

PhD

3.º
CICLO
FCUP
2018

U. PORTO

Estimating galaxy cluster properties using
Hierarchical Bayesian Models

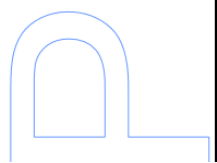
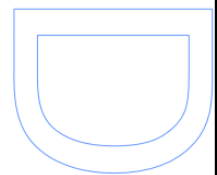
Leyla Seyed Ebrahimpour

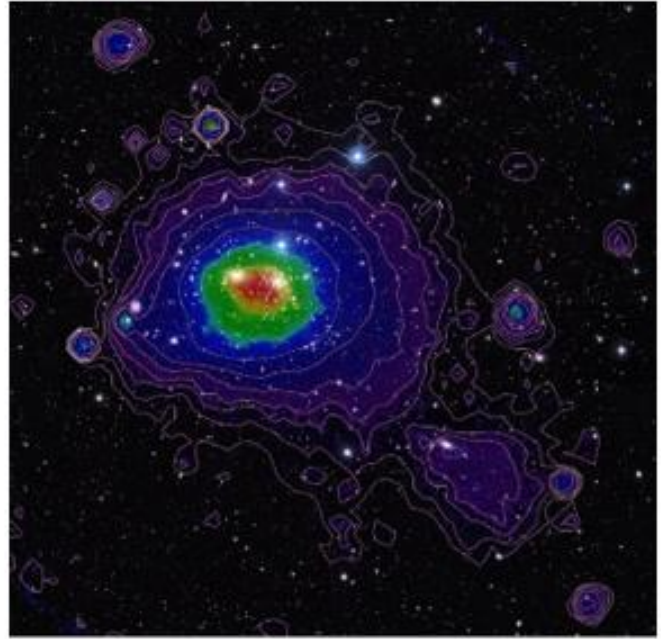
FC

Estimating galaxy cluster properties using Hierarchical Bayesian Models

Leyla Seyed Ebrahimpour

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto
Departamento de Física e Astronomia
2018





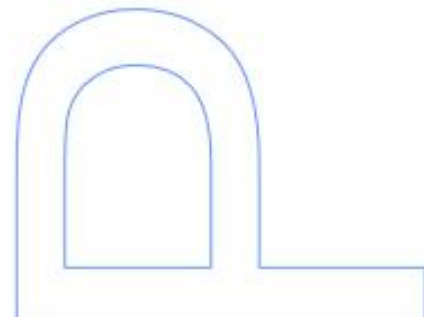
Estimating galaxy cluster properties using Hierarchical Bayesian Models

Leyla Seyed Ebrahimpour

Doutoramento em Astronomia
Departamento de Física e Astronomia
2018

Orientador

Pedro T. P. Viana, Professor Auxiliar,
Faculdade de Ciências



Leyla Seyed Ebrahimpour

Estimating galaxy cluster properties using Hierarchical Bayesian Models



Centro de Astrofísica da Universidade do Porto

Departamento de Física e Astronomia

Faculdade de Ciências

Tese submetida à Faculdade de Ciências da Universidade do Porto para obtenção do
grau de Doutor em Astronomia

Orientador:

Pedro T. P. Viana

December 2018

Acknowledgments

I would like to acknowledge my supervisor, Pedro who accompanied me throughout this project with his patience, understanding, and support. He helped me from the moment I arrived at Porto and did not hesitate to help me whenever I needed. I learned a lot from him and this project undoubtedly would not have been fulfilled without his supervising. I also thank IA and CAUP for their support, in particular, my office-mates who made this place an ideal place to work.

Hodjat, my dear husband, you are a real companion, who have always encouraged me and accompanied me in each step of the difficult path of science and life. I would say from the bottom of my heart that without you being able to make this path was impossible for me.

Sotoudeh, my little sweetheart and artist, thank you for standing my absence for all the moments that I should have been with you and I was not. Your art and your scientific progress at school inspired me to continue this path.

I would like to thank everyone who made me arrive at this stage of my life, people unrelated to the thesis: every single member of my energetic family for their endless support from far away even though they never wanted me to leave home: my parents, my sisters, my brothers and their lovely families, Hodjat's parents and his lovely family who always cared a lot about me. My special thanks to my father, Abbas, and my mother, Fatemeh, who always wished for my academic achievements; "Asheghetam Maman Nasiri-e-Man". I am grateful to have a brother like Mahmoud, who helped me a lot during these years and cared about me like a father. Finally, I would like to thank my dear friends in Portugal, who always have been reasons for happiness in this distance away from hometown.

I also acknowledge the support by the fellowship SFRH/BD/52138/2013 funded by Fundação para a Ciência e Tecnologia (FCT) and the support of the International Doctorate Network in Particle Physics, Astrophysics and Cosmology (IDPASC) for giving me the opportunity of this fellowship.

Abstract

The main goal of this thesis was to develop parametric and non-parametric statistical tools to infer galaxy cluster scaling relations and individual cluster properties in different wavelengths. Several properties of clusters and groups of galaxies, like the X-ray temperature and luminosity, seem to show different scaling behaviors depending on whether one considers the cluster or the group regime. Their study sheds light on the distinct assembly histories of these structures, and leads to a better understanding of the physical processes involved. Further, the precise knowledge of the cluster mass is needed to constrain cosmological parameters, namely through the cluster mass function. But direct measurement of cluster masses is not feasible for the thousands of objects required. Hence, indirect methods for cluster mass estimation have to be used, based on the correlations that exist between cluster properties.

In Chapter 2, we introduce the Bayesian statistical framework used for the data analysis, where both parametric and non-parametric methods were considered. Chapter 3 reports the results obtained by applying a parametric hierarchical Bayesian algorithm to a sample of 353 clusters and groups of galaxies from the XMM-Newton Cluster Survey (XCS) with X-ray temperatures in excess of 1 keV , which have also been identified in Sloan Digital Sky Survey (SDSS) data using the redMaPPer algorithm. We allow for the normalization and intrinsic scatter of the $L_X - T$ relation to evolve with time, as well as for the possibility of a temperature dependent change-point in the exponent of such relation. We do not find strong statistical support for any deviation, from the usual modelling of the $L_X - T$ relation as a single power-law. However, there is some suggestion of a transition between the group and cluster regimes, slightly below 2 keV . Our results also point towards a possible increase in the slope of the $\log(L_X) - \log(T)$ relation when moving from the group to the cluster regime, and faster evolution in the former with respect to the later, driving the temperature dependent change point towards higher values with redshift. Chapter 4 describes the results obtained by applying Gaussian Processes (GPs) to samples of Planck SZ2, SDSS and MCXC clusters, in order to forecast their mass given, respectively, their Sunyaev- Zel'dovich (SZ) flux, optical richness or X-ray luminosity. A sample of known mass galaxy clusters, the Literature Catalogs of Weak Lensing Clusters of galaxies (LC2), was used as training data for the algorithm. The results show that as long as there are enough clusters with known Weak Lensing (WL) mass, GPs tend to predict higher masses than the linear regression method, as implemented by Sereno & Ettori 2017.

Resumo

O objetivo principal desta tese foi desenvolver ferramentas estatísticas paramétricas e não-paramétricas para inferir relações de escala para enxames de galáxias e propriedades de enxames individuais em diferentes comprimentos de onda. Diversas propriedades de grupos e grupos de galáxias, como a temperatura e a luminosidade nos raios X, parecem relacionar-se de modo diferente dependendo se considerarmos grupos ou enxames. O seu estudo lança luz sobre as distintas histórias evolutivas destas estruturas e leva a uma melhor compreensão dos processos físicos envolvidos. Além disso, o conhecimento preciso da massa de um enxame é necessário para restringir parâmetros cosmológicos, nomeadamente através da função de massa dos enxames. Mas a medição direta da massa de um enxame não é viável para os milhares de objetos necessários. Portanto, métodos indiretos para estimativa da massa de um enxame precisam de ser usados, com base nas correlações existentes entre as propriedades dos enxames.

No Capítulo 2, introduzimos o procedimento estatístico Bayesiana utilizado para a análise dos dados, onde foram considerados métodos paramétricos e não-paramétricos. O Capítulo 3 relata os resultados obtidos pela aplicação de um algoritmo bayesiano hierárquico paramétrico a uma amostra de 353 enxames e grupos de galáxias, parte da XMM-Newton Cluster Survey (XCS), com temperaturas de raios-X superiores a 1 keV , que também foram identificados em dados da Sloan Digital Sky Survey (SDSS) usando o algoritmo redMaPPer. Permitimos que a normalização e a dispersão intrínseca da relação $L_X - T$ evoluam com o tempo, bem como a possibilidade de um ponto de mudança dependente da temperatura no expoente dessa relação. Não encontramos forte suporte estatístico para qualquer desvio da modelação usual da relação $L_X - T$ como uma única lei de potência. No entanto, há sugestão de uma transição entre grupos e enxames, ligeiramente abaixo de 2 keV . Os nossos resultados também apontam para um possível aumento no declive da relação $\log(L_X) - \log(T)$ ao passarmos de grupos para enxames, e evolução mais rápida nos primeiros em relação ao segundos, conduzindo a um aumento da temperatura associada ao ponto de mudança para valores mais altos do desvio para o vermelho. O Capítulo 4 descreve os resultados obtidos pela aplicação de Processos Gaussianos (GPs) a amostras de enxames presentes nos catálogos Planck SZ2, SDSS e MCXC, com o objetivo de prever sua massa tendo em conta, respectivamente, a amplitude do efeito de Sunyaev-Zel'dovich (SZ), a riqueza ótica ou luminosidade nos raios-X. Uma amostra de enxames de galáxias com massa conhecida, os Literature Catalogs of Weak Lensing Clusters of galaxies (LC2), foi usada para treino do algoritmo. Os resultados mostram que, contanto que existam enxames suficientes com massa conhecida, os GPs tendem a prever massas mais altas do que o método de regressão linear, conforme implementado por Sereno & Ettori 2017.

Contents

Abstract	1
Resumo	3
Contents	5
List of Figures	7
List of Tables.....	13
1 Introduction.....	15
1.1 The beginning.....	15
1.2 Galaxy clusters.....	15
1.2.1 Properties	16
1.2.2 Scaling relations	19
1.2.3 Importance for Extragalactic Astronomy and Cosmology	22
1.3. Surveys of galaxy clusters.....	26
1.3.1 Sloan Digital Sky Survey.....	26
1.3.2 Planck Survey	27
1.3.3 ROSAT All Sky Survey.....	27
1.3.4 XMM Newton Cluster Survey	28
2 Bayesian statistical data analysis	30
2.1 Introduction.....	30
2.2 How to begin the Bayesian analysis?	31
2.2.1 Using Markov Chain Monte Carlo	31
2.3 Linear regression	32
2.4 Gaussian Processes	34
3 Joint modelling of the $L_X - T$ scaling relation.....	39
3.1 Introduction.....	39
3.2 Data	41
3.3 Analysis.....	44

3.3.1 Bayesian framework.....	44
3.3.2 Sample selection effects.....	49
3.3.3 Model comparison	51
3.4 Results	53
3.4.1 Without taking into account sample selection effects	53
3.4.2 Taking into account sample selection effects	73
3.5 Discussion	73
3.6 Conclusions.....	81
4 Mass estimation of galaxy clusters	83
4.1 Introduction.....	83
4.2 The datasets	84
4.2.1 The Planck SZ catalogue	84
4.2.2 The SDSS redMaPPer catalogue	85
4.2.3 Meta-Catalogue of X-ray detected Clusters of galaxies	86
4.2.4 Weak lensing data	87
4.3 Results and discussion.....	89
5 Conclusions and future work.....	99
5.1 Conclusions.....	99
5.2 Future work	100
Appendix	104
Bibliography.....	110

List of Figures

Figure 1. The image of Coma cluster observed through the SZ effect (left) obtained by Planck and X-rays (right) obtained by ROSAT. Both images are overlaid on the image of the cluster taken in visible light by DSS. Taken from Giodini et al. 2013.
..... 16

Figure 2. The dependency of the mass function on the cosmological models shown by Vikhlinin et al. 2009a. While in the left panel, the measured mass function and predicted models are shown for a cosmology with $\Lambda = 0.75$, the right panel shows a different model prediction and data measurements in the case of $\Lambda = 0$ at high redshifts. Taken from Giodini et al. 2013..... 24

Figure 3. The upper panel shows the estimated cluster counts for a survey that could detect halos more massive than $2 \times 10^{14} M_{\odot}$, for three cosmological models with fixed $\Omega_M = 0.3$ and $\sigma_8 = 0.9$. The difference between models relative to the statistical errors has been shown in the lower panel. Taken from Mohr 2005..... 25

Figure 4. View of the XMM-Newton spacecraft mirror modules (to the left) and the backend of the instrument platform with the radiators (to the right). Taken from Jansen et al. 2001. 28

Figure 5. Three different trace plots that imply using longer chains in MCMC. While the chain in the left panel shows a slow mixing, the central panel shows a chain, which stays in a fixed value for too long and the right panel shows two chains, which are still not converging given that the final values of each chain, is different. Taken from Andreon & Weaver 2015..... 32

Figure 6. The + symbols in each panel show data generated from Gaussian Processes with $(l, \sigma_f) = (1, 1)$ (top), $(l, \sigma_f) = (0.3, 1)$ (lower left) and $(l, \sigma_f) = (3, 1)$ (lower right). A 95%

confidence region has been obtained for the underlying function f , which is shown in grey for all the cases. Taken from Rasmussen & Williams 2006. 35

Figure 7. In panel (a), three random functions drawn from GPs prior have been shown. In panel (b), three random functions from GPs prior that are conditioned on the five indicated observational data have been shown. The grey area corresponds to the 95% confidence region in both plots. Taken from Rasmussen & Williams 2006... 37

Figure 8. The distributions of most probable values for the redshift, z , temperature, T (in keV), and X-ray luminosity, L_X (in erg/s), given the data, for the 353 XCS-DR2-SDSS-DR8-redMaPPer groups and clusters used in this work. 43

Figure 9. Most probable values for temperature, T (in keV , upper panel), and X-ray luminosity, L_X (in erg/s , lower panel), versus redshift, given the data, for the 353 XCS-DR2-SDSS-DR8-redMaPPer groups and clusters used in this work..... 44

Figure 10. The $L_X - T$ relation assuming the simplest model, with (top, model 0) and without (bottom, model 0u) sample selection effects taken into account. Each point indicates the most probable values for the temperature, T , in keV , and re-scaled bolometric X-ray luminosity, L_X , with respect to each of the 353 systems in our sample, given the X-ray data. The self-similar re-scaling of L_X with the cluster redshift, z , is performed dividing L_X by $E(z)$. The error bars associated with each point identify the 1σ uncertainty intervals. The solid regression line was determined by fixing the model hyper-parameters to their expected (i.e. mean) values. The amplitude of the intrinsic vertical scatter about the regression line is indicated by the dashed lines (1σ , inner blue; 2σ , outer red). 54

Figure 11. The $L_X - T$ relation assuming model 1, with (top, model 1) and without (bottom, model 1u) sample selection effects taken into account. In this case, the re-scaling of L_X with the cluster redshift, z , is performed dividing L_X by $E(z)$, where γ is the expected (i.e. mean) value for γ 56

Figure 12. The $L_X - T$ relation assuming model 2. The prior distribution for the temperature is modelled through a time-evolving Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only

at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts..... 57

Figure 13. The $L_X - T$ relation assuming model 3. The prior distribution for the temperature is modelled through a time-evolving Gaussian..... 57

Figure 14. Comparison of the one- and two-dimensional marginalised posterior distributions for model 0 vs 1. Note that the horizontal axis that is associated with each one-dimensional marginalised posterior distribution is below it, at the bottom of the plot. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability..... 58

Figure 15. Comparison of the one and two-dimensional marginalised posterior distributions for model 0 vs 2. Note that the horizontal axis that is associated with each one-dimensional marginalised posterior distribution is below it, at the bottom of the plot. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability..... 59

Figure 16. Comparison of the one- and two-dimensional marginalised posterior distributions for model 0 vs 3. Note that the horizontal axis that is associated with each one-dimensional marginalised posterior distribution is below it, at the bottom of the plot. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability..... 60

Figure 17. The $L_X - T$ relation assuming model 4. The prior distribution for the temperature is modelled through a time-evolving Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts..... 61

Figure 18. The $L_X - T$ relation assuming model 5. The prior distribution for the temperature is modelled through a time-evolving Gaussian..... 61

Figure 19. The $L_X - T$ relation assuming model 6. The prior distribution for the temperature is modelled through a time-evolving Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only

at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts.....	62
Figure 20. The $L_X - T$ relation assuming model 7. The prior distribution for the temperature is modelled through a Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts.....	62
Figure 21. Comparison of the one- and two-dimensional marginalised posterior distributions for model 1 vs. 5. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.	63
Figure 22. Comparison of the one- and two-dimensional marginalised posterior distributions for model 2 vs. 6. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.	64
Figure 23. Comparison of the one- and two-dimensional marginalised posterior distributions for model 4 vs. 7. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.	65
Figure 24. Comparison, for all 353 systems in our sample, of the expected values and associated standard deviations associated with: the observed temperatures and X-ray luminosities, T_{obs} and L_{obs} ; the temperatures and X-ray luminosities, T_{model} and L_{model} , obtained assuming a priori the $L_X - T$ relation to be described by model 4.	71
Figure 25. Comparison of the sample distributions of most probable values for the observed temperatures and X-ray luminosities, with those obtained through LIRA, assuming a priori the $L_X - T$ relation to be described by model 4.	72
Figure 26. The self-similar scaled comptonisation parameter versus redshift for 926 clusters of PSZ2 with available redshift.....	84
Figure 27. The richness versus redshift for clusters of the SDSS redMaPPer catalogue. Taken from Rykoff et al. 2013.	86

Figure 28. The luminosity versus redshift for clusters of MCXC catalogue. Taken from Piffaretti et al. 2011.	87
Figure 29. Redshift distribution of 485 WL clusters in the LC2-single catalogue. Taken from CoMaLit III.	87
Figure 30. Mass distribution of 485 WL clusters in the LC2-single catalogue. Taken from CoMaLit III.	88
Figure 31. The distribution of 485 WL clusters in the LC2-single catalogue in a mass-redshift plane. The mass is in units of $10^{14} M_{\odot}$. Taken from CoMaLit III.....	88
Figure 32. The predicted mass for 676 clusters of Planck-SZ2 (blue dots) using the integrated Compton parameter and WL mass of 135 clusters with available WL mass in LC2 (black dots) versus the integrated Compton parameter. The green points show the estimated mass in CoMaLit V and the red points show the estimated mass(upper panel) or estimated biased mass (lower panel) by Planck collaboration.	90
Figure 33. Comparison of the predicted masses for 676 clusters of Planck-SZ2 by GPs (blue circles) and Planck collaboration (red circles) to the masses estimated in CoMaLit V.....	91
Figure 34. Relative difference of masses predicted by Planck collaboration (red circles) and GPS (blue circles) with respect to the mass estimates by CoMaLit V versus the integrated Compton parameter.	92
Figure 35. The predicted mass for 2496 clusters of SDSS redMaPPer (blue diamonds) using the richness and WL mass of 144 clusters with available WL mass in LC2 (black dots). The green diamonds show the estimated mass in CoMaLit V.....	93
Figure 36. Comparison of the predicted masses for 2496 clusters of SDSS redMaPPer by GPs to the masses estimated in CoMaLit V.	94

Figure 37. Relative difference of GPs masses with respect to the mass estimates by CoMaLit V versus the richness.	94
Figure 38. The predicted mass for 1382 clusters of MCXC (blue diamonds) using the X-ray luminosity and WL mass of 196 clusters with available WL mass in LC2 (black points). The green diamonds show the estimated mass in CoMaLit V and the red diamonds show the MCXC masses.	96
Figure 39. Comparison of the predicted masses for 1382 clusters of MCXC by GPs (blue circles) and MCXC (red circles) to the masses estimated in CoMaLit V.	97
Figure 40. Relative difference of masses predicted by MCXC (red circles) and GPs (blue circles) with respect to the mass estimates by CoMaLit V versus the X-ray luminosity.....	97
Figure 41. Comparison of the expected values and associated standard deviations for the X-ray temperatures and luminosities of 204 groups and clusters considered in Hilton et al. 2012, using the previous and current X-ray analysis methodology. .	107
Figure 42. The $L_X - T$ relation assuming model 1, without (top, model 1u) and with (down, model 1) sample selection effects taken into account. Each point indicates the temperature, T , in keV , and re-scaled bolometric X-ray luminosity, L_X , expected for each of the 204 systems in the Hilton et al. 2012 sample considered here, given the X-ray data. The re-scaling of L_X with the cluster redshift, z , is performed dividing L_X by $E(z)^{1+\gamma}$, where γ is the expected (i.e. mean) value for γ . The error bars associated with each point identify the 1σ uncertainty intervals. The solid regression line was determined by fixing the model hyper-parameters to their expected (i.e. mean) values. The amplitude of the intrinsic vertical scatter about the regression line is indicated by the dashed lines (1σ , inner blue; 2σ , outer red).	108

