster and *recognised* radio-sources. $\mathbf{C}_{\nu,\nu'} \equiv \langle \mathbf{n}_{\nu} \mathbf{n}_{\nu'}^T \rangle$ is the covariance matrix of the visibilities. Assuming instrumental (Section 2.5.2.1), CMB (Section 2.5.2.2), and confusion (Section 2.5.2.3) noise are independent of each other, $\mathbf{C}_{\nu,\nu'}$ can be written as

$$\mathbf{C}_{\nu,\nu'} = \mathbf{C}_{\nu,\nu'}^{\text{ins}} + \mathbf{C}_{\nu,\nu'}^{\text{CMB}} + \mathbf{C}_{\nu,\nu'}^{\text{conf}}. \quad (2.51)$$

Note that the instrumental noise associated with AMI observations is measured, and so does not need to be predicted. For further information on all three sources of noise, see FF09 Section 5.3 and [Hobson & Maisinger \(2002\)](#). Z_N is a normalisation factor given by

$$Z_N = (2\pi)^{N_{\text{vis}}} |\mathbf{C}|^{\frac{1}{2}}, \quad (2.52)$$

where N_{vis} is the total number of visibilities observed over all six frequency channels.

PHYSICAL MODELLING OF CLUSTERS DETECTED BY PLANCK

YP15 present the results of the AMI follow-up of clusters detected by Planck – this follow-up is analysed using the ‘observational model’, which parameterises a cluster in terms of its integrated Comptonisation parameter Y and angular scale θ . YP15 find that these AMI estimates for Y are consistently lower than the values obtained from Planck data, and conclude that this may indicate that the cluster pressure profiles are deviating from the ‘universal’ one. I use the physical model described in Section 2.4 with data obtained from AMI of clusters detected by Planck (including ones which were detected after the analysis in YP15 was carried out). I also consider the cluster mass estimates given in the PSZ2 Planck cluster catalogue (Planck Collaboration [XXVII 2016](#)) and compare them with the values obtained using AMI data. Furthermore I use the PSZ2 mass estimates as inputs to simulations which are then analysed in the same way as real AMI observations. The work discussed in this Chapter has been published in MNRAS ([Javid et al. 2019](#)), and has been modified post-referee comments.

3.1 Selection and observation of the cluster sample

PSZ2 contains 1653 cluster candidates detected in the all-sky 29 month mission. The initial cluster selection criteria for AMI closely resembles that described in YP15, with a few modifi-

Parameter	Minimum value	Maximum value
Declination	20.31°	86.24°
z	0.045	0.83
S/N	4.50	28.40
M_{SZ} ($\times 10^{14} M_{\text{Sun}}$)	1.83	10.80

Table 3.1: Minimum and maximum values for a selection of parameters taken from PSZ2 for the AMI sample of 199 clusters.

ations as follows.

- The lower z limit $z \leq 0.100$ was relaxed here, to see how well AMI data can constrain physical model parameters at low redshift. However it is important to realise that the sample at $z \leq 0.100$ were not observed specifically for the purpose of this work, but were part of other observation projects.
- The Planck signal-to-noise ratio (S/N) lower bound was reduced to 4.5.
- The automatic radio-source environment rejection remained the same. However the manual rejection was done on a map-by-map basis— see Section 3.5.
- Note that the observation declination limits $20^\circ < \delta < 87^\circ$ were kept.

This led to an initial sample size of 199 clusters. The maximum and minimum values of some key parameters for this sample from the Planck catalogue are given in Table 3.1. Note that M_{SZ} is taken in PSZ2 as the hydrostatic equilibrium mass $M(r_{500})$, assuming the best-fit $Y - M$ relation.

The pointing strategy for each cluster was as follows. Clusters were observed using a single pointing centre on the SA, which has a primary beam of size ≈ 20 arcmin FWHM, to noise levels of $\lesssim 120 \mu\text{Jy beam}^{-1}$. To cover the same area with the LA, which has a primary beam of size ≈ 6 arcmin FWHM, the cluster field was observed as a 61-point hexagonal raster. The noise level of the raster was $\lesssim 100 \mu\text{Jy beam}^{-1}$ in the central 19 pointings, and slightly higher in the outer regions. The observations for a given cluster field were carried out simultaneously on both arrays, and the average observation time per cluster was ≈ 30 hours. The observations were carried out between 2013 and 2015, and so they began before the PSZ2 catalogue was published. This means that the AMI pointing centre coordinates in general were not the same as those published in the final Planck catalogue which was released in 2015. This is discussed in the context of the cluster centre offset parameters in Section 3.2. Data from both arrays were flagged for interference and calibrated using the AMI in-house software package REDUCE. Flux calibration was applied using contemporaneous observations of the primary calibration sources 3C 286, 3C 48, and 3C 147. The assumed flux densities for 3C 286 were converted from

Parameter	Prior distribution
x_c	$\mathcal{N}(0'', 60'')$
y_c	$\mathcal{N}(0'', 60'')$
z	$\delta(z_{\text{Planck}})$
$M(r_{200})$	$\mathcal{U}[\log(0.5 \times 10^{14} M_{\text{Sun}}), \log(50 \times 10^{14} M_{\text{Sun}})]$
$f_{\text{gas}}(r_{200})$	$\mathcal{N}(0.13, 0.02)$

Table 3.2: Cluster parameter prior distributions. δ denotes a Dirac delta function, \mathcal{U} is a uniform distribution and \mathcal{N} is a normal distribution (parameterised by its mean and standard deviation).

Very Large Array total-intensity measurements (Perley & Butler 2013) and are consistent with the Rudy et al. (1987) model of Mars transferred onto an absolute scale, using results from the Wilkinson Microwave Anisotropy Probe. The assumed flux densities for 3C 48 and 3C 147 were based on long-term monitoring with the SA using 3C 286 for flux calibration. Phase calibration was applied using interleaved observations of a nearby bright source selected from the VLBA Calibrator survey (Petrov et al. 2008); in the case of the LA, a secondary amplitude calibration was also applied using contemporaneous observations of the phase calibration source on the SA.

3.2 AMI data analysis

The likelihood function given by equation 2.49, along with all the preceding calculational steps covered in Chapter 2 are calculated using our AMI Bayesian data analysis pipeline, McADAM. Referring back to the prior distributions defined in Section 2.7, the cluster sampling parameters for the physical model are

$$\pi(\boldsymbol{\Theta}_{\text{cl}}) = \pi(M(r_{200}))\pi(f_{\text{gas}}(r_{200}))\pi(z)\pi(x_c)\pi(y_c). \quad (3.1)$$

x_c and y_c are the cluster centre offsets from the SA pointing centre, measured in arcseconds. The prior distributions assigned to the cluster parameters are the same as the ones used in Olamaie, Hobson, & Grainge (2013), but with an alteration to the mass limits. Upon running McADAM on data from a few of the Planck clusters, it was found that the posterior distributions of $M(r_{200})$ were hitting the lower bound $1 \times 10^{14} M_{\text{Sun}}$ used in Olamaie, Hobson, & Grainge (2013). Hence for this analysis the lower limit on $M(r_{200})$ was decreased. Table 3.2 lists the type of prior used for each cluster parameter and the probability distribution parameters.

I note here that $M(r_{500})$ (the AMI mass estimate I compare with those obtained in PSZ2) is not a sampling parameter of the physical model, but it can be calculated by evaluating equa-

tion 2.22 at $r = r_{500}$. r_{500} is calculated as part of the steps to determine the pressure profile given by equation 2.21, and so this does not cause any calculation overheads.

3.3 PSZ2 redshift values

The values of z_{Planck} used for each cluster's z prior distribution were taken to be the values stated in PSZ2. Catalogue z values are measured in the optical / infrared or X-ray, with major input from the Sloan Digital Sky Survey (York et al. 2000). A number of cluster catalogues have been extracted from these data (see e.g. Hao et al. 2010, Wen, Han, & Liu 2012, and Rykoff et al. 2014), providing estimates of both spectroscopic and photometric z values, the reliability of the latter values falls as z increases. In the X-ray part of the spectrum, the Meta-Catalogue of X-ray detected Clusters of galaxies, or MCXC (Piffaretti et al. 2011) has a substantial number of matches with the Planck-catalogue clusters. The MCXC is from the available catalogues based on the ROSAT All-Sky Survey (Voges et al. 1999) as well as serendipitous X-ray catalogues (see e.g. Gioia et al. 1990). MCXC contains only clusters with measured z , but does not state the redshift type or source. Further sources of Planck catalogue clusters candidate z s are the Russian-Turkish Telescope (Planck Collaboration Int. XXVI 2015) and the ENO telescopes in the Canary Islands (Planck Collaboration Int. XXXVI 2016); for each z these state whether it was obtained photometrically or spectroscopically.

3.4 PSZ2 methodology for deriving cluster mass estimates

For comparison with the mass values obtained with AMI data, I look at the PSZ2 mass estimates obtained from Planck data and the requisite scaling relations. The mass values published in PSZ2 are derived from data from one of three detection algorithms: MMF1, MMF3 (both of which are extensions of the matched multi-filter algorithm suitable for SZ studies (MMF, see Haehnelt & Tegmark 1996, Herranz et al. 2002 and Melin, Bartlett, & Delabrouille 2006), over the whole sky) & PowellSnakes (PwS, Carvalho et al. 2012). The former two rely on multi-frequency matched-filter detection methods, whilst PwS is a fully Bayesian method. Since the PwS methodology most closely matches the Bayesian analysis pipeline used for AMI data, I focus on the cluster parameter values from PwS. PwS will be described in more detail in Section 8.1.3.2 where I carry out Bayesian analysis on AMI and Planck datasets simultaneously, which requires extensive use of the algorithm.

The observable quantity measured by Planck is the integrated Comptonisation parameter Y . As described in Section 5 of the PSZ2 paper (Planck Collaboration XXVII 2016), for each cluster candidate there is a two-dimensional posterior of the integrated Comptonisation parameter within the radius $5r_{500}$, $Y(5r_{500})$ and the angular scale radius of the GNFW pressure, θ_p ($= r_p/D_A$). The values for $Y(5r_{500})$ published in PSZ2 are obtained by marginalising over θ_p and then taking the expected value of $Y(5r_{500})$. I refer to this value as $Y_{\text{marg}}(5r_{500})$. As described in Sections 5.2 and 5.3 of Planck Collaboration XXVII (2016), this ‘blind’ measurement of the integrated Comptonisation parameter may not be reliable when the underlying cluster pressure distribution deviates from that given by the GNFW model. To overcome this, a function relating $Y(5r_{500})$ and θ_p is derived in an attempt to provide prior information on the angular scale of the cluster based on X-ray measurements and earlier Planck mission samples. I refer to this function as the slicing function.

3.4.1 Derivation of the slicing function

The scaling relations considered here are given in Planck Collaboration XX (2014). Of particular importance to deriving the slicing function, are the $Y(r_{500}) - M(r_{500})$ and $\theta_{500} - M(r_{500})$ relations. The first of these is given by

$$E(z)^{-2/3} \left[\frac{D_A^2 Y(r_{500})}{10^{-4} \text{Mpc}^2} \right] = 10^{-0.19 \pm 0.02} \left[\frac{(1-b)M(r_{500})}{6 \times 10^{14} M_{\text{Sun}}} \right]^{1.79 \pm 0.08}, \quad (3.2)$$

where $E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_A}$ and is equal to the ratio of the Hubble parameter evaluated at redshift z to its value now for a flat Λ CDM Universe. The factor in the exponent $-2/3$ arises from the scaling relations between mass, temperature and Comptonisation parameter given by equations 1–5 in Kravtsov, Vikhlinin, & Nagai (2006). $(1-b)$ represents a bias factor, which is assumed in Planck Collaboration XX (2014) to contain four possible observational biases of departure from hydrostatic equilibrium, absolute instrument calibration, temperature inhomogeneities and residual selection bias. Its value is calculated to be $(1-b) = 0.80^{+0.02}_{-0.01}$ from numerical simulations as described in Appendix A.4 of Planck Collaboration XX (2014). Equation 3.2 uses the fitting parameters from the relation between Y_X (the X-ray ‘analogue’ of the integrated Comptonisation parameter see e.g. Kravtsov, Vikhlinin, & Nagai 2006, $Y_X(r_{500}) \equiv M_g(r_{500})T_X$ where M_g is the cluster gas mass within r_{500} and T_X is the spectroscopic temperature in the range $[0.15, 0.75]r_{500}$) and the X-ray hydrostatic mass, $M_{\text{HE}}(r_{500})$ (which is equal to $(1-b)M(r_{500})$), established for 20 local *relaxed* clusters by Arnaud et al. (2010) to give the relation between the X-ray mass proxy $M_{Y_X}(r_{500})$ and $M(r_{500})$. Finally,

the fitting parameters for the $Y(r_{500}) - M_{Y_X}(r_{500})$ relation are obtained empirically from a 71-cluster sample consisting of SZ data from the Planck Early SZ clusters (Planck Collaboration XI 2011), of Planck-detected LoCuSS clusters (Planck Collaboration Int. III. 2013) and from the XMM-Newton validation programme (Planck Collaboration IX 2011), all with X-ray data taken from XMM-Newton observations (Willis et al. 2013 and Mehrtens et al. 2012).

The $\theta_{500} - M(r_{500})$ relation is based on the equation $M(r_{500}) = 500 \times \frac{4\pi}{3} \rho_{\text{crit}}(z) r_{500}^3$ and is given by

$$\theta_{500} = 6.997 \left[\frac{h}{0.7} \right]^{-2/3} \left[\frac{(1-b)M_{500}}{3 \times 10^{14} M_{\text{Sun}}} \right]^{1/3} E(z)^{-2/3} \left[\frac{D_A}{500 \text{ Mpc}} \right]. \quad (3.3)$$

Equations (3.2) and (3.3) can be solved for $(1-b)M(r_{500})$ and equated to give $Y(r_{500})$ as a function of θ_{500}

$$Y(r_{500}) = \left[\frac{\theta_{500}}{6.997} \right]^{5.4 \pm 0.2} \left[\frac{h}{0.7} \right]^{3.60 \pm 0.13} \left[\frac{E(z)^{4.26 \pm 0.13} D_A^{3.4 \pm 0.2}}{10^{19.29 \pm 0.54} \text{ Mpc}^{3.4 \pm 0.2}} \right], \quad (3.4)$$

where $Y(r_{500})$ is in sr. Assuming a GNFW pressure profile, $Y(r_{500})$ can be converted to the corresponding value of $Y(5r_{500})$, through the relation

$$\frac{Y(r_{500})}{Y(5r_{500})} = \frac{B \left(\frac{(c_{500})^a}{1+(c_{500})^a}; \frac{3-c}{a}, \frac{b-3}{a} \right)}{B \left(\frac{(5c_{500})^a}{1+(5c_{500})^a}; \frac{3-c}{a}, \frac{b-3}{a} \right)}, \quad (3.5)$$

where $B(x, y, z) = \int_0^x t^{y-1} (1-t)^{z-1} dt$ is the incomplete beta function. For the GNFW parameter values used in equation 2.21, equation 3.5 gives a value of 0.55. Similarly, θ_{500} can be related to θ_p through the relation $\theta_p = \theta_{500}/c_{500}$.

3.4.2 Mass estimates

For a given cluster, the resulting $Y(5r_{500})$ function is used to ‘slice’ the posterior, and the value where the function intersects the posterior ‘ridge’ is taken to be the most reliable estimate of $Y(5r_{500})$, given the external information. The posterior ridge (see Figure 3.1) is defined to be the value of $Y(5r_{500})$ which gives the highest probability density for a given θ_p . The error estimates are obtained by considering where the slicing function intersects with the ridges defined by the 68% maximum likelihood confidence intervals for $Y(5r_{500})$ at each θ_p . $Y(5r_{500})$ is then converted to $Y(r_{500})$ using the reciprocal of the value given by equation 3.5, and this is used to derive a value for $M(r_{500})$ using equation 3.2, but with the $(1-b)$ term excluded. The bias term is not included in the $M(r_{500})$ calculation because it has already been accounted for in the slicing function. Note that this value of $M(r_{500})$ is what is referred to as M_{SZ} in PSZ2.

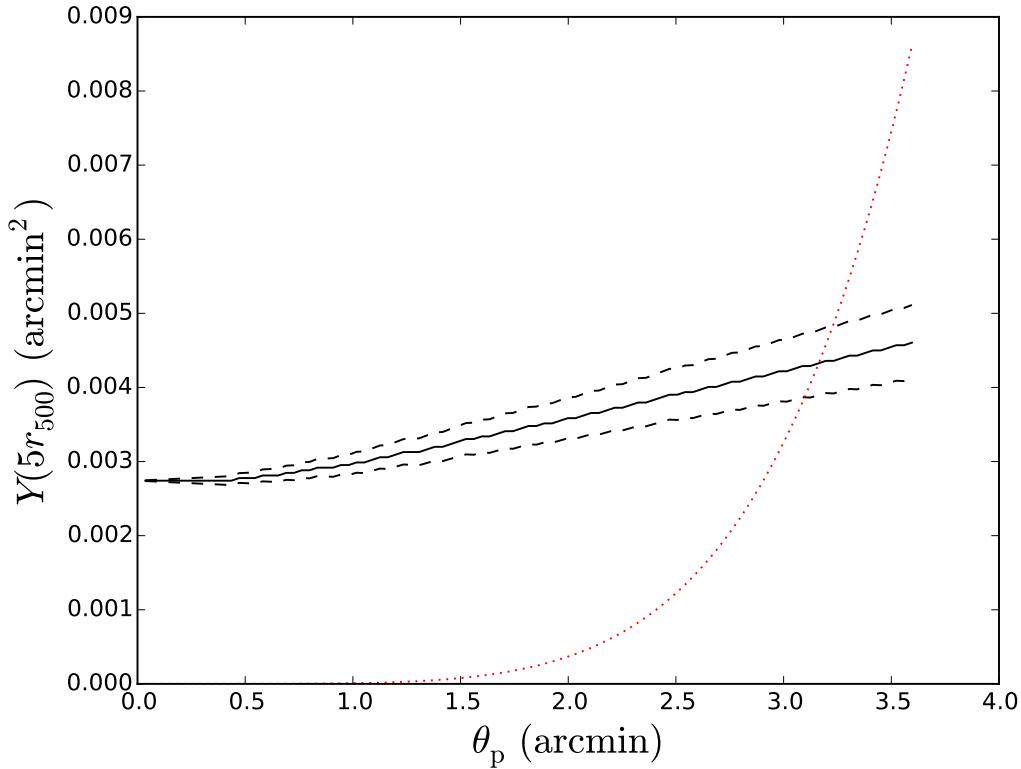


Figure 3.1: Example of the posterior slicing methodology for cluster PSZ2G228.16+75.20. The black solid line represents the ‘ridge’ (i.e. the most probable value of $Y(5r_{500})$ for each θ_p) of the posterior. The upper dashed curve represents the upper boundaries of the 68% maximum likelihood confidence interval on $Y(5r_{500})$ for each value of θ_p , and the lower dashed curve corresponds to the lower boundaries. The red dotted curve is the slicing function.

3.5 Obtaining AMI mass estimates

First I describe how I arrived at a final sample of clusters for which the AMI mass estimates are compared with those derived from Planck data.

3.5.1 Final cluster sample

3.5.1.1 Well constrained posterior sample

McADAM was used on data from the initial sample of 199 clusters. MULTINEST failed to produce posterior distributions for two clusters. These clusters were surrounded by high flux, extended radio-sources. Of the 197 clusters for which posterior distributions were produced, 73 clusters show good constraints (adjudged by physical inspection) on the sampling parameters $M(r_{200})$, $f_{\text{gas}}(r_{200})$, x_c and y_c ; with z_s ranging from 0.089 to 0.83.

I illustrate a ‘well constrained’ posterior distribution (for cluster PSZ2G184.68+28.91) in the first half of Figure 3.2, plotted using GETDIST^a (a kernel density estimation algorithm, which is described in Section 9.5.2). In contrast the second half of Figure 3.2 is an example of a cluster (PSZ2G121.77+51.75) which shows poor constraints on mass as the posterior distribution is peaked at the lower boundary of the mass sampling range ($5 \times 10^{13} M_{\text{Sun}}$) which could not be classed as a detection within our mass prior range. I also note that in the latter case the mass posterior largely resembles the prior distribution.

3.5.1.2 Moderate radio-source environment sample

For the 197 cluster sample, AMI data maps were produced using the software package AIPS^b using the automated CLEAN procedure with a limit determined using IMEAN. Source-finding was carried out at four σ on the LA continuum map, as described in Davies et al. (2011) and Franzen et al. (2011). For each cluster both a non-source-subtracted and a source-subtracted map was produced. The values used to subtract the sources from the maps were the mean values of the one-dimensional marginalised posterior distributions of the sources’ position, flux and spectral index produced by McADAM. Maps of the 73 cluster sample were inspected in detail. It was found that for seven of these clusters, even though the posterior distributions were well constrained, that the radio-source and primordial CMB contamination could bias the cluster parameter estimates in an unpredictable way. In these cases it was found that the subtracted maps contained residual flux close to the cluster centre, from either radio-sources (some of which were extended), radio-frequency interference, or CMB. PSZ2G125.37-08.67 is an example of one of these clusters and its non-source-subtracted and source-subtracted maps are shown in Figure 3.3. I thus arrived at a 66 cluster sample.

^a<http://getdist.readthedocs.io/en/latest/>.

^b<http://aips.nrao.edu/>.

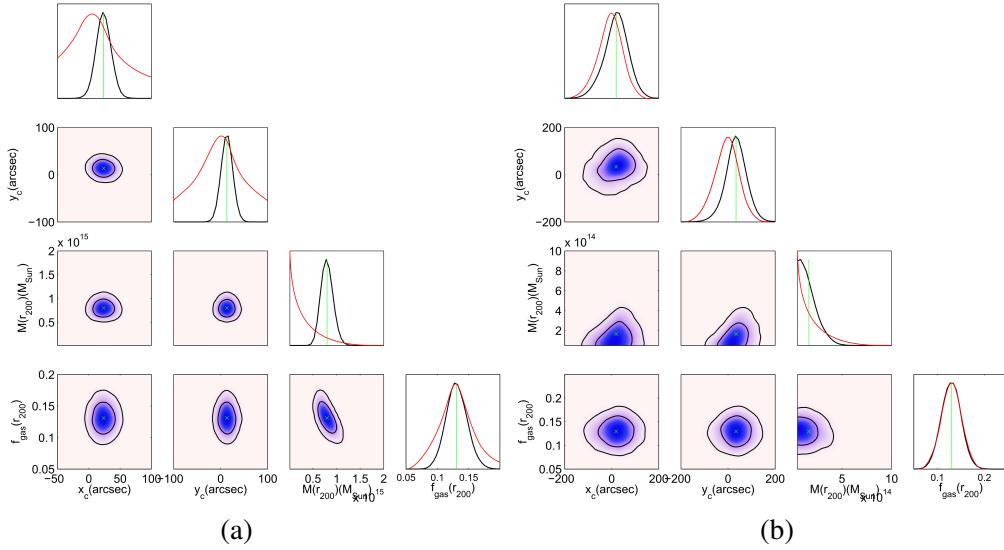


Figure 3.2: Posterior distributions derived from AMI data for the sampling parameters: $M(r_{200})$; $f_{\text{gas}}(r_{200})$; x_c & y_c . The contoured maps show the two-dimensional posteriors for the different pairs of parameters. The contours represent the 95% and 68% mean confidence intervals, with the green crosses denoting the expected value of the joint distributions. The four one-dimensional plots are the marginalised posteriors corresponding to the variable given at the bottom of the respective column. The red curves are the prior distributions on the relevant parameters. Each green line is the expected value of the distribution. Posterior distributions in (a) show narrow distributions on the cluster mass, with the domain spanning feasible mass values for a galaxy cluster (cluster PSZ2G184.68+28.91). In such cases the posteriors are said to be well constrained. The mass posteriors in (b) show that the data imply unphysical values for its mass, as the posterior distribution is hitting the lower bound of the prior ($5 \times 10^{13} M_{\text{Sun}}$) at almost its peak value (cluster PSZ2G121.77+51.75). The distribution also resembles the uniform in log-space prior assigned to $M(r_{200})$. In such cases the posteriors are said to be poorly constrained with respect to the mass estimates.

3.5.1.3 Well defined cluster-centre sample

The posteriors of x_c and y_c give the position of the modelled cluster centre relative to the actual SA pointing centre used for the observation. For seven of the 66 cluster sample, it was found that the mean posterior values of x_c and y_c changed dramatically between different runs of McADAM (on the same cluster data), by up to 70 arcseconds in either direction, leading to differences in mass estimates of up to 70%. The estimates for these clusters are not reliable, since the model was creating a completely different cluster between runs, and so these clusters were excluded leaving a 59 cluster sample. For the remaining clusters, the change in $M(r_{200})$ between runs was

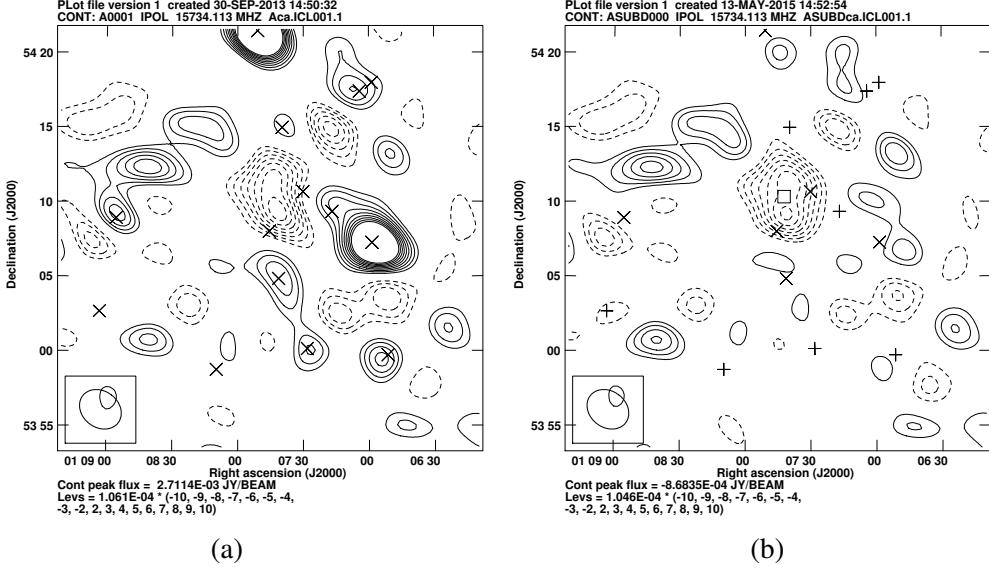


Figure 3.3: (a) Unsubtracted map produced from AMI observation. Contours are plotted at $\pm(2, 3, 4, \dots, 10) \times$ the r.m.s. noise level, and dashed contours are negative. (b) Source subtracted map produced from AMI observation. The \square denotes the McADAM-determined centre of the cluster (posterior mean values for x_c and y_c). Here ‘+’ signs denote radio-source positions as measured by the LA which were assigned delta priors on their parameters, whilst ‘x’ denote sources which were assigned priors as described in Section 2.7.1.

much smaller than the standard deviation of the corresponding posterior distributions. Figure 3.4 shows the subtracted map for PSZ2G183.90+42.99, which we consider to be an example of a cluster with an ill-defined centre. The map shows three flux decrement peaks close to the cluster centre. Movement of the centre between these peaks with the current source environment modelling would lead to a change in the size of the predicted cluster, and consequently different mass estimates each time.

3.5.1.4 PwS detected cluster sample

For five of the 59 cluster sample, the data available on the Planck website^c did not contain a detection using the PwS algorithm, and so no mass estimates based on PwS data could be calculated. Hence the final sample size for which I present the mass estimates from both AMI and

^c<https://pla.esac.esa.int/pla/catalogues>.

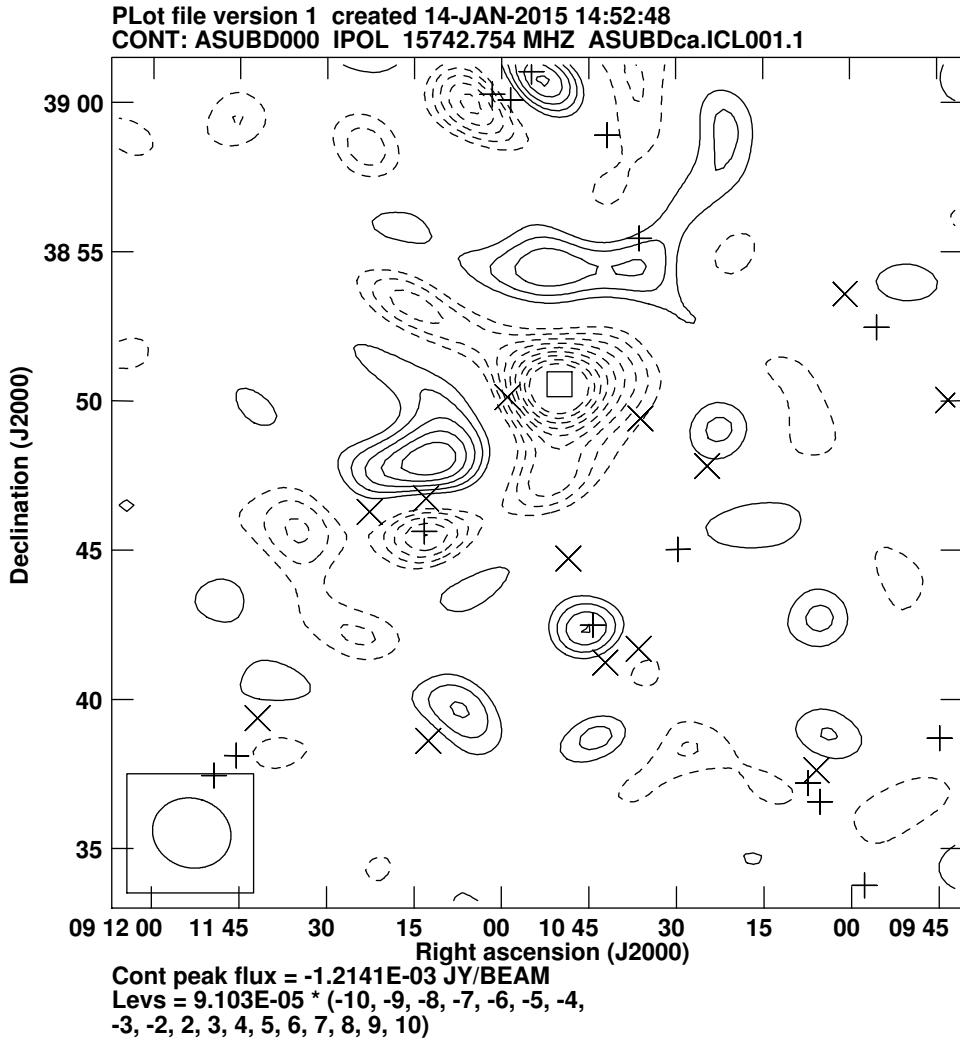


Figure 3.4: Subtracted map of cluster with ill-defined centre. The cluster is clearly offset from the observation pointing centre (middle of the map), and the lobes to the bottom and the top left of the cluster cause the centre position to be ambiguous.

Planck data is 54.

It is important to realise that selection biases are introduced in reducing the sample size from 199 to 54. In particular, selecting only the clusters which showed good AMI posterior constraints means that clusters corresponding to a signal too faint for AMI to detect, clusters with large enough angular size for AMI's shortest baselines not to be able to measure the signal

from the outskirts of the cluster ("resolved clusters"), and clusters where the radio-source and CMB contamination dwarfs the signal of the cluster, are all likely to have been excluded from the sample to some extent. In addition, removing the seven clusters with an ill defined centre likely removes some unrelaxed clusters from the sample.

3.6 AMI and PSZ2 mass estimates

The AMI and PSZ2 parameter estimates for the 54 clusters are given in Table A.1 in Appendix A. The clusters are listed in ascending order of z . Note that whether a redshift is photometric or spectroscopic is stated in the fifth column. All AMI values are the mean values of the corresponding parameter posterior distributions, with the error taken as the standard deviation. The estimates of the sampling parameters are included for comparison with each other, and with the sampling prior ranges and associated parameters given in Table 3.2. The AMI values for $M(r_{500})$ are given for comparison with the corresponding PSZ2 estimates. Two values for the PSZ2 mass estimates are given, $M_{\text{Pl, marg}}(r_{500})$ and $M_{\text{Pl, slice}}(r_{500})$. $M_{\text{Pl, marg}}(r_{500})$ corresponds to the mass given by the $Y(r_{500}) - M(r_{500})$ relation when the marginalised integrated Comptonisation parameter is used as described in Section 3.4. The uncertainties associated with these Y values are taken as the standard deviations of the marginalised posteriors. $M_{\text{Pl, slice}}(r_{500})$ is detailed in Section 3.4.2; its associated errors are calculated from the $Y(5r_{500})$ values where the slicing function intersects with the two ridges formed by the 68% maximum likelihood confidence interval values of the $Y(5r_{500})$ probability densities over the posterior domain of θ_p . Figure 3.5 shows $M(r_{200})$ as a function of z . Excluding the clusters at $z = 0.089, 0.4$ and 0.426 , there is a steepening in mass between $0.1 \lesssim z \lesssim 0.5$ before it flattens off at higher z . This result is consistent with the PSZ2 mass estimates presented in [Planck Collaboration XXVII \(2016\)](#).

I now focus on the comparison between AMI and Planck mass estimates. Note that [Planck Collaboration XXVII \(2016\)](#) do not provide any means for estimating $M(r_{200})$ from their data, as r_{200} is the distance related to the scale radius ($r_{200} = c_{200} \times r_s$) for the NFW dark matter profile given by equation 2.19, which they do not incorporate into their modelling process. Figure 3.6 gives the AMI and two Planck estimates for $M(r_{500})$ vs the row number, in Table A.1. I have not used z as the independent variable in this plot for clarity. The row number is monotonically related to z , as Table A.1 is sorted by ascending z . From Figure 3.6 it is clear that AMI underestimates the mass relative to both PSZ2 values. In fact $M(r_{500})$ is lower than $M_{\text{Pl, slice}}(r_{500})$ in 37 out of 54 cases. $M(r_{500})$ is lower than $M_{\text{Pl, marg}}(r_{500})$ in 45 out of 54 cases. 31 of the AMI masses are within one combined standard deviation of $M_{\text{Pl, slice}}(r_{500})$, while 46 are within two.

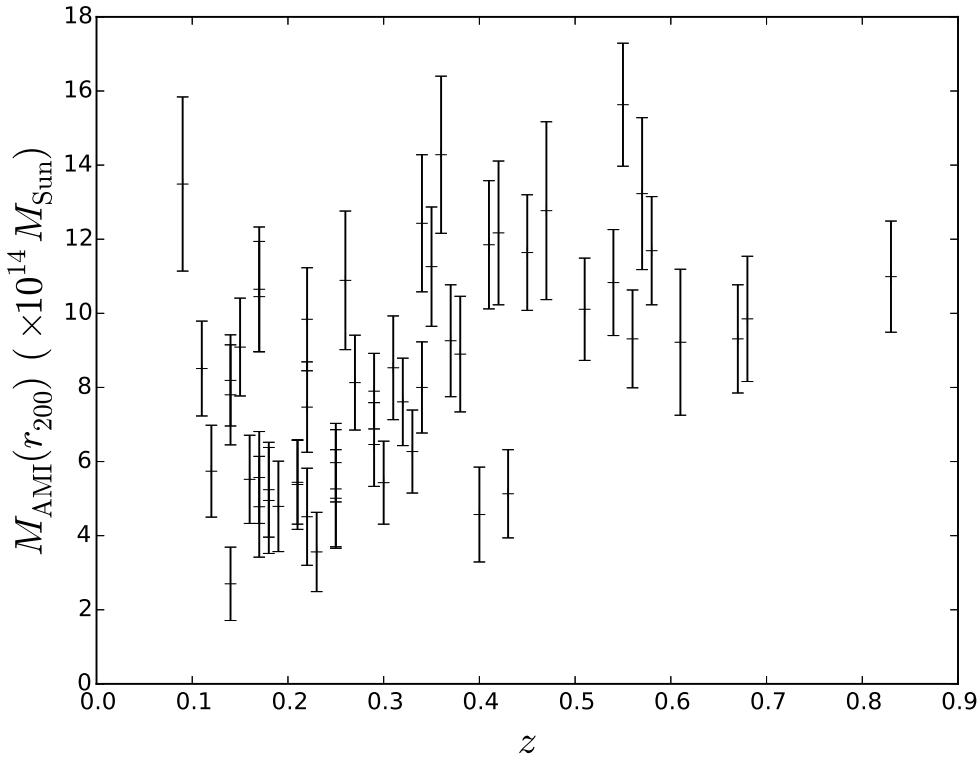


Figure 3.5: Plot of $M(r_{200})$ derived from AMI data using physical modelling vs redshift for the sample of 54 clusters.

Four clusters have discrepancies larger than three combined standard deviations. Three of these clusters are at relatively low redshift (≤ 0.25), whilst one is at $z = 0.43$.

It is also noteworthy that $M_{\text{Pl, marg}}(r_{500})$ is larger than $M_{\text{Pl, slice}}(r_{500})$ in 47 out of 54 cases. This implies that the additional information obtained from X-ray data incorporated in the slicing function consistently predicts a lower mass cluster than from the Planck SZ data alone.

Figure 3.7 shows the ratios of the mass estimates between the three different methods. The most obvious thing to note is that the ratio of PSZ2 masses is consistently greater than one, which again emphasises the fact that the marginalisation method attributes a much higher mass to the clusters than the slicing method. Furthermore, the ratio of AMI mass to the marginalised mass is small at medium redshift, which suggests that the marginalised mass is systematically high in this range. This graph also emphasises that the AMI mass and the slicing methodology mass are

the most consistent with one another.

3.7 AMI simulations with PSZ2 mass inputs

To investigate further the discrepancies between the mass estimates, it was decided to create simulated data based on the PSZ2 mass estimates obtained from the slicing methodology, which were then ‘observed’ by AMI. The data from these simulated observations were analysed the same way as the real data. The simulations were carried out using the in-house AMI simulation package PROFILE, which has been used in various forms in e.g. [Grainge et al. \(2002\)](#), [Davies et al. \(2011\)](#), [Olamaie et al. \(2012\)](#) and [Olamaie, Hobson, & Grainge \(2013\)](#). The input parameters for the simulation— which uses the physical model to create the cluster— are the sampling parameters of the model. Since [Planck Collaboration XXVII \(2016\)](#) does not give a method for calculating $M(r_{200})$ it was calculated as follows. First r_{500} was calculated by solving equation 2.22 with $M(r_A) = M_{\text{SZ}}$ and $r_A = r_{500}$. r_{200} can be determined from r_{500} , but we note that the function mapping from r_{200} to r_{500} is non-invertible, thus r_{200} had to be calculated by solving equation 2.25 iteratively. $M(r_{200})$ can then be calculated by evaluating equation 2.22 at r_{200} .

As well as the values of $M(r_{200})$ derived from PSZ2 mass estimates, values for the other inputs were also required. I used $f_{\text{gas}}(r_{200}) = 0.13$, $z = z_{\text{Planck}}$, and $x_c = y_c = 0$ arcsec.

The objective of these simulations was to see whether we could recover the mass input into the simulation to create a cluster using the physical model, ‘observed’ by AMI and then analysed using the same model. I tried this for the four sets of simulations described below.

For each simulation different noise / canonical radio-source environment realisations (where relevant) were used each time. Due to the large sample size this should not affect any systematic trends seen in the results, and it avoids having to pick a particular realisation to be used in all the simulations.

3.7.1 Simulations of clusters plus instrumental noise

For each cluster, $M(r_{200})$ was calculated and Gaussian instrumental noise (Section 2.5.2.1) was added to the sky. The RMS of the noise added was 0.7 Jy per channel per baseline per second, a value typical of an AMI cluster observation. Figure 3.8 shows the map produced from the simulated data of cluster PSZ2G044.20+48.66 plus this instrumental noise. The mass estimate derived from the Bayesian analysis of this cluster is 0.56 standard deviations above the input value.

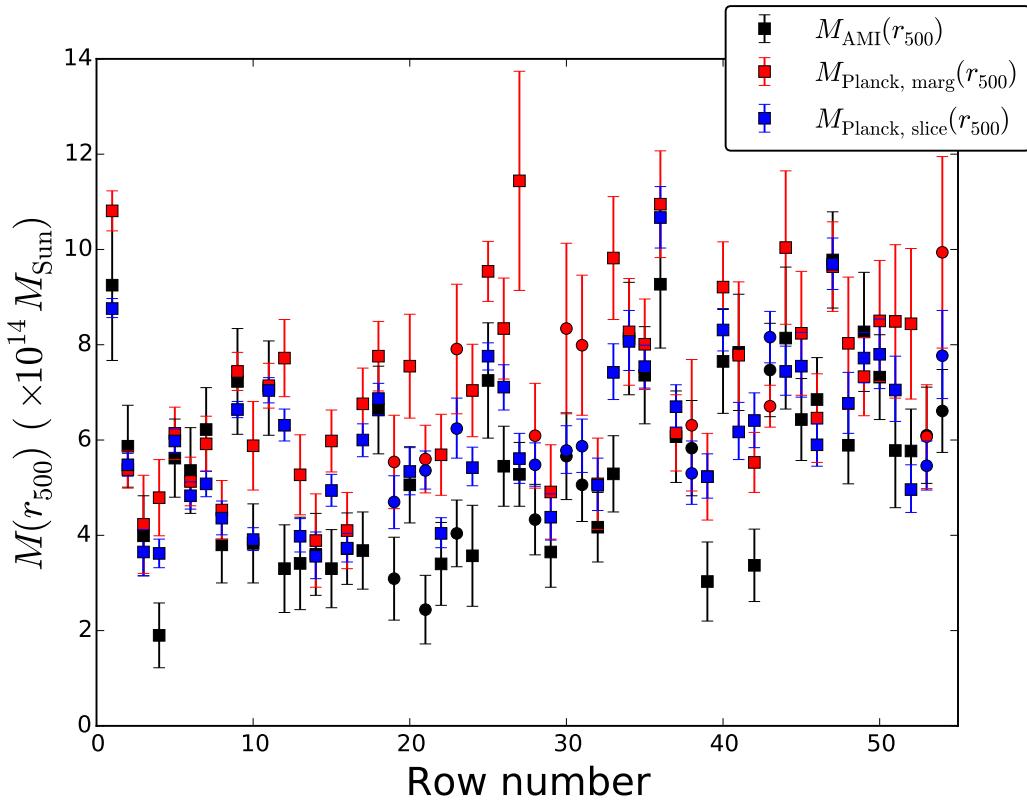


Figure 3.6: Plot of $M(r_{500})$ vs row number of Table A.1 for three different cases: the value derived from AMI data using the physical model, $M_{\text{AMI}}(r_{500})$; the value derived from Planck data using the marginalised value for $Y(5r_{500})$, $M_{\text{Pl, marg}}(r_{500})$ and the value derived from Planck data using the slicing function value for $Y(5r_{500})$, $M_{\text{Pl, slice}}(r_{500})$. The row number is monotonically related to z , as Table A.1 is sorted by ascending z . The points with circular markers correspond to clusters whose redshifts were measured photometrically (as listed in Table A.1).

Figure 3.9 shows the difference between the input masses and the ones recovered from running the simulated observations through McADAM, visualised using a histogram. All but three of the clusters lie within one standard deviation of the input mass, and even these clusters (PSZ2G154.13+40.19, PSZ2G207.88+81.31 and PSZ2G213.39+80.59) give an output mass 1.01, 1.26 and 1.08 standard deviations below the input mass.

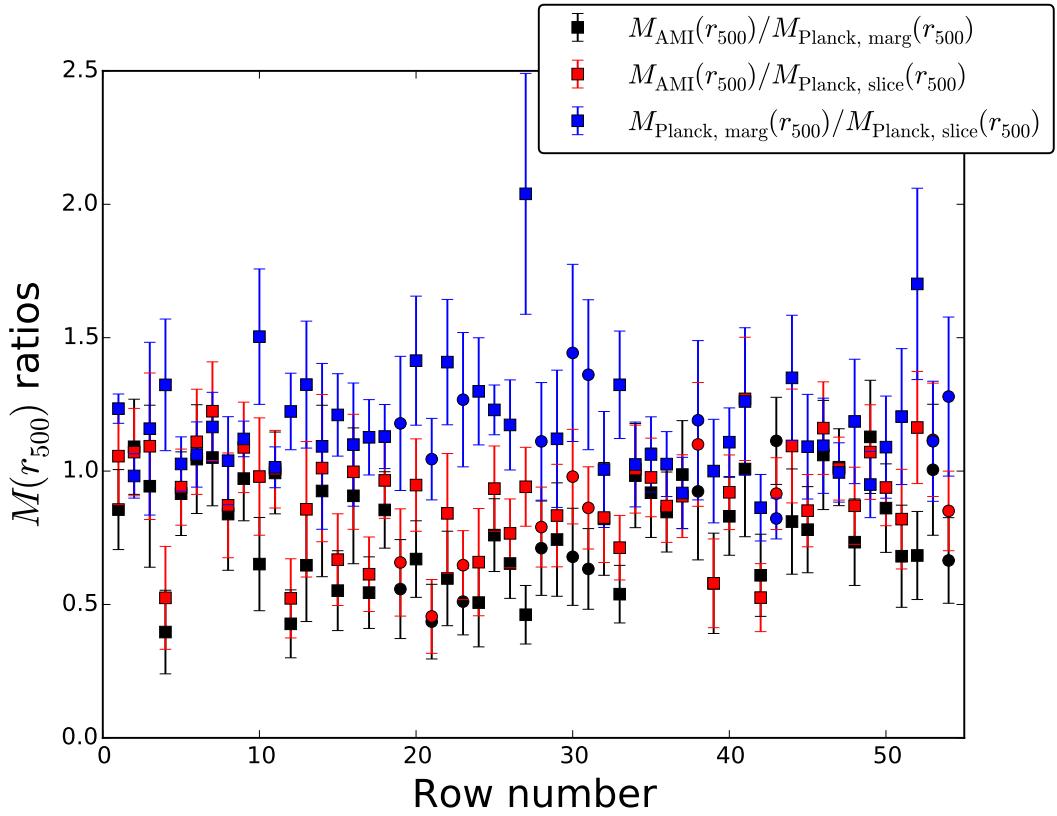


Figure 3.7: Plot of $M(r_{500})$ ratios vs row number of Table A.1 for three different cases: $M_{\text{AMI}}(r_{500})/M_{\text{Pl, marg}}(r_{500})$; $M_{\text{AMI}}(r_{500})/M_{\text{Pl, slice}}(r_{500})$ and $M_{\text{Pl, marg}}(r_{500})/M_{\text{Pl, slice}}(r_{500})$. The points with square markers correspond to clusters whose redshifts were measured spectroscopically, and the circular markers correspond to photometric redshifts (as listed in Table A.1).

3.7.2 Simulations further adding confusion noise and primordial CMB

Confusion noise is defined to be the flux from radio-sources below a certain limit (see Section 2.5.2.3, here $S_{\text{conf}} = 0.3$ mJy). In this Section all radio-source realisations only contribute to the confusion noise. However in Sections 3.7.3 and 3.7.4 sources above S_{conf} are included. The confusion noise contributions (see e.g. Section 5.3 of FF09) were sampled from the probability density function corresponding to the 10C source counts given in Davies et al. (2011), and placed at positions chosen at random. Similarly, the primordial CMB (Section 2.5.2.2) realisations were sampled from an empirical distribution (Hinshaw et al. 2013), and randomly added to the maps.

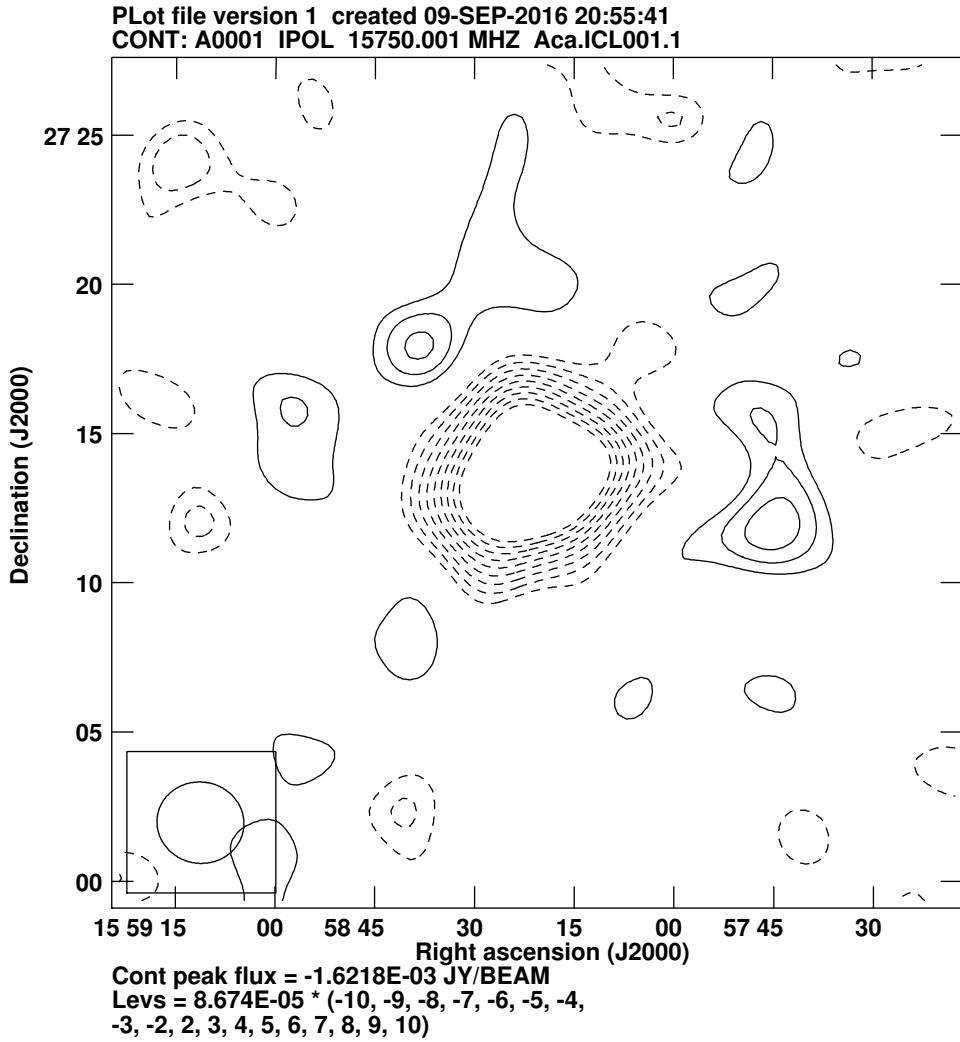


Figure 3.8: Unsubtracted map produced from simulated AMI data of cluster PSZ2G044.20+48.66, including instrumental noise.

Figure 3.10 shows the map produced from the simulated data of cluster PSZ2G044.20+48.66, including the three noise contributions. The mass estimate derived from the Bayesian analysis of this cluster is 0.22 standard deviations above the input value. The differences between output and input masses are shown in Figure 3.11. This time eight out of the 54 clusters cannot recover the input mass to within one standard deviation. In all eight of these cases, the mass is underestimated with respect to the input value. Five of the outlier values correspond to clusters

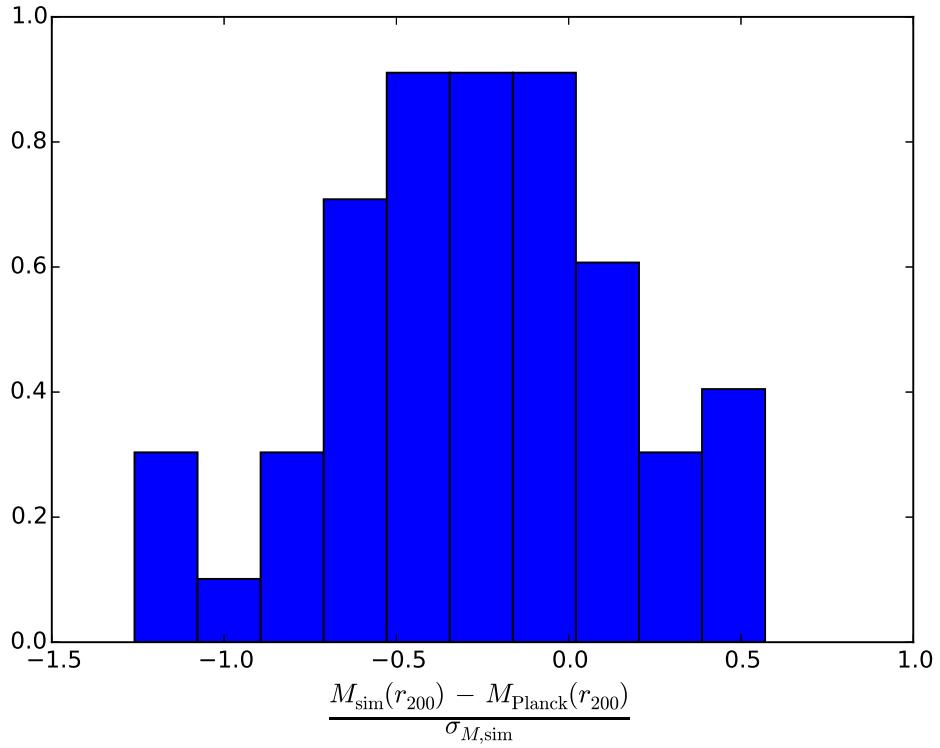


Figure 3.9: Normalised histogram of the differences between the input and output masses of the AMI simulations including the cluster and instrumental noise only, in units of standard deviations of the output mass.

at low redshift ($z < 0.2$). This suggests that the confusion and CMB noise may be causing AMI to systematically underestimate the cluster masses, and may explain why AMI mass estimates were consistently lower than those obtained by Planck for the real data.

3.7.3 Simulations further adding a canonical radio-source environment

The third set of simulations included detectable radio-sources (Section 2.5.1, which formed a canonical radio-source environment. They were created in the same way as with the confusion noise described above, but with higher flux limits so that in reality, the LA would have been able to detect them. The upper flux limit was set to 25 mJy.

Figure 3.12 shows the map produced from the simulated data of cluster PSZ2G044.20+48.66,

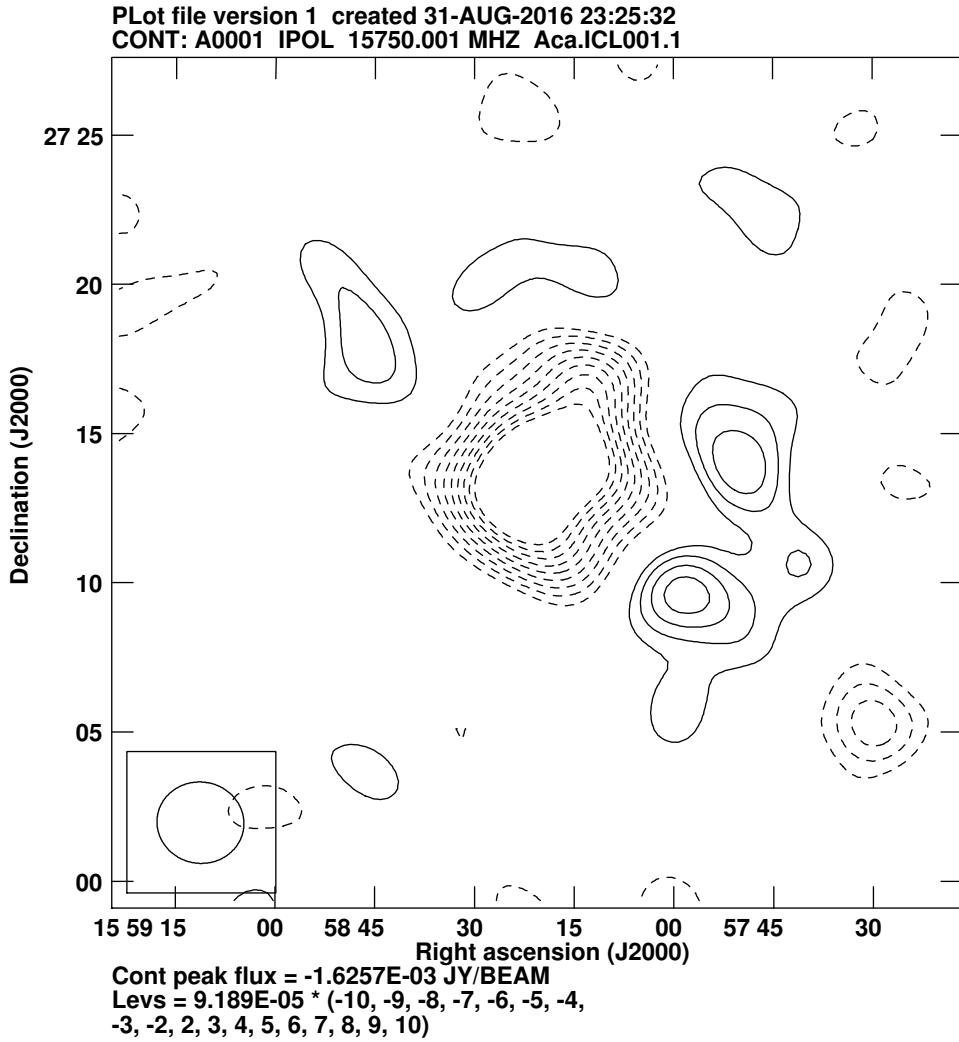


Figure 3.10: Unsubtracted map produced from simulated AMI data of cluster PSZ2G044.20+48.66, including instrumental, confusion and CMB noise.

including a canonical source environment and background noise. The mass estimate derived from the Bayesian analysis of this cluster is 0.51 standard deviations below the input value. Figure 3.13 shows that the canonical radio-source environment have little effect on the mass estimation relative to Section 3.7.2, as there are still 8 clusters which give mass estimates greater than one standard deviation away from the input value. Note that in this case, the outliers occurred across the entire range of redshifts, which suggests that in Section 3.7.2 the low redshift

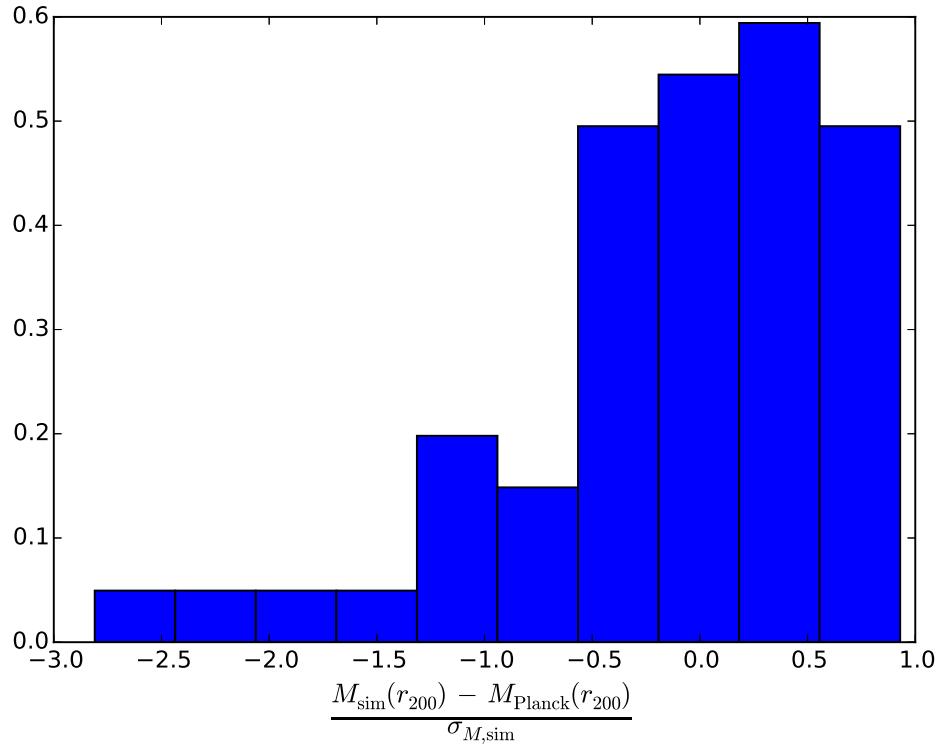


Figure 3.11: Normalised histogram of the differences between the input and output masses of the AMI simulations, in units of standard deviations of the output mass. This is the case for instrumental, confusion and CMB noise contributions.

trend was just a coincidence.

3.7.4 Simulations with LA observed radio-source environment plus instrumental, confusion and CMB noise

The final set of simulations included the radio-source environment measured by the LA during the real observation for each cluster. These are only estimates of the actual source environments, and are only as reliable as the LA's ability to measure them. Figure 3.14 shows the maps produced from the real & simulated data of cluster PSZ2G044.20+48.66. The mass estimate derived from the Bayesian analysis of the simulated dataset is just 0.08 standard deviations above the input value.

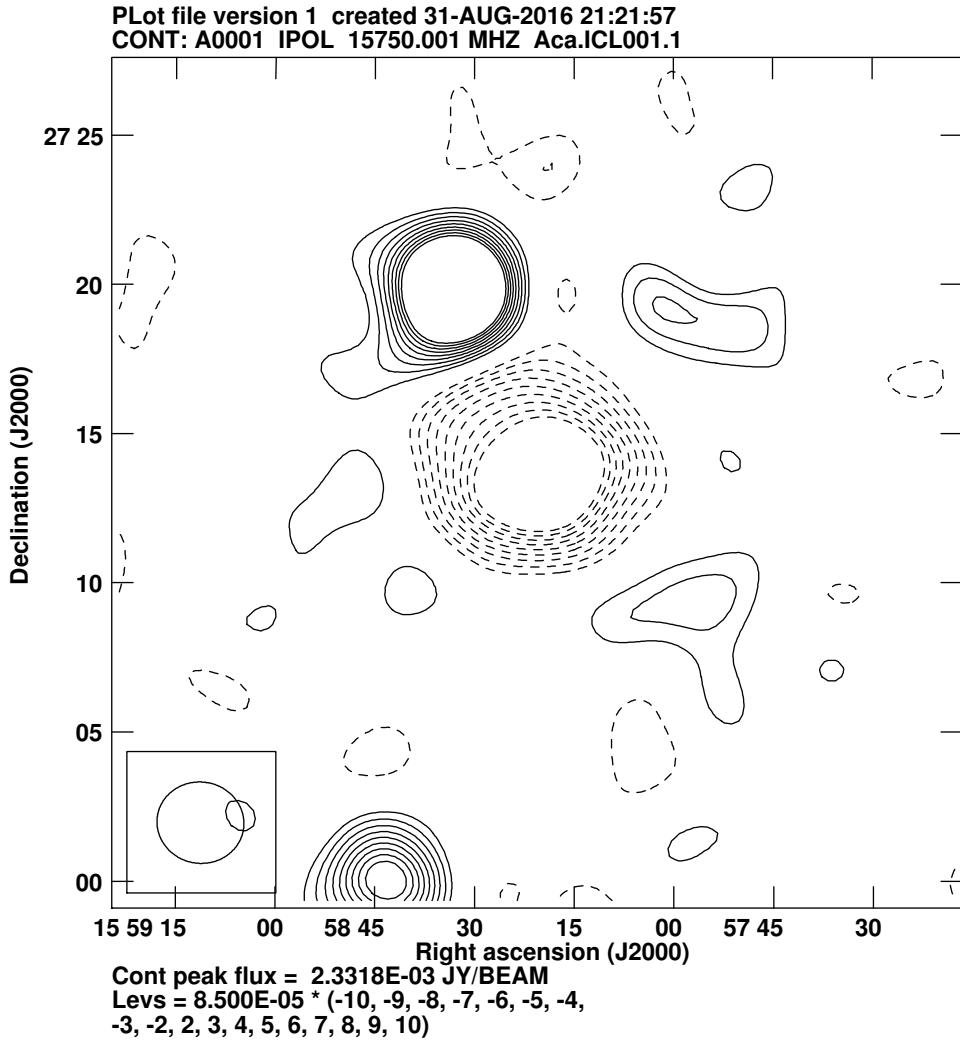


Figure 3.12: Unsubtracted map produced from simulated AMI data of cluster PSZ2G044.20+48.66, including a canonical radio-source environment as well as instrumental, confusion and CMB noise.

Figure 3.15 shows that including the LA observed radio-source environment has a large effect on the results, as this time there are 16 clusters which are more than one standard deviation away from the input mass. Furthermore, three of these overestimated the mass relative to the input, the first time we have seen this occur in any of the simulations. A possible source of bias could be due to for example, the empirical prior on the spectral index incorrectly modelling some radio-sources. Another source of bias could be the position of a source relative to the cluster, and the

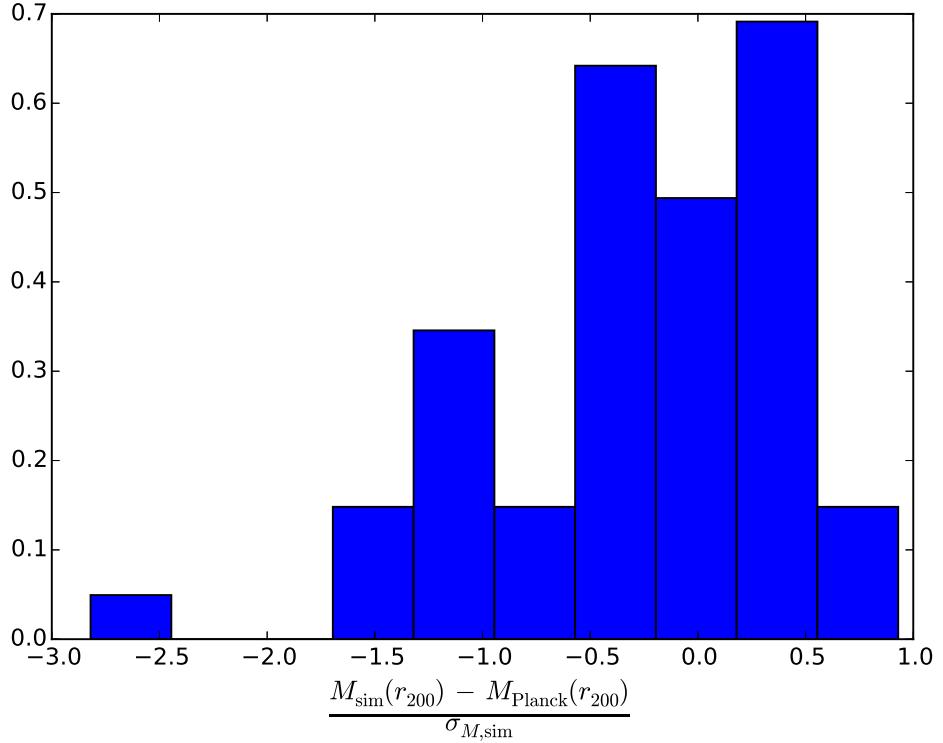


Figure 3.13: Normalised histogram of the differences between the input and output masses of the AMI simulations, in units of standard deviations of the output mass. This is the case for a canonical radio-source environment as well instrumental, confusion and CMB noise contributions.

magnitude of the source flux. For example, if a high flux radio-source is close to the centre of the galaxy cluster, then even a slight discrepancy between the real and the modelled values for the source could have a large effect on the cluster parameter estimates.