List of Tables

Table 1. The first 10 rows of a supplementary table available with the online edition of this article and at http://risa.stanford.edu/redmapper , providing the XCS ID, SDSS-DR8-redMaPPer redshift (Rykoff et al. 2014), z , most probable values for the temperature, T (in units of keV), and X-ray bolometric luminosity within R_{500} , L_X (in units of $10^{44} \text{ erg s}^{-1}$), as well as the associated 68% confidence intervals derived through our X-ray analysis pipeline, for all 353 sample systems.....	41
Table 2. The median and symmetric credible intervals, centred on the median, that contain 68.3% and 95.4% of the probability for the marginal posterior distributions of the interesting hyper-parameters associated with each model. After each median value, are represented the numbers that need to be added and subtracted from it, in order to obtain the limits associated with the respective credible interval.	66
Table 3. The mean, standard deviation and skewness for the marginal posterior distributions of the interesting hyper-parameters associated with each model.	67
Table 4. Estimates for WBIC, its sample standard deviation, σ_{WBIC} , WAIC, and its sample standard deviation, σ_{WAIC} , with respect to each model.....	68
Table 5. The first 10 rows of a supplementary table available with the online edition of this article, providing the XCS ID, most probable values for the temperature, $Mo_{T,model}$ (in units of keV), and X-ray bolometric luminosity within R_{500} , $Mo_{L_X,model}$ (in units of $10^{44} \text{ erg s}^{-1}$), as well as the associated expected values, $\mu_{T,model}$ and $\mu_{L_X,model}$, and standard deviations, $\sigma_{T,model}$ and $\sigma_{L_X,model}$, obtained through LIRA, assuming a priori the $L_X - T$ relation to be described by model 4, for all 353 sample systems.	70

Table 6. The exponent of the $L_X - T$ relation estimated by Bharadwaj et al. 2015 and Zou et al. 2016 (using $Y X$ regression, private communication), compared with the expected values for the exponent we obtain assuming different models and samples.	78
Table 7. The first 10 entries of the catalogue of the comparison of GPs masses to CoMaLit V and Planck masses, <i>HFI_PCCS_SZ-MMF3_R2.08.fits</i> . The forecast by CoMaLit V is available here. Col. 1: index. Col. 2: name of the cluster. Col. 3: GPs mass and associated error, Col. 4: CoMaLit V mass and associated error, Col. 5: Planck mass and associated error. Masses are within R_{500} and in units of $10^{14}M_{\odot}$	91
Table 8. The first 10 entries of the catalogue of the comparison of GPs masses to CoMaLit V masses for a subsample of SDSS DR8 redMaPPer clusters, <i>redmapper_dr8_public_v6.3_catalog.fits</i> . The forecast by CoMaLit V is available here. Col. 1: index. Col. 2: name of the cluster. Col. 3: GPs mass and associated error, Col. 4: CoMaLit V mass and associated error. Masses are within R_{500} and in units of $10^{14}M_{\odot}$	95
Table 9. The first 10 entries of the catalogue of the comparison of GPs masses to CoMaLit V and MCXC masses, <i>J/A+A/534/A109/mcxc.dat</i> . The forecast by CoMaLit V is available here. Col. 1: index. Col. 2: name of the cluster. Col. 3: GPs mass and associated error, Col. 4: CoMaLit V mass and associated error, Col. 5: MCXC mass. Masses are within R_{500} and in units of $10^{14}M_{\odot}$	95
Table 10. The mean and standard deviation for the marginal posterior distributions of the interesting hyper-parameters associated with each model, given 204 groups and clusters in the Hilton et al. 2012 sample. The expected values and associated standard deviations for the observed X-ray temperatures and luminosities were obtained with the current X-ray analysis methodology.	105

Chapter 1

بیرون از تو نیست هر چه در عالم هست
در خود بطلب هر آنچه خواهی که تویی

مولانا رومی

This universe is not outside of you
Look inside yourself
everything that you want
you are already that

Maulana Rumi

Introduction

1.1 The beginning

Galaxy clusters, as extremely dense environments in the universe, are one of the important objects to study the growth of structure in the Universe through constraining cosmological parameters. Their properties can lead us to a better understanding of the physical processes that have been responsible for the formation and evolution of the large-scale structures in the Universe as well as its overall dynamics. For all those reasons, interest in galaxy clusters has been increased steadily in recent years.

In this introduction, we will present the background and framework that is necessary for a better understanding of the work presented in this thesis. We will give some background on the most important properties of galaxy clusters. Then, since an important part of this work has been done on the $L_X - T$ relation, we will present the scaling relations of galaxy clusters. We also will introduce the surveys whose data have been used in this work.

1.2 Galaxy clusters

Galaxy clusters typically consist of tens to hundreds of luminous galaxies, and thousands of fainter galaxies, which are located in a region with a range of virial radii $r_{\text{vir}} \approx 1 - 3$ Mpc for virial radii (Sarazin 1988). In terms of shape, they are typically elliptical. In

terms of mass, clusters cover a range of $10^{14}M_{\odot}$ to $10^{15}M_{\odot}$ with $M_{\odot}$ as the mass of the sun. Clusters are fair samples of the material in the Universe, so they can help us obtain a better understanding of the growth of structure and the history of the Universe. The spectrum of initial fluctuations, the growth of structures over time and the dynamics of the collapse of halos can be studied through galaxy clusters, which can lead us to constrain the cosmological parameters and that is what makes them an excellent probe of the growth of structure in the Universe (Euclid Red Book 2011).

1.2.1 Properties

There are different ways that galaxy clusters can be observed through. While the galaxy populations can be imaged in optical/infrared (IR) and the bremsstrahlung radiation from the hot cluster gas can be observed in X-ray, the cluster can also be imaged through the Sunyaev-Zeldovich (SZ) effect (Planck Collaboration 2014). See Figure 1.

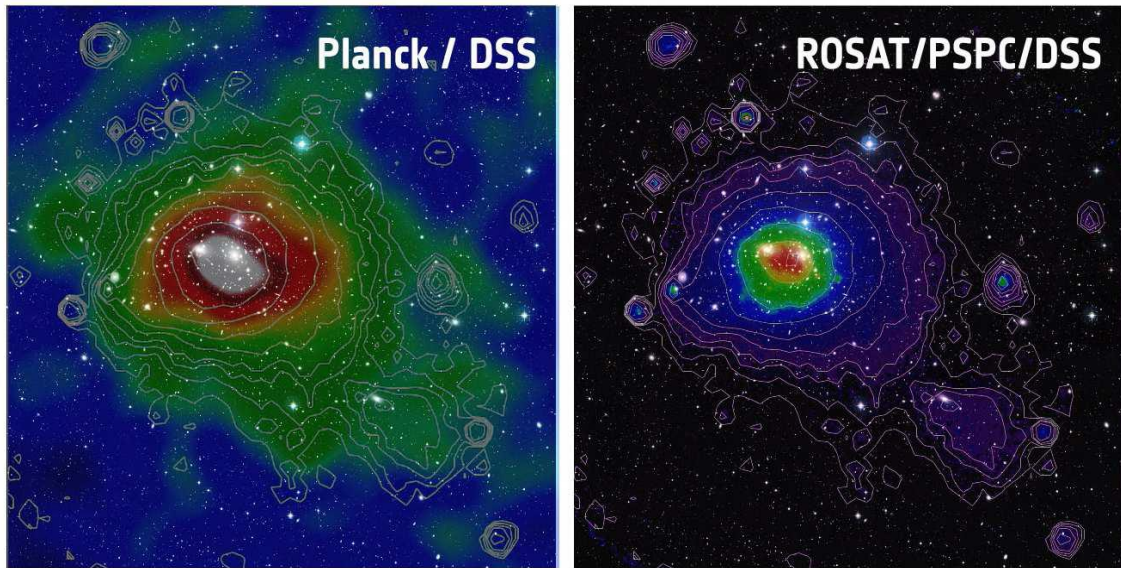


Figure 1. The image of Coma cluster observed through the SZ effect (left) obtained by Planck and X-rays (right) obtained by ROSAT. Both images are overlaid on the image of the cluster taken in visible light by DSS. Taken from Giodini et al. 2013.

Optical observables

Looking at clusters, galaxies are the component that are observed first, but they have a very small contribution to the mass of clusters. The total number of galaxies in a cluster is usually denoted as “richness”. In terms of the population of galaxies in a cluster, the fainter galaxies are more common than the bright and luminous galaxies. “Early-type” galaxies, which are elliptical (E), and lenticular (S0) galaxies, are found in rich clusters

primarily, while the lack of spiral (Sp) galaxies can be seen in clusters particularly near their centers. Since the elliptical galaxies in clusters consist of old, low mass $\lesssim 1 M_{\odot}$, red stars with very low rates of star formation, they are often expressed as “red and dead”. The “Brightest Cluster Galaxies” (BCGs) are these elliptical galaxies that exist very close by the centers of relaxed clusters. There are also cD galaxies which are extremely luminous $L_{\text{opt}} \gtrsim 10L_*$, where L_* is a characteristic luminosity of the galaxies (Sarazin 1988). The cD galaxies are formed when many small galaxies are merged with the central brightest galaxy of the cluster. Some of the BCGs are cD galaxies with very extensive outer envelopes.

X-ray observables

Clusters of galaxies consist of dark matter (partially baryonic and partially non-baryonic) and a diffuse, hot gas known as the intra-cluster medium (ICM) in addition to galaxies. The temperature of ICM is of order $T \sim 10^7 - 10^8$ K (thermal energies of $kT \sim 1 - 10$ keV) which is as a result of compression by the gravitational forces. Due to the fact that this high temperature makes ICM to shine in bremsstrahlung emission, X-ray observations help us to detect the emission and confirm that the huge space which seems to be empty between galaxies in clusters is not empty indeed. Instead, clusters are filled with a diffuse, hot gas that consists of a mix of magnetic fields, thermal plasma, and relativistic particles. The thermal origin of the emission of ICM can be demonstrated by detailed study of cluster X-ray spectra. Two important quantities that characterize the X-ray emission of clusters are the total (bolometric) luminosity and the temperature. For typical cluster temperatures ($kT \geq 2$ keV), the thermal bremsstrahlung emission overcomes the emission lines while for groups of galaxies ($kT < 2$ keV) the emission lines are dominant (Voit 2005). Beside flux, spatial extent and spectral hardness, which are primary X-ray observables, the spectra of clusters, can be determined very precisely through the observations by modern X-ray satellites through which the temperature, density and metallicity profiles of the ICM and some other thermodynamical quantities can be measured (Planck Collaboration 2014). The luminosity is defined as an integral over the cluster volume:

$$L_X \propto \int dV n_g^2 T^{1/2} \propto f_g M \langle n_g T^{1/2} \rangle_p \quad (1)$$

The first proportionality of the Equation (1) is valid at high temperature systems, where the luminosity is dominated by bremsstrahlung emission. The second proportionality of the Equation (1) expresses the relation between X-ray luminosity, L_X , and the virial mass of cluster, M , while f_g is the gas fraction and $\langle n_g T^{1/2} \rangle_p$ is the particle averaged quantity

with n_g as the gas density. If we assume that the temperature of the gas has reached to the virial temperature of the gravitational potential, T , during the collapse, then we have

$$T \propto \frac{M}{R} \propto M^{2/3}(1+z), \quad (2)$$

Wherein R referred to the virial radius of the distribution of the cluster mass. It can be expressed as the radius beyond which matter is infalling but has not been dynamically integrated into the cluster system. In terms of the second proportionality in Equation (2), it comes from $R \sim (M/\rho)^{1/3} \sim M^{1/3}(1+z)^{-1}$ (Bartlett 1997). X-ray surveys are very effective at finding cluster candidates because of this fact that surface-brightness profiles of clusters are centrally concentrated. As stated in Voit 2005: “Because X-ray emission depends on density squared, clusters of galaxies strongly stand out against regions of lesser density, minimizing the complications of projection effects”.

Sub-mm observables

Here we introduce one of the most effective ways to observe clusters of galaxies, which is again led by the hot intracluster gas. As Cosmic Microwave Background (CMB) photons pass through the hot ICM electrons of a galaxy cluster, a phenomenon called Sunyaev-Zel’dovich (SZ) effect occurs, which is the inverse of Compton scatter. What happens in this effect is that the CMB photons energy increases through scattering while they are passing through clusters and it makes a shift in the CMB spectrum, which can be observed through SZ surveys of galaxy clusters. This is an excellent mass proxy as both observations and numerical simulations confirm it (Planck Collaboration 2014). This effect has been detected around rich galaxy clusters due to the high temperature of the hot ionized gas (Planck Collaboration 2014 & Maughan 2014). The difference of the induced change in sky brightness, i_ν , towards the cluster to the mean of CMB intensity can be used to quantify the effect:

$$\delta i_\nu = y j_\nu(x), \quad (3)$$

where the Compton y -parameter specifies the amplitude in terms of an integral of the product of the gas density and temperature along the line-of-sight

$$y = \int dl \frac{kT_e}{m_e c^2} n_e \sigma_T, \quad (4)$$

and the spectral dependence is given by

$$j_\nu(x) = 2 \frac{(kT_0)^3}{(h_p c)^2} \frac{x^4 e^x}{(e^x - 1)^2} \left[\frac{x}{\tanh(x/2)} - 4 \right], \quad (5)$$

where T_e , n_e and m_e referred to the electron temperature, density and mass, respectively. In Equation (4), $\sigma_T = 6.65 \times 10^{-25} \text{cm}^2$ is the Thompson cross-section. In Equation (5), $x \equiv h\nu/kT$ referred to the dimensionless frequency of observation in terms of the CMB temperature $T = 2.728 \text{ K}$ (Bartlett 1997). The fact that the thermal SZ effect, unlike optical and X-ray surface brightness, is independent of distance makes it a very strong probe for cosmological applications. However, how this effect is a mass proxy of clusters? An SZ cluster survey, which observes clusters through the SZ effect, could reveal them out to arbitrarily high redshifts. The distortion parameter which is measured through the SZ surveys can be integrated and expressed as

$$Y = \int y dA \propto \int n_e T dV, \quad (6)$$

In which the first integral is over the projected surface of a cluster while the second one is over the volume of a cluster. The integrated Comptonisation parameter, Y , describes the overall thermal energy of the electrons, which is related to the total gas mass. Assuming the gas mass proportional to the mass of cluster, we can say that the parameter Y is a good proxy of cluster's mass, while the relation between Y and mass is calibrated.

1.2.2 Scaling relations

The abundance of galaxy clusters as a function of mass, known as the mass function (MF), is a well-known powerful cosmological probe to constrain cosmological parameters, which needs a precise knowledge of the total mass of galaxy clusters as a crucial ingredient (Rozo 2010). There are different ways that can lead us to the masses of clusters. For instance, weak lensing and strong lensing methods in the framework of virial equilibrium theorem, X-ray gas/SZ measurements under hydrostatic equilibrium assumption and applying the velocity dispersion of the galaxies can help us to have an estimation of the mass of clusters. However, these kind of telescope-based measurements are time consuming. For this reason, studying scaling relations that are able to relate mass to the basic properties of clusters is of interest. Cluster properties like optical richness, temperature, luminosity and SZ flux that are more easily derived from optical, infrared, X-ray, sub-millimeter and radio observations are related to cluster mass (Giodini et al. 2013). Further, the scientific community working on cosmological simulations are also interested in the characterization of scaling relations between fundamental properties of galaxy clusters because scaling relations carry information on

non-gravitational processes that one needs to include in order to reproduce the observed scaling relations. In this direction, modeling the forms of the X-ray scaling relations, in particular the X-ray luminosity-temperature ($L_X - T$) relation, attracts interests because they not only improve our understanding of the physical processes that form the ICM and heat it over cluster lifetimes (Maughan 2014) but also can be efficiently applied to estimate the mass of clusters when we do not have access to detailed data that are needed to estimate the mass.

1.2.2.1 Self-similar scaling relations

It is always useful to start by a simple model and then discuss about possible observational constraints on the scaling relations. This simple model, known as self-similar model, considers what to expect if the gravitational heating is the only source of energy. The results of this so-called self-similar model are simple scaling relations which relate the masses of galaxy clusters to the observable quantities, inferred from optical, infrared, submillimeter and X-ray observations, in a power law form as shown in Kaiser 1986. The word “self-similar” is applied to an object that each part of it can be considered a reduced-scale image of the entire (Sarazin 1998). For this reason, in a hierarchical scenario, as the building blocks for the larger structures are provided through the formed small structures, the small structures are also expected to be scaled down versions of the big ones (Giodini et al. 2013)

There are three main X-ray observable properties of the ICM that are correlated to the cluster mass: The luminosity (L), temperature (T) and mass (M). By assuming the self-similar model and making some simple assumptions, one can infer various simple scaling laws for the X-ray properties of clusters. The fact of being the structures self-similar in time brings us to this conclusion that two halos with the same birth time necessarily have the same density. Consequently

$$\frac{M_{\Delta z}}{R_{\Delta z}^3} = cte, \quad (7)$$

where $R_{\Delta z}$ refers to the radius in which the density contrast is Δ_z . The parameter Δ_z is defined with respect to the critical density at the redshift of cluster, $\rho_{\text{crit},0} = 3H^2/8\pi G$. It is called the critical density because if the average density of the universe is greater than it, the expansion can in future reverse to a contraction but if it is less, the universe continue to expand forever. The redshift dependence of Δ_z comes from this fact that the critical density and the background are evolved by time, $\Delta_z = \Delta(z=0) \frac{\Delta_{\text{vir}}(z)}{\Delta_{\text{vir}}(z=0)}$. This way,

it is possible to relate the clusters with different sizes that have formed at different epochs to each other.

$M_{\Delta z}$, which refers to the mass inside a sphere of radius $R_{\Delta z}$ is defined as:

$$M_{\Delta z} = \frac{4\pi}{3} \Delta z \rho_{crit,0} E_z^2 R_{\Delta z}^3, \quad (8)$$

in which $E_z = H_z/H_0 = [(\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda)]^{1/2}$ expresses the redshift evolution of the Hubble parameter at redshift z in a universe with Ω_m , as the matter density, and Ω_Λ , as the dark energy density parameters. When the gravitational force is balanced by a pressure gradient force, a cluster is said to be in hydrostatic equilibrium. For a cluster which is in hydrostatic equilibrium, the gas temperature could be an efficient representative of the depth of the potential well and, as a result, of the cluster's virial mass.

$$T_{gas} \propto \frac{GM}{R} \propto R_{vir}^2, \quad (9)$$

where R_{vir} is the virial radius. By substituting Equation (7) into Equation (9) we have

$$M_{\Delta z} \propto T_{gas}^{3/2}, \quad (10)$$

which expresses the relation between the cluster mass and temperature ($M - T$).

In order to achieve a scaling relation between X-ray luminosity and temperature ($L_X - T$), first we should have an assumption on the emission mechanism. We know that a massive system emits mainly by thermal bremsstrahlung when the ICM temperature is $10^7 - 10^8 \text{ K}$ due to a gravitational collapse. Hence, the total emissivity ϵ , which is defined as luminosity per unit volume, can be related to the temperature as follows

$$\epsilon \cong 3.0 \times 10^{-27} T_{gas}^{1/2} \rho_{gas}^2 \text{ erg cm}^{-3} \text{ s}^{-1}. \quad (11)$$

The electrons have been assumed to have the same temperature of the ions because we are implicitly assuming thermal equilibrium. Then, from Equations (10) and (11) the X-ray luminosity can be related to the total mass:

$$L_X \approx \epsilon R_{\Delta z}^3 \approx T_{gas}^{1/2} \rho_{gas}^2 R_{\Delta z}^3 \approx T_{gas}^{1/2} f_{gas}^2 M_{tot} \approx f_{gas}^2 T_{gas}^2, \quad (12)$$

where f_{gas} is the gas fraction defined as M_{gas}/M_{tot} . The last scaling relation is obtained using the second proportionality in Equation (9). Considering the gas fraction as a constant in the self-similar scenario (e.g. Ponman et al. 1999), we have

$$L_X \propto T_{gas}^2. \quad (13)$$

The principal X-ray self-similar scaling relations in galaxy clusters are those expressed in Equations (10) and (13). These relations hold for halo masses where any process other than gravitational processes can be ignored (Giodini et al. 2013).

1.2.2.2 Evolution of scaling relations

The redshift-dependency of the X-ray scaling relations is one of the consequences of the evolution of the background matter density caused by the cosmological expansion. It appears in the normalization factor through the redshift dependent Hubble parameter, E_z (or $F_z = E_z \times (\Delta_z/\Delta_{z=0})^{1/2}$). We do not assume any redshift evolution in the slope β because it is not important in the self-similar scenario. Therefore, the scaling relation between X and Y can be written as:

$$Y(X, z) = \alpha X^\beta F_z^\gamma \quad (14)$$

Scenarios that are more complicated can also be considered in which the slope β also depends on redshift but then additional physics is required. In order to use clusters for cosmology one should have a good understanding of the evolution of the scaling relations, which can also be applied for a better comprehension of the redshift evolution of the mass function. So far, we have only obtained well-calibrated mass-observable scaling relations at low redshifts. Hence, it is obvious how important the study of the high redshift objects is. The X-ray data of high redshift clusters obtained through long exposure observations makes the measurement of thermodynamical properties of clusters possible as well as the determination of the scaling relations at high redshifts.

1.2.3 Importance for Extragalactic Astronomy and Cosmology

Clusters of galaxies have long been suggested as cosmological probes. They are sensitive to the cosmological parameters in three ways. First, their distribution in redshift is sensitive to cosmology. Models of modified gravity are limited strongly by the observed redshift distribution of galaxy clusters (Lombriser et al. 2012). Secondly, they trace the underlying dark matter power spectrum through their spatial distribution. Density fluctuations can be converted to power spectrum. Hence, the shape of the power spectrum depends on the composition of dark matter dominating the universe, which is led by the

spatial distribution of galaxy clusters. Thirdly, individual clusters can be considered, under certain conditions, as fair samples of the matter content in the Universe. The nature of the mechanisms responsible for the formation of large-scale structure in the Universe, which is still the central problem in modern cosmology, can be revealed by the characteristics of galaxy clusters. The cluster abundance and its evolution are useful tools to study structure formation, which can tell us about the theoretical models. According to the scenario of “formation by gravitational instability”, when the density contrast, δ , is large enough so that the surrounding matter may collapse and separate from the general expansion, galaxies and clusters are formed. Hence, an immediate conclusion is that the abundance of the objects that are collapsed depends on the amplitude of the density perturbations. The density perturbations follow a probability distribution, $P(\delta)$ (e.g., a Gaussian: $P(\delta) = 1/\sqrt{2\pi\sigma^2} e^{-\delta^2/2\sigma^2}$). The variance of this distribution, $\sigma(R)$, expresses the amplitude of the perturbations on a scale of R . It is related to the power spectrum, $P(k)$, which is a fundamental quantity. When the scale is increased, the amplitude, $\sigma(R)$, and as a result, the density needed to form a large object decreases. It implies that the probability of forming a large object for instance a galaxy cluster lies on the tail of the statistical distribution. As a consequence, any small change in $P(k)$ affects the present abundance of clusters. In addition, the density parameter, Ω , controls the rate of cluster evolution. Hence, the importance of studying the cluster abundance and its evolution, which can constrain both, $P(k)$ and Ω (Bartlett 1997) becomes clear.

There is a variety of methods to constrain cosmological parameters using galaxy clusters. For example:

1. The relation between the power spectrum and the abundance of collapsed objects of a given mass leads us to the mass function of galaxy clusters (MF) at redshift z , $n(M, z)$, which is defined as the number density of virialized halos found at that redshift with mass in the range $[M, M + dM]$. Figure 2 shows how the number density of clusters as a function of mass is sensitive to the underlying cosmological parameters. Equation (15) states the expression for the Press–Schechter mass function.

$$\frac{d(n(M, z))}{dM} = \frac{2}{V_M} \frac{\partial(p > \delta_c(M, z))}{\partial M} = \sqrt{\frac{2}{\pi}} \frac{\bar{p}}{M^2} \frac{\delta_c}{\sigma_M(z)} \left| \frac{d(\log \sigma_M(z))}{d \log M} \right| \exp\left(-\frac{\delta_c^2}{2\sigma_M(z)^2}\right), \quad (15)$$

in where $p > \delta_c(M, z)$ is the probability for the linearly-evolved smoothed field δ_M to exceed at redshift z the critical density contrast δ_c . Equation (15) highlights the importance of the mass function as a strong tool for cosmological models. The connection of the mass function to cosmology is through $\sigma_M(z)$, the mass variance, which

is dependent to the cosmological density parameters and to the power spectrum, through the linear perturbation growth factor and through δ_c which is the critical density contrast. For massive objects (i.e., rich galaxy clusters), the exponential tail dominates the MF shape. For this reason, the cosmological parameters have a very important role in the exponential shape of the mass function for galaxy clusters. It demonstrates the reason for which clusters are powerful probes of dark energy. In other words, one can put tight constraints on cosmological parameters through a very well observationally determined mass function of clusters.

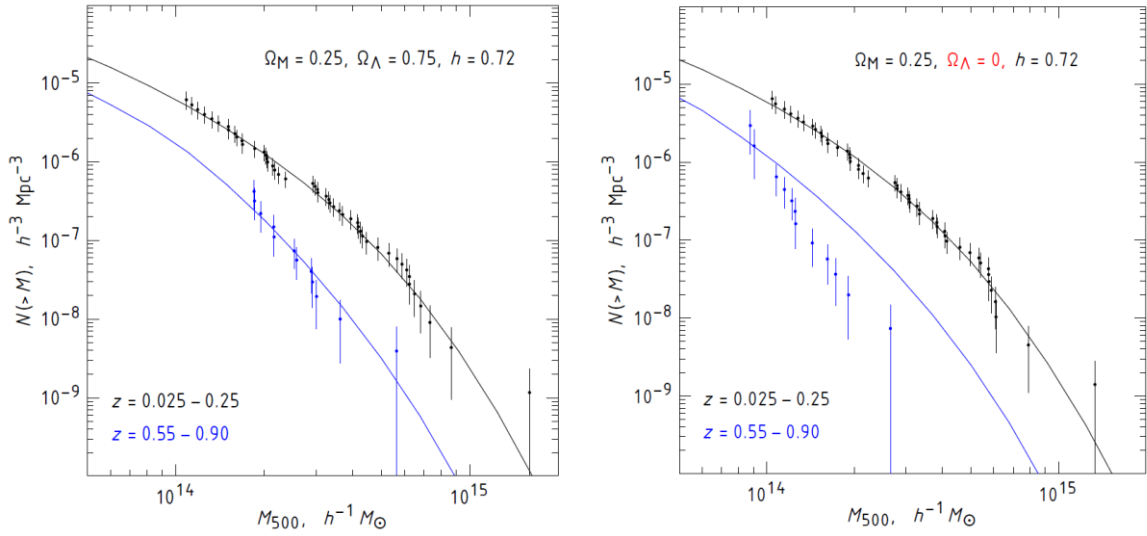


Figure 2. The dependency of the mass function on the cosmological models shown by Vikhlinin et al. 2009a. While in the left panel, the measured mass function and the predicted models shown for a cosmology with $\Lambda = 0.75$, the right panel shows a different model prediction and data measurements in the case of $\Lambda = 0$ at high redshifts. Taken from Giodini et al. 2013.

At the same time, the linear growth rate of density perturbations is constrained by the redshift evolution of mass function, which leads to dynamical constraints on the matter and dark energy parameters. Assuming O as the observable and $f(O, z)$ as a redshift-dependent selection function in a survey, the redshift distribution of clusters can be given by

$$\frac{d^2 N(z)}{dz d\Omega} = \frac{r^2(z)}{H(z)} \int_0^\infty f(O, z) dO \int_0^\infty p(O|M, z) \frac{dn(z)}{dM} dM, \quad (16)$$

where $r(z)$ is the radial coordinate appearing in the Friedmann–Robertson–Walker metric: $r(z) = \int_0^z dz E^{-1}(z)$, $dn(z)/dM$ is the space density of dark halos in comoving coordinates, and $p(O|M, z)$ which is the probability that a halo of mass M at redshift z is observed as a cluster with observable property O , is the mass-observable relation. It is

clear that this probe is useful as long as the relation between observable properties of clusters like richness, X-ray luminosity or temperature, SZ effect parameter, or weak lensing shear, and cluster mass are determined (e.g., Borgani S., 2008). The term multiplying the integral in Equation (16), and the growth of structure, which is dependent to the evolution of density perturbations, make the cluster counts a powerful probe of dark energy.

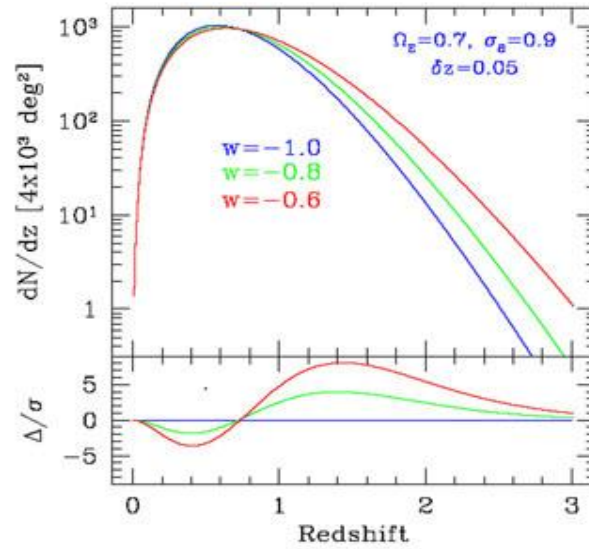


Figure 3. The upper panel shows the estimated cluster counts for a survey that could detect halos more massive than $2 \times 10^{14} M_{\odot}$, for three cosmological models with fixed $\Omega_M = 0.3$ and $\sigma_8 = 0.9$. The difference between models relative to the statistical errors has been shown in the lower panel. Taken from Mohr 2005.

Figure 3 shows how sensitive the cluster counts are to the dark energy equation-of-state parameter for the data obtained through the South Pole Telescope and the Dark Energy Survey, two surveys aimed at unveiling the nature of the dark energy. The sensitivity of the cluster counts to the growth rate of perturbations is obvious at redshifts $z > 0.6$. Since the systematic concerns arise from the uncertainties in the selection function and in the mass-observable relation, the stronger cluster observables correlate with mass and the better selection function is determined, the stronger cosmological constraints we have.

2. The shape and amplitude of the power spectrum for the dark matter distribution can be determined through applying the correlation function and power spectrum of the distribution of clusters as the clustering properties. Moreover, the density parameters can also be constrained by clustering parameters because they affect clustering

parameters through the linear growth rate perturbations (e.g., Borgani & Guzzo 2001 and Moscardini et al. 2001).

3. If we have the mean luminosity density of the Universe and we assume that light is traced by mass with the same efficiency inside and outside of clusters, the matter density parameter, Ω_m , can be constrained by the mass-to-light ratio in the optical band (e.g., Bahcall et al. 2000, Girardi et al. 2000, Carçberg et al. 1996).

4. The matter density parameter, Ω_M , can also be constrained through the baryon fraction in nearby clusters with this assumption that galaxy clusters fairly contain baryons and once the cosmic baryon density parameter is known, (e.g., Fabian 1991, White et al. 1993). Furthermore, the baryonic fraction of distant clusters can put constraints on the dark energy content and equation of state if we assume that there is no evolution for the baryon fraction inside clusters (e.g., Allen et al. 2002, Ettori et al. 2003).

1.3. Surveys of galaxy clusters

As mentioned before, we can make benefit of the possibility of observing galaxy clusters at different wavelengths (Giodini et al. 2013). Below we describe briefly the surveys whose data are used in this thesis.

1.3.1 Sloan Digital Sky Survey

The first light was observed by The Sloan Digital Sky Survey (SDSS; York et al. 2000) in May 1998. A telescope with the length of 2.5 m and a 3° field of view (Gunn et al. 2006) is utilized by SDSS at Apache Point Observatory (APO) in Southern New Mexico. SDSS-I and SDSS-2 are the first and second phases of the survey, which were completed with two instruments: The first one was a drift-scan imaging camera (Gunn et al. 1998) having 30 CCDs imaging in five filters u, g, r, I and z (Fukugita et al. 1996) and the second one was a pair of double spectrographs which were fed through 640 optical fibers. About $11,600 \text{ deg}^2$ were observed by SDSS-I/II out of more than $14,500 \text{ deg}^2$, which are currently covered by the imaging data of SDSS. The photometrically (Padmanabhan et al. 2008, Tucker et al. 2006; see also Smith et al. 2002) and astrometrically (Pier et al. 2003) analyzed and calibrated data led to catalogues which include about 500 million detected objects. Among sample of galaxies (Strauss et al. 2002; Eisenstein et al. 2001), stars (Yanny et al. 2009), quasars (Richards et al. 2002b) and other objects that were selected for spectroscopy, about 1.8 million spectra has been obtained by the survey as of Summer 2009. The data are publicly available in a series of data releases EDR, and

DR1 to DR14. In this thesis, we use the redMaPPer algorithm selected clusters from SDSS DR8, which we explain later in section 4.1.2.

1.3.2 Planck Survey

Planck was launched on 14 May 2009. Although at first the mission was to complete two full-sky surveys by the spacecraft, Planck terminated its mission after 30 months, about twice the time supposed to be employed, and completed five whole surveys of the sky. Moreover, even more data was provided in 2013 because the survey of the sky was continued by Low Frequency Instrument (LFI), which was able to work at slightly higher temperatures than HFI and it improved the Planck achievements. Planck was switched off on 23 October 2013. The data provided by Planck has high quality which are still explored scientifically and will be analyzed in the future years. In this thesis, we use the second Planck Catalogue of Sunyaev-Zel'dovich Sources (PSZ2, Planck Collaboration 2016) which we explain about later in section 4.1.1.

1.3.3 ROSAT All Sky Survey

Before launching of the ROSAT satellite numerous all-sky X-ray catalogues existed in view of collimated counter surveys. One of the fundamental scientific goals of ROSAT was to obtain the first all-sky survey in X-rays with an imaging telescope prompting a noteworthy increment in sensitivity and source position accuracy (Trumper 1983, Aschenbach 1988). The satellite was launched on June 1, 1990 and saw first light on June 16, 1990 (Trumper et al. 1991). The ROSAT All-Sky Survey (RASS) was led in 1990/91, soon after the two-month switch-on and execution confirmation phase. The fundamental strategy of the Survey was to scan the sky in super circles whose planes had been orientated kind of perpendicular to the solar direction. The first processing of the RASS occurred in 1991–1993 bringing about around 50,000 sources. ROSAT has directed the primary All Sky Surveys in soft X-rays($0.1 - 2.4 \text{ keV}$; $100 - 5 \text{ \AA}$) and the extreme ultra-violet ($0.025 - 0.2 \text{ keV}$; $500 - 60 \text{ \AA}$) bands utilizing imaging telescopes. It was turned off on 12 February 1999. In this thesis, we utilize the MCXC, a Meta-Catalogue of X-ray detected Clusters of galaxies, which have been derived from ROSAT observations (Piffaretti et al. 2011) and we will explain more about it in section 4.1.3.

1.3.4 XMM Newton Cluster Survey

The *XMM Cluster Survey* (XCS¹; Romer et al. 2001) is a serendipitous search for galaxy clusters and groups in the *XMM-Newton Science Archive* (XSA²). XSA provides simple and flexible access to data from the XMM-Newton mission.

The X-ray Multi-Mirror Mission, *XMM-Newton*, is one of the ESA's space observatories, which was launched in 1999. It is the largest scientifically built satellite in Europe. Moreover, its powerful telescope mirrors and sensitive cameras make it unique comparing to previous X-ray satellites. It includes two large payload modules, which are connected by a long carbon fiber tube forming the telescope optical bench (Figure 4). Two sets of data known as the Observation Data File (ODFs), which contains the necessary data for the analysis, and the Current Calibration File (CCF), which contains Calibrated data of ODFs, are provided for the user to perform the scientific analysis. These data are offered as the best currently known calibration pertaining to the subject observation (Jansen et al. 2001).

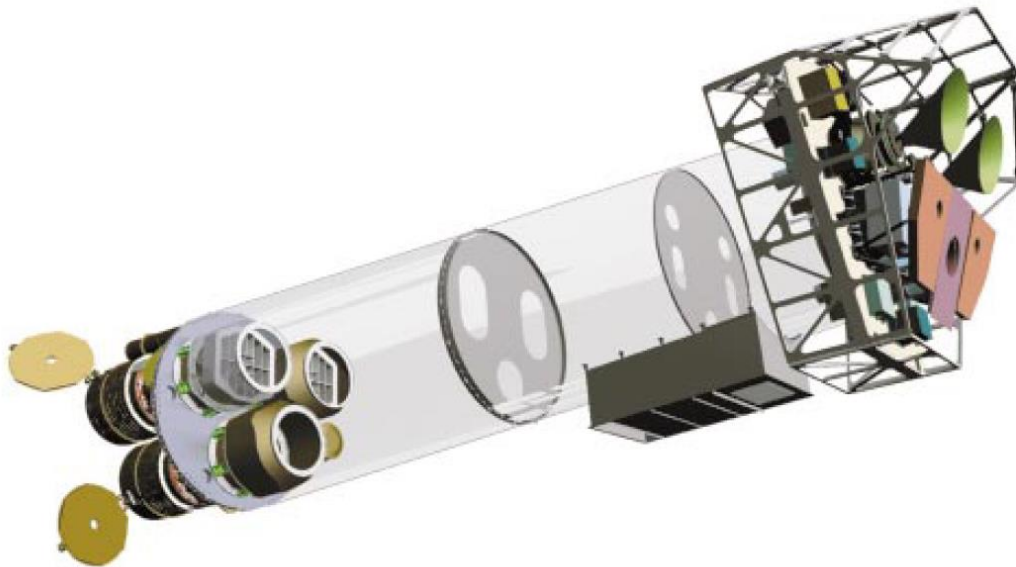


Figure 4. View of the XMM-Newton spacecraft mirror modules (to the left) and the backend of the instrument platform with the radiators (to the right). Taken from Jansen et al. 2001.

XCS is aimed for first constraining cosmological parameters using clusters and their properties and then obtaining scaling relations; understanding the astrophysical processes and the evolution of galaxies in clusters. The methodology of the X-ray analysis was first described by Lloyd-Davies et al. 2011, next updated in Rooney 2015

¹ <http://www.xcs-home.org>

² <http://nxsa.esac.esa.int>

and more recently in Bermeo-Hernandez 2017 and Manolopoulou et al. 2019. XCS-DR1 was the first data release of XCS (Mehrtens et al. 2012) which contained a total of 401 X-ray selected groups and clusters with estimated redshifts and temperatures covering a unique (non-overlapping) area of 276 deg^2 (Lloyd-Davies et al. 2011). Since then, we have been expanding the XCS, by analysing newer XSA data and using more recent optical cluster/group catalogues, namely those that are based on Sloan Digital Sky Survey (SDSS; York et al. 2000) data, to confirm our X-ray candidates. In the Appendix of the thesis, we use the sample considered in Hilton et al. 2012 to illustrate the impact on the estimation of X-ray luminosities and temperatures of the changes described in Rooney 2015, Bermeo-Hernandez 2017 and Manolopoulou et al. 2019, as well as on the characterization of the $L_X - T$ relation. At the time of writing, the XCS-DR2 image archives covers more than 1,050 square degrees in total (Rooney 2015).

Chapter 2

من از نهایت شب حرف می زنم

من از نهایت تاریکی

و از نهایت شب حرف می زنم

فروغ فرخزاد

I speak out of the deep of night

out of the deep of darkness

and out of the deep of night I speak

Forough Farrokhzad

Bayesian statistical data analysis

2.1 Introduction

Bayesian statistics or Bayesian probability theory have attracted researchers in many branches of science increasingly. Bayesian analysis is able to improve model parameter estimates. It provides a simple approach to all data analysis problems, allowing the experimenters to update their beliefs in the evidence of new data on the basis of the current state of knowledge (Gregori 2005).

Bayes' Theorem follows from the basic laws of probability. If for two propositions A, B we have

$$P(A|B)P(B) = P(A, B) = P(B|A)P(A) \quad (17)$$

So that

$$P(A|B) = P(B|A)P(A)/P(B). \quad (18)$$

This Theorem is simply a rule to invert the order of conditioning of propositions. By changing the name of the variables in Equation (18) and rewrite it we have:

$$p(\theta|data) = \frac{p(data|\theta) \times p(\theta)}{p(data)}, \quad (19)$$

where θ are the parameters, which we are interested to estimate their distribution. Each term in the expression above is given a name:

$$Posterior = \frac{Likelihood \times prior}{evidence} \quad (20)$$

Here, $p(\theta)$, the prior, is the probability of having the parameters of interest.

$p(data|\theta)$, the likelihood, is the probability of observing our result given the distribution for θ .

$p(data)$, the evidence, is the probability of data as determined by summing (or integrating) across all possible values of θ .

$p(\theta|data)$, the posterior, is the probability of having our parameters of interest after observing the data. This is the basic rule for parameter estimation in Bayesian statistics (Andreon S. & Weaver S. 2015).

2.2 How to begin the Bayesian analysis?

In order to begin the Bayesian analysis as with any statistical analysis, first a model must be specified. The process of determining a model consists of specifying the likelihood and priors. After that, we may characterize the posterior joint distribution through applying the Bayes theorem. What the posterior probability distribution tell us is what we know about a quantity after observing the data. Therefore, a narrower distribution means a better measurement of the quantity. In order to compute this distribution in practice, software packages like WinBUGS, OpenBUGS and JAGS would help. Markov Chain Monte Carlo, known as MCMC, which is a technique for Monte Carlo computation, and Metropolis-Hasting algorithm or the Gibbs sampler particularly, are used by these software packages to sample the posterior, which is introduced in the next section.

2.2.1 Using Markov Chain Monte Carlo

Metropolis et al. 1953 introduced MCMC algorithm for the first time as a method for fluids simulation. After about 35 years, the MCMC methods started to be utilized in (Bayesian) statistics analysis (Hitchcock 2003) in which complicated models were used by MCMC techniques in Hitchcock 2003, Tanner and Wong 1987, Gelfand et al. 1990 and Gelfand and Smith 1990. Thereafter, MCMC methods became the main computational tools in Bayesian inference because the inference of integrals with multi dimensions and parameters, which are required in the Bayesian analysis, were possible through them. Using MCMC the posterior distribution can be numerically sampled. What MCMC does simply is to draw samples from the posterior distribution. These samples are called chains. The more samples we have the more precise posterior we get. The mean, median, mode, standard deviation and credible regions of any parameter of interest can

also be calculated from the posterior distribution. The first thing required for the MCMC calculations to get started is to set initial values. It means that when MCMC begins, the sampler is not sampling from the desired function. MCMC needs time to draw samples from posterior distribution. Hence, usually a huge number of drawn samples are discarded from the beginning of the chain, known as “burn-in”. Although setting a burn-in increases our chance to obtain a chain sampling from the posterior, but like all numerical computations, there is always a possibility that things do not work correctly which means sometimes we have the problem of trapping the chain in a local maximum or we have a late/none convergence for the parameters of our interest. Hence, one should run several chains, which begin MCMC with different initial values. The higher number of chains one has, the higher is the probability of the convergence for the parameters of interest and the narrower is the posterior distribution at the end. Thus for models that are more complex, more chains are needed. Figure 5 indicates three example of chains, each of them drawn from 5000 samples that indicate the need of using longer chains in the MCMC. In the left-hand panel, the chain is wandering to explore the posterior while in the central panel it slowly moves from one value to another. The right panel is an example of two chains that are still not converging. In such cases, longer chains or even more efficient sampling schemes are needed (Andreon & Weaver 2015).