I now compare these results to the simulations in YP15 (which concluded that the underestimation of the simulation input values could be due to deviation from the ‘universal’ profile, see Figure 23a in the paper). The results of the large cluster simulations (total integrated Comptonisation parameter = 7×10^3 arcmin 2 and $\theta_p = 7.4$ arcmin) in YP15 seem biased low at a more significant level than those in Figure 3.15, as in the former case less than half of the clusters recover the true value within two standard deviations. For the smaller clusters however, YP15 found a slight upward bias in the simulation results, but this is probably smaller in magnitude

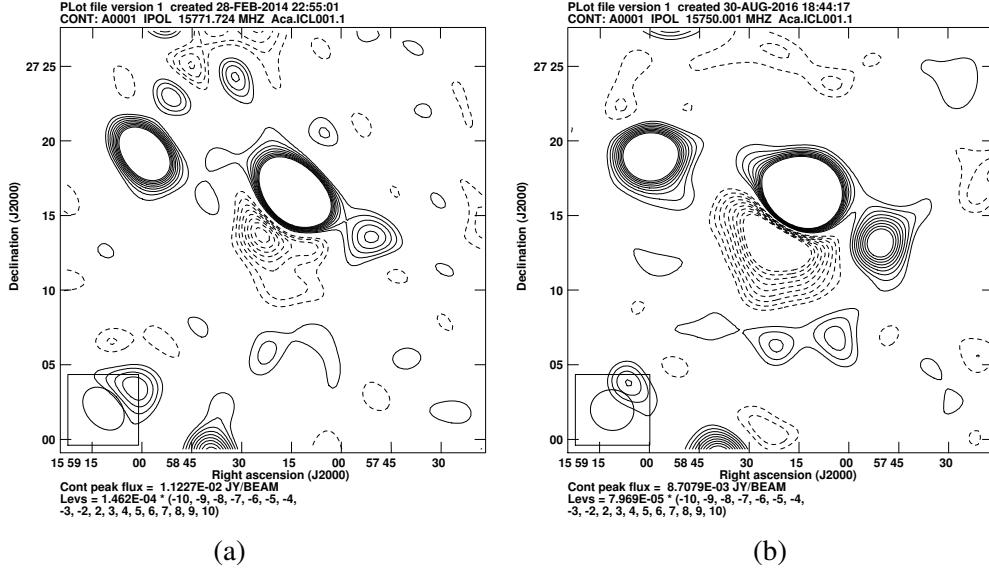


Figure 3.14: (a) Unsubtracted map produced from real AMI data of cluster PSZ2G044.20+48.66. (b) Unsubtracted map produced from simulated AMI data of PSZ2G044.20+48.66, including the real source environment (as measured by the LA) as well as instrumental, confusion and CMB noise. The peak flux in the simulation has been underestimated relative to the real observation by $\approx 25\%$. This could be due to the source sitting on a negative decrement caused by background noise, or it could be from the cluster decrement.

than the bias found in this Section.

3.7.5 Statistics of results of real and simulated data

Looking at the histograms produced in Sections 3.7.1, 3.7.2, 3.7.3, and 3.7.4, in the last three cases it is apparent that there is a negative skew in the data, i.e. the output masses are negatively biased relative to the input masses. The skews calculated from the samples associated with the four histograms are -0.17 , -1.30 , -0.91 , and -0.96 respectively in units of standard deviations of the output mass. This suggests that the inclusion of confusion and CMB noise bias the cluster mass. I also calculate the median values of these histograms, and compare them with the medians corresponding to the real AMI and PSZ2 masses given in Figure 3.6. The median values for the four histograms are -0.24 , 0.09 , -0.27 and -0.34 respectively in units of standard deviations of the output mass. For the real data the median values for $(M_{\text{AMI}}(r_{500}) - M_{\text{Pl, marg}}(r_{500})) / \sigma_{\text{AMI}}$ and

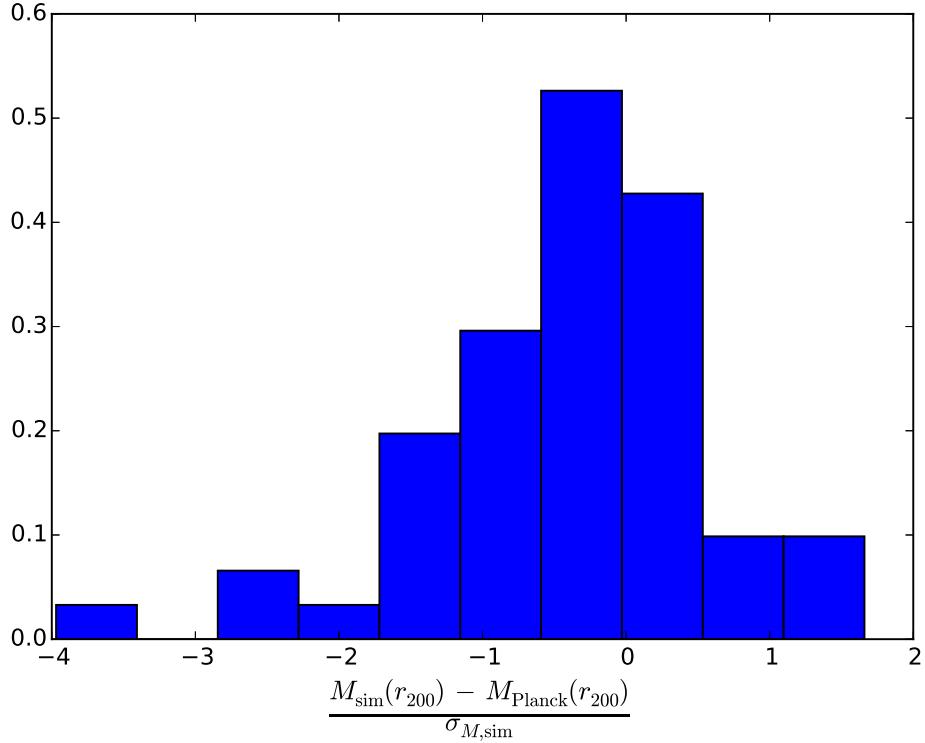


Figure 3.15: Normalised histogram of the differences between the input and output masses of the AMI simulations, in units of standard deviations of the output mass. This is the case for the real radio-source environment as measured by the LA, with instrumental, confusion and CMB noise contributions.

$(M_{\text{AMI}}(r_{500}) - M_{\text{Pl, slice}}(r_{500})) / \sigma_{\text{AMI}}$ are -1.57 and -0.56 . It makes sense to compare the second of these real data values with those obtained from the simulations, as it was $M_{\text{Pl, slice}}(r_{500})$ which was used to derive the input masses. The fact that the median from the real data is greater in magnitude than the values from the simulations implies in general, our simulations can recover their input values with better agreement than that obtained between real AMI estimates and those obtained from Planck data using the slicing function methodology. This seems plausible as you would expect that inferring results from data which was created using the same model used in the inference would be more accurate than results from data taken from two different telescopes, which use different models in their inference. Furthermore the simulation medians tell us that AMI is capable of inferring the masses derived with the slicing methodology, if the cluster is

created using the model used in the inference and assuming there are no large discrepancies between the real and simulated AMI observations.

3.8 Conclusions

We have made observations of galaxy clusters detected by the Planck space telescope, with the Arcminute Microkelvin Imager (AMI) radio interferometer system in order to compare mass estimates obtained from their data. I analysed this data using the physical model described in Section 2.4, following largely the data analysis method outlined in Feroz et al. (2009). This allowed us to derive physical parameter estimates for each cluster, in particular the total mass out to a given radius. I have also calculated two mass estimates for each cluster from Planck’s PowellSnakes detection algorithm (Carvalho et al. 2012) data following Planck Collaboration XVII (2016) (PSZ2), and found the following.

- For the AMI mass estimates of Planck selected clusters there is generally a steeping in the mass of galaxy clusters as a function of redshift, which flattens out at around $z \approx 0.5$.
- AMI $M(r_{500})$ estimates are within one combined standard deviation of the PSZ2 slicing function mass estimates for 31 out of the final sample of 54 clusters. However, the AMI masses are lower than both PSZ2 estimates for 37 out of the 54 cluster sample.
- The PSZ2 mass estimates derived from the marginalised $Y - \theta$ posteriors are larger than those which use the slicing function in 47 out of 54 cases. This suggests that the X-ray data which form the basis of the slicing procedure predict lower cluster masses relative to what the SZ Planck data alone find.

To investigate further the possible biasing of AMI mass estimates, I created simulations of AMI data with input mass values from the PSZ2 slicing methodology. I considered four different cases for the simulations: 1) galaxy cluster plus instrumental noise; 2) galaxy cluster plus instrumental plus confusion & CMB noise; 3) galaxy cluster plus instrumental, confusion & CMB noise, plus a randomly positioned radio-source environment; 4) galaxy cluster plus instrumental, confusion & CMB noise, plus the radio-source environment recognised by the LA in the real observations. These simulated datasets were analysed in the same way as the real datasets, and I found the following.

- For case 1), the physical model recovered the input mass to within one standard deviation for 51 of the 54 clusters. The three which did not give an underestimate relative to the masses input to the simulation.

- For case 2), eight of the simulations gave results which were more than one standard deviation lower than the input values. This highlights the effect of incorporating the noise sources into the error covariance matrix rather than trying to model the associated signals explicitly.
- Case 3) shows similar results to case 2), which implies that ‘ideal’ radio-sources placed randomly in the sky have little effect on cluster mass estimates.
- However in case 4) with real source environments, 16 simulations did not recover the input mass to within one standard deviation. This suggests that real radio-source environments, which can include sources with high flux values, and often sources which are located very close to the cluster centre, introduce biases in the cluster mass estimates. In real observations there are also additional issues (the sources are not ‘ideal’), such as sources being extended and emission not being circularly symmetric on the sky.
- Cases 2), 3) and 4) give distributions of output – input mass which are negatively skewed. Thus AMI mass estimates are expected to be systematically lower than the PSZ2 slicing methodology values.
- The median values of the distributions of output – input mass of the simulations in each of the four cases are smaller in magnitude than those obtained from comparing AMI and PSZ2 estimates from real data. This is expected as I used the same model to simulate and analyse the clusters in all four cases.
- Compared to the results of simulations of large clusters carried out in [Perrott et al. \(2015\)](#), which test the robustness of the ‘universal’ pressure profile, the case 4) bias appears relatively small in magnitude, and in the same direction (downward). When comparing the case 4) results with the small cluster simulations of [Perrott et al. \(2015\)](#), the latter shows a relatively small bias in the opposite direction.
- The simulated and real data medians also indicate that while the simulations have shown that AMI mass estimates are systematically low, this does not fully accommodate for the discrepancies in the results obtained from the real data. This suggests that there is a systematic difference between the AMI & Planck data and / or the cluster models used to determine the mass estimates (which generally leads to PSZ2 estimates being higher than those obtained from AMI data).

CHAPTER 4

COMPARISON OF PHYSICAL AND OBSERVATIONAL GALAXY CLUSTER MODELLING USING AMI DATA

This Chapter provides a follow-up to Chapter 3 in which I performed Bayesian inference on data obtained with the Arcminute Microkelvin Imager (AMI) array to derive estimates of physical properties of clusters that have been detected by Planck. I now focus on the observational properties of clusters obtained from telescopes such as AMI and Planck which measure the SZ effect: the angular radius θ , and the integrated Comptonisation parameter Y . For the sample considered in the previous Chapter, we compare observational parameters derived from the physical model with those obtained from two observational models similar to the one described in YP15 and [Olamaie et al. \(2012\)](#), using data from AMI. I also compare the different models using Bayesian analysis as described in Section 2.6.2, as well as with another technique presented here (see Section 4.3.2). The work discussed in this Chapter has been submitted to MNRAS and is under review ([Javid et al. 2018](#)).

4.1 Physical model estimates of observational parameters

Y can be calculated using the physical model (PM from here on in this chapter) presented in Section 2.4, by first calculating $P_e(r)$ and then calculating $Y(r)$ using equation 2.18. θ and r are related through $\theta = r/D_A$. The prior distributions used are the same as the ones used in the

previous Chapter.

4.2 Observational models

Here I consider two observational models, observational model I (OM I) and observational model II (OM II). They are based on the model used in YP15. They use the same GNFW profile (given by equation 2.21) to model the gas content, but with the slope parameters stated in Section 2.4; they take into account only the cluster gas – they do not explicitly model the dark matter component. They work in *angular* rather than physical sizes. Like the PM, they also use equation 2.18 to calculate Y . However, the calculation steps are different. We start by evaluating equation 2.18 in the limit that $r \rightarrow \infty$. It can be shown that for the GNFW pressure profile this gives (see Appendix B.1 for a derivation of this result)

$$\lim_{r \rightarrow \infty} Y_{\text{sph}}(r) \equiv Y_{\text{tot}} = \lim_{r \rightarrow \infty} \frac{\sigma_T}{m_e c^2} \int_0^r P_e(r') 4\pi r'^2 dr' = \frac{4\pi P_{\text{ei}} D_A \theta_p^3 \sigma_T}{m_e c^2} \frac{\Gamma\left(\frac{3-c}{a}\right) \Gamma\left(\frac{b-3}{a}\right)}{a \Gamma\left(\frac{b-c}{a}\right)} \quad (4.1)$$

where $\Gamma(x)$ is the Gamma function and $\theta_p = r_p/D_A$. Note that for finite r (and thus θ)

$$Y_{\text{sph}}(\theta) = \frac{4\pi P_{\text{ei}} D_A \theta_p^2 \sigma_T}{m_e c^2} \int_0^\theta \left(\frac{\theta'}{\theta_p}\right)^{2-c} \left(1 + \left(\frac{\theta'}{\theta_p}\right)^a\right)^{(c-b)/a} d\theta'. \quad (4.2)$$

Both equations have a common (unknown) factor $D_A P_{\text{ei}}$. Hence for given (i.e. input) values of Y_{tot} and θ_p , equation 4.1 can be solved for $D_A P_{\text{ei}}$ and then equation 4.2 can be solved for finite θ numerically. Furthermore the OMs assume that the cluster is spherically symmetric and that the cluster gas can be described by the equation of state of an ideal gas. The OMs have four cluster input parameters: Y_{tot} , θ_p , x_c and y_c . They differ only in the prior distributions they use.

4.2.1 Observational model I prior

The priors used on Y_{tot} and θ_p are the same as the ‘new’ priors used in YP15. These were derived from the Planck completeness simulations (Planck Collaboration XXIX 2014) as follows. The simulations were produced by drawing a cluster population from the Tinker mass function (Tinker et al. 2008) and using the scaling relations in Planck Collaboration XI (2011) to obtain observable quantities. This cluster population was injected into the real Planck data and a simulated union catalogue was created by running the Planck detection pipelines on this simulated dataset. An elliptical Gaussian function was then fitted to the posterior of Y_{tot} and θ_p in log space. Hence the prior has the Planck selection function implicitly included in it.

Parameter	Prior distribution
x_c	$\mathcal{N}(0'', 60'')$
y_c	$\mathcal{N}(0'', 60'')$
$\log(Y_{\text{tot}}), \log(\theta_p)$	$\mathcal{N}((-2.7, 0.62), (0.29, 0.12), 40.2^\circ)$

Table 4.1: Observational model I input parameter prior distributions. Note that the Gaussian elliptical function on $\log(Y_{\text{tot}}) - \log(\theta_p)$ is parameterised in terms of the mean in both dimensions, the respective standard deviations and the offset of the principle axes from the vertical and horizontal axes measured clockwise.

Parameter	Prior distribution
x_c	$\mathcal{N}(0'', 60'')$
y_c	$\mathcal{N}(0'', 60'')$
θ_p	$\mathcal{U}[\log(\theta_p, \min(z)), \log(\theta_p, \max(z))]$
Y_{tot}	$\mathcal{U}[\log(Y_{\text{tot}, \min(z)}), \log(Y_{\text{tot}, \max(z)})]$

Table 4.2: Observational model II input parameter prior distributions.

For consistency, the same cluster centre priors were used in both observational models as in the PM. The priors for OM I are summarised in Table 4.1.

4.2.2 Observational model II

The priors on Y_{tot} and θ_p in OM II incorporate the spectroscopic or photometric redshift of each cluster. From the z and $M(r_{200})$ priors of the PM and for $f_{\text{gas}}(r_{200}) = 0.13$, upper and lower bounds on Y_{tot} and θ_p are calculated using the PM. Note that Y_{tot} and θ_p are assumed to be a-priori *uncorrelated*, unlike in OM I. For the lowest redshift cluster ($z = 0.0894$), these limits are $\theta_{p, \min} = 4.24$ arcmin, $\theta_{p, \max} = 19.04$ arcmin, $Y_{\text{tot}, \min} = 1.06 \times 10^{-4}$ arcmin 2 and $Y_{\text{tot}, \max} = 0.19$ arcmin 2 ; for the highest redshift ($z = 0.83$) cluster these limits are $\theta_{p, \min} = 0.67$ arcmin, $\theta_{p, \max} = 3.01$ arcmin, $Y_{\text{tot}, \min} = 5.7 \times 10^{-6}$ arcmin 2 and $Y_{\text{tot}, \max} = 0.01$ arcmin 2 . It clear that z has a large effect on the PM calculations, as it is used to calculate the angular scale from r through $\theta = r/D_A(z)$ where $D_A(z)$ is the angular diameter distance of the cluster at redshift z , and to convert the units of Y . It is also used to calculate c_{200} which affects the scale of the self-similar dark matter density profile, and the normalisation constant ρ_s in equation 2.19 is proportional to $\rho_{\text{crit}}(z)$. The priors for OM II are summarised in Table 4.2. Note that in using the PM calculations to calculate the prior limits, we have made the assumptions underlying the PM that OM I is not subject to (i.e. hydrostatic equilibrium up to radius r_{200} and f_{gas} is much less than unity up to the same radius).

4.3 AMI model comparisons

I now use AMI data to compare the PM, OM I and OM II, and begin by comparing their observational parameter estimates. Secondly I introduce a metric which measures the ‘distance’ between probability distributions. In this context the distance is measured between the $(Y(r_{500}), \theta_{500})$ posterior distributions of the three models. Finally the models are compared using the evidence ratios introduced in Section 2.6.2. The results obtained from these analyses are given in Appendix B, which lists the values obtained for the 54 cluster sample in ascending order of z .

I emphasise the notation used for Y . For consistency I parameterise Y by r for all three models ($Y \equiv Y(r)$). For the PM, $Y(r)$ has units [length²]; to convert this to the more conventional [angle²] we divide by D_A^2 : $Y(r) \rightarrow Y(r)/D_A^2$ as mentioned in Section 2.2. The Y value given by an OM is naturally in units of [angle²]; when I refer to $Y(r)$ in the context of the OMs I equivalently mean $Y(\theta)$.

4.3.1 Physical and observational models Y values comparison

Figure 4.1 shows the posterior mean values for $Y(r_{500})$ for the three models used on the same AMI datasets. I first note that the errors associated with the OM estimates are generally larger than those with the PM. Secondly it appears that the OM I Y are less strongly correlated with z than those from the PM and OM II. This may be because OM I contains no explicit z -information, and in fact its only reliance on z is from the simulated and empirical datasets used to fit its prior distribution, but the same prior is used for all clusters, and so the dependence on redshift is very weak.

I now compare the results from the three models pairwise. Note that when we refer to the dispersion between values in units of standard deviations, we are referring to the combined standard deviation of the two Y values. When comparing PM and OM I values of Y , just 15 clusters are within one standard deviation, 27 within two and 18 are more than three standard deviations away from each other. The same comparison between PM and OM II gives corresponding values of 23, 40 and 5. This implies that the dispersion between OM II and PM is much smaller (especially in the extreme cases), and shows the importance in the choice of priors. Table 4.3 gives a summary of the dispersion of the PM with respect to the OMs. Figure 4.2 shows the fractional difference between the Y values for the three models, and shows that the PM estimates are generally much higher than both OM values at low z . However, in general the PM yields lower Y estimates compared to the OMs (PM underestimates Y relative to OM I and OM II 35

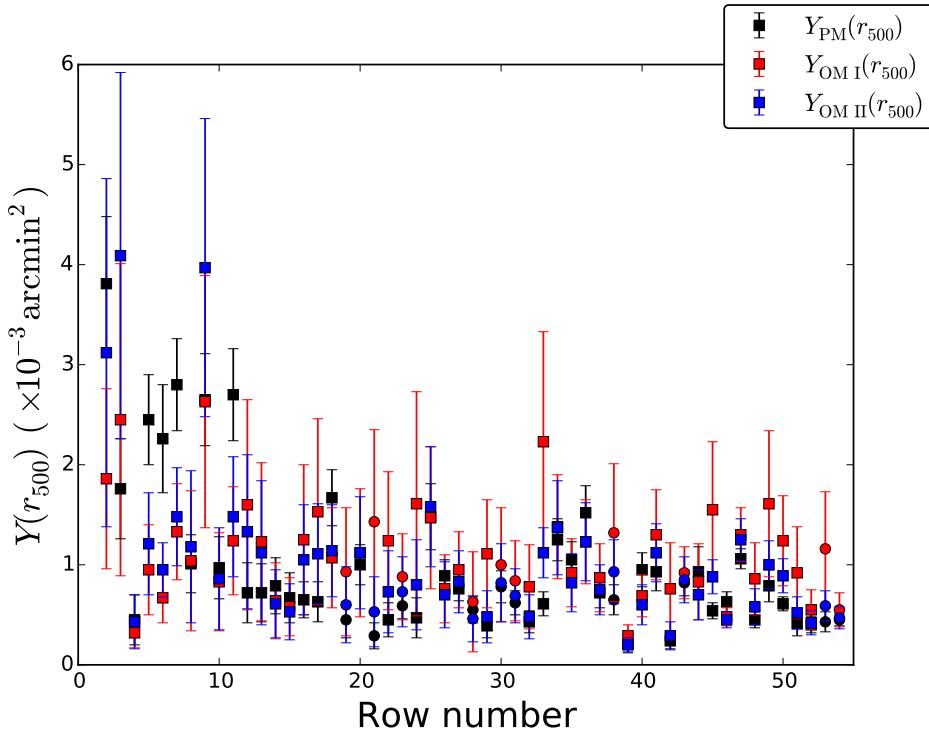


Figure 4.1: Plot of $Y(r_{500})$ obtained from AMI data using the physical and observational models vs row number of Table B.1. The points with circular markers correspond to clusters whose redshifts were measured photometrically as opposed to spectroscopically. For clarity purposes the first row is not plotted due to its relatively large value ($Y(r_{500}) \approx 10 \text{ arcmin}^2$).

Model comparison ($Y_{\mathcal{M}_i} \equiv$)	$ Y_{\text{PM}} - Y_{\mathcal{M}_i} / \sigma_{\text{PM} \& \mathcal{M}_i} < 1$	$ Y_{\text{PM}} - Y_{\mathcal{M}_i} / \sigma_{\text{PM} \& \mathcal{M}_i} < 2$	$ Y_{\text{PM}} - Y_{\mathcal{M}_i} / \sigma_{\text{PM} \& \mathcal{M}_i} > 3$
$Y_{\text{OM I}}$	15	27	18
$Y_{\text{OM II}}$	23	40	5

Table 4.3: Difference between physical model mean values for $Y(r_{500})$ & observational model mean values, measured in units of the physical model $Y(r_{500})$ standard deviation. The numbers in the columns correspond to the number of clusters out of the sample of 54 which satisfy the criterion specified in the respective header.

and 36 times respectively).

Looking at the dispersion between OM I and OM II, 36 clusters are within one standard deviation, four within two and just four are more than three standard deviations away from each other. This implies that OM II seems to be in reasonable agreement with the two other models (usually in between the values from the other models).

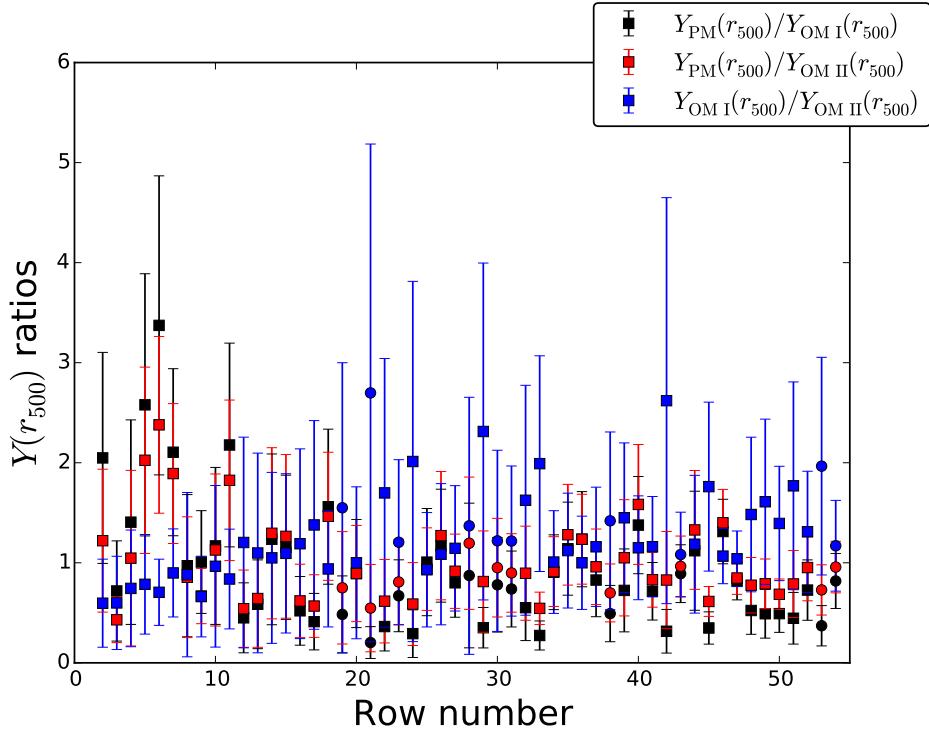


Figure 4.2: Plot of $Y(r_{500})$ ratio vs row number of Table B.1 for three different cases: $Y_{\text{PM}}(r_{500})/Y_{\text{OM I}}(r_{500})$; $Y_{\text{PM}}(r_{500})/Y_{\text{OM II}}(r_{500})$ and $Y_{\text{OM I}}(r_{500})/Y_{\text{OM II}}(r_{500})$. The points with square markers correspond to clusters whose redshifts were measured spectroscopically, and the circular markers photometrically (as listed in Table B.1).

4.3.2 Earth Mover's distance

The Earth Mover's distance (EMD), first introduced in [Rubner, Tomasi, & Guibas \(1998\)](#) is a "distance" function defined between two distributions. In the case where these distributions integrate over all space to the same value (e.g. they are probability distributions), the EMD is given in terms of the first Wasserstein distance ([Levina & Bickel 2001](#)). A common analogy used to describe the EMD is the following: if the probability distributions are interpreted as two different ways of piling up a certain amount of earth, and the amount of earth at position \mathbf{x}_i and \mathbf{x}_j belonging to each probability distribution at those points are $P_1(\mathbf{x}_i)$ and $P_2(\mathbf{x}_j)$, then the EMD is the minimum cost of moving one pile into the other, where the cost of moving each "spadeful" is taken to be the mass of each spadeful (f_{ij}) \times the distance by which it is moved ($|\mathbf{x}_i - \mathbf{x}_j|$). For discrete two-dimensional probability distributions P_1 & P_2 , with two-dimensional

domains \mathbf{x}_i & \mathbf{y}_j , then the EMD between these probability distributions $d_{\text{EMD}}(P_1, P_2)$ is defined to be the minimum value of

$$W(P_1, P_2) = \sum_{i=1}^m \sum_{j=1}^n f_{ij} |\mathbf{x}_i - \mathbf{y}_j| \quad (4.3)$$

with respect to distance and f_{ij} . Here m and n are the number of values in the domains of P_1 and P_2 respectively and f_{ij} are the ‘flow’ of probability density from $P_1(\mathbf{x}_i)$ to $P_2(\mathbf{y}_j)$. Different implementations of the algorithm use different distance measures, but we use the Euclidean distance in equation 4.3. The f_{ij} are subject to the following constraints

$$f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n; \quad (4.4)$$

$$\sum_{j=1}^n f_{ij} = P_1(\mathbf{x}_i), 1 \leq i \leq m; \quad (4.5)$$

$$\sum_{i=1}^m f_{ij} = P_2(\mathbf{y}_j), 1 \leq j \leq n; \quad (4.6)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \sum_{i=1}^m P_1(\mathbf{x}_i) = \sum_{j=1}^n P_2(\mathbf{y}_j) = 1. \quad (4.7)$$

For a more detailed account of the EMD see [Levina & Bickel \(2001\)](#).

4.3.3 Application of EMD

The EMD metric is applied to the different pairs of models using Gary Doran’s wrapper^a for Yossi Rubner’s algorithm ([Rubner, Tomasi, & Guibas 1998](#)). Before running the algorithm the $(Y(r_{500}), \theta_{500})$ posteriors are normalised so that the metric is not skewed towards θ_{500} (the use of Euclidean distances in the EMD algorithm, are obviously misrepresentative if the dimensions are not normalised). Each dimension is normalised to the range $[0, 1]$ by performing the following transformations

$$\theta_{500} \rightarrow \frac{\theta_{500} - \theta_{500, \min}}{\theta_{500, \max} - \theta_{500, \min}}; Y(r_{500}) \rightarrow \frac{Y(r_{500}) - Y_{\min}(r_{500})}{Y_{\max}(r_{500}) - Y_{\min}(r_{500})}. \quad (4.8)$$

The values for $\theta_{500, \min}$, $\theta_{500, \max}$, $Y_{\min}(r_{500})$ and $Y_{\max}(r_{500})$ are deduced by considering all of the values of $Y(r_{500})$ and θ_p from the posteriors obtained from the three models at once, to ensure that all posterior values are normalised by the same factor. The larger the value of the

^a<https://github.com/garydoranjr/pyemd>.

Statistic	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM I}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM II}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{OM I}}, \mathcal{P}_{\text{OM II}})$	union
mean	0.093	0.067	0.057	0.072
standard deviation	0.057	0.050	0.077	0.064
median	0.076	0.051	0.027	0.051
min	0.020	0.013	0.006	0.006
max	0.225	0.297	0.514	0.514

Table 4.4: Summary of EMD values calculated between the $Y(r_{500}) - \theta_{500}$ posterior distributions from all three model pairs, and their union.

EMD, the ‘further away’ the distributions are from each other. The EMD was calculated for each cluster with each pair of models (giving $3 \times 54 = 162$ distances in total). The full set of EMD values calculated can be found in Table B.2 in Appendix B. Table 4.4 provides a summary of $d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM I}})$, $d_{\text{EMD}}(\mathcal{P}_{\text{OM I}}, \mathcal{P}_{\text{OM II}})$, $d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM II}})$, and the union of the three. Concerning both mean and median, the posteriors are most discrepant between the PM and OM I, followed by PM and OM II. However it is interesting to note that the two largest EMD values come from $d_{\text{EMD}}(\mathcal{P}_{\text{OM II}}, \mathcal{P}_{\text{OM I}})$ and $d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM II}})$ cases, with values 0.514 and 0.297 respectively. Furthermore these are from the same cluster, which is at the lowest z ($= 0.0894$). This suggests that incorporating z information into an observational model for very low redshift clusters has a significant effect. Ignoring the lowest redshift cluster (or by looking at the median value, which is skewed less by outliers), it is clear that of the three models, OM I and OM II posteriors are most in agreement with each other. Figure 4.3 shows the $Y(r_{500}), \theta_{500}$ posterior distributions created using GETDIST (with the 95% and 68% confidence intervals plotted), for the highest and lowest EMD values obtained from the 162 values calculated. Both of these come from OM II – OM I comparisons.

Figure 4.4 shows $d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM II}})$ vs z from which it is apparent that there is a negative correlation between d_{EMD} and z .

4.3.4 Physical and observational models comparison

As described in Section 2.6.2, one can perform a model comparison, by comparing the Bayesian evidence values calculated when the models were applied to the same (AMI) datasets. We can also define the detection ratio of a model as the ratio of the evidences of the ‘data’ and ‘null-data’ runs. The first of these corresponds to modelling the cluster, background and detectable radio-sources. The null-data run models everything but the cluster. The ratio of these evidences therefore gives a measure of the significance that the cluster has in modelling the data. Note that

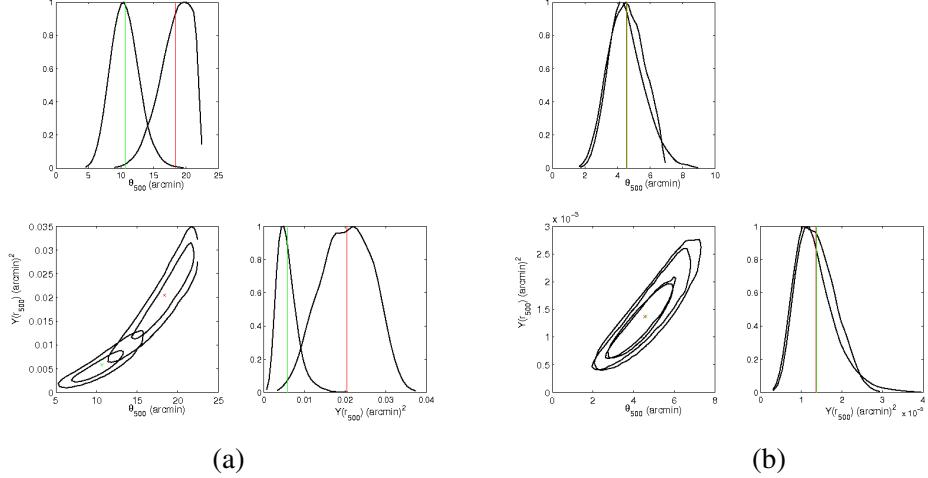


Figure 4.3: (a) Highest d_{EMD} value $Y(r_{500}) - \theta_{500}$ posteriors for cluster PSZ2G044.20+48.66 at $z = 0.0894$. (b) Lowest d_{EMD} value $Y(r_{500}) - \theta_{500}$ posteriors for cluster PSZ2G132.47-17.27 at $z = 0.341$. For both triangle plots, the top graph shows the marginalised θ_{500} posteriors for OM II and OM I. The bottom right graph shows the marginalised $Y(r_{500})$ posteriors. The bottom left graph shows the two-dimensional $Y(r_{500}) - \theta_{500}$ posteriors from which the EMD is calculated. The contours represent the 95% and 68% confidence intervals. Note that the parameters in the plots are not normalised, but the ones in the distance calculations are normalised by transforming the parameters as discussed in the text. For all of the plots, the green crosses / lines are the mean values of the OM I posteriors (the smaller values in (a)) and the red crosses / lines are the mean values of the OM II posteriors (the larger values in (a)). For Figure (b), the mean values for $Y(r_{500})$ are so close together that the lines cannot be distinguished.

the null-data run is the same for all three models considered here, as they only differ in the way they model the galaxy cluster itself. Table B.2 in Appendix B gives the log of a detection ratio, $\ln(\mathcal{Z}_i/\mathcal{Z}_{\text{null}})$ for each of the three models, and the ratios between the different pairs of models, $\ln(\mathcal{Z}_i/\mathcal{Z}_j)$ where \mathcal{Z}_i and \mathcal{Z}_j are one of \mathcal{Z}_{PM} , \mathcal{Z}_{OMI} or $\mathcal{Z}_{\text{OMII}}$, for each cluster.

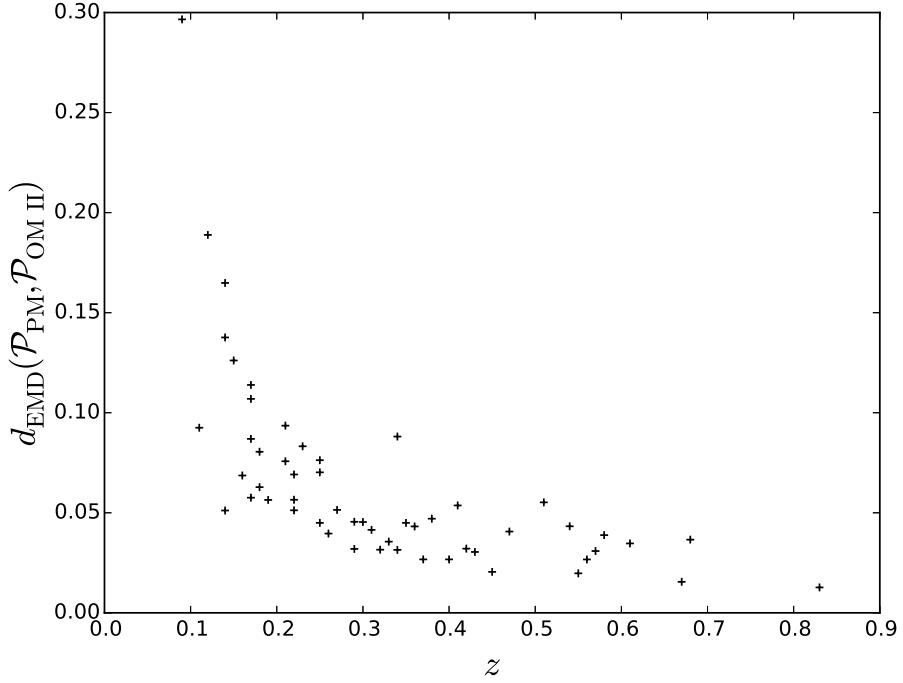


Figure 4.4: Earth Mover's distance calculated between $Y(r_{500}) - \theta_{500}$ posteriors for PM and OM II, versus z for the 54 clusters. The crosses indicate the point– they are not error bars.

4.3.4.1 Physical model and observational model I

The data favour OM I over the PM for 50 of the 54 clusters. Though in 36 of the 50 cases $\log(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}})$ is between minus one and zero, which according to the Jeffreys scale means "more data are needed to come to a meaningful conclusion". (see Table 2.1). A further 12 of these had $\log(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}})$ values between -2.5 and -1 which can be interpreted as "weak preference" in favour of OM I, whilst no clusters had a value of $\log(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}})$ less than minus five ("strong preference" in favour of OM I). The largest absolute value for the ratio was actually in favour of the PM with $\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}}) = 4.73 \pm 0.23$ (for the lowest z cluster) which suggests "moderate preference" towards the PM. There is no correlation between $\log(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}})$ and z . Figure 4.5 shows the prior space for the observational parameters corresponding to the PM with the lowest and highest z values in the sample.

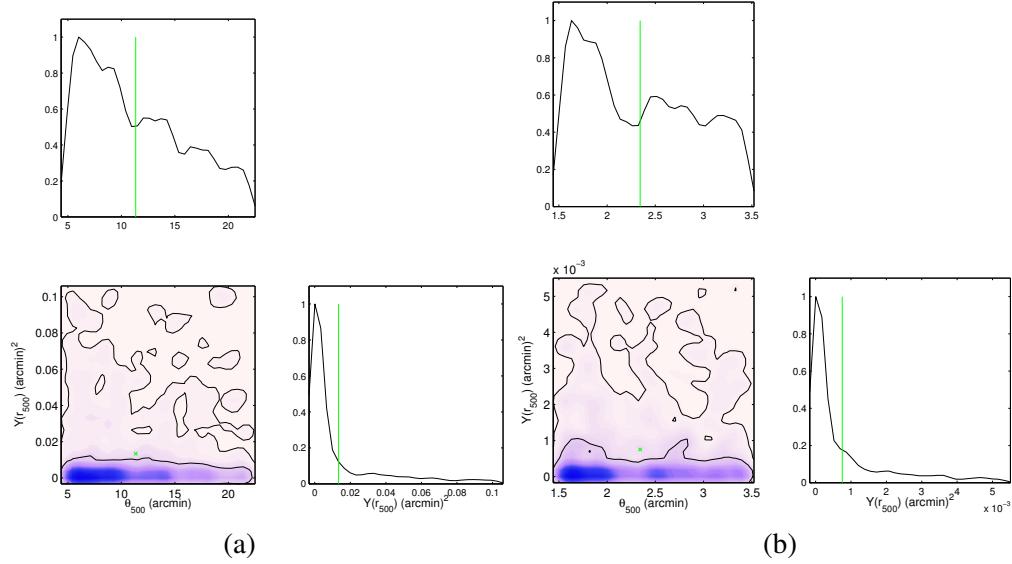


Figure 4.5: (a) Lowest z ($= 0.0894$) prior parameter space for $Y(r_{500}) - \theta_{500}$ using the PM. (b) Highest z ($= 0.83$) prior parameter space for $Y(r_{500}) - \theta_{500}$ for the PM and OM II. Note the scales on the axes are different for each plot, and the green vertical lines represent the mean values.

4.3.4.2 Observational models I & II

Similarly, OM I is favoured over OM II for 53 clusters, but with 14 cases having $0 \leq \log(\mathcal{Z}_{\text{OMI}}/\mathcal{Z}_{\text{OMII}}) \leq 1$. Again the highest absolute value came from the lowest redshift cluster, highlighting the importance of z information at such a low z value. Since these models have the same input parameters, it is easier to compare their sampling parameter spaces. Figure 4.6 shows the prior range of $(Y(r_{500}), \theta_{500})$ for OM I. Around 68% of the prior mass (i.e. the inner contour in the Figure) is bounded roughly by $Y(r_{500}) = 2 \times 10^{-3}$ arcmin 2 and $\theta_{500} = 10$ arcmin. The 95% contour gives upper bounds of $Y(r_{500}) \approx 4 \times 10^{-3}$ arcmin 2 and $\theta_{500} \approx 15$ arcmin. In comparison the OM II prior ranges for the lowest redshift cluster are $\theta_{500} = [4.9, 19.0]$ arcmin and $Y(r_{500}) = [0.006, 1.0] \times 10^{-1}$ arcmin 2 , and for the highest redshift cluster are $\theta_{500} = [0.8, 3.5]$ arcmin, $Y(r_{500}) = [0.003, 5.0] \times 10^{-3}$ arcmin 2 . The ratio of the upper and lower limits for θ and Y are approximately 4.5 and 1.8×10^3 across all clusters. This suggests that the ratio of the bounds of the parameter space for each cluster does not change for the OM II, but that the sampling space is shifted depending on z . Note that even though the sampling parameters for the observational models are Y_{tot} and θ_p , these are related to $Y(r_{500})$ and θ_{500} by constant factors, and so

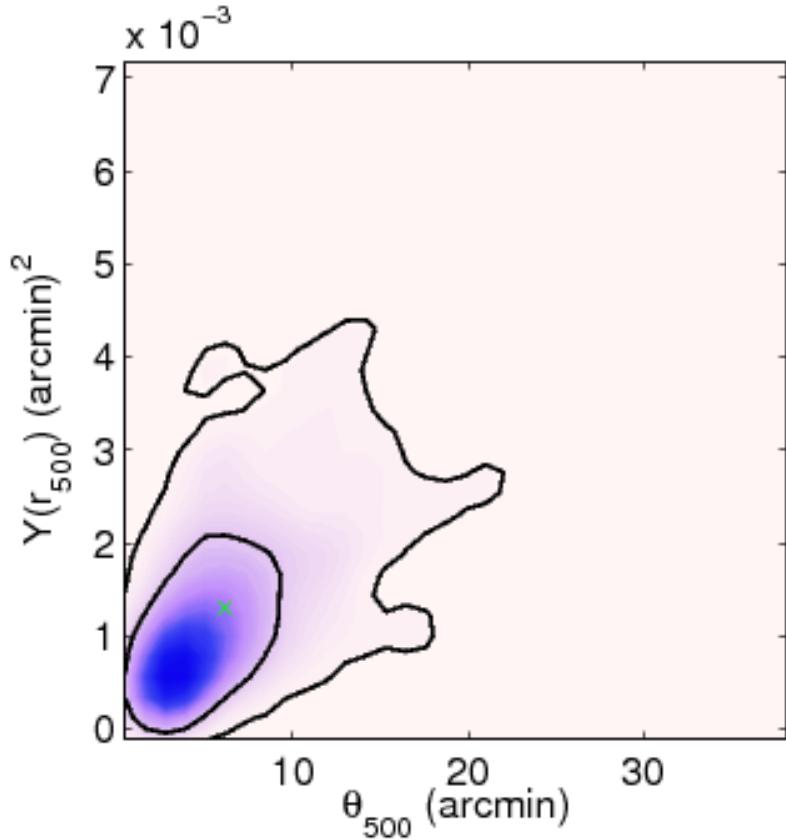


Figure 4.6: Two-dimensional prior probability distribution of $Y(r_{500})$ and θ_{500} for OM I, which is based on Planck data as detailed in Section 4.2.1.

comparisons made on both are equivalent.

4.3.4.3 Physical model and observational model II

Comparison of PM and OM II, the models which incorporate redshift information into their priors leads to interesting results. For 43 clusters, the PM is preferred over OM II. However for all of these clusters $\log(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMII}})$ is less than one, meaning that none of them give "conclusive" model preference. There are only three clusters which give "weak evidence" in favour of a model (OM II). These are the clusters at redshift $z = 0.144, 0.341, 0.5131$ with ratio values

-1.88 , -1.06 , -1.16 respectively. The fact that data from 51 clusters do not provide any "conclusive" preference between PM and OM II, suggests that these models are equally well suited for the current data, even though their parameter estimates are often not in such agreement.

4.4 Conclusions

For the cluster sample analysed in the previous Chapter, I compare the parameter estimates obtained from different physical and observational models applied to AMI data using Bayesian analysis. The physical model (PM) used is as described in Section 2.4, and the observational models (OM I and OM II) are based on the one described in Perrott et al. (2015). I have focused on comparisons of $Y(r_{500})$ and found the following.

- The PM generally yields lower estimates of Y relative to the observational models, apart from at low z where the reverse is true.
- For two thirds of the sample, the OM I and OM II estimates are within one combined standard deviation of each other.

To investigate further the discrepancies between the three models, we computed the Earth Mover's distance between the two-dimensional posterior distributions in $Y(r_{500})$, θ_{500} space, for each model pair. This gives a measure of the 'distance' between the respective probability distributions. I then compared the evidence values obtained from the Bayesian analysis of the AMI data using the different models, referring to the Jeffreys scale to form conclusions on model preference, and found the following.

- Based on the Earth Mover's distances calculated for each cluster, the posteriors are most discrepant between the PM and OM I models when the sample was considered as a whole, followed by PM and OM II.
- The two largest discrepancies come from the lowest- z cluster, one between PM & OM I and one between OM II & OM I, suggesting that z information at very low z can have a large effect on the different models.
- The distance between posteriors from PM and OM II clearly decreases with increasing z . This suggests that the difference between physical and observational model parameter estimates, provided the latter also includes z information, is reduced at higher z .
- When comparing Bayesian evidence values, OM I is preferred over PM for 50 of the clusters, although only 14 of these showed either "weak" or "moderate" preference to OM I (the remaining 36 being "inconclusive"); however the highest log(evidence ratio) actually favours the PM ("moderate" preference) and occurs for the lowest- z cluster.

- Similarly, OM I is preferred to OM II in 53 of the cases. 14 suggested more data are needed to come to a "meaningful" conclusion, while the remaining 39 clusters showed "weak" or "moderate" preference for OM I. This suggests that OM I is the preferred model in more cases relative to OM II than when OM I is compared with PM.
- For 43 of the clusters, PM is preferred over OM II; however in all of these cases, the Jeffreys scale suggests "no conclusion can be made without more data", and only three clusters give any "conclusive" preference (a "weak" preference in favour for OM II).

CHAPTER 5

PHYSICAL MODELLING OF GALAXY CLUSTERS USING EINASTO DARK MATTER PROFILES

This Chapter provides an alternative to the physical model presented in Section 2.4. The physical model described previously uses an NFW profile (Navarro, Frenk, & White 1995) for the dark matter component of the galaxy cluster, which is derived from N-body simulations of galaxy clusters. Einasto (1965) gives an empirical profile for dark matter halos. Previous investigations comparing the two dark matter profiles using simulated data (see e.g. Dutton & Macciò 2014, Meneghetti et al. 2014, Klypin et al. 2016 and Sereno, Fedeli, & Moscardini 2016) have shown that the Einasto model provides a better fit. In particular, Sereno, Fedeli, & Moscardini (2016) showed for weak lensing analysis of clusters that the NFW profile can overestimate virial masses of very massive halos ($\geq 10^{15} M_{\text{Sun}}/h$ where M_{Sun} is units of solar mass and h is the reduced Hubble constant) by up to 10%.

It is these previous analyses which have motivated us to derive a physical galaxy cluster model for interferometric SZ data which uses the Einasto profile to model the dark matter component of the cluster. I also compare the parameter estimates and fits of the NFW & Einasto models for the cluster A611 with data obtained with AMI, and with simulations created with both Einasto and NFW profiles. The work discussed in this Chapter is currently being published in MNRAS (Javid et al. 2018). Note the paper includes post-referee changes.

5.1 Einasto physical model

The physical model presented here (PM II) follows the same calculational steps as the model presented in Section 2.4 (PM I) to calculate $\delta I_{\text{cl},\nu}$, but with an Einasto profile replacing the NFW one used for the dark matter component. Below we derive the relevant equations for the Einasto case. Furthermore PM II is subject to the same assumptions as PM I listed in Section 2.4.

The three input parameters required to calculate $\delta I_{\nu,\text{cl}}$ for either PM are $M(r_{200})$, $f_{\text{gas}}(r_{200})$, and z . A fourth input parameter is required for the PM II which we call the Einasto parameter α_{Ein} , which is also described below.

5.1.0.4 Dark matter profile

Assuming an Einasto profile (Einasto 1965), the dark matter density profile for a cluster $\rho_{\text{dm,PM II}}$ is given by

$$\rho_{\text{dm,PM II}} = \rho_{-2} \exp \left[-\frac{2}{\alpha_{\text{Ein}}} \left(\left(\frac{r}{r_{-2}} \right)^{\alpha_{\text{Ein}}} - 1 \right) \right], \quad (5.1)$$

where α_{Ein} is a shape parameter, r_{-2} is the scale radius where the logarithmic derivative of the density is -2 (analogue to r_s in the NFW model, but note that in general $r_{-2} \neq r_s$), and ρ_{-2} is the density at this radius. The parameter α_{Ein} controls the degree of curvature of the profile. The larger its value, the more rapidly the slope varies with respect to r . In the limit that $\alpha_{\text{Ein}} \rightarrow 0$, the logarithmic derivative is -2 for all r . It is tempting to assume that the Einasto profile is capable of providing a better fit due to the fact that the Einasto profile has an extra degree of freedom (three for the Einasto profile, two for the NFW), the shape parameter. However Klypin et al. (2016) claims that this is not strictly true, as the Einasto profile was seen to give a better fit to simulated dark matter haloes even with α_{Ein} fixed. The asymptotic values of the logarithmic slope for the two profiles are as follows: as $r \rightarrow 0$ then $d \ln \rho_{\text{dm,PM I}}(r) / d \ln r \rightarrow -1$ and $d \ln \rho_{\text{dm,PM II}}(r) / d \ln r \rightarrow 0$. As $r \rightarrow \infty$ then $d \ln \rho_{\text{dm,PM I}}(r) / d \ln r \rightarrow -3$ and $d \ln \rho_{\text{dm,PM II}}(r) / d \ln r \rightarrow -\infty$. The magnitude of α_{Ein} determines how quickly the slope changes between the two asymptotic values. Throughout this work when I refer to the NFW or Einasto model, I really mean the physical model which uses the NFW or Einasto model when considering the dark matter density profile.

Referring back to equation 5.1, the ratio r_{200}/r_{-2} is defined as the concentration parameter c_{200} . Dutton & Macciò (2014) determines an analytical form for c_{200} as a function of total mass and redshift for Einasto profiles based on simulations similar to those described in Macciò et al.

(2007) and Macciò, Dutton, & van den Bosch (2008)

$$\log_{10}(c_{200}) = j(z) + k(z) \log_{10} \left[\frac{M(r_{200})}{10^{12} h^{-1} M_{\text{Sun}}} \right], \quad (5.2)$$

where $j(z) = 0.459 + 0.518 \exp(-0.49z^{1.303})$ and $k(z) = -0.13 + 0.029z$. The fitting is said to be accurate in the redshift range [0, 5]. To calculate ρ_{-2} we must make the assumption stated for PM I, that the total mass enclosed at r_{200} is approximately equal to the enclosed dark matter mass. That is

$$M(r_{200}) = M_{\text{dm}}(r_{200}) + M_{\text{g}}(r_{200}) \approx M_{\text{dm}}(r_{200}), \quad (5.3)$$

where $M_{\text{dm}}(r_{200})$ and $M_{\text{g}}(r_{200})$ are the dark matter and gas masses. With this assumption we can say that for any $r \leq r_{200}$

$$\begin{aligned} M(r) &\approx \int_0^r 4\pi r'^2 \rho_{\text{dm, PM II}}(r') dr' \\ &= \frac{4\pi \rho_{-2} r_{-2}^3}{\alpha_{\text{Ein}}} \exp(2/\alpha_{\text{Ein}}) \left(\frac{\alpha_{\text{Ein}}}{2} \right)^{3/\alpha_{\text{Ein}}} \\ &\quad \times \gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right], \end{aligned} \quad (5.4)$$

where $\gamma[a, x] = \int_0^x t^{a-1} e^{-t} dt$ is the incomplete lower gamma function. The steps taken to get this result are given in Appendix C.1. Equation 2.22 can be evaluated at r_{200} and equated with equation 5.4 evaluated at the same radius to obtain the following solution for ρ_{-2}

$$\begin{aligned} \rho_{-2} &= \frac{200}{3} \left(\frac{r_{200}}{r_{-2}} \right)^3 \rho_{\text{crit}}(z) \times \frac{1}{\left[1/\alpha_{\text{Ein}} \exp(2/\alpha_{\text{Ein}}) \left(\frac{\alpha_{\text{Ein}}}{2} \right)^{3/\alpha_{\text{Ein}}} \right]} \\ &\quad \times \frac{1}{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r_{200}}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]}. \end{aligned} \quad (5.5)$$

Equivalently, equation 5.4 can be evaluated at r_{200} and set equal to the known value of $M(r_{200})$ to determine ρ_{-2} . Figure 5.1 shows the logarithmic dark matter density profiles as a function of r for a cluster at $z = 0.15$ with $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$ and $f_{\text{gas}}(r_{200}) = 0.12$ for PM I and PM II for the α_{Ein} values: 0.05, 0.2, 2.0. It is clear that the Einasto profiles diverge the most from each other at low r and for the high α_{Ein} value at high r as well.

5.1.0.5 Gas density and pressure profiles

Calculating the pressure normalisation constant (defined below) again requires the assumption that the cluster is in hydrostatic equilibrium up to radius r_{200} . This means at any radius equal to

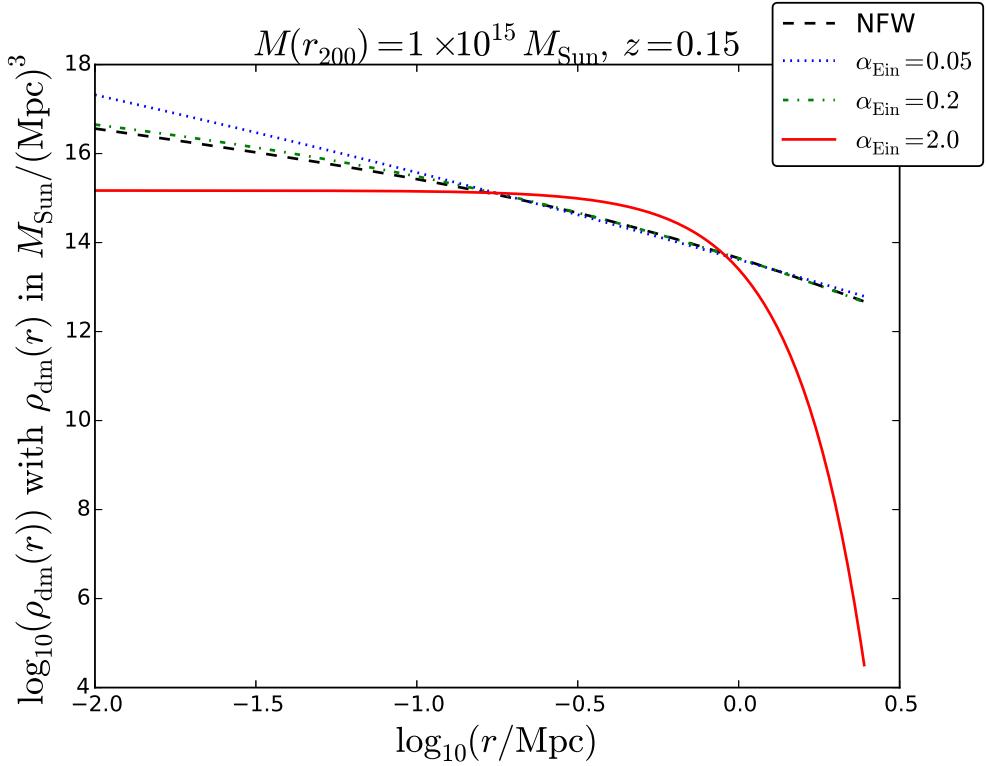


Figure 5.1: Logarithmic dark matter density profiles as a function of log cluster radius using NFW and Einasto models. Three values of the Einasto profile are used: 0.05, 0.2, and 2.0. The additional input parameters used to generate these profiles are: $z = 0.15$, $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$ and $f_{\text{gas}}(r_{200}) = 0.12$.

or below r_{200} the outward pushing pressure force created by the pressure differential at that point must be equal to the gravitational binding force due to the mass enclosed within that radius, i.e. that equation 2.30 holds. Furthermore I follow Nagai, Kravtsov, & Vikhlinin (2007) and assume the GNFW model given by equation 2.21 for the pressure profile, as in PM I. However, for all analysis presented in this Chapter (both PMs), the GNFW slope parameters are taken to be $a = 1.0510$, $b = 5.4905$ and $c = 0.3081$. These ‘universal’ values were taken from Arnaud et al. (2010) and are the best fit GNFW slope parameters derived from the REXCESS sub-sample (observed with XMM-Newton, Böhringer et al. 2007), as described in Section 5 of Arnaud et al.. I also take the Arnaud et al. value of c_{500} which is 1.177. Note that in the previous Chapters (as well as in MO12) slightly different values derived for the standard self-similar case (Appendix B of Arnaud et al.) were used ($a = 1.0620$, $b = 5.4807$, $c = 0.3292$ and $c_{500} = 1.156$). It was

shown in [Olamaie, Hobson, & Grainge \(2013\)](#) that PM I is not affected by which of these two sets of parameters is used.

The analytical function used to convert from r_{200} to r_{500} in PM I is specific to the NFW dark matter profile case and so is not applicable to PM II. I have not found an analytic fitting function for the conversion in the case of an Einasto dark matter profile and so I obtain r_{500} iteratively as described in [Appendix C.2](#). As in PM I, the pressure profile can be substituted into the hydrostatic equilibrium equation to derive an expression for the gas density. Using equation [5.4](#) for $M(r)$ gives

$$\begin{aligned} \rho_g(r) = & \frac{\mu_e}{\mu_g} \frac{P_{ei}}{4\pi G \rho_{-2} r_{-2}^3} \frac{1}{\left[(1/\alpha_{Ein}) \exp(2/\alpha_{Ein}) (\alpha_{Ein}/2)^{3/\alpha_{Ein}} \right]} \\ & \times \frac{r}{\gamma \left[\frac{3}{\alpha_{Ein}}, \frac{2}{\alpha_{Ein}} \left(\frac{r_{200}}{r_{-2}} \right)^{\alpha_{Ein}} \right]} \\ & \times \left(\frac{r}{r_p} \right)^{-c} \left[1 + \left(\frac{r}{r_p} \right)^a \right]^{-\left(\frac{a+b-c}{a} \right)} \left[b \left(\frac{r}{r_p} \right)^a + c \right]. \end{aligned} \quad (5.6)$$

Note that like PM I, the gas mass $M_g(r)$ given by

$$M_g(r) = \int_0^r 4\pi \rho_g(r') r'^2 dr' \quad (5.7)$$

must be integrated numerically. Hence $f_{\text{gas}}(r) = M_g(r)/M(r)$ does not have a closed form solution. Nevertheless, we can use equations [5.6](#) and [5.7](#) to determine P_{ei} since we know $M(r_{200})$, $f_{\text{gas}}(r_{200})$ and r_{200} . Evaluating equations [5.6](#) and [5.7](#) at r_{200} and solving for P_{ei} gives the following expression

$$\begin{aligned} P_{ei} = & \left(\frac{\mu_g}{\mu_e} \right) (G \rho_{-2} r_{-2}^3) \left[\frac{\exp(2/\alpha_{Ein})}{\alpha_{Ein}} (\alpha_{Ein}/2)^{3/\alpha_{Ein}} \right] M_g(r_{200}) \\ & \times \frac{1}{\int_0^{r_{200}} r'^3 \frac{\left[b \left(\frac{r'}{r_p} \right)^a + c \right]}{\gamma \left[\frac{3}{\alpha_{Ein}}, \frac{2}{\alpha_{Ein}} \left(\frac{r'}{r_{-2}} \right)^{\alpha_{Ein}} \right] \left(\frac{r'}{r_p} \right)^c \left[1 + \left(\frac{r'}{r_p} \right)^a \right]^{\left(\frac{a+b-c}{a} \right)}} dr'}, \end{aligned} \quad (5.8)$$

which must be evaluated numerically. Once P_{ei} and r_p have been calculated, the Comptonisation parameter and therefore $\delta I_{\nu, \text{cl}}$ can be calculated the same way as in PM I.

5.1.0.6 Additional cluster parameters

As stated in Section 2.4, the radial profile of the electron number density is given by $n_e(r) = \rho_g(r)/\mu_e$. Using the ideal gas assumption, the electron temperature is therefore given by

$$T_e(r) = \left(\frac{4\pi\mu_g G \rho_{-2} r_{-2}^3}{k_B} \right) \left[(1/\alpha_{\text{Ein}}) \exp(2/\alpha_{\text{Ein}}) (\alpha_{\text{Ein}}/2)^{3/\alpha_{\text{Ein}}} \right] \times \frac{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]}{r} \times \left[1 + \left(\frac{r}{r_p} \right)^a \right] \left[b \left(\frac{r}{r_p} \right)^a + c \right]^{-1} \quad (5.9)$$

which also equals $T_g(r)$.

The gas mass can be determined numerically from equation 5.7,

$$M_g(r) = \left(\frac{\mu_e}{\mu_g} \right) \frac{1}{G} \frac{P_{\text{ei}}}{\rho_{-2}} \frac{1}{\left[(1/\alpha_{\text{Ein}}) \exp(2/\alpha_{\text{Ein}}) (\alpha_{\text{Ein}}/2)^{3/\alpha_{\text{Ein}}} \right] r_{-2}^3} \times \int_0^r r'^3 \frac{\left[b \left(\frac{r'}{r_p} \right)^a + c \right]}{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r'}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]} \times \left(\frac{r'}{r_p} \right)^c \left[1 + \left(\frac{r'}{r_p} \right)^a \right]^{\left(\frac{a+b-c}{a} \right)} dr'. \quad (5.10)$$

5.1.0.7 Prior probability distributions

For both PM I and PM II I adopt the following approach (excluding any mention of α_{Ein} in the former case).

As in Section 3.2, the cluster parameters are assumed to be independent of one another, so that

$$\pi(\boldsymbol{\Theta}_{\text{cl}}) = \pi(\alpha_{\text{Ein}}) \pi(M(r_{200})) \pi(f_{\text{gas}}(r_{200})) \pi(z) \pi(x_c) \pi(y_c). \quad (5.11)$$

Table 5.1 lists the type of prior used for each cluster parameter and the probability distribution parameters. The values used for z and α_{Ein} will be specified on a case by case basis in Section 5.2.2.

Parameter	Prior distribution
x_c	$\mathcal{N}(0'', 60'')$
y_c	$\mathcal{N}(0'', 60'')$
z	$\delta(z)$
$M(r_{200})$	$\mathcal{U}[\log(0.5 \times 10^{14} M_{\text{Sun}}), \log(50 \times 10^{14} M_{\text{Sun}})]$
$f_{\text{gas}}(r_{200})$	$\mathcal{N}(0.12, 0.02)$
α_{Ein}	$\delta(\alpha_{\text{Ein}})$

Table 5.1: Cluster parameter prior distributions, where the normal distributions are parameterised by their mean and standard deviations.

5.2 Results

5.2.1 Cluster parameter profiles

I first present the results of using the Einasto model in the profiling of cluster dark matter for a range of different cluster input parameters, along with the equivalent results from PM I.