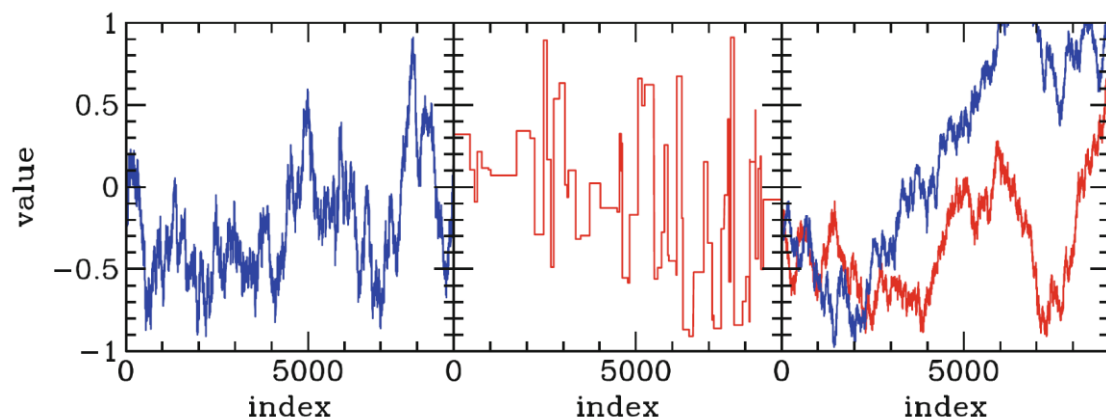


Figure 5. Three different trace plots that imply using longer chains in MCMC. While the chain in the left panel shows a slow mixing, the central panel shows a chain, which stays in a fixed value for too long and the right panel shows two chains, which are still not converging given that the final values of each chain, is different. Taken from Andreon & Weaver 2015.

2.3 Linear regression

Linear regression can be simply defined as drawing a line, which nicely interpolates a distribution of points but it is not as simple as it seems. By linear regression we are

looking for the best values of our interested parameters given the data points with associate uncertainties. Although there are tools to help us find the best line, for instance the ordinary least square estimator, the results may not be valid if we apply it out of its range of validity. There are some aspects that influence regression results and make us think of possible appropriate statistical tools that could tackle them. For instance:

- Observational errors that overshadow the independent and dependent parameters.
- The prior distribution on the independent parameters that may not be uniform.
- The intrinsic uncertainty that lies in the linear regression between the independent and dependent parameters.
- The independent and dependent parameters may not be observable parameters so that other parameters, which are proxies of them, are measured.
- The sample that we observe always is affected by selection effects, which make it not to be a real representative of the population we aim to study.

All of these effects have been considered broadly (Kelly 2007, Akritas & Bershadsky 1996, Hogg et al. 2010, Isobe et al. 1990, Feigelson & Babu 2012, and references therein). The intrinsic scatter or the observational errors have been assumed Gaussian in most cases. In 2007, Kelly described a Bayesian method that was able to take the intrinsic scatter, observational errors and multi-dimensional independent parameter into account as well as the selection effects in the independent parameter. He noticed that in order to get unbiased parameters for a regression, one must model the independent parameter's prior distribution, which was proposed to be a Gaussian mixture. A Gaussian mixture has enough flexibility when the true values of the independent parameter are estimated. Kelly's algorithm was extended recently by Mantz 2015 to a case which included a multi-dimensional dependent parameter. He also discussed about modeling the prior distribution of the independent parameter by a Dirichlet process instead of a mixture of Gaussian functions. Hogg et al. 2010 and Robotham et al. 2015 have also proposed other approaches for the prior distribution on the independent parameter.

We developed an existing Bayesian inference R package, LIRA (Linear Regression in Astronomy) (Sereno 2016) which mostly takes Kelly's method into account, in order to apply the linear regression model in our Bayesian statistical analysis and to infer the $L_X - T$ scaling relation for a sample of groups and clusters of galaxies which we will describe in detail in section 3.3. We modified LIRA in a way that the effect of any arbitrary

survey selection function (in our case XCS selection function, which is a bivariate sigmoid function) can be taken into account in the analysis.

2.4 Gaussian Processes

In this section, we summarize the Gaussian processes algorithm and we mainly follow the book by Rasmussen and Williams 2006. We use Gaussian Processes as a non-parametric Bayesian method to predict the mass of clusters, which will be described in detail in Chapter 4.

A Gaussian process is the generalization of a Gaussian distribution. To make a simple statement, while the distribution of a random variable is described by a Gaussian distribution, the Gaussian Process expresses a distribution over *functions*. If we assume that f is a function which is formed by a Gaussian Process, we can say that the value of f at a point x , has a Gaussian distribution with mean $\mu(x)$ and covariance

$$\text{cov}(f(x), f(\hat{x})) = k(x, \hat{x})$$

From the definition above, what is clear is that the function value at x is dependent to the function value at some other point $\hat{x}$ (especially when x and $\hat{x}$ are close to each other) because the covariance function connects them to each other.

The above definition implies that the function value at x is not independent of the function value at some other point $\hat{x}$ (especially when x and $\hat{x}$ are close to each other) because they are related by the covariance function. The mean function, $\mu(x)$, the covariance function, $K(x, \hat{x})$, and the variance, $\text{Var}(x)$, of a Gaussian process are defined as

$$\mu(x) = E[f(x)], \tag{21}$$

$$K(x, \hat{x}) = E[(f(x) - \mu(x))(f(\hat{x}) - \mu(\hat{x}))] \tag{22}$$

$$\text{Var}(x) = k(x, x) \tag{23}$$

and the way we can write the Gaussian Process is

$$f(x) \sim N(\mu(x), k(x, \hat{x})). \tag{24}$$

There are many possible covariance functions that can be applied in a Gaussian Process. The mostly used covariance functions are the ones which are defined based on the distance between the input points, although it is not necessarily required. There are some free parameters in the covariance functions that we use, called hyperparameters, which control the smoothness and stationarity of the predicted function. Hence, when we are talking about doing Gaussian Processes, in fact we are trying to find suitable hyperparameters for the covariance function.

Throughout this work in Chapter 4, we use the squared exponential covariance function

$$k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{1}{2}(x_i - x_j)^T W^{-1}(x_i - x_j)\right), \quad (25)$$

where the hyperparameters σ_f and W , the diagonal matrix of characteristic length-scale, l , control the overall variance and the characteristic length-scale of the predicted function. One thing that make this function useful is that it is infinitely differentiable, which is good to reconstruct the derivative of a function. What hyperparameters do is to characterize the “bumpiness” of the function. Figure 6 simply shows the effects of varying the hyperparameters on Gaussian Processes predictions. Panel (a), (b) and (c) have been generated from Gaussian Processes with the squared exponential covariance function (Equation (25)) with $l = 1$, $l = 0.3$ and $l = 3$ respectively. If we want to have a simple understanding of the characteristic length scale, l , and the signal variance, σ_f , from Figure 6 it is clear that the former controls the distance one should travel in x -direction to see a significant change in $f(x)$, while the latter controls the change in y -direction.

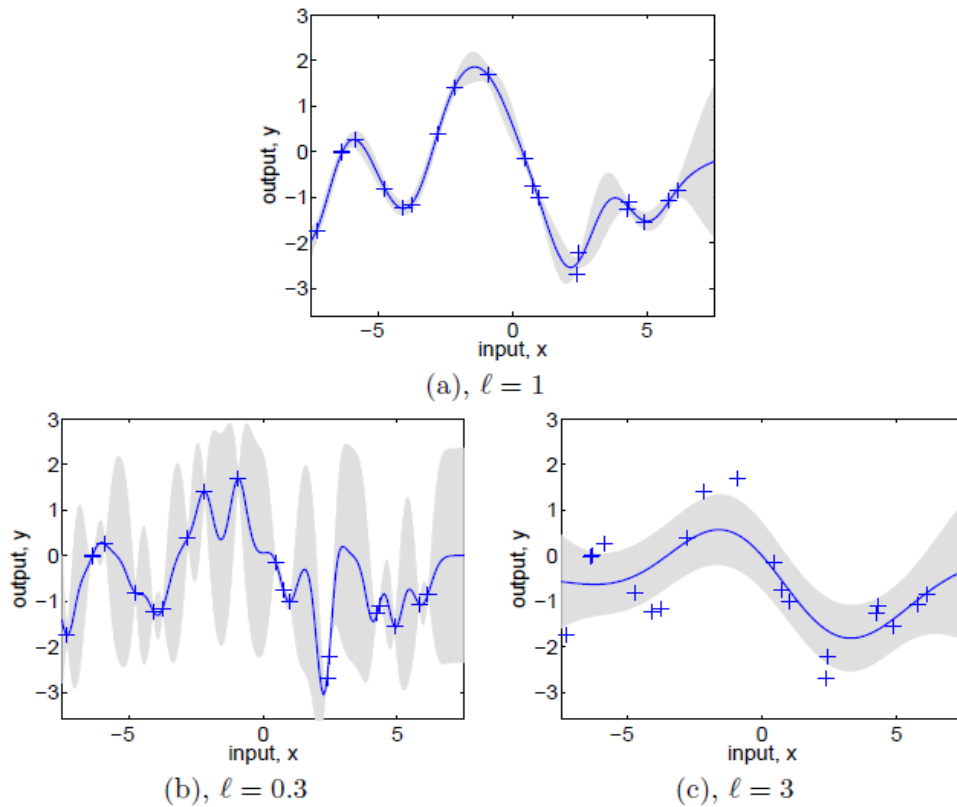


Figure 6. The + symbols in each panel show data generated from Gaussian Processes with $(l, \sigma_f) = (1, 1)$ (top), $(l, \sigma_f) = (0.3, 1)$ (lower left) and $(l, \sigma_f) = (3, 1)$ (lower right). A 95 % confidence region has been obtained for the underlying function f , which is shown in grey for all the cases. Taken from Rasmussen & Williams 2006.

Since we include the input noise of data in our analysis based on what is proposed by

Dallaire et al. 2011, we use the squared exponential covariance function including the uncertainties on our data that is computed with

$$k((\mu_i, \Sigma_i), (\mu_j, \Sigma_j)) = \frac{\sigma_f^2 \exp\left(-\frac{1}{2}(\mu_i - \mu_j)^T (W + \Sigma_i + \Sigma_j)^{-1} (\mu_i - \mu_j)\right)}{|I + W^{-1}(\Sigma_i + \Sigma_j)(1 - \delta_{ij})|^{1/2}}, \quad (26)$$

which is again a squared exponential form. It is assumed that x_i and x_j are normally distributed, $x_i \sim N(\mu_i, \Sigma_i)$ and $x_j \sim N(\mu_j, \Sigma_j)$. This is the first time that the effect of including input uncertainties in Gaussian Processes is considered when it is applied to observational cluster scaling's data. Usually one assumes a zero mean function, $\mu(x) \equiv 0$, for doing Gaussian Processes.

The set of n observational data can be given by $D = \{(x_i, y_i) | i = 1, \dots, n\}$ as the training data, where x is the input data with dimension of d and y is the response (target) data. We assume that the input data are collected in a $d \times n$ matrix denoted by X . The actual observations are assumed to be scattered around the underlying function, i.e. $y_i = f(x_i) + \epsilon_i$, where noise ϵ_i is assumed to have a Gaussian distribution with variance σ_i^2 . We should add this variance to the covariance matrix:

$$y \sim N(\mu, K(X, X) + C), \quad (27)$$

where $K(X, X) + C$ is the covariance matrix of the data including the noise effect induced by noisy observations. For the data which are uncorrelated we have simply $C = \text{diag}(\sigma_i^2)$. We are interested in predicting the function (true) values of the dependent variable, f^* , corresponding to test inputs, collected in a $d \times n^*$ matrix, X^* . Assuming $f_i^* = f(x_i^*)$ and considering f^* as the Gaussian vector of function values at X^* we have

$$f^* \sim N(\mu^*, K(X^*, X^*)), \quad (28)$$

where μ^* is the *a priori* assumed mean for f^* . Equation (28) means that the Gaussian Processes are evaluated at specific points x^* , where $f(x^*)$ has been randomly sampled from a normal distribution with mean of μ^* .

We can combine the two Gaussian Processes for f^* and y in the joint distribution

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu^* \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right). \quad (29)$$

The values of y are given by the observations while we can sample the values for f^* from the joint distribution of the observed target values and the function values at test points as shown in Equation (29) under the prior

$$p(f^* | X, y, X^*) \quad (30)$$

The principal predictive equations for Gaussian Processes regression can be derived from the conditional distribution in Equation (30) as

$$p(f_*|X, y, X_*) \sim N(\bar{f}^*, \text{cov}(f^*)),$$

where

$$\bar{f}^* \triangleq E[f^*|X, y, X^*] = \mu^* + (K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}(y - \mu)), \quad (31)$$

$$\text{cov}(f^*) = K(X^*, X^*) - K(X^*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(X, X^*). \quad (32)$$

$K(X, X^*)$ describes an $n \times n$ covariance matrix which has been evaluated at all possible pairs of training and test points and this applies for $K(X, X)$, $K(X^*, X^*)$ and $(K(X^*, X))$ as well. In order to get the posterior distribution over functions, the joint prior distribution is constrained so that only the functions which agree with the observed data remain. This is shown very well in Figure 7. In panel (a), three functions are generated randomly from the GPs prior, and then those functions that agree with the observational data are kept in panel (b). They are functions that are conditioned on the five noise free observations.

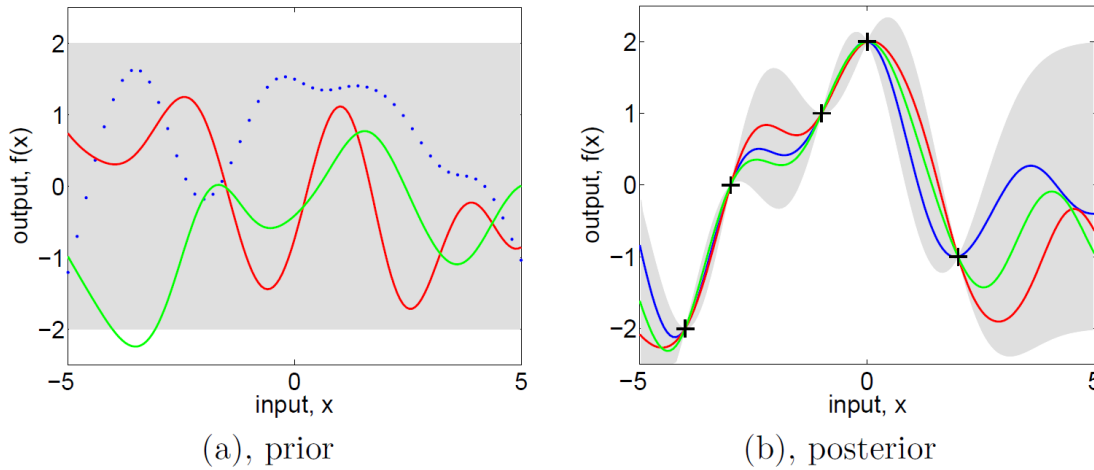


Figure 7. In panel (a), three random functions drawn from GPs prior have been shown. In panel (b), three random functions from GPs prior that are conditioned on the five indicated observational data have been shown. The grey area corresponds to the 95% confidence region in both plots. Taken from Rasmussen & Williams 2006.

It is worth to highlight that although it is common to use a constant value which is often a zero function as the mean function, $\mu(x)$, we use instead a linear mean function with a priori unknown parameters in our analysis. In fact, we are modeling the residuals by the Gaussian Processes while we are assuming that the data obeys a linear model. In order to find the suitable hyperparameters for the covariance function and the mean function with the objective of estimating the function values at arbitrary points, one can either optimize the hyperparameters by the Maximum A Posteriori (MAP) procedure or

marginalize over them by using MCMC techniques to produce samples of the hyperposterior. Optimization has the positive aspect of being faster but has the negative aspect of underestimating the uncertainty in the reconstruction of the function. In our analysis, which will be described in Chapter 4, we estimate the function values by marginalizing over the hyperparameters of the covariance function and the mean function.

Chapter 3

پس از دشواری آسانیت ناپار
ولیکن آدمی را صبر باید

سعدی

Have patience

All things are difficult before they become easy

Saadi

Joint modelling of the $L_X - T$ scaling relation

3.1 Introduction

Clusters and groups of galaxies are the largest gravitationally bound structures in the universe. They are very complex systems, within which galaxies and the intracluster/intergroup medium (ICM/ICG) interact and evolve with time, in a background of gravitationally dominant dark matter. Clusters and groups can be detected and characterised in multiple regions of the electromagnetic spectrum: the ICM/ICG emits X-rays, mostly through the bremsstrahlung effect, while the Sunyaev-Zel'dovich effect produced by the ICM/ICG leaves an imprint in the microwave background radiation; their galaxy population is most clearly seen in the optical and infrared wavebands. Several of the quantities that can be estimated by observing clusters and groups in those regions of the electromagnetic spectrum, like the temperature and X-ray luminosity, the integrated comptonization parameter, and richness-related measures, seem to obey different scaling behaviours depending on whether one considers the cluster or the group regime (e.g. Sun 2012; Giodini et al. 2013). This suggests these structures have formed and/or evolved with time in distinct ways. Thus, the study of the different behaviours

exhibited by cluster and group scaling relations can shed light on the disparate physical processes behind the assembly of those structures. It can also help define the distinction between the group and cluster regimes. This could prove to be particularly useful when it comes to using cluster properties to impose constraints on cosmological parameters (e.g. Vikhlinin et al. 2009; Mantz et al. 2010b; Rozo et al. 2010; Allen et al. 2011; Böhringer et al. 2014; Mantz et al. 2014; de Haan et al. 2016; Planck Collaboration 2016). The reliability of these depends on the correct modelling of cluster scaling relations, used to link cluster mass to more directly observable cluster properties, as well as on the assembly of cluster samples devoid of objects, like groups, that may follow different scaling relations.

For example, it is expected that clusters and groups exhibit a different scaling relation between X-ray luminosity (L_X) and temperature (T) in the context of the self-similar model (e.g. Sun 2012; Zou et al. 2016). Although the contrary has been recently claimed for observed clusters and groups (Bharadwaj et al. 2015; Lovisari et al. 2015; Zou et al. 2016), in none of these analysis the $L_X - T$ relation in the cluster and group regimes was characterised simultaneously. In fact, every study of the $L_X - T$ relation either made no distinction between clusters and groups, or ad hoc assumptions were used to define cluster-only or group-only samples. In particular, it has been common practice to assume that clusters have temperatures exceeding a value somewhere between 2 and 3 keV, contrary to groups. In either case, biases may have been introduced in the derivation of the behaviour of the $L_X - T$ relation for the objects dominating the sample considered. This may be one of the reasons for some inconsistency among results obtained in the many studies already devoted to this subject, although inadequate or absent corrections for sample selection effects have surely also played a role. Namely, contradictory results have been obtained when using the behaviour of the $L_X - T$ relation to test whether the self-similar model Kaiser 1986 provides a reasonable description of how clusters and groups have formed and evolved. Most noticeably, while the self-similar model predicts $L_X \propto T^2$, observations suggest instead a steeper relation, $L_X \propto T^{2-3}$ (e.g. see review by Giodini et al. 2013). A consensus has been even more difficult to reach with respect to the evolution of the cluster and group $L_X - T$ relations. While the results of some analysis are consistent with the prediction of self-similarity (e.g. Mantz et al. 2010a; Maughan et al. 2012), whereby the normalisation of the $L_X - T$ relation is positively correlated with redshift, z , being proportional to $E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$, other studies have found evidence for either zero or even a negative correlation (e.g. Reichert et al. 2011; Hilton et al. 2012; Clerc et al. 2014). Given that the self-similar model assumes that the properties of the ICM/ICG are determined by the action of gravitational collapse

alone, any departure from self-similarity would indicate that additional sources of energy have been heating the ICM/ICG. The most probable are the Active Galactic Nuclei (AGN) known to reside in the centres of most clusters, and in a significant fraction of groups (e.g. Gitti et al. 2012).

In this section, we start by reviewing the characteristics of the XCS and SDSS data considered in the analysis. Next, we describe the method used to characterise the $L_X - T$ relation in the cluster and group regimes simultaneously. We then present the results obtained, discuss them, and present our conclusions.

We assume throughout the universe to be geometrically flat, with fractional densities of matter and dark energy (as a cosmological constant) to be $\Omega_m = 0.27$ and $\Omega_\Lambda = 0.73$, respectively, with the Hubble constant being $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

3.2 Data

The first data release of XCS (XCS-DR1; Mehrrens et al. 2012) contained a total of 401 X-ray selected clusters and groups with temperature and redshift estimates. Since then, we have been expanding the XCS, by analysing newer XSA data and using more recent optical cluster/group catalogues, namely those that are based on Sloan Digital Sky Survey (SDSS; York et al. 2000) and Dark Energy Survey (DES; Flaugher et al. 2005) data, to confirm our X-ray candidates.

Table 1. The first 10 rows of a supplementary table available with the online edition of this article and at <http://risa.stanford.edu/redmapper>, providing the XCS ID, SDSS-DR8-redMaPPer redshift (Rykoff et al. 2014), z , most probable values for the temperature, T (in units of keV), and X-ray bolometric luminosity within R_{500} , L_X (in units of $10^{44} \text{ erg s}^{-1}$), as well as the associated 68% confidence intervals derived through our X-ray analysis pipeline, for all 353 sample systems.

XCS ID	z	T	L_X
XMMXCS J140101.9 + 025238.3	0.26	$6.63^{+0.03}_{-0.02}$	$55.39^{+0.09}_{-0.09}$
XMMXCS J142601.0 + 374937.0	0.17	$8.65^{+0.10}_{-0.10}$	$31.88^{+0.20}_{-0.20}$
XMMXCS J172009.8 + 263725.8	0.17	$6.15^{+0.04}_{-0.04}$	$21.50^{+0.08}_{-0.08}$
XMMXCS J224321.4 - 093550.2	0.44	$7.22^{+0.09}_{-0.07}$	$33.11^{+0.08}_{-0.09}$
XMMXCS J212939.7 + 000516.9	0.25	$5.29^{+0.05}_{-0.05}$	$22.74^{+0.13}_{-0.12}$
XMMXCS J075124.2 + 173057.7	0.19	$3.28^{+0.05}_{-0.05}$	$2.86^{+0.02}_{-0.02}$
XMMXCS J172227.0 + 320758.0	0.23	$7.21^{+0.09}_{-0.09}$	$27.58^{+0.17}_{-0.18}$
XMMXCS J104044.2 + 395711.1	0.14	$3.74^{+0.04}_{-0.04}$	$7.39^{+0.04}_{-0.04}$
XMMXCS J111253.4 + 132640.2	0.17	$4.89^{+0.06}_{-0.06}$	$6.12^{+0.04}_{-0.04}$
XMMXCS J024803.3 - 033143.4	0.19	$3.76^{+0.04}_{-0.04}$	$9.64^{+0.06}_{-0.06}$

In Manolopoulou et al. 2019 we present XCSDR2-SDSS, the subset of the XCS Data Release 2 within the SDSS DR13 footprint (Albareti et al. 2017), comprising more than one thousand optically-confirmed groups and clusters, about two hundred of which are new discoveries. However, here we concentrate on a sample of groups and clusters that includes only 353 of those objects, which were also independently detected by applying the redMaPPer algorithm to SDSS-DR8 (Rykoff et al. 2014).