I consider two input masses, $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ and $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$, which roughly span the range of galaxy cluster masses. I use z -values of 0.15 and 0.9, take $f_{\text{gas}}(r_{200}) = 0.12$ following [Komatsu et al. \(2011\)](#), and consider α_{Ein} values of 0.05, 0.2, and 2.0 – see Figure 5.1. I note that the same r range ($-2 \leq \log_{10}(r) \leq 0.5$ where r is in units of Mpc) is considered for each cluster, and thus even though each parameter profile is self-similar in r with respect to mass and redshift, they are different for each cluster over the range of r considered here.

5.2.1.1 Dark matter mass profiles

Figure 5.2 shows the dark matter mass profiles. The Einasto profiles are calculated using equation 5.4 and the NFW profile from the equivalent relation given by equation 2.19. Note that even though the notation in these equations corresponds to the total mass, this is in fact just the dark matter mass as we have used the approximation $M(r) \approx M_{\text{dm}}(r)$ in deriving them. The $\alpha_{\text{Ein}} = 2$ case always converges quickly as the density rapidly falls to zero, while the other three profiles including the NFW show divergent behaviour at the largest radii considered here. The high mass inputs result in similar profiles for the $\alpha_{\text{Ein}} = 0.05$, $\alpha_{\text{Ein}} = 0.2$ and NFW cases, whereas the low mass inputs result in the $\alpha_{\text{Ein}} = 0.05$ case diverging somewhat more rapidly than the others.

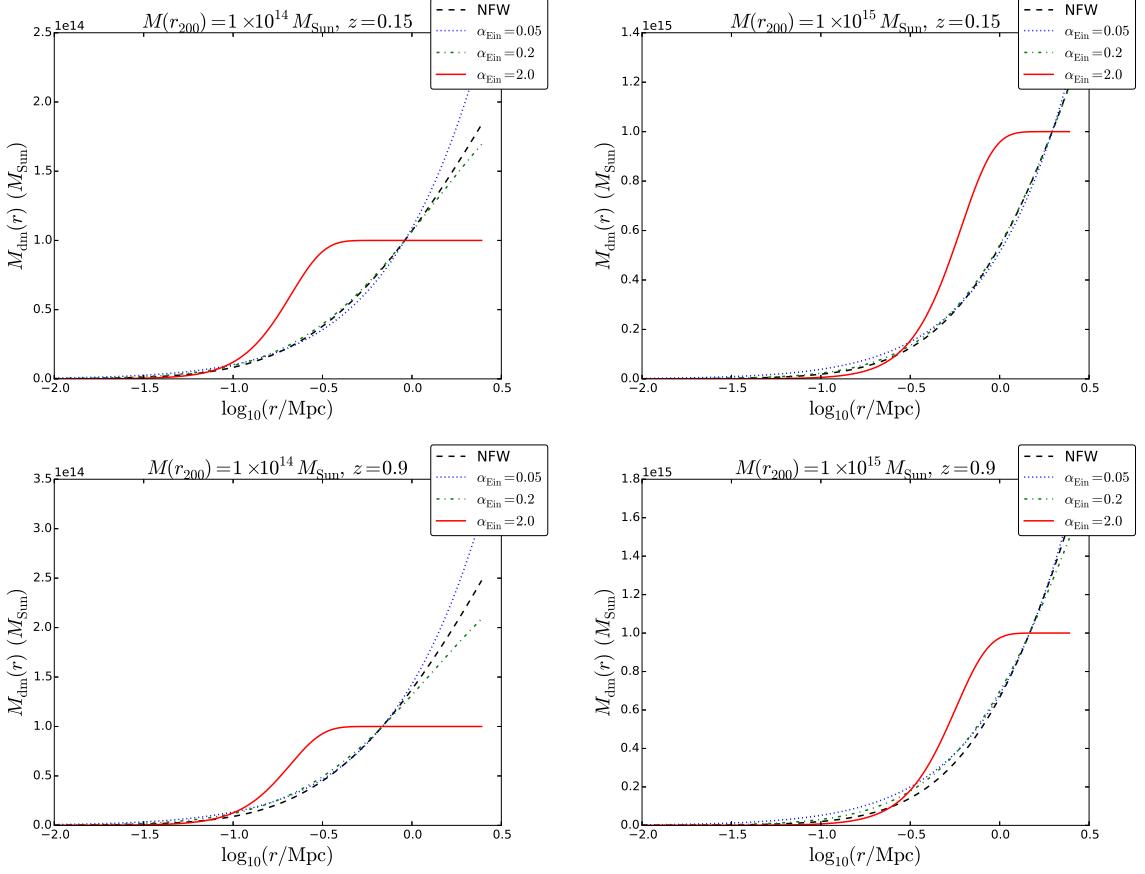


Figure 5.2: Dark matter mass profiles as a function of log cluster radius using NFW and Einasto models. Values of $\alpha_{\text{Ein}} = 0.05, 0.2$, and 2.0 are used as inputs. Top row has $z = 0.15$, bottom row has $z = 0.9$. Left column has $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$, right column has $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$.

5.2.1.2 Gas density profiles

Figure 5.3 shows the gas density profiles. The Einasto profiles are calculated using equation 5.6 and the NFW profile from the equivalent relation given in MO12 (equation 6). Note that when calculating $\rho_g(r)$ for arbitrary r , we are assuming hydrostatic equilibrium at that radius so that equation 2.30 holds, and we have to assume that $f_{\text{gas}}(r') \approx 0$ for all $r' \leq r$ so that $M(r) \approx M_{\text{dm}}(r)$ at this radius. The plots show that the profiles are similar for all inputs of mass and redshift, with the $\alpha_{\text{Ein}} = 0.2$ Einasto profile again most resembling the NFW profile. However, the $\alpha_{\text{Ein}} = 2.0$ profile has the highest gas density at high r for both masses and both z values.

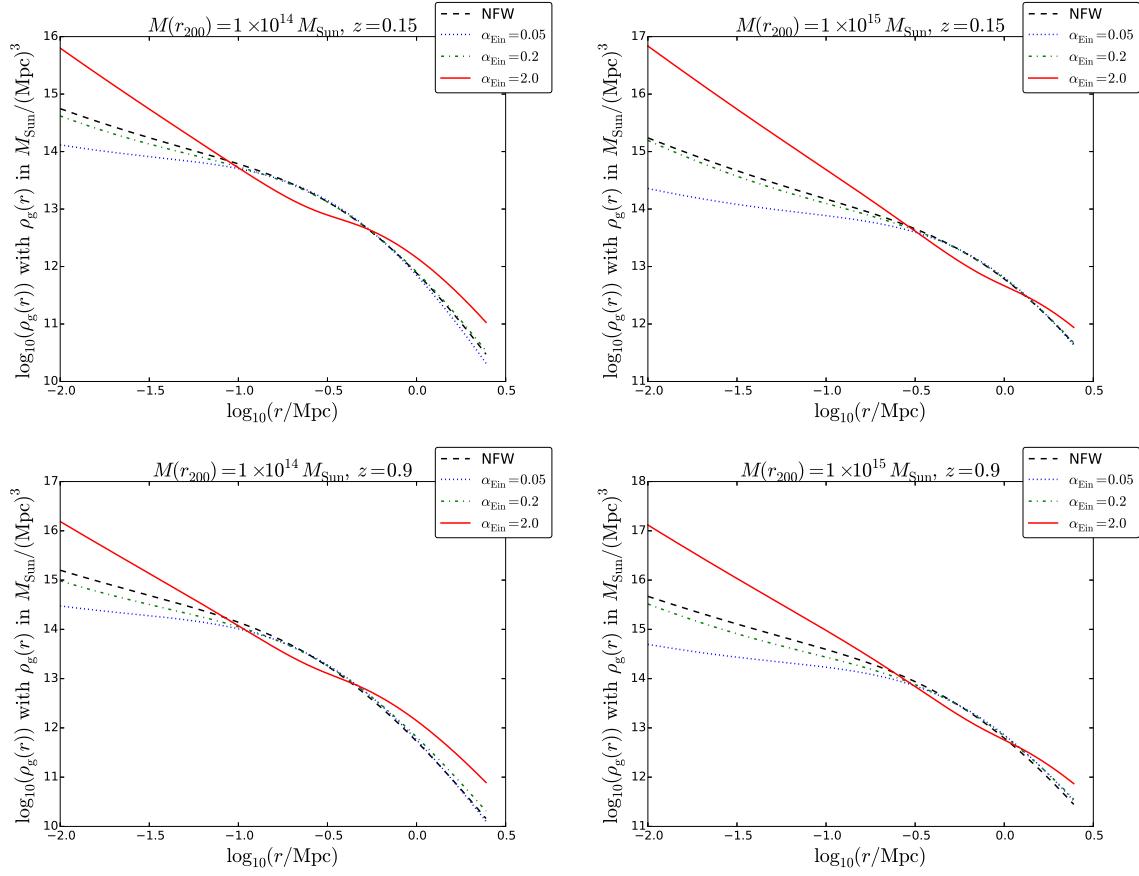


Figure 5.3: Logarithmic gas density profiles as a function of log cluster radius using NFW and Einasto models. Values of $\alpha_{\text{Ein}} = 0.05, 0.2$, and 2.0 are used as inputs. Top row has $z = 0.15$, bottom row has $z = 0.9$. Left column has $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$, right column has $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$.

5.2.1.3 Gas mass profiles

Figure 5.4 shows $M_g(r)$ as a function of cluster radius. As in Figure 5.2 with the dark matter mass profiles, the high mass inputs correspond to divergent behaviour at large r . But for $\alpha_{\text{Ein}} = 2.0$ the profile of $M_g(r)$ also shows a more noticeable such divergence. Furthermore, in all four input parameter cases, $\alpha_{\text{Ein}} = 2.0$ shows more divergent behaviour than other values of α_{Ein} and the NFW profile in gas mass, which is in contrast to the dark matter mass profiles.

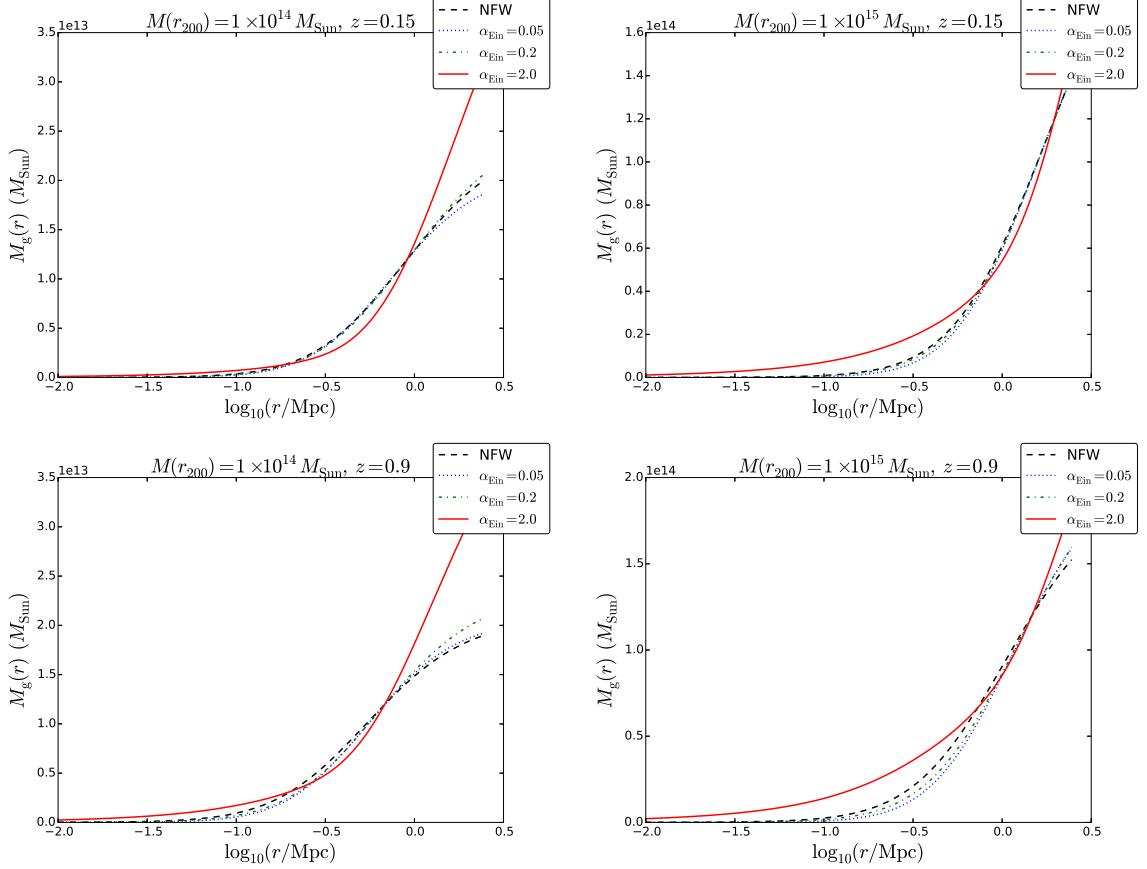


Figure 5.4: Gas mass profiles as a function of log cluster radius using NFW and Einasto models. Values of $\alpha_{\text{Ein}} = 0.05, 0.2$, and 2.0 are used as inputs. Top row has $z = 0.15$, bottom row has $z = 0.9$. Left column has $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$, right column has $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$.

5.2.1.4 Gas temperature profiles

Gas temperature profiles are shown in Figure 5.5. The $\alpha_{\text{Ein}} = 2.0$ is very distinctive, always peaking at much higher r than the other three and also always much more sharply.

5.2.2 Bayesian analysis of AMI data

I now focus on applying the PM II to real and simulated AMI data, to compare the parameter estimates and Bayesian evidences with those obtained from the PM I.

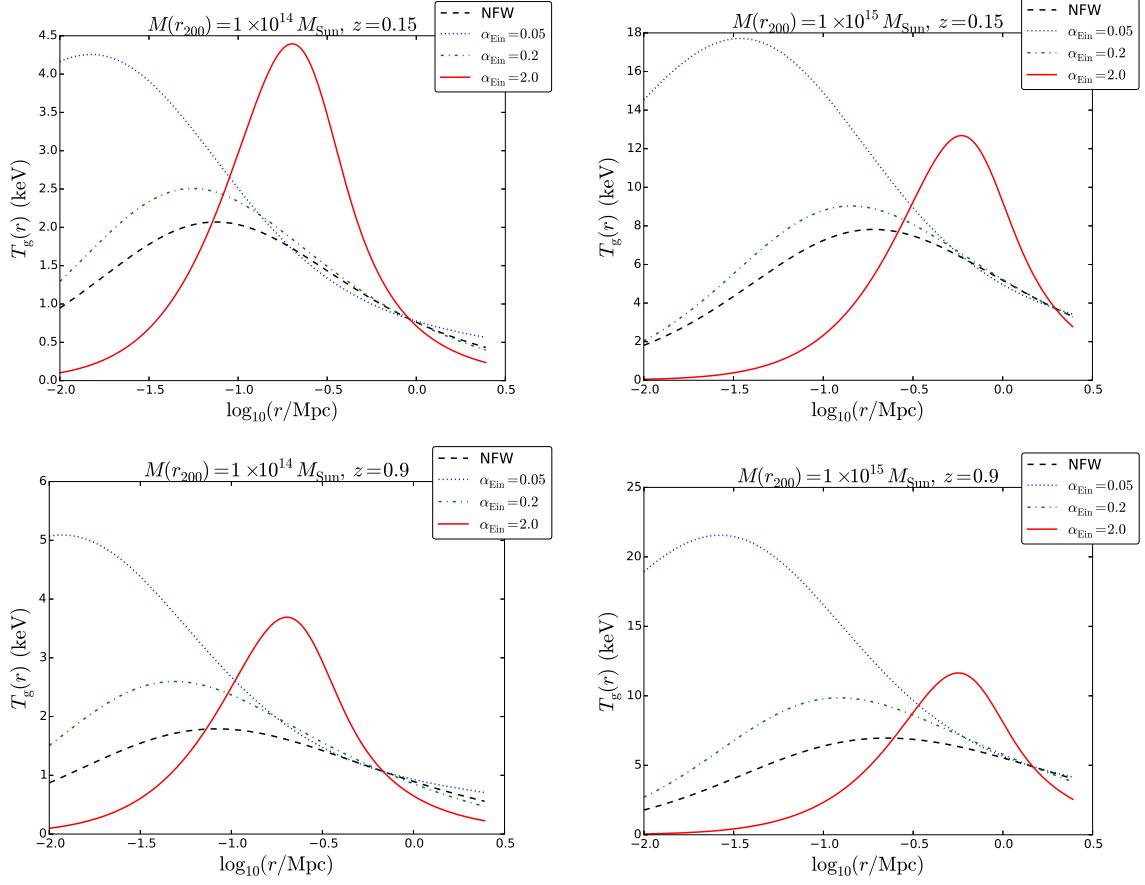


Figure 5.5: Gas temperature profiles as a function of log cluster radius using NFW and Einasto models. Values of $\alpha_{\text{Ein}} = 0.05, 0.2$, and 2.0 are used as inputs. Top row has $z = 0.15$, bottom row has $z = 0.9$. Left column has $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$, right column has $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$.

5.2.2.1 Analysis of real AMI observations of A611

I conduct Bayesian analysis on data from observations with AMI of the cluster A611 at $z = 0.288$, which has been studied through its X-ray emission, strong lensing, weak lensing and SZ effect (see [Schmidt & Allen 2007](#), [Donnarumma et al. 2011](#), [Romano et al. 2010](#) and [Rumsey et al. 2016](#) respectively). These studies suggest that there is no significant contamination from radio-sources and that the cluster is close to the T_X – T_{SZ} relation for clusters in hydrostatic equilibrium.

I first compare the posterior distributions for the input parameters (except those with δ -function

Model	x_c (arcsec)	y_c (arcsec)	$M(r_{200}) (\times 10^{14} M_{\text{Sun}})$	$f_{\text{gas}}(r_{200})$	$\ln(\mathcal{Z})$
NFW	24.7 ± 12.4	13.9 ± 11.5	7.84 ± 1.24	0.129 ± 0.020	$3.862944 \times 10^4 \pm 0.25$
$\alpha_{\text{Ein}} = 0.05$	22.7 ± 12.5	13.1 ± 12.6	7.45 ± 1.24	0.130 ± 0.019	$3.862921 \times 10^4 \pm 0.25$
$\alpha_{\text{Ein}} = 0.2$	25.5 ± 12.8	14.9 ± 13.0	7.67 ± 1.27	0.127 ± 0.017	$3.862967 \times 10^4 \pm 0.24$
$\alpha_{\text{Ein}} = 2.0$	24.3 ± 12.4	14.3 ± 13.2	6.17 ± 1.12	0.130 ± 0.017	$3.862924 \times 10^4 \pm 0.24$

Table 5.2: Marginalised posterior distribution mean values and standard deviations of physical model input parameters and Bayesian evidences associated with each model, applied to real A611 data.

priors). The means and standard deviations of the four analyses are given in Table 5.2. As in Section 5.2.1, $\alpha_{\text{Ein}} = 0.05$ and $\alpha_{\text{Ein}} = 0.2$ show similar results to PM I. $\alpha_{\text{Ein}} = 2$ gives a different estimate for $M(r_{200})$, and its posterior distribution is shown in Figure 5.6 along with that obtained with the NFW profile. These posterior distributions are plotted using GETDIST and the contours on the two-dimensional plots represent the 95% and 68% confidence intervals. The mean mass estimates are within one combined standard deviation away from each other. However, as seen in Table 5.2 the value of $\ln(\mathcal{Z}_{\text{Ein}}/\mathcal{Z}_{\text{NFW}})$ imply that ‘no model is favoured by the data’ according to the Jeffreys scale.

5.2.2.2 Simulated AMI data

[Sereno, Fedeli, & Moscardini \(2016\)](#) study the errors associated with fitting NFW profiles to Einasto dark matter halos and vice versa for weak lensing studies. I conduct similar work in the context of simulated SZ observations. The simulations were carried out using the in-house AMI simulation package PROFILE, which has been used in various forms in e.g. [Grainge et al. \(2002\)](#) and [Olamaie, Hobson, & Grainge \(2013\)](#).

As before I consider Einasto profiles with the α_{Ein} values 0.05, 0.2, and 2.0 plus an NFW profile, each with $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ or $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$, $z = 0.15$ or $z = 0.90$ and $f_{\text{gas}}(r_{200}) = 0.12$. These 16 simulations, were analysed as in Section 5.2.2.1. Note for all of these simulations no radio-sources, primordial CMB or confusion noise were included, and instrumental noise was set to a negligible level. Table C.1 in Appendix C.3 summarises the input and output values of the 16 simulations. The first column gives the model used to simulate the cluster, with the following two columns giving the mass and z input values. For each simulation, I analysed the data using two models, one using the NFW profile and one using an Einasto profile. For data simulated using an NFW profile, when analysing the data with an

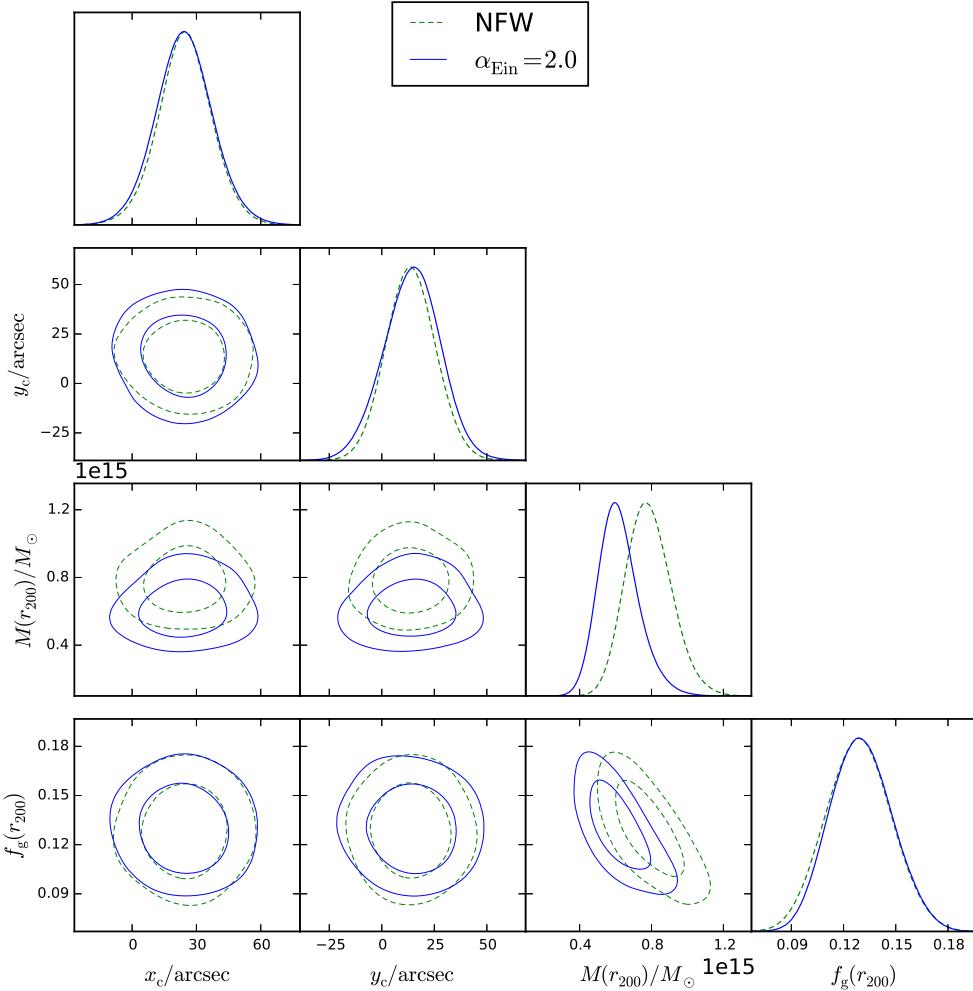


Figure 5.6: Marginalised posterior distributions of physical model input parameters for the NFW and $\alpha_{\text{Ein}} = 2.0$ models applied to real A611 data. The contour plots are the two dimensional marginalised plots of the parameters named in the corresponding row / column. The line plots are the fully marginalised posterior distributions.

Einasto profile I used $\alpha_{\text{Ein}} = 0.2$. For data simulated using an Einasto profile, when analysing the data with an Einasto profile I set α_{Ein} equal to the value used as the input for the simulation.

In all but one of the simulations (NFW simulated with $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ and $z = 0.9$), the Einasto posterior mean mass value was closer to the input value than the corresponding NFW value. It's worth noting that a more thorough statistical treatment would involve repeating the Bayesian analyses many times to see if these results held consistently, but this was not

considered here. In 11 out of 16 cases the Einasto profile recovers the input mass to within 10% (interestingly, it does so for all the NFW simulated clusters). However, in only two of 16 cases does the Einasto model recover the input value within three standard deviations. This could be due to errors associated with the simulated ‘observing’ of the cluster on a pixelated grid, binning the data in u-v space/ frequency and then modelling the data by creating another pixelated grid. These effects are not accounted for in the Bayesian analysis, thus leading to an underestimate in the associated errors. Furthermore, the fact that the Einasto model recovers the NFW simulated clusters better than when those simulations are analysed with the NFW profile for three of the four NFW simulated clusters, could be down to the fact that the Einasto model is more robust to the imperfections associated with the generation of the simulations. Another source of error underestimation could be the sampling errors being underestimated in the nested sampling algorithm as studied in [Higson et al. \(2017\)](#). Looking at the individual evidence values for both Einasto and NFW models, the value is considerably lower for the high mass simulations, $\ln(\mathcal{Z}_{\text{low mass}}/\mathcal{Z}_{\text{high mass}}) \approx 3000$ suggests the models fit the low mass datasets much better when averaged over the (same) parameter sampling spaces. It is crucial to note that when comparing evidences calculated from different datasets (specifically their ratio), we are not looking at $Pr(\mathcal{M}|\mathcal{D}_1)/Pr(\mathcal{M}|\mathcal{D}_2)$, since the $Pr(\mathcal{D})$ -like terms on the right hand side of equation 2.44 do not cancel in this case. Nevertheless for the same model, the evidence ratio between two different datasets does give a measure of the relative goodness of fit of the datasets to the model. Looking at the evidence ratios between the Einasto and NFW models for a given simulation, more data is needed to come to a conclusive decision over model preference in 10 of the simulations. Three simulations lead to ‘substantial preference’ in favour of the Einasto model ($\ln(\mathcal{Z}_{\text{Ein}}/\mathcal{Z}_{\text{NFW}}) \geq 5$). In two of these cases ($\alpha_{\text{Ein}} = 0.2$ with $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$ and $z = 0.9$, and $\alpha_{\text{Ein}} = 2.0$ with $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$ and $z = 0.9$) the posteriors show reasonable constraints in both the Einasto and NFW analyses (Figure 5.7 shows posterior distributions for $\alpha_{\text{Ein}} = 2.0$ with $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$ and $z = 0.9$), with the former giving better estimates of mass and $f_{\text{gas}}(r_{200})$. The third case however ($\alpha_{\text{Ein}} = 2$ simulated with $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ and $z = 0.9$) leads to low estimates of $f_{\text{gas}}(r_{200})$ in both cases (Figure 5.8), and a very high mass estimate in the case of the NFW model. The two cases where the NFW model is preferred over the Einasto also produce posteriors similar to those in Figure 5.8.

Finally, I tried running the Bayesian analysis on eight of the Einasto simulated clusters with uniform analysis priors on α_{Ein} . These clusters corresponded to the simulations with input values of either $\alpha_{\text{Ein}} = 0.2$ or $\alpha_{\text{Ein}} = 2.0$. For the former value of α_{Ein} I assigned the uniform

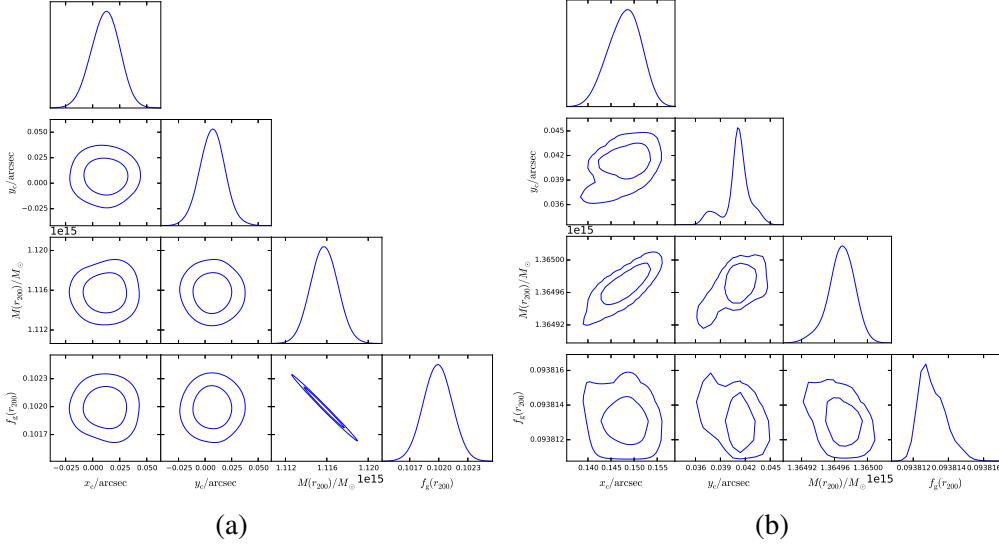


Figure 5.7: Posterior distributions for cluster simulated with $\alpha_{\text{Ein}} = 2.0$, $M(r_{200}) = 1 \times 10^{15} M_\odot$ and $z = 0.9$, modelled with: (a) Einasto dark matter profile, and (b) NFW dark matter profile.

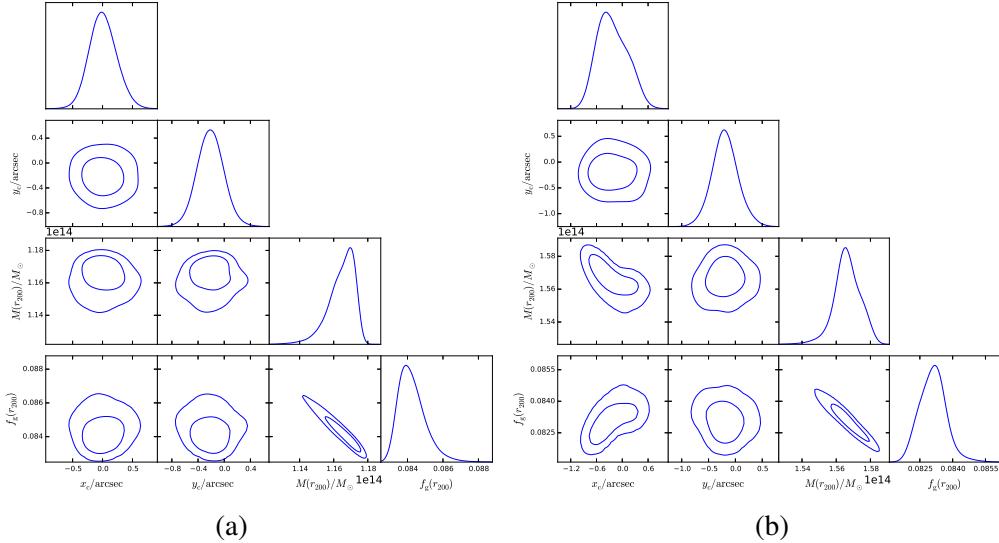


Figure 5.8: Posterior distributions for cluster simulated with $\alpha_{\text{Ein}} = 2.0$, $M(r_{200}) = 1 \times 10^{14} M_\odot$ and $z = 0.9$, modelled with: (a) Einasto dark matter profile, and (b) NFW dark matter profile.

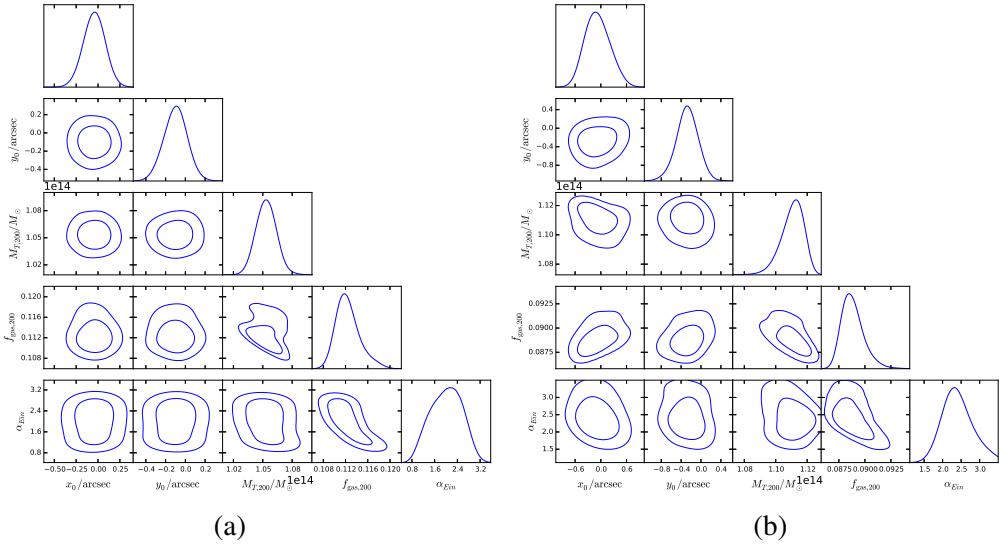


Figure 5.9: Posterior distributions of Einasto model input parameters for: (a) $\alpha_{\text{Ein}} = 2.0$, $M(r_{200}) = 1 \times 10^{14} M_{\odot}$ and $z = 0.15$ simulated cluster, and (b) $\alpha_{\text{Ein}} = 2.0$, $M(r_{200}) = 1 \times 10^{14} M_{\odot}$ and $z = 0.9$ simulated cluster.

prior $\mathcal{U}[0.05, 0.35]$ and $\mathcal{U}[0.5, 3.5]$ for the latter. For two of these simulations the posterior distributions did not show much degeneracy between any of the input parameters, including α_{Ein} . Both of these clusters had $\alpha_{\text{Ein}} = 2$, $M(r_{200}) = 1 \times 10^{14} M_{\odot}$ and $z = 0.15$ or $z = 0.9$ as inputs. Their posterior distributions are shown in Figure 5.9. Both posteriors give a mean value for the shape parameter within one standard deviation of the input value (2.01 ± 0.54 and 2.39 ± 0.40), but looking at the distributions they are not sharply peaked, meaning the errors on the estimates are quite large. Nevertheless these simulations do show the Einasto profile is capable of recovering some information about α_{Ein} , in contrast to the efforts in MO12 to recover c_{200} which led to large $c_{200} - M(r_{200})$ degeneracies (although c_{200} relates to the scale of the dark matter profile, not its shape).

5.3 Conclusions

Based on the physical model introduced in Section 2.4 (PM I) which uses an NFW profile (Navarro, Frenk, & White 1995) to model the dark matter content of galaxy clusters, I derive a new physical model (PM II) which models the dark matter with an Einasto profile (Einasto 1965). The Einasto profile has an additional degree of freedom compared to the NFW profile, which

dictates the shape of the dark matter density as a function of radius. For different values of α_{Ein} we have investigated the profiles of several physical properties of a cluster, namely the dark matter density, dark matter mass, gas density, gas mass and gas temperature. I have also provided the equivalent profiles in the NFW case. From this I found the following.

- Of the three values of α_{Ein} considered, $\alpha_{\text{Ein}} = 0.2$ gave the most similar profile to that given by the NFW model (as discussed in [Dutton & Macciò 2014](#)), with the main discrepancy between the two arising in the peak amplitude of the gas temperature.
- $\alpha_{\text{Ein}} = 2.0$ showed the most convergent behaviour in $M_{\text{dm}}(r)$ at high r , but the most divergent in $M_{\text{g}}(r)$ in the same limit.
- The gas temperature profiles were somewhat different for the α_{Ein} values considered here. This suggests that if one can carefully measure the temperature profile of a cluster, then one could infer α_{Ein} and use this in the model presented here (though one has to be aware of cooling flow and merger activity).

Next we applied Bayesian analysis to real and simulated AMI datasets using PM I and PM II, to compare the models' parameter estimates and fits to the data. Using real data from cluster A611 I found the following.

- The $\alpha_{\text{Ein}} = 0.05$ and $\alpha_{\text{Ein}} = 0.2$ models gave very similar results to the NFW model; the $\alpha_{\text{Ein}} = 2$ model however underestimates $M(r_{200})$ relative to the other three models.
- The Bayesian evidence values calculated from these four analyses were roughly equal, suggesting no model provided a statistically significant fit relative to the others.

Simulating clusters with either NFW or Einasto dark matter profiles, which were then 'observed' by AMI, I found the following.

- For 15 out of 16 clusters, the Einasto model recovered the input mass better than the NFW model. The only cluster where this was not the case (NFW simulated with $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ and $z = 0.9$), the posterior distributions do not show good constraints on the sampling parameters, and so the parameter estimates should not be used.
- The evidence values of both Einasto and NFW models are considerably lower for the high mass simulations.
- Considering the evidence ratios between the Einasto and NFW models for a given simulation, more data is needed to come to a conclusive decision over model preference in 10 of the cases. However according to the Jeffreys scale ([Jeffreys 1961](#)), three of the simulations gave 'substantial' preference towards the Einasto model; and in two of these cases the NFW analysis did not constrain the sampling parameters as well as the Einasto analysis. In the third case neither analysis constrained the parameters well.

- The two clusters where the evidence ratio was in favour of the NFW model also showed poor posterior distribution constraints.
- When allowing α_{Ein} to vary in the analysis, in two out of eight of the Einasto simulations used the posterior distributions showed some constraints on the value of α_{Ein} which gave estimates close to the input values.

CHAPTER 6

ENHANCED PHYSICAL MODELLING I: RELAXING THE f_{gas} ASSUMPTION

As stated in Section 2.4, one of the key assumptions of the physical model (for both PM I and PM II) is that the gas mass fraction $f_{\text{gas}}(r)$ is much smaller than unity up to r_{200} , so that we can say the total mass at r_{200} is equal to dark matter mass enclosed up to this radius. In this Chapter we relax this assumption for both models, so that the total mass is the sum of the dark matter and gas contributions. We refer to these new models as PMT I and PMT II which respectively use NFW and Einasto profiles to model the dark matter content.

6.1 Total mass equations

Dropping the assumption that $f_{\text{gas}}(r) \ll 1$ we can no longer assume that $M(r) \approx \int_0^r 4\pi\rho_{\text{dm}}(r')r'^2 \, dr'$, but instead

$$M(r) = \int_0^r 4\pi\rho_{\text{dm}}(r')r'^2 \, dr' + \int_0^r 4\pi\rho_{\text{g}}(r')r'^2 \, dr'. \quad (6.1)$$

Using the hydrostatic equilibrium assumption given by equation 2.30 to substitute for $\rho_{\text{g}}(r')$, we get the following integral equation

$$M(r) = \int_0^r 4\pi\rho_{\text{dm}}(r')r'^2 \, dr' - \frac{4\pi}{G} \int_0^r \frac{dP_{\text{g}}(r')}{dr'} \frac{r'^4}{M(r')} \, dr'. \quad (6.2)$$

Differentiating equation 6.2 with respect to r gives the differential equation

$$\frac{dM(r)}{dr} = 4\pi\rho_{\text{dm}}(r)r^2 - \frac{4\pi}{G} \frac{dP_g(r)}{dr} \frac{r^4}{M(r)}. \quad (6.3)$$

Assuming a GNFW profile (equation 2.21) for P_e , and relating it to P_g using equation 2.31 the second term on the RHS of equation 6.3 becomes

$$\frac{\mu_e}{\mu_g} \frac{4\pi P_{\text{ei}}}{G} \frac{r^3}{M(r)} \left(\frac{r}{r_p}\right)^{-c} \left[1 + \left(\frac{r}{r_p}\right)^a\right]^{-(1+(b-c)/a)} \left[b\left(\frac{r}{r_p}\right)^a + c\right] \equiv \frac{K(r)}{M(r)}. \quad (6.4)$$

Hence for PMT I (NFW dark matter profile)

$$\frac{dM(r)}{dr} = 4\pi\rho_s r_s^3 \frac{r}{(r + r_s)^2} + \frac{K(r)}{M(r)}, \quad (6.5)$$

and for PMT II (Einasto dark matter profile)

$$\frac{dM(r)}{dr} = 4\pi\rho_{-2} r^2 \exp\left[-\frac{2}{\alpha_{\text{Ein}}} \left(\left(\frac{r}{r_{-2}}\right)^{\alpha_{\text{Ein}}} - 1\right)\right] + \frac{K(r)}{M(r)}. \quad (6.6)$$

6.2 Determining cluster profile parameters

Equations 6.5 and 6.6 are first order non-linear differential equations with dependent variable M and independent variable r . They are subject to the boundary condition that $M(r_{200})$ = the value input to the model. Each equation has four unknown parameters: r_s for PMT I (r_{-2} for PMT II), ρ_s for PMT I (ρ_{-2} for PMT II), r_p and P_{ei} . r_s (r_{-2}) can be calculated the same way as previously. ρ_s (ρ_{-2}) can be calculated in a similar way to previously (i.e. as in Section 2.4 for PM I and Section 5.1.0.4 for PM II), but we now solve

$$M_{\text{dm}}(r) = (1 - f_{\text{gas}}(r)) \int_0^r 4\pi\rho_{\text{dm}}(r')r'^2 dr', \quad (6.7)$$

at $r = r_{200}$ for known $M(r_{200})$ and $f_{\text{gas}}(r_{200})$. However r_p can no longer be determined, since the mapping from r_{200} to r_{500} explicitly requires the assumption $M(r) = M_{\text{dm}}(r)$ for both dark matter models. Thus P_{ei} cannot be uniquely determined from the ODEs, as there is a family of solutions of (r_p, P_{ei}) which satisfy the ODEs, and therefore the pressure profile is no longer uniquely defined for a given set of cluster input parameters. I have thought of three ways to overcome this issue, only one of which I pursue. Nevertheless I now give a brief note on all three ideas.

6.2.1 Determining r_p and P_{ei} directly from constraints on M and its derivative

If we knew the value of $M(r)$ and $dM(r)/dr$ at two different radii then we would be able to determine unique values of r_p and P_{ei} directly from the ODEs. However I have not been able to think of any sensible conditions to impose on $dM(r)/dr$ other than $dM(r \rightarrow \infty)/dr = 0$. Furthermore, evaluating equation 6.3 asymptotically (i.e. $r \rightarrow 0$ and $r \rightarrow \infty$) does not yield any useful results. I therefore have not been able to use this method successfully in determining r_p and P_{ei} .

6.2.2 Determining r_p and P_{ei} using Lagrange multipliers

Consider the function

$$g(P_{ei}, r_p, r) = M(r) - M_{dm}(r) - M_g(P_{ei}, r_p, r), \quad (6.8)$$

which tells us that (P_{ei}, r_p) must satisfy $g(P_{ei}, r_p, r) = 0$ for all r . Since the ODEs in Section 6.1 are derived from $g(P_{ei}, r_p, r) = 0$, they share the same family of solutions of (P_{ei}, r_p) . Thus finding values of (P_{ei}, r_p) which satisfy the ODEs (subject to their boundary condition on $M(r_{200})$) also satisfies $g = 0$ (subject to the same boundary condition). We can formulate a constrained optimisation problem using Lagrange multipliers

$$f(P_{ei}, r_p) - \lambda g(P_{ei}, r_p, r) \quad (6.9)$$

to find stationary points in $f(P_{ei}, r_p)$ subject to the constraint $g = 0$ for arbitrary λ . The form of $f(P_{ei}, r_p)$ dictates the nature of (P_{ei}, r_p) at which the stationary point(s) of equation 6.8 are observed. For example $f(P_{ei}, r_p) = (P_{ei}r_p)^2$ would find the minimum value of the product $P_{ei}r_p$ which satisfies $g = 0$.

I do not pursue this idea any further however, since I cannot justify using a particular form for $f(P_{ei}, r_p)$, and because I suspect that finding the stationary points of equation 6.9 is difficult numerically.

6.2.3 Determining r_p and P_{ei} using approximate methods

Since ρ_s (ρ_{-2}) can be (correctly) calculated from equation 6.7 for the PMTs, we can use it in the calculational steps given by the PMs to get approximate values for r_p and P_{ei} . The issue with this method is that it is difficult to quantify the assumptions made, as we start off considering dark matter and gas contributions to the total mass to calculate ρ_s (ρ_{-2}), but then have to resort to

the $M(r_{200}) \approx M_{\text{dm}}(r_{200})$ to calculate r_p and P_{ei} . Despite this issue, I have adopted this method (due to its simplicity) to plot the mass profiles of clusters with a range of input parameters for illustrative purposes below. Note however that I have not implemented the PMTs into the Bayesian analysis software McADAM, since not being able to quantify the assumptions of the models invalidates their use in Bayesian inference.

6.3 Mass profile plots

We now compare the mass profiles of PM I and PM II (calculated using equations 2.23 and 5.4 respectively), with those obtained from PMT I and PMT II (calculated using equations 6.5 and 6.6 respectively), using values for r_p and P_{ei} obtained using the method outlined in Section 6.2.3 for the PMTs. As in Section 5.2.1 we consider two input masses, $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ and $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$, which roughly span the range of galaxy cluster masses. We use z -values of 0.15 and 0.9, and take $f_{\text{gas}}(r_{200}) = 0.12$ following Komatsu et al. (2011). For PM II and PMT II we consider α_{Ein} values of 0.05, 0.2, and 2.0.

Figure 6.1 shows the profiles for low $M(r_{200})$ and z . All four profiles are similar up to $r \approx 1 \text{ Mpc}$ (which is also $\approx r_{200}$), after which the NFW and Einasto profiles diverge. The two NFW profiles (PM I and PMT I) have roughly the same shape, but start to diverge slightly at high r (10 Mpc) with PM I taking higher values than PMT I. In the case of $\alpha_{\text{Ein}} = 0.05$, both PM II and PMT II diverge to large mass values at high r , with PMT II taking smaller values than PM II. $\alpha_{\text{Ein}} = 0.2$ shows a similar relationship between PM II and PMT II, but with the two taking lower values than PM I and PMT I at high r . $\alpha_{\text{Ein}} = 2$ presents an interesting result as PMT II does not appear to converge at high r like PM II does. Note that for the Einasto dark matter profile,

$$\lim_{r \rightarrow \infty} M_{\text{dm}}(r) = 4\pi\rho_{-2} \exp(2/\alpha_{\text{Ein}}) r_{-2}^3 \left(\frac{\alpha_{\text{Ein}}}{2}\right)^{\frac{3}{\alpha_{\text{Ein}}}} \frac{1}{\alpha_{\text{Ein}}} \Gamma\left(\frac{3}{\alpha_{\text{Ein}}}\right), \quad (6.10)$$

and so the first term on the right hand side of equation 6.2 is roughly constant at high r , meaning the increase in mass must be from the gas component. It seems unphysical that the gas content would contribute so much to the total mass at high r and thus questions the validity of the model (at least for the values of r_p and P_{ei} used here). Figures 6.2, 6.3, and 6.4 show the profiles for the other three combinations of $M(r_{200})$ and z inputs: low mass & high z ; high mass & low z and high mass & high z , respectively. All three cases show similar results between the approximate and full mass results to the previous case, which implies that the desparity between the two sets of results is not dependent on the input parameters (boundary conditions imposed on the ODEs).

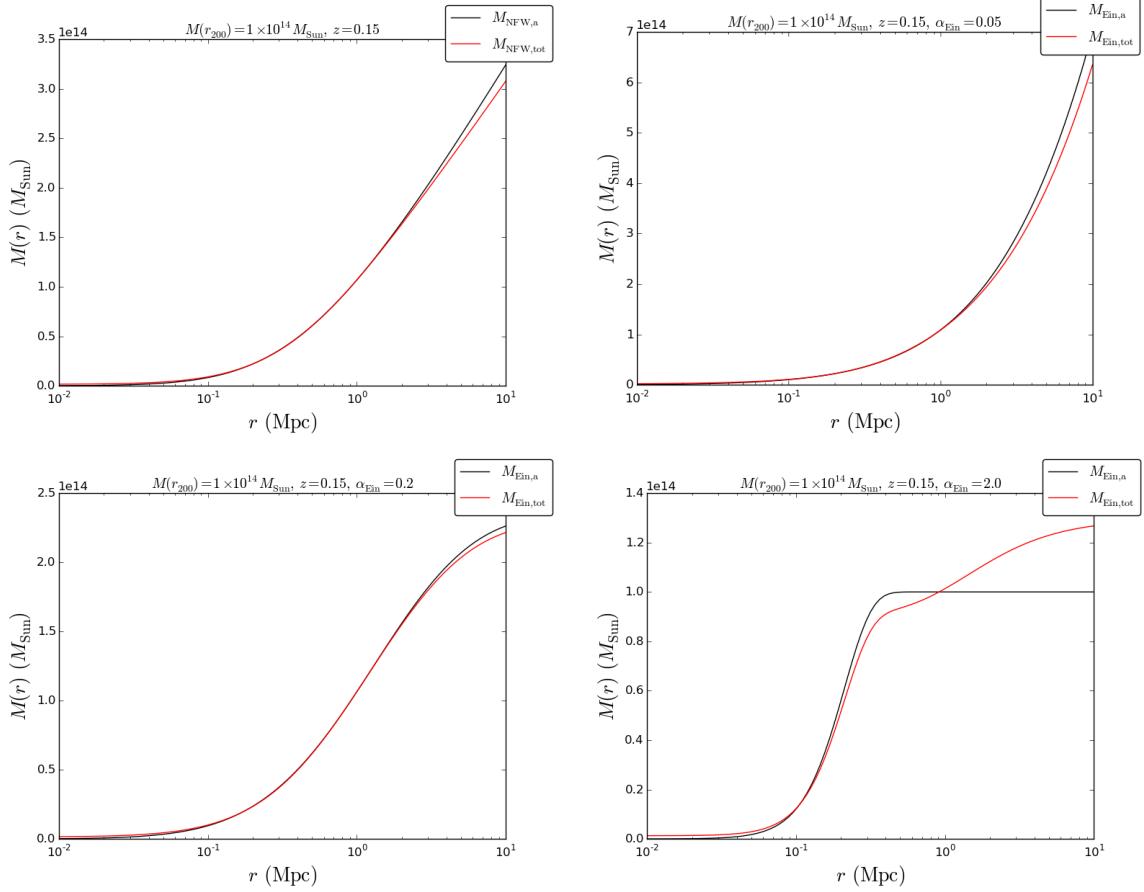


Figure 6.1: Mass profiles of cluster with input parameters given in titles. PM I (NFW dark matter profile, $M(r) \approx M_{\text{dm}}(r)$ approximation) and PMT I (NFW dark matter profile, $M(r) = M_{\text{dm}}(r) + M_g(r)$) are shown in the top left graph by black and red curves respectively. The other three graphs plot PM II (Einasto dark matter profile, $M(r) \approx M_{\text{dm}}(r)$ approximation), and PMT II (Einasto dark matter profile, $M(r) = M_{\text{dm}}(r) + M_g(r)$) in black and red respectively, for α_{Ein} values of 0.05 (top right), 0.2 (bottom left), and 2.0 (bottom right).

6.4 Conclusions

This Chapter relaxes the $M(r_{200}) \approx M_{\text{dm}}(r_{200})$ assumption present in the physical model presented in Section 2.4 (PM I) and the equivalent Einasto physical model (Section 5.1, PM II), to see if this would produce more physically plausible models for clusters. I derive two new models PMT I and PMT II based on the equation $M(r) = M_{\text{dm}}(r) + M_g(r)$. Both PMTs require non-linear ordinary differential equations in $M(r)$ to be solved. But to do this, values for r_p and P_{ei}

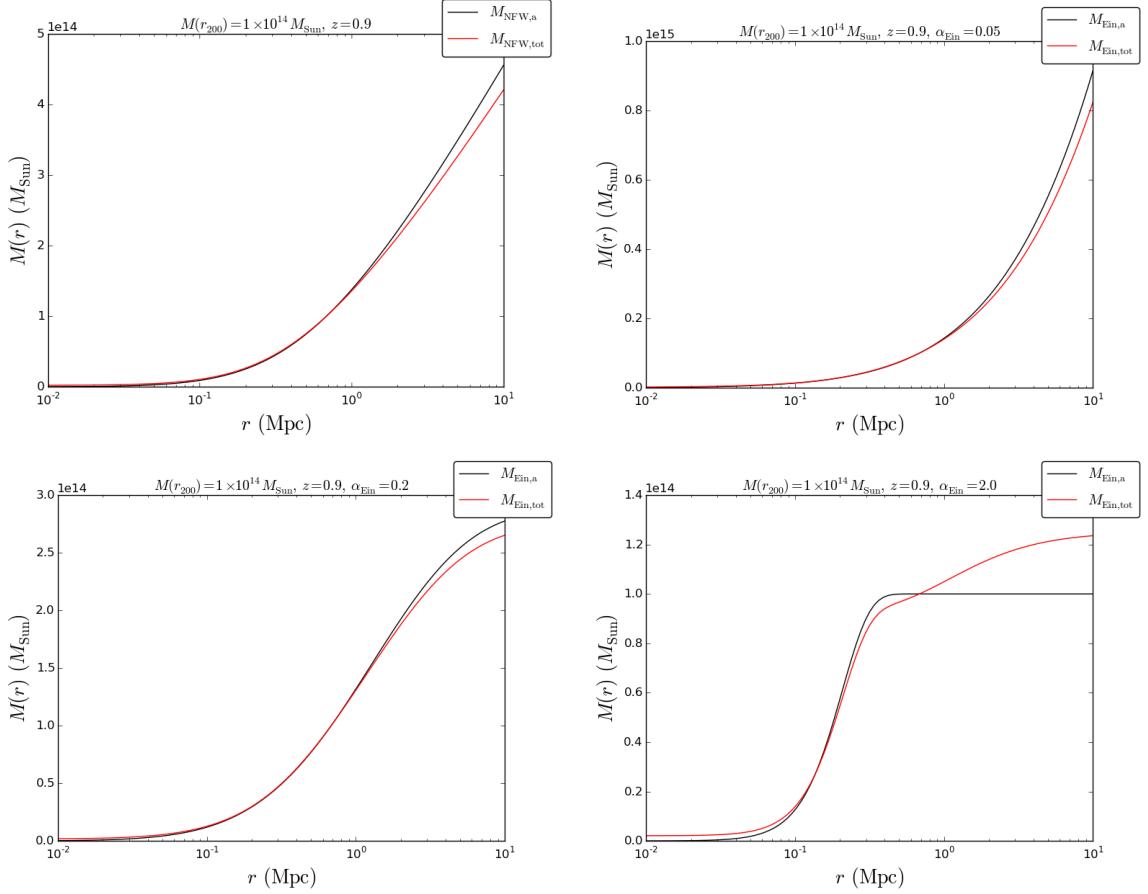


Figure 6.2: Mass profiles of cluster with input parameters given in title, for models given in Figure 6.1.

need to be determined and this turns out to be a non-trivial process. I investigated three possible ways of calculating r_p & P_{ei} and found the following.

- Determining r_p and P_{ei} directly from constraints on M and its derivative seems implausible as we could not justify assigning a particular value to the derivative at any finite value of r .
- Treating the problem of finding a value of r_p and P_{ei} from the family of solutions as a constrained optimisation problem (with a function $f(r_p, P_{\text{ei}})$ dictating the nature of the values of r_p and P_{ei} obtained, and the differential equations in $M(r)$ providing the constraints) seems promising in theory. However in practice, justifying a particular form for $f(r_p, P_{\text{ei}})$ isn't straightforward, and I anticipate that the optimisation is difficult numerically.

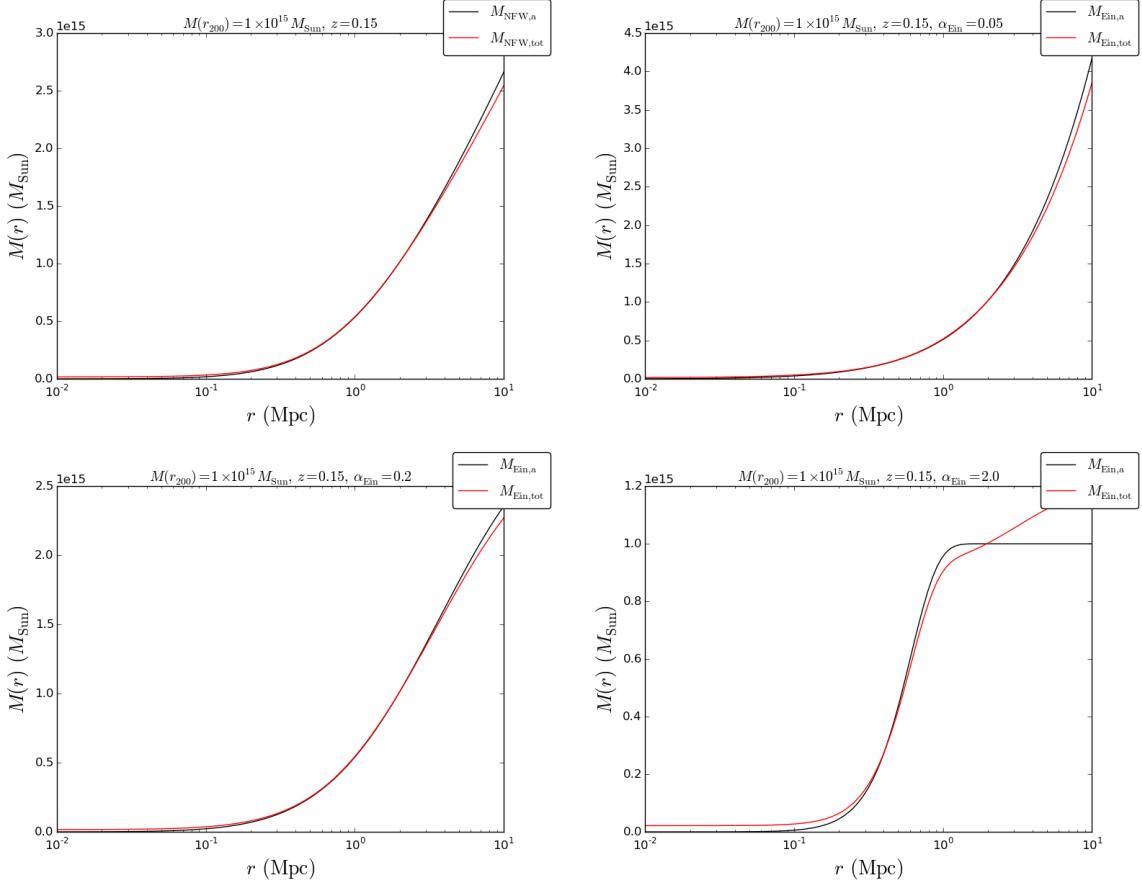


Figure 6.3: Mass profiles of cluster with input parameters given in title, for models given in Figure 6.1.

- The third method relied on using the calculational steps of PM I and PM II to determine ‘approximate’ values for r_p and P_{ei} from ‘true’ values of r_s for PMT I (r_{-2} for PMT II) and ρ_s for PMT I (ρ_{-2} for PMT II) calculated without the assumption that $\rho_g \ll \rho_{dm}$.

The third method was by far the simplest and the one I used to plot the mass profiles for the PMTs to compare with the equivalent PM profiles. From plots of the profiles we found the following.

- The values of the input parameters $M(r_{200})$ and z had very little effect on the shape or scale of the PMT I or PMT II profiles when compared with the corresponding PM profiles.
- For the $\alpha_{Ein} = 2$ case PMT II does not show the convergence in mass at high r that PM II does. Since $M_{dm}(r)$ asymptotically converges as $r \rightarrow \infty$ this implies that it is the gas

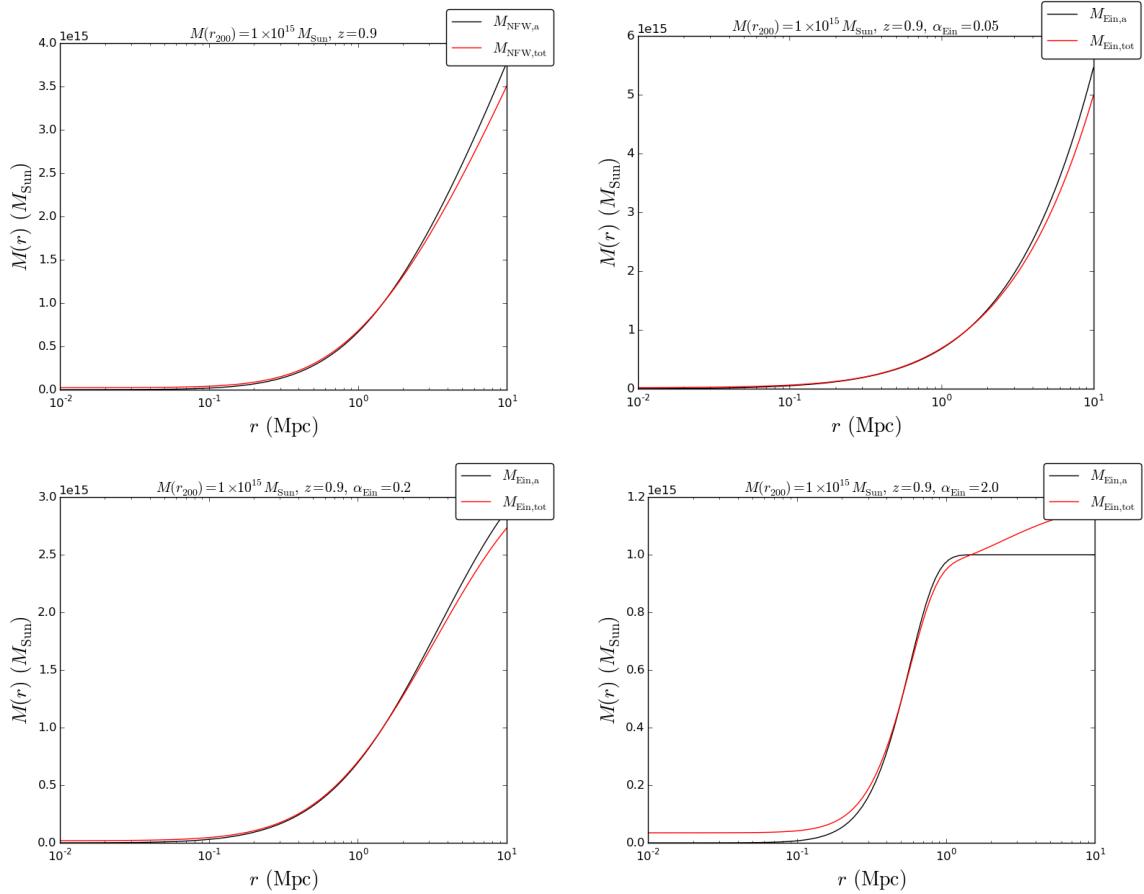


Figure 6.4: Mass profiles of cluster with input parameters given in title, for models given in Figure 6.1.

which is contributing to the mass increase, which seems unphysical for large r .

ENHANCED PHYSICAL MODELLING II: INCLUSION OF NON-THERMAL PRESSURE

All physical models presented so far in this thesis assume that the cluster gas pressure comes solely from the *thermal* gas pressure. Cosmological simulations have long predicted that magnetic fields, gas bulk motion and turbulence contribute to pressure support (see e.g. [Lau, Kravtsov, & Nagai 2009](#), [Battaglia et al. 2010](#), [Battaglia et al. 2011](#), [Nagai & Lau 2011](#), [Nagai 2011](#), [Battaglia et al. 2012](#), and [Parrish et al. 2012](#)).

Observational studies of clusters using the Chandra, Suzaku and XMM-Newton satellites have long invoked (see e.g. [Bautz et al. 2009](#) for cluster A1795, [Reiprich et al. 2009](#) A2204, [George et al. 2009](#) PKS0745-191, [Hoshino et al. 2010](#) A1413, [Kawaharada et al. 2010](#) A1689, [Urban et al. 2011](#) Virgo and [Simionescu et al. 2011](#) Perseus) these additional pressure sources to explain their observations. So including a non-thermal contribution to the hydrostatic equilibrium (HSE) relation given by equation 2.30, and altering the succeeding calculational steps of the PMs accordingly should be interesting.

In this Chapter I first give an overview of the contributors to non-thermal pressure. I then derive physical models for both NFW and Einasto (dark matter) models and incorporate non-thermal pressure into the HSE equation. We refer to these two models as PMN I and PMN II. We then plot the cluster parameter profiles of PMN I & PMN II and compare with those already obtained for PM I and PM II. Note that I do not include any modifications discussed in Chapter 6

here.

7.1 Non-thermal gas pressure

Galaxies orbiting or infalling onto clusters not only stir the gas, but also make the ICM clumpier. In the dense inner regions of clusters, these clumps only exist on short timescales as the ram pressure (pressure exerted on a body as it moves through a fluid medium) acting on the gas is high. At higher radii where the average cluster density is lower, orbital times are longer and accretion of new cluster material is ongoing, clumpiness can have significant effects on the total pressure profile. The clumpiness of the ICM depends on a number of physical processes, such as efficient feedback, which removes gas from merging structures, and thermal conduction, which homogenises the ICM temperature (see e.g. [Dolag et al. 2004](#)). Cosmic rays can originate from accretion shocks and supernova explosions, active galactic nuclei (AGN), and radio galaxies (see [Brunetti & Jones 2014](#) for a review).

7.2 Modelling non-thermal gas pressure

7.2.1 Analytic expression for non-thermal gas pressure

[Martizzi & Agrusa \(2016\)](#) (from here on DM16) derive an analytic expression for the non-thermal pressure P_{nt} component in galaxy clusters. They derive the function $P_{\text{nt}}(r)$ by considering a subset of ten cosmological hydrodynamical zoom-in simulations of galaxy clusters from the sample of [Martizzi et al. \(2014\)](#). The ten simulations were performed using the `RAMSES` code ([Teyssier 2002](#)) and have total masses $> 10^{14} M_{\text{Sun}}$. Half of the subsample are relaxed according to the criteria outlined in Section 2.1 of DM16 (based on the ratio of the velocity dispersion of dark matter particles to the velocity dispersion of an equivalent virialised system). These simulations do not include non-thermal contributions from cosmic rays and magnetic fields.

DM16 derive an expression for P_{nt} by evolving a cluster from high z and measuring its $\rho_g(r)$, $M(r)$, thermal pressure P_{th} and thermal mass $M_{\text{th}}(r)$. From these four quantities the form of P_{nt} can be determined from the HSE relation (equation 2.30) (assuming that $P_g = P_{\text{th}} + P_{\text{nt}}$). The following analytic expression is obtained by fitting to the simulated data using a least squares regression

$$P_{\text{nt}}(r) = 5.388 \times 10^{13} \left(\frac{r_{200,\text{m}}}{\text{Mpc}} \right)^3 \left(\frac{\rho_g(r)}{\text{g/cm}^3} \right) \text{erg/cm}^3, \quad (7.1)$$

where $r_{200,m}$ is the radius at which the average cluster density is $200 \times$ the *average* matter density $\rho_m(z) = 3H_0^2/(8\pi G)Q_M(1+z)^3$. Here H_0 is the Hubble parameter evaluated at $z = 0$. $r_{200,m}$ can be calculated from r_{200} in a similar way to how r_{500} is (for the NFW case, see equations 2.25 through to 2.29, and for Einasto see Section C.2).

Expressed in ‘astronomical’ units ($M_{\text{Sun}} \text{Mpc}^{-1} \text{s}^{-2}$), equation 7.1 can be written as

$$P_{\text{nt}}(r) = \beta \left(\frac{r_{200,m}}{\text{Mpc}} \right)^3 \rho_g(r), \quad (7.2)$$

where $\beta = 5.658 \times 10^{-36} \text{ Mpc}^2 \text{s}^{-2}$.

7.2.2 Incorporating non-thermal pressure into the physical models

Redefining $P_g(r)$ from equation 2.31 as

$$P_g(r) \equiv P_{\text{th}}(r) + P_{\text{nt}}(r), \quad (7.3)$$

where

$$P_{\text{th}}(r) = \frac{\mu_e}{\mu_g} \frac{P_{\text{ei}}}{\left(\frac{r}{r_p} \right)^c \left(1 + \left(\frac{r}{r_p} \right)^a \right)^{(b-c)/a}}, \quad (7.4)$$

and re-evaluating the HSE relation with the new form of $P_g(r)$ gives

$$\frac{dP_g(r)}{dr} = \frac{d}{dr} [P_{\text{th}}(r) + P_{\text{nt}}(r)] = -\frac{G\rho_g(r)M(r)}{r^2}. \quad (7.5)$$

Equation 7.5 can be rearranged to give

$$\frac{d\rho_g(r)}{dr} + \frac{G(\text{Mpc})^3 M(r)}{\beta r_{200,m}^3} \frac{1}{r^2} \rho_g(r) = \frac{\mu_e}{\mu_g} \left(\frac{\text{Mpc}}{r_{200,m}} \right)^3 \frac{P_{\text{ei}}}{\beta} \left[\frac{1}{r} \left(\frac{r}{r_p} \right)^{-c} \left[1 + \left(\frac{r}{r_p} \right)^a \right]^{-(1+(b-c)/a)} \left[b \left(\frac{r}{r_p} \right)^a + c \right] \right], \quad (7.6)$$

which is an inhomogeneous first order linear ODE with dependent variable $\rho_g(r)$ and independent variable r . Since equation 7.6 includes a $M(r)$ term, its final form depends on the dark matter profile considered. Note that when calculating $M(r)$ for either PMN I or PMN II we assume $M(r) \approx M_{\text{dm}}(r)$ as we did when profiling PM I and PM II.

For PMN I we have the expression for $M(r)$ given by equation 2.23 and so the differential equation becomes

$$\begin{aligned} \frac{d\rho_g(r)}{dr} + \frac{4\pi G(\text{Mpc})^3 \rho_s r_s^3}{\beta r_{200,m}^3} \frac{\left[\ln \left(1 + \frac{r}{r_s} \right) - \left(1 + \frac{r_s}{r} \right)^{-1} \right]}{r^2} \rho_g(r) \\ = \frac{\mu_e}{\mu_g} \left(\frac{\text{Mpc}}{r_{200,m}} \right)^3 \frac{P_{\text{ei}}}{\beta} \left[\frac{1}{r} \left(\frac{r}{r_p} \right)^{-c} \left[1 + \left(\frac{r}{r_p} \right)^a \right]^{-(1+(b-c)/a)} \left[b \left(\frac{r}{r_p} \right)^a + c \right] \right]. \end{aligned} \quad (7.7)$$

For PMN II we have the expression for $M(r)$ given by equation 5.4 and so

$$\begin{aligned} \frac{d\rho_g(r)}{dr} + \frac{4\pi G(\text{Mpc})^3 \rho_{-2} r_{-2}^3 \exp(2/\alpha_{\text{Ein}}) \left(\frac{\alpha_{\text{Ein}}}{2}\right)^{3/\alpha_{\text{Ein}}} \gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r}{r_{-2}}\right)^{\alpha_{\text{Ein}}}\right]}{r^2 \beta r_{200,m}^3 \alpha_{\text{Ein}}} \rho_g(r) \\ = \frac{\mu_e}{\mu_g} \left(\frac{\text{Mpc}}{r_{200,m}}\right)^3 \frac{P_{\text{ei}}}{\beta} \left[\frac{1}{r} \left(\frac{r}{r_p}\right)^{-c} \left[1 + \left(\frac{r}{r_p}\right)^a\right]^{-(1+(b-c)/a)} \left[b \left(\frac{r}{r_p}\right)^a + c\right]\right]. \end{aligned} \quad (7.8)$$

For brevity we define the following constants

$$\begin{aligned} \omega_{\text{NFW}} &\equiv \frac{4\pi G \rho_s r_s^3 (\text{Mpc})^3}{\beta r_{200,m}^3} \\ \omega_{\text{Ein}} &\equiv \frac{4\pi G \rho_{-2} r_{-2}^3 (\text{Mpc})^3}{\beta r_{200,m}^3} \frac{1}{\alpha_{\text{Ein}}} \left(\frac{\alpha_{\text{Ein}}}{2}\right)^{3/\alpha_{\text{Ein}}} \exp(2/\alpha_{\text{Ein}}) \\ \sigma &\equiv \frac{\mu_e}{\mu_g} \frac{1}{\beta} \left(\frac{\text{Mpc}}{r_{200,m}}\right)^3. \end{aligned} \quad (7.9)$$

In fact the inhomogeneous ODEs derived above can be transformed into homogeneous ODEs as follows. Consider a general ODE of the form

$$\frac{d\rho_g(r)}{dr} + g(r)\rho_g(r) = f(r), \quad (7.10)$$

then using an integrating factor defined by

$$I(r) = \exp\left(\int^r g(r') dr'\right), \quad (7.11)$$

equation 7.10 can be transformed into a homogeneous separable ODE which gives the result

$$\rho_g(r)I(r) - \rho_g(r_0)I(r_0) = \int_{r_0}^r I(r')f(r')dr', \quad (7.12)$$

where r_0 and $\rho_g(r_0)$ are dependent on the input parameters of the problem.

For PMN I this gives

$$\begin{aligned} \rho_g(r) \left(1 + \frac{r}{r_s}\right)^{-\omega_{\text{NFW}}/r} - \rho_g(r_0) \left(1 + \frac{r_0}{r_s}\right)^{-\omega_{\text{NFW}}/r_0} \\ = P_{\text{ei}} \sigma \int_{r_0}^r \left(1 + \frac{r'}{r_s}\right)^{-\omega_{\text{NFW}}/r'} \left[\frac{1}{r'} \left(\frac{r'}{r_p}\right)^{-c} \left[1 + \left(\frac{r'}{r_p}\right)^a\right]^{-(1+(b-c)/a)} \left[b \left(\frac{r'}{r_p}\right)^a + c\right]\right] dr'. \end{aligned} \quad (7.13)$$

For PMN II the integrating factor does not have an analytical form, hence the homogeneous form can only be simplified to

$$\begin{aligned} \rho_g(r) \exp \left(\int_{r_0}^r \omega_{\text{Ein}} \frac{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r'}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]}{r'^2} dr' \right) - \rho_g(r_0) \exp \left(\int_{r_0}^{r_0} \omega_{\text{Ein}} \frac{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r'}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]}{r'^2} dr' \right) \\ = P_{\text{ei}} \sigma \int_{r_0}^r \left[\exp \left(\omega_{\text{Ein}} \int_{r'}^{r''} \frac{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r''}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]}{r''^2} dr'' \right) \right] \left[\frac{1}{r'} \left(\frac{r'}{r_p} \right)^{-c} \left[1 + \left(\frac{r'}{r_p} \right)^a \right]^{-(1+(b-c)/a)} \left[b \left(\frac{r'}{r_p} \right)^a + c \right] \right] dr'. \end{aligned} \quad (7.14)$$

It is also interesting to see if the non-thermal only pressure term provides a solution to the HSE (I have already verified this is the case for $P_{\text{th}}(r)$ in PM I and PM II, by deriving the relevant expressions for $\rho_g(r)$). Putting the expression for $P_{\text{nt}}(r)$ into the HSE gives

$$\int_{r_0}^{r_1} \frac{1}{\rho_g(r')} \frac{d\rho_g(r')}{dr'} dr' = \left(\frac{\text{Mpc}}{r_{200,\text{m}}} \right)^3 \frac{G}{\beta} \int_{r_0}^{r_1} \frac{M(r')}{r'^2} dr'. \quad (7.15)$$

For the NFW dark matter profile this gives

$$\rho_g(r_1) = \rho_g(r_0) \frac{\left(1 + \frac{r_1}{r_s} \right)^{\alpha_{\text{NFW}}/r_1}}{\left(1 + \frac{r_0}{r_s} \right)^{\alpha_{\text{NFW}}/r_0}}. \quad (7.16)$$

As was the case with the integrating factor in the full solution for the thermal and non-thermal pressure, the Einasto dark matter profile does not give an analytic solution the non-thermal only case

$$\rho_g(r_1) = \rho_g(r_0) \exp \left[\omega_{\text{Ein}} \int_{r_0}^{r_1} \frac{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r'}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]}{r'^2} dr' \right]. \quad (7.17)$$

Note that since the HSE is a inhomogeneous differential equation, the solutions associated with P_{th} and P_{nt} do not sum to solution associated with $P_g = P_{\text{th}} + P_{\text{nt}}$.

7.2.3 Determining P_{ei} for the non-thermal case

Equation 7.7 (equation 7.8) has four unknown parameters: r_s (r_{-2}), ρ_s (ρ_{-2}), r_p and P_{ei} . The first three of these can be calculated in the same way as in PM I and PM II. However, as was the case with the full mass modelling in Chapter 6, P_{ei} cannot be calculated trivially from the input parameters and calculations derived above for the PMNs. Hence we consider the methods described in Sections 6.2.1, 6.2.2, and 6.2.3 which we denote method I, method II and method III respectively.

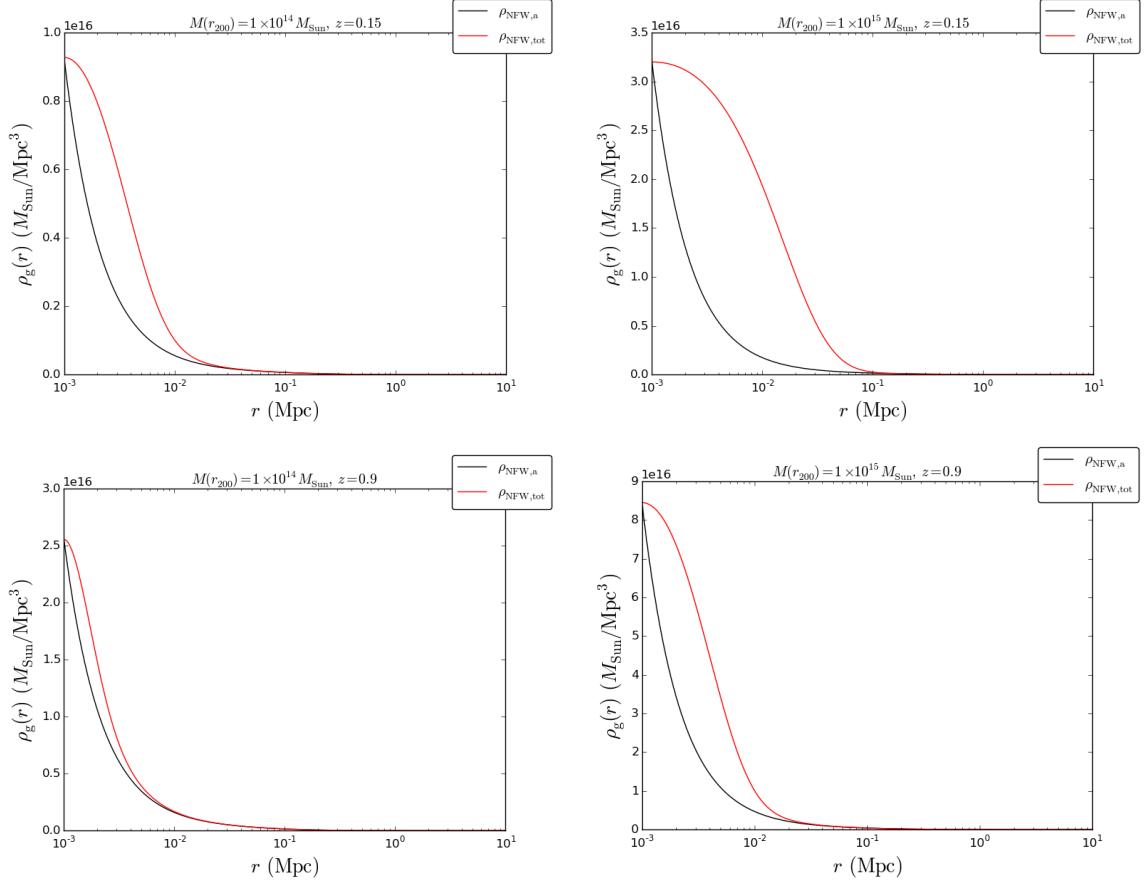


Figure 7.1: $\rho_g(r)$ profiles for PM I and PMN I. Each graph features both profiles for one of the four different input parameter sets. Top row has $z = 0.15$, bottom row has $z = 0.9$. Left column has $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$, right column has $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$

For method I, since there is only one unknown parameter we only need to know the value of $\rho_g(r)$ and its derivative at one point. It makes sense to consider the asymptotic case ($r \rightarrow \infty$) in which case $\rho_g(r)$ and its derivative tend to zero. However, since there is no constant term in the ODEs, this gives us (using equation 7.7 or 7.8) $0 = P_{\text{ei}} \times 0$ and thus P_{ei} cannot be determined. I have not been able to come up with any physically justified estimates for $\rho_g(r)$ and its derivative at finite r , and so I do not pursue this method any further.

Method II presents the same potential difficulties as in the Chapter 6, and so I do not pursue it here.

Method III would require us to get an approximate value for P_{ei} from the calculational steps of

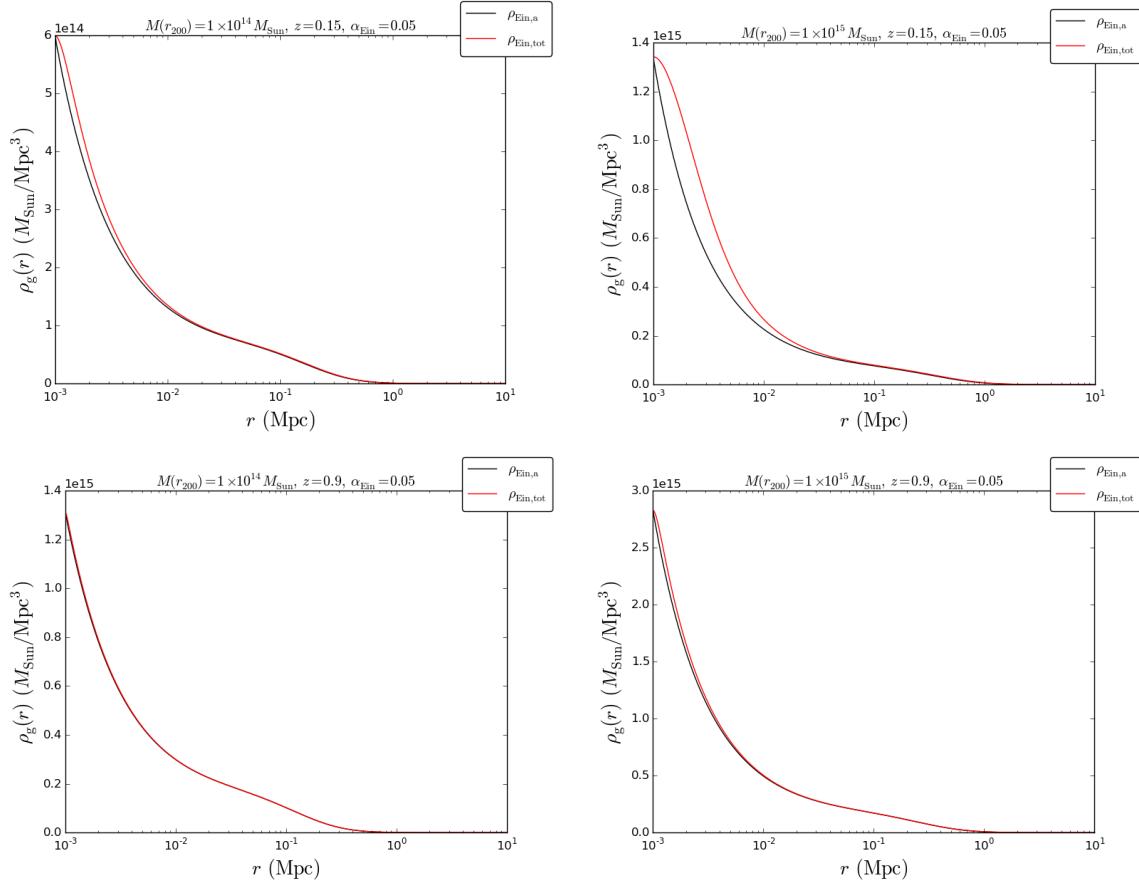


Figure 7.2: $\rho_g(r)$ profiles for PM II and PMN II with $\alpha_{\text{Ein}} = 0.05$. Graphs are laid out as described in Figure 7.1.

PM I and PM II. This requires us to ignore the non-thermal contribution in the HSE equation and determine an analytic form for $\rho_g(r)$. As was the case in Chapter 6, this is by far the simplest way of determining P_{ei} , I therefore use it to obtain cluster parameter profiles for PMN I & PMN II and compare them with those from PM I & PM II for illustrative purposes.

7.2.4 Boundary conditions for $\rho_g(r)$

I first tried setting $\rho_g(r_{\text{max}}) = 0$ where r_{max} is the upper limit on r used in the ODE solver. However, this failed to generate a sensible profile for $\rho_g(r)$. This is expected, as such an initial condition surely provides ‘too little’ information on the form of $\rho_g(r)$ to constrain its profile at low r . I next applied the initial condition $\rho_g(r_{\text{min}}) = \rho_{g,a}(r_{\text{min}})$, where $\rho_{g,a}$ is the value obtained

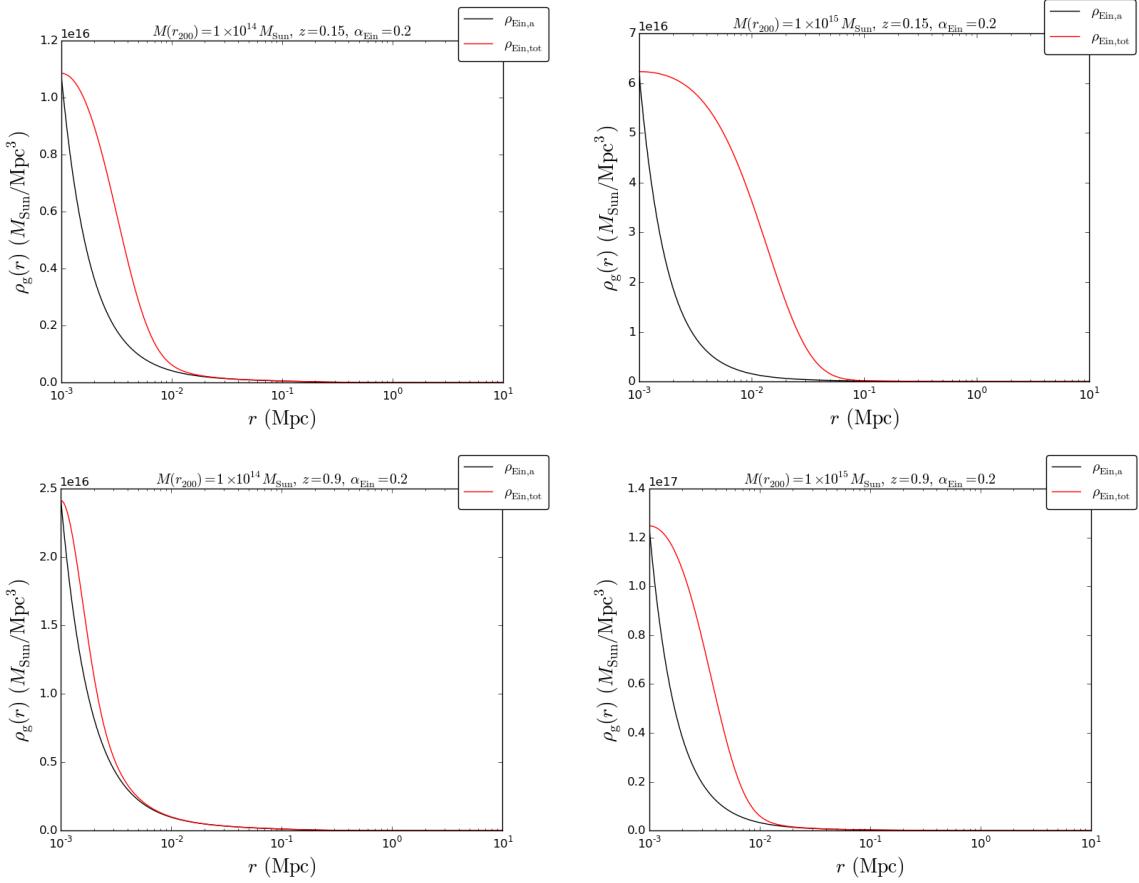


Figure 7.3: $\rho_g(r)$ profiles for PM II and PMN II with $\alpha_{\text{Ein}} = 0.2$. Graphs are laid out as described in Figure 7.1.

from PM I / PM II. I think this assumption is sensible, given that the non-thermal contributions are generally thought to be less and less significant at smaller radii as pointed out in Section 7.1. I generally found that the latter initial condition produced solution curves for $\rho_g(r)$ when solving the ODEs given by equations 7.7 and 7.8.

7.3 Non-thermal pressure profiling

As in the previous Sections which focus on cluster profiling, I create plots for clusters with input values of $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$ & $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$, and $z = 0.15$ & $z = 0.9$. For PM II and PMN II we consider α_{Ein} values of 0.05, 0.2, and 2.0.

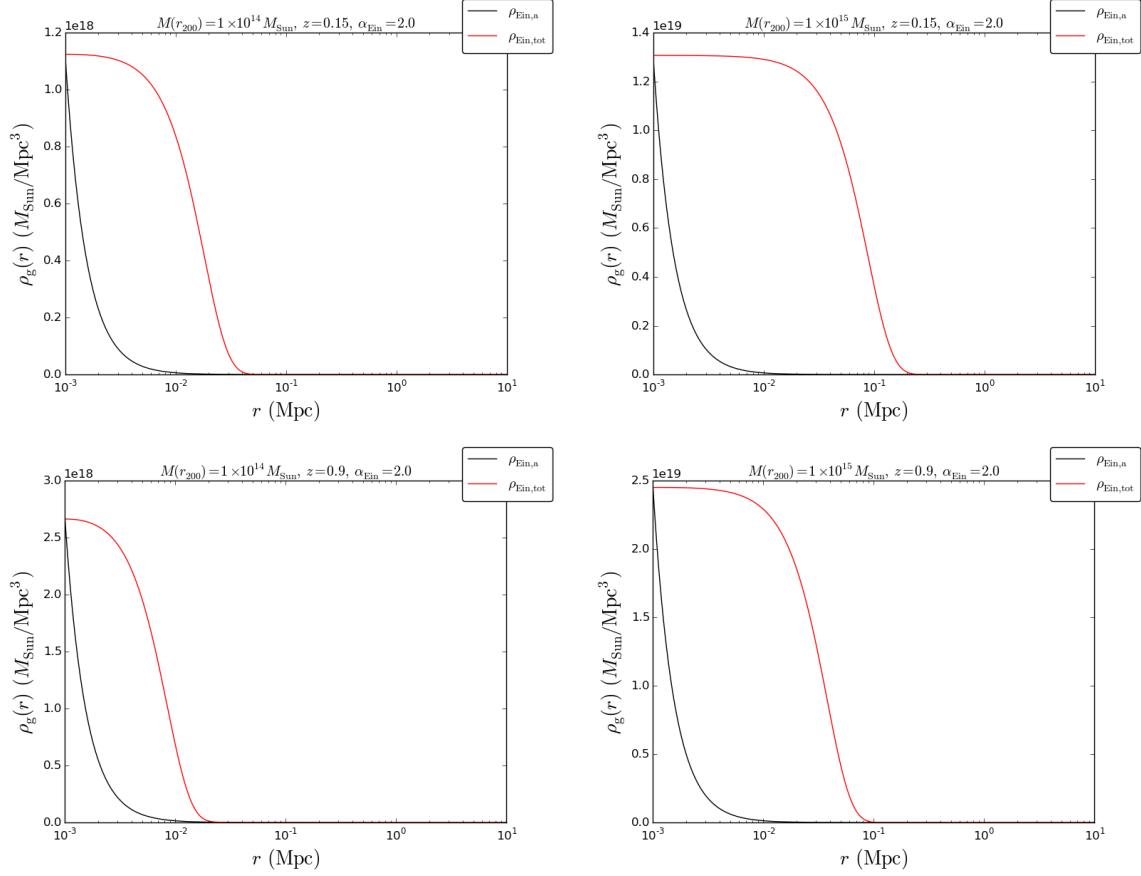


Figure 7.4: $\rho_g(r)$ profiles for PM II and PMN II with $\alpha_{\text{Ein}} = 2$. Graphs are laid out as described in Figure 7.1.

7.3.1 Gas density profiles

Figures 7.1, 7.2, 7.3, and 7.4 compare the PM and PMN profiles for the NFW, $\alpha_{\text{Ein}} = 0.05$, $\alpha_{\text{Ein}} = 0.2$, and $\alpha_{\text{Ein}} = 2$ cases respectively. The most striking feature of these graphs is the fact that the PMN profiles have higher gas densities than their PM equivalent for radii $> r_0$, until they decay to ≈ 0 at high r . As was the case with the PMT profiles, changing the mass / z input parameters does not seem to effect the shape of the PMN gas density profiles. However, unlike the comparison between the PM and PMT models, changing the input parameters here does seem to have an effect on the level of disparity between the PM and PMN profiles.

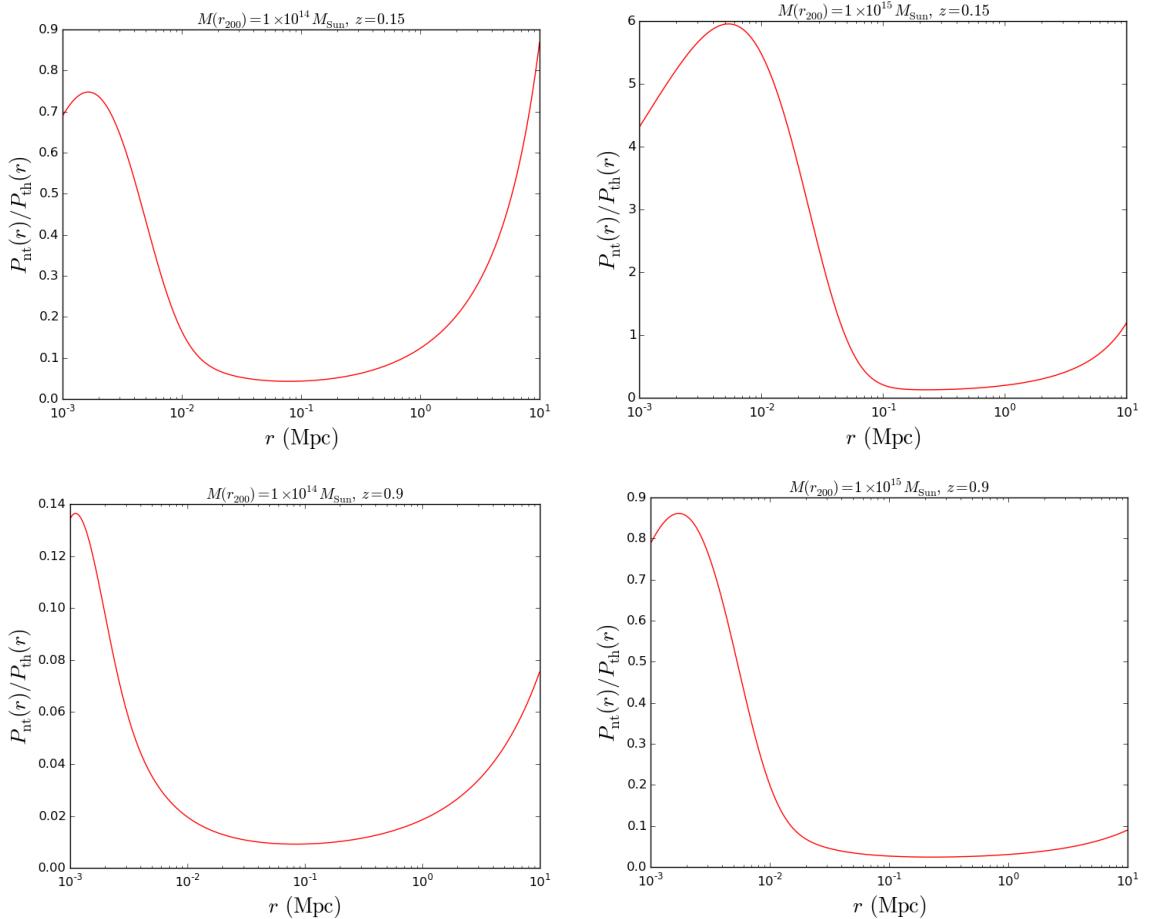


Figure 7.5: Ratio of P_{nt} to P_{th} profiles for PM I and PMN I. Each graph features both profiles for one of the four different input parameter sets. Top row has $z = 0.15$, bottom row has $z = 0.9$. Left column has $M(r_{200}) = 1 \times 10^{14} M_{\text{Sun}}$, right column has $M(r_{200}) = 1 \times 10^{15} M_{\text{Sun}}$

7.3.2 Thermal and non-thermal pressure profiles

Once $\rho_g(r)$ has been determined, P_{nt} can be calculated from equation 7.2. P_{th} is given by equation 7.4, and so is the same as the profiles of P_g calculated for the PMs. Furthermore the mass is still calculated using the approximation $M \approx M_{\text{dm}}$ and so it has identical values between the PMs and PMNs. Figures 7.5, 7.6, 7.7, and 7.8 show the ratio of non-thermal to thermal pressure (calculated from the PMNs) for the NFW, $\alpha_{\text{Ein}} = 0.05$, $\alpha_{\text{Ein}} = 0.2$, and $\alpha_{\text{Ein}} = 2$ cases respectively. Given that in simulations (Rasia, Tormen, & Moscardini 2004, Ameglio et al.

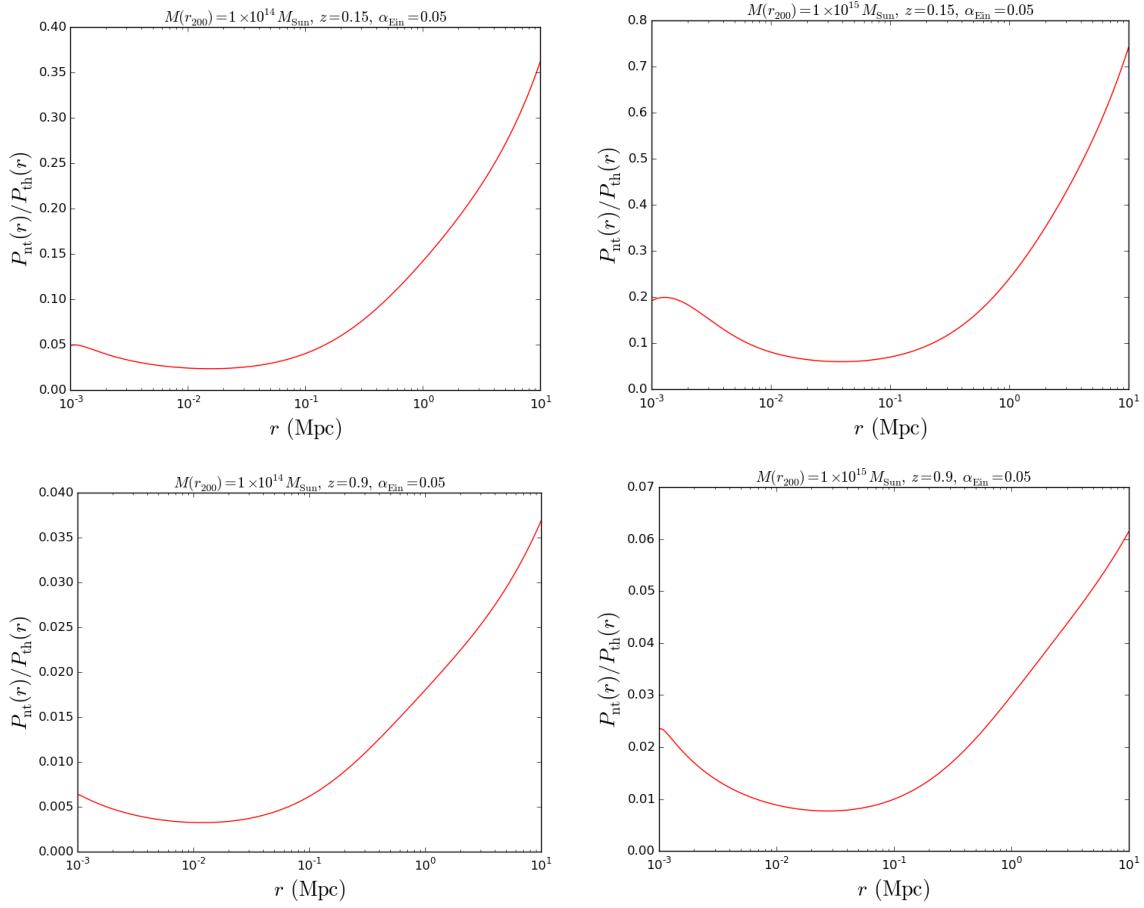


Figure 7.6: Ratio of P_{nt} to P_{th} profiles for PM II and PMN II with $\alpha_{\text{Ein}} = 0.05$. Graphs are laid out as described in Figure 7.5.

2009, Piffaretti & Valdarnini 2008, and Biffi, Dolag, & Böhringer 2011), non-thermal pressure was found to be at a maximum $\approx 15\%$ of the thermal pressure, these Figures show that the PMNs considered here are unphysical, particularly as the value of α_{Ein} increases. The only profiles which give sensible values are the $\alpha_{\text{Ein}} = 0.05$ cases. Here, the non-thermal pressure does go above 20%, but only at high r , where both types of pressure should take negligibly small values. Even though the ratio profiles look sensible for $\alpha_{\text{Ein}} = 0.05$, the fact they are off by such a large amount for the other clusters implies the models formulated here are probably unfeasible (including the validity of the method used to determine P_{el}), and that the case of one good result has probably been obtained by chance. However we do note that in DM16 the ratio approaches

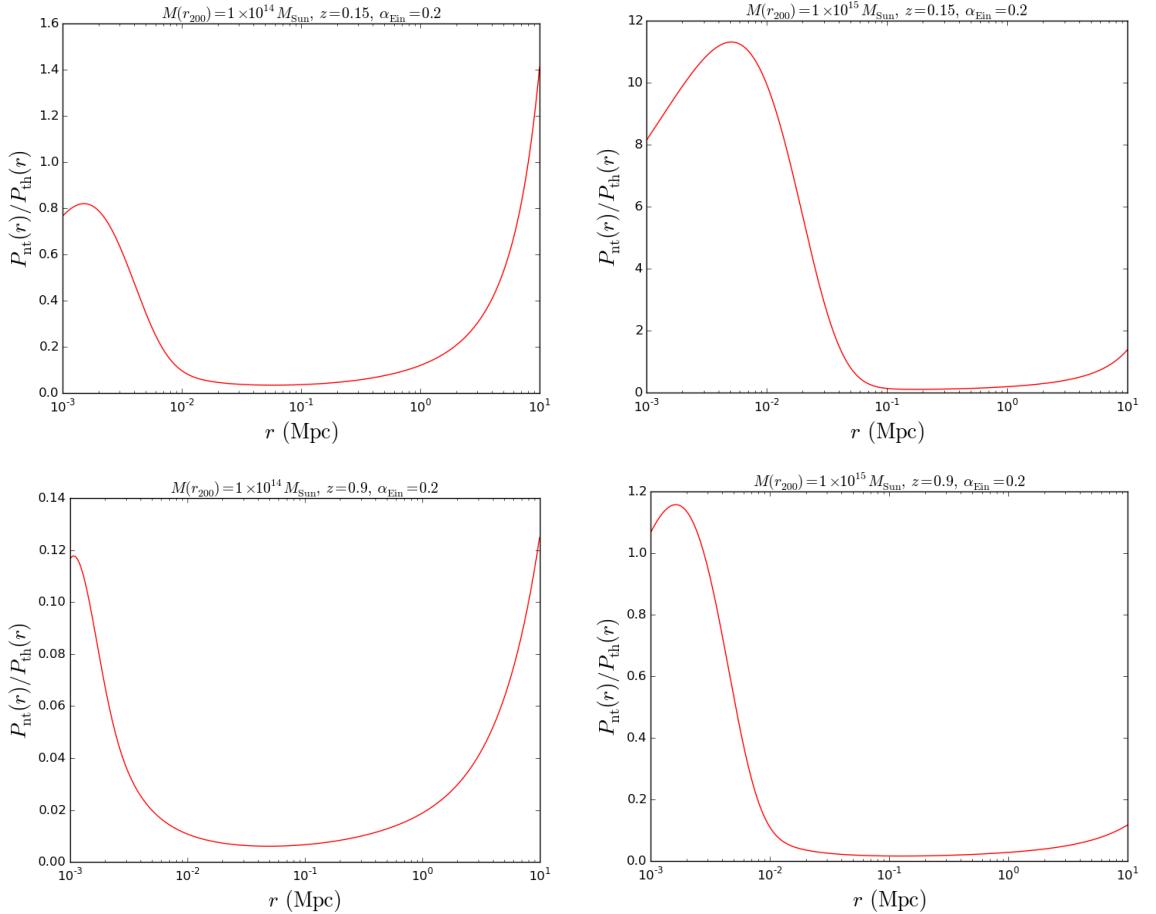


Figure 7.7: Ratio of P_{nt} to P_{th} profiles for PM II and PMN II with $\alpha_{\text{Ein}} = 0.2$. Graphs are laid out as described in Figure 7.5.

unity for five of the ten cluster sample of simulations at $r \approx r_{200,\text{m}}$ (see Figure 5 of DM16). Whilst this doesn't add any validity to the results, it does suggest that non-thermal pressure can contribute greatly (up to \approx the majority) towards the total pressure, and thus further work on incorporating its effect into cluster SZ models is important in improving their performance.

7.4 Conclusions

In this Chapter I incorporated non-thermal pressure into the physical models presented in Section 2.4 (PM I) and the equivalent Einasto physical model (Section 5.1, PM II) to see if this

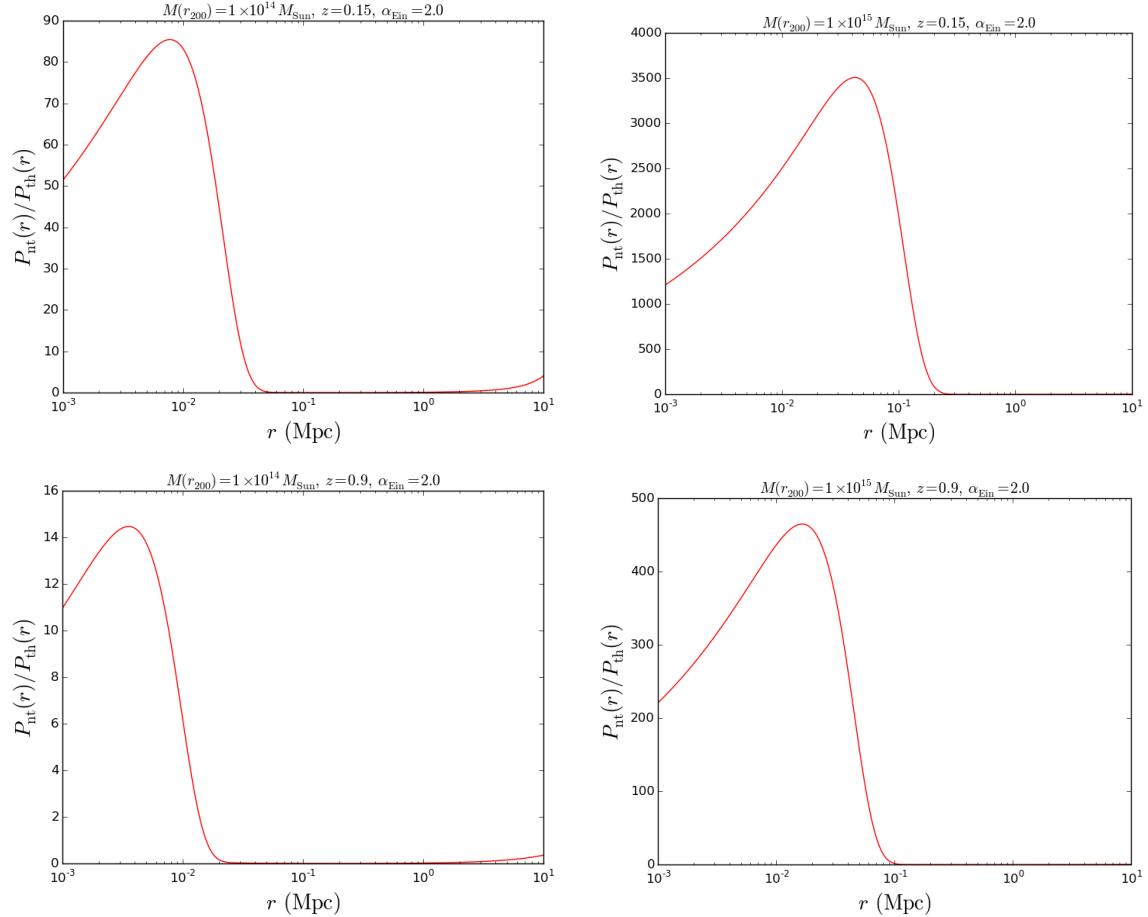


Figure 7.8: Ratio of P_{nt} to P_{th} profiles for PM II and PMN II with $\alpha_{\text{Ein}} = 2$. Graphs are laid out as described in Figure 7.5.

would produce more physically plausible models for clusters. I derive two new models PMN I and PMN II based on the analytical equation for non-thermal gas pressure in a cluster derived in [Martizzi & Agrusa \(2016\)](#). Both PMNs require linear inhomogeneous ordinary differential equations in $\rho_g(r)$ to be solved. However to do so, values for P_{ei} need to be determined as in Chapter 6. Due to its simplicity, I used the method outlined in Section 6.2.3 to determine an approximate value for P_{ei} . I then solved the ODEs in $\rho_g(r)$ for various cluster input parameters and found the following.

- The PMN profiles have higher gas densities than their PM equivalent, until they decay to ≈ 0 at high r .

- As was the case with the PMT profiles in Chapter 6, changing the mass or z input parameters does not seem to effect the shape of the PMN gas density profiles. However, unlike the comparison between the PM and PMT models, changing the input parameters here does seem to have an effect on the level of disparity between the PM and PMN profiles.

I then plotted the ratio of non-thermal to thermal pressure for different cluster inputs, to see how the ratio compared with those obtained from simulations in the literature, and found the following.

- For all but the $\alpha_{\text{Ein}} = 0.05$ clusters, the ratio of non-thermal to thermal pressure was unphysical, as it exceeded values well over 100% (which seems unfeasible in isolation, and even more unreasonable when compared to the values of $\approx 15\%$ obtained in simulations).
- Even though the ratio profiles looked sensible for $\alpha_{\text{Ein}} = 0.05$, the fact they were off by such a large amount for the other clusters implies the models formulated here are probably unfeasible, and that the case of one good result was probably obtained by chance.

CHAPTER 8

JOINT ANALYSIS OF AMI AND PLANCK DATA

Analysis of data obtained from different telescopes allows one to compare and verify inferences from measurements of different quantities, are subject to different systematic errors, and are obtained from different wavebands and on different angular scales. Simultaneous analysis of multiple datasets can lead to results different from those obtained in the individual cases, and can be used to investigate problems which cannot be resolved by the individual analyses.

In the context of galaxy clusters [LaRoque et al. \(2006\)](#) have used joint X-ray–SZ data in an attempt to constrain the dark energy equation of state parameter w . Similarly, cosmological constants have been estimated from X-ray analyses (see e.g. [Vikhlinin et al. 2009](#) and [Mantz et al. 2010](#)), SZ measurements (see e.g. [Muchovej et al. 2011](#), [Williamson et al. 2011](#), and [Marriage et al. 2011](#)) and a joint X-ray–SZ analysis ([Hasler et al. 2012](#)).

Joint analysis of data from galaxy clusters is not restricted to telescopes which measure different quantities. [Adam et al. \(2015\)](#), [Romero et al. \(2015\)](#), [Adam et al. \(2016\)](#), and [Romero et al. \(2017\)](#) all use SZ measurements from instruments including the Planck satellite ([The Planck Collaboration 2006](#)), Bolocam ([Sayers et al. 2013](#)) and ([Czakon et al. 2015](#)), Green Bank telescope ([Korngut et al. 2011](#)), and IRAM 30-metre telescope ([Monfardini et al. 2014](#)), that probe different angular scales and operate over different frequency ranges, to infer profiles of cluster parameters such as pressure, temperature and mass.

In this Chapter I carry out joint analysis of SZ data from AMI and from the Planck satellite. Note that I conduct *separate* analyses on these data in Chapter 3. I apply Bayesian analysis

using a joint likelihood for data from both instruments, to simulated cluster data generated with observational and physical models (largely based on the ones introduced in Chapters 2, 4, and 5). I analyse the resulting posterior distributions and compare them with results obtained from analysing the likelihoods for AMI and Planck separately.

I also apply the joint analysis to real data from the Planck detected cluster PSZ2G063.80+11.42, whose mass estimates derived from AMI and Planck data in Chapter 3 showed discrepancies with one another. Note the work in this Chapter has been published as a paper in MNRAS (Perrott et al. 2019), which I am a lead author of. The paper includes more information on how the Planck simulations were generated, and presents results of analyses where the simulated data was much better understood (and less prone to bugs).

8.1 Joint likelihood analysis

The key aspects of Bayesian inference have already been highlighted in Chapter 2. Nevertheless it is useful to highlight how we evaluate the joint likelihood function of datasets which have previously been analysed in isolation and with different analysis pipelines.

8.1.1 AMI data analysis

As previously, McADAM is used to calculate the posterior distribution for AMI data (see Section 8.1.3.1).

8.1.2 Planck detection algorithms

The Y and M values published in the Planck catalogue PSZ2 are derived from data from one of three detection algorithms: MMF1, MMF3 (Staniszewski et al. 2009; Marriage et al. 2011) and PowellSnakes (PwS, Carvalho et al. 2012). The mass estimates presented in Chapter 3 that are based on Planck data were calculated from the outputs of the PwS algorithm. Similarly the joint AMI-Planck analysis here uses PwS to process the data for the Planck part of the analysis (see Section 8.1.3.2).

8.1.3 Joint likelihood function

If one has an AMI dataset \mathbf{d}_{AMI} and a Planck dataset \mathbf{d}_{Pl} , then the *joint* likelihood function for the data is given by

$$\mathcal{L}(\boldsymbol{\Theta}) = \mathcal{L}(\mathbf{d}_{\text{AMI}}, \mathbf{d}_{\text{Pl}} | \boldsymbol{\Theta}, \mathcal{M}). \quad (8.1)$$

In this analysis we treat \mathbf{d}_{AMI} and \mathbf{d}_{Pl} as being independent (see Section 8.4.1 for justification), and since the Planck-predicted data only rely on the cluster parameters we can write

$$\mathcal{L}(\boldsymbol{\Theta}) = \mathcal{L}_{\text{AMI}}(\mathbf{d}_{\text{AMI}}|\boldsymbol{\Theta}, \mathcal{M}) \mathcal{L}_{\text{Pl}}(\mathbf{d}_{\text{Pl}}|\boldsymbol{\Theta}_{\text{cl}}, \mathcal{M}). \quad (8.2)$$

8.1.3.1 AMI likelihood function

The form of the AMI likelihood function used here is exactly the same as the one presented in Section 2.8. Note also that the AMI covariance matrix $\mathbf{C}_{\text{AMI},\nu,\nu'}$ is comprised of the same components as noted in Section 2.8 (which are described in Sections 2.5.2.1, 2.5.2.2, and 2.5.2.3), and recognised radio-sources are also treated the same way as previously described (Sections 2.5.1 and 2.7.1) for *AMI* data. For clarity I note that the predicted AMI data are denoted $\mathbf{d}_{\text{AMI},\nu'}^{\text{p}}(\boldsymbol{\Theta})$

8.1.3.2 PwS likelihood function

For a single source and given observing frequency, PwS treats the data observed by Planck as a superposition of background sky emission (including foreground emission and primordial CMB) \mathbf{b}_ν , instrumental noise \mathbf{n}_ν , and signal from the source \mathbf{s}_ν . The model for the predicted data vector is thus

$$\mathbf{d}_{\text{Pl},\nu}^{\text{p}}(\boldsymbol{\Theta}_{\text{cl}}) = \mathbf{s}_\nu(\boldsymbol{\Theta}_{\text{cl}}) + \mathbf{b}_\nu + \mathbf{n}_\nu. \quad (8.3)$$

PwS works with patches of sky sufficiently small such that it can be assumed the noise contributions are statistically homogeneous. In this limit it is more convenient to work in Fourier space, as the Fourier modes are uncorrelated assuming the noise contributions are Gaussian. This assumption is fair in the case of instrumental noise, but more questionable for \mathbf{b}_ν . The deviations from Gaussianity of \mathbf{b}_ν are discussed in Section 4.3 of the second PwS paper (Carvalho et al. 2012). Since PwS is a detection algorithm, it calculates the ratio of the likelihood of detecting a cluster parameterised by $\boldsymbol{\Theta}_{\text{cl}}$ and the likelihood of the data with no cluster signal ($\mathbf{s}_\nu(\boldsymbol{\Theta}_{\text{cl},0}) = 0$). Thus the log-likelihood ratio of the Fourier transformed quantities is

$$\begin{aligned} \ln \left[\frac{\mathcal{L}_{\text{Pl}}(\boldsymbol{\Theta}_{\text{cl}})}{\mathcal{L}_{\text{Pl}}(\boldsymbol{\Theta}_{\text{cl},0})} \right] &= \sum_{\nu,\nu'} \tilde{\mathbf{d}}_{\text{Pl},\nu}^{\text{p}}(\boldsymbol{\Theta}_{\text{cl}})^T \mathbf{C}_{\text{Pl},\nu,\nu'}^{-1} \tilde{\mathbf{d}}_{\text{Pl},\nu}^{\text{p}}(\boldsymbol{\Theta}_{\text{cl}}) \\ &\quad - \frac{1}{2} \tilde{\mathbf{d}}_{\text{Pl},\nu}^{\text{p}}(\boldsymbol{\Theta}_{\text{cl}})^T \mathbf{C}_{\text{Pl},\nu,\nu'}^{-1} \tilde{\mathbf{d}}_{\text{Pl},\nu}^{\text{p}}(\boldsymbol{\Theta}_{\text{cl}}), \end{aligned} \quad (8.4)$$

where tildes denote the Fourier transform of a quantity, and $\mathbf{C}_{\text{Pl},\nu,\nu'}$ is the covariance matrix of the data in Fourier space.

A full specification of the PwS likelihood ratio is given in [Carvalho, Rocha, & Hobson \(2009\)](#) and [Carvalho et al. \(2012\)](#).

8.2 Joint likelihood analysis hyperparameters

[Lahav et al. \(2000\)](#) and [Hobson, Bridle, & Lahav \(2002\)](#) (MH02 from here on) introduced a Bayesian method for determining the relative weighting of two or more *independent* datasets when analysed simultaneously, while [Ma & Berndsen \(2014\)](#) built on this work to develop a method which works for datasets *correlated* with one another. The basic idea behind the approach is to introduce additional *hyperparameters* α into the Bayesian inference problem. In other words we extend our parameter space to include not only the parameters of interest (Θ), but also the hyperparameters α . Thus we have

$$\mathcal{P}(\Theta, \alpha) = \frac{\mathcal{L}(\Theta, \alpha)\pi(\Theta, \alpha)}{\mathcal{Z}}, \quad (8.5)$$

where \mathcal{Z} is now given by

$$\mathcal{Z} = \int \int \mathcal{L}(\Theta, \alpha)\pi(\Theta, \alpha)d\alpha d\Theta. \quad (8.6)$$

Equations 8.5 and 8.6 tell us that to obtain the quantities of interest ($\mathcal{P}(\Theta)$ and \mathcal{Z}) we have to marginalise over the hyperparameters.

It is reasonable to assume that the parameters of the original problem and those affecting the weighting of each likelihood are independent of one another, so the priors can be written as

$$\pi(\Theta, \alpha) = \pi(\Theta)\pi(\alpha). \quad (8.7)$$

For more information on the typical priors used for α we refer the reader to Section 4.1 of MH02. To see how α are incorporated into $\mathcal{L}(\Theta)$ we consider two independent datasets, so that $\mathcal{L}(\Theta, \alpha)$ can be written as

$$\mathcal{L}(\Theta, \alpha) = \mathcal{L}_1(\Theta)^{\alpha_1} \mathcal{L}_2(\Theta)^{\alpha_2}. \quad (8.8)$$

Note we have chosen for the likelihoods to have such dependence on α so that if \mathcal{L}_1 and \mathcal{L}_2 are Gaussian (equation 2.49)

$$\mathcal{L}_1(\Theta)^{\alpha_1} = \frac{1}{Z_{N,1}} e^{-\frac{1}{2}\alpha_1 \chi_1^2} \quad (8.9)$$

(and similarly for \mathcal{L}_2), then we can write

$$\chi_{\text{joint}}^2 = \alpha_1 \chi_1^2 + \alpha_2 \chi_2^2, \quad (8.10)$$

where the χ^2 quantities are defined by equation 2.50. Thus α_1 and α_2 control the relative weighting of the goodness-of-fit metrics of the data.

8.2.1 Effects of likelihood hyperparameters

The effects of including α in Bayesian analysis are best illustrated through examples. Here I provide a very brief overview and quote the results of the toy model considered in MH02, to emphasise how the inclusion of hyperparameters affects evidence and posterior estimates of joint analyses compared with not using them (i.e. $\alpha_1 = \alpha_2 = 1$). I refer to the results obtained from including hyperparameters as HP and those from the ‘vanilla’ method as V. The toy problem consists of fitting a straight line through two (independent) sets of data points, and thus is a two-likelihood (one for each set of data), two-parameter problem of inferring the gradient (m) and intercept (c) of the line. The likelihood thus takes the form given by equation 8.8 with $\Theta = (m, c)$. Several versions of the problem are considered which vary in the standard deviations used for the likelihood functions and how the two datasets are generated.

8.2.1.1 Correct likelihood standard deviations and consistent datasets

The first example considered involves drawing points for each of the two datasets from the same distribution, namely Gaussian distributions with standard deviations $\sigma_{d,1} = \sigma_{d,2} = 0.1$, and mean values corresponding to the line with $m_d = c_d = 1$; the same deviations are used for the likelihood functions: $\sigma_{l,1} = \sigma_{l,2} = 0.1$. The resulting posterior distributions for the HP and V cases are shown in Figure 1 of MH02. The two methods recover the true values of m and c equally well, but the V run leads to a higher evidence estimate. This is to be expected for simple problems (for which the methods provide an equivalent fit to the data), as the added complexity of the HP method decreases the Bayesian evidence according to Occam’s razor.

8.2.1.2 Incorrect likelihood standard deviations and consistent datasets

The second example generates the two datasets in the same way, but the standard deviations used in the likelihoods are incorrect: $\sigma_{l,1} = 0.02$, $\sigma_{l,2} = 0.1$. Thus the predicted errors on the first dataset are much smaller than the true values used to generate it. In this case (Figure 2 of MH02) the V posterior underestimates the errors on m and c such that the true value is outside the 99% probability interval; whereas the HP method results in much larger error estimates, leading to the correct value being within the 95% confidence interval. This suggests that α_1 on average took relatively small values to accommodate for $\sigma_{l,1}$ being underestimated in the analysis. Furthermore the evidence ratio between the HP and V analyses is greatly in favour of the former, suggesting the data are fit sufficiently better by the HP model to overcome its additional complexity.

8.2.1.3 Correct likelihood standard deviations with inconsistent datasets

The final scenario considered for the toy problem in MH02 involves sampling the two datasets from different distributions i.e. sampling two sets of data which represent different lines. This means that there are two ‘true’ values for m and c corresponding to each dataset, and so a good inference of the data should produce a bimodal distribution with peaks at these values. They first test this out by sampling one dataset from a distribution corresponding to $m_{d,1} = c_{d,1} = 1$ and the other from $m_{d,2} = 0, c_{d,2} = 1.5$. The resultant posterior distributions shown in Figure 3 of MH02 show that the V distribution is unimodal and does not contain either of the true values within its 99% probability contours, while the HP distribution is bimodal with the peaks occurring close to the true values.

They repeat this analysis but sample from distributions corresponding to $m_{d,1} = c_{d,1} = 1$ and $m_{d,2} = 0.7, c_{d,2} = 0.7$ and find again that the V posterior distribution is unimodal and centred far from the true values, while HP results in a bimodal distribution with peaks in the vicinity of the true values (but not as close as in the previous case). The evidence ratio between the V and HP analyses suggests the latter is a more suitable model in both cases.

8.2.2 Incorporating the likelihood hyperparameters into AMI-Planck analysis

From the examples reviewed above, it is clear that inclusion of the likelihood hyperparameters leads to inferences more representative of the data in the cases that the errors in the analysis are underestimated or the datasets are systematically different from one another. Thus it makes sense to include them in analyses of data obtained from telescopes operating at different frequencies and angular scales and that are subject to different systematic errors.

However the log-ratio given by equation 8.4 is not a probability density due to the fact that it is missing a normalisation factor proportional to $\tilde{\mathbf{d}}_{\text{Pl},\nu}(\boldsymbol{\Theta}_{\text{cl}})^T \mathbf{C}_{\text{Pl},\nu,\nu}^{-1} \tilde{\mathbf{d}}_{\text{Pl},\nu}(\boldsymbol{\Theta}_{\text{cl}}) \equiv C$. Inclusion of the likelihood hyperparameters means that the normalisation factor of a likelihood function is dependent on α , since it is marginalised over to obtain $\mathcal{L}(\boldsymbol{\Theta})$, so $C \equiv C(\alpha) = C^\alpha$.

To test whether the inclusion of $C(\alpha)$ was strictly needed for the hyperparameter methodology, I replicated the toy model example considered in Section 6.1 of MP02 (reviewed in Section 8.2.1.1), ran the analysis using the ‘full’ hyperparameter likelihood functions (equation 8.9) and also conducted the analysis using hyperparameter likelihood ratios (i.e. using the likelihoods given by equation 8.9 but excluding the C^α factors present in the χ^2 s). The full likelihood analysis produced a a posterior distribution similar to the one obtained in MP02 (left plot of Figure 8.1) while the likelihood ratio analysis failed to produce posterior samples. The reason

why likelihood ratios are incompatible with the hyperparameter method is shown graphically in the right plot of Figure 8.1. From this plot it is clear that C^α dictates the shape of the likelihood function as well as its peak. For example around $\chi^2 = 0$ the normalised $\alpha = 2$ curve is above the $\alpha = 1$, while the un-normalised $\alpha = 2$ curve is below it. This inconsistency generalises to all χ^2 and α values and thus one cannot reliably evaluate the effect of α on the analysis without knowing C^α and hence the hyperparameters cannot be used with likelihood ratios such as the one used by PwS. As a result, the two likelihoods had to be weighted equally (i.e. I set $\alpha_1 = \alpha_2 = 1$).

8.3 Cluster models

As described in Section 2.2, a radio interferometer measure signal that is the Fourier transform of a quantity proportional to the Comptonisation parameter y . Similarly the Planck satellite is also sensitive SZ effect and thus measures a signal $\propto y$. Thus the cluster models introduced previously in this thesis which calculate a map of y can be used to calculate $\mathbf{d}_{\text{AMI},\nu}^p(\boldsymbol{\Theta})$ and $\mathbf{d}_{\text{Pl},\nu}^p(\boldsymbol{\Theta}_{\text{cl}})^T$.

8.3.1 Observational model

The observational model used in this Chapter (OM III) is the same as the ones introduced in Chapter 4 other than the priors it uses.

Here I assign non-informative, independent priors to Y_{tot} and θ_p (see Table 8.1), to get a better idea of how much the joint likelihood function can constrain the parameters. The priors used for a and b vary throughout the analysis (Table 8.1); they are either fixed at some specific value (as was the case in OM I and OM II) or allowed to vary uniformly.

8.3.2 Physical models

The physical models used here are the same as the ones presented in Chapters 2 and 4 (PM I and PM II) i.e. they model the cluster dark matter content using NFW and Einasto profiles respectively. The prior distributions the PMS are also given in Table 8.1.

All three models can be used to calculate the profile of $P_e(r)$ which can be used to produce a y map using equation 2.16.

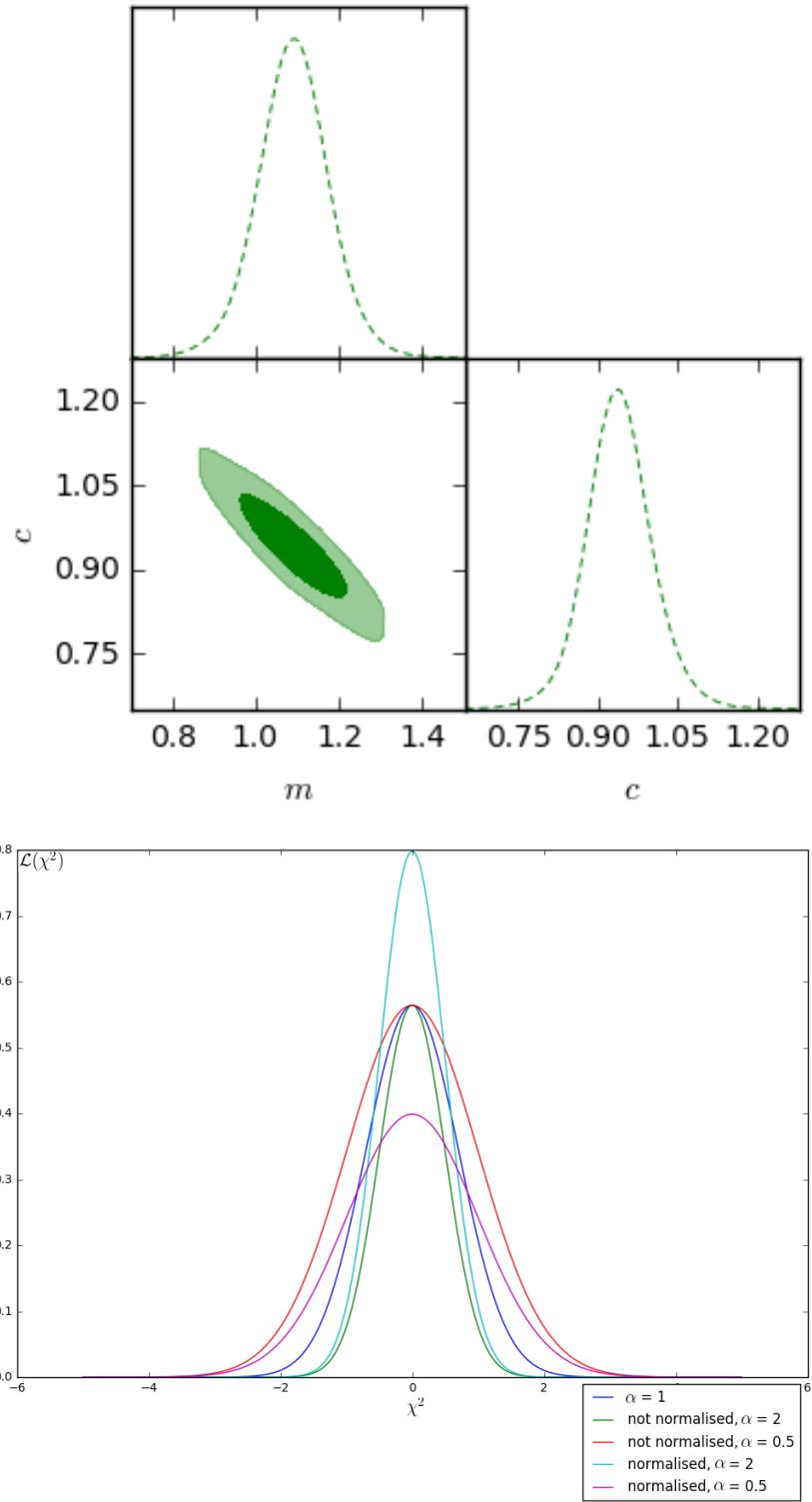


Figure 8.1: Top: Two-dimensional posterior distribution obtained from application of likelihood hyperparameter method on toy model considered in Section 6.1 of MP02. m and c are the gradient and intercept parameters of the toy model respectively. These results were obtained using the likelihood functions given by equation 8.9. Bottom: Gaussian likelihood for a range of α values, including (normalised) or excluding (un-normalised) C^α . Note for the $\alpha = 1$ case the normalisation doesn't depend on α since $C(\alpha = 1) = C$.