We decided to consider only such objects in order to make the underlying selection function of our sample as simple as possible, just the result of the convolution of the XCS selection function with that for the SDSS-DR8-redMaPPer catalogue. Note that more objects than the 353 considered appear both in the XCS and in the SDSS-DR8-redMaPPer catalogue. The smaller size of our sample results mainly from retaining only the objects with a coefficient of variation for the temperature lower than 0.25, thus ensuring that our temperature estimates are reliable (for further details, see Manolopoulou et al. 2019). We also restricted our sample to groups and clusters with a redshift equal or greater than 0.1, in order to avoid having to deal with the more complicated nature of the XCS selection function at lower redshifts, and with measured temperatures in excess of 1 *keV*, given that the XCS selection function has not been characterised for lower temperatures (Sahlén et al. 2009).

The XCS identification (ID), SDSS-DR8-redMaPPer redshift (Rykoff et al. 2014), most probable values (i.e. modes) for the temperature and X-ray luminosity, as well as the associated 68% confidence intervals derived through our X-ray analysis pipeline, for all 353 sample systems, are listed in a supplementary table available with the online edition of this article. The first 10 rows are shown in Table 1.

The sample distributions of most probable values for the redshift, temperature and X-ray luminosity, given the data, are represented in Figures 8 and 9. The groups and clusters span the redshift range 0.10 – 0.59 (median 0.33), temperature range 1.06 – 11.53 *keV* (median 3.39 *keV*), and luminosity range 2.36×10^{42} – 8.90×10^{45} *erg s⁻¹* (median 1.19×10^{44} *erg s⁻¹*).

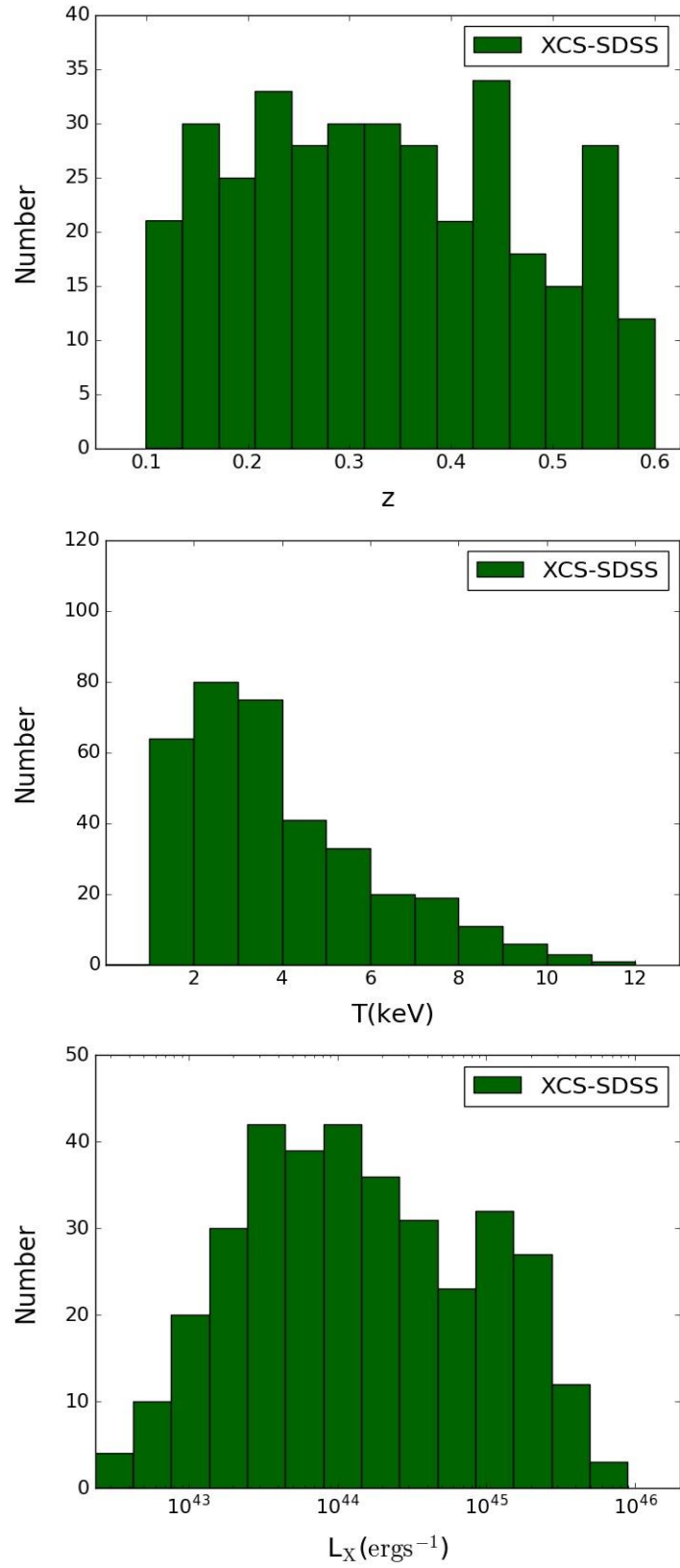


Figure 8. The distributions of most probable values for the redshift, z , temperature, T (in keV), and X-ray luminosity, L_X (in erg/s), given the data, for the 353 XCS-DR2-SDSS-DR8-redMaPPer groups and clusters used in this work.

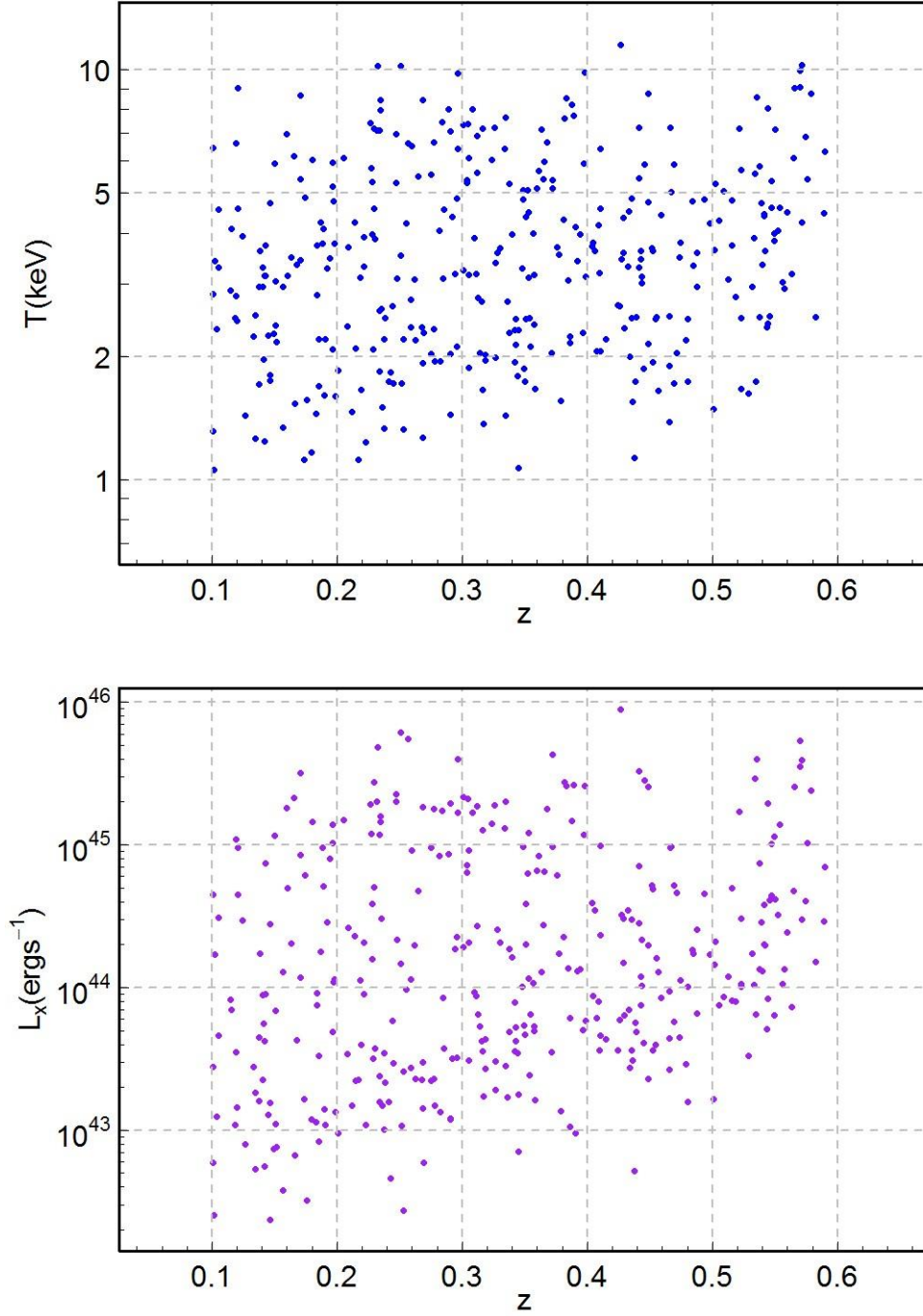


Figure 9. Most probable values for temperature, T (in keV , upper panel), and X-ray luminosity, L_X (in erg/s , lower panel), versus redshift, given the data, for the 353 XCS-DR2-SDSS-DR8-redMaPPer groups and clusters used in this work.

3.3 Analysis

3.3.1 Bayesian framework

Cluster and group scaling relations have been most often modelled as a single power-law with a time-evolving normalization, following the self-similar expectation (Kaiser

1986). Under such assumptions, the general form of the relation between two properties, e.g., the dependent (response), Y , and the independent (covariate), X , variables is

$$Y = \alpha X^\beta F_z^\gamma, \quad (33)$$

which in logarithmic scale, is expressed as a linear relation,

$$\log(Y) = \alpha + \beta \log(X) + \gamma \log(F_z) + \varepsilon, \quad (34)$$

where the factor depending on F_z allows for evolution with redshift, and ε represents the intrinsic scatter in $\log(Y)$ about the regression relationship. We will assume the later to be described by a Gaussian distribution with mean zero and a priori unknown dispersion σ . This may be allowed to evolve with time according to

$$\sigma(z) = \sigma_0 F_z^{\gamma_\sigma} \quad (35)$$

where σ_0 is the zero redshift dispersion, and γ_σ accounts for the time-evolution of the dispersion.

Our interest lies in modelling the scaling relation between X-ray luminosity, L_X , and temperature, T . Therefore, we will re-write Equation (34) as

$$\log[L_X/E(z)] = \alpha + \beta \log(T/5) + \gamma \log E(z) + \varepsilon \quad (36)$$

where the logarithms are base 10, L_X is the bolometric X-ray luminosity and T is the temperature, in units of erg s^{-1} and keV respectively. Both are estimated iteratively using only data that falls within R_{500} , the radius enclosing an average density 500 times larger than the critical density $3H^2/8\pi G$. Note that L_X and T are estimated using all data that falls within R_{500} , i.e. they are not core-excised quantities (for further details, see Manolopoulou et al. 2019). Since the luminosity has been scaled by $E^{-1}(z)$ in Equation (36), the self-similar model predicts $\gamma = 0$, as well as $\beta = 2$ (Kaiser 1986). We set F_z equal to $E(z)$, instead of $(1+z)$, to enable a more straightforward comparison with the expectation for selfsimilar evolution when the exponent for the $L_X - T$ relation deviates from the self-similar prediction (Maughan 2014). The pivot temperature is fixed at 5 keV , close to the mean of the sample (3.4 keV), to allow for comparison of our results with those of Hilton et al. 2012. It is assumed that z is not affected by measurement uncertainties, i.e. its actual value is perfectly known, given that typically $\sigma_z \sim 0.01(1+z)$ for the clusters and groups in our sample (Rozo & Rykoff 2014; Rykoff et al. 2014).

However, because the properties of clusters and groups may obey different scaling relations, we will also consider the possible existence of an abrupt transition or change-point in the scaling relation between T and L_X . This is achieved by introducing two extra

parameters, T_{break} and β_{break} in our generic model, where the later takes the role of β for $T < T_{break}$. All other generic model parameters discussed so far, i.e. α, γ, σ and γ_σ , are assumed to be universal across the range of L_X and T values that we will consider, in order to ensure continuity between the cluster and group scaling regimes.

We will be using a Bayesian hierarchical modelling framework to constrain T_i and $L_{X,i}$, for each i cluster or group in our sample, as well as the hyper-parameters associated with each particular model we will consider, conditioned on all the X-ray and redshift data, $D \equiv \{D_i\}$, we have gathered regarding those objects (for a similar approach see e.g. Mantz et al. 2010a; Maughan 2014). This will be achieved through the characterization of the marginal posterior probability distributions for all T_i and $L_{X,i}$, as well as for the hyper-parameters, in the context of each of those particular models. All these, numbered from 0 to 7, are nested inside the generic model (labelled as 7). Each is uniquely identified through its associated set of hyper-parameters, although models 0 and 1 have variants, denoted 0u and 1u, which we will later differentiate from the main model:

$$\theta_0 \equiv (\alpha, \beta, \sigma)$$

$$\theta_1 \equiv (\alpha, \beta, \sigma, \gamma)$$

$$\theta_2 \equiv (\alpha, \beta, \sigma, \gamma_\sigma)$$

$$\theta_3 \equiv (\alpha, \beta, \sigma, T_{break}, \beta_{break})$$

$$\theta_4 \equiv (\alpha, \beta, \sigma, \gamma, \gamma_\sigma)$$

$$\theta_5 \equiv (\alpha, \beta, \sigma, \gamma, T_{break}, \beta_{break})$$

$$\theta_6 \equiv (\alpha, \beta, \sigma, \gamma_\sigma, T_{break}, \beta_{break})$$

$$\theta_7 \equiv (\alpha, \beta, \sigma, \gamma, \gamma_\sigma, T_{break}, \beta_{break}).$$

In practice, it is computationally easier to work with the (generic) linear model represented by Equation (36), where $X_i \equiv \log(T_i/5)$ and $Y_i \equiv \log[L_{X,i}/E(z_i)]$. Therefore, we will first characterize the marginal posterior probability distributions for each X_i and Y_i , and then transform them into marginal posterior probability distributions for each T_i and $L_{X,i}$. Marginal distributions are obtained by integrating the joint posterior probability distribution, which can be evaluated, up to a normalization constant, through the expressions

$$P(\{X_i, Y_i\}, \theta | D) \propto P(D | \{X_i, Y_i\}, \theta) P(\{X_i, Y_i\}, \theta) \quad (37)$$

$$\propto P(\{D_i\} | \{X_i, Y_i\}, \theta) P(\{X_i, Y_i\} | \theta) P(\theta) \quad (38)$$

$$\propto \prod_{i=1}^n P(D_i|X_i, Y_i, \theta)P(Y_i|X_i, \theta)P(X_i|\theta)P(\theta) \quad (39)$$

where n is the number of objects in our sample. We assumed, as usual, that the data pertaining to each cluster or group, D_i , does not hold information about the properties of any other object. In the last expression, the first term is known as the likelihood, and in our case it will be independent of the values that any particular set of hyper-parameters, generically labelled θ , may take. The product of the second and third terms is the prior probability distribution for each set of hidden (sometimes also called latent) variables, X_i and Y_i , conditioned on the hyper-parameters. We assumed all such sets to be mutually independent. The second term is just the Gaussian distribution that describes the stochastic relation between X and Y , characterized through Equation (36), and thus is in fact independent of the object i under consideration. The last term in the expression is the joint prior probability distribution for the hyper-parameters (also known as hyper-prior).

The R package for Bayesian Linear Regression in Astronomy (LIRA), described in Sereno 2016, and available both at CRAN³ and GitHub⁴, was used to efficiently explore, using Gibbs sampling, the joint posterior probability distribution. The Monte Carlo Markov Chains (MCMC) thus assembled was then used to characterize the marginal posterior distributions for each set of the hidden variables T_i and $L_{X,i}$, as well as for the hyper-parameters, θ . We assessed chain convergence using the Gelman-Rubin diagnostic (Gelman & Rubin 1992), applied independently to each hidden variable and hyper-parameter of the model under consideration.

The likelihood, $P(D_i|X_i, Y_i)$, is affected by the presence of sample selection effects. In our case, these change the relative abundance of objects in the sample with respect to what is found in the population, as a function of both the covariate (X) and response (Y). In the former case, this comes about from our selection of only objects with a measured temperature in excess of 1 keV , while in the later case it results from the XCS selection function (Sahlén et al. 2009), which we model as a bivariate sigmoid function and favours the detection of objects with higher X-ray luminosity at any given redshift or temperature. Unfortunately, LIRA is unable to take into account the impact of both sample selection effects on the likelihood. We will describe in the next sub-section how this problem can be addressed. The raw likelihood (i.e. in the absence of sample selection effects) with respect to each set of the hidden variables T_i and $L_{X,i}$ is assumed to take the form of a

³ <https://cran.r-project.org/web/packages/lira/>

⁴ <https://github.com/msereno/lira/>

bivariate split Gaussian probability distribution, with the co-variance terms in all quadrants set to zero. Such a distribution is equivalent to the product of two independent univariate split Gaussian distributions (Villani et al. 2006), which is the type of distribution used to approximate the results of our X-ray analysis pipeline (Lloyd-Davies et al. 2011; Manolopoulou et al. 2019), based on XSPEC⁵. We then follow Sereno & Ettori 2015 and use expressions (27) and (28) provided by D’Agostini & Raso 2000 to estimate the expected values and terms of the covariance matrix associated with the bivariate Gaussian distribution we assume describes the likelihood with respect to each set of the hidden variables, X_i and Y_i , in the absence of sample selection effects.

The prior probability distribution for each hidden variable X_i , conditioned on the hyper-parameters, θ , i.e. $P(X_i|\theta)$, is assumed to be well described by a time-evolving mixture of Gaussian distributions (Sereno 2016). This allows the analysis to take into account: (1) the effect of the underlying temperature function, which describes the relative abundance of groups and clusters in the overall population as a function of their temperature; (2) the SDSSDR8-redMaPPer selection function, which make the incorporation of an object in our sample less probable as its temperature decreases and redshift increases. We explored the cases where the number of Gaussians in the mixture was fixed a priori to either one or two. However, in the case of a mixture of two Gaussians we were unable to ascertain convergence of the MCMC for all the hyper-parameters associated with the more complex models. Therefore, we chose instead to restrict the mixture to just one Gaussian. Nevertheless, the marginal posterior distributions thus obtained are very similar to those that result from assuming a mixture of two Gaussians. Every particular set of hyper-parameters is henceforth expanded to include: μ , the Gaussian (mixture) mean; σ_μ , the Gaussian (mixture) dispersion; and γ_μ , the parameter that allows for time-evolution of the Gaussian (mixture) mean. Note that LIRA can allow for a possible evolution of σ_μ with time, but we did not consider a priori the possibility of it being different from zero. We are not directly interested in characterizing the marginal probability distributions for the three Gaussian (mixture) hyper-parameters, and therefore from here on we will omit any explicit reference to them, except when discussing the results.

All hyper-parameters are assumed to be independent a priori, and thus for every particular model the hyper-prior is just the product of the prior probability distributions for each hyper-parameter associated with the model. We set these to the default choices in LIRA (Sereno 2016), adapted to the characteristics of our models and sample. Namely,

⁵ <https://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/>

uniform prior distributions in the ranges [40,50] for α , [-0.7,0.3] for $\log(T_{break}/5)$, and [-2.0,1.0] for μ , were assumed, in order to bracket, within a conservative margin, the sample values for $\log[L_X/E(z)]$, in the first case, and $\log(T/5)$, in later two cases. We checked a posteriori that changing these limits did not have a significant impact on the results. Further, the prior distributions for β , β_{break} , γ , γ_σ and γ_μ , were set to Student's t distributions with one degree of freedom (also known as standard Cauchy distributions). This choice ensures that the results are invariant under a rotation of the coordinate system used to identify the location of the sample objects, i.e. it ensures that all angles the possible regression lines make with, for example, the horizontal coordinate axis are equally probable a priori (see e.g. von Toussaint 2011; Sharma 2017). Finally, a Γ prior distribution, with shape and rate parameters equal to 0.0001, was assumed for both $1/\sigma^2$ and $1/\sigma_\mu^2$. This choice makes the prior distribution with respect to the variances, σ^2 and σ_μ^2 , nearly scale-invariant.