Parameter	Model(s) featured in	Prior distribution(s)
x_c	OM III, PM I, and PM II	$\mathcal{N}(0'', 60'')$
y_c	OM III, PM I, and PM II	$\mathcal{N}(0'', 60'')$
Y_{tot}	OM III	$\mathcal{U}[0.00 \text{ arcmin}^2, 0.02 \text{ arcmin}^2]$
θ_p	OM III	$\mathcal{U}[1.3', 15']$
z	PM I and PM II	$\delta(z)$
$M(r_{200})$	PM I and PM II	$\mathcal{U}[\log(0.5 \times 10^{14} M_{\text{Sun}}), \log(50 \times 10^{14} M_{\text{Sun}})]$
$f_{\text{gas}}(r_{200})$	PM I and PM II	$\mathcal{N}(0.12, 0.02)$
α_{Ein}	PM II	$\delta(\alpha_{\text{Ein}})$ or $\mathcal{U}[0.05, 0.3]$
a	OM III, PM I, and PM II	$\delta(a)$ or $\mathcal{U}[0.3, 3.5]$
b	OM III, PM I, and PM II	$\delta(b)$ or $\mathcal{U}[3.5, 7.5]$
c	OM III, PM I, and PM II	$\delta(c)$

Table 8.1: Cluster parameter prior distributions. \mathcal{N} denotes a normal distribution parameterised by its mean and standard deviation, \mathcal{U} denotes a uniform distribution, and δ is a Dirac delta function. In the cases where the latter is used, the values used for the function’s argument will be stated when the analyses are carried out.

8.4 Cluster simulations

The cluster simulations were generated using the in-house package PROFILE (used in Chapters 3 and 5). For all simulations the y map of a single cluster is generated with either OM III, PM I, or PM II, and primordial CMB noise is sampled from an empirical distribution (Hinshaw et al. 2013) and added at random positions to the data. At this point the data are duplicated so that additional noise contributions specific to each telescope can be added.

For the AMI simulated data, confusion noise is added as described in Section 3.7 using the 10C source counts given in Davies et al. (2011). Instrumental noise with an RMS value of 0.379 Jy per channel per baseline per second is also added.

For the Planck simulated data, foreground emission and instrumental noise are added. For more information on the Planck simulations, see Perrott et al. (2019). Finally, the data are ‘observed’ by AMI and Planck separately to generate $\mathbf{d}_{\text{AMI},\nu}$ and $\mathbf{d}_{\text{Pl},\nu}$.

8.4.1 Testing the independence of the AMI and Planck datasets

In Section 8.1.3 we made the assumption that $\mathbf{d}_{\text{AMI},\nu}^p$ and $\mathbf{d}_{\text{Pl},\nu}^p$ are not correlated with each other, so that the likelihoods for the two datasets can be separated. The instrumental noises associated with each telescope can safely be assumed to be independent. Due to the telescopes operating at different angular scales and frequencies, the confusion noise present in AMI data

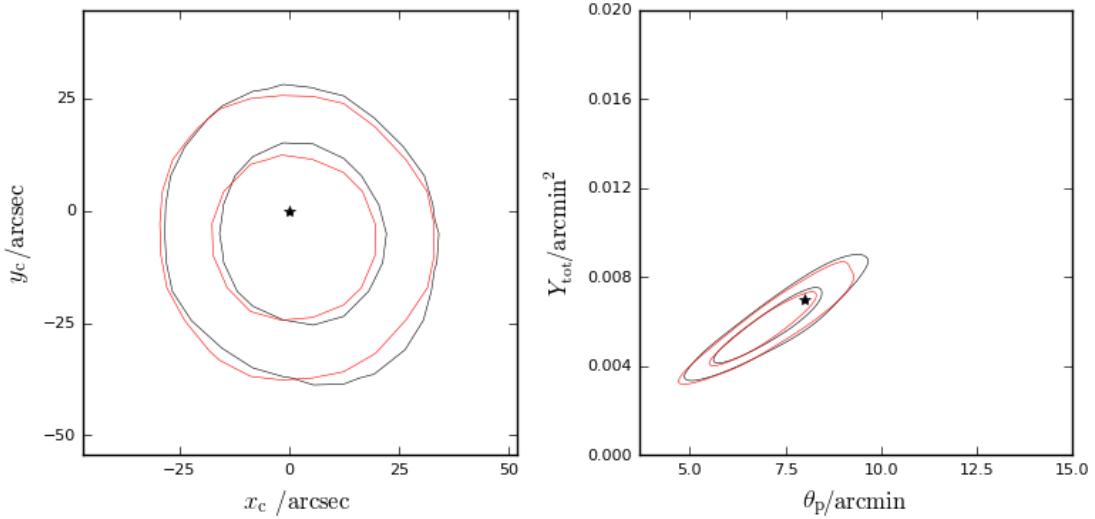


Figure 8.2: Two-dimensional marginalised $x_c - y_c$ and $Y_{\text{tot}} - \theta_p$ posterior distributions for a high SNR (see Section 8.5.1) cluster simulation generated using OM III. The red contours correspond to the posterior distribution associated with the AMI and Planck datasets which had different CMB realisations to each other, while the black ones correspond to datasets generated with the same realisation. The star symbols indicate the values input when generating the simulations.

and foreground emission present in Planck data are assumed to be independent of one another. A similar argument can be applied for primordial CMB noise, nevertheless I carried out a simple test to see if this is the case. For a given set of cluster parameters, I ran the joint analysis on Planck and AMI datasets which had different CMB realisations to one another. I found that the resultant parameter constraints were not affected by this when compared with the results obtained using AMI and Planck data which had the same CMB realisations as one another (Figure 8.2). I thus concluded that the covariance between the datasets was negligible.

8.5 Cluster simulation results

In the following analysis I generate cluster simulations for different noise realisations and cluster parameter values (and models). I apply the joint analysis to these simulated clusters, and compare results with analyses which use (the same) AMI or Planck data alone. Note that for all examples considered, the model used to *simulate* the cluster was also used to *analyse* the data.

Parameter	Simulation input	
	low SNR	high SNR
x_c		0 arcsec
y_c		0 arcsec
Y_{tot}	0.001 arcmin ²	0.007 arcmin ²
θ_p	2 arcmin	8 arcmin
a		1.051
b		5.4905
c		0.3081
z		0.17

Table 8.2: Cluster simulation inputs using an observational model and the ‘universal’ GNFW shape parameters calculated in Arnaud et al. (2010). Although z isn’t an input parameter for observational models, it is still required to generate simulations of clusters.

8.5.1 Observational model with ‘universal’ shape parameters

I generate simulations using OM III, with GNFW shape parameter values $a = 1.0510$, $b = 5.4905$, and $c = 0.3081$ (i.e. the same ones used in Chapter 5). As shown in Table 8.2 I consider a ‘low’ and a ‘high’ signal-to-noise ratio (SNR) cluster, which correspond to input values of $Y_{\text{tot}} = 0.001 \text{ arcmin}^2$ and $\theta_p = 2 \text{ arcmin}$ and $Y_{\text{tot}} = 0.007 \text{ arcmin}^2$ and $\theta_p = 8 \text{ arcmin}$ respectively. I generate 10 simulations for each of these clusters, each of which has a different noise realisation. I then analyse these simulations using the priors given in Table 8.1, with delta priors on a , b and c centred on their ‘true’ values (the ones used as inputs to the simulations), and plot the resulting posterior distributions using GETDIST.

8.5.1.1 Low SNR simulation analyses

Figure 8.3 shows the two-dimensional marginalised $x_c - y_c$ and $Y_{\text{tot}} - \theta_p$ posterior distributions of the joint, AMI-only, and Planck-only analyses of the low SNR cluster. Note that each plot contains the posteriors of the 10 different simulations, each of which is represented by a contour (68% confidence interval). Looking at the AMI data only analyses, in two of the simulations the correct values for x_c and y_c are not recovered within a 68% confidence interval. The plot shows that the constraints in $Y_{\text{tot}} - \theta_p$ are generally tight, but three contours do not encompass the input value.

The Planck-only analyses generally recover the correct values for x_c and y_c but the contours are much wider. There is a large degeneracy in θ_p , suggesting that in this case Planck cannot

constrain the geometric size of the clusters very well.

The joint analysis shows similar results to the AMI-only analyses for the $x_c - y_c$ distributions, but the constraints on Y_{tot} and θ_p are very tight (sharper distributions than in the case of AMI-only), which suggests that even though the Planck data in isolation was degenerate, when combined with AMI it can help infer the correct size of a cluster.

8.5.1.2 High SNR simulation analyses

Figure 8.4 shows the contours of the posterior distributions obtained from the high SNR simulations.

The AMI-only $x_c - y_c$ posterior contours are similar to the low SNR case, but are generally more offset from the correct value in this instance. The $Y_{\text{tot}} - \theta_p$ posteriors show large degeneracies along the line of changing Y_{tot} and θ_p (i.e. a large positive covariance between the two parameters). The Planck-only data results show tighter constraints on x_c and y_c relative to the low SNR simulations, but still wider than the other two analysis methods. The $Y_{\text{tot}} - \theta_p$ posteriors show that Planck arguably does a better job than AMI in recovering the true values, as the contours are generally tighter in the former case, and both analyses give a similar number of distributions where the correct value lies in the proximity of the contours.

The joint analysis shows that the cluster offset inferences are driven almost entirely by the AMI data, as they strongly resemble the results of the AMI runs. In contrast the $Y_{\text{tot}} - \theta_p$ posteriors suggest Planck data is dominating the inferences, and that the joint data distributions provide the tightest constraints on $Y_{\text{tot}} - \theta_p$ estimates. However, five of these distributions fail to recover the true values within their 68% confidence intervals.

8.5.1.3 Variable shape parameter analysis

I next consider the same simulations described in Section 8.5.1, but allowing the GNFW shape parameters a and b to vary in the *analysis*. I thus assign the uniform priors stated in Table 8.2 to a and b . I note that throughout the analysis I found that the cluster model used to analyse the data did not affect the posterior constraints on x_c and y_c , and so I do not discuss them in the subsequent analyses.

Figure 8.5 shows two-dimensional posterior distributions of pairs of the parameters: Y_{tot} , θ_p , a , and b , resultant from six low SNR simulations. The $\theta_p - a$ posteriors show that the AMI-only and Planck-only analyses fail to produce good constraints, as the former has a large degeneracy in θ_p which misses the simulation input while the latter is almost completely uninformative

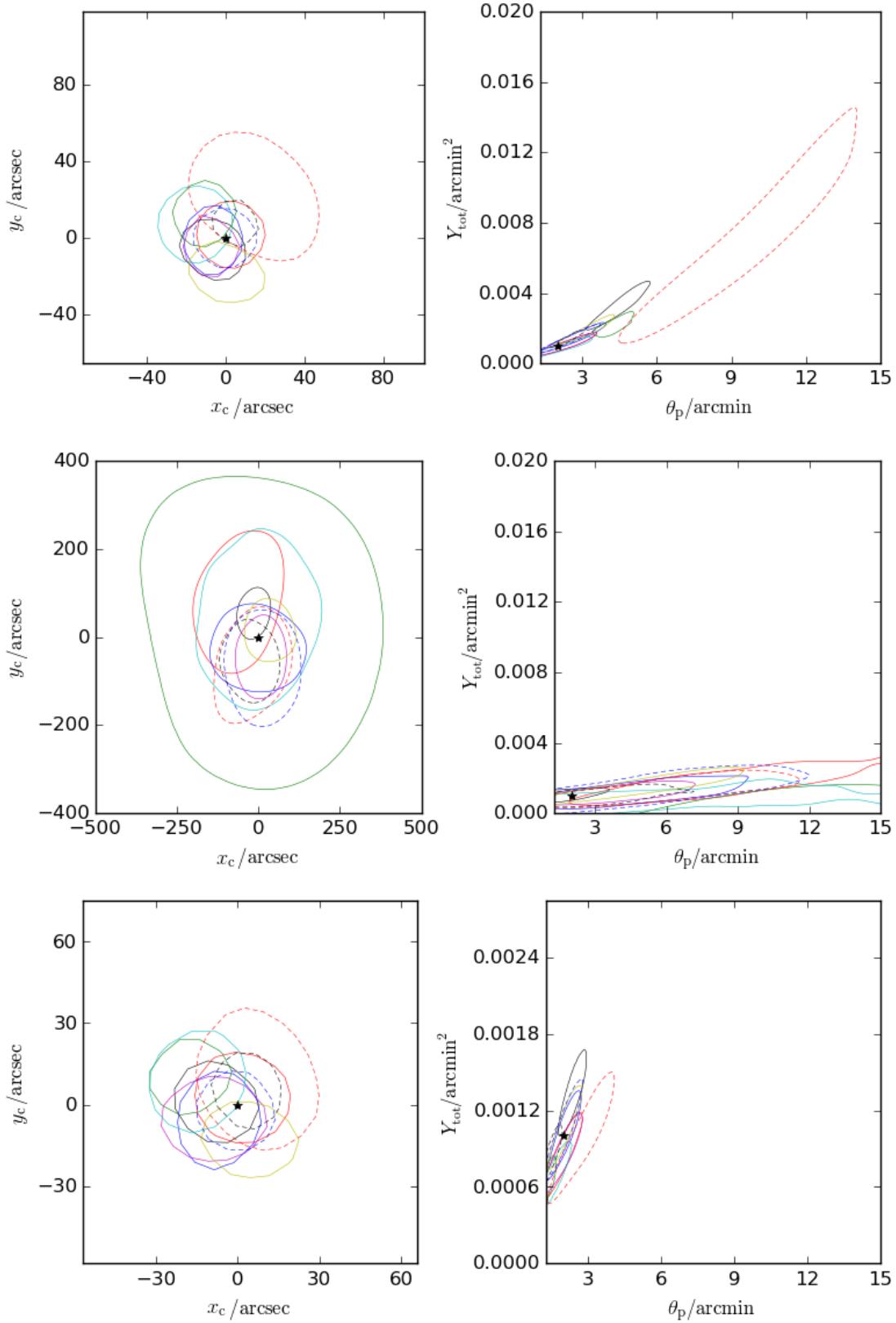


Figure 8.3: Two-dimensional marginalised $x_c - y_c$ and $Y_{\text{tot}} - \theta_p$ posterior distributions for the 10 OM III low SNR cluster simulations obtained from: AMI data (top row), Planck data (middle row), and AMI and Planck data combined (bottom row). The contours in each plot represent the 68% confidence intervals of the separate posterior distributions obtained from each of the 10 simulations. The star symbols indicate the values input when generating the simulations.

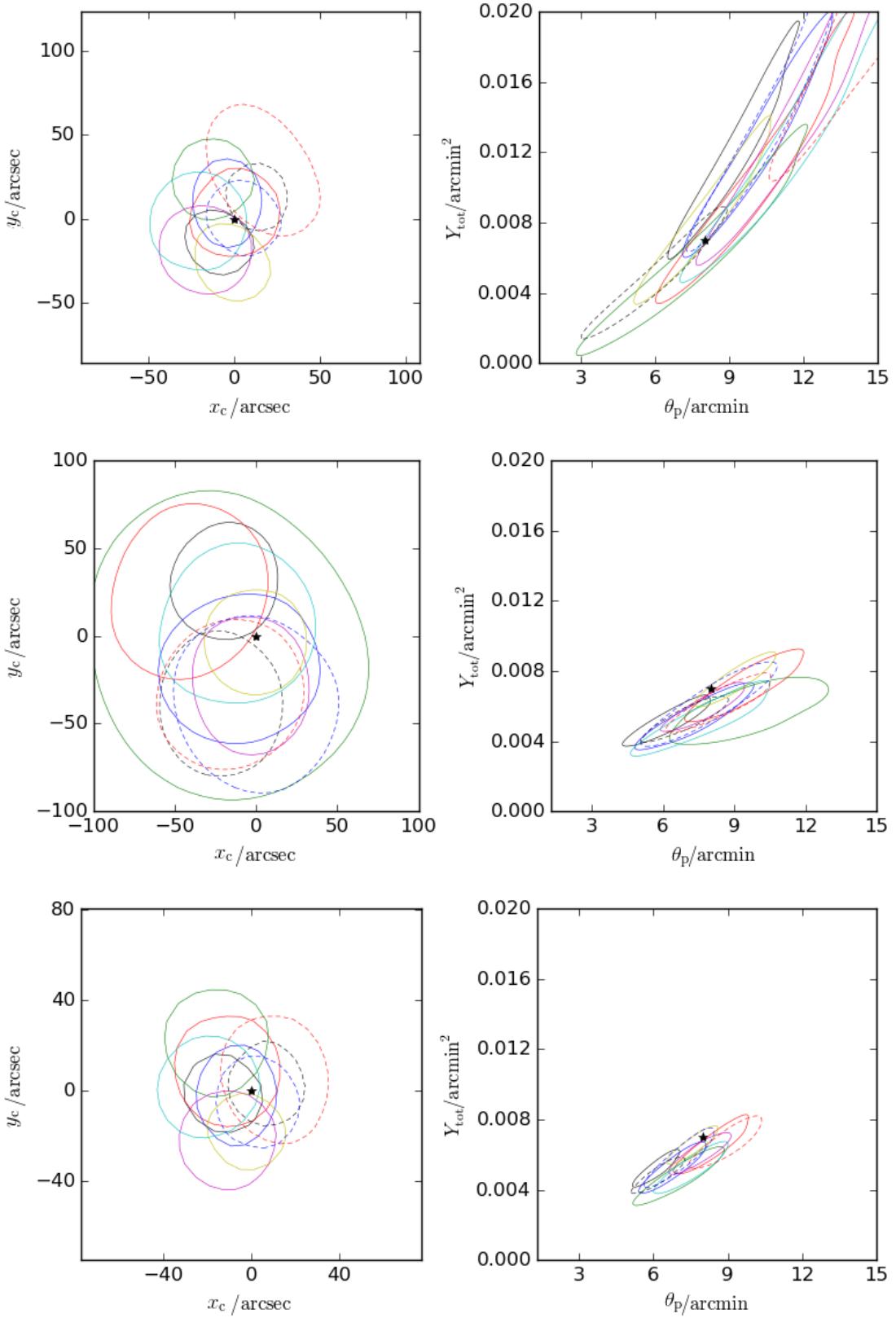


Figure 8.4: Two-dimensional marginalised $x_c - y_c$ and $Y_{\text{tot}} - \theta_p$ posterior distributions for the 10 OM III high SNR cluster simulations. The Figure layout is as described in Figure 8.3.

(resembles the prior). The joint analysis leads to results that encompass the true value within the 68% contour, albeit with large degeneracies in a at low θ_p (where the true θ_p value lies) and in θ_p at low a . The $Y_{\text{tot}} - a$ posterior plots shows similar results for AMI, Planck recovers Y_{tot} well but has a large degeneracy in a . The joint analysis gives similar results to Planck-only, but with a tighter constraint on Y_{tot} (as was the case in the fixed a and b low SNR analyses). Posteriors in the $Y_{\text{tot}} - b$ plane show similar results, but in this case the joint analysis produces contours which are less degenerate in b than the Planck-only results. The $\theta_p - b$ plots show that all three analyses fail to produce informative (well constrained) posteriors.

While the joint analysis tends to show degeneracy in a and b , it does produce superior constraints on Y_{tot} and θ_p relative to the single data analyses for marginalised posteriors considered here.

Figure 8.6 shows the two-dimensional posterior distributions for the high SNR simulations, in which case the AMI posterior distributions recover a relatively well (with the one clear exception). The AMI posteriors for b are quite wide but generally peak around the input value of b .

The Planck-only distributions also show some improvement over the low SNR case.

The joint analysis gives slightly worse results for a than the AMI-only case (though the exceptionally bad AMI distribution improves), while the posteriors for b arguably improve in the joint case for five of the six simulations.

8.5.2 Cluster simulations using physical models

I repeat the simulation procedure described in Section 8.5.1, but this time using PM I and PM II in the cluster simulation and analysis. Table 8.3 shows the input parameters used for PM simulations; the low SNR simulations have $M(r_{200}) = 5 \times 10^{14} M_{\text{Sun}}$ while the high SNR use $10 \times 10^{14} M_{\text{Sun}}$.

8.5.2.1 PM I low SNR posteriors

The one-dimensional posterior distributions for x_c , y_c , $M(r_{200})$, and $f_{\text{gas}}(r_{200})$ for the 10 low SNR simulations are shown in Figure 8.7. Four of the AMI mass posteriors replicate the shape of the prior distribution (which has a $1/M(r_{200})$ dependence in linear space), indicating that the likelihood is negligible for these analyses. $f_{\text{gas}}(r_{200})$ is recovered very well by AMI for all ten simulations (and also takes the same shape as the prior).

In the case of the Planck mass estimates, the modes of the posteriors overestimate the input value by a factor of at least two. The same statistic slightly underestimates $f_{\text{gas}}(r_{200})$ in some cases,

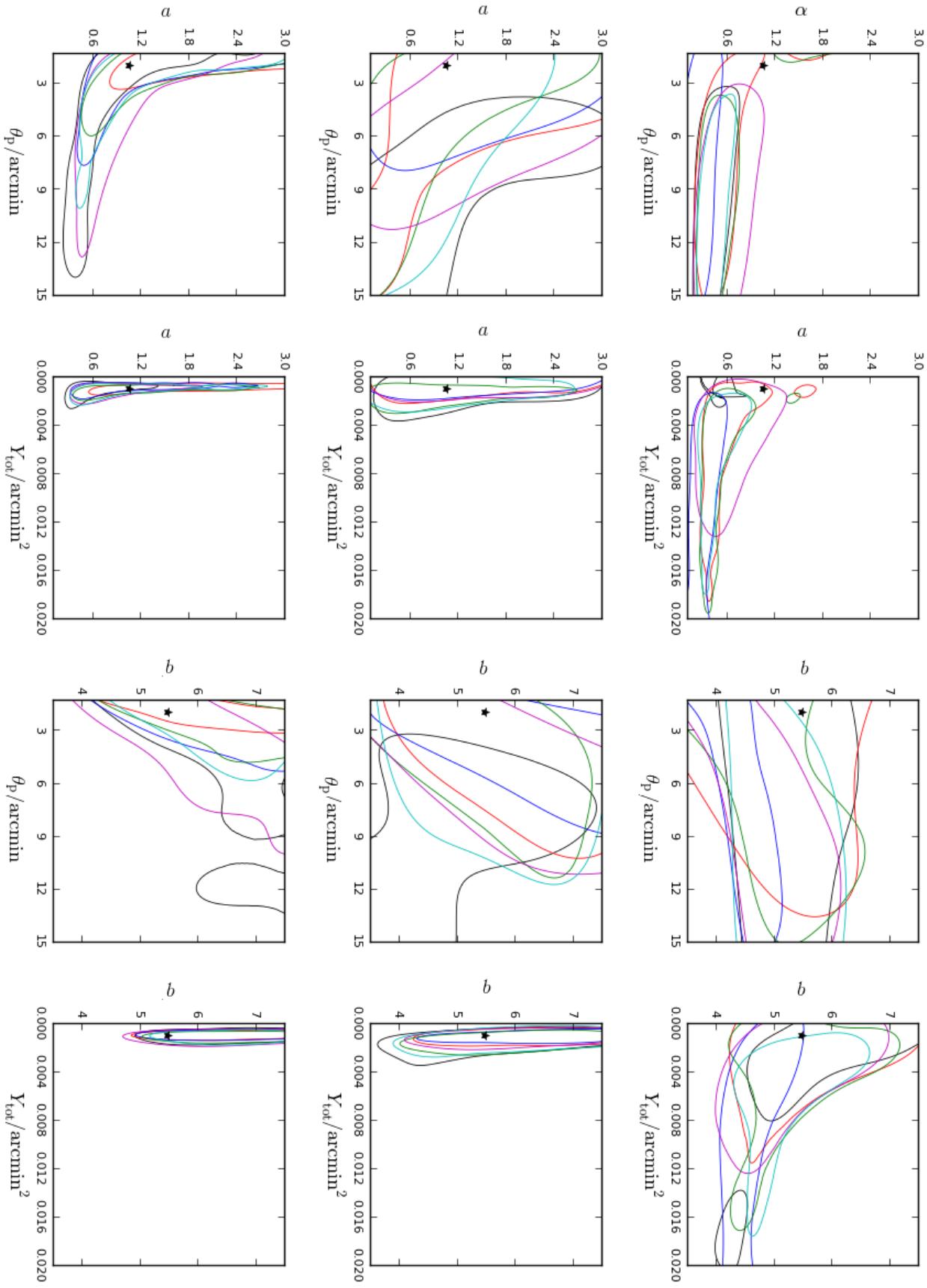


Figure 8.5: Two-dimensional marginalised posterior distributions for six OM III low SNR cluster simulations obtained from: AMI data (top row), Planck data (middle row), and AMI and Planck data combined (bottom row). In the Bayesian analysis of the data, the values of the GNFW shape parameters a and b were allowed to vary (had uniform priors). The black stars indicate the values input when generating the simulations.

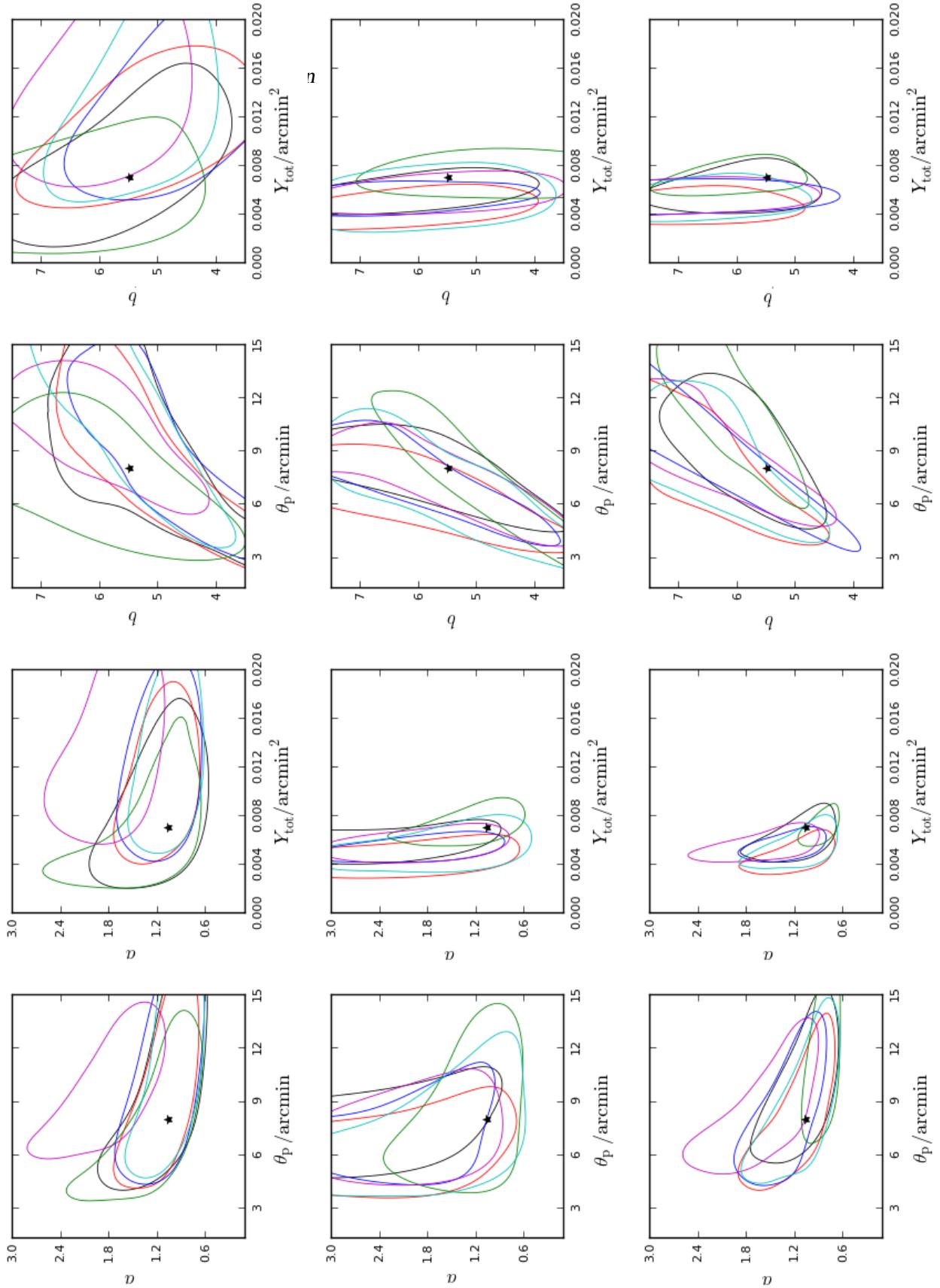


Figure 8.6: Two-dimensional marginalised posterior distributions for six OM III high SNR cluster simulations where a and b were allowed to vary in the analysis. The plot layout is as described in Figure 8.5.

Parameter	Simulation input			
	PM I low SNR	PM I high SNR	PM II low SNR	PM II high SNR
x_c			0 arcsec	
y_c			0 arcsec	
$M(r_{200})$	$5 \times 10^{14} M_{\text{Sun}}$	$10 \times 10^{14} M_{\text{Sun}}$	$5 \times 10^{14} M_{\text{Sun}}$	$10 \times 10^{14} M_{\text{Sun}}$
$f_{\text{gas}}(r_{200})$			0.12	
α_{Ein}	—			0.2
a		1.051		
b		5.4905		
c		0.3081		
z		0.17		

Table 8.3: Cluster simulation inputs for PM I and PM II. The cluster centre, GNFW shape parameters, and redshift inputs are the same for all four models. The Einasto shape parameter is only an input for PM II.

but not to the same degree as the $M(r_{200})$ values are overestimated.

The combined data also overestimates $M(r_{200})$, with the modes ranging between $\approx 1.75 - 4$ times the true values. What is also striking is the values of the modes of the $f_{\text{gas}}(r_{200})$ posteriors, which in some cases (which correspond to the larger mass estimates) occur around $f_{\text{gas}}(r_{200}) \approx 0.8$. The overestimation of mass and underestimation of $f_{\text{gas}}(r_{200})$ suggests that in the joint analysis, it is the composition of the clusters which have been incorrectly inferred, whilst in the Planck-only case it appears that the physical size of the clusters is overestimated.

8.5.2.2 PM I high SNR posteriors

For the high SNR cluster simulations (Figure 8.8) the AMI mass estimates on average peak on the true mass value. The Planck mass modal values generally underestimate the input mass, which is in stark contrast to the low SNR case where they massively overestimated it. The Planck estimates of $f_{\text{gas}}(r_{200})$ are extremely accurate, which again suggests that it is the size rather than the composition of the cluster that Planck has difficulty with. The joint estimates perform similarly well to the separate analyses.

8.5.2.3 PM II cluster simulations

Cluster simulations were generated with the PM II setting $\alpha_{\text{Ein}} = 0.2$. Note this value for α_{Ein} corresponds to a profile similar to that given by the NFW profile (as discussed in [Dutton & Macciò 2014](#) and Chapter 5). The clusters were analysed with a uniform prior on α_{Ein} (given

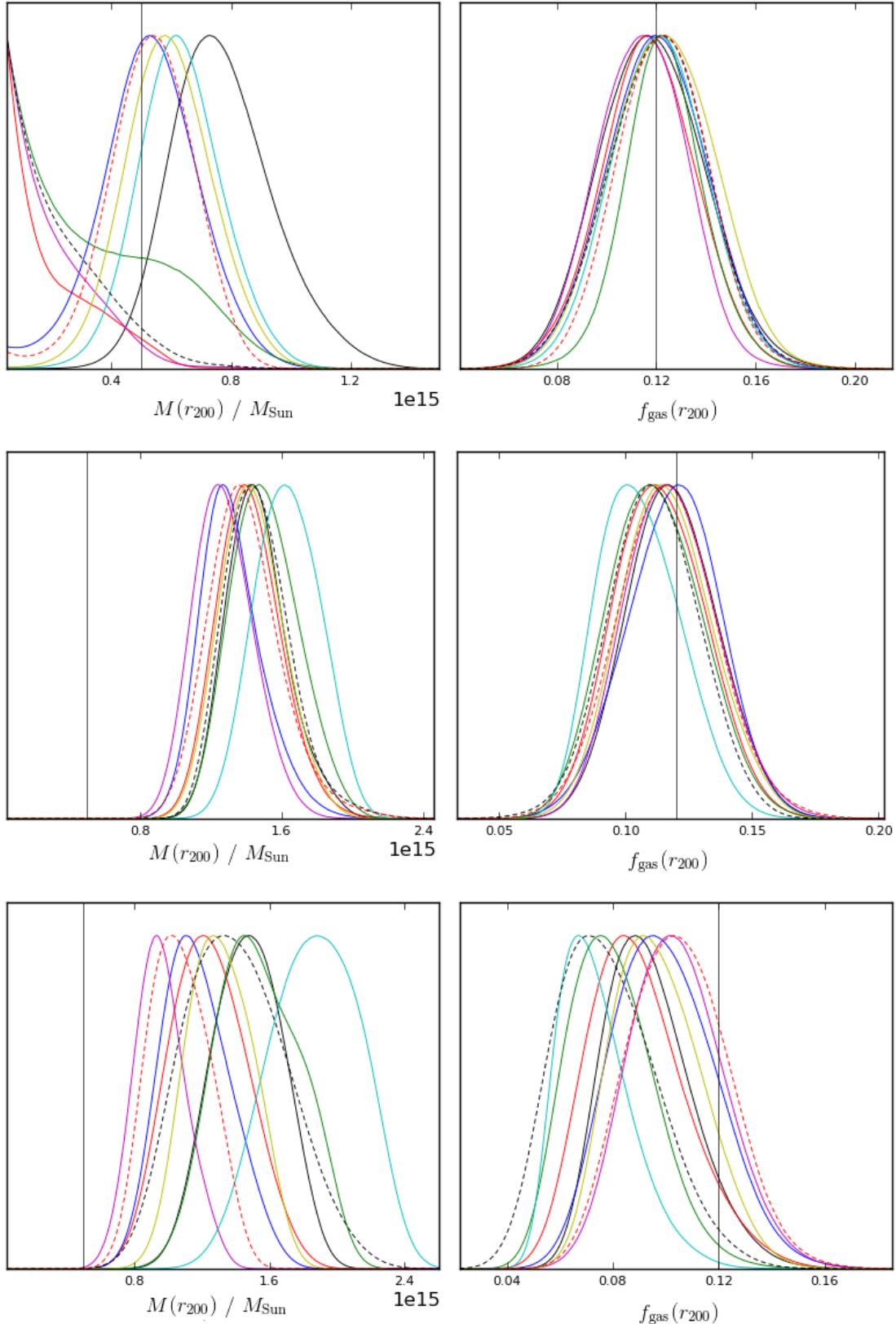


Figure 8.7: One-dimensional marginalised posterior distributions for the 10 PM I low SNR cluster simulations obtained from: AMI data (top row), Planck data (middle row), and AMI and Planck data combined (bottom row). The black vertical lines indicate the values input when generating the simulations.

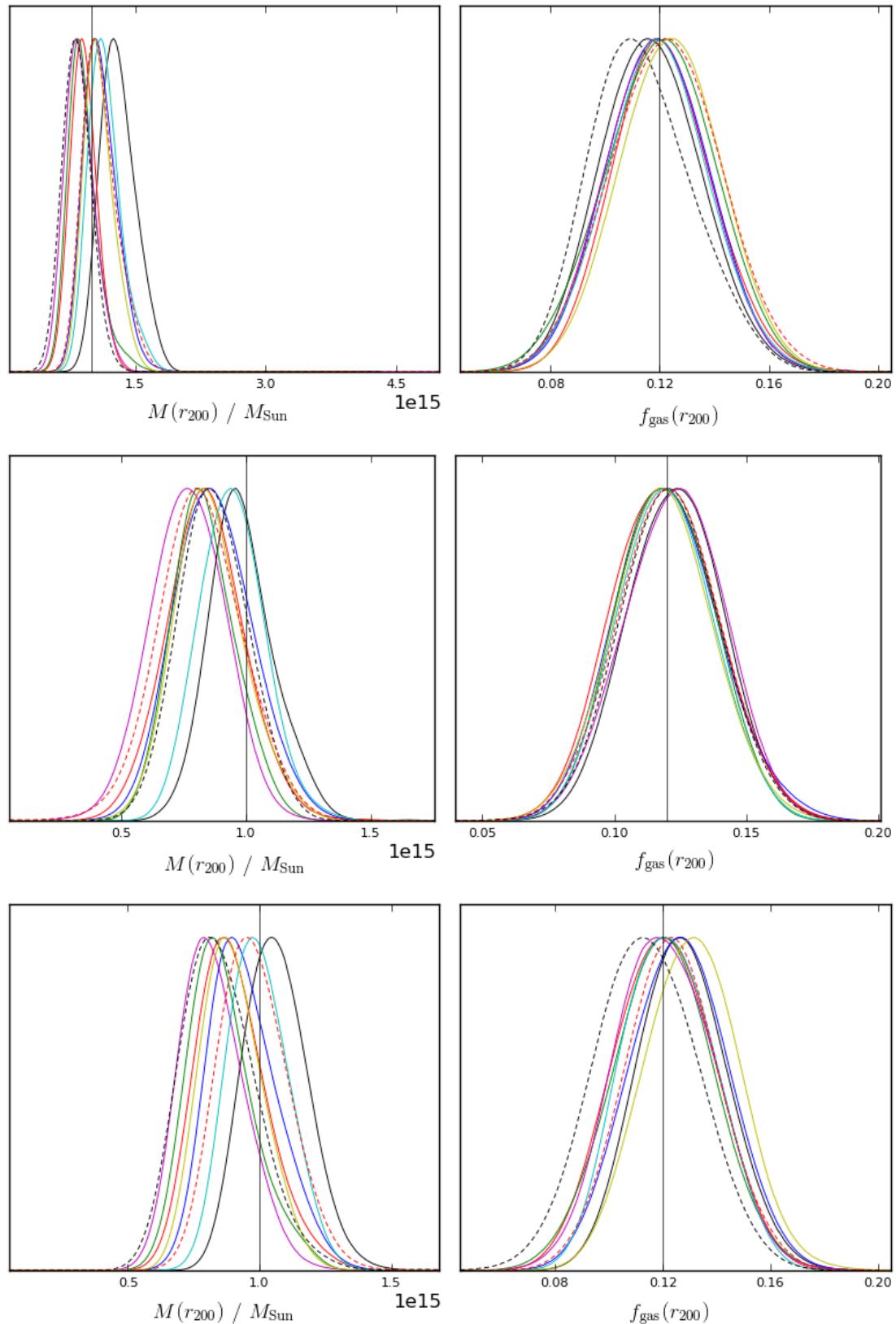


Figure 8.8: One-dimensional marginalised posterior distributions for the 10 PM I high SNR cluster simulations. The plots are laid out as described in Figure 8.7.

in Table 8.1), but GETDIST failed to plot distributions from the resultant posterior samples. This suggests that the marginalised α_{Ein} posterior distributions are not ‘compatible’ with Gaussian kernel density estimation techniques used in the program. Nevertheless GETDIST still produced posterior distributions of other parameters (by marginalising over α_{Ein}), and gave results similar to the PM I simulations. The posteriors obtained from analysis of the high SNR PM II clusters are shown in Figure 8.9. Likewise Bayesian analysis of the cluster simulations with a delta prior on α_{Ein} resulted in posterior distributions similar to those obtained from PM I simulations and analysis.

Note that the overestimation of cluster parameters has been resolved in [Perrott et al. \(2019\)](#) by understanding the Planck simulations better (and correcting a couple of associated bugs), but the paper focuses on observational models rather than physical.

8.6 Application of joint analysis to real cluster data

8.6.1 Mass estimates of cluster PSZ2G063.80+11.42

I apply the joint analysis to a cluster featured in PSZ2 (PSZ2G063.80+11.42) and the 54 cluster sample analysed in Chapters 3 and 4. Note that in these Chapters slightly different values for a , b , and c were used, which were derived in [Arnaud et al. \(2010\)](#) for the standard self-similar case (Appendix B of Arnaud et al.). It was shown in [Olamaie, Hobson, & Grainge \(2013\)](#) that PM I is not affected by which of these two sets of parameters is used. In Chapter 3 I calculated the AMI mass estimate to be $M_{\text{AMI}}(r_{500}) = (3.37 \pm 0.76) \times 10^{14} M_{\text{Sun}}$ and the PwS mass estimate (using the slicing function methodology introduced in PSZ2 and detailed in Section 3.4) to be $M_{\text{Pl, slice}}(r_{500}) = (6.41 \pm 0.57) \times 10^{14} M_{\text{Sun}}$. I chose to run the joint analysis on this cluster due to the fact that its AMI and Planck masses were quite discrepant, despite the AMI radio-source environment not appearing to be problematic on the map of the observation. The cluster redshift is taken from PSZ2 as $z = 0.426$, and the coordinates of the Planck patch centre are within 0.01 arcmin of the AMI SA pointing centre of the observation.

I run the joint analysis with PM I using the priors given in Table 8.1 (assigning delta priors to the GNFW shape parameters). The marginalised posterior distribution for $M_{\text{Joint, PMI}}(r_{500})$ (Figure 8.10) gives a mean estimate of $M_{\text{Joint, PMI}}(r_{500}) = (5.74 \pm 0.70) \times 10^{14} M_{\text{Sun}}$. Hence the joint analysis gives a value within one combined standard deviation of the value obtained from Planck data using the PSZ2 slicing function methodology, and within three combined standard deviations of the value obtained from AMI data alone.

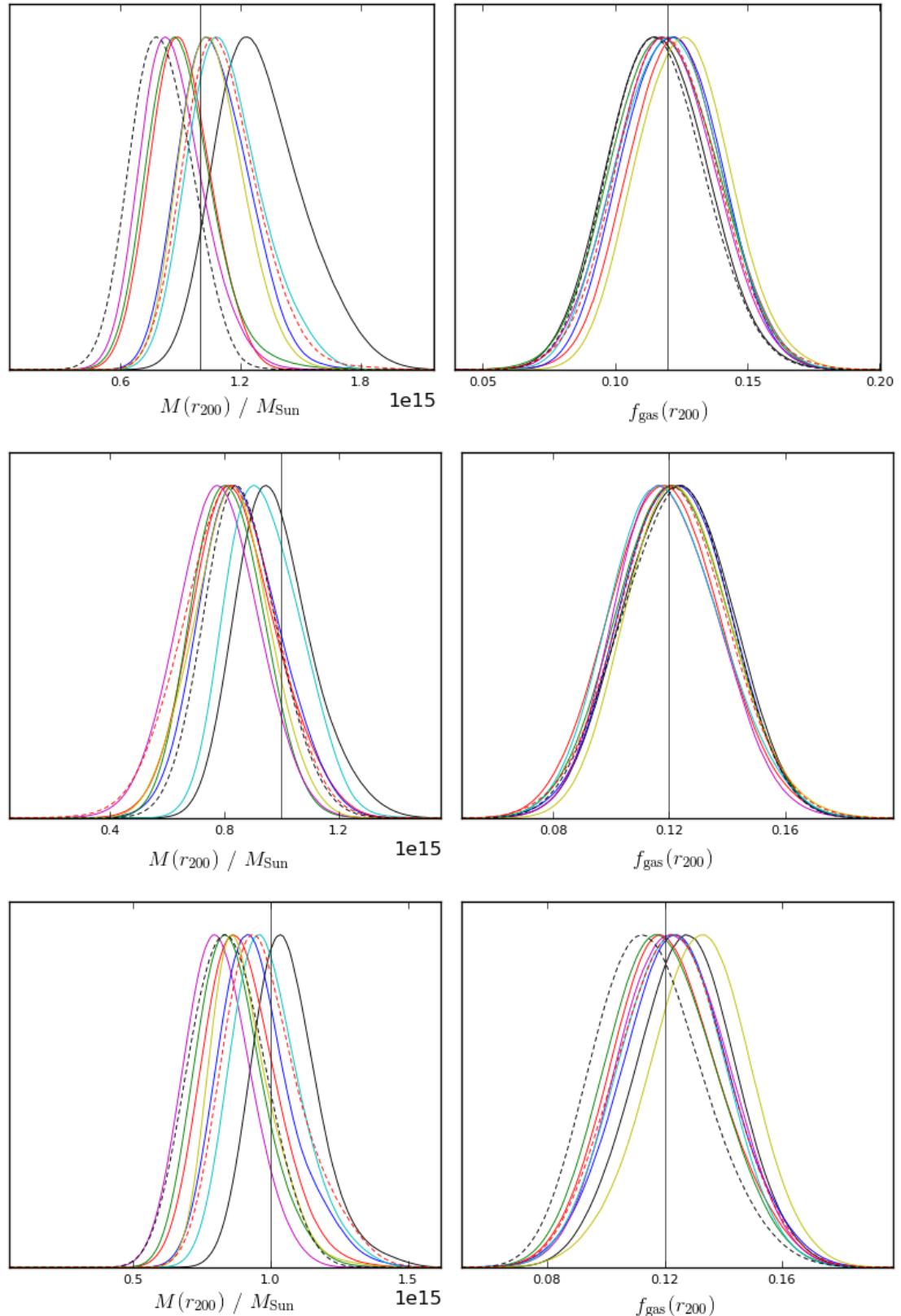


Figure 8.9: One-dimensional marginalised posterior distributions for the 10 PM II high SNR cluster simulations obtained by marginalising over α_{Ein} . The plots are laid out as described in Figure 8.7.

For further comparison I run the Planck-only analysis for the same cluster using the same model, and find that $M_{\text{Pl, PMI}}(r_{500}) = (6.98 \pm 1.02) \times 10^{14} M_{\text{Sun}}$. For clarity I note that $M_{\text{Pl, slice}}(r_{500})$ and $M_{\text{Pl, PMI}}(r_{500})$ are obtained from the same data using the same PwS algorithm, but the former uses the scaling relations and slicing function methodology to obtain a mass estimate, whereas the latter uses PM I in the Bayesian analysis to directly infer mass posterior distributions.

In Chapter 3 I found that PSZ2 mass estimates were generally higher than those obtained by AMI. In this Chapter the low SNR PM I simulations show similar results, as Planck data analyses gives large overestimates of the true values, whereas AMI underestimates it on average. The real data analysed here suggest the same – although we do not know the ‘true’ mass value in this case. The fact that both estimates from Planck data are relatively high suggests the data themselves are causing this, not the model being applied. I note however that this is based on just one real cluster, and that the Planck-only analysis of high SNR simulations did not produce mass overestimates.

8.6.2 Variable a and b analysis

For comparison with the results obtained from simulated data in Section 8.5.1.3, I analysed the PSZ2G063.80+11.42 data using OM III while allowing a and b to vary. Figure 8.12 shows the resulting posterior distributions for the three analysis methods. As was the case in the simulations, the joint analysis gives a tighter constraint on the Y_{tot} and θ_p parameters, but does show a degeneracy in a and b .

8.7 Conclusions

I have introduced a joint likelihood function for data obtained from Planck and AMI in order to compare inferences obtained using it with those from the individual likelihood functions. The Bayesian analysis of Planck data was carried out using PowellSnakes (PwS, [Carvalho et al. 2012](#)) and AMI data were analysed in a way similar to the method outlined in [Feroz et al. \(2009\)](#) (and used in the preceding Chapters); the joint analysis ran both of these simultaneously. I tried implementing the likelihood hyperparameter method introduced in [Lahav et al. \(2000\)](#) and [Hobson, Bridle, & Lahav \(2002\)](#). I showed that likelihood ratios cannot be used with the hyperparameter method by implementing the toy model considered in [Hobson, Bridle, & Lahav \(2002\)](#). Therefore since PwS evaluates a likelihood ratio it is not compatible with this method. I generated simulations of clusters using an observational model (OM III, similar to the ones

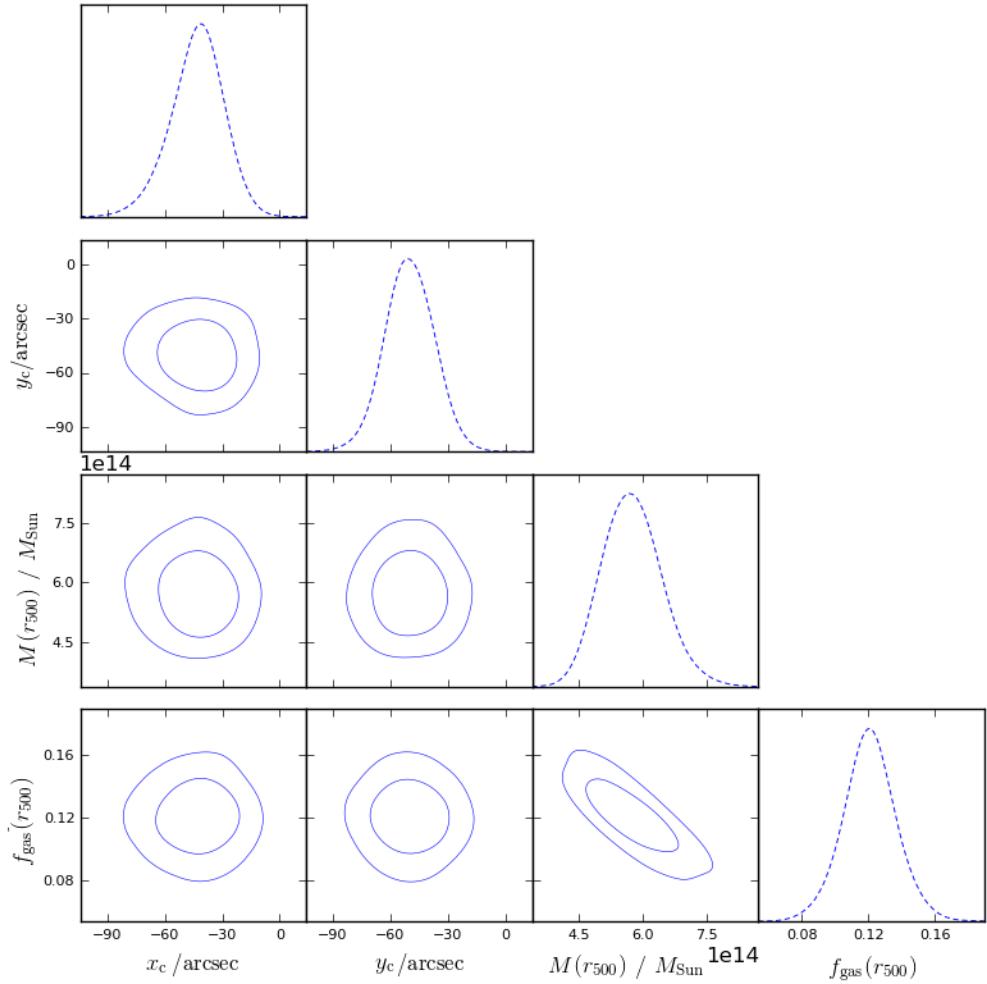


Figure 8.10: Marginalised posterior distributions obtained from joint AMI-Planck analysis of cluster PSZ2G063.80+11.42, using PM I. The dashed line plots are fully marginalised posterior distributions, while the contour plots are two-dimensional marginalised distributions. The inner contours correspond to the region of 68% confidence, while the outer contours corresponds to 95%.

used in Chapter 4) for 10 different noise realisations, and analysed the data using the same model. From looking at the resulting posterior distributions I found the following.

- For low signal-to-noise ratio (SNR) clusters, AMI data alone could be used to constrain values for the integrated Comptonisation parameter Y and angular radius θ rather well, but Planck data showed large degeneracies in θ . The joint analysis however showed the tightest constraints in $Y - \theta$ space (generally centred around the simulation input values).

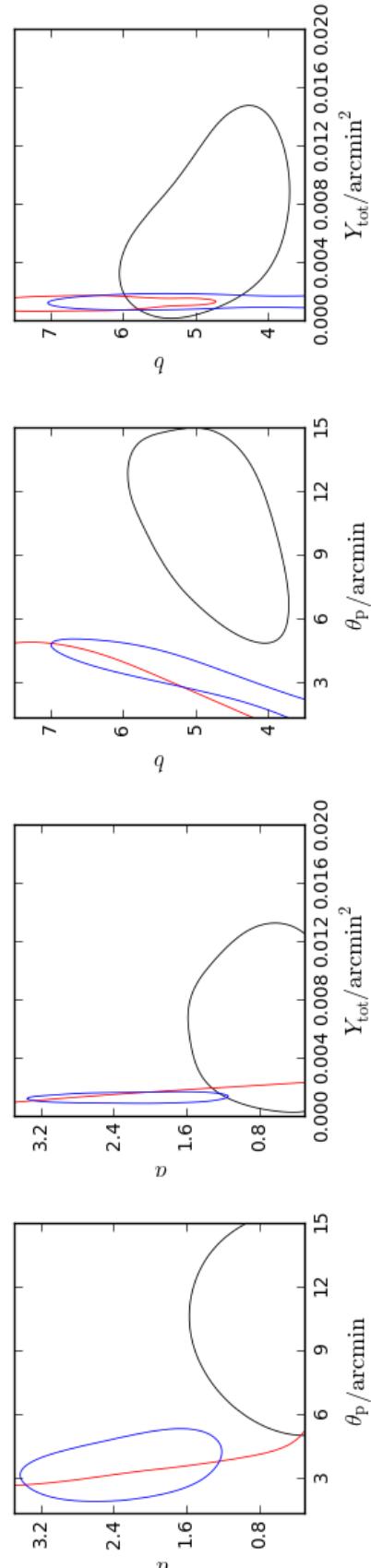


Figure 8.12: Two-dimensional posterior distributions for parameter pairs: (a, θ_p) , (a, Y_{tot}) , (b, θ_p) , and (b, Y_{tot}) . The black, red (in the first three plots, the region of higher probability is to the left of the curve), and blue contours correspond to the 68% confidence intervals of the AMI, Planck, and AMI-Planck posterior distributions respectively.

- For high SNR clusters, the Planck-only analyses gave moderate constraints on θ and good results for Y , while the AMI-only analyses showed large covariance between Y and θ . The joint analysis results gave similar results to the former, but with tighter constraints. For all three analyses, the true value was often in the proximity of the 68% confidence interval contours, rather than close to their centres.
- When allowing the shape parameters a and b of the generalised NFW model (Nagai, Kravtsov, & Vikhlinin 2007, used to parameterise the electron pressure) to vary in the Bayesian analysis, it was found that the joint analysis could generally constrain the Y and θ parameters better than the individual analyses, but showed degeneracies in a and b .

Using physical models derived in Chapters 2 and 5 I generated cluster simulations and analysed them with the three likelihood functions to infer cluster mass estimates. From this I found the following.

- For low SNR clusters I found that AMI underestimated cluster masses on average, but did recover the true value for some noise realisations. Planck systematically overestimated the masses by factors of at least two, while the joint analysis also led to overestimations (but generally to a smaller extent), suggesting it was the Planck likelihood dominating the joint posterior inferences.
- The gas fraction estimates from the joint analysis for low SNR clusters are consistently lower than the simulation input values, which suggests that the joint analysis is struggling to correctly infer the composition of the cluster, which is probably the cause of the mass overestimates.
- Analysis of high SNR clusters with AMI data gave accurate estimates of the input mass, while Planck data led to slight underestimates. Application of the joint analysis gave results similar to the individual analyses.

Finally, I applied the joint analysis to real data for the cluster PSZ2G063.80+11.42 which is part of the sample of 54 clusters considered in Chapters 3 and 4. I compared the mass estimates obtained with those obtained from AMI and Planck data and found the following.

- The AMI estimates and joint analysis mass estimates are $M_{\text{AMI}}(r_{500}) = (3.37 \pm 0.76) \times 10^{14} M_{\text{Sun}}$ (obtained in Chapter 4) and $M_{\text{Joint}}(r_{500}) = (5.74 \pm 0.70) \times 10^{14} M_{\text{Sun}}$ respectively. The two estimates derived from Planck data are $M_{\text{Pl, PMI}}(r_{500}) = (6.98 \pm 1.02) \times 10^{14} M_{\text{Sun}}$ and $M_{\text{Pl, slice}}(r_{500}) = (6.41 \pm 0.57) \times 10^{14} M_{\text{Sun}}$. The former of these was inferred directly from the PM I posterior distributions. The latter was obtained from the slicing function method introduced in Planck Collaboration XXVII (2016) and detailed in Section 3.4.

- The joint analysis estimate is sandwiched in between the other three values, but is closer to $M_{\text{Pl, PMI}}(r_{500})$ than it is $M_{\text{AMI}}(r_{500})$, suggesting that the Planck likelihood has a large effect on the joint analysis posterior distribution.
- The fact that both Planck data-only mass estimates are higher than the AMI value suggests that it is the data which are causing the relatively high estimates, at least for the real example considered here.
- When allowing the GNFW shape parameters a and b to vary, the joint analysis generally provides much tighter parameter constraints than the individual analyses.

CHAPTER **9**

MONTE CARLO SAMPLING METHODS

For most astrophysical problems, calculating the Bayesian evidence numerically is unfeasible, especially for high dimensional problems. Likewise, attempting to calculate parameter probability distributions exactly is computationally impossible. Thus one usually resorts to statistical sampling to make estimates of these quantities.

Monte Carlo sampling methods are a broad class of computational algorithms that rely on repeated random sampling of some distribution to obtain a numerical approximation of the true results. In the context of Bayesian inference, this amounts to representing a posterior distribution via a set of n_s weighted samples

$$\mathcal{S} = \{(\boldsymbol{\theta}_1, \mathcal{P}_1), \dots, (\boldsymbol{\theta}_{n_s}, \mathcal{P}_{n_s})\}, \quad (9.1)$$

where \mathcal{P}_i is the weight of each sample and $\sum_{i=1}^{n_s} \mathcal{P}_i = 1$. In this Chapter I give a brief review of how these samples can be obtained and used to plot approximations of the true posterior distribution. It serves as a reference to astrophysicists who are new to sampling, and refers to methods which are well known in the field of statistics.

9.1 Inverse transform sampling

Assuming we can draw independent, identically distributed random variables u that are uniformly distributed on $[0, 1]$, and provided we can calculate the inverse of the cumulative distri-

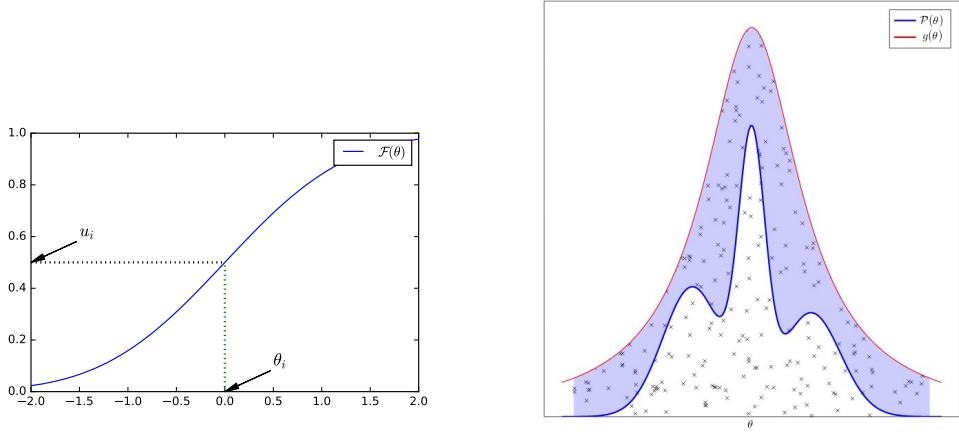


Figure 9.1: Left: Illustrating inverse transform sampling for a one-dimensional distribution $\mathcal{P}(\theta)$. Once a value for u_i is obtained, one draws a horizontal line from $(0, u)$ until it intersects with $\mathcal{F}(\theta)$ (black dotted line). The value of θ at the point of intersection is the point sampled from the distribution $(\theta_i$, green dotted line). Clearly the steeper $\mathcal{F}(\theta)$ is over an interval $\delta\theta$, the higher the chance of it intersecting with the horizontal lines corresponding to the uniform $[0, 1]$ samples. Right: Illustrating rejection sampling, taken from [Handley \(2016\)](#). The crosses correspond to samples from the distribution $g(\theta)$. Provided $g(\theta) > \mathcal{P}(\theta)$ for the domain of interest, then the samples beneath the blue curve (white area) can be regarded as samples from $\mathcal{P}(\theta)$.

bution function of the posterior \mathcal{F}^{-1} , then we can draw random samples from $\mathcal{P}(\boldsymbol{\Theta})$. We can interpret u as being a probability, and thus by evaluating $\mathcal{F}^{-1}(u)$ we are finding the value of $\boldsymbol{\Theta}$ which satisfies

$$u = \mathcal{F}(\boldsymbol{\Theta}) = \int_{\boldsymbol{\Theta}_m}^{\boldsymbol{\Theta}} \mathcal{P}(\boldsymbol{\Theta}') d\boldsymbol{\Theta}', \quad (9.2)$$

where $\boldsymbol{\Theta}_m$ is the component-wise minimum value of $\boldsymbol{\Theta}$ over which $\mathcal{P}(\boldsymbol{\Theta})$ is defined. The steepness of \mathcal{F} at a given point is proportional to the value of $\mathcal{P}(\boldsymbol{\Theta})$ and thus regions of higher probability density will be sampled from more often as shown in the left plot of Figure 9.1. Consequently the weights of the samples are proportional to the number of times a value of $\boldsymbol{\Theta}$ is sampled. The difficulty in inverse transform sampling arises when \mathcal{F}^{-1} is hard to evaluate.

9.2 Rejection sampling

Rejection sampling involves sampling from a proposal distribution $g(\theta)$ to ultimately draw samples from the distribution of interest $\mathcal{P}(\theta)$. The only requirement on $g(\theta)$ is that $g(\theta) > \mathcal{P}(\theta)$ for the domain of interest. The method works as follows.

- 1) Sample a value of θ ($\equiv \theta_g$) from $g(\theta)$ by using e.g. inverse transform sampling.
- 2) Sample a random variable u_g uniformly from the range $[0, g(\theta_g)]$.
- 3) If $\mathcal{P}(\theta_g) > u_g$, *accept* the point θ_g as a sample from \mathcal{P} and *reject* it otherwise. The sample weights are thus once again proportional to the number of times a value is sampled.

Rejection sampling is demonstrated graphically in the right plot of Figure 9.1. Note that rejection sampling can be inefficient (reject a lot of samples) when g and \mathcal{P} are dissimilar. The similarity between the two can be quantified by some distribution distance metric such as the Earth Mover's Distance used in Chapter 4.

9.3 Markov Chain Monte Carlo sampling

Before talking about Markov Chain Monte Carlo (MCMC) sampling methods I give a primer on Markov chains and state some of their key properties relevant to MCMC.

9.3.1 Markov chains

9.3.1.1 Types of Markov chain

A Markov chain is a sequence of random variables for which the probability of outcomes for a particular element of the chain depends only on the state attained in the previous step of the chain. A Markov chain X can be continuous in time i.e. $X \equiv X(t)$ for $t \geq 0$ or discrete, $X \equiv \{X_0, X_1, \dots, X_{n-1}\}$. In the case of the former we are saying that the chain can be measured at any time t , while for the latter we are saying X can only be measured at discrete times defined by the index n . The possible values that X can take (often referred to as the *state space*, Θ) can also be continuous or discrete. A continuous state space refers to one in which X can take any of the (uncountably infinite) values defined on the space. A discrete state space can include a finite or a countably infinite number of states.

9.3.1.2 Discrete time discrete state space Markov chains

For the properties considered here we will consider discrete time discrete state space Markov chains only, but note that these ideas generalise to the continuous cases. For more information on continuous Markov chains we refer the reader to MacKay (2002), Robert & Casella (2004), and Johansen & Evers (2007). A Markov chain X with discrete time domain and discrete state

space Θ can be stated mathematically as

$$P(X_n = \theta_i | X_{n-1} = \theta_j, \dots, X_0 = \theta_k) = P(X_n = \theta_i | X_{n-1} = \theta_j) \equiv T_{n,ji}, \quad (9.3)$$

where the $\theta_l \in \Theta$. $T_{n,ji}$ is the transition probability from θ_j to θ_i between steps $n-1$ and n . For a **homogeneous** Markov chain the transition probability between two states is independent of time, thus we can write $T_{n,ji} \equiv T_{ji}$.

9.3.1.3 State properties

We will now focus on *homogeneous* Markov chains and introduce some of their properties relevant to Monte Carlo sampling.

A state θ_i is said to be **accessible** from state θ_j (denoted $\theta_j \rightarrow \theta_i$) if

$$\inf\{n : P(X_n = \theta_i | X_0 = \theta_j) > 0\} < \infty, \quad (9.4)$$

or equivalently $\inf\{n : T_{ji}^n > 0\} < \infty$. \inf refers to the infimum of the set (greatest lower bound of the set). If this condition is satisfied it means that there is a finite probability of moving from state θ_j to state θ_i after a finite number of steps n . The definition of **communication** follows from accessibility: two states θ_j and θ_i are said to *communicate* with each other ($\theta_j \leftrightarrow \theta_i$) if they are *accessible* from one another

$$\theta_j \leftrightarrow \theta_i \Leftrightarrow \theta_j \rightarrow \theta_i \text{ and } \theta_i \rightarrow \theta_j. \quad (9.5)$$

A Markov chain is said to be **irreducible** if all states *communicate* with each other, that is $\theta_j \leftrightarrow \theta_i$ for all $\theta_i, \theta_j \in \Theta$. This is important in the context of MCMC as a chain with this property can explore the entire state space without being confined to some portion of it (which could be determined by the chain's initial state). The chain is said to be **strongly irreducible** if any state can be reached from any other state in a single step i.e. if $T_{ji} > 0$ for all i and j .