3.3.2 Sample selection effects

The sample that is being considered is a sub-set of the groups and clusters of galaxies identified by the redMaPPer algorithm in SDSS DR8 data. Thus, we are in fact working with a censored sub-sample of redMaPPer objects. In this situation, the full likelihood, taking into account both detections and non-detections, is given by (Equation (42), Kelly 2007):

$$p(x_{obs}, y_{obs}, D | \theta, |\psi_{obs}) \propto \prod_{i \in A_{det}} p(x_i, y_i | \theta, \psi_{obs}) \prod_{j \in A_{non_det}} p(x_j | \psi_{obs}) \times \int p(D_j = 0 | y_j, x_j) p(y_j | x_j, \theta, \psi_{obs}) dy_j. \quad (40)$$

In our case, the integral is over the intrinsic X-ray temperature and luminosity, because the XCS selection function is defined with respect to these. So, we have instead:

$$p(x_{obs}, y_{obs}, D | \theta, |\psi_{obs}) \propto \prod_{i \in A_{det}} p(x_i, y_i | \theta, \psi_{obs}) \prod_{j \in A_{non_det}} \iint p(D_j = 0 | y_j, x_j) p(y_j | x_j, \theta, \psi_{obs}) p(x_j | \theta, \psi_{obs}) dy_j dx_j. \quad (41)$$

Thus the full likelihood, taking into account the sample selection effects, is just the result of multiplying the likelihood defined in the previous sub-section, i.e. the first set of multiplications in the expression above, with the probability of non-detection of each object not in our sample but known to be in the redMaPPer SDSS DR8 catalogue overlapping with the XCS footprint, as a function of the regression parameters as well as those that parametrize the (assumed Gaussian) function that describes the underlying distribution of intrinsic X-ray temperatures.

Therefore, the joint posterior probability distribution, with respect to all parameters and taking into account the sample selection effects, can be obtained simply by multiplying the joint posterior probability distribution obtained from LIRA with the second product in the Equation (41). However, LIRA does not provide the joint posterior probability distribution directly, but MCMC samples from it. Therefore, one can either: (1) change the MCMC samples so that they become effectively samples of the joint posterior probability distribution that takes into account the sample selection effects; (2) or use the LIRA MCMC samples to approximate the associated joint posterior probability distribution, and then multiply it by the second product in the Equation (41).

The first approach can be implemented using (self-normalized) Importance Sampling (IS, e.g. Vehtari et al. 2017). If for each s^{th} draw of the MCMC, we call the full posterior probability distribution for parameters θ , the target distribution, $p(\theta_s)$, and the LIRA posterior probability distribution for those parameters, the proposed distribution, $g(\theta_s)$, then for each s^{th} draw,

$$p(\theta_s) = g(\theta_s)r(\theta_s), \quad (42)$$

where the $r(\theta_s)$ are called importance ratios and, in this case, are equal to the second product in Equation (41) given the values of θ associated with the s^{th} draw of the MCMC. Any function h of θ and $p(\theta)$, like the mean, median and credible intervals, can then be estimated using the LIRA MCMC draws from $g(\theta)$ by averaging over them, weighted by the importance ratios

$$\mu_\theta = \frac{1}{S} \sum_{s=1}^S h(\theta^s) = \frac{\frac{1}{S} \sum_{s=1}^S h(\theta^s) r(\theta^s)}{\frac{1}{S} \sum_{s=1}^S r(\theta^s)} \quad (43)$$

However, if the $p(\theta_s)$ and $g(\theta_s)$ have very different values, i.e. the $r(\theta_s)$ are large, even if only within a small region in parameter space, then, in general, the IS ratios will not

allow for an accurate estimation of a given function of θ and $p(\theta)$. In order to verify if this is the case or not, diagnostics have to be used (e.g. Vehtari et al. 2017).

For the second approach to be implemented, a parametric family of probability distribution functions needs to be chosen, within which we will identify that which best approximates, according to some criterium, the joint posterior probability distribution for the parameters θ that was sampled by the LIRA MCMC output. A standard approach is to find the Gaussian mixture which best approximates such output.

3.3.3 Model comparison

There are several approaches to the issue of model comparison, even within Bayesian statistics. The most prominent ones attempt either to identify the true underlying model or to find the most useful model. Useful in this context is usually considered to mean the ability of a given model to make predictions about future or yet undisclosed observations. In the first approach, the true underlying model is assumed a priori to be within the (finite) set of models considered in the analysis, contrary to the second approach. This has led, for example Vehtari & Ojanen 2012, to propose labelling those two contrasting views of model comparison as M-closed and M-open, respectively.

Model comparison usually proceeds through the evaluation of the probability of each model given the data, also known as the marginal likelihood or evidence, in the M-closed approach, or using cross-validation in the M-open approach (e.g. Piironen & Vehtari 2017). However, both evidence estimation and cross-validation are computationally expensive to perform, which has led to several proposed approximations [see e.g. section 2.5 of Sharma 2017 for a short review]. There is a growing consensus that the Widely applicable (or Watanabe) Bayesian Information Criterion (WBIC) and the Widely Applicable (or Watanabe-Akaike) Information Criterion (WAIC) are among the most accurate and theoretically justified methods to approximate the results of model comparison based on, respectively, evidence estimation and cross-validation (Gelman et al. 2014; Friel et al. 2017; Piironen & Vehtari 2017; Sharma 2017; Vehtari et al. 2017).

In our case, WBIC is given by (Watanabe 2013; Sharma 2017)

$$WBIC = E^{1/\ln n}[-\ln P(D|\{X_i, Y_i\}, \theta)] \quad (44)$$

or

$$WBIC = \frac{\int -\ln P(D|\{X_i, Y_i\}, \theta) [P(D|\{X_i, Y_i\}, \theta)]^{1/\ln n} P(\{X_i, Y_i\}, \theta) dX_i dY_i d\theta}{\int [P(D|\{X_i, Y_i\}, \theta)]^{1/\ln n} P(\{X_i, Y_i\}, \theta) dX_i dY_i d\theta} \quad (45)$$

where n is the number of independent data points, i.e. $n = 353$ for the sample under consideration. The WBIC is an approximation to the Bayes free energy, i.e. $-\ln P(D|M_\theta)$, where $P(D|M_\theta)$ is the evidence associated to a particular model M_θ , with hyperparameters θ . The difference between the expectation values for WBIC and the Bayes free energy is $O(\ln(\ln n))$ (Watanabe 2013). Comparing WBIC values associated with different particular models is thus equivalent to comparing their posterior probability distributions under the assumption of equality for their prior probability distributions, which under no reason to assume otherwise, is what we will be doing.

The expectation value for some parameter-dependent function can be readily estimated using the output of a MCMC used to characterize the joint posterior probability distribution of the parameter(s) the function depends upon. This is achieved by averaging over the values such function takes given the values of the parameter(s) at each point in the MCMC. In our case, this means that

$$WBIC \simeq -\frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n \ln P(D_i | X_{i,j}, Y_{i,j}, \theta_j) \quad (46)$$

where j identifies a point in a given MCMC of length m used to characterize the joint posterior probability distribution that is proportional to $[P(D|\{X_i, Y_i\}, \theta)]^{1/\ln n} P(\{X_i, Y_i\}, \theta)$. Given that the summations in the MCMC approximation to WBIC can be interchanged, its estimation can be thought as resulting from the sum of the n independent components $WBIC_i \equiv -\frac{1}{m} \sum_{j=1}^m \ln P(D_i | X_{i,j}, Y_{i,j}, \theta_j)$. Hence, the product of n by their variance equals the sample variance associated with the WBIC estimate (Vehtari et al. 2017).

On the other hand, WAIC is given by (Watanabe 2010; Piironen & Vehtari 2017)

$$WAIC = T + V/n \quad (47)$$

where, in our case,

$$T = \sum_{i=1}^n T_i \quad (48)$$

is the training utility, loss or error, with

$$T_i \equiv -\frac{1}{n} \ln E[P(D_i | X_i, Y_i, \theta)] \quad (49)$$

and

$$V = \sum_{i=1}^n V_i \quad (50)$$

is the functional variance, with

$$V_i \equiv [(\ln P(D_i | X_i, Y_i, \theta))^2] - \{E[\ln P(D_i | X_i, Y_i, \theta)]\}^2. \quad (51)$$

In these expressions, the expectations are calculated with respect to the joint posterior probability distribution of $\{X_i, Y_i\}$ and θ . The WAIC is asymptotically equal to the result of Bayesian leave-oneout cross-validation (LOO-CV), as well as to the generalization utility or loss, with the difference between their expectation values being $O(1/n^2)$ (Watanabe 2010, 2013). In our case, the estimation of WAIC can be also thought as the sum of the n independent components $WAIC_i \equiv T_i + V_i/n$. Therefore, the product of n by their variance equals the sample variance associated with the WAIC estimate (Vehtari et al. 2017).

The expectations that need to be calculated to estimate each component $WAIC_i$ can be obtained using MCMC output, in a similar way to what was done with regards to the estimation of WBIC. Therefore, in our case

$$T_i \equiv \ln \left[\frac{1}{m} \sum_{j=1}^m P(D_i | X_{i,j}, Y_{i,j}, \theta_j) \right] \quad (52)$$

and

$$\frac{1}{m} \sum_{j=1}^m [\ln P(D_i | X_{i,j}, Y_{i,j}, \theta_j)]^2 - \left[\frac{1}{m} \sum_{j=1}^m \ln P(D_i | X_{i,j}, Y_{i,j}, \theta_j) \right]^2. \quad (53)$$

where again j identifies a point in a given MCMC of length m , but now used to characterize the joint posterior probability distribution that is proportional to $P(D | \{X_i, Y_i\}, \theta)P(\{X_i, Y_i\}, \theta)$.

3.4 Results

3.4.1 Without taking into account sample selection effects

Let us start by considering the simplest model, with hyperparameters θ_0 , where only the normalisation, α , slope, β , and scatter, σ , of the $\log(L_X) - \log(T)$ relation are free to vary within the respective prior distributions. The time-evolution parameters γ and γ_σ are set to zero, and there is no change-point. In the upper panel of Figure 10, the regression line obtained assuming the expected (i.e. mean) values for the relevant model hyperparameters, can be compared with the distribution of data points. In order to allow for comparison with previous results, we also show (lower panel) the regression line obtained when the XCS selection function and the imposition of a lower threshold of 1 keV in observed temperature for inclusion in the sample are not considered in the modelling, and the prior distribution for the temperature is (implicitly) assumed to be

uniform. This version of model 0, we label as 0u, hence does not take into account any type of sample selection effect.

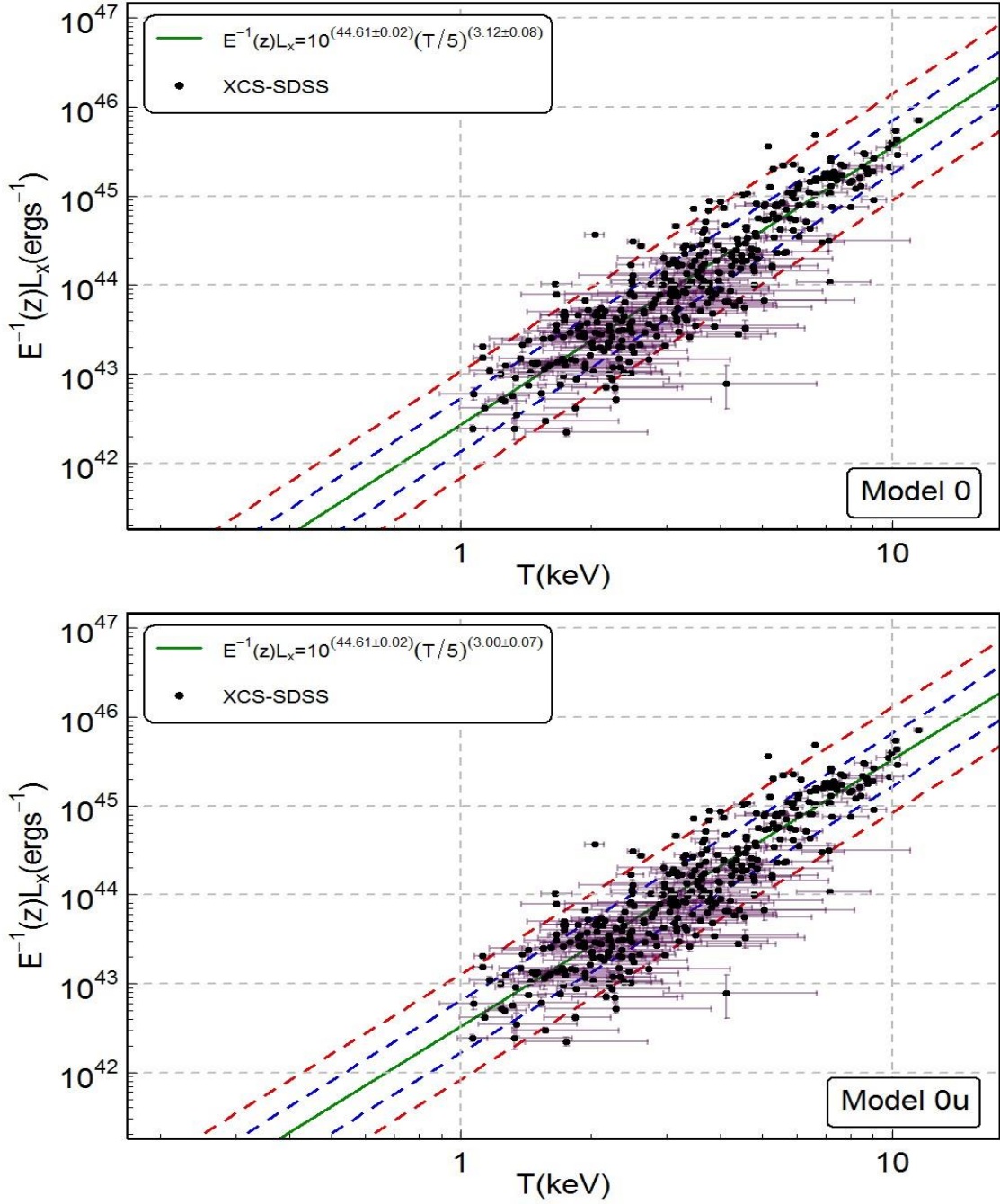


Figure 10. The $L_X - T$ relation assuming the simplest model, with (top, model 0) and without (bottom, model 0u) sample selection effects taken into account. Each point indicates the most probable values for the temperature, T , in keV , and re-scaled bolometric X-ray luminosity, L_X , with respect to each of the 353 systems in our sample, given the X-ray data. The self-similar re-scaling of L_X with the cluster redshift, z , is performed dividing L_X by $E(z)$. The error bars associated with each point identify the 1σ uncertainty intervals. The solid regression line was determined by fixing the model hyper-parameters to their expected (i.e. mean) values. The amplitude of the intrinsic vertical scatter about the regression line is indicated by the dashed lines (1σ , inner blue; 2σ , outer red).

We will now see what happens when model complexity is increased, by adding to the simplest model one of the following sets of hyper-parameters: γ , which allows the normalisation to evolve with time (model 1); γ_σ , which permits time-evolution of the scatter (model 2); T_{break} and β_{break} , respectively, the transition temperature and the value of β below T_{break} (model 3), associated with a possible change-point. The resulting regression lines can be seen in Figures 11 (upper panel), 12 and 13, respectively. In Figure 11 we also show (lower panel) the results obtained when no sample selection effects are considered, which we label as model version 1u. In Figures 14, 15 and 16 we compare the one- and two-dimensional marginalised posterior distributions for models 1, 2 and 3, with those for model 0 respectively.

The even more complex models 4, 5, 6 and 7, include all possible combinations of the hyper-parameters introduced in the preceding models. The respective regression lines are represented in Figures 17 to 20. The one- and two-dimensional marginalised posterior distributions for models 1 vs. 5, 2 vs. 6, and 4 vs. 7, are compared in Figures 21, 22 and 23, thus highlighting the impact of introducing a change-point in the $L_X - T$ relation. The posterior distributions for the hyper-parameters T_{break} and β_{break} rapidly decrease towards zero when approaching 1 keV and 4, respectively. This happens for all models, i.e. 3, 5, 6, and 7, which include these hyper-parameters, as can be seen in Figures 14 to 16 and in Figures 21 to 23. The lower limit of 1 keV on T_{break} is driven by the prior distribution assumed for this hyper-parameter, which imposed a prior probability of zero for T_{break} smaller than 1 keV. It would not make sense to assume otherwise, because all objects in our sample have observed temperatures greater than 1 keV. Only the addition to the sample of objects with temperatures below 1 keV would allow for a better understanding of the behaviour of the posterior distribution for T_{break} close to and below 1 keV. Nevertheless, given that its mode is close to 1.5 keV, adding such objects to the sample should not change significantly the location of the highest posterior probability region associated with T_{break} , and hence its expected value. On the other hand, the upper limit of 4 on β_{break} is mostly driven by the data. Given that the normalization of the $\log(L_X) - \log(T)$ relation is strongly constrained, values for β_{break} approaching 4 incur a rapidly growing penalty due to a regression line that lies increasingly far away from objects with low T and relatively high L_X . These have T values with much smaller measurement uncertainties than objects with low T and relatively low L_X , thus regression lines that lie equally far away from the later, i.e. characterised by β_{break} values approaching 0, are not as penalized, leading to a slower decrease in the posterior probability for β_{break} left of its mode.

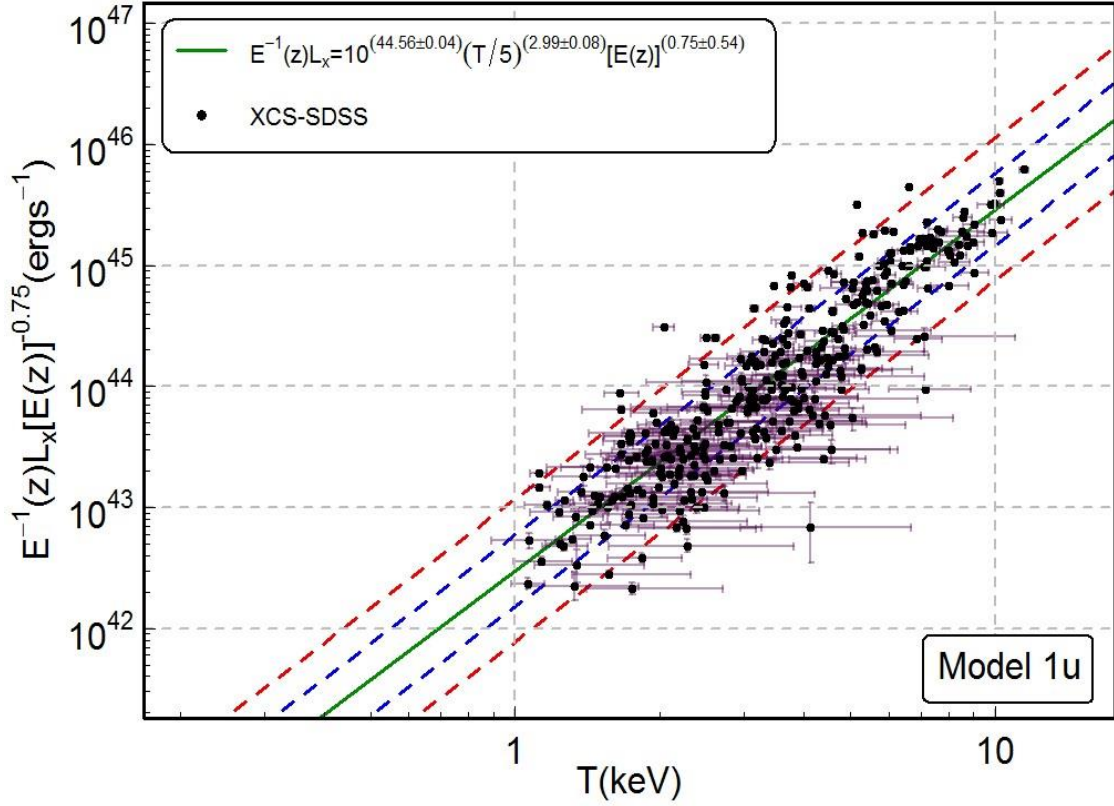
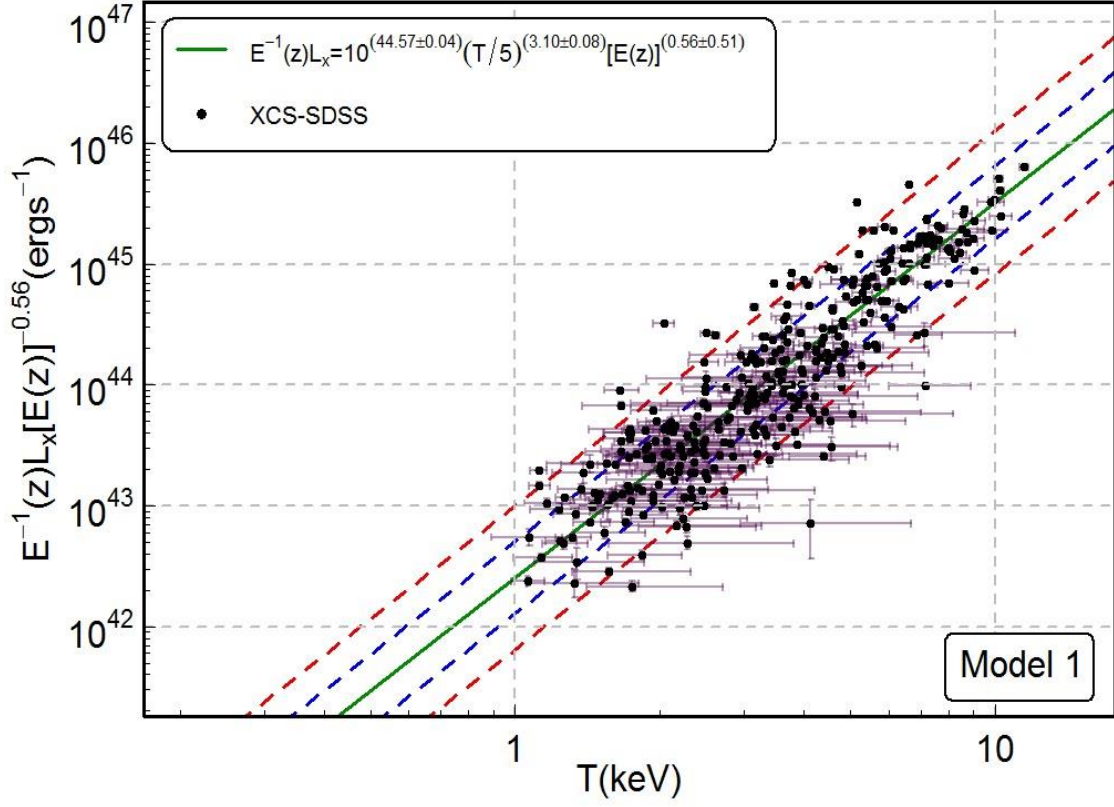


Figure 11. The $L_X - T$ relation assuming model 1, with (top, model 1) and without (bottom, model 1u) sample selection effects taken into account. In this case, the re-scaling of L_X with the cluster redshift, z , is performed dividing L_X by $E(z)$, where γ is the expected (i.e. mean) value for γ .