It is also important to consider the number of paths can take from a state θ_i before the chain returns to θ_i , as this will tell us something about the presence of long-range correlation between the states of the chain. A state θ_i has period $d(\theta_i)$ which is given by

$$d(\theta_i) = \gcd\{n \geq 1 : T_{i,i}^n > 0\}, \quad (9.6)$$

where \gcd denotes the greatest common denominator of the set. It can be shown that all states which communicate have the same period, hence for an *irreducible* Markov chain all states have the same period. An irreducible Markov chain with $d(\theta_i) = 1$ is said to be **aperiodic**. This

essentially means that the Markov chain can transition back into the same state that it was in at the previous step. In the context of MCMC this means that the same value can be sampled consecutively.

Another quantity relevant to MCMC is the number of times a state is visited, n_{θ_i} , in the asymptotic limit $n \rightarrow \infty$. We define this as

$$n_{\theta_i} = \sum_{j=0}^{\infty} \mathcal{I}(X_j = \theta_i), \quad (9.7)$$

where $\mathcal{I}(Y)$ equals one for $Y = 0$ and zero otherwise. The introduction of n_{θ_i} allows us to introduce two more properties of Markov chains: **transience** and **recurrence**. A state is said to be *transient* if

$$\mathbb{E}(n_{\theta_i}) < \infty, \quad (9.8)$$

while it is said to be *recurrent* if

$$\mathbb{E}(n_{\theta_i}) = \infty, \quad (9.9)$$

where the expectations are taken in the asymptotic limit. In the case of irreducible chains, transience and recurrence are properties of the chain itself rather than its individual states, so we can say that for such a chain all states are either transient or they are all recurrent. If the Markov chain is recurrent then the samples from MCMC can take any value in Θ an infinite number of times. Another notion of recurrence can be defined with respect to time rather than frequency of transitions to a state: if the ‘time’ (number of steps) between a chain moving to state θ_i and revisiting the state, $\tau_{\theta_i, \theta_i}$ has a finite first moment, then the state is said to be **positive recurrent**. Note that positive recurrence is also a property of the whole Markov chain in the case that it is irreducible.

9.3.1.4 Stationarity and reversibility of Markov chains

A distribution μ defined on Θ is said to be **stationary** if

$$\mu \mathbf{T} = \mu, \quad (9.10)$$

where μ is a row vector of the values of $\mu(\theta_i) \equiv \mu_i$ for all $\theta_i \in \Theta$ and \mathbf{T} is a matrix of transition probabilities T_{ij} for all valid i and j . If at any step along the Markov chain its marginal distribution $P(X_i)$ is distributed according to its stationary distribution μ , then it stays distributed according to μ since $\mu \mathbf{T}^n = \mu$ for arbitrary n .

A *stationary* stochastic process is said to be **reversible** if the statistics of the time-reversed version of the process match those obtained in the original. An alternative way of interpreting this is that the distribution of any collection of future states given the past states must match the conditional distribution of the past states given the future states. This means that we require

$$P(X_0 = \theta_i | X_{-1} = \theta_j) = P(X_0 = \theta_i | X_1 = \theta_j). \quad (9.11)$$

It can be shown that if a Markov chain satisfies the **detailed balance** relation given by

$$T_{ij}\mu_i = T_{ji}\mu_j, \quad (9.12)$$

then the chain is *reversible*. Note that satisfying equation 9.11 is a *sufficient* condition for a Markov chain to converge to its stationary distribution (μ). The reversibility property can be shown by substituting equation 9.12 into 9.11

$$\begin{aligned} P(X_0 = \theta_i | X_{-1} = \theta_j) &= T_{ji} \\ &= \frac{T_{ij}\mu_i}{\mu_j} \\ &= \frac{P(X_1 = \theta_j | X_0 = \theta_i)P(X_0 = \theta_i)}{P(X_1 = \theta_j)} \\ &= P(X_0 = \theta_i | X_1 = \theta_j). \end{aligned} \quad (9.13)$$

Note that the *necessary* conditions for a Markov chain (with a discrete state space) to converge on the target distribution are for it to be irreducible, aperiodic, and for the stationary distribution to be the target distribution.

9.3.2 Examples of MCMC algorithms

To use MCMC to sample from continuous probability distributions, we must assume that our Markov chain has a continuous state space for θ , but we still work in discrete time. In this case the detailed balance relation between steps k and $k + 1$ along the chain is given by

$$T(\theta_{k+1}, \theta_k)\mu(\theta_k) = T(\theta_k, \theta_{k+1})\mu(\theta_{k+1}), \quad (9.14)$$

where $T(\theta_{k+1}, \theta_k)\mu(\theta_k) = P(X_{k+1} \in \Theta | X_k \in \Theta)$ and $\mu(\theta_k) = \mu(X_k \in \Theta)$. In the context of Bayesian inference, the posterior distribution should be the target distribution of the Markov chain and so we want $\mu(\theta_k) \equiv \mathcal{P}(\theta_k)$. All that is left is to find a form for the transition distribution that satisfies equation 9.14 (a *sufficient* condition for the Markov chain to converge to $\mathcal{P}(\theta)$).

9.3.2.1 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm (MH, [Hastings 1970](#)) generates samples from $\mathcal{P}(\theta)$ using a relatively simple trial distribution q . For a step along the Markov chain from k to $k + 1$ the algorithm operates as follows.

- 1) Sample a trial point θ' from the trial distribution $q(\theta'|\theta_k)$.
- 2) Calculate the acceptance probability $\alpha(\theta',\theta_k) = \min\left(\frac{q(\theta_k|\theta')\mathcal{P}(\theta')}{q(\theta'|\theta_k)\mathcal{P}(\theta_k)}, 1\right)$.
- 3) Draw a uniform random variable u from $[0, 1]$. If $u < \alpha(\theta',\theta_k)$ set $\theta_{k+1} \rightarrow \theta'$. Otherwise $\theta_{k+1} \rightarrow \theta_k$.

In Section [D.1](#) we show that the MH algorithm satisfies detailed balance, that the MH acceptance probability can be derived from the detailed balance relation, and that the MH acceptance probability is optimal in the sense that it permits the most steps along the chain without violating detailed balance. The Appendix also gives the relation between T and α .

Like the previous sampling techniques considered, MH produces posterior samples with weights proportional to the number of times each state is visited.

9.3.2.2 Metropolis algorithm

When the trial distribution $q(\theta'|\theta_k)$ is symmetric in its arguments, i.e. $q(\theta'|\theta_k) = q(\theta_k|\theta')$, then the trial acceptance probability simplifies to ([Metropolis et al. 1953](#))

$$\alpha(\theta',\theta_k) = \min\left(\frac{\mathcal{P}(\theta')}{\mathcal{P}(\theta_k)}\right). \quad (9.15)$$

This form for α still satisfies detailed balance (for suitable \mathcal{P} and q) and can be useful when calculating the trial distribution (not necessarily sampling from it) is difficult, as is the case in Chapter [10](#).

9.4 Nested sampling

[Skilling \(2004\)](#) introduced a novel sampling method referred to as nested sampling. This algorithm focuses on calculating the evidence, but also generates samples from the *posterior* probability distribution. The key computational expense associated with nested sampling is the constraint that newly generated samples must be above a certain likelihood value which increases at each iteration.

Initially, [Sivia & Skilling \(2006\)](#) suggested satisfying this constraint by evolving a Markov chain starting at one of the pre-existing samples and evaluating an acceptance ratio based on the one

used by the Metropolis algorithm (Metropolis et al. 1953) used in Markov Chain Monte Carlo (MCMC) sampling (see e.g. MacKay 2002 for a review). A variant of the nested sampling algorithm which focused on sampling from ellipsoids which approximate the region in which the likelihood constraint is satisfied was also developed (Mukherjee, Parkinson, & Liddle 2006). A major breakthrough in the applicability of nested sampling to highly multi-modal distributions came with the invention of clustering nested sampling algorithms (Shaw, Bridges, & Hobson 2007, Feroz & Hobson 2008, and MULTINEST). The latter of these was used extensively in the preceding Chapters to carry out Bayesian inference). These algorithms effectively sample from multiple ellipsoids determined by some clustering algorithm, with the aim of approximating the likelihood constraint for each mode of the distribution. More recently, the slice sampling algorithm POLYCHORD (Handley, Hobson, & Lasenby 2015, Handley, Hobson, & Lasenby 2015) has been introduced and is effective at navigating high dimensional spaces, due to the fact that it is not a rejection sampling algorithm. Section 4.1 of Handley, Hobson, & Lasenby (2015) gives further examples of nested sampling algorithms which have different ways of satisfying the likelihood constraint.

9.4.1 Overview of the nested sampling algorithm

Nested sampling exploits the relation between the likelihood and ‘prior volume’ to transform the N -dimensional integral given by equation 2.43 into a one-dimensional integral. The prior volume X is defined by $dX = \pi(\boldsymbol{\Theta}) d\boldsymbol{\Theta}$ for parameter space $\boldsymbol{\Theta}$, thus X is defined on $[0, 1]$ and we can set

$$X(\lambda) = \int_{\mathcal{L}(\boldsymbol{\Theta}) > \lambda} \pi(\boldsymbol{\Theta}) d\boldsymbol{\Theta}. \quad (9.16)$$

The integral extends over the region(s) of the parameter space contained within the iso-likelihood contour $\mathcal{L}(\boldsymbol{\Theta}) = \lambda$ (see Figure 9.2). Assuming that the inverse of equation 9.16 ($\lambda(X) = X^{-1}(\lambda) \equiv \mathcal{L}(X)$) exists which is the case when π is strictly positive, then the evidence integral can be written as (see Section D.2)

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX. \quad (9.17)$$

Thus, if one can evaluate $\mathcal{L}(X)$ at n_s values of X , the integral given by equation 9.17 can be approximated by standard quadrature methods

$$\mathcal{Z} \approx \sum_{i=1}^{n_s} \mathcal{L}_i (X_{i-1} - X_i), \quad (9.18)$$

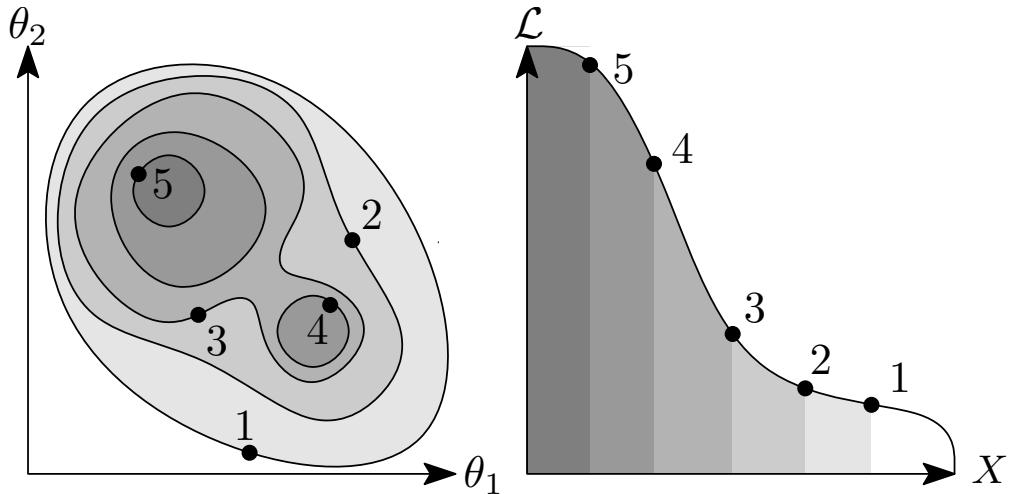


Figure 9.2: Left plot: Five iso-likelihood contours of a two-dimensional, multi-modal likelihood $\mathcal{L}(\theta_1, \theta_2)$. Each contour encloses some fraction of the prior X , with the colourscale indicating the value of X (darkest: smallest X). Right: Corresponding \mathcal{L} as a function of X plot (not to scale). The area under the curve is equal to \mathcal{Z} .

where

$$0 < X_{n_s} < \dots < X_1 < X_0 = 1. \quad (9.19)$$

Note that one can use more accurate approximations to the integral 9.17 such as the trapezium rule (which has numerical error $\mathcal{O}\left(\frac{1}{n_s^2}\right)$, compared with $\mathcal{O}\left(\frac{1}{n_s}\right)$ for the sum given above)

$$\mathcal{Z} \approx \sum_{i=1}^{n_s} \frac{1}{2} (\mathcal{L}_{i-1} + \mathcal{L}_i) (X_{i-1} - X_i). \quad (9.20)$$

However, I use the method given by equation 9.18 in our implementation of the geometric nested sampler (Chapter 10) for simplicity. Note further that the first inequality in equation 9.19 follows from the fact that there could always remain some tiny prior volume containing a larger likelihood value than \mathcal{L}_{n_s} , unless that can be ruled out by some *a-priori* knowledge of the maximum value of \mathcal{L} .

9.4.1.1 Determining the nested sampling sum

The nested sampling algorithm performs the summation 9.18 as follows. At initiation n_l ‘live-points’ are sampled from the prior $\pi(\boldsymbol{\Theta})$ which are uniformly distributed in the region X_{i-1} ($= 1$ upon initiation). Note also that $\mathcal{L}_0 = 0$ (relevant when the trapezium rule is used). \mathcal{L} is calcu-

lated for each of these points, and the livepoint corresponding to the lowest likelihood value \mathcal{L}_i is removed from the livepoint set. This ‘deadpoint’ is replaced by a point drawn from $\pi(\boldsymbol{\Theta})$, say $\boldsymbol{\Theta}_t$, subject to the constraint that $\mathcal{L}_t > \mathcal{L}_i$. Once this constraint has been satisfied, $\boldsymbol{\Theta}_t$ is added to the livepoint set. As noted in [Skilling \(2004\)](#), it is intuitive to assume that the shrinkage in X associated with each iso-likelihood contour is geometric. Hence we can write

$$X_1 = t_1 X_0, X_2 = t_2 X_1, \dots, X_{n_{s-1}} = t_{n_{s-1}} X_{n_{s-2}}, X_{n_s} = t_{n_s} X_{n_{s-1}}, \quad (9.21)$$

where each t_i lies between zero and one, and can be thought of as the shrinkage factor between successive shells of the prior volume. In practice it is difficult to determine the exact values of t_i , as the amount of prior volume shrinkage between iso-likelihood contours \mathcal{L}_i and \mathcal{L}_{i-1} is in general, non-trivial to calculate. Nevertheless, we can estimate t_i statistically as follows. Since at each iteration of shrinking the prior volume, there are n_l livepoints uniformly distributed in X_{i-1} , then we can take t_i to be the largest of n_l uniformly distributed numbers between zero and one, since the lowest likelihood should be attributed with the smallest volume shrinkage. This gives the following distribution for the shrinkage factor (derived in Section [D.3](#))

$$P(t_i) = n_l t_i^{n_l-1}. \quad (9.22)$$

This statistical treatment of the t_i can be used to calculate the expected value of \mathcal{Z} as well as its error, as detailed in [Keeton \(2011\)](#). Once t_i has been calculated, X_i can be determined and one is left with n_l livepoints uniformly distributed in the range $[0, X_i]$. For the next iteration of the algorithm the process is repeated from the step of determining the livepoint with the lowest likelihood.

As explained in [Skilling \(2004\)](#), the geometric uncertainty associated with the X_i leads to the idea that $\log(\mathcal{Z})$ rather than \mathcal{Z} is a normally distributed variable. Assuming the latter to be normally distributed can result in distributions of \mathcal{Z} with variances that suggest \mathcal{Z} can take negative values, which is unphysical. This is the case with the likelihood describing gravitational wave detection used in Section [10.4.4.3](#). The mean and variance of a log-normally distributed random variable, $\mathbb{E}[\log(\mathcal{Z})]$ and $\text{var}[\log(\mathcal{Z})]$, can be calculated from the moments of the non-logarithmic variables as

$$\mathbb{E}[\log(\mathcal{Z})] = 2 \log(\mathbb{E}[Z]) - \frac{1}{2} \log(\mathbb{E}[Z^2]), \quad (9.23)$$

$$\text{var}[\log(\mathcal{Z})] = \log(\mathbb{E}[Z^2]) - 2 \log(\mathbb{E}[Z]). \quad (9.24)$$

Hence our geometric nested sampling algorithm calculates the moments of the linear variables following Keeton (2011) (in log-space to avoid numerical difficulties, see Section D.4) but the final evidence estimate and its associated error are calculated using equations 9.23 and 9.24.

9.4.1.2 Stopping criterion

The nested sampling algorithm can be terminated based on an estimate of how precisely the evidence value has been calculated up to the current iteration. One measure of this is to look at the ratio of the current estimate of \mathcal{Z} to its value plus an estimate of the ‘remaining’ evidence associated with the current livepoints. Since after iteration n_s the livepoints are uniformly distributed in the range $[0, X_{n_s}]$, we can approximate their final contribution to the evidence as

$$\mathcal{Z}_f \approx \frac{X_{n_s}}{n_l} \sum_{i=1}^{n_l} \mathcal{L}_i, \quad (9.25)$$

where \mathcal{L}_i is the likelihood value of the i^{th} remaining livepoint. The stopping criterion can then be quantified as

$$\frac{\mathcal{Z}_f}{\mathcal{Z}_f + \mathcal{Z}} < \epsilon. \quad (9.26)$$

ϵ is a user defined parameter, which I set to 0.01 in the nested sampling implementations used in Chapter 10. The final estimate of \mathcal{Z} is then updated to be $\mathcal{Z} \rightarrow \mathcal{Z} + \mathcal{Z}_f$. Note that after a large number of iterations of the nested sampling algorithm, we can be fairly confident that the remaining contribution to the evidence is small. Referring back to equation 9.18, as the sampling progresses the value of $(X_{i-1} - X_i)$ gets smaller and there will be a point part way through the process, where its value decreases at a rate faster than \mathcal{L}_i increases. Thus after this point, the contribution to the evidence at each iteration becomes smaller, until at some point it becomes negligible (see Figure 9.3).

9.4.1.3 Posterior inferences

Once \mathcal{Z} has been determined, posterior inferences can easily be generated using the deadpoints and final livepoints from the nested sampling process to give a total of $n_s + n_l$ samples (and we set $n_s \rightarrow n_s + n_l$). Each such point is assigned the weight

$$\mathcal{P}_i = \frac{\mathcal{L}_i (X_{i-1} - X_i)}{\mathcal{Z}}. \quad (9.27)$$

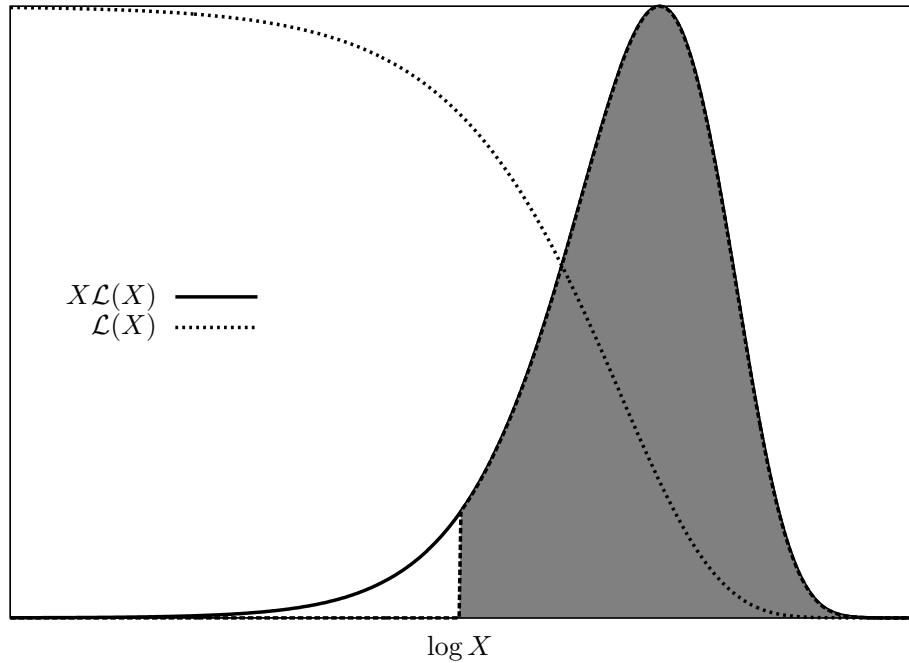


Figure 9.3: Plots of $\mathcal{L}(X)$ and $\mathcal{L}(X)X$ for typical likelihood functions. The area under the $\mathcal{L}(X)$ curve corresponds to \mathcal{Z} . The height of the curve $\mathcal{L}(X)X$, gives an indication of the contribution to \mathcal{Z} for a small fractional change in X . After a number of nested sampling iterations, this contribution becomes negligible.

Note that for the n_l samples obtained from the final set of livepoints $X_{i-1} - X_i = \frac{X_{n_s}}{n_l}$. The weights (along with the corresponding values of $\boldsymbol{\theta}$) can be used to calculate statistics of the posterior distribution, or plot it using software such as GETDIST or CORNER^a.

9.5 Plotting posterior samples

The set of discrete samples \mathcal{S} can be used to determine functional approximations to \mathcal{P} . Histograms and kernel density estimation (KDE) are two popular methods deployed to obtain distribution approximations from samples.

^a<https://pypi.python.org/pypi/corner>.

9.5.1 Histograms

Histograms provide a quick way to generate a piecewise discontinuous approximation of $\mathcal{P}(\boldsymbol{\Theta})$. The sample weights are ‘binned’ into a series of intervals separating $\boldsymbol{\Theta}$. The new sample weight for each bin b , $\mathcal{P}(\boldsymbol{\Theta}_b)$, is simply the sum of the $\mathcal{P}(\boldsymbol{\Theta}_i)$ associated with that bin, and the value of $\boldsymbol{\Theta}_b$ is defined as some function of the corresponding $\boldsymbol{\Theta}_i$ (e.g. their average). For each bin \mathcal{P} is constant over the corresponding interval on $\boldsymbol{\Theta}$ and so the function approximation is discontinuous. using a small number of bins reduces the noise associated with the sampling process, but can lead to key features of the true $\mathcal{P}(\boldsymbol{\Theta})$ being missed, while a large number of bins will tend to overfit to the samples & produce a very ‘peaky’ approximation. Figure 9.4 shows an example of a histogram with a moderate number of bins, which catches the main features of \mathcal{P} but also includes a noticeable amount of sampling noise.

9.5.2 Kernel density estimation

KDE is a non-parametric method for estimating probability densities from samples, which ‘improves’ on simple histograms by smoothing the resulting curve. A continuous function approximation for the posterior, $\hat{\mathcal{P}}(\boldsymbol{\Theta})$ is given by

$$\hat{\mathcal{P}}(\boldsymbol{\Theta}) = \sum_{i=1}^{n_s} \mathcal{P}(\boldsymbol{\Theta}_i) K_h(\boldsymbol{\Theta} - \boldsymbol{\Theta}_i), \quad (9.28)$$

where K_h is a smoothing kernel with width parameter h . K_h must integrate over its domain to one (i.e. be a probability density function) to ensure that $\hat{\mathcal{P}}(\boldsymbol{\Theta})$ is also normalised. h determines the variance of the smoothing kernel and thus how smooth $\hat{\mathcal{P}}(\boldsymbol{\Theta})$ is. Figure 9.4 illustrates the use of KDE with a Gaussian smoothing kernel and either $h = 0.1$, $h = 0.3$, or $h = 1$ ^b. The latter value corresponds to a $\hat{\mathcal{P}}(\boldsymbol{\Theta})$ which is a poor estimation of $\mathcal{P}(\boldsymbol{\Theta})$ (due to ‘oversmoothing’ $\hat{\mathcal{P}}(\boldsymbol{\Theta})$ does not reveal the bimodality of $\mathcal{P}(\boldsymbol{\Theta})$). The $\hat{\mathcal{P}}(\boldsymbol{\Theta})$ corresponding to $h = 0.1$ and $h = 0.3$ capture the bimodality of $\mathcal{P}(\boldsymbol{\Theta})$, but include a lot of small peaks not present in the true distribution (‘undersmoothing’). GetDist uses a truncated Gaussian for K_h with the determination of h based on minimisation of the mean integrated square error ^c

$$\min_h \left[\int \mathbb{E} \left[(\mathcal{P}(\boldsymbol{\Theta}) - \hat{\mathcal{P}}(\boldsymbol{\Theta}))^2 \right] d\boldsymbol{\Theta} \right]. \quad (9.29)$$

^bExample inspired by <https://jakevdp.github.io/blog/2013/12/01/kernel-density-estimation/>.

^cFor more information on the specifics of the implementation of KDE used in GetDist, see <https://cosmologist.info/notes/GetDist.pdf>.

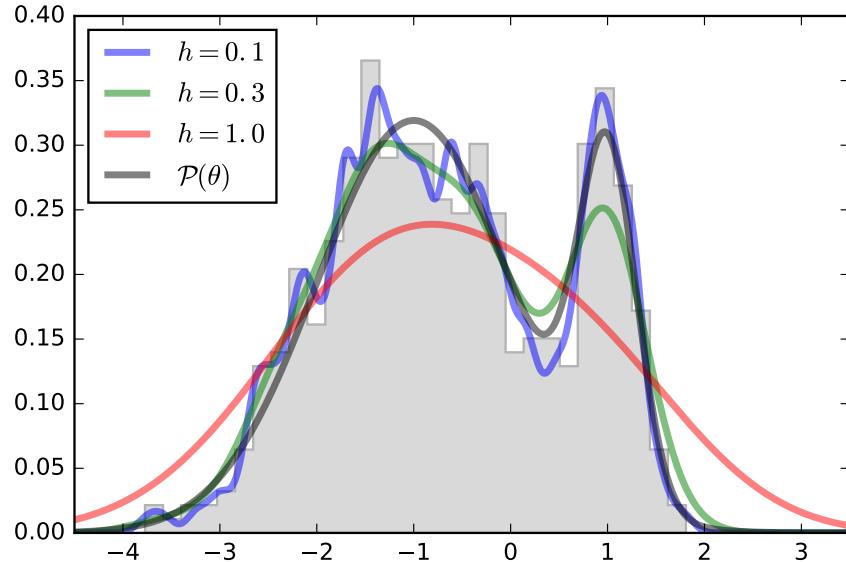


Figure 9.4: Illustration of approximating a one-dimensional posterior function $\mathcal{P}(\theta)$ using a histogram or KDE. $\mathcal{P}(\theta)$ is a Gaussian mixture model (parameterised in terms of means and standard deviations): $\mathcal{P}(\theta) = 0.8 \times \mathcal{N}(-1, 1) + 0.2 \times \mathcal{N}(1, 0.3)$. The samples $\mathcal{S} = \{(\theta_1, \mathcal{P}_1), \dots, (\theta_{n_s}, \mathcal{P}_{n_s})\}$ are generated by drawing 100 samples from $\mathcal{N}(1, 0.3)$ & 400 samples from $\mathcal{N}(-1, 1)$, and are assigned to the θ_{n_i} while the weights are set to unity. The histogram is generated by binning the samples into 30 bins of uniform width over the range $[-4.5, 3.5]$. The KDE estimates are generated using a Gaussian function for K_h , with either $h = 0.1$, $h = 0.3$, or $h = 1$. The black curve represents the ‘true’ form of the function $\mathcal{P}(\theta)$, the grey region represents the (discrete) histogram approximation, while the blue, green and red curves correspond to the $\hat{\mathcal{P}}(\theta)$ obtained for the different values of h .

CHAPTER 10

GEOMETRIC NESTED SAMPLING

Here I present a nested sampling algorithm which provides a new method for satisfying the nested sampling likelihood constraint (see Section 9.4) based on the Markov method used in [Sivia & Skilling \(2006\)](#) (and also applied in [Feroz & Hobson 2008](#)). Certain parameters relevant to astrophysics exhibit special properties which mean they naturally parameterise points on geometric objects such as circles, tori and spheres. The algorithm we introduce here which we refer to as the geometric nested sampler, exploits these properties to generate samples efficiently and enables mobile exploration of distributions which are defined on such geometries. My implementation of the algorithm can be found at https://github.com/SuperKam91/nested_sampling. A paper corresponding to the work carried out in this Chapter is going to be submitted to MNRAS ([Javid 2018](#)), and contains several more motivating toy examples for the geometric nested sampler.

10.1 Nested sampling prior distributions

Bayesian inference has been reviewed in Section 2.6 and nested sampling in Section 9.4. Here we make a note about the form of the prior distribution π of the parameter set Θ used throughout this Chapter.

In general for nested sampling, $\pi(\Theta)$ can take any form as long as the distribution integrates to one and has a connected support ([Chopin & Robert 2008](#); this roughly means that the parts of the

domain at which $\pi(\boldsymbol{\Theta}) \neq 0$ is not ‘separated’ by the parts at which $\pi(\boldsymbol{\Theta}) = 0$). For simplicity, in all examples considered here I assume that each component of the N -dimensional vector $\boldsymbol{\Theta} = (\theta_1, \dots, \theta_N)$ is independent of one another, and that each $\pi(\theta_i)$ is a uniform probability distribution, so that

$$\pi(\boldsymbol{\Theta}) = \prod_{i=1}^N \pi_i(\theta_i) = \prod_{i=1}^N \frac{1}{\theta_{\max,i} - \theta_{\min,i}}, \quad (10.1)$$

where $\theta_{\max,i}$ and $\theta_{\min,i}$ are respectively the upper and lower bounds on θ_i . Values for $\theta_{\max,i}$ and $\theta_{\min,i}$ used in the examples presented here will be stated in the following Sections.

10.2 Satisfying the likelihood constraint

At each step of the nested sampling iteration, one needs to sample a new point which satisfies $\mathcal{L}_t > \mathcal{L}_i$. As mentioned in Section 9.4, considerable work has been put into increasing the efficiency of this process, as it is by far the most computationally expensive step of the nested sampling algorithm. I now give a review of the Metropolis nested sampling method used by [Sivia & Skilling \(2006\)](#) and [Feroz & Hobson \(2008\)](#), which forms the basis of the method used in geometric nested sampling.

10.2.1 Metropolis nested sampling

The Metropolis nested sampling method is an adaption of the Metropolis algorithm used in MCMC sampling of a posterior distribution (see Sections 9.3 and D.1). The acceptance ratio for the Metropolis nested sampling algorithm takes the form

$$\alpha = \begin{cases} \min [\pi(\boldsymbol{\Theta}_t) / \pi(\boldsymbol{\Theta}_l), 1] & \text{if } \mathcal{L}_t > \mathcal{L}_i, \\ 0 & \text{otherwise.} \end{cases} \quad (10.2)$$

Here $\boldsymbol{\Theta}_l$ is obtained by picking one of the current livepoints at random, and using its value of $\boldsymbol{\Theta}$. The value for $\boldsymbol{\Theta}_t$ is sampled from a trial distribution $q(\boldsymbol{\Theta}_t | \boldsymbol{\Theta}_l)$. [Sivia & Skilling](#) and [Feroz et al.](#) use symmetric Gaussian distributions centred on $\boldsymbol{\Theta}_l$ for $q(\boldsymbol{\Theta}_t | \boldsymbol{\Theta}_l)$. The trial point is accepted to be a new livepoint (replacing the deadpoint associated with \mathcal{L}_i) with probability α . Note that equation 10.2 implicitly assumes that the proposal distribution is symmetric in its arguments, that is $q(\boldsymbol{\Theta}_t | \boldsymbol{\Theta}_l) = q(\boldsymbol{\Theta}_l | \boldsymbol{\Theta}_t)$. In the case that the proposal distribution is asymmetric, the acceptance ratio includes an additional factor $q(\boldsymbol{\Theta}_l | \boldsymbol{\Theta}_t) q(\boldsymbol{\Theta}_t | \boldsymbol{\Theta}_l)$ (in which case the algorithm is referred to as the Metropolis-Hastings algorithm, see Section 9.3). The fact that the Metropolis nested

sampling method uses the current livepoints as a ‘starting point’ for selecting $\boldsymbol{\Theta}_t$, means that the autocorrelation between the livepoints is high, which in turn leads to biased sampling. This can be prevented by increasing the variance of the trial distribution used, or by requiring that multiple trial points must be accepted before the final one is accepted as a livepoint, i.e. after the first accepted trial point is found, set $\boldsymbol{\Theta}_l \rightarrow \boldsymbol{\Theta}_t$ and use this to sample a new $\boldsymbol{\Theta}_t$ from $q(\boldsymbol{\Theta}_t|\boldsymbol{\Theta}_l)$. This can be repeated an arbitrary number of times, but in general more iterations leads to a lower correlation between the livepoint used at the beginning of the chain and the final accepted trial point which is added to the livepoint set. Sivia & Skilling suggest that at each nested sampling iteration, the number of trial points generated n_t to get a new livepoint should be ≈ 20 . In my implementation I set this number to $20 \times N$ where N is the dimensionality of the parameter estimation problem. Note that n_t includes both accepted and rejected trial points. Sivia and Skilling also suggest that the acceptance rate for the trial points at each nested sampling iteration should be $\approx 50\%$. This is because a high acceptance rate usually suggests high auto-correlation between the successive trial points, whilst a low acceptance rate can suggest high correlation between the final accepted trial point and the one used to initialise the chain, as too few steps have been made between the two. In the extreme case that the acceptance rate is zero, the process of picking a new livepoint has failed, as one cannot have two livepoints corresponding to the same $\boldsymbol{\Theta}$. The acceptance rate is affected by the variance of the trial distribution, a large variance usually results in more trial points being rejected (especially near the peaks of the posterior). Sivia & Skilling suggest updating the trial standard deviation as

$$\sigma_t \rightarrow \begin{cases} \sigma_t \exp(1/N_a) & \text{if } N_a > N_r, \\ \sigma_t \exp(-1/N_r) & \text{if } N_a \leq N_r, \end{cases} \quad (10.3)$$

where N_a and N_r are the number of accepted and rejected trial points in the current nested sampling iteration respectively. Note however that I determine the variance using different methods (see Sections 10.3.3 and 10.3.6).

Feroz et al. incorporate the Metropolis likelihood sampling into their clustering nested sampling algorithm rather than use it in isolation. The geometric likelihood sampling I introduce in the next Section is a modified version of the Metropolis algorithm used in isolation.

10.3 Geometric nested sampling

One key issue with Metropolis nested sampling is that at each nested sampling iteration, if too many trial points are rejected, then the livepoints will be highly correlated with each other after

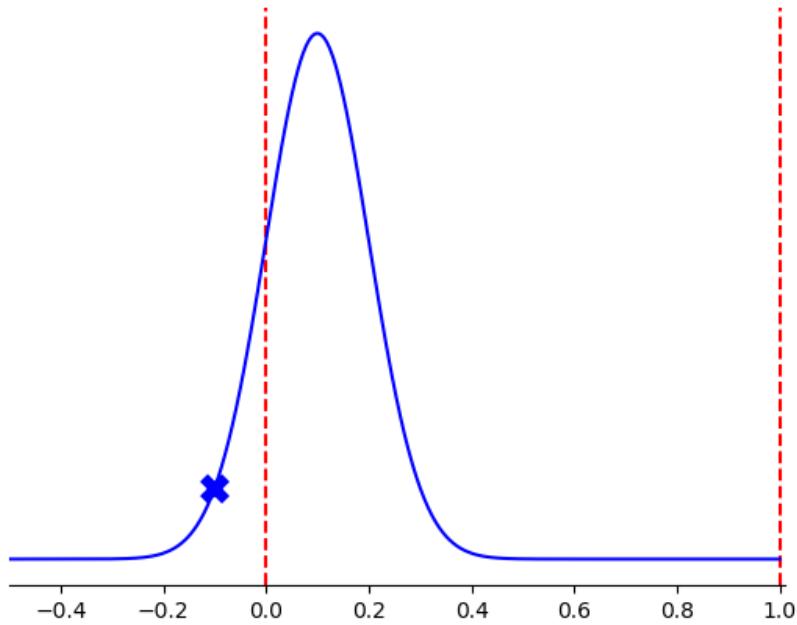


Figure 10.1: ‘Vanilla’ non-wrapped trial distribution. The blue curve represents a Gaussian ‘vanilla’ trial distribution $q(\theta'|\theta)$ with starting point $\theta = 0.1$, and sampled trial point $\theta' = -0.1$ shown by the blue cross. The support of π is indicated by the red dashed lines ($[0, 1]$). Since θ' lies outside the support of π , it would automatically be rejected by the Metropolis algorithm.

a number of nested sampling iterations. To prevent this one must sample a large number of trial points in order to increase the number of acceptances and decrease the auto-correlation of the trial point chain. This solution can be problematic if computing the likelihood is computationally expensive. One particular case in which the sampled point is guaranteed to be rejected, is if the point lies outside of the domain of \mathcal{P} (support of π). Such a case is illustrated in Figure 10.1 for parameter θ . Of course, this can be avoided by adapting $q(\boldsymbol{\theta}_t|\boldsymbol{\theta}_l)$ so that it is truncated to fit the support of π , but in high dimensions this can be tedious, and inefficient in itself. Hence one desires an algorithm which does not sample outside the support of π , without having to truncate q .

Another issue which most sampling algorithms are subject to occurs when the modes of the posterior distribution are far away from each other in $\boldsymbol{\theta}$ space, e.g. when they are at ‘opposite

ends' of the domain of π . In the context of nested sampling this can result in one or more of the modes not being sampled accurately, particularly in the case of low livepoint runs. Thus a sampling algorithm should be able to efficiently manoeuvre between well separated modes which lie at the 'edges' of π 's support.

Geometric nested sampling attempts to solve these two issues by interpreting parameter values as points on geometric objects, namely on circles, tori and spheres.

10.3.1 Wrapping the trial distribution

A relatively straightforward way of ensuring that the trial points sampled from q are in the support of π is to 'wrap' q . This is illustrated in Figure 10.2, where we consider a one-dimensional uniform prior on $[0, 1]$. For any point θ , there will be a non-zero probability of sampling a value of θ' from the trial distribution $q(\theta'|\theta)$ that lies outside $[0, 1]$. If the point sampled has a value of say $\theta' = -0.1$, then if we consider q to be wrapped around the support this can be interpreted as sampling a point at value $\theta' = 0.9$. More generally, if θ' is outside the support of π defined by upper and lower bounds θ_{\max} and θ_{\min} it will be transformed as

$$\theta' = \begin{cases} \theta_{\max} - W(\theta') & \text{if } \theta' > \theta_{\max}, \\ \theta_{\min} + W(\theta') & \text{if } \theta' < \theta_{\min}, \end{cases} \quad (10.4)$$

where

$$W(\theta) = \begin{cases} (\theta - \theta_{\max}) \bmod (\theta_{\max} - \theta_{\min}) & \text{if } \theta > \theta_{\max}, \\ (\theta_{\min} - \theta) \bmod (\theta_{\max} - \theta_{\min}) & \text{if } \theta < \theta_{\min}. \end{cases} \quad (10.5)$$

Assuming the support of π is connected (a requirement of nested sampling, as stated in Section 10.1), then this operation will be well defined for all π with bounded supports, of arbitrary dimension. Using this transformation does not affect the argument symmetry of q , thus the value of α given by equation 10.2 still holds. Furthermore, this symmetry ensures that the detailed balance relation given by equation 9.14 is still satisfied.

10.3.2 Circular parameters

As well as ensuring that none of the sampled trial points lie outside the support of π , the wrapped trial distribution can also improve the manoeuvrability of the sampling process, since the trial point chain can always 'move in either direction' without stepping outside of the support of π . This proves to be particularly useful for 'circular parameters'. Here I define circular parameters

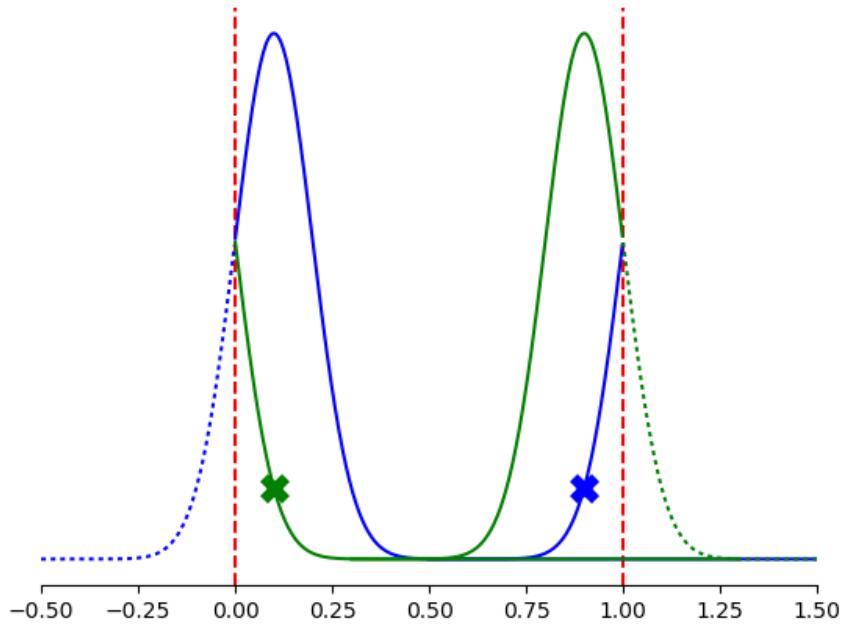


Figure 10.2: Wrapped trial distributions. The solid blue curve represents a Gaussian trial distribution $q(\theta'|\theta)$ as in previous Figure, but now incorporating the wrapping methodology. As a result of the wrapping, θ' (blue cross) is at 0.9, and so won't be automatically rejected by the Metropolis algorithm. The green curve shows the same trial distribution $q(\theta|\theta')$ centred on 0.9. The fact that $\theta = 0.1$ (green cross) is sampled from $q(\theta|\theta')$ with the same probability as θ' is from $q(\theta'|\theta)$ shows that the wrapped trial distribution is still symmetric with respect to its arguments (provided $q(a|b)$ is a symmetric function about the point b).

to be those whose value at θ_{\max} and θ_{\min} correspond physically to the same point. Examples of circular parameters include angles (which are circular at e.g. zero and 2π) and time periods (e.g. 00:00 and 24:00). Often, circular parameters have probability distributions associated with them which are also circular. An example of a circular distribution is the von Mises distribution, an example of which is shown in Figure 10.3 (and defined in Section 10.4.1). This particular example shows that the function's peak(s) may be split by the wrapping, so that when plotted linearly, they appear to have 'half peaks' about θ_{\max} and θ_{\min} . Such half peaks would be classified as two separate peaks by clustering nested sampling algorithms. Thus in general, the number of livepoints would need to be increased to accommodate for the higher number

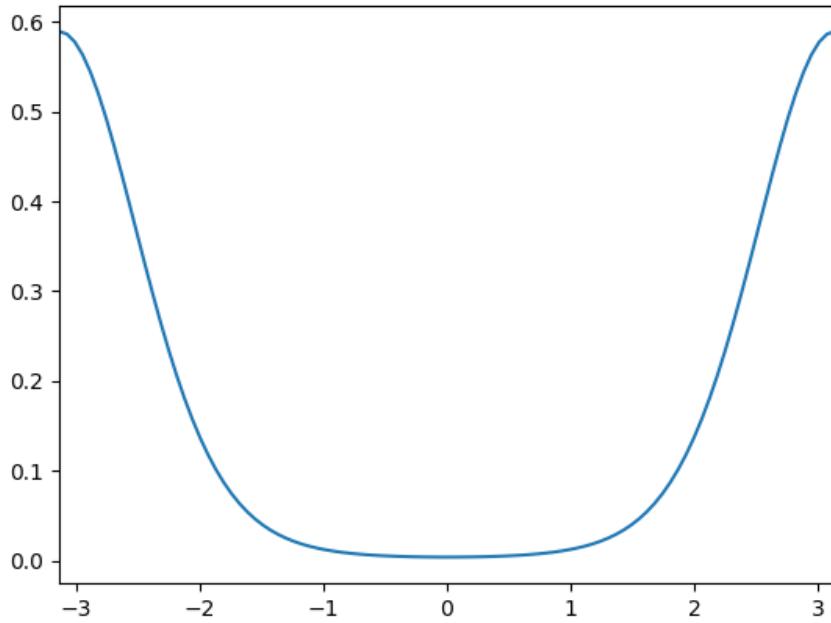


Figure 10.3: von Mises distribution with domain $[-\pi, \pi]$, centred on π . The peak wraps around at edges of domain, so that it appears as two half peaks on a linear space.

of modes, to ensure both half peaks are sampled adequately without one cluster ‘dying out’. Furthermore, the two half peaks occur at opposite ends of the domain of a linear space, making it more difficult for a sampler to explore the regions of higher probability efficiently. The wrapped trial distribution resolves both of these issues, as the two half peaks in linear space are treated as one full peak as far as the sampling (and allocation of livepoints) is concerned. Consequently, the second issue of the half peaks being far away from each other is automatically eradicated. The wrapped trial distribution methodology can thus be applied to problems which involve sampling on non-Euclidean spaces. I apply the method to toy models with distributions defined on circles and tori in Sections 10.4.1 and 10.4.2 respectively. Furthermore, I apply the methodology to a practical example in Section 10.4.4.

10.3.3 Variance of the trial distribution

As with any sampling procedure which relies on a trial distribution, picking a variance for the distribution is difficult without a-priori knowledge of the posterior distribution you are sampling from. A low variance results in a lot of trial points being accepted, but a high auto correlation between these points. A high variance gives a lot of trial rejections, but when these points are accepted, their correlation with the starting point is often low. Since picking the trial variance can in itself be a mammoth task, I use a simplistic approach and take it to be

$$0.1 \times \left| \max_{\text{livepoints}} (\theta_i) - \min_{\text{livepoints}} (\theta_i) \right|, \quad (10.6)$$

for each component i of $\boldsymbol{\Theta}$. I use this approach to avoid the sampler from taking large steps when the livepoints are close together. However, I acknowledge that this method is far from optimal when the livepoints are compactly located at the edges of the domain of $\mathcal{P}(\boldsymbol{\Theta})$.

10.3.4 Non-Euclidean sampling via coordinate transformations

The wrapped trial distribution introduced in Section 10.3.1 can in theory be used in Metropolis nested sampling to sample effectively from circular and toroidal spaces parameterised in terms of circular variables. However, it is not particularly effective at sampling from spherical spaces, since wrapping around the zenith angle (usually defined on $[0, \pi]$) would result in discontinuous jumps between the poles of the sphere. One could of course just wrap the trial distribution in the dimension representing the azimuthal angle (usually defined on $[0, 2\pi]$), rather than in both angles. However, this would re-introduce the issues stated in Section 10.3, i.e. wasting samples and inefficient exploration of the parameter space. I therefore propose an alternative method for exploring spherical spaces which I incorporate in the geometric nested sampling algorithm.

10.3.5 Spherical coordinate transformations

Assuming the surface of a unit sphere is parameterised by azimuthal angle ϕ on $[0, 2\pi]$ and zenith angle θ on $[0, \pi]$, then the corresponding Cartesian coordinates are

$$\begin{aligned} x &= r \cos(\phi) \sin(\theta), \\ y &= r \sin(\phi) \sin(\theta), \\ z &= r \cos(\theta), \end{aligned} \quad (10.7)$$

with $r = 1$. Note that ϕ is the angle measured anti-clockwise from the positive x -axis in the x - y plane and θ is the angle measured from the positive z -axis. Thus a trial point ϕ_t, θ_t can be sampled

as follows. Starting from a point ϕ_l, θ_l , calculate x_l, y_l, z_l , from which a trial point x', y', z' can be sampled from $q(x', y', z' | x_l, y_l, z_l)$. We use a three-dimensional spherically symmetric Gaussian distribution for $q(x', y', z' | x_l, y_l, z_l)$. In general, the point x', y', z' will not lie on the unit sphere. Nevertheless the point is implicitly projected onto it by solving the equations given by 10.7 simultaneously for ϕ and θ , where we set $x = x'$, $y = y'$, $z = z'$, and $r = r'$ (see Figure 10.4). The resulting values are ϕ_t and θ_t , from which the acceptance ratio given by equation 10.2 can be evaluated as normal. There are a few things to note about sampling the trial point in the Cartesian space. Firstly, for equation 10.2 to hold we must have $q(\phi_t, \theta_t | \phi_l, \theta_l) = q(\phi_l, \theta_l | \phi_t, \theta_t)$, which is equivalent to

$$\int_{\mathbf{x}' \in \{\mathbf{x}_{t,\phi,\theta}\}} q(\mathbf{x}' | \mathbf{x}) d\mathbf{x}' = \int_{\mathbf{x} \in \{\mathbf{x}_{l,\phi,\theta}\}} q(\mathbf{x} | \mathbf{x}') d\mathbf{x}, \quad (10.8)$$

where $\mathbf{x}' = (x', y', z')$ and $\mathbf{x} = (x, y, z)$. $\{\mathbf{x}_{t,\phi,\theta}\}$ are the set of Cartesian coordinates which satisfy 10.7 for $\phi = \phi_t$, $\theta = \theta_t$, & all $r \neq 0$. Similarly $\{\mathbf{x}_{l,\phi,\theta}\}$ are the \mathbf{x} which satisfy 10.7 for $\phi = \phi_l$ & $\theta = \theta_l$ (see Figure 10.4). Due to the symmetry of the spherical coordinate system, these sets of vectors lie along the lines given by (ϕ_t, θ_t) and (ϕ_l, θ_l) respectively. The only additional requirement for equation 10.8 to hold is that $q(x', y', z' | x, y, z)$ is symmetric in its arguments, which it is provided that $q(a|b)$ is a symmetric function about the point b . As in Section 10.3.1, the symmetry of the trial distribution ensures that the detailed balance relation given by equation 9.14 is still satisfied.

Sampling in Cartesian coordinates eliminates the risk of sampling points which are automatically rejected (due to being outside the support of $\pi(\phi, \theta)$) to a negligible level, since the only points in Cartesian coordinates which are ill-defined in spherical coordinates are $x = y = 0$ for all z . How the coordinate transformation improves the manoeuvrability of the sampler relative to sampling in the original parameter space is less clear-cut. For the latter, when the variance is fixed the step sizes taken by the sampler along the surface of the sphere depend on where you start from. For example, at $\theta \approx 0$, large moves in ϕ will result in relatively small steps along the sphere whereas at $\theta \approx \pi/2$ such moves in ϕ would result in large steps along the sphere. However when sampling in a Cartesian coordinate system, for a constant variance (see below), the trial points sampled will have the same average step size in Euclidean space regardless of the starting point. Furthermore due to the symmetry of a sphere, when the sampled point (x', y', z') is projected back onto the sphere (implicitly when determining ϕ_t and θ_t), the variance of the steps along the sphere is still independent of the starting point. In either the original parameter space or the transformed space, the variance of the trial distribution can be tweaked to adjust

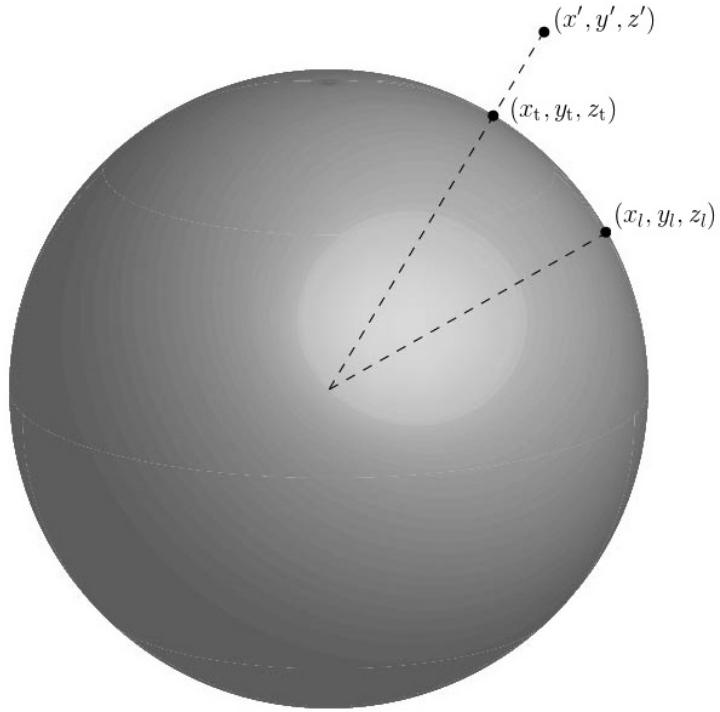


Figure 10.4: Sampling points on the surface of a sphere in Cartesian coordinates. The three-dimensional trial distribution is centred at the point (x_l, y_l, z_l) , which corresponds to (ϕ_l, θ_l) . The point (x', y', z') sampled from q in general will not lie on the surface of the sphere, however the point is implicitly projected onto the sphere at (x_t, y_t, z_t) when calculating (ϕ', θ') [$\equiv (\phi_t, \theta_t)$].

the average step size of the sampler. Nevertheless, it seems more intuitive to me to perform the sampling in the space in which adjusting the variance has an effect which is independent of where you are sampling from.

A spherical distribution is used in the toy model presented in Section 10.4.2, and also features in the gravitational wave detection likelihood function in Section 10.4.4.

10.3.6 Variance of the Cartesian trial distribution

For given variances of ϕ and θ : σ_ϕ^2 & σ_θ^2 , the variance corresponding to a function of these two variables is given by

$$\sigma_f^2 = \left(\frac{\partial f}{\partial \phi} \right)^2 \sigma_\phi^2 + \left(\frac{\partial f}{\partial \theta} \right)^2 \sigma_\theta^2 + 2 \frac{\partial f}{\partial \phi} \frac{\partial f}{\partial \theta} \sigma_{\phi, \theta}, \quad (10.9)$$

where $\sigma_{\phi,\theta}$ is the covariance between ϕ and θ . Hence one can calculate the corresponding variance in Cartesian coordinates, σ_x^2 , σ_y^2 , and σ_z^2 by substituting the equations given by 10.7 into equation 10.9. Using these values for $q(x',y',z'|x,y,z)$ however, leads to an asymmetric trial distribution in its arguments, since the variance is now a function of θ and ϕ . Our entire formulation of the geometric nested sampling algorithm requires q to be symmetric in order for equations 10.2 and 9.14 to hold. Thus we set $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 4/100$ to ensure q is symmetric.

10.3.7 Non-spherical coordinate transformations

The transformation of the trial sampling problem introduced in the previous Section need not be unique to the case of a sphere. Indeed, our implementation of geometric nested sampling includes the option to transform to Cartesian coordinates from circular or toroidal parameters. This is done in the same way as described for the spherical case, but with the relations given by 10.7 replaced with the equivalent transformations for a circle or torus.

10.3.7.1 Circular coordinate transformations

For a parameter which can be interpreted as representing points on a circle e.g. $\phi \in [0, 2\pi]$, we can transform ϕ into the Cartesian coordinates of a unit circle,

$$\begin{aligned} x &= r \cos(\phi), \\ y &= r \sin(\phi), \end{aligned} \tag{10.10}$$

with $r = 1$. A trial point can be sampled as described for the spherical case but working in two dimensions instead. The symmetry of a circle ensures that the trial distribution $q(\phi_t|\phi_I)$ is symmetric in its arguments as long as the Cartesian trial distribution $q(x',y'|x,y)$ adheres to the same symmetry. This is indeed true when a circularly symmetric Gaussian distribution is used for $q(x',y'|x,y)$. The circular transformation and sampling process is illustrated in Figure 10.5.

10.3.7.2 Toroidal coordinate transformations

In the case of two parameters representing points on a circle e.g. $\phi \in [0, 2\pi]$ and $\theta \in [0, 2\pi]$, either we can apply separate circular coordinate transformations to each parameter, or we can say that together they parameterise points on the surface of a torus (Figure 10.6). In the latter

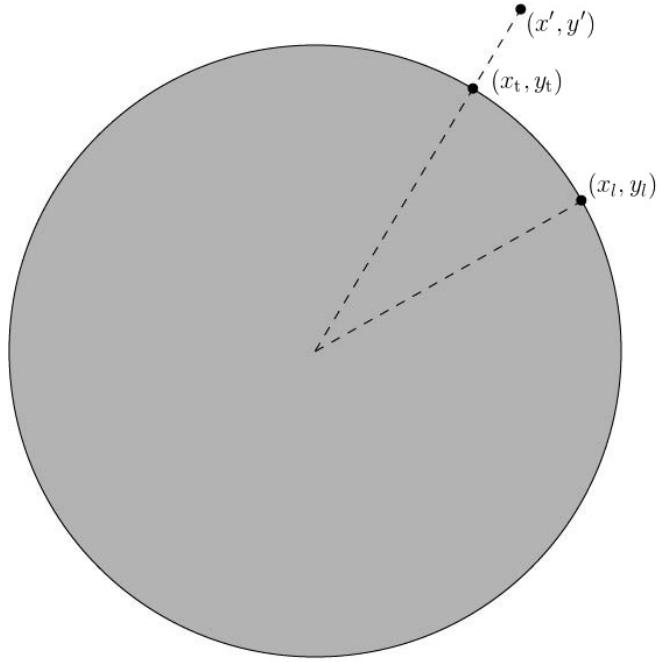


Figure 10.5: Sampling points on the perimeter of a circle in Cartesian coordinates. The two-dimensional trial distribution is centred at the point (x_l, y_l) , which corresponds to ϕ_l . The point (x', y') sampled from q in general will not lie on the perimeter of the circle, however the point is implicitly projected onto it at (x_t, y_t) when calculating $\phi' [\equiv \phi_t]$.

case ϕ and θ can be expressed in terms of Cartesian coordinates through

$$\begin{aligned} x &= (R + r \cos(\theta)) \cos(\phi), \\ y &= (R + r \cos(\theta)) \sin(\phi), \\ z &= r \sin(\theta), \end{aligned} \tag{10.11}$$

where: R is the distance from the centre of the tube to the centre of the torus and r is the radius of the tube; ϕ is the angle between the positive x -axis and the line from the centre of the torus to the point (x, y) , measured anti-clockwise; and θ is the angle between (a) the line in the $x - y$ plane pointing ‘outwards relative to the centre of the torus’ from the centre of the tube, and (b) the line from the centre of the tube to point (x, y, z) (also measured anti-clockwise).

In the case of a torus the Cartesian sampling has an additional complication compared with the circular and spherical cases with regards to $q(\phi_t, \theta_t | \phi_l, \theta_l)$ being symmetric in its arguments.

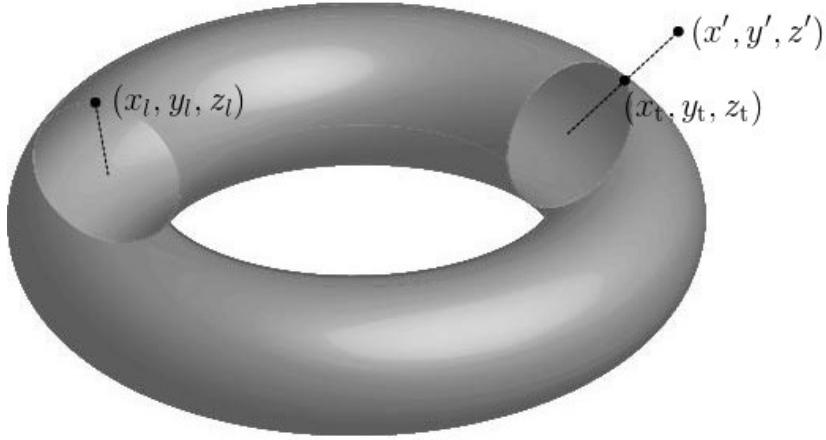


Figure 10.6: Sampling points on the surface of a torus with major and minor radii R and r in Cartesian coordinates. R corresponds to the distance from the centre of the torus (centre of the whitespace in the middle of the grey tube) to the centre of the cross section (depicted as grey circles) of the torus, while r is the radius of the torus' cross section. The three-dimensional trial distribution is centred at the point (x_l, y_l, z_l) , which corresponds to (ϕ_l, θ_l) . The point (x', y', z') sampled from q in general will not lay on the surface of the torus, however the point is implicitly projected onto it at (x_t, y_t, z_t) when calculating (ϕ', θ') [$\equiv (\phi_l, \theta_l)$]. The projection of a general point (x, y, z) onto a torus will be such that the distance between the point and the centre of the torus cross section (corresponding to the point it is projected to) is minimised.

If we first restrict our thinking to the two-dimensional half-plane defined by $\phi = \phi_p$ for arbitrary ϕ_p , the torus maps out a circle with radius r at a distance R from the origin (note that this is just the cross-section of the torus at $\phi = \phi_p$, see Figure 10.7). If we consider sampling (in two dimensions) from a point on this circle, if the sampled point is at $\theta' = \pi$ and the distance between this point and the centre of the circle is $> R$, then the sampled point is not on the half-plane corresponding to ϕ_p but is instead on the one defined by $\phi = \phi_p + \pi$. Consequently when the trial point is projected back onto the torus, it is projected onto a point corresponding to $\phi = \phi_p + \pi$. This implies that there is an asymmetry in the probability of sampling a point which is projected onto the part of the torus corresponding to $\pi/2 < \theta \leq 3\pi/2$ relative to sampling a point which projects onto the part corresponding to $0 \geq \theta \leq \pi/2$ plus $3\pi/2 < \theta \leq 2\pi$; the probability of picking a point in the region given by the latter is higher for an unrestricted trial distribution since the half-plane extends out to infinity. This can be avoided by restricting the range in which (x', y', z') is sampled from such that the shortest distance between the point (x', y', z') and the centre of the tube of the torus is $\leq R$. This ensures that for a symmetric q the

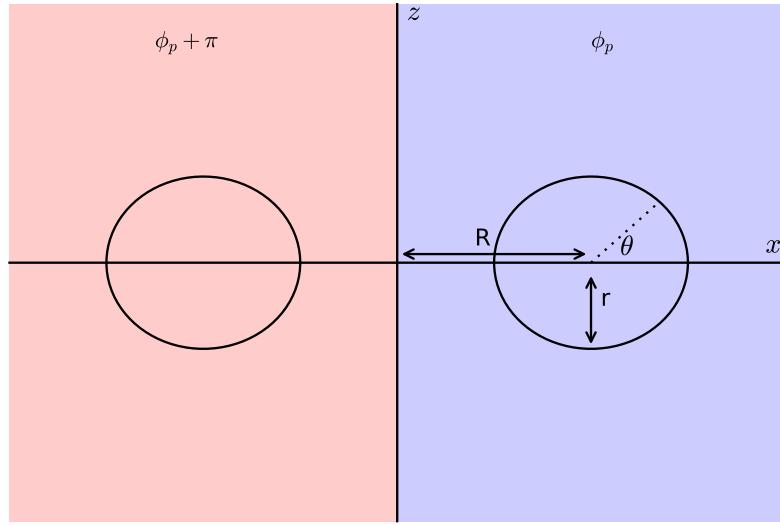


Figure 10.7: Torus cross section at $\phi = \phi_p$ and $\phi = \phi_p + \pi$ in $x-z$ plane. If the sampled point lies in the half-plane (shaded blue) defined by $\phi = \phi_p$, it will be projected onto the circle in this half-plane, otherwise it will be projected onto the circle in the $\phi = \phi_p + \pi$ half-plane (shaded pink).

probability of sampling a point from the range $\pi/2 < \theta \leq 3\pi/2$ is the same as from $0 \geq \theta \leq \pi/2$ plus $3\pi/2 < \theta \leq 2\pi$, and thus $q(x', y', z' | x, y, z)$ is symmetric in its arguments for fixed ϕ .

A similar thought experiment can be applied to the case when θ is fixed and ϕ is allowed to vary. For arbitrary θ this maps out two-dimensional surfaces in the three-dimensional sampling space, for which the restricted sampling stated above results in q being symmetric in its arguments as long as θ remains fixed.

When varying ϕ and θ simultaneously during (three-dimensional) sampling (as you would in the real implementation of the algorithm) there is no trivial way to truncate the trial distribution to ensure $q(x', y', z' | x, y, z)$ is symmetric in its arguments. Thus one is required to evaluate the set of integrals given by 10.8 (but over integration domains which satisfy 10.11 for given ϕ_l, θ_l and ϕ_t, θ_t) to determine $q(\phi_l, \theta_l | \phi_l, \theta_l)$ and $q(\phi_l, \theta_l | \phi_t, \theta_t)$. Using the truncated trial distribution (introduced when considering fixed ϕ) with the variance stated in Section 10.3.6 I found that $q(\phi_l, \theta_l | \phi_l, \theta_l)$ and $q(\phi_l, \theta_l | \phi_t, \theta_t)$ vary by no more than $O(10^{-6})$ and on average by $O(10^{-8})$.

Given the circular nature of the variables parameterising the points on a circle / torus, I do not think that performing coordinate transformations for these objects will give any advantages over using the wrapped trial distributions in the original parameter spaces. Hence in the applications considered in this thesis, parameters which exhibit circular or toroidal properties will be sampled using the wrapped trial distribution, whilst those of a spherical nature will be sampled using the coordinate transformation methodology. The coordinate transformation methodology can be applied to arbitrary geometries. However geometries which lack symmetry will in general be much more difficult to sample from without breaking the trial distribution symmetry requirement of the Metropolis acceptance ratio. In this case the Metropolis-Hastings acceptance ratio for nested sampling must be used

$$\alpha_{\text{MH}} = \begin{cases} \min \left[\frac{\pi(\boldsymbol{\Theta}_t)q(\boldsymbol{\Theta}_t|\boldsymbol{\Theta}_i)}{\pi(\boldsymbol{\Theta}_i)q(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_t)}, 1 \right] & \text{if } \mathcal{L}_t > \mathcal{L}_i, \\ 0 & \text{otherwise.} \end{cases} \quad (10.12)$$

One can assume that such unsymmetrical geometries mean the integrals associated with calculating the trial distributions distributions in Euclidean space become non-trivial to evaluate. Failure to evaluate equation 10.12 correctly would likely lead to violation of detailed balance which is a *sufficient* condition for a Markov chain to asymptotically converge to the target distribution.