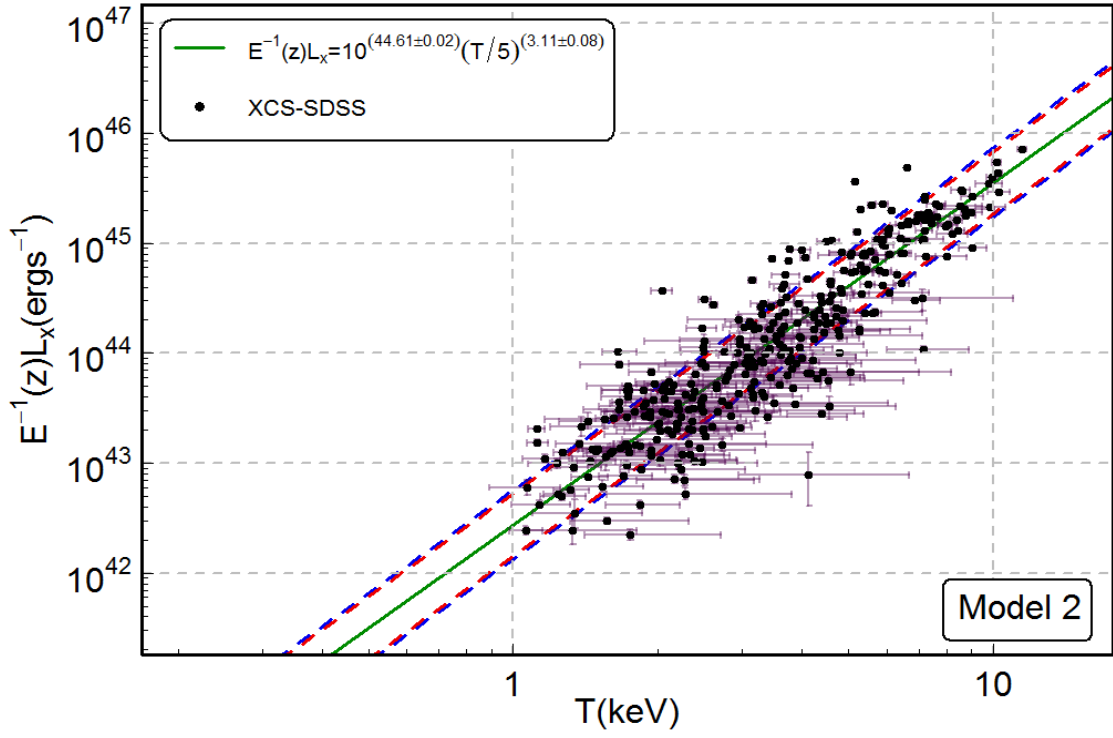


Figure 12. The $L_X - T$ relation assuming model 2. The prior distribution for the temperature is modelled through a time-evolving Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts.

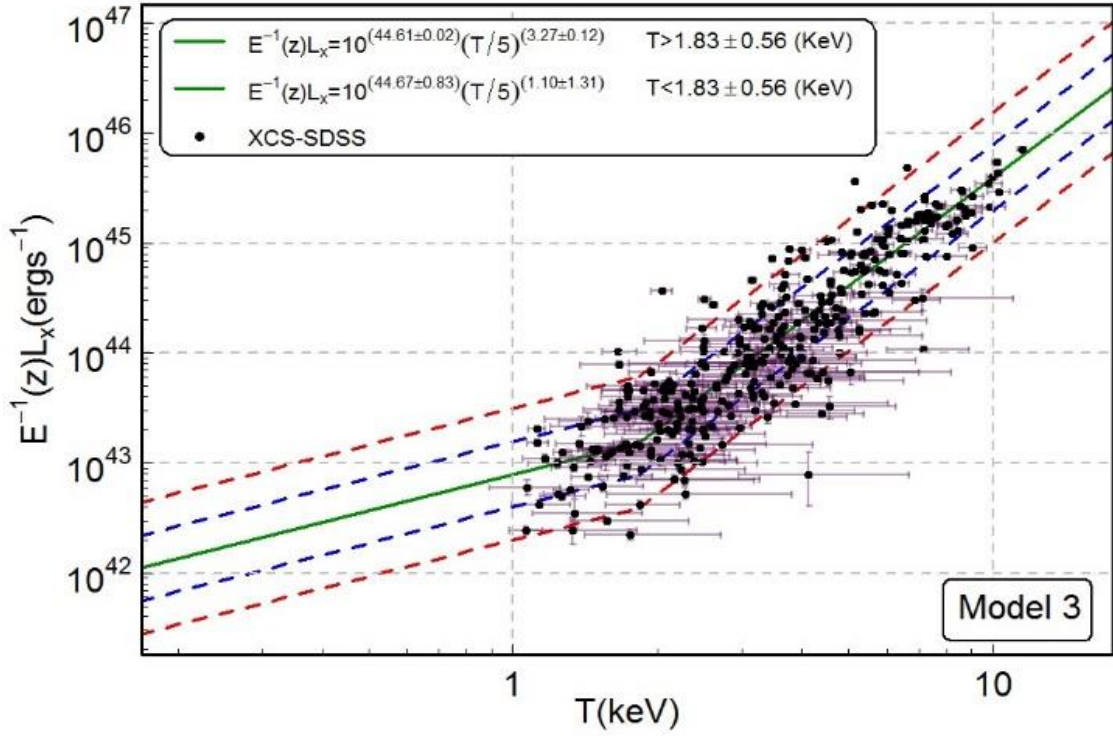


Figure 13. The $L_X - T$ relation assuming model 3. The prior distribution for the temperature is modelled through a time-evolving Gaussian.

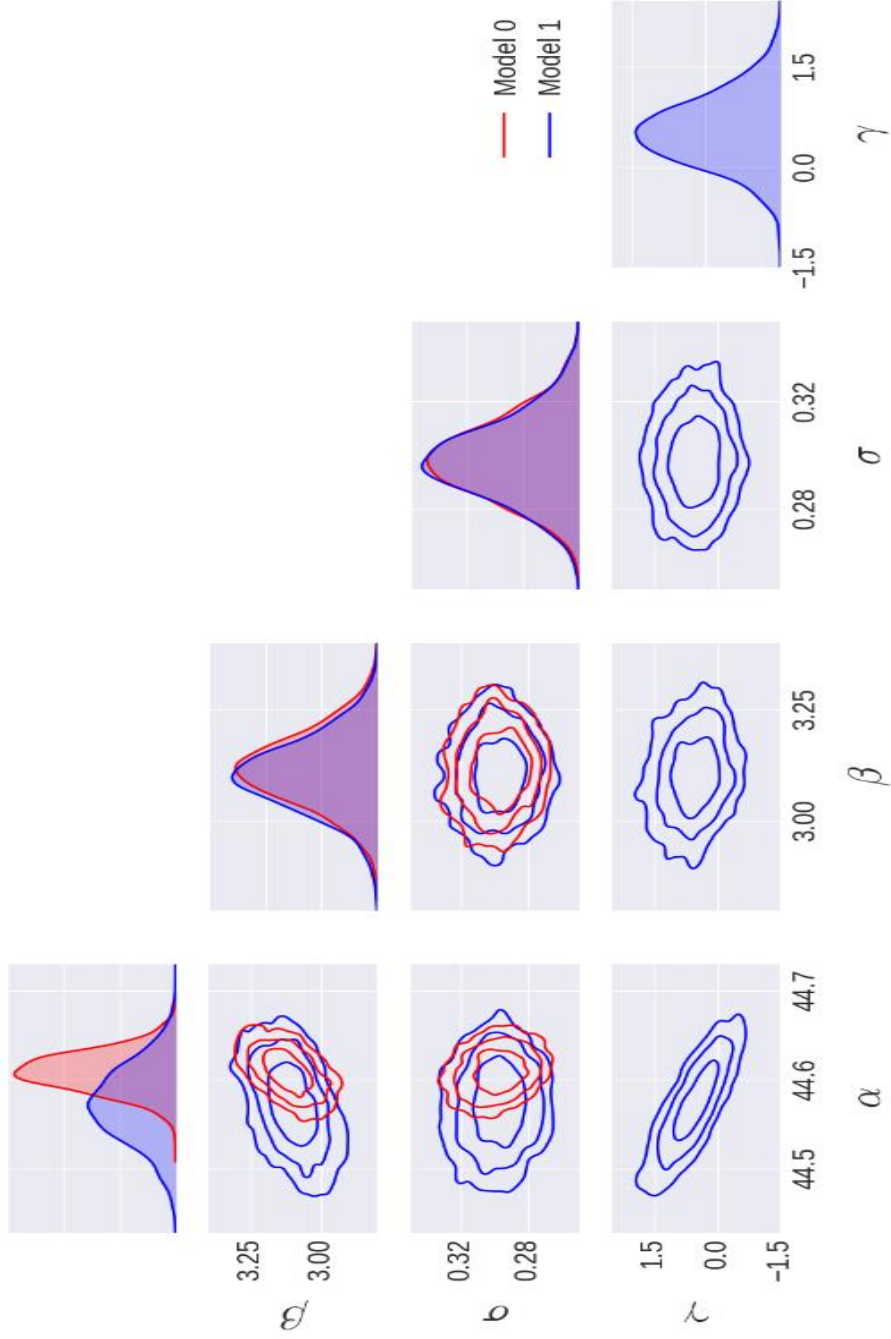


Figure 14. Comparison of the one- and two-dimensional marginalised posterior distributions for model 0 vs 1. Note that the horizontal axis that is associated with each one-dimensional marginalised posterior distribution is below it, at the bottom of the plot. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.

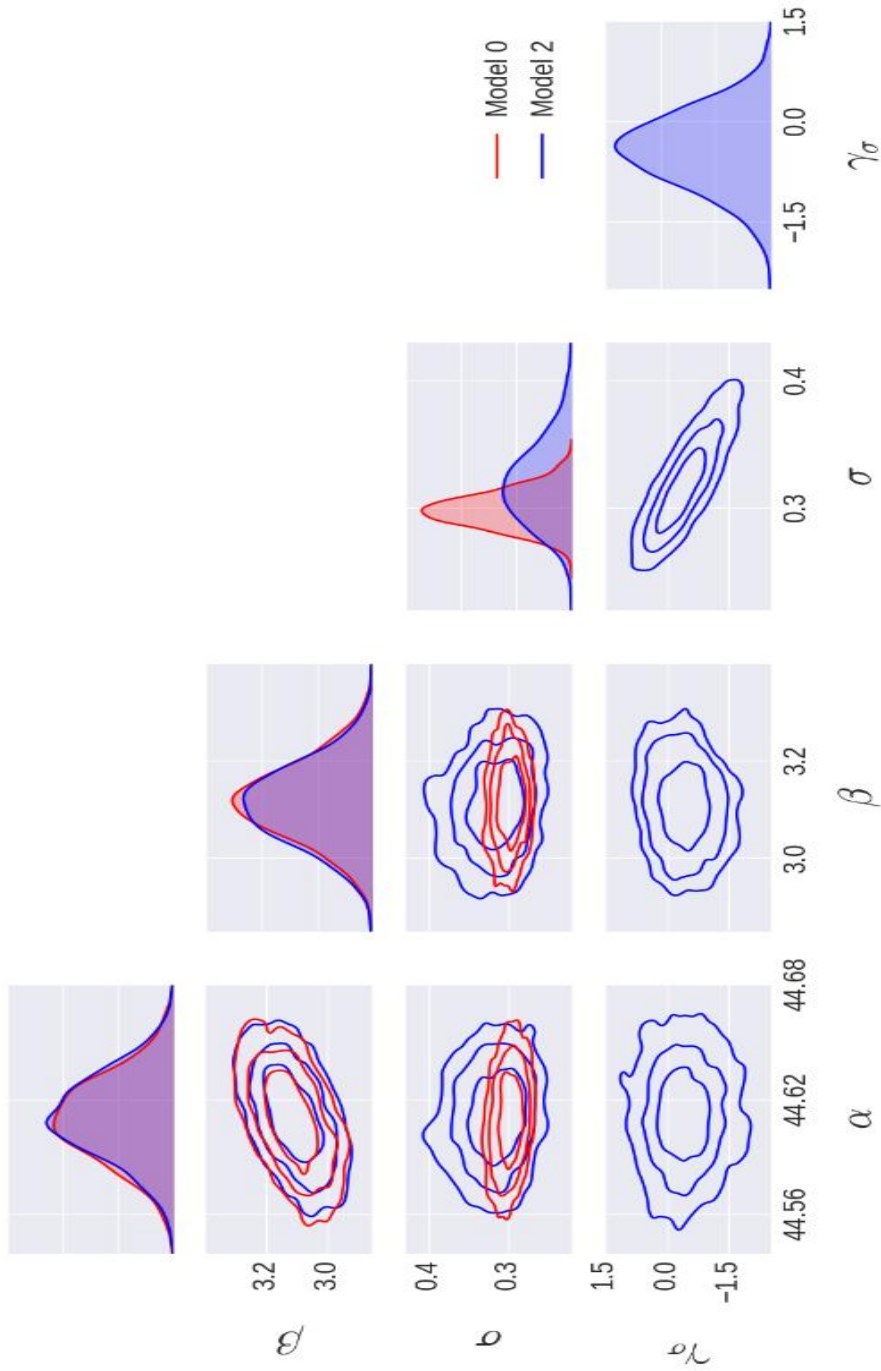


Figure 15. Comparison of the one- and two-dimensional marginalised posterior distributions for model 0 vs 2. Note that the horizontal axis that is associated with each one-dimensional marginalised posterior distribution is below it, at the bottom of the plot. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.

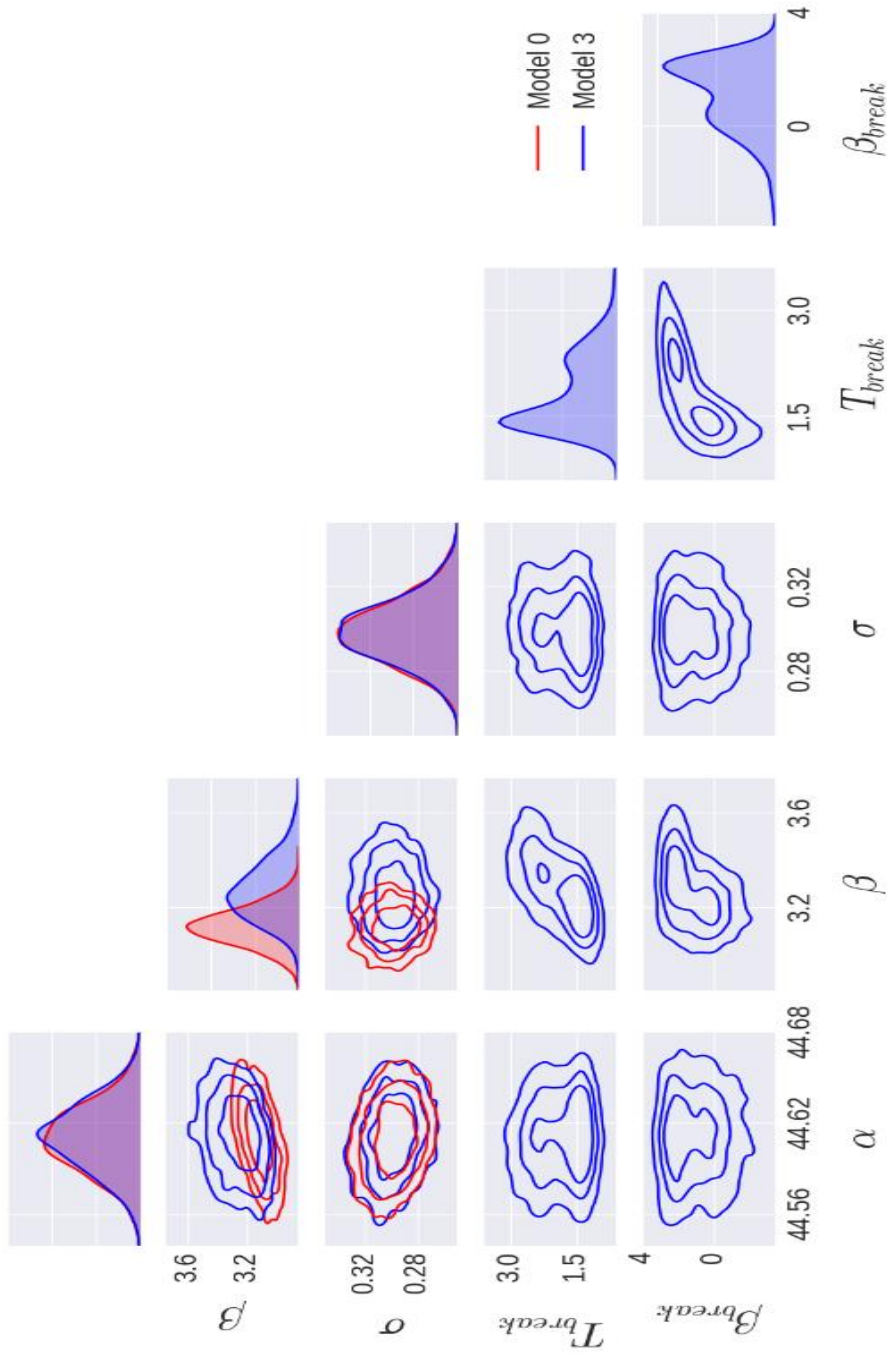


Figure 16. Comparison of the one- and two-dimensional marginalised posterior distributions for model 0 vs 3. Note that the horizontal axis that is associated with each one-dimensional marginalised posterior distribution is below it, at the bottom of the plot. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.

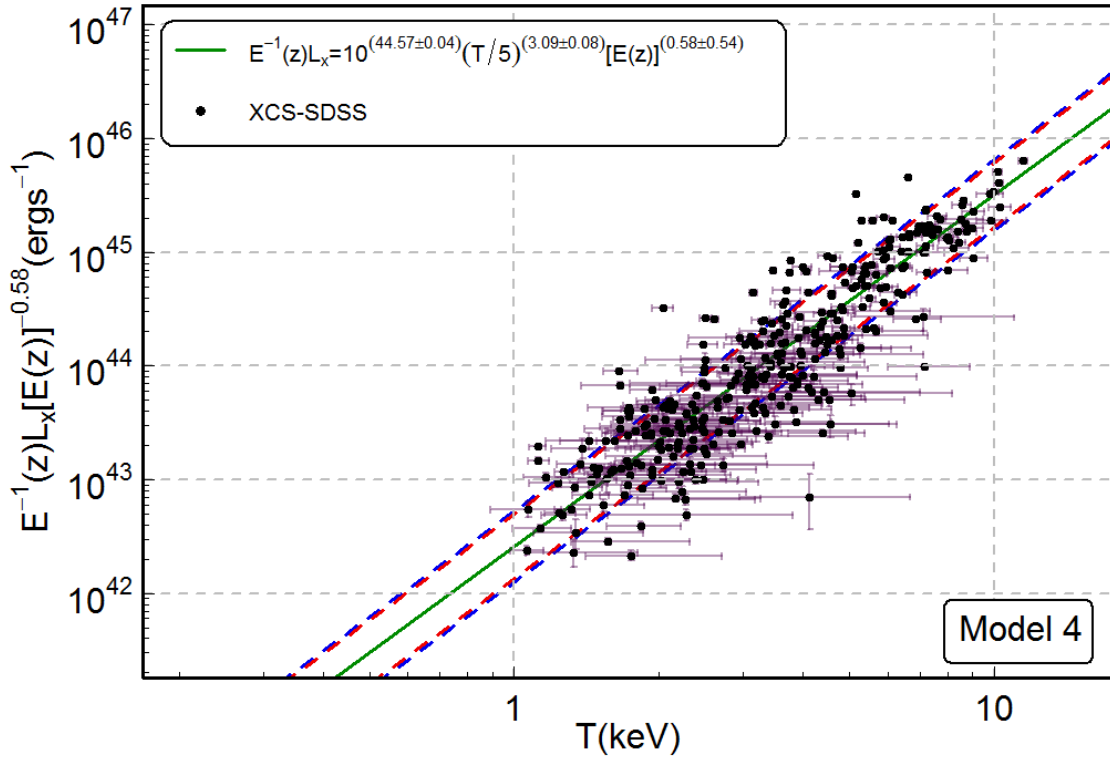


Figure 17. The $L_X - T$ relation assuming model 4. The prior distribution for the temperature is modelled through a time-evolving Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts.