10.4 Applications of geometric nested sampling

I now apply the geometric nested sampling algorithm to models which include circular, toroidal and spherical parameters. I evaluate the algorithm's performance by plotting the posterior samples using CORNER. I also conduct the analysis with the 'vanilla' Metropolis nested sampling algorithm. For circular and toroidal parameters, the vanilla algorithm doesn't use a wrapped trial distribution. In the case of spherical parameters, the vanilla algorithm does not transform to Cartesian coordinates before sampling from the trial distribution. For further comparison, I calculate posterior samples using MULTINEST (Feroz, Hobson, & Bridges 2009) (i.e. the algorithm I have used for all Bayesian inferences done in the preceeding Chapters), a state of the art clustering nested sampling algorithm, effective in low dimensional problems.

I refer to the samples / distributions obtained from the geometric nested sampler as MG (Metropolis geometric nested sampling), those obtained from the vanilla Metropolis nested sampler as M, and those obtained from MULTINEST as MN.

For all applications I run the algorithms twice, once with a low number of livepoints (50), and once with a high number of livepoints (500).

10.4.1 Toy model I: circular distribution

I first consider the problem of a one-dimensional circular distribution from which we would like to sample from. The model is parameterised by one variable ϕ , which is defined on $[0, 2\pi]$. Referring back to Section 10.1 I take $\pi(\phi)$ to be uniform on $[0, 2\pi]$. For the likelihood function, I use the von Mises distribution introduced in Section 10.3.2 and defined by

$$\mathcal{L}(\phi | \mu, \sigma^2) = \frac{\exp(\cos(\phi - \pi - \mu)/\sigma^2)}{2\pi I_0\left(\frac{1}{\sigma^2}\right)}, \quad (10.13)$$

where μ and σ are the mean and standard deviation of the distribution, and $I_0(x)$ is the zeroth order modified Bessel function. Here I set $\mu = 0$ so that the peak of the posterior distribution is wrapped around $[0, 2\pi]$, and appears as two half peaks. I set the variance equal to 0.25.

Since the problem involves the circular parameter ϕ , the geometric nested sampling algorithm uses a wrapped trial distribution.

10.4.1.1 Low livepoint runs

Figure 10.8 shows the posterior distribution obtained for toy model I from the three samplers using a low number of livepoints. Note that the Figure also includes a curve plotted from samples which were obtained by evaluating the posterior distribution analytically over a uniform range of ϕ values. I refer to this curve as the theoretical (T) result. The three samplers obtain similar results in the central bins where the probability density is low. However the distributions become asymmetric towards the edges of the domain when compared with the T curve. Overall the MG and MN samplers marginally outperform the M sampler, given the latter has a large asymmetry between the first ($\phi \approx 0$) and final ($\phi \approx 2\pi$) bins.

10.4.1.2 High livepoint runs

Figure 10.9 shows the results when a high number of livepoints is used for the nested sampling algorithms. The plot shows that all three algorithms do a much better job of replicating the T curve than when they were used with a low number of livepoints, with the MG samples giving the curve most similar to the T result.

10.4.2 Toy model II: toroidal distribution

I next consider a two-dimensional problem where each parameter is circular. I refer to this as a toroidal model, as it is equivalent to sampling from the surface of a torus parameterised by

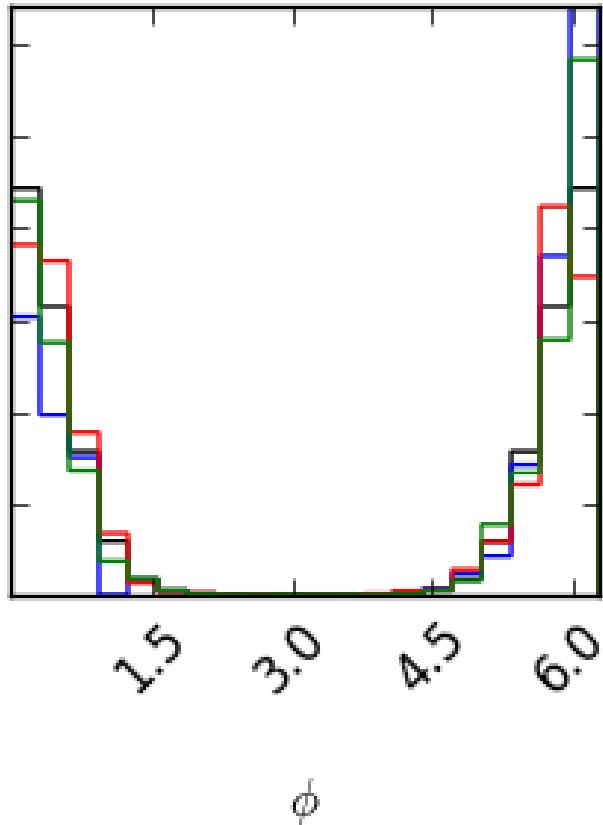


Figure 10.8: Posterior distributions of the circular toy model defined in Section 10.4.1. The black curve corresponds to samples obtained from the analytical expression for $\mathcal{P}(\phi)$ evaluated over a range of ϕ values. The blue, red and green curves correspond to the samples obtained from the M, MG and MN algorithms respectively. All three samplers were run with 50 livepoints.

two angles ϕ and θ . We take both $\pi(\phi)$ and $\pi(\theta)$ to be uniform on $[0, 2\pi]$. For the likelihood function, I again use the von Mises distribution, and take the likelihood functions for ϕ and θ to be independent so that

$$\mathcal{L}(\phi, \theta | \mu_\phi, \sigma_\phi^2, \mu_\theta, \sigma_\theta^2) = \mathcal{L}(\phi | \mu_\phi, \sigma_\phi^2) \mathcal{L}(\theta | \mu_\theta, \sigma_\theta^2), \quad (10.14)$$

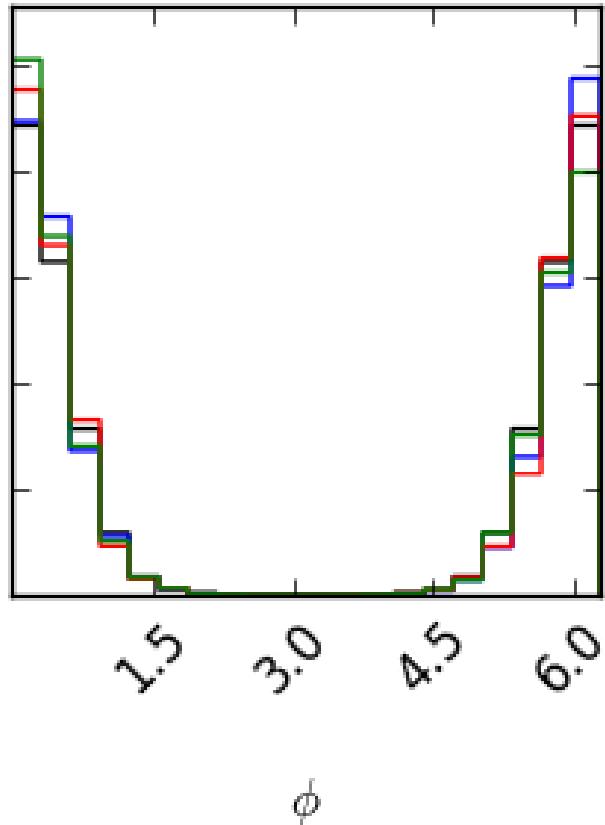


Figure 10.9: Posterior distributions of the circular toy model with the number of livepoints set to 500. The colour coding of the plot is as described in Figure 10.8.

where the likelihood for each individual parameter takes the form of equation 10.13. I set $\mu_\phi = \mu_\theta = 0$ so that the two-dimensional posterior contains four ‘quarter peaks’ at the corners of its domain. I also take $\sigma_\phi^2 = \sigma_\theta^2 = 0.25$.

Since this model represents a toroidal distribution (or two circular distributions), the geometric nested sampling algorithm uses wrapped trial distributions to sample ϕ and θ .

10.4.2.1 Low livepoint runs

Figure 10.10 shows the posterior distributions obtained for toy model II from the three samplers using a low number of livepoints. As in Section 10.4.1, samples of the analytical posterior are included for comparison. Looking at the one-dimensional marginalised posteriors for ϕ and θ , the M algorithm does a poor job at recovering the true distribution, overestimating the half peaks at low values of ϕ, θ and overestimating them at high ϕ, θ . The MG algorithm does a relatively good job of replicating the T distribution, and looking at the marginalised posteriors, outperforms MN at three of the four half peaks (MN does better at the $\theta \approx 0$ peak). One may expect MN to struggle with such a distribution, using a low number of livepoints. Since the four quarter peaks will appear to a clustering algorithm as four separate peaks, MN will on average assign 12.5 livepoints to each of these peaks, which may not be enough to sample each peak adequately. The MG algorithm on the other hand treats these four quarter peaks as one, and so you would expect it to be able to use all 50 livepoints to sample this peak relatively well.

10.4.2.2 High livepoint runs

The high livepoint run results for the toroidal distribution are shown in Figure 10.11. All three samplers recover the true distribution well, with the M and MG giving marginally better results than MN. This is perhaps surprising since one would expect MN to easily be able to cope with four modes using 500 livepoints. It does however, highlight the possibility that it is not the number of peaks that MN is struggling with, it is their shape that is causing it to underperform relative to the other two samplers.

10.4.3 Toy model III: spherical distribution

For the final toy model I consider the posterior distribution of two angles which parameterise the surface of a sphere. As in Section 10.3.5, ϕ and θ represent the azimuthal and zenith angles respectively. I take $\pi(\phi)$ to be uniform on $[0, 2\pi]$, and $\pi(\theta)$ to be uniform on $[0, \pi]$. I use a von Mises distribution for $\mathcal{L}(\phi | \mu_\phi, \sigma_\phi^2)$ with $\mu_\phi = 0$ and $\sigma_\phi^2 = 0.25$. For $\mathcal{L}(\theta | \mu_\theta, \sigma_\theta^2)$ I use a truncated Gaussian (defined on $[0, \pi]$) with $\mu_\theta = \pi/2$ and $\sigma_\theta^2 = 0.25$.

For this model the geometric nested sampling algorithm uses the spherical transformation sampling procedure detailed in Section 10.3.5 to sample ϕ_t and θ_t .

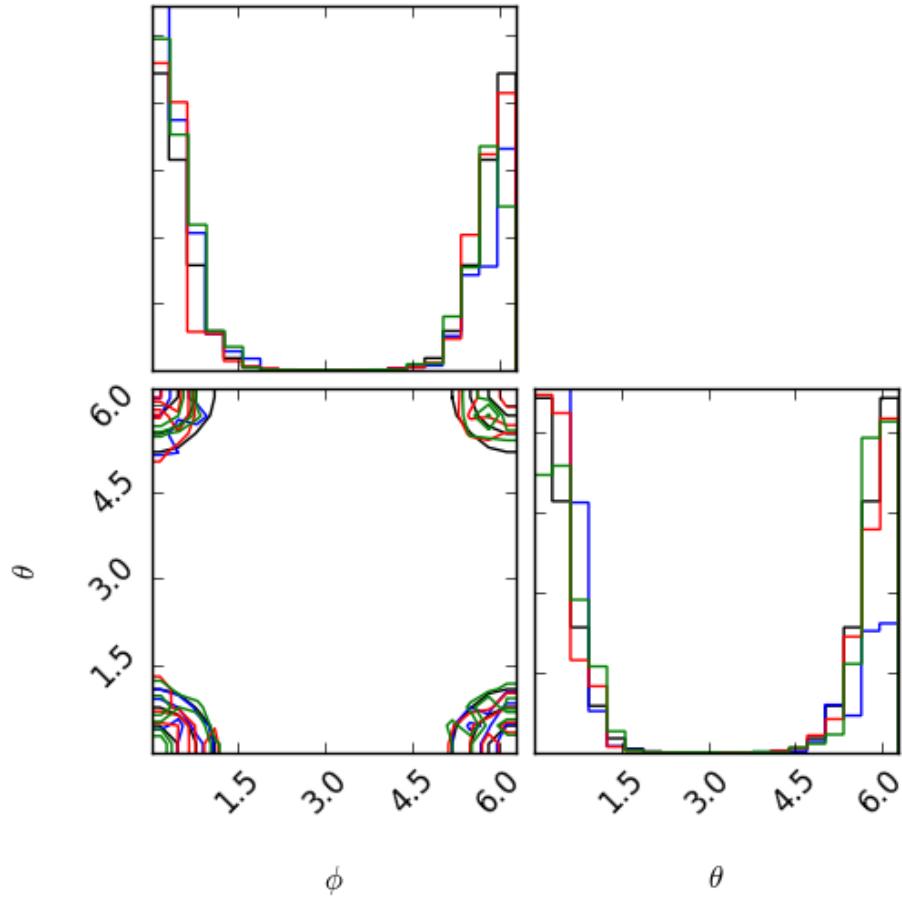


Figure 10.10: Posterior distributions of the toroidal toy model defined in Section 10.4.2, with the number of livepoints set to 50. The colour coding is as described in Figure 10.8. The plots along the diagonal show the one-dimensional marginalised posteriors for ϕ and θ . The centre plot shows the joint two-dimensional posterior.

10.4.3.1 Low livepoint runs

Figure 10.12 shows the posterior distributions obtained for toy model III from the three samplers using a low number of livepoints, plus the T samples. The circular distribution of ϕ is well recovered by the M and MG algorithms, but less so by MN. All three samplers do a relatively poor job of recovering the truncated Gaussian distribution of θ , with M probably giving the best results due to the symmetry of its distribution.

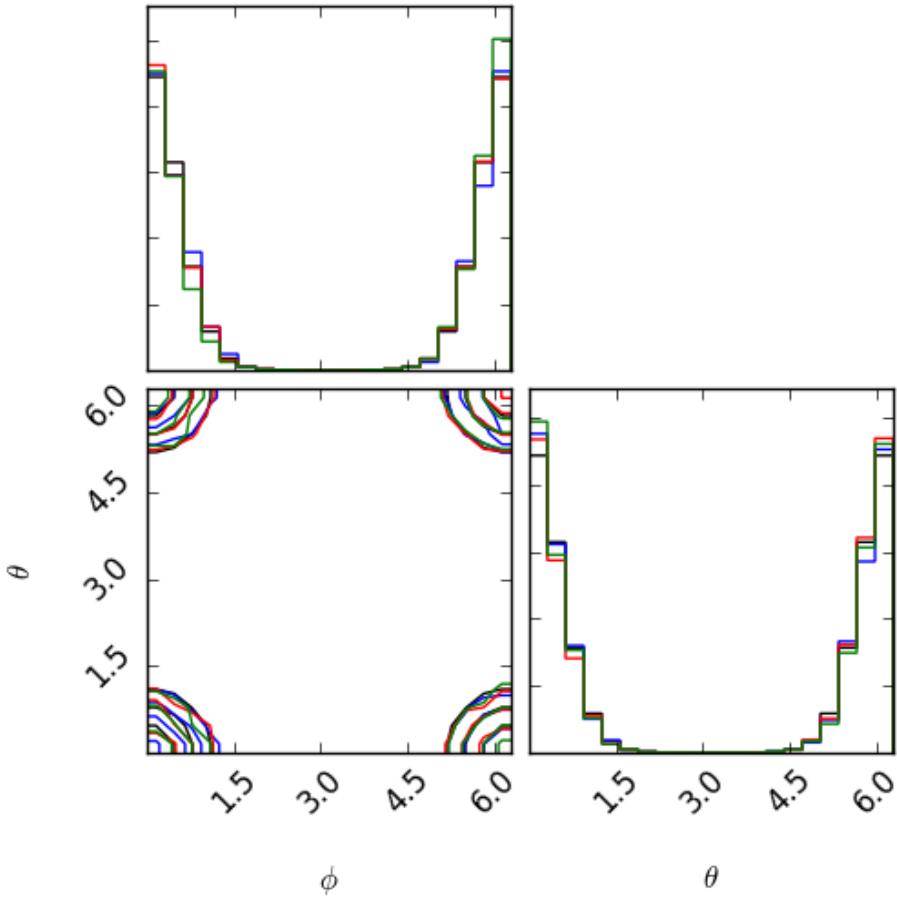


Figure 10.11: Posterior distributions of toroidal toy model with the number of livepoints set to 500. The colour coding and layout of the plots is as explained in Figure 10.10.

10.4.3.2 High livepoint runs

When 500 livepoints are used for the samplers (Figure 10.13), the MG and MN algorithms recover the ϕ profile similarly well. However, the MG sampler seems to slightly overestimate $\mathcal{P}(\theta)$ at high probability densities, and underestimate it to a similar extent at low densities.

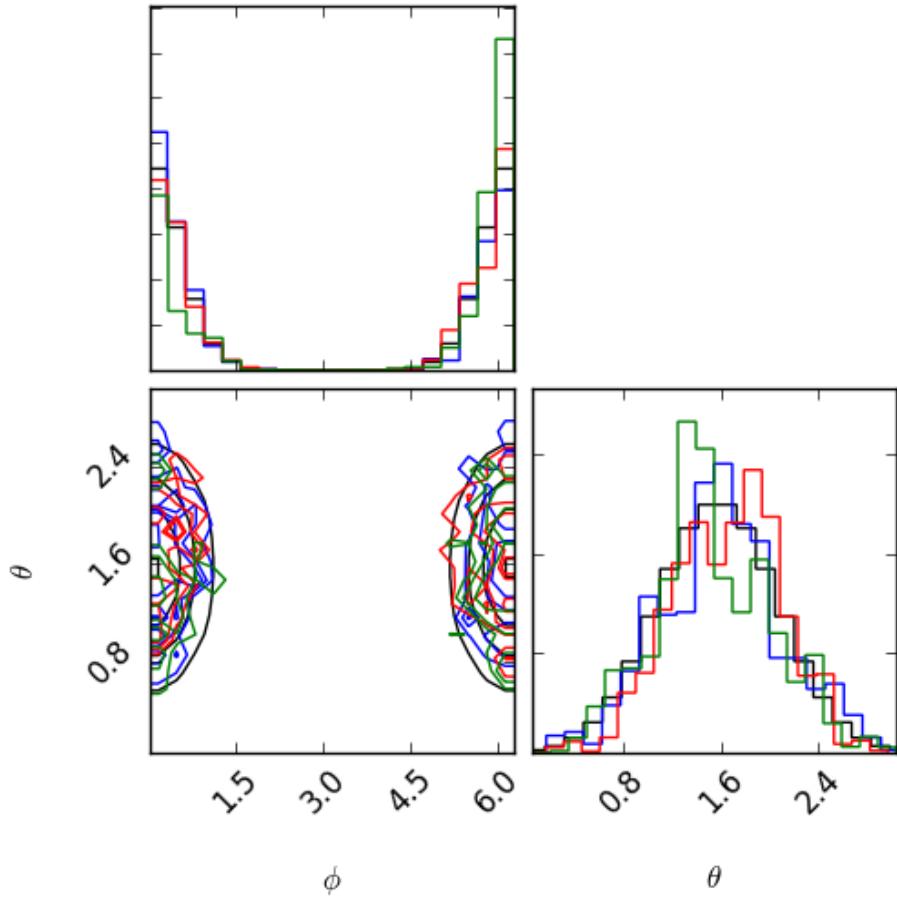


Figure 10.12: Posterior distributions of the spherical toy model with the number of livepoints set to 50. The colour coding and layout of the plots is as explained in Figure 10.10.

10.4.4 Practical example: gravitational wave emission from binary black hole mergers

I now consider a likelihood function which corresponds to detecting gravitational waves from (binary) black hole mergers. The data for the likelihood are obtained from the LIGO^a and Virgo^b interferometers (see e.g. Abbott et al. 2016 and Lange & LIGO-Virgo Collaboration

^a<https://www.ligo.caltech.edu/page/ligo-gw-interferometer>.

^b<http://www.virgo-gw.eu/>.

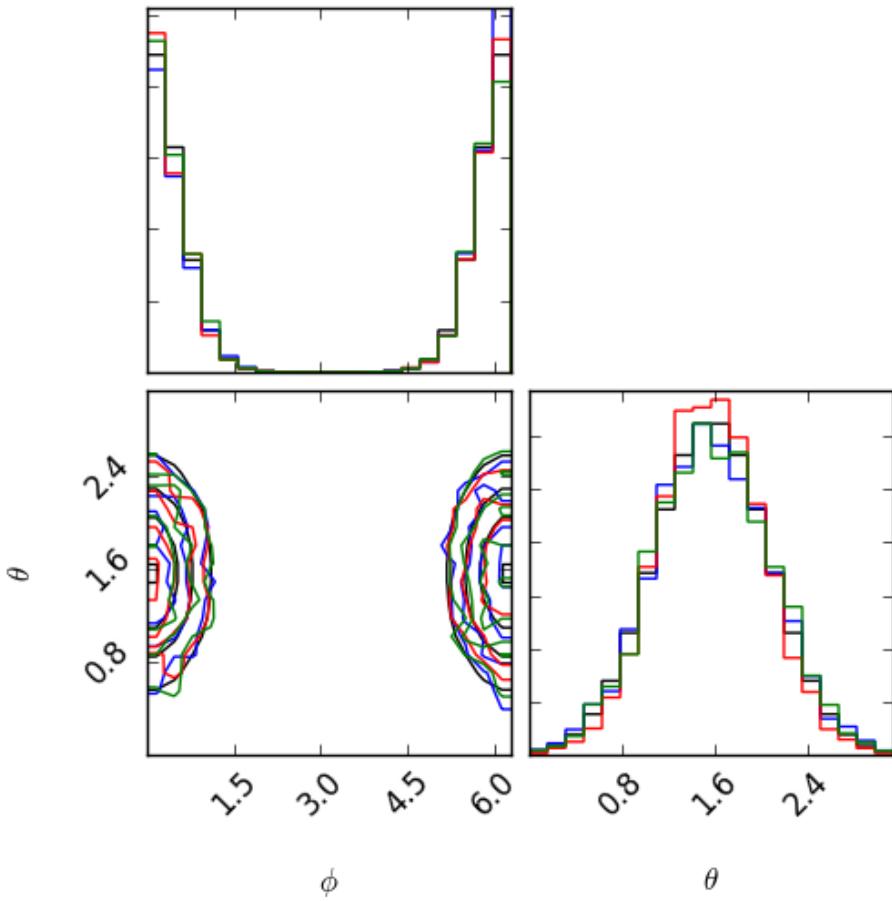


Figure 10.13: Posterior distributions of the spherical toy model with the number of livepoints set to 500. The colour coding and layout of the plots is as explained in Figure 10.10.

2018). I now give a brief overview on gravitational waves and how they are detected, but for more thorough analysis see e.g. Hobson, Efstathiou, & Lasenby (2006), Kokkotas (2008), or Blanchet (2014).

10.4.4.1 Origin of gravitational waves

For an observer lying in a region of spacetime satisfying the Minkowski metric $\eta_{\mu,\nu}$, fluctuations in the metric can be described by a linear perturbation

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (10.15)$$

where the perturbations are assumed to be small ($|h_{\mu\nu} \ll 1|$). By solving Einstein's field equations using the metric given by equation 10.15, it can be shown that the tensor which represents the gravitational field

$$\tilde{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h_{\alpha}^{\alpha}, \quad (10.16)$$

satisfies the wave equation for a vacuum, and hence has a solution

$$\tilde{h}_{\mu\nu} = A_{\mu\nu} \exp(ik^{\alpha}x_{\alpha}), \quad (10.17)$$

where $A_{\mu\nu}$ describe the wave's polarisation and amplitude. It can be shown that by setting an appropriate gauge (the Transverse-Traceless gauge) that $A_{\mu\nu}$ can be defined in terms of two polarisation states h_+ and h_{\times} . For a gravitational wave travelling in the z direction, the tensor h_+ causes simultaneous expansion (contraction) in the x direction and contraction (expansion) in the y direction. h_{\times} acts similarly at an angle $\pi/4$ to the x - y axes.

Exact solutions of Einstein's field equations have not yet been found, leading to the development of analytic approximations such as the Post-Newtonian (PN) approximation (see e.g. [Asada & Futamase 1997](#)) to determine h_+ and h_{\times} . Here we consider the PN approximation up to second order for inspiralling black hole binary systems as described in [Blanchet et al. \(1996\)](#).

10.4.4.2 Detection of gravitational waves

Laser beam interferometers such as LIGO and Virgo detect gravitational waves by measuring the differential arm length between perpendicular arms of the interferometers. The differential measured is proportional to the gravitational strain h , which describes the fractional change in proper space caused by the gravitational perturbation. h can be written as a linear combination of the two polarisation states h_+ and h_{\times}

$$h(t) = F_+h_+(t) + F_{\times}h_{\times}(t), \quad (10.18)$$

where t denotes the time at which the strain is measured, and F_+ & F_{\times} are functions dependent on the geometry of the detector. Here we consider three detectors: LIGO Hanford, LIGO Livingston and Virgo. The geometries used in this analysis for these detectors can be found at <https://www.ligo.org/scientists/GW100916/GW100916-geometry.html>.

10.4.4.3 Likelihood function for gravitational wave detection

Assuming we have n_d data points $\{x_{i,j}\}$ recorded at times $\{t_i\}$ for each detector j , then the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\Theta}) = \prod_{i=1}^{n_d} \prod_{j=1}^3 \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_{i,j} - h_j(t_i, \boldsymbol{\Theta}))^2}{2\sigma^2}\right), \quad (10.19)$$

where $h_{i,j}(t_i, \boldsymbol{\Theta})$ is the theoretical strain and is dependent on the model parameters $\boldsymbol{\Theta}$ (defined below). In the analysis presented here we consider data which are simulated by evaluating $h_j(t_i, \boldsymbol{\Theta})$ for fixed model parameters (say $\boldsymbol{\theta}_r$), i.e. we set

$$x_{i,j} \equiv h_{j,\boldsymbol{\theta}_r}(t_i). \quad (10.20)$$

Furthermore, we set $\sigma = 1 \times 10^{-21}$ and $n_d = 1000$.

10.4.4.4 Model parameters

$\boldsymbol{\Theta}$ is a nine-dimensional vector with components

$$\boldsymbol{\Theta} = (m_1, m_2, r, t_c, \phi_c, \phi, \theta, p, i) \quad (10.21)$$

here m_1 and m_2 are the masses of the individual black holes, r is the luminosity distance to the centre of the binary system, and t_c is the time of coalescence of the two black holes (i.e. the time at which they merge). ϕ_c is the orbital phase of the binary system at time t_c (and is defined on $[0, 2\pi]$), and ϕ & θ are the angular location of the merger system in the sky (as observed from a detector). The inclination angle i is the angle between the line of sight from the binary system to a detector, and the normal to the orbital plane. The normal is chosen to be right-handed with respect to the sense of motion so that i is defined on $[0, \pi]$. p is the corresponding azimuthal angle as observed from the binary system. Table 10.1 gives the values of these parameters used in the simulated data, and how they are sampled using the geometric nested sampler. Notice that I only vary the angular parameters $(\phi_c, \phi, \theta, p, i)$ in the Bayesian analysis, making it a five-dimensional parameter estimation problem. All five parameters are assigned uniform priors over the ranges they are defined on.

Referring back to equation 10.19, the time values t_i are spaced uniformly between $t_c - t$ and $t_c + t$, where

$$t = \frac{1000G(m_1 + m_2)}{c^3}. \quad (10.22)$$

Here G is Newton's gravitational constant and c is the speed of light in a vacuum.

Parameter	Simulation input value	Sampling procedure
m_1	$35 M_{\text{Sun}}$	fixed
m_2	$25 M_{\text{Sun}}$	fixed
r	390 Mpc	fixed
t_c	0	fixed
ϕ_c	0	circular (wrapped trial distribution)
ϕ	0	spherical coordinate transformation (azimuthal angle)
θ	$\pi/2$	spherical coordinate transformation (zenith angle)
p	0	spherical coordinate transformation (azimuthal angle)
i	$\pi/2$	spherical coordinate transformation (zenith angle)

Table 10.1: Gravitational wave detection model parameters, their simulation input values, and how the parameters are sampled by the geometric nested sampler. The parameters ‘fixed’ sampling procedures were not sampled from, instead their true (simulation input) value was used in each evaluation of $h_j(t_i, \boldsymbol{\theta})$. ϕ_c is interpreted as a circular quantity by the geometric nested sampler, and the pairs of angles (ϕ, θ) , & (p, i) are treated as two independent sets of spherical coordinates (and thus are transformed independently).

10.4.4.5 Posterior sampling

For the toy models I calculated the posterior distributions analytically over uniform grids so that I could benchmark the sampling algorithms’ performance with the ‘true’ distributions. However, since we are sampling from a five-dimensional parameter space in this example, obtaining samples analytically is no longer feasible. We thus run the MN algorithm with a very high number of livepoints (2000) and refer to this as the mega MultiNest run (MMN). We use the MMN result as a reference distribution for our low and high livepoint runs of the MG and MN algorithms (we do not include the M algorithm in our comparison here). I note however, in the toy model applications I found evidence to suggest that MN struggles recovering quarter / half peaks even with 500 livepoints, and when sampling from low dimensional & low number of mode models. Thus I can make no guarantees that the MMN distribution is the ‘true’ posterior distribution.

10.4.4.6 Low livepoint run

Figure 10.14 shows the posterior distributions for the angular parameters obtained from the low livepoint run. Looking at the one-dimensional posterior for ϕ , it is clear that MG picks up on the two half peaks at 0 and 2π , but overestimates them compared to the values obtained with MMN. It also underestimates the middle peak ($\phi \approx \pi$) compared to MMN. In fact, one could

argue that it doesn't really infer this peak at all. The MN run does the opposite, it overestimates the middle peak, but completely misses the half peaks. Looking at θ , MG finds a peak around $\theta = \pi/3$, whereas MMN puts the peak at slightly lower θ . The MMN curve shows a flat, high probability density region around $\theta = 2\pi/3$, but MG misses this. The MN run puts the biggest peak at $\theta \approx 2\pi/3$, and a smaller one at $\theta \approx 0.75$. Both MG and MN do a relatively poor job at constraining p correctly, as the former misses the fourth peak present on the MMN curve at $p \approx 3\pi/2$ (and instead overestimates the first peak at $\approx \pi/2$). The MN algorithm more or less gets the correct number of peaks when compared with MMN, but systematically gets their shape wrong. MG does a better job than MN in recovering the distribution of i relative to MMN. MG and MN recover similar profiles for ϕ_c , and roughly get the shape of the distribution correct when comparing with the MMN result.

10.4.4.7 High livepoint run

Figure 10.15 shows the posterior distributions for the angular parameters obtained from the 500 livepoint run. In this case MG and MN do a reasonable job of recovering the MMN profile for ϕ , but still underestimate / overestimate in the same way they did in the low livepoint case. For θ MG does a good job at replicating the MMN result. MG and MN have similar levels of success in recovering the MMN profiles of p , i and ϕ_c .

Overall the MG algorithm performs well relative to MN for the example considered here, given the relative simplicity of the algorithm. To make a statement on which algorithm obtained more accurate inferences of \mathcal{P} for this multi-modal ($\mathcal{O}(10)$ modes), five-dimensional distribution, I believe that a more thorough comparison than the visual inspection conducted here is required. One possible solution to this would be to calculate a distance metric between the MG and MMN posteriors (e.g. Earth mover's distance or the Kullback-Leibler divergence) and compare it with the corresponding value between the MN and MMN distributions. However, I do not make this comparison here. It would also be interesting to see if anyone else has acquired results for this set of simulations, I do not consider this here, however.

10.5 Geometric nested sampling implementation

The implementation of the geometric nested sampler (and the vanilla Metropolis nested sampler) used in this paper, along with the toy models and the gravitational wave likelihood function can be found at https://github.com/SuperKam91/nested_sampling. The algorithm is written in

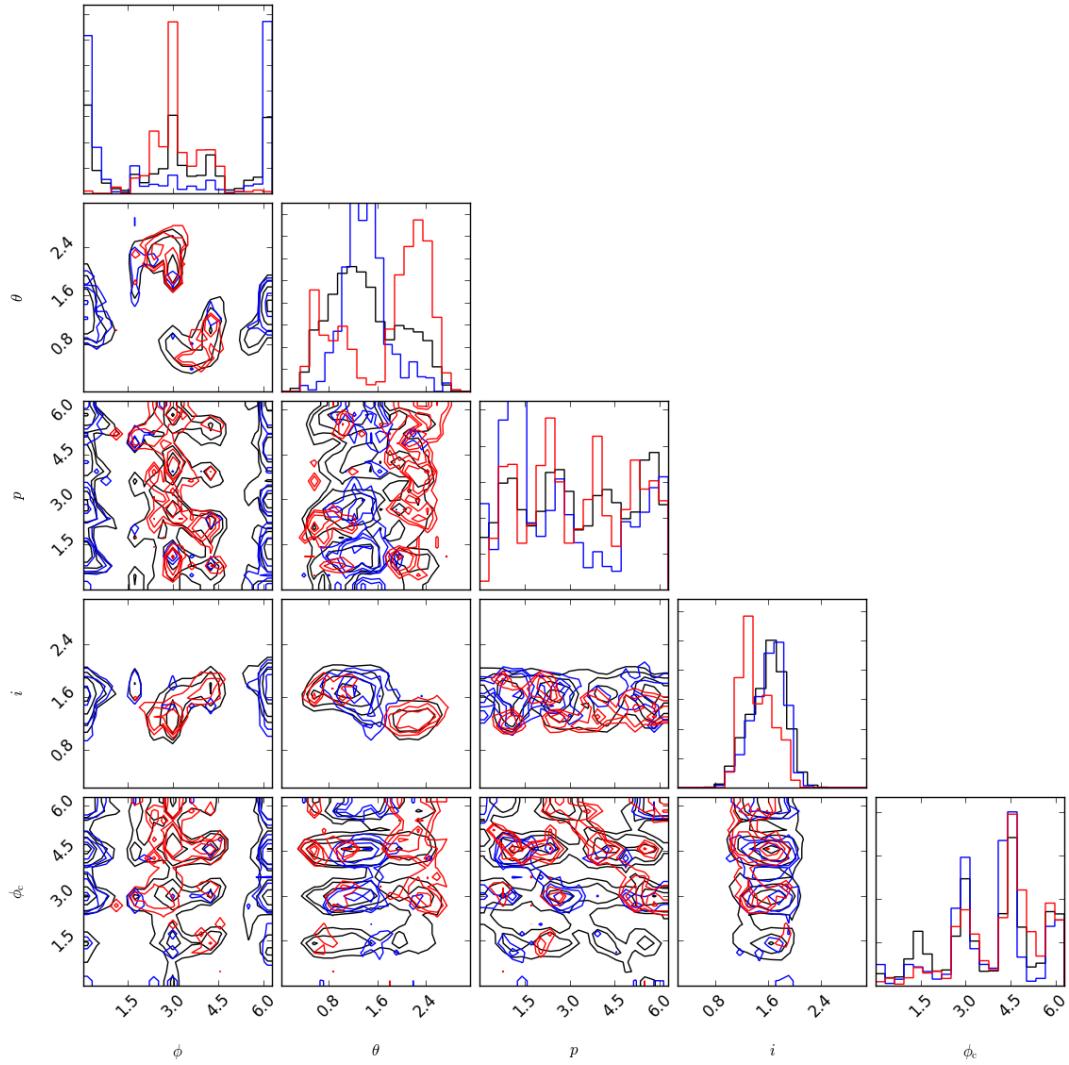


Figure 10.14: Marginalised one- and two-dimensional posterior distributions for the five angular parameters, ϕ_c, ϕ, θ, p , and i . The black curves are the results from the 2000 livepoint MN run. The blue and red curves are plotted using the samples of the MG and MN algorithms respectively, which are obtained from runs with 50 livepoints.

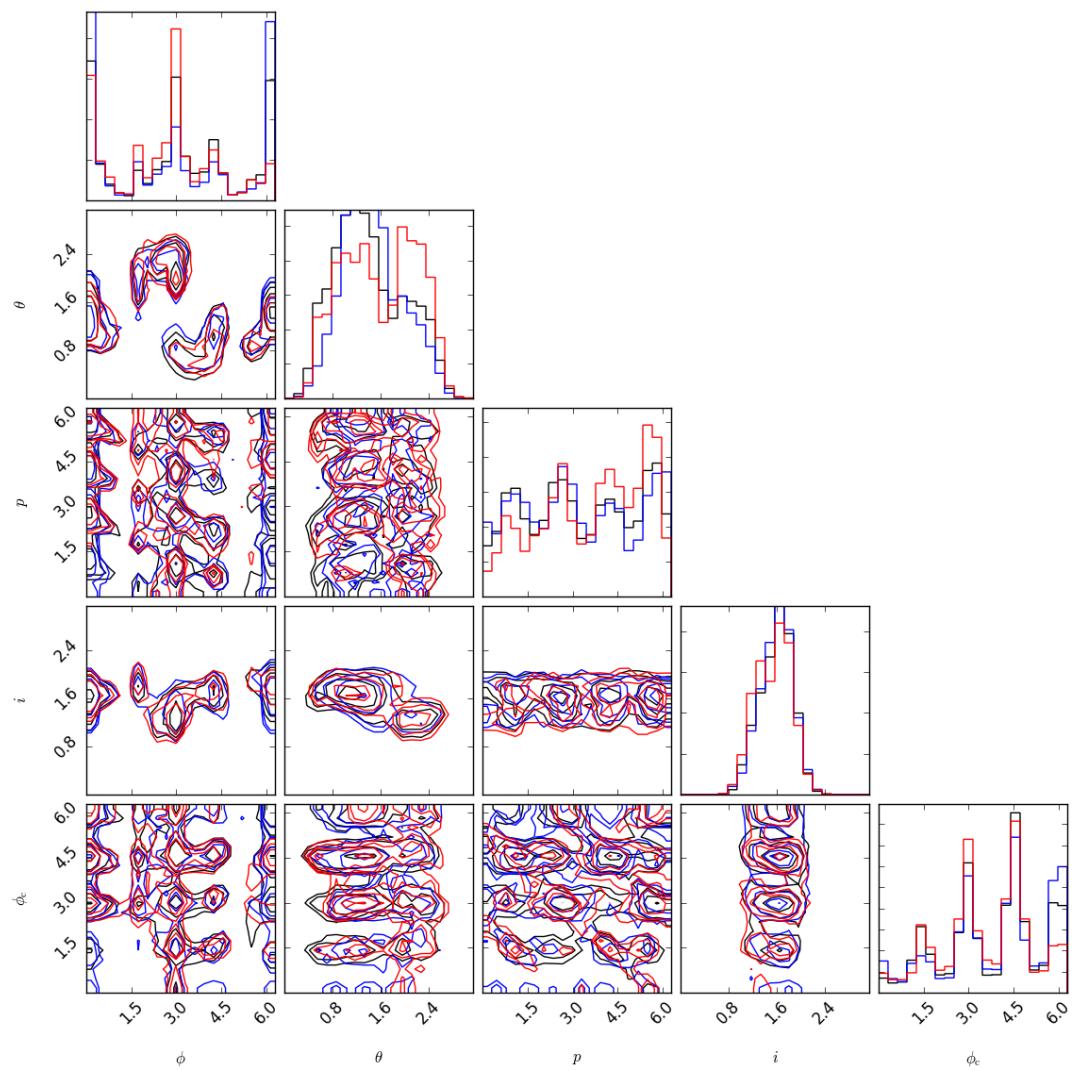


Figure 10.15: Same plot as Figure 10.14 but the blue and red curves show the MG and MN runs with 500 livepoints.

PYTHON 2.7, hence our *implementation* of the algorithm cannot match that of the state of the art nested sampling algorithms such as MULTINEST or POLYCHORD (Handley, Hobson, & Lasenby 2015). These algorithms are implemented in FORTRAN 90, and parallelised using a master-slave paradigm (see Section 5.4 of Handley, Hobson, & Lasenby). Nevertheless there is no reason why geometric nested sampling cannot be implemented more efficiently and parallelised using this method. Furthermore as already mentioned in Section 10.2.1, Feroz & Hobson (2008) incorporate the Metropolis likelihood sampler into a livepoint clustering algorithm. This same idea could be applied to the geometric nested sampling algorithm. However, in the case of circular parameters, the clustering would also need to be wrapped around the domain of \mathcal{P} along with the trial distribution. This could be avoided by instead performing coordinate transformations (Section 10.3.7) for circular and toroidal parameters before sampling from the ellipsoids resultant from the livepoint clustering. The clustering could be performed in either the original parameter or the transformed Euclidean space, but it is important to note that in either case samples could still be automatically rejected if they lie outside the ellipsoid. Nevertheless the algorithm would still provide the benefit of sampling in the ‘natural’ topology of the problem as discussed in Section 10.3.5.

10.6 Conclusions

I have presented a new nested sampling algorithm based on the Metropolis nested sampler proposed in Sivia & Skilling (2006) and applied in Feroz & Hobson (2008). The algorithm exploits the geometric properties of certain kinds of parameters which describe points on circles, tori and spheres, to sample the parameters more efficiently in the context of nested sampling. The algorithm should be more mobile in sampling distributions defined on such geometries.

The algorithm consists of two key sampling modes which can be summarised as follows.

- For circular and toroidal problems, the trial distribution used in the sampling process is wrapped around the support of the prior distribution π (domain of the posterior distribution \mathcal{P}).
- This wrapping ensures that no trial points are automatically rejected when evaluating the Metropolis acceptance ratio as a consequence of the point being outside the sampling space of the model.
- The wrapped trial distribution also makes the sampling more mobile at the edges of the domain of \mathcal{P} , meaning that circular and toroidal distributions should be easier to sample, particularly in the case of posteriors with high probability densities at these edges.

- For spherical problems, parameters specifying the coordinates on a sphere are transformed to Cartesian coordinates and sampled from the corresponding Euclidean space.
- This again ensures that no trial points are automatically rejected because they are outside the domain of \mathcal{P} .
- It also enhances the mobility of the sampler, whose average step size along the surface of the sphere is not dependent on the location at which the trial distribution is centred.

I applied the geometric nested sampling algorithm (MG) to three toy models, which respectively represented models on a circle, torus and a sphere. I compared the posterior plots with those obtained from a ‘vanilla’ Metropolis nested sampler (M) based on the one used in [Feroz & Hobson \(2008\)](#), and with the distributions obtained with the livepoint clustering nested sampling algorithm MULTINEST (MN, [Feroz, Hobson, & Bridges 2009](#)). For each model, all three samplers were run twice, once with a low number of livepoints (50), and once with a high number of livepoints (500). I included the distributions obtained from evaluating \mathcal{P} analytically as means of reference to the ‘correct’ distribution (T). The results can be summarised as follows.

- For the low livepoint run on the circular toy model (von Mises distribution centred on the origin), the MG and MN samplers marginally outperform the M sampler.
- For the high livepoint run on the circular toy model, all samplers perform similarly, with the MG algorithm giving slightly superior results with respect to the T distribution.
- The low livepoint run for the toroidal model (two-dimensional von Mises distribution centred on the origin) the MG outperforms both M and MN. One would maybe expect MN to struggle on a four-mode problem with only 50 livepoints, whereas the MG effectively treats these four modes as one given their location in the domain of \mathcal{P} .
- The high livepoint run gives better results for all three samplers, but the MN distribution seems the least accurate. This highlights the potential issues which clustering algorithms face with modes which occur at the edges of \mathcal{P} , independent of the number of livepoints used.
- The spherical toy model which consists of a von Mises distribution on the azimuthal angle ϕ and a truncated Gaussian on the zenith angle θ shows that in the case of low livepoint runs, the M algorithm surprisingly performs the best, as it does a better job at recovering the profile of θ than the MG algorithm.
- For the 500 livepoint run the MG and MN algorithms recover ϕ similarly well, but the former systematically overestimates the probability density in θ around its peak, and underestimates it at low densities.

I then applied the MG and MN sampling algorithms to a model representing the detection of gravitational waves generated by binary black hole mergers and detected with the LIGO and Virgo instruments (Lange & LIGO-Virgo Collaboration 2018). Using simulated datasets, we obtained inferences of a five-dimensional (all circular / spherical parameters), multi-modal ($O(10)$ modes) posterior distribution. For this example my ‘correct’ reference distribution was a MN run with 2000 livepoints. I found the following.

- Overall for the low livepoint run, both algorithms struggle to correctly infer all the peaks of the distribution (of the 2000 livepoint MN run).
- However, this is to be expected for MN since it can only attribute \approx a few livepoints to each mode. Furthermore, the locations of the modes, which occur not just at the edges of \mathcal{P} , mean that the MG algorithm must also allocate its livepoints separately to different modes, a task which it is not designed to cope well with.
- With 500 livepoints the MG algorithm recovers all the modes inferred from the 2000 livepoint MN run. MN performs similarly well, but slightly overestimates the number of modes; further quantitative work is needed to home in on this.

APPENDIX



**APPENDIX A: RESULTS OF PHYSICAL MODELLING OF
PLANCK CLUSTERS**

A.1 Results table

Table A.1: Summary of values for final sample of 54 clusters. The redshift types correspond to S: spectroscopically measured and P: photometrically measured. z , $M(r_{200})$, x_c , y_c and $f_{\text{gas}}(r_{200})$ are the physical model sampling parameters. $M_{\text{AMI}}(r_{500})$, $M_{\text{Pl,marg}}(r_{500})$ and $M_{\text{Pl,slice}}(r_{500})$ are the $M(r_{500})$ estimates obtained from the AMI and Planck data respectively. All masses are given in units of $\times 10^{14} M_{\text{Sun}}$ and all cluster centre coordinates are measured in arcseconds.

Row	Planck ID	Alias	z	z type	$M_{\text{AMI}}(r_{200})$	x_c	y_c	$f_{\text{gas}}(r_{200})$	$M_{\text{AMI}}(r_{500})$	$M_{\text{Pl,marg}}(r_{500})$	$M_{\text{Pl,slice}}(r_{500})$
1	PSZ2G044.20+48.66	ACO2142	0.0894	S	13.49 ± 2.35	9.14 ± 18.20	8.80 ± 15.08	0.13 ± 0.02	9.25 ± 1.58	10.81 ± 0.42	8.76 ± 0.19 $_{-0.21}^{+0.19}$
2	PSZ2G053.53+59.52	ACO2034	0.113	S	8.51 ± 1.28	-1.80 ± 13.10	19.39 ± 9.86	0.13 ± 0.02	5.87 ± 0.86	5.38 ± 0.39	5.48 ± 0.24 $_{-0.24}^{+0.24}$
3	PSZ2G151.90+11.63	CIZAJ0515.3+5845	0.12	S	5.74 ± 1.24	67.58 ± 27.09	68.01 ± 18.58	0.13 ± 0.02	3.99 ± 0.84	4.23 ± 1.03	3.65 ± 0.47 $_{-0.47}^{+0.50}$
4	PSZ2G218.59+71.31	ACO1272	0.137	S	2.70 ± 0.99	2.82 ± 25.21	-16.62 ± 25.98	0.13 ± 0.02	1.90 ± 0.68	4.79 ± 0.80	3.62 ± 0.30 $_{-0.30}^{+0.30}$
5	PSZ2G226.18+76.79	ACO1413	0.1427	S	8.19 ± 1.23	-35.33 ± 10.98	-1.13 ± 13.44	0.13 ± 0.02	5.62 ± 0.82	6.14 ± 0.55	5.98 ± 0.25 $_{-0.25}^{+0.25}$
6	PSZ2G165.06+54.13	ACO990	0.144	S	7.80 ± 1.35	32.43 ± 13.21	-27.57 ± 15.52	0.14 ± 0.02	5.36 ± 0.90	5.13 ± 0.51	4.83 ± 0.28 $_{-0.29}^{+0.28}$
7	PSZ2G077.90-26.63	ACO2409	0.147	S	9.09 ± 1.32	-26.87 ± 10.89	18.00 ± 11.85	0.14 ± 0.02	6.22 ± 0.88	5.92 ± 0.58	5.08 ± 0.27 $_{-0.27}^{+0.27}$
8	PSZ2G050.40+31.17	ACO2259	0.164	S	5.52 ± 1.19	35.72 ± 21.77	9.31 ± 19.56	0.13 ± 0.02	3.80 ± 0.80	4.53 ± 0.62	4.36 ± 0.35 $_{-0.36}^{+0.35}$
9	PSZ2G097.72+38.12	ACO2218	0.1709	S	10.65 ± 1.68	31.99 ± 15.25	-0.95 ± 13.52	0.13 ± 0.02	7.23 ± 1.11	7.44 ± 0.40	6.64 ± 0.17 $_{-0.17}^{+0.17}$
10	PSZ2G099.30+20.92	MCXJC1935.3+6734	0.171	S	5.57 ± 1.24	-37.19 ± 19.92	-24.50 ± 21.16	0.13 ± 0.02	3.83 ± 0.83	5.88 ± 0.93	3.91 ± 0.23 $_{-0.25}^{+0.23}$
11	PSZ2G067.17+67.46	ACO1914	0.1712	S	10.45 ± 1.49	31.39 ± 12.81	-33.15 ± 11.99	0.13 ± 0.02	7.09 ± 0.99	7.14 ± 0.47	7.04 ± 0.26 $_{-0.27}^{+0.26}$
12	PSZ2G167.67+17.63	RXJ0638.1+4747	0.174	S	4.78 ± 1.36	-28.70 ± 31.24	10.76 ± 28.64	0.13 ± 0.02	3.30 ± 0.92	7.72 ± 0.81	6.31 ± 0.33 $_{-0.34}^{+0.33}$
13	PSZ2G066.68+68.44	ACO1902	0.181	S	4.95 ± 1.43	56.07 ± 25.47	8.14 ± 33.23	0.13 ± 0.02	3.41 ± 0.97	5.27 ± 0.84	3.98 ± 0.33 $_{-0.37}^{+0.37}$
14	PSZ2G065.28+44.53	ACO2187	0.183	S	5.24 ± 1.28	-16.66 ± 22.61	-16.54 ± 21.65	0.13 ± 0.02	3.60 ± 0.86	3.89 ± 0.98	3.56 ± 0.47 $_{-0.51}^{+0.47}$
15	PSZ2G084.47+12.63	MCXJC1948.3+5113	0.185	S	4.79 ± 1.22	-73.73 ± 31.17	-16.97 ± 20.93	0.13 ± 0.02	3.30 ± 0.82	5.98 ± 0.65	4.94 ± 0.33 $_{-0.34}^{+0.34}$
16	PSZ2G100.04+23.73	ACO2317	0.21	S	5.44 ± 1.13	20.24 ± 19.02	-22.73 ± 20.90	0.13 ± 0.02	3.72 ± 0.75	4.10 ± 0.80	3.73 ± 0.29 $_{-0.31}^{+0.29}$
17	PSZ2G180.60+76.65	SDSSCGB26344.3	0.2138	S	5.38 ± 1.21	37.81 ± 15.59	-66.98 ± 19.41	0.13 ± 0.02	3.68 ± 0.81	6.76 ± 0.75	6.00 ± 0.35 $_{-0.34}^{+0.35}$
18	PSZ2G166.09+43.38	ACO773N	0.2172	S	9.84 ± 1.39	-5.35 ± 10.66	-3.98 ± 9.70	0.13 ± 0.02	6.63 ± 0.92	7.76 ± 0.73	6.87 ± 0.34 $_{-0.32}^{+0.34}$
19	PSZ2G125.30-27.99	N/A	0.223	P	4.51 ± 1.31	-8.08 ± 26.99	8.82 ± 30.24	0.13 ± 0.02	3.09 ± 0.87	5.54 ± 0.98	4.70 ± 0.56 $_{-0.55}^{+0.56}$
20	PSZ2G060.13+11.44	N/A	0.224	S	7.47 ± 1.22	-64.79 ± 12.50	-49.27 ± 14.16	0.13 ± 0.02	5.06 ± 0.80	7.55 ± 1.09	5.34 ± 0.49 $_{-0.50}^{+0.49}$
21	PSZ2G166.62+42.13	ACO746	0.232	P	3.56 ± 1.07	-38.98 ± 29.87	-38.09 ± 37.84	0.13 ± 0.02	2.44 ± 0.72	5.60 ± 0.71	5.36 ± 0.39 $_{-0.41}^{+0.39}$
22	PSZ2G097.94+19.43	4C 65.28	0.25	S	5.01 ± 1.31	-114.76 ± 22.50	-13.64 ± 34.07	0.13 ± 0.02	3.40 ± 0.87	5.69 ± 0.85	4.04 ± 0.30 $_{-0.33}^{+0.30}$
23	PSZ2G164.29+08.94	N/A	0.251	P	5.97 ± 1.06	-62.17 ± 14.03	18.12 ± 17.06	0.13 ± 0.02	4.04 ± 0.70	7.91 ± 1.36	6.24 ± 0.62 $_{-0.64}^{+0.64}$
24	PSZ2G133.60+69.04	RXJ1229.0+4737	0.254	S	5.26 ± 1.60	5.87 ± 25.04	59.40 ± 37.35	0.13 ± 0.02	3.57 ± 1.06	7.04 ± 0.97	5.42 ± 0.38 $_{-0.43}^{+0.43}$
25	PSZ2G086.47+15.31	MCXJC1938.3+5409	0.26	S	10.89 ± 1.87	-39.65 ± 13.24	19.83 ± 12.61	0.13 ± 0.02	7.25 ± 1.21	9.54 ± 0.63	7.76 ± 0.29 $_{-0.28}^{+0.29}$

Table A.1 – continued from previous page

Row	Planck ID	Alias	z	z type	$M_{\text{AMI}}(r_{200})$	x_c	y_c	$f_{\text{gas}}(r_{200})$	$M_{\text{AMI}}(r_{500})$	$M_{\text{Pl,marg}}(r_{500})$	$M_{\text{Pl,slice}}(r_{500})$
26	PSZ2G139.62+24.18	N/A	0.2671	S	8.13 ± 1.28	36.66 ± 11.64	-12.58 ± 10.80	0.13 ± 0.02	5.45 ± 0.84	8.34 ± 1.06	7.11 ± 0.48
27	PSZ2G184.68+28.91	ACO611	0.288	S	7.90 ± 1.02	22.61 ± 10.45	13.48 ± 9.97	0.13 ± 0.02	5.28 ± 0.67	11.44 ± 2.30	5.61 ± 0.52
28	PSZ2G154.13+40.19	ACO747	0.29	P	6.46 ± 1.13	70.99 ± 14.72	-42.86 ± 13.25	0.13 ± 0.02	4.33 ± 0.74	6.09 ± 1.10	5.48 ± 0.45
29	PSZ2G095.49+16.41	N/A	0.3	S	5.43 ± 1.12	-24.47 ± 19.10	-102.18 ± 18.33	0.13 ± 0.02	3.65 ± 0.74	4.91 ± 0.99	4.38 ± 0.49
30	PSZ2G109.52-19.16	N/A	0.3092	P	8.53 ± 1.40	-30.38 ± 13.77	-15.21 ± 15.15	0.13 ± 0.02	5.66 ± 0.91	8.34 ± 1.79	5.78 ± 0.48
31	PSZ2G198.90+18.16	[SPD2011] 298	0.3184	P	7.61 ± 1.18	26.76 ± 14.62	-58.07 ± 11.95	0.13 ± 0.02	5.06 ± 0.77	7.99 ± 1.47	5.87 ± 0.55
32	PSZ2G152.33+81.28	MCXCJ1230.7+3439	0.333	S	6.27 ± 1.12	-52.81 ± 20.78	44.11 ± 14.62	0.13 ± 0.02	4.17 ± 0.73	5.08 ± 0.96	5.05 ± 0.53
33	PSZ2G108.17-11.56	N/A	0.336	S	8.00 ± 1.23	35.19 ± 13.14	-70.15 ± 19.09	0.13 ± 0.02	5.29 ± 0.80	9.82 ± 1.29	7.42 ± 0.57
34	PSZ2G132.47-17.27	MCXCJ0142.9+4438	0.341	S	12.43 ± 1.85	31.87 ± 10.19	15.27 ± 12.93	0.13 ± 0.02	8.13 ± 1.18	8.27 ± 1.12	8.07 ± 0.61
35	PSZ2G207.88+81.31	ACO1489	0.353	S	11.26 ± 1.61	68.55 ± 8.44	62.56 ± 11.55	0.13 ± 0.02	7.36 ± 1.02	8.01 ± 0.95	7.54 ± 0.45
36	PSZ2G157.32-26.77	MCSJ0308.9+2645	0.356	S	14.28 ± 2.12	0.33 ± 8.12	17.65 ± 11.53	0.13 ± 0.02	9.27 ± 1.34	10.95 ± 1.12	10.67 ± 0.64
37	PSZ2G071.21+28.86	RXSJ175201.5+444046	0.366	S	9.26 ± 1.51	-29.82 ± 9.95	-12.58 ± 13.26	0.13 ± 0.02	6.07 ± 0.96	6.15 ± 0.80	6.70 ± 0.44
38	PSZ2G194.98+54.12	MCSJ1006.9+3200	0.375	P	8.90 ± 1.56	32.58 ± 12.17	-0.22 ± 19.18	0.13 ± 0.02	5.83 ± 1.00	6.31 ± 1.38	5.30 ± 0.65
39	PSZ2G109.86+27.94	N/A	0.4	S	4.57 ± 1.28	3.98 ± 22.50	7.39 ± 18.70	0.13 ± 0.02	3.03 ± 0.83	5.23 ± 0.91	5.23 ± 0.48
40	PSZ2G083.29-31.03	MCXCJ2228.6+2036	0.412	S	11.85 ± 1.73	81.05 ± 13.29	-3.42 ± 12.73	0.13 ± 0.02	7.65 ± 1.09	9.21 ± 0.95	8.31 ± 0.45
41	PSZ2G063.38+53.44	NSCJ1537+392702	0.422	S	12.17 ± 1.94	46.13 ± 12.01	46.02 ± 9.37	0.13 ± 0.02	7.84 ± 1.22	7.78 ± 1.54	6.17 ± 0.58
42	PSZ2G063.80+11.42	N/A	0.426	S	5.13 ± 1.19	-36.41 ± 22.22	-47.14 ± 19.79	0.13 ± 0.02	3.37 ± 0.76	5.53 ± 0.63	6.41 ± 0.58
43	PSZ2G157.43+30.34	RXJ0748.6+5940	0.45	P	11.64 ± 1.56	-61.32 ± 7.38	4.53 ± 8.27	0.13 ± 0.02	7.47 ± 0.98	6.71 ± 0.44	8.16 ± 0.54
44	PSZ2G150.56+58.32	CLGJ1115+5319	0.47	S	12.77 ± 2.40	10.18 ± 13.31	34.06 ± 18.57	0.13 ± 0.02	8.14 ± 1.49	10.04 ± 1.61	7.44 ± 0.50
45	PSZ2G170.98+39.45	[SPD2011] 16774	0.5131	S	10.11 ± 1.38	31.48 ± 10.20	-30.87 ± 12.67	0.12 ± 0.02	6.43 ± 0.86	8.24 ± 1.30	7.55 ± 0.65
46	PSZ2G094.56+51.03	N/A	0.5392	S	10.83 ± 1.43	81.61 ± 8.09	52.86 ± 8.80	0.13 ± 0.02	6.85 ± 0.88	6.46 ± 0.93	5.90 ± 0.45
47	PSZ2G228.16+75.20	CLGJ1149+2223	0.545	S	15.63 ± 1.66	-15.49 ± 5.32	17.11 ± 4.75	0.13 ± 0.01	9.78 ± 1.01	9.64 ± 0.94	9.69 ± 0.53
48	PSZ2G213.39+80.59	SDSSCGB41791	0.5586	S	9.31 ± 1.32	-9.73 ± 11.90	69.37 ± 12.14	0.13 ± 0.02	5.89 ± 0.81	8.03 ± 1.39	6.77 ± 0.65
49	PSZ2G066.41+27.03	N/A	0.5699	S	13.23 ± 2.05	-33.18 ± 11.12	97.03 ± 11.32	0.13 ± 0.02	8.27 ± 1.25	7.33 ± 0.82	7.72 ± 0.52
50	PSZ2G144.83+25.11	CLGJ0647+7015	0.584	S	11.69 ± 1.46	4.15 ± 7.87	-1.21 ± 8.54	0.13 ± 0.02	7.32 ± 0.89	8.50 ± 1.27	7.80 ± 0.72
51	PSZ2G045.87+57.70	N/A	0.611	S	9.22 ± 1.97	11.71 ± 14.87	24.21 ± 12.21	0.13 ± 0.02	5.78 ± 1.20	8.49 ± 1.61	7.05 ± 0.66
52	PSZ2G108.27+48.66	N/A	0.674	S	9.31 ± 1.46	9.99 ± 11.34	35.79 ± 11.45	0.13 ± 0.02	5.77 ± 0.88	8.44 ± 1.58	4.96 ± 0.52
53	PSZ2G086.93+53.18	N/A	0.6752	P	9.85 ± 1.69	-47.72 ± 14.38	27.69 ± 10.67	0.13 ± 0.02	6.10 ± 1.01	6.07 ± 1.09	5.46 ± 0.51
54	PSZ2G141.77+14.19	N/A	0.83	P	10.99 ± 1.50	-4.36 ± 8.54	-19.02 ± 8.85	0.13 ± 0.02	6.61 ± 0.87	9.94 ± 2.01	7.77 ± 0.90

APPENDIX B: Y_{tot} DERIVATION AND RESULTS OF PHYSICAL AND OBSERVATIONAL MODELLING COMPARISON

B.1 GNFW Y_{tot} analytical solution

We wish to solve the integral

$$\int_0^\infty r^2 \left(\frac{r}{r_p}\right)^{-c} \left[1 + \left(\frac{r}{r_p}\right)^a\right]^{\frac{c-b}{a}} dr. \quad (\text{B.1})$$

First we use the substitution $s = r/r_p$ to get

$$r_p^3 \int_0^\infty s^2 s^{-c} [1 + s^a]^{\frac{c-b}{a}} ds. \quad (\text{B.2})$$

Next we use the substitution $t = s^a$ which gives

$$\begin{aligned} & \frac{r_p^3}{a} \int_0^\infty s^{3-a-c} [1 + t]^{\frac{c-b}{a}} dt \\ &= \frac{r_p^3}{a} \int_0^\infty \frac{t^{\frac{3-c}{a}-1}}{(1+t)^{\frac{b-c}{a}}} dt. \end{aligned} \quad (\text{B.3})$$

Now using the following form for the beta function

$$\beta(x, y) = \int_0^\infty \frac{t^{x-1}}{(1+t)^{x+y}} dt, \quad (\text{B.4})$$

and the relation between the beta and Gamma functions given by $\beta(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ then we get the result

$$\frac{r_p^3}{a} \frac{\Gamma\left(\frac{3-c}{a}\right) \Gamma\left(\frac{b-3}{a}\right)}{\Gamma\left(\frac{b-c}{a}\right)}, \quad (\text{B.5})$$

as required.

Table B.1: Summary of parameter estimates for final sample of 54 clusters. All Y values are given in units of $\times 10^{-3}$ (arcmin 2), and all cluster centre coordinates are given in arcseconds. The cluster centre estimates from the physical model are omitted here but can be found in the results Table A.1 in Appendix A, which is ordered in the same order as this Table. Note the Table in Appendix A also gives external names associated with these clusters, as well as the method used to measure the respective redshifts (i.e. spectroscopic or photometric).

Row	Planck ID	z	$Y_{\text{PM}}(r_{500})$	$Y_{\text{OMI}}(r_{500})$	$x_{0,\text{OMI}}$	$y_{0,\text{OMI}}$	$Y_{\text{OM II}}(r_{500})$	$x_{0,\text{OM II}}$	$y_{0,\text{OM II}}$
1	PSZ2G044.20+48.66	0.0894	11.59 ± 2.28	6.77 ± 3.32	6.53 ± 18.56	8.93 ± 14.41	20.48 ± 6.19	10.36 ± 18.38	8.32 ± 15.32
2	PSZ2G053.53+59.52	0.113	3.81 ± 0.67	2.02 ± 0.90	-1.77 ± 12.69	23.19 ± 9.38	3.12 ± 1.74	-1.07 ± 12.67	20.89 ± 9.88
3	PSZ2G151.90+11.63	0.12	1.76 ± 0.50	2.55 ± 1.56	63.93 ± 28.11	67.61 ± 18.86	4.09 ± 1.83	59.05 ± 27.67	67.19 ± 19.36
4	PSZ2G218.59+71.31	0.137	0.45 ± 0.25	0.35 ± 0.15	-8.85 ± 14.58	-17.72 ± 14.59	0.43 ± 0.27	0.04 ± 23.62	-16.95 ± 24.66
5	PSZ2G226.18+76.79	0.1427	2.45 ± 0.45	0.91 ± 0.45	-45.20 ± 10.61	6.46 ± 12.25	1.21 ± 0.51	-42.92 ± 10.66	3.80 ± 12.00
6	PSZ2G165.06+54.13	0.144	2.26 ± 0.54	0.70 ± 0.25	29.82 ± 10.17	-29.36 ± 12.22	0.95 ± 0.27	31.51 ± 10.76	-29.04 ± 12.83
7	PSZ2G077.90-26.63	0.147	2.80 ± 0.46	1.35 ± 0.48	-27.99 ± 9.91	20.12 ± 11.23	1.48 ± 0.49	-28.06 ± 10.13	19.93 ± 11.07
8	PSZ2G050.40+31.17	0.164	1.01 ± 0.29	1.07 ± 0.70	37.21 ± 20.82	9.59 ± 19.09	1.18 ± 0.76	36.11 ± 22.25	9.30 ± 19.70
9	PSZ2G097.72+38.12	0.1709	2.65 ± 0.46	2.72 ± 1.26	29.79 ± 15.43	-2.59 ± 13.68	3.97 ± 1.49	32.13 ± 15.62	-1.56 ± 13.81
10	PSZ2G099.30+20.92	0.171	0.97 ± 0.31	0.79 ± 0.49	-35.09 ± 19.11	-24.57 ± 21.53	0.86 ± 0.51	-36.16 ± 19.13	-25.55 ± 21.67
11	PSZ2G067.17+67.46	0.1712	2.70 ± 0.46	1.30 ± 0.54	34.00 ± 11.65	-30.54 ± 10.97	1.48 ± 0.60	33.18 ± 11.61	-31.32 ± 11.16
12	PSZ2G167.67+17.63	0.174	0.72 ± 0.30	1.69 ± 1.05	-24.86 ± 32.03	10.55 ± 28.11	1.33 ± 0.77	-23.41 ± 33.17	11.93 ± 29.04
13	PSZ2G066.68+68.44	0.181	0.72 ± 0.29	1.24 ± 0.79	55.97 ± 25.19	9.20 ± 32.13	1.12 ± 0.72	56.41 ± 26.70	7.31 ± 32.63
14	PSZ2G065.28+44.53	0.183	0.79 ± 0.28	0.65 ± 0.38	-21.13 ± 20.72	-15.63 ± 18.96	0.61 ± 0.34	-19.57 ± 22.13	-16.08 ± 20.64
15	PSZ2G084.47+12.63	0.185	0.67 ± 0.25	0.58 ± 0.29	-67.12 ± 29.59	-23.26 ± 18.01	0.53 ± 0.28	-69.03 ± 30.83	-20.78 ± 20.07
16	PSZ2G100.04+23.73	0.21	0.65 ± 0.18	1.28 ± 0.75	17.47 ± 19.11	-22.73 ± 22.21	1.05 ± 0.55	17.93 ± 19.85	-23.27 ± 22.53
17	PSZ2G180.60+76.65	0.2138	0.63 ± 0.20	1.73 ± 0.93	36.57 ± 16.66	-73.38 ± 20.39	1.11 ± 0.50	35.90 ± 17.29	-70.57 ± 22.18
18	PSZ2G166.09+43.38	0.2172	1.67 ± 0.28	1.10 ± 0.50	-4.29 ± 10.57	-6.54 ± 9.54	1.14 ± 0.46	-4.73 ± 10.32	-6.66 ± 9.63
19	PSZ2G125.30-27.99	0.223	0.45 ± 0.18	0.99 ± 0.64	-8.12 ± 26.53	2.49 ± 30.79	0.60 ± 0.38	-9.03 ± 28.36	6.48 ± 31.71
20	PSZ2G060.13+11.44	0.224	1.00 ± 0.20	1.17 ± 0.64	-64.93 ± 12.76	-49.60 ± 15.02	1.12 ± 0.56	-64.67 ± 12.69	-49.56 ± 14.77
21	PSZ2G166.62+42.13	0.232	0.29 ± 0.13	1.57 ± 0.92	-36.13 ± 30.51	-54.22 ± 32.52	0.53 ± 0.35	-34.92 ± 31.92	-40.79 ± 38.30

Table B.1 – continued from previous page

Row	Planck ID	z	$Y_{\text{PM}}(r_{500})$	$Y_{\text{OMI}}(r_{500})$	$x_{0,\text{OMI}}$	$y_{0,\text{OMI}}$	$Y_{\text{OM II}}(r_{500})$	$x_{0,\text{OM II}}$	$y_{0,\text{OM II}}$
22	PSZ2G097.94+19.43	0.25	0.45 ± 0.17	1.24 ± 0.69	-121.19 ± 21.52	-2.42 ± 32.74	0.73 ± 0.41	-115.20 ± 27.60	-5.84 ± 34.15
23	PSZ2G164.29+08.94	0.251	0.59 ± 0.13	0.87 ± 0.43	-62.15 ± 13.92	20.46 ± 17.35	0.73 ± 0.35	-62.23 ± 13.90	18.67 ± 17.99
24	PSZ2G133.60+69.04	0.254	0.47 ± 0.20	1.60 ± 1.12	0.13 ± 24.80	66.74 ± 35.89	0.80 ± 0.45	3.35 ± 25.98	63.00 ± 37.13
25	PSZ2G086.47+15.31	0.26	1.48 ± 0.33	1.70 ± 0.71	-41.40 ± 14.66	19.45 ± 13.73	1.58 ± 0.60	-40.08 ± 14.39	20.08 ± 13.75
26	PSZ2G139.62+24.18	0.2671	0.89 ± 0.16	0.77 ± 0.34	35.74 ± 11.80	-13.45 ± 11.11	0.70 ± 0.33	35.83 ± 11.49	-13.78 ± 10.76
27	PSZ2G184.68+28.91	0.288	0.76 ± 0.12	0.95 ± 0.38	22.66 ± 10.55	12.19 ± 10.37	0.83 ± 0.31	22.58 ± 10.48	13.03 ± 10.41
28	PSZ2G154.13+40.19	0.29	0.55 ± 0.13	0.72 ± 0.50	71.59 ± 15.07	-42.78 ± 13.41	0.46 ± 0.23	69.88 ± 14.52	-42.45 ± 13.20
29	PSZ2G095.49+16.41	0.3	0.39 ± 0.12	0.87 ± 0.54	-19.80 ± 21.12	-94.58 ± 19.43	0.48 ± 0.26	-22.58 ± 20.72	-98.75 ± 20.62
30	PSZ2G109.52-19.16	0.3092	0.78 ± 0.16	1.00 ± 0.57	-31.66 ± 14.34	-15.21 ± 15.68	0.82 ± 0.39	-31.16 ± 14.43	-15.23 ± 15.95
31	PSZ2G198.90+18.16	0.3184	0.62 ± 0.12	0.86 ± 0.40	27.42 ± 15.36	-59.55 ± 12.35	0.69 ± 0.27	27.03 ± 15.25	-57.65 ± 12.64
32	PSZ2G152.33+81.28	0.333	0.43 ± 0.11	0.78 ± 0.42	-49.96 ± 20.35	44.73 ± 15.45	0.48 ± 0.22	-53.60 ± 20.93	43.79 ± 15.28
33	PSZ2G108.17-11.56	0.336	0.61 ± 0.12	2.24 ± 1.10	27.48 ± 14.92	-36.56 ± 20.44	1.12 ± 0.25	30.62 ± 13.89	-51.07 ± 19.77
34	PSZ2G132.47-17.27	0.341	1.25 ± 0.21	1.38 ± 0.52	32.53 ± 10.83	16.82 ± 12.65	1.37 ± 0.47	32.34 ± 10.66	16.61 ± 12.56
35	PSZ2G207.88+81.31	0.353	1.05 ± 0.18	0.90 ± 0.34	67.45 ± 8.46	61.30 ± 11.45	0.82 ± 0.29	66.90 ± 8.21	59.84 ± 11.43
36	PSZ2G157.32-26.77	0.356	1.52 ± 0.27	1.25 ± 0.42	-0.28 ± 8.01	19.15 ± 11.86	1.23 ± 0.39	-1.07 ± 7.59	17.73 ± 11.58
37	PSZ2G071.21+28.86	0.366	0.72 ± 0.15	0.91 ± 0.34	-29.47 ± 10.86	-12.29 ± 14.04	0.75 ± 0.25	-29.64 ± 10.48	-12.13 ± 13.74
38	PSZ2G194.98+54.12	0.375	0.65 ± 0.15	1.28 ± 0.69	32.85 ± 12.59	-5.89 ± 18.85	0.93 ± 0.32	32.71 ± 12.45	-3.46 ± 19.90
39	PSZ2G109.86+27.94	0.4	0.21 ± 0.09	0.30 ± 0.11	8.03 ± 16.29	-1.95 ± 14.87	0.20 ± 0.07	7.15 ± 21.69	2.87 ± 17.97
40	PSZ2G083.29-31.03	0.412	0.95 ± 0.17	0.66 ± 0.21	75.26 ± 13.22	-0.29 ± 12.25	0.60 ± 0.20	72.16 ± 13.03	2.13 ± 11.88
41	PSZ2G063.38+53.44	0.422	0.93 ± 0.19	1.28 ± 0.45	39.37 ± 14.20	49.33 ± 10.77	1.12 ± 0.29	41.65 ± 13.30	48.43 ± 10.17
42	PSZ2G063.80+11.42	0.426	0.24 ± 0.08	0.80 ± 0.46	-42.04 ± 23.06	-44.32 ± 20.40	0.29 ± 0.14	-36.98 ± 23.28	-45.28 ± 20.74
43	PSZ2G157.43+30.34	0.45	0.82 ± 0.13	0.91 ± 0.26	-61.41 ± 7.56	4.85 ± 8.34	0.85 ± 0.23	-61.63 ± 7.29	4.79 ± 8.26
44	PSZ2G150.56+58.32	0.47	0.93 ± 0.25	0.86 ± 0.38	9.81 ± 14.03	35.97 ± 18.29	0.70 ± 0.25	8.34 ± 12.93	36.51 ± 18.01
45	PSZ2G170.98+39.45	0.5131	0.54 ± 0.08	1.62 ± 0.68	23.91 ± 12.09	-18.32 ± 13.31	0.88 ± 0.17	26.68 ± 11.52	-22.95 ± 12.68
46	PSZ2G094.56+51.03	0.5392	0.63 ± 0.10	0.50 ± 0.09	82.24 ± 7.64	50.61 ± 8.76	0.45 ± 0.08	81.87 ± 7.67	50.51 ± 8.62
47	PSZ2G228.16+75.20	0.545	1.06 ± 0.10	1.35 ± 0.27	-14.53 ± 5.57	16.35 ± 5.31	1.25 ± 0.21	-14.39 ± 5.59	16.50 ± 5.08
48	PSZ2G213.39+80.59	0.5586	0.45 ± 0.08	0.89 ± 0.36	-5.34 ± 12.49	65.15 ± 12.29	0.58 ± 0.18	-8.19 ± 12.21	68.13 ± 12.60
49	PSZ2G066.41+27.03	0.5699	0.79 ± 0.16	1.76 ± 0.73	-37.37 ± 11.95	100.92 ± 13.20	1.00 ± 0.24	-34.28 ± 11.21	97.77 ± 11.89

Table B.1 – continued from previous page

Row	Planck ID	z	$Y_{\text{PM}}(r_{500})$	$Y_{\text{OMI}}(r_{500})$	$x_{0,\text{OMI}}$	$y_{0,\text{OMI}}$	$Y_{\text{OM II}}(r_{500})$	$x_{0,\text{OM II}}$	$y_{0,\text{OM II}}$
50	PSZ2G144.83+25.11	0.584	0.61 ± 0.07	1.34 ± 0.45	1.55 ± 9.00	-3.86 ± 8.95	0.89 ± 0.17	3.09 ± 8.57	-2.97 ± 8.79
51	PSZ2G045.87+57.70	0.611	0.41 ± 0.12	0.93 ± 0.46	20.59 ± 17.97	16.79 ± 15.76	0.52 ± 0.16	16.61 ± 16.65	20.54 ± 14.20
52	PSZ2G108.27+48.66	0.674	0.40 ± 0.08	0.55 ± 0.20	8.45 ± 11.83	35.26 ± 11.93	0.42 ± 0.12	9.91 ± 12.03	35.53 ± 11.69
53	PSZ2G086.93+53.18	0.6752	0.43 ± 0.10	1.28 ± 0.57	-40.06 ± 16.39	30.84 ± 12.08	0.59 ± 0.15	-44.92 ± 15.26	29.36 ± 11.53
54	PSZ2G141.77+14.19	0.83	0.45 ± 0.06	0.56 ± 0.17	-3.40 ± 8.77	-18.18 ± 9.36	0.47 ± 0.11	-4.11 ± 8.78	-18.97 ± 9.37

Table B.2: Summary of model comparison statistics for final sample of 54 clusters.

The Planck IDs are omitted but are the same as in Table B.1.

Row	z	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OMI}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{OM II}}, \mathcal{P}_{\text{OMI}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM II}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{OMI}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{OM II}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}})$	$\ln(\mathcal{Z}_{\text{OM II}}/\mathcal{Z}_{\text{OMI}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OM II}})$	
1	0.0894	0.222		0.514	0.297	33.90 ± 0.16	29.17 ± 0.16	33.38 ± 0.16	4.73 ± 0.23	4.21 ± 0.23	0.52 ± 0.22
2	0.113	0.152		0.091	0.093	30.94 ± 0.17	31.06 ± 0.17	30.01 ± 0.17	-0.12 ± 0.24	-1.05 ± 0.24	0.93 ± 0.24
3	0.12	0.083		0.123	0.189	10.40 ± 0.13	10.54 ± 0.13	10.00 ± 0.14	-0.14 ± 0.19	-0.53 ± 0.19	0.39 ± 0.19
4	0.137	0.132		0.115	0.051	1.71 ± 0.17	3.41 ± 0.17	1.76 ± 0.17	-1.70 ± 0.24	-1.65 ± 0.24	-0.05 ± 0.24
5	0.1427	0.170		0.033	0.138	23.01 ± 0.15	24.85 ± 0.15	23.50 ± 0.15	-1.84 ± 0.21	-1.35 ± 0.21	-0.49 ± 0.21
6	0.144	0.210		0.045	0.165	13.68 ± 0.13	17.82 ± 0.13	15.56 ± 0.14	-4.14 ± 0.18	-2.26 ± 0.19	-1.88 ± 0.19
7	0.147	0.140		0.014	0.126	32.94 ± 0.12	34.76 ± 0.12	33.50 ± 0.12	-1.82 ± 0.17	-1.26 ± 0.17	-0.56 ± 0.17
8	0.164	0.065		0.026	0.069	9.61 ± 0.08	10.32 ± 0.08	9.10 ± 0.08	-0.71 ± 0.11	-1.23 ± 0.11	0.51 ± 0.12
9	0.1709	0.049		0.082	0.087	33.10 ± 0.16	33.00 ± 0.16	32.62 ± 0.16	0.10 ± 0.22	-0.37 ± 0.22	0.47 ± 0.23
10	0.171	0.058		0.012	0.058	7.73 ± 0.15	8.46 ± 0.15	7.08 ± 0.15	-0.73 ± 0.21	-1.38 ± 0.21	0.65 ± 0.21
11	0.1712	0.135		0.022	0.114	26.98 ± 0.10	28.19 ± 0.10	27.08 ± 0.11	-1.21 ± 0.14	-1.11 ± 0.15	-0.10 ± 0.15
12	0.174	0.132		0.029	0.107	3.67 ± 0.11	4.53 ± 0.11	3.56 ± 0.11	-0.86 ± 0.15	-0.97 ± 0.16	0.11 ± 0.16
13	0.181	0.084		0.015	0.080	4.42 ± 0.13	5.00 ± 0.12	4.06 ± 0.13	-0.58 ± 0.18	-0.95 ± 0.18	0.36 ± 0.18
14	0.183	0.068		0.010	0.063	5.57 ± 0.13	6.52 ± 0.13	5.35 ± 0.13	-0.94 ± 0.18	-1.16 ± 0.18	0.22 ± 0.19
15	0.185	0.062		0.010	0.056	3.57 ± 0.18	4.28 ± 0.18	3.47 ± 0.18	-0.71 ± 0.25	-0.80 ± 0.25	0.09 ± 0.25
16	0.21	0.094		0.026	0.076	7.98 ± 0.14	8.67 ± 0.14	7.51 ± 0.14	-0.69 ± 0.20	-1.15 ± 0.20	0.46 ± 0.20
17	0.2138	0.143		0.051	0.094	4.68 ± 0.18	5.67 ± 0.18	4.38 ± 0.18	-0.99 ± 0.25	-1.29 ± 0.25	0.30 ± 0.25
18	0.2172	0.072		0.006	0.069	27.82 ± 0.12	28.93 ± 0.12	27.64 ± 0.13	-1.11 ± 0.17	-1.29 ± 0.17	0.18 ± 0.18
19	0.223	0.097		0.054	0.057	4.36 ± 0.10	4.84 ± 0.10	3.95 ± 0.10	-0.48 ± 0.14	-0.89 ± 0.14	0.41 ± 0.14
20	0.224	0.049		0.009	0.051	16.34 ± 0.13	17.23 ± 0.13	15.79 ± 0.13	-0.89 ± 0.18	-1.44 ± 0.19	0.55 ± 0.19

Table B.2 – continued from previous page

Row	z	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OM1}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{OMII}}, \mathcal{P}_{\text{OM1}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OMII}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{OM1}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{OMII}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OM1}})$	$\ln(\mathcal{Z}_{\text{OMII}}/\mathcal{Z}_{\text{OM1}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMII}})$
21	0.232	0.225		0.147	3.02 ± 0.15	4.37 ± 0.15	2.54 ± 0.15	-1.35 ± 0.21	-1.82 ± 0.21	0.48 ± 0.21
22	0.25	0.136		0.071	3.03 ± 0.15	3.96 ± 0.15	2.26 ± 0.15	-0.93 ± 0.21	-1.70 ± 0.21	0.77 ± 0.21
23	0.251	0.055		0.024	12.67 ± 0.16	13.45 ± 0.16	11.69 ± 0.17	-0.78 ± 0.23	-1.76 ± 0.23	0.99 ± 0.23
24	0.254	0.180		0.110	3.80 ± 0.11	5.27 ± 0.11	3.86 ± 0.11	-1.47 ± 0.15	-1.41 ± 0.15	-0.06 ± 0.15
25	0.26	0.041		0.009	13.18 ± 0.17	13.79 ± 0.16	12.32 ± 0.17	-0.60 ± 0.23	-1.46 ± 0.23	0.86 ± 0.23
26	0.2671	0.043		0.012	28.23 ± 0.14	29.05 ± 0.14	27.67 ± 0.14	-0.81 ± 0.20	-1.38 ± 0.20	0.56 ± 0.20
27	0.288	0.038		0.018	22.61 ± 0.14	23.45 ± 0.14	21.90 ± 0.14	-0.85 ± 0.19	-1.55 ± 0.19	0.71 ± 0.20
28	0.29	0.045		0.034	9.72 ± 0.18	10.64 ± 0.18	9.42 ± 0.18	-0.92 ± 0.26	-1.23 ± 0.26	0.31 ± 0.26
29	0.3	0.138		0.115	0.45 ± 0.20	5.26 ± 0.20	5.94 ± 0.19	4.44 ± 0.20	-0.68 ± 0.28	-1.51 ± 0.28
30	0.3092	0.047		0.027	14.83 ± 0.12	15.62 ± 0.12	14.13 ± 0.12	-0.80 ± 0.17	-1.49 ± 0.17	0.70 ± 0.17
31	0.3184	0.042		0.025	14.64 ± 0.11	15.36 ± 0.10	13.88 ± 0.11	-0.72 ± 0.15	-1.48 ± 0.15	0.76 ± 0.15
32	0.333	0.071		0.058	0.36 ± 0.15	9.30 ± 0.15	9.89 ± 0.15	8.59 ± 0.15	-0.58 ± 0.21	-1.30 ± 0.21
33	0.336	0.209		0.122	0.88 ± 0.20	10.98 ± 0.20	14.24 ± 0.19	12.05 ± 0.20	-3.26 ± 0.28	-2.19 ± 0.28
34	0.341	0.032		0.006	0.31 ± 0.14	32.32 ± 0.14	33.03 ± 0.14	31.53 ± 0.14	-0.71 ± 0.20	-1.51 ± 0.20
35	0.353	0.036		0.016	0.45 ± 0.16	20.74 ± 0.16	21.70 ± 0.15	20.26 ± 0.16	-0.96 ± 0.22	-1.44 ± 0.22
36	0.356	0.039		0.007	0.43 ± 0.13	25.23 ± 0.13	25.70 ± 0.13	24.79 ± 0.14	-0.47 ± 0.19	-0.91 ± 0.19
37	0.366	0.037		0.018	0.27 ± 0.13	11.84 ± 0.13	12.47 ± 0.13	11.00 ± 0.13	-0.62 ± 0.19	-1.47 ± 0.19
38	0.375	0.093		0.050	0.47 ± 0.14	16.17 ± 0.14	17.58 ± 0.14	15.83 ± 0.14	-1.41 ± 0.20	-1.74 ± 0.20
39	0.4	0.023		0.013	0.27 ± 0.15	3.36 ± 0.15	2.77 ± 0.15	2.75 ± 0.15	0.59 ± 0.22	-0.02 ± 0.22
40	0.412	0.040		0.015	0.54 ± 0.16	26.82 ± 0.16	27.58 ± 0.16	26.56 ± 0.16	-0.76 ± 0.23	-1.01 ± 0.23
41	0.422	0.058		0.027	0.32 ± 0.22	14.70 ± 0.22	15.84 ± 0.22	14.37 ± 0.22	-1.14 ± 0.31	-1.48 ± 0.31
42	0.426	0.126		0.106	0.30 ± 0.15	4.48 ± 0.15	4.89 ± 0.14	4.24 ± 0.15	-0.41 ± 0.20	-0.66 ± 0.20
43	0.45	0.025		0.010	0.20 ± 0.16	31.61 ± 0.16	32.30 ± 0.15	30.87 ± 0.16	-0.69 ± 0.22	-1.43 ± 0.22
44	0.47	0.032		0.023	0.41 ± 0.10	8.28 ± 0.10	8.74 ± 0.10	8.14 ± 0.11	-0.46 ± 0.14	-0.60 ± 0.15
45	0.5131	0.133		0.078	0.55 ± 0.14	23.66 ± 0.14	27.24 ± 0.13	24.82 ± 0.14	-3.58 ± 0.19	-2.42 ± 0.19
46	0.5392	0.036		0.007	0.43 ± 0.18	23.74 ± 0.18	24.69 ± 0.18	24.49 ± 0.18	-0.95 ± 0.25	-0.20 ± 0.25
47	0.545	0.028		0.010	0.20 ± 0.19	110.33 ± 0.19	110.78 ± 0.19	109.81 ± 0.19	-0.45 ± 0.26	-0.97 ± 0.26
48	0.5586	0.064		0.041	0.27 ± 0.20	21.75 ± 0.20	22.86 ± 0.20	21.54 ± 0.20	-1.11 ± 0.28	-1.31 ± 0.28
49	0.5699	0.101		0.071	0.31 ± 0.17	14.90 ± 0.17	16.67 ± 0.17	14.44 ± 0.17	-1.77 ± 0.24	-2.23 ± 0.24
50	0.584	0.080		0.041	0.39 ± 0.17	43.03 ± 0.17	45.57 ± 0.17	43.52 ± 0.17	-2.54 ± 0.24	-2.05 ± 0.24
51	0.611	0.112		0.079	0.35 ± 0.14	8.60 ± 0.14	10.46 ± 0.14	8.54 ± 0.14	-1.86 ± 0.20	-1.92 ± 0.20
52	0.674	0.032		0.026	0.15 ± 0.16	13.43 ± 0.16	13.61 ± 0.16	12.69 ± 0.16	-0.18 ± 0.23	-0.92 ± 0.23

Table B.2 – continued from previous page

Row	z	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OMI}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{OMII}}, \mathcal{P}_{\text{OMI}})$	$d_{\text{EMD}}(\mathcal{P}_{\text{PM}}, \mathcal{P}_{\text{OMII}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{OMI}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{OMII}}/\mathcal{Z}_{\text{null}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMI}})$	$\ln(\mathcal{Z}_{\text{OMII}}/\mathcal{Z}_{\text{OMI}})$	$\ln(\mathcal{Z}_{\text{PM}}/\mathcal{Z}_{\text{OMII}})$
53	0.6752	0.126		0.090		0.037		13.17 ± 0.13	15.96 ± 0.13	
54	0.83	0.020		0.014		0.013		35.45 ± 0.12	35.38 ± 0.11	

APPENDIX C: EINASTO MODEL DERIVATIONS AND RESULTS OF PM I AND PM II COMPARISONS

C.1 Einasto mass integral

From equations 5.1 and 5.4 we have that

$$\begin{aligned} M(r) &= \int_0^r 4\pi r'^2 \rho_{\text{dm, PMII}}(r') dr' \\ &= 4\pi \rho_{-2} \exp(2/\alpha_{\text{Ein}}) \int_0^r r'^2 \exp\left[\frac{-2}{\alpha_{\text{Ein}}} \left(\frac{r'}{r_{-2}}\right)^{\alpha_{\text{Ein}}}\right] dr'. \end{aligned} \quad (\text{C.1})$$

Using the substitution

$$u = \frac{2^{3/\alpha_{\text{Ein}}} r'^3}{\alpha_{\text{Ein}}^{3/\alpha_{\text{Ein}}} r_{-2}^3} \Rightarrow du = \frac{3 \times 2^{3/\alpha_{\text{Ein}}} r'^2}{\alpha_{\text{Ein}}^{3/\alpha_{\text{Ein}}} r_{-2}^3} dr' \quad (\text{C.2})$$

equation C.1 becomes

$$\begin{aligned} M(r) &= \frac{4\pi \rho_{-2} \exp(2/\alpha_{\text{Ein}}) \alpha_{\text{Ein}}^{3/\alpha_{\text{Ein}}} r_{-2}^3}{3 \times 2^{3/\alpha_{\text{Ein}}}} \\ &\times \int_{u=0}^{u=\frac{2^{3/\alpha_{\text{Ein}}} r^3}{\alpha_{\text{Ein}}^{3/\alpha_{\text{Ein}}} r_{-2}^3}} \exp(-u^{\alpha_{\text{Ein}}/3}) du. \end{aligned} \quad (\text{C.3})$$

Finally, using the substitution $t = u^{\alpha_{\text{Ein}}/3}$ so that $dt = \frac{\alpha_{\text{Ein}}}{3} u^{\alpha_{\text{Ein}}/3-1} du$, then the integral in equation C.3 (ignoring the constant factor) becomes

$$\begin{aligned} \frac{3}{\alpha_{\text{Ein}}} \int_0^{u^{\alpha_{\text{Ein}}/3}} u^{1-\alpha_{\text{Ein}}/3} e^{-t} dt &= \frac{3}{\alpha_{\text{Ein}}} \int_0^{\frac{2r^{\alpha_{\text{Ein}}}}{\alpha_{\text{Ein}} r_{-2}^{\alpha_{\text{Ein}}}}} t^{3/\alpha_{\text{Ein}}-1} e^{-t} dt \\ &= \gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right], \end{aligned} \quad (\text{C.4})$$

where the last equality follows from the definition of the incomplete lower gamma function $\gamma[a, x] = \int_0^x t^{a-1} e^{-t} dt$. Including the constant factor in equation C.3 leads to the result

$$M(r) = \frac{4\pi\rho_{-2}}{\alpha_{\text{Ein}}} \exp(2/\alpha_{\text{Ein}}) \left(\frac{\alpha_{\text{Ein}}}{2} \right)^{3/\alpha_{\text{Ein}}} \gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]. \quad (\text{C.5})$$

C.2 Determining r_{500} iteratively

Evaluating equations 2.22 and 5.4 at r_{500} and equating we get

$$\begin{aligned} \frac{4\pi}{3} 500 \rho_{\text{crit}}(z) r_{500}^3 &= 4\pi \rho_{-2} 1/\alpha_{\text{Ein}} \exp(2/\alpha_{\text{Ein}}) \left(\frac{\alpha_{\text{Ein}}}{2} \right)^{3/\alpha_{\text{Ein}}} r_{-2}^3 \\ &\quad \times \gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r_{500}}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]. \end{aligned} \quad (\text{C.6})$$

If we let $R = r_{500}/r_{-2}$, then we can determine r_{500} by solving the following for R

$$\begin{aligned} \frac{R^3}{\gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} R^{\alpha_{\text{Ein}}} \right]} \\ - \frac{1}{\rho_{\text{crit}}(z)} \frac{3\rho_{-2}}{500} \left(\frac{\alpha_{\text{Ein}}}{2} \right)^{3/\alpha_{\text{Ein}}} \frac{\exp(2/\alpha_{\text{Ein}})}{\alpha_{\text{Ein}}} = 0 \end{aligned} \quad (\text{C.7})$$

by some iterative root finding method e.g. Newton-Raphson. We use the starting point $R_0 = \frac{2r_{200}}{3r_{-2}}$ which usually results in the algorithm converging in $\mathcal{O}(10)$ iterations.

We now show that equation C.7 only has one solution for a given r_{-2} . We start by considering both sides of equation C.6 as two different functions, and ignore constant terms for simplicity (this doesn't affect the truth of the final result), i.e. we consider the two functions

$$f(r_{500}) = r_{500}^3, \quad g(r_{500}) = \gamma \left[\frac{3}{\alpha_{\text{Ein}}}, \frac{2}{\alpha_{\text{Ein}}} \left(\frac{r_{500}}{r_{-2}} \right)^{\alpha_{\text{Ein}}} \right]. \quad (\text{C.8})$$

We first note that $f(0) = g(0) = 0$, and differentiate both functions with respect to r_{500}

$$\frac{df}{dr_{500}} \propto r_{500}^2, \quad \frac{dg}{dr_{500}} \propto r_{500}^2 \exp \left[-\frac{2}{\alpha_{\text{Ein}}} \left(\left(\frac{r_{500}}{r_{-2}} \right)^{\alpha_{\text{Ein}}} - 1 \right) \right]. \quad (\text{C.9})$$

Setting these two derivatives equal to each other yields one solution at $r_{500} = r_{-2}$ for all $\alpha_{\text{Ein}} \neq 0$, meaning the derivatives only intersect once. Furthermore $\frac{dg}{dr_{500}}$ tends to zero for large r_{500} whilst $\frac{df}{dr_{500}}$ is a monotonically increasing function, meaning the former must be larger before the two intersect. This coupled with the fact that $f(0) = g(0) = 0$ means that $g(r_{500}) > f(r_{500})$ until some point (which has to be after the derivatives intersect) when the two intersect, after which $f(r_{500}) > g(r_{500})$ as $g(r_{500})$ flattens off. This proves that equation C.7 only has one root and that equation C.6 only has one solution in r_{500} for fixed r_{-2} .

C.3 Simulations results table

Table C.1: Input and output values of simulations using NFW and Einasto dark matter profiles. The first column is what dark matter profile was used to *simulate* the cluster. Input $M(r_{200})$ and Input z are the input values used to create the simulation for the given model. Ein out $M(r_{200})$ is the mean and standard deviation of the posterior distribution obtained inferred using an Einasto profile to *model* the cluster. Ein $\ln(\mathcal{Z})$ is the log Bayesian evidence corresponding to the inference. NFW... is as before but using an NFW profile in the *modelling*. $\ln(\mathcal{Z}_{\text{Ein}}/\mathcal{Z}_{\text{NFW}})$ is the log ratio of the two evidences obtained.

Model	Input $M(r_{200})$ ($\times 10^{14} M_{\text{Sun}}$)	Input z	Ein out $M(r_{200})$ ($\times 10^{14} M_{\text{Sun}}$)	NFW out $M(r_{200})$ ($\times 10^{14} M_{\text{Sun}}$)	Ein $\ln(\mathcal{Z})$	NFW $\ln(\mathcal{Z})$	$\ln(\mathcal{Z}_{\text{Ein}}/\mathcal{Z}_{\text{NFW}})$
$\alpha_{\text{Ein}} = 0.2$	1	0.15	1.05 ± 0.01	1.08 ± 0.01	47104.4 ± 0.4	47104.6 ± 0.4	-0.2
$\alpha_{\text{Ein}} = 2.0$	1	0.15	1.05 ± 0.01	1.48 ± 0.01	47181.4 ± 0.4	47180.7 ± 0.5	0.6
$\alpha_{\text{Ein}} = 0.05$	1	0.15	1.05 ± 0.01	0.97 ± 0.01	47175.0 ± 0.4	47175.3 ± 0.4	-0.3
NFW	1	0.15	1.03 ± 0.01	1.05 ± 0.01	47064.3 ± 0.4	47063.6 ± 0.4	0.7
$\alpha_{\text{Ein}} = 0.2$	1	0.9	1.10 ± 0.01	1.21 ± 0.01	47173.2 ± 0.5	47174.5 ± 0.5	-1.3
$\alpha_{\text{Ein}} = 2.0$	1	0.9	1.16 ± 0.01	1.57 ± 0.01	47100.8 ± 0.5	47095.1 ± 0.5	5.7
$\alpha_{\text{Ein}} = 0.05$	1	0.9	1.02 ± 0.01	1.18 ± 0.01	47094.2 ± 0.4	47094.8 ± 0.5	-0.6
NFW	1	0.9	0.95 ± 0.01	1.05 ± 0.01	47105.2 ± 0.4	47106.7 ± 0.4	-1.5
$\alpha_{\text{Ein}} = 0.2$	10	0.15	10.23 ± 0.02	10.33 ± 0.01	46814.8 ± 0.5	46815.0 ± 0.5	-0.2
$\alpha_{\text{Ein}} = 2.0$	10	0.15	10.18 ± 0.01	15.06 ± 0.01	46640.7 ± 0.5	46638.3 ± 0.6	2.4
$\alpha_{\text{Ein}} = 0.05$	10	0.15	10.21 ± 0.02	9.61 ± 0.01	46844.5 ± 0.5	46844.9 ± 0.5	-0.4
NFW	10	0.15	10.13 ± 0.01	10.23 ± 0.01	46873.9 ± 0.5	46873.0 ± 0.5	0.9
$\alpha_{\text{Ein}} = 0.2$	10	0.9	11.47 ± 0.02	12.12 ± 0.01	46837.2 ± 0.5	46829.7 ± 0.6	7.5
$\alpha_{\text{Ein}} = 2.0$	10	0.9	11.16 ± 0.01	13.65 ± 0.01	46835.0 ± 0.5	46829.9 ± 0.7	5.1
$\alpha_{\text{Ein}} = 0.05$	10	0.9	11.48 ± 0.01	12.26 ± 0.02	46833.8 ± 0.6	46834.5 ± 0.5	-0.7
NFW	10	0.9	10.48 ± 0.02	11.57 ± 0.02	46926.4 ± 0.6	46926.6 ± 0.5	-0.2

APPENDIX

APPENDIX D: SUPPLEMENTARY STATISTICAL DERIVATIONS AND RESULTS

D.1 Metropolis-Hastings and detailed balance

D.1.1 Metropolis-Hastings satisfies detailed balance proof

Following [Hauser \(2013\)](#) we show that for a Markov chain whose values are sampled from target distribution P using proposal distribution q & the MH acceptance ratio, satisfies detailed balance (a *sufficient* condition for the chain to asymptotically converge to the target distribution).

Consider an arbitrary point along the Markov chain, θ_k , then the proceeding step θ_{k+1} can lead to one of two possible scenarios which we denote scenario I and scenario II

$$\alpha(\theta_{k+1}, \theta_k) \Rightarrow \begin{cases} \theta_{k+1} \neq \theta_k & \text{scenario I,} \\ \theta_{k+1} = \theta_k & \text{scenario II.} \end{cases} \quad (\text{D.1})$$

Scenario I only occurs when an MH step is accepted, which occurs with probability

$$T(\theta_{k+1} | \theta_k) = \alpha(\theta_{k+1}, \theta_k) q(\theta_{k+1} | \theta_k). \quad (\text{D.2})$$

Substituting α for the MH acceptance ratio gives

$$\begin{aligned} T(\theta_{k+1}|\theta_k) &= \min \left(1, \frac{p(\theta_{k+1})q(\theta_k|\theta_{k+1})}{p(\theta_k)q(\theta_{k+1}|\theta_k)} \right) q(\theta_{k+1}|\theta_k) \\ &= \frac{1}{p(\theta_k)} \min (p(\theta_k)q(\theta_{k+1}|\theta_k), p(\theta_{k+1})q(\theta_k|\theta_{k+1})) , \end{aligned} \quad (\text{D.3})$$

where the second equality follows from taking a factor of $1/p(\theta_k)$ out of both terms in the minimisation and multiplying the factor $q(\theta_{k+1}|\theta_k)$ into the function (both actions are allowed if P and q are strictly positive).

Observe that the arguments of the the minimisation function are invariant to the relabelling $\theta_{k+1} \rightarrow \theta_k$ & $\theta_k \rightarrow \theta_{k+1}$ (except their ordering is switched). Thus $T(\theta_k|\theta_{k+1})$ can be written as

$$T(\theta_k|\theta_{k+1}) = \frac{1}{p(\theta_{k+1})} \min (p(\theta_k)q(\theta_{k+1}|\theta_k), p(\theta_{k+1})q(\theta_k|\theta_{k+1})) . \quad (\text{D.4})$$

Substituting these into the detailed balance equation gives

$$\begin{aligned} T(\theta_{k+1}|\theta_k)p(\theta_k) &= \frac{1}{p(\theta_k)} \min (p(\theta_k)q(\theta_{k+1}|\theta_k), p(\theta_{k+1})q(\theta_k|\theta_{k+1})) p(\theta_k) \\ &= \frac{1}{p(\theta_{k+1})} \min (p(\theta_k)q(\theta_{k+1}|\theta_k), p(\theta_{k+1})q(\theta_k|\theta_{k+1})) p(\theta_{k+1}) \\ &= T(\theta_k|\theta_{k+1})p(\theta_{k+1}), \end{aligned} \quad (\text{D.5})$$

and thus the relation is satisfied for scenario I.

Scenario II occurs when either the MH step is accepted and the sampled point happens to be θ_k , or when the MH step is rejected

$$T(\theta_k|\theta_k) = \alpha(\theta_k, \theta_k)q(\theta_k|\theta_k) + \int_{\hat{\theta}} q(\theta'|\theta_k)(1 - \alpha(\theta', \theta_k))d\theta', \quad (\text{D.6})$$

where $\hat{\theta}$ is the domain of p . Note that substituting $T(\theta_k|\theta_k)$ into the detailed balance equation gives $T(\theta_k|\theta_k)p(\theta_k) = T(\theta_k|\theta_k)p(\theta_k)$ and so scenario II trivially satisfies the relation.

D.1.2 Deriving the Metropolis-Hastings acceptance ratio from the detailed balance relation

We can use the same scenario analysis to derive the MH acceptance ratio from the detailed balance relation. For scenario I

$$\begin{aligned} \alpha(\theta_{k+1}, \theta_k)q(\theta_{k+1}|\theta_k)p(\theta_k) &= T(\theta_{k+1}|\theta_k)p(\theta_k) \\ &= T(\theta_k|\theta_{k+1})p(\theta_{k+1}) \\ &= \alpha(\theta_k, \theta_{k+1})q(\theta_k|\theta_{k+1})p(\theta_{k+1}), \end{aligned} \quad (\text{D.7})$$

which can be rearranged to give

$$\frac{\alpha(\theta_{k+1}, \theta_k)}{\alpha(\theta_k, \theta_{k+1})} = \frac{q(\theta_k | \theta_{k+1})p(\theta_{k+1})}{q(\theta_{k+1} | \theta_k)p(\theta_k)}. \quad (\text{D.8})$$

For α to be a probability it must be bounded by $[0, 1]$, thus we need a form for α which satisfies this constraint and equation D.8. If we consider the case where $q(\theta_k | \theta_{k+1})p(\theta_{k+1}) > q(\theta_{k+1} | \theta_k)p(\theta_k)$ then we are saying if $\alpha(\theta_{k+1}, \theta_k) = \alpha(\theta_k, \theta_{k+1})$ then the system moves from θ_{k+1} to θ_k more often than the reverse process happens, and thus detailed balance is violated. This asymmetry suggests that we should maximise $\alpha(\theta_{k+1}, \theta_k)$ i.e. set it equal to one, in which case we can say that

$$\frac{1}{\alpha(\theta_k, \theta_{k+1})} = \frac{q(\theta_k | \theta_{k+1})p(\theta_{k+1})}{q(\theta_{k+1} | \theta_k)p(\theta_k)} > 1, \quad (\text{D.9})$$

which is satisfied by the equality

$$\alpha(\theta_k, \theta_{k+1}) = \min \left(\frac{q(\theta_{k+1} | \theta_k)p(\theta_k)}{q(\theta_k | \theta_{k+1})p(\theta_{k+1})} \right). \quad (\text{D.10})$$

Similarly in the case that $q(\theta_k | \theta_{k+1})p(\theta_{k+1}) < q(\theta_{k+1} | \theta_k)p(\theta_k)$ we can set $\alpha(\theta_k, \theta_{k+1}) = 1$ and deduce the value of $\alpha(\theta_{k+1}, \theta_k)$. Substituting both of these expressions back in equation D.8 satisfies the equality, and thus we have verified the MH acceptance ratio satisfies detailed balance for scenario I.

Scenario II satisfies detailed balance trivially for any α and so does not place any additional constraints on its form. Thus the relation derived from equation D.8 is valid for both scenarios.

Note that the way we have derived the MH acceptance probability implies that there is no alternative acceptance probability β that satisfies $\beta(\theta', \theta_k) > \alpha(\theta', \theta_k)$, that does not violate either $\beta \in [0, 1]$ or the detailed balance relation.

D.2 Evidence integral transformation

We now show that equations 2.43 and 9.17 are equivalent for the case of a one-dimensional parameter problem: $\boldsymbol{\theta} \equiv \theta$, but note that this proof holds for $\boldsymbol{\theta}$ of arbitrary dimension.

The expression for X can be re-written as

$$X(\lambda) = \int_{\hat{\theta}} \pi(\theta) H(\mathcal{L}(\theta) - \lambda) d\theta, \quad (\text{D.11})$$

where $\hat{\theta}$ is the support of $\pi(\theta)$ and $H(x)$ is the Heaviside step function which satisfies $\frac{d}{dx} H(x) = \delta(x)$. Taking the derivative of equation D.11 with respect to λ gives

$$\frac{dX}{d\lambda} = - \int_{\hat{\theta}} \pi(\theta) \delta(\mathcal{L}(\theta) - \lambda) d\theta. \quad (\text{D.12})$$

Equation 9.17 can be re-expressed as an integral over \mathcal{L} by making a change of variables from X to \mathcal{L} . Recalling that $\lambda(X) \equiv \mathcal{L}(X)$,

$$\begin{aligned}\mathcal{Z} &= \int_0^1 \mathcal{L}(X) dX \\ &= \int \lambda \cdot \frac{dX}{d\lambda} d\lambda \\ &= - \int \mathcal{L} \int_{\hat{\theta}}^{\lambda} \pi(\theta) \delta(\mathcal{L}(\theta) - \lambda) d\theta d\lambda.\end{aligned}\tag{D.13}$$

Since the integrals are over all possible values of θ and λ and the limits are not a function of the other variable of integration (i.e. $\text{limits}(\theta) \neq g(\lambda)$ and vice versa), the order of integration can be switched according to Fubini's theorem

$$\begin{aligned}\mathcal{Z} &= \int_{\hat{\theta}} \pi(\theta) \left(- \int \lambda \delta(\mathcal{L}(\theta) - \lambda) d\lambda \right) d\theta \\ &= \int_{\hat{\theta}} \mathcal{L}(\theta) \pi(\theta) d\theta,\end{aligned}\tag{D.14}$$

which is the form for \mathcal{Z} given by equation 2.43.

D.3 Probability density function of the largest of n numbers

Let u_1, \dots, u_n be n random variables from the uniform distribution on $[0, 1]$ and let u_m be their maximum. We can derive $P(t)$ by considering the cumulative distribution function of u_m , $F(u_m)$. u_m is less than some value t if and only if all u_1, \dots, u_n are less than t . Therefore

$$F(u_m = t) = F(u_1 = t, \dots, u_n = t).\tag{D.15}$$

Since the u_j 's are independent this is the same as

$$F(u_1 = t) \dots F(u_n = t).\tag{D.16}$$

If t is also defined on $[0, 1]$, then since u_j 's are uniformly distributed we get

$$F(u_m = t) = t \dots t = t^n.\tag{D.17}$$

Hence the probability density function for u_m is

$$F'(u_m = t) = P(t) = nt^{n-1}.\tag{D.18}$$

D.4 Determining $\log(\mathbb{E}[\mathcal{Z}])$ and $\log(\mathbb{E}[\mathcal{Z}^2])$ from $\log(\mathcal{L})$ and $\log(X)$.

For the standard quadrature approximation of \mathcal{Z} (equation 9.18) and statistical treatment of t given by equation 9.22, Keeton (2011) derives expressions for $\mathbb{E}[\mathcal{Z}]$ and $\text{var}[\mathcal{Z}]$. The form of $\text{var}[\mathcal{Z}]$ derived incorporates the covariance between the value of \mathcal{Z} obtained from the main nested sampling algorithm loop and that obtained from the final contribution to the evidence after the main loop has terminated (\mathcal{Z}_f , see Section 9.4.1.2). When deriving these equations Keeton works in linear space, which is valid as long as $\sqrt{\text{var}[\mathcal{Z}]}/\mathbb{E}[\mathcal{Z}] \ll 1$, as in this limit \mathcal{Z} is normally distributed. As stated in Section 9.4.1.1 \mathcal{Z} is log-normally distributed in general and thus we should quote the statistics given by equations 9.23 and 9.24. Furthermore, working in linear space can lead to numerical difficulties if \mathcal{L} and X are sufficiently small / large, as is the case in the nested sampling example considered in Section 10.4.4. We can adapt the equations derived by Keeton to calculate $\log(\mathbb{E}[\mathcal{Z}])$ & $\log(\mathbb{E}[\mathcal{Z}^2])$ from $\log(\mathcal{L})$ & $\log(X)$ to obtain estimates of $\mathbb{E}[\log(\mathcal{Z})]$ & $\text{var}[\log(\mathcal{Z})]$ (which hopefully avoid numerical under / overflow issues), as follows.

We first define a function L which takes a vector \mathbf{x} as an input, exponentiates this vector component-wise, adds together the resultant values and then takes the logarithm of this sum (known as the LogSumExp function)^a

$$L(\mathbf{x}) = \log \left(\sum_{i=1}^{i=n} \exp(x_i) \right). \quad (\text{D.19})$$

We also define $\log(\mathbf{x}) \equiv (\log(x_1), \dots, \log(x_n))$. $\log(\mathbb{E}[\mathcal{Z}])$ (c.f. equation 17 of Keeton) can then be calculated as

$$\log(\mathbb{E}[\mathcal{Z}]) = L(\log(\mathcal{L}) + \log(\delta X)) - \log(n_l), \quad (\text{D.20})$$

where $\log(\mathcal{L})$ is the vector of $\log \mathcal{L}_i$ values obtained in the main nested sampling loop and $\log(\delta X) = (\log(\mathbb{E}[t]), \dots, n_s \log(\mathbb{E}[t]))$. Note that $\log(\mathcal{L})$ and $\log(\delta X)$ are both vectors of length n_s .

$\log(\mathbb{E}[\mathcal{Z}^2])$ (c.f. equation 22 of Keeton) is given by

$$\log(\mathbb{E}[\mathcal{Z}^2]) = \log \left(\frac{2}{(n_l(n_l + 1))} \right) + L(\log(\mathcal{L}) + \log(\delta X) + \log(\mathbf{I})), \quad (\text{D.21})$$

^aUnderflow and / or overflow issues can be avoided to some extent using the trick given in <https://hips.seas.harvard.edu/blog/2013/01/09/computing-log-sum-exp/>.

where $\log(\mathbf{I}) = (\log(I_1), \dots, \log(I_k), \dots, \log(I_{ns}))$ and

$$\log(I_k) = L \left(\log(\mathcal{L}_k) + \log \left(\mathbb{E}[\mathbf{t}^2]_k \right) - \log(\delta X_k) \right). \quad (\text{D.22})$$

Here the vector quantities denoted \mathbf{x}_k each have length k , and $\log(\mathbb{E}[\mathbf{t}^2]_k) = (\log(\mathbb{E}[t^2]), \dots, k \log(\mathbb{E}[t^2]))$. The expected contribution to the evidence after the nested sampling algorithm loop terminates, $\log(\mathbb{E}[\mathcal{Z}_f])$ can be determined from the log-likelihood values of the final set of livepoints $\log(\mathcal{L}_f) = (\mathcal{L}_1, \dots, \mathcal{L}_l)$, and the remaining prior volume X_{ns} through

$$\log(\mathbb{E}[\mathcal{Z}_f]) = \log(X_{ns}) - \log(n_l) + L(\log(\mathcal{L}_f)). \quad (\text{D.23})$$

Similarly the log of the second moment of \mathcal{Z}_f (equation 28 of Keeton) is given by

$$\log(\mathbb{E}[\mathcal{Z}_f^2]) = L(\log(\mathcal{L}_f)) - 2 \log(n_l) + n_s \log(\mathbb{E}[t^2]). \quad (\text{D.24})$$

Finally, the log of the cross term $\mathbb{E}[\mathcal{Z}\mathcal{Z}_f]$ (Keeton equation 32) can be calculated as

$$\log(\mathbb{E}[\mathcal{Z}\mathcal{Z}_f]) = L(\log(\mathcal{L}_f)) + \log(\delta X_{ns}) - \log(n_l(n_l + 1)) + \log(I_{ns}). \quad (\text{D.25})$$

$\log(\mathbb{E}[\mathcal{Z}])$ & $\log(\mathbb{E}[\mathcal{Z}^2])$ are then updated as

$$\log(\mathbb{E}[\mathcal{Z}]) \rightarrow L((\log(\mathbb{E}[\mathcal{Z}]), \log(\mathbb{E}[\mathcal{Z}_f]))), \quad (\text{D.26})$$

$$\log(\mathbb{E}[\mathcal{Z}^2]) \rightarrow L((\log(\mathbb{E}[\mathcal{Z}^2]), \log(\mathbb{E}[\mathcal{Z}_f^2]), \log(2) + \log(\mathbb{E}[\mathcal{Z}\mathcal{Z}_f]))), \quad (\text{D.27})$$

and used in equations 9.23 & 9.24 to calculate $\mathbb{E}[\log(\mathcal{Z})]$ & $\text{var}[\log(\mathcal{Z})]$.

REFERENCES

Abbott B. P., et al., 2016, *PhRvL*, 116, 061102

Abbott B. P., et al., 2017, *PhRvL*, 119, 141101

Adam R., et al., 2016, *A&A*, 586, A122

Adam R., et al., 2015, *A&A*, 576, A12

Allen S. W., Evrard A. E., Mantz A. B., 2011, *ARA&A*, 49, 409

Allen S. W., Rapetti D. A., Schmidt R. W., Ebeling H., Morris R. G., Fabian A. C., 2008, *MNRAS*, 383, 879

Ameglio S., Borgani S., Pierpaoli E., Dolag K., Ettori S., Morandi A., 2009, *MNRAS*, 394, 479

Anders E., Grevesse N., 1989, *GeCoA*, 53, 197

Arnaud K. A., et al., 1994, *ApJ*, 436, L67

Arnaud M., Pratt G. W., Piffaretti R., Böhringer H., Croston J. H., Pointecouteau E., 2010, *A&A*, 517, A92

Asada H., Futamase T., 1997, *PThPS*, 128, 123

Böhringer H., et al., 2007, *A&A*, 469, 363

Bahcall J. N., Sarazin C. L., 1977, *ApJ*, 213, L99

Barry J. R., Lee E. A., Messerschmitt D. G., 2004, *Digital Communications*, Springer

Bartlett J. G., Silk J., 1994, *ApJ*, 423, 12

Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012, *ApJ*, 758, 74

Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., Sijacki D., 2010, *ApJ*, 725, 91

Battaglia N., Bond J., Sievers J., Pfrommer C., 2011, *AAS*, 43, 109.03

Bautz M. W., et al., 2009, *PASJ*, 61, 1117

Bennett C. L., et al., 2013, *ApJS*, 208, 20

Biffi V., Dolag K., Böhringer H., 2011, *MNRAS*, 413, 573

Birkinshaw M., 1999, *PhR*, 310, 97

Blanchet L., 2014, *LRR*, 17, 2

Blanchet L., Iyer B. R., Will C. M., Wiseman A. G., 1996, CQGra, 13, 575

Brunetti G., Jones T. W., 2014, IJMPD, 23, 1430007-98

Carlstrom J. E., Holder G. P., Reese E. D., 2002, ARA&A, 40, 643

Carvalho P., Rocha G., Hobson M. P., 2009, MNRAS, 393, 681

Carvalho P., Rocha G., Hobson M. P., Lasenby A., 2012, MNRAS, 427, 1384

Cavaliere A., Fusco-Femiano R., 1978, A&A, 70, 677

Cavaliere A., Fusco-Femiano R., 1976, A&A, 49, 137

Challinor A., Lasenby A., 1998, ApJ, 499, 1

Chevallier M., Polarski D., 2001, IJMPD, 10, 213

Chib S., Greenberg E., 1995, American Statistical Association, 49, 4

Chopin N., Robert C., 2008, arXiv, arXiv:0801.3887

Corless V. L., King L. J., Clowe D., 2009, MNRAS, 393, 1235

Czakon N. G., et al., 2015, ApJ, 806, 18

AMI Consortium, et al., 2011, MNRAS, 415, 2708

Davies J. I., et al., 2011, MNRAS, 415, 1883

Dolag K., Bykov A. M., Diaferio A., 2008, SSRv, 134, 311

Dolag K., Jubelgas M., Springel V., Borgani S., Rasia E., 2004, ApJ, 606, L97

Donnarumma A., et al., 2011, A&A, 528, A73

Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

Einasto J., 1965, TrAlm, 5, 87

Ettori S., Morandi A., Tozzi P., Balestra I., Borgani S., Rosati P., Lovisari L., Terenziani F., 2009, A&A, 501, 61

Feroz F., Hobson M. P., 2008, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2011, ascl.soft, ascl:1109.006

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Feroz F., Hobson M. P., Zwart J. T. L., Saunders R. D. E., Grainge K. J. B., 2009, MNRAS, 398, 2049

Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, ApJ, 473, 576

AMI Consortium, et al., 2011, MNRAS, 415, 2699

Furlanetto S. R., Loeb A., 2001, ApJ, 556, 619

Gawiser E., Silk J., 2000, PhR, 333, 245

George M. R., Fabian A. C., Sanders J. S., Young A. J., Russell H. R., 2009, MNRAS, 395, 657

Giodini S., Lovisari L., Pointecouteau E., Ettori S., Reiprich T. H., Hoekstra H., 2013, SSRv, 177, 247

Gioia I. M., Maccacaro T., Schild R. E., Wolter A., Stocke J. T., Morris S. L., Henry J. P., 1990, ApJS, 72, 567

Giovannini G., Bonafede A., Feretti L., Govoni F., Murgia M., Ferrari F., Monti G., 2009, A&A, 507, 1257

Grainge K., Jones M. E., Pooley G., Saunders R., Edge A., Grainger W. F., Kneissl R., 2002, MNRAS, 333, 318

Grainger W. F., Das R., Grainge K., Jones M. E., Kneissl R., Pooley G. G., Saunders R. D. E., 2002, MNRAS, 337, 1207

Gunn J. E., 1978, obco.meet, 1

Haehnelt M. G., Tegmark M., 1996, MNRAS, 279, 545

Hancock S., Davies R. D., Lasenby A. N., de La Cruz C. M. G., Watson R. A., Rebolo R., Beckman J. E., 1994, Natur, 367, 333

Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 453, 4384

Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 450, L61

Handley W. J., 2016, PhD Thesis, University of Cambridge

Hao J., et al., 2010, ApJS, 191, 254

Hasler N., et al., 2012, ApJ, 748, 113

Hastings W. K., 1970, Bimka, 57, 97

Hauser K., 2013, "Why does the Metropolis-Hastings procedure satisfy the detailed balance criterion?", URL:<http://people.duke.edu/~kh269/teaching/notes/MetropolisExplanation.pdf>

Herranz D., Sanz J. L., Hobson M. P., Barreiro R. B., Diego J. M., Martínez-González E., Lasenby A. N., 2002, MNRAS, 336, 1057

Hickish J., et al., 2018, MNRAS, 475, 5677

Higson E., Handley W., Hobson M., Lasenby A., 2017, arXiv, arXiv:1703.09701

Hinshaw G., et al., 2013, ApJS, 208, 19

Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377

Hobson M. P., Efstathiou G. P., Lasenby A. N., 2006, General relativity: an introduction for physicists, Cambridge University Press

Hobson M. P., Maisinger K., 2002, MNRAS, 334, 569

Hoshino A., et al., 2010, PASJ, 62, 371

Hu W., Kravtsov A. V., 2003, *ApJ*, 584, 702

Itoh N., Kohyama Y., Nozawa S., 1998, *ApJ*, 502, 7

Javid K., Perrott Y. C., Hobson M. P., Olamaie M., Rumsey C., Saunders R. D. E., 2018, arXiv, arXiv:1805.01968

Javid K., Perrott Y. C., Rumsey C., Saunders R. D. E., 2018, arXiv, arXiv:1809.03325

Javid K., Olamaie M., Perrott Y. C., Carvalho P., Grainge K. J. B., Hobson M. P., Rumsey C., Saunders R. D. E., 2019, *MNRAS*, 483, 3529

Javid K., 2018, arXiv, arXiv:1905.09110

Jeffreys H., 1961, *Theory of Probability*, Oxford University Press

Johansen A. M., Evers L., 2007, "Monte Carlo Methods", URL:<https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/johansen/teaching/mcm-2007.pdf>

Kawaharada M., et al., 2010, *ApJ*, 714, 423

Keeton C. R., 2011, *MNRAS*, 414, 1418

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340

Koester B. P., et al., 2007, *ApJ*, 660, 221

Kogure T., Leung K. C., 2007, *The astrophysics of emission-line stars*, Springer

Kokkotas K. D., 2008, *RvMA*, 20, 140

Komatsu E., et al., 2011, *ApJS*, 192, 18

Kompaneets A. S. 1957, *Soviet Physics JETP*, 4, 730

Korngut P. M., et al., 2011, *ApJ*, 734, 10

Kravtsov A. V., Vikhlinin A., Nagai D., 2006, *ApJ*, 650, 128

Lahav O., Bridle S. L., Hobson M. P., Lasenby A. N., Sodré L., 2000, *MNRAS*, 315, L45

Lange J., LIGO-Virgo Collaboration, 2018, *AAS*, 231, #325.01

LaRoque S. J., Bonamente M., Carlstrom J. E., Joy M. K., Nagai D., Reese E. D., Dawson K. S., 2006, *ApJ*, 652, 917

Larson D., et al., 2011, *ApJS*, 192, 16

Lau E. T., Kravtsov A. V., Nagai D., 2009, *ApJ*, 705, 1129

Lay O. P., Halverson N. W., 2000, *ApJ*, 543, 787

Levina E., Bickel P., 2001, *ICCV*, 251-256

Ma Y.-Z., Berndsen A., 2014, *A&C*, 5, 45

Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, *MNRAS*, 391, 1940

Macciò A. V., Dutton A. A., van den Bosch F. C., Moore B., Potter D., Stadel J., 2007, *MNRAS*, 378, 55

Maciel T. R., 2014, PhD Thesis, University of Cambridge

MacKay D. J. C., 2002, Information theory, inference & learning algorithms, Cambridge University Press

Mantz A., Allen S. W., Ebeling H., Rapetti D., 2008, MNRAS, 387, 1179

Mantz A., Allen S. W., Rapetti D., Ebeling H., 2010, MNRAS, 406, 1759

Markevitch M., Mushotzky R., Inoue H., Yamashita K., Furuzawa A., Tawara Y., 1996, ApJ, 456, 437

Marriage T. A., et al., 2011, ApJ, 737, 61

Marriage T. A., et al., 2011, ApJ, 731, 100

Marshall P. J., Hobson M. P., Slosar A., 2003, MNRAS, 346, 489

Martizzi D., Agrusa H., 2016, arXiv, arXiv:1608.04388

Martizzi D., Mohammed I., Teyssier R., Moore B., 2014, MNRAS, 440, 2290

Mason B. S., Myers S. T., 2000, ApJ, 540, 614

Mehrtens N., et al., 2012, MNRAS, 423, 1024

Melin J.-B., Bartlett J. G., Delabrouille J., 2006, A&A, 459, 341

Meneghetti M., et al., 2014, ApJ, 797, 34

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, JChPh, 21, 1087

Monfardini A., et al., 2014, JLTP, 176, 787

Muchovej S., et al., 2011, ApJ, 732, 28

Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJ, 638, L51

Nagai D., 2011, MmSAI, 82, 594

Nagai D., Kravtsov A. V., Vikhlinin A., 2007, ApJ, 668, 1

Nagai D., Lau E. T., 2011, ApJ, 731, L10

Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720

Neto A. F., et al., 2007, MNRAS, 381, 1450

Nozawa S., Itoh N., Kohyama Y., 1998, ApJ, 508, 17

Nyquist H., 1928, PhRv, 32, 110

Olamaie M., Hobson M. P., Grainge K. J. B., 2013, MNRAS, 430, 1344

Olamaie M., Hobson M. P., Grainge K. J. B., 2012, MNRAS, 423, 1534

AMI Consortium, et al., 2012, MNRAS, 421, 1136

Parrish I. J., McCourt M., Quataert E., Sharma P., 2012, MNRAS, 419, L29

Perley R. A., Butler B. J., 2013, ApJS, 204, 19

Perrott Y. C., et al., 2015, *A&A*, 580, A95

Perrott Y. C., Javid K., Carvalho P., Elwood P. J., Hobson M. P., Lasenby A. N., Olamaie M., Saunders R. D. E., 2019, *MNRAS*, 486, 2116

Petrov L., Kovalev Y. Y., Fomalont E. B., Gordon D., 2008, *AJ*, 136, 580

Pfrommer C., Enßlin T. A., Springel V., Jubelgas M., Dolag K., 2007, *MNRAS*, 378, 385

Piffaretti R., Arnaud M., Pratt G. W., Pointecouteau E., Melin J.-B., 2011, *A&A*, 534, A109

Piffaretti R., Valdarnini R., 2008, *A&A*, 491, 71

Planck Collaboration XI, 2011, *A&A*, 536, A11

Planck Collaboration VIII, 2011, *A&A*, 536, A8

Planck Collaboration IX, 2011, *A&A*, 536, A9

Planck Collaboration Int. III., 2013, *A&A*, 550, A129

Planck Collaboration XX, 2014, *A&A*, 571, A20

Planck Collaboration XXIX, 2014, *A&A*, 571, A29

Planck Collaboration Int. XXVI, 2015, *A&A*, 582, A29

Planck Collaboration XXXII, 2015, *A&A*, 581, A14

Planck Collaboration XXVII, 2016, *A&A*, 594, A27

Planck Collaboration Int. XXXVI, 2016, *A&A*, 586, A139

Pointecouteau E., Giard M., Barret D., 1998, *A&A*, 336, 44

Rasia E., Tormen G., Moscardini L., 2004, *MNRAS*, 351, 237

Reiprich T. H., et al., 2009, *A&A*, 501, 899

Rephaeli Y., 1995, *ARA&A*, 33, 541

Robert C., Casella G., 2002, *Monte Carlo statistical methods*, Springer

Romano A., et al., 2010, *A&A*, 514, A88

Romero C., et al., 2017, arXiv, arXiv:1707.06113

Romero C. E., et al., 2015, *ApJ*, 807, 121

Rosati P., Borgani S., Norman C., 2002, *ARA&A*, 40, 539

Rozo E., Rykoff E. S., Bartlett J. G., Melin J.-B., 2015, *MNRAS*, 450, 592

Rubner Y., Tomasi C., Guibas L. J., 1998, *IEEE ICCV*, 59-66

Rudy D. J., Muhleman D. O., Berge G. L., Jakosky B. M., Christensen P. R., 1987, *Icar*, 71, 159

Rumsey C., et al., 2016, *MNRAS*, 460, 569

Rykoff E. S., et al., 2014, *ApJ*, 785, 104

Ryle M., Vonberg D. D., 1948, *RSPSA*, 193, 98

Sayers J., et al., 2013, *ApJ*, 768, 177

Scheuer P. A. G., 1957, PCPS, 53, 764

Schmidt R. W., Allen S. W., 2007, MNRAS, 379, 209

Schramm T., Kayser R., 1994, A&A, 289, L5

Sehgal N., et al., 2011, ApJ, 732, 44

Seljak U., Slosar A., McDonald P., 2006, JCAP, 10, 014

Sereno M., Fedeli C., Moscardini L., 2016, JCAP, 1, 042

Shaw J. R., Bridges M., Hobson M. P., 2007, MNRAS, 378, 1365

AMI Consortium, et al., 2011, arXiv, arXiv:1101.5590

Silk J., 1968, ApJ, 151, 459

Silk J., 1967, Natur, 215, 1155

Simionescu A., et al., 2011, Sci, 331, 1576

Sivia D. S., Skilling J., 2006, Data analysis: a Bayesian tutorial, Oxford University Press

Skilling J., 2004, AIPC, 735, 395

Smoot G. F., et al., 1992, ApJ, 396, L1

Staniszewski Z., et al., 2009, ApJ, 701, 32

Sunyaev R. A., Zeldovich Y. B., 1970, CoASP, 2, 66

Teyssier R., 2002, A&A, 385, 337

The Planck Collaboration, 2006, astro, arXiv:astro-ph/0604069

Thompson R. A., Moran J., Swenson G. W., 2007, Interferometry and Synthesis in Radio Astronomy, Springer

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709-728

Urban O., Werner N., Simionescu A., Allen S. W., Böhringer H., 2011, MNRAS, 414, 2101

Völk H. J., Atoyan A. M., 2000, ApJ, 541, 88

Vazza F., Brüggen M., Gheller C., Brunetti G., 2012, MNRAS, 421, 3375

Venturi T., Giacintucci S., Dallacasa D., Cassano R., Brunetti G., Bardelli S., Setti G., 2008, A&A, 484, 327

Vikhlinin A., Kravtsov A., Forman W., Jones C., Markevitch M., Murray S. S., Van Speybroeck L., 2006, ApJ, 640, 691

Vikhlinin A., et al., 2009, ApJ, 692, 1060

Vladimirov A., Ellison D. C., Bykov A., 2006, ApJ, 652, 1246

Voges W., et al., 1999, A&A, 349, 389

Waldram E., Bolton R., Pooley G. G., Riley J. M., 2007, mru..conf, 140

Waldrum E. M., Bolton R. C., Pooley G. G., Riley J. M., 2007, yCat, 837,

Watson R. A., et al., 2003, MNRAS, 341, 1057

Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2001, astro, arXiv:astro-ph/0111069

Weisskopf M. C., Brinkman B., Canizares C., Garmire G., Murray S., Van Speybroeck L. P., 2002, PASP, 114, 1

Wen Z. L., Han J. L., Liu F. S., 2012, ApJS, 199, 34

Williamson R., et al., 2011, ApJ, 738, 139

Willis J. P., et al., 2013, MNRAS, 430, 134

Wright E. L., et al., 1992, ApJ, 396, L13

Voit G. M., 2005, RvMP, 77, 207

York D. G., et al., 2000, AJ, 120, 1579

Zaroubi S., 2013, ASSL, 396, 45

Zeldovich Y. B., Sunyaev R. A., 1969, Ap&SS, 4, 301

Zwart J. T. L., et al., 2008, MNRAS, 391, 1545

Zwicky F., 1933, AcHPh, 6, 110

Zwicky F., 1937, ApJ, 86, 217