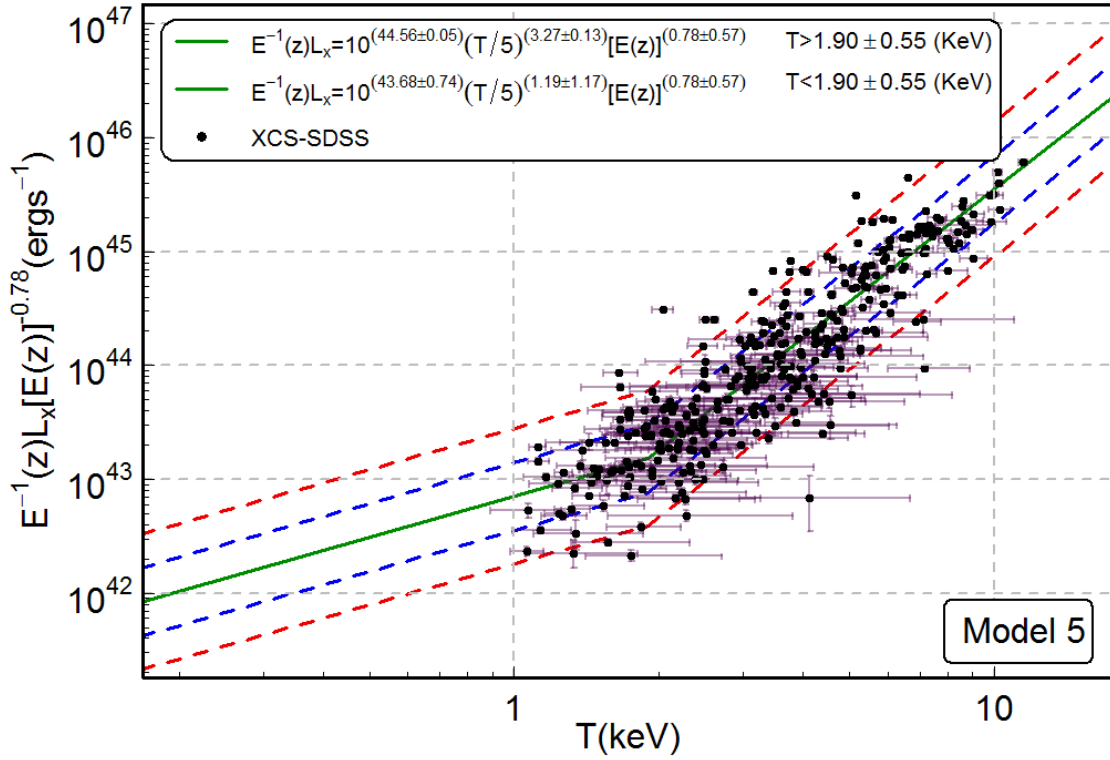


Figure 18. The $L_X - T$ relation assuming model 5. The prior distribution for the temperature is modelled through a time-evolving Gaussian.

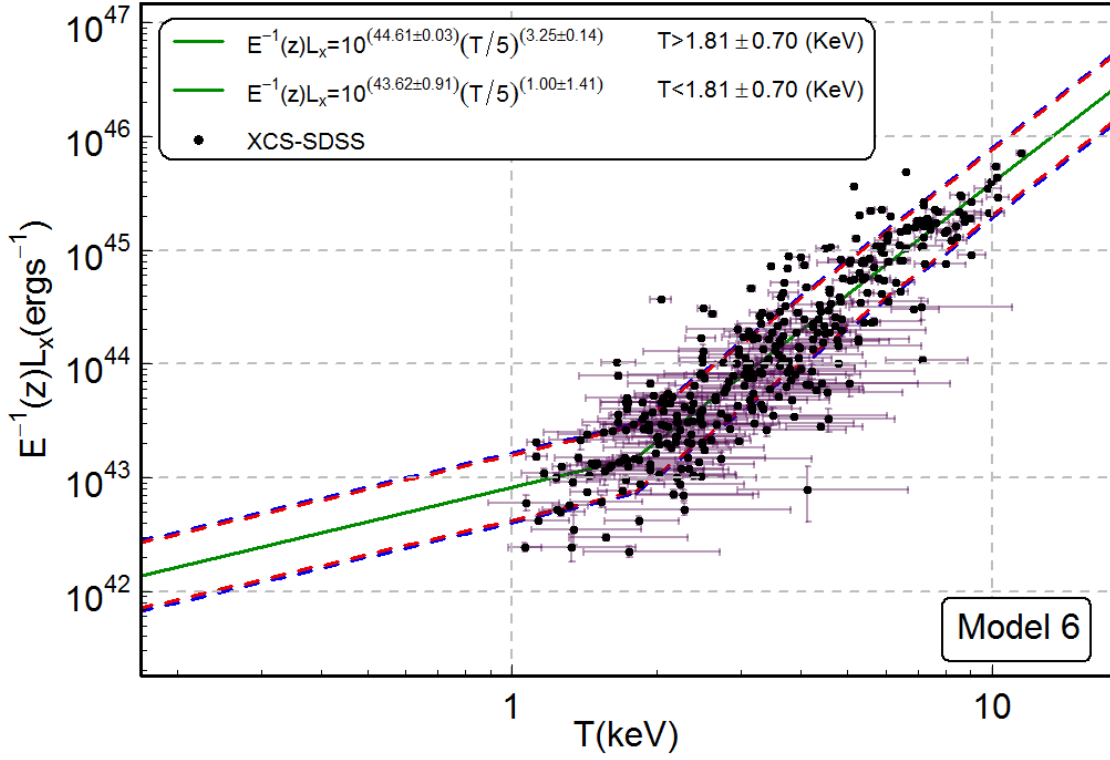


Figure 19. The $L_X - T$ relation assuming model 6. The prior distribution for the temperature is modelled through a time-evolving Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts.

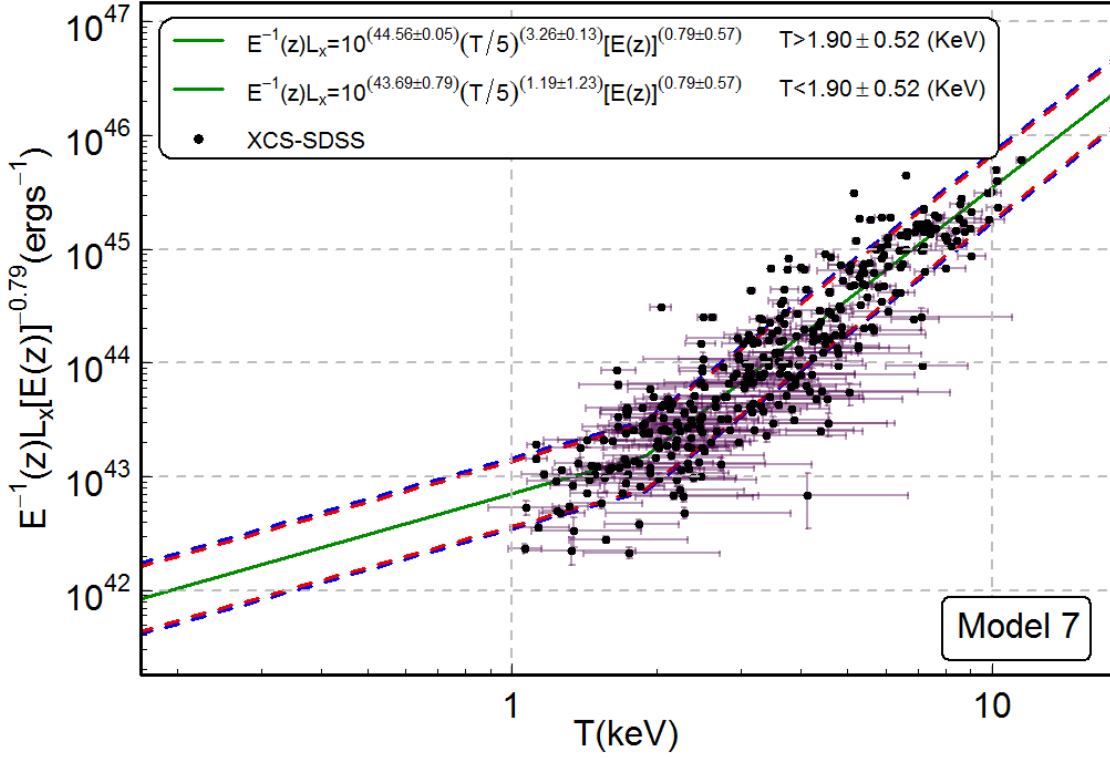


Figure 20. The $L_X - T$ relation assuming model 7. The prior distribution for the temperature is modelled through a Gaussian. The dashed lines now indicate the amplitude of the intrinsic vertical scatter about the regression line only at the 1σ level, for the minimum (0.1, outer blue) and maximum (0.6, inner red) sample redshifts.

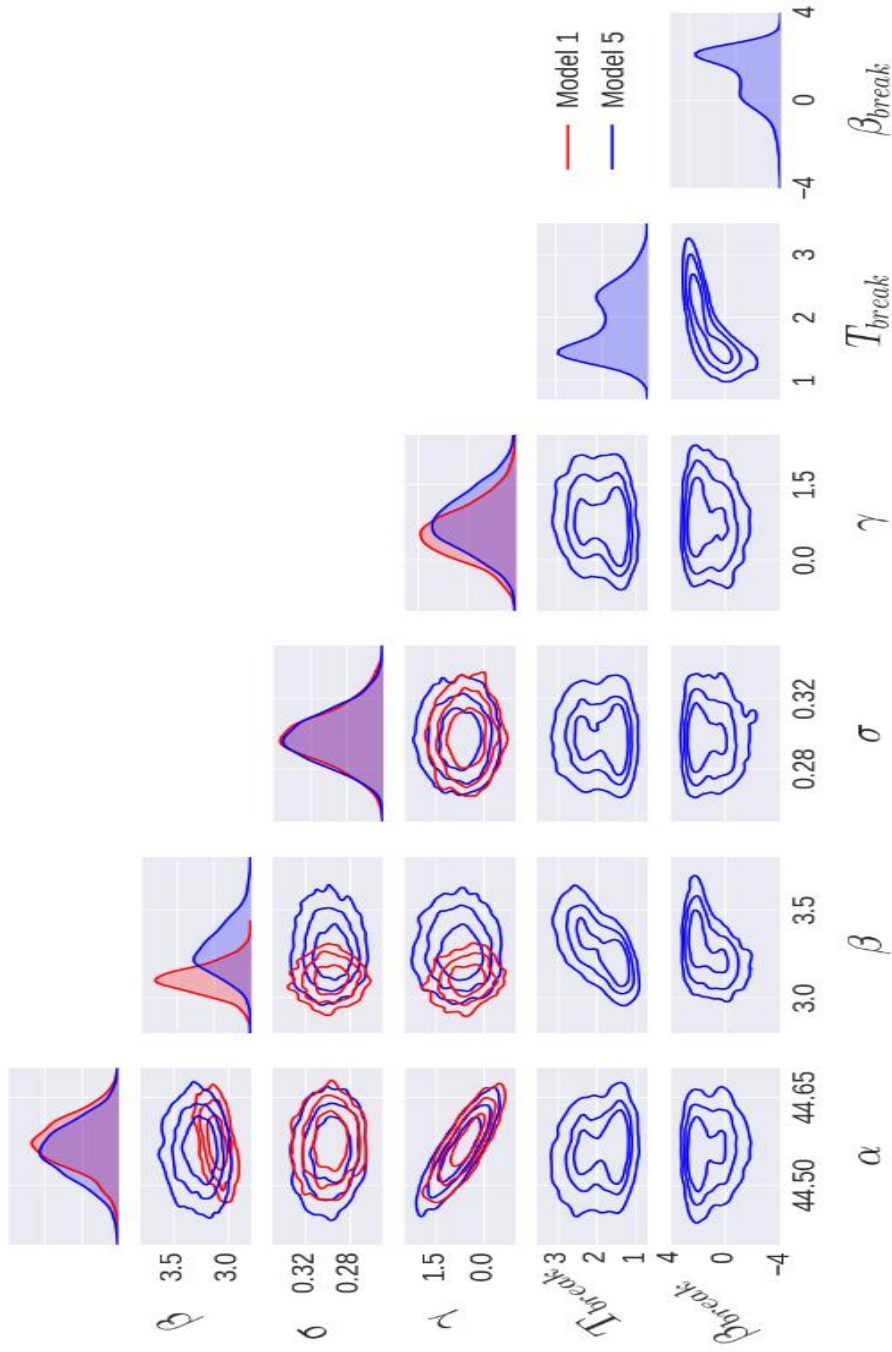


Figure 21. Comparison of the one- and two-dimensional marginalised posterior distributions for model 1 vs. 5. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.

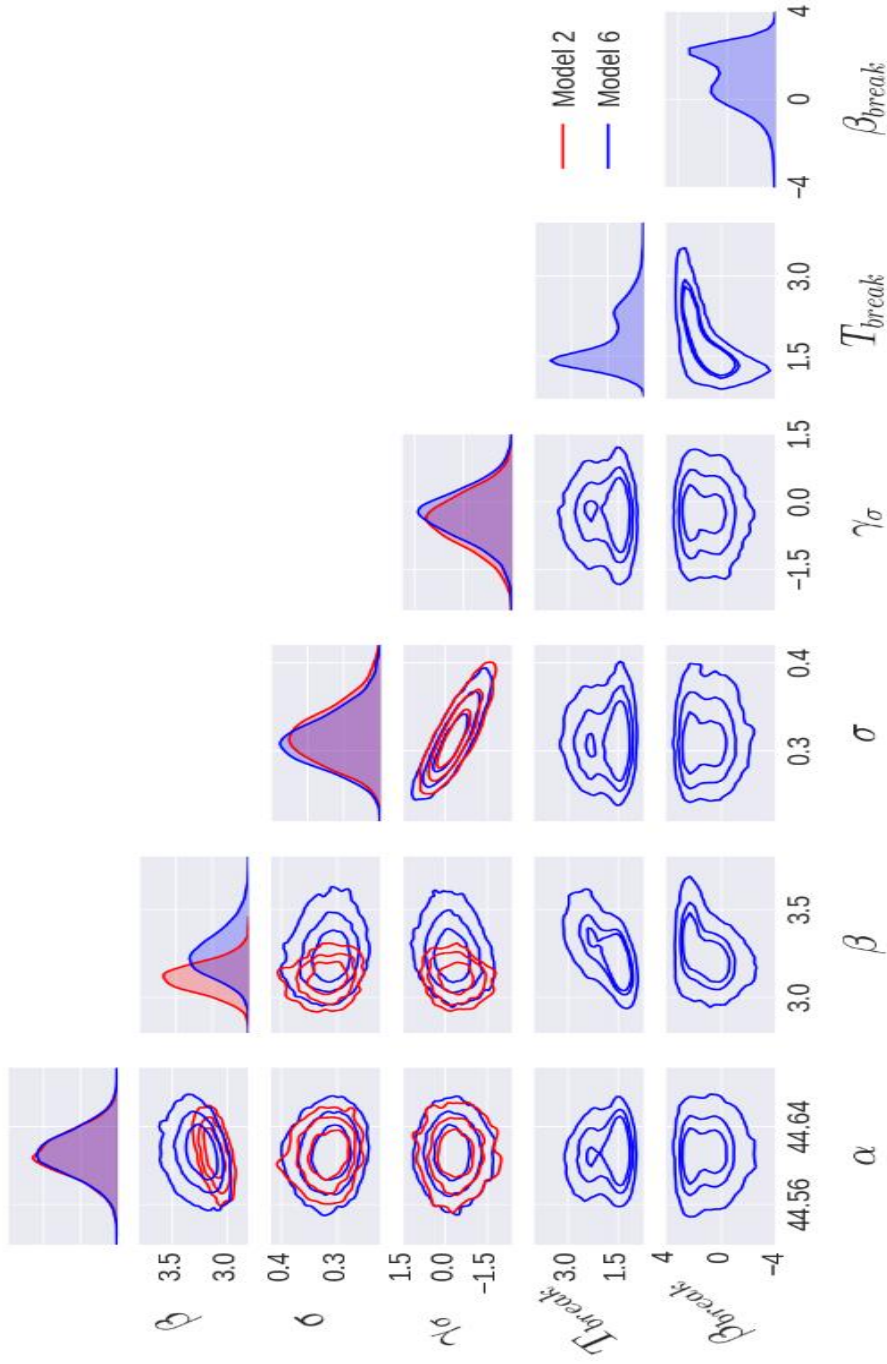


Figure 22. Comparison of the one- and two-dimensional marginalised posterior distributions for model 2 vs. 6. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.

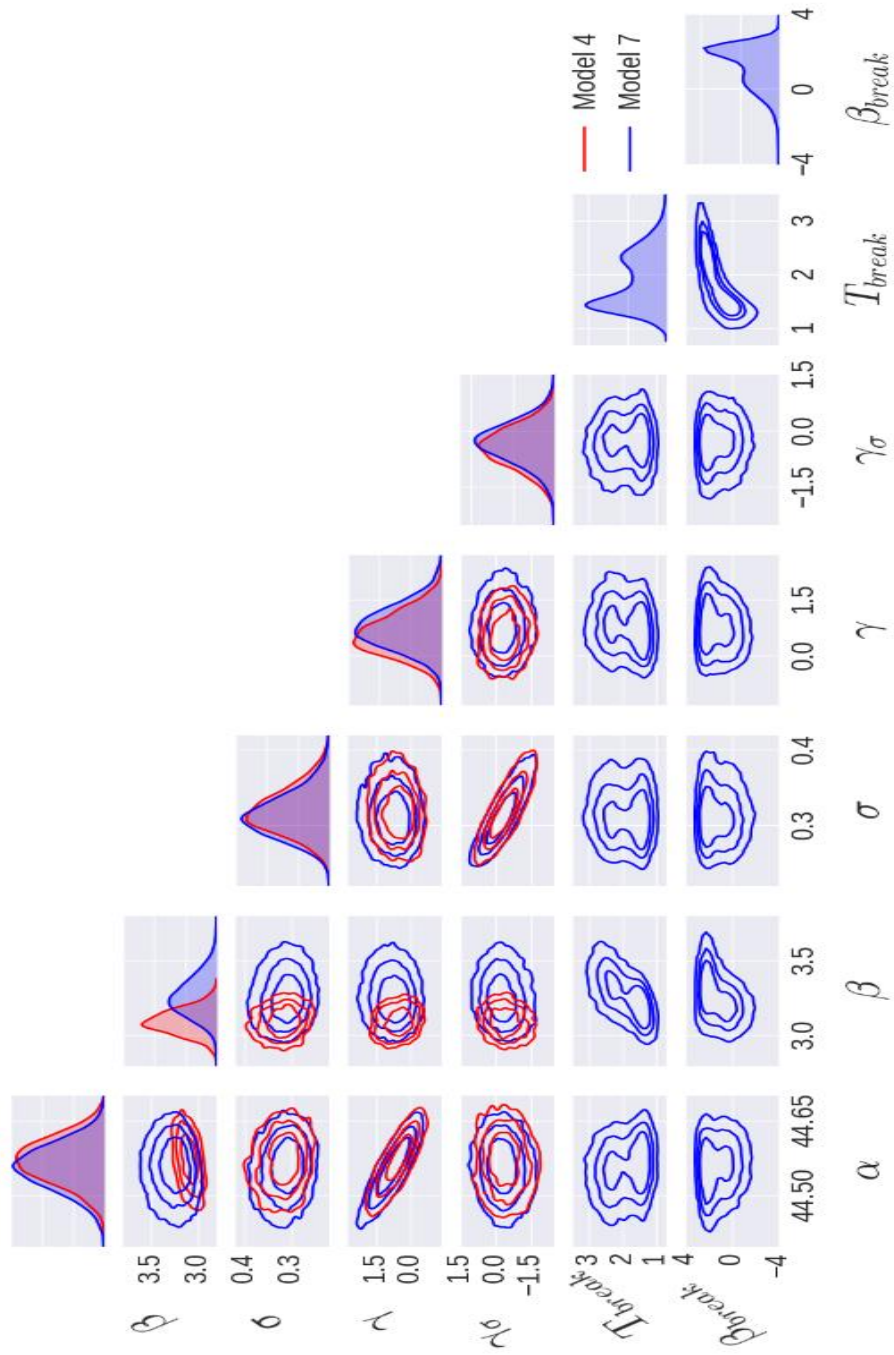


Figure 23. Comparison of the one- and two-dimensional marginalised posterior distributions for model 4 vs. 7. The two-dimensional contours enclose 0.50 (inner), 0.80 (middle) and 0.95 (outer) of the total probability.

In Table 2, we provide the estimated median value and symmetric (or equal-tailed) credible intervals, centred on the median, for the marginal posterior distributions of each hyper-parameter of interest associated with the models being considered.

Table 2. The median and symmetric credible intervals, centred on the median, that contain 68.3% and 95.4% of the probability for the marginal posterior distributions of the interesting hyper-parameters associated with each model. After each median value, are represented the numbers that need to be added and subtracted from it, in order to obtain the limits associated with the respective credible interval.

Model	α	β	σ	γ	$\gamma\sigma$	T_{break}	β_{break}
0u	44.61 ^{+0.02+0.04} _{-0.02-0.04}	3.00 ^{+0.07+0.15} _{-0.07-0.14}	0.30 ^{+0.01+0.03} _{-0.01-0.03}	0	0	—	—
0	44.61 ^{+0.02+0.05} _{-0.02-0.05}	3.12 ^{+0.08+0.16} _{-0.08-0.16}	0.30 ^{+0.02+0.03} _{-0.01-0.03}	0	0	—	—
1u	44.56 ^{+0.04+0.08} _{-0.04-0.09}	2.99 ^{+0.08+0.15} _{-0.08-0.16}	0.30 ^{+0.01+0.03} _{-0.02-0.03}	0.73 ^{+0.56+1.16} _{-0.53-0.96}	0	—	—
1	44.57 ^{+0.04+0.08} _{-0.04-0.09}	3.10 ^{+0.08+0.17} _{-0.08-0.17}	0.30 ^{+0.01+0.03} _{-0.01-0.03}	0.55 ^{+0.52+1.09} _{-0.49-1.01}	0	—	—
2	44.61 ^{+0.02+0.04} _{-0.02-0.04}	3.11 ^{+0.08+0.16} _{-0.08-0.17}	0.32 ^{+0.03+0.07} _{-0.03-0.05}	0	-0.41 ^{+0.55+1.07} _{-0.58-1.21}	—	—
3	44.61 ^{+0.02+0.04} _{-0.02-0.05}	3.26 ^{+0.14+0.29} _{-0.11-0.21}	0.30 ^{+0.01+0.03} _{-0.01-0.03}	0	0	1.65 ^{+0.77+1.34} _{-0.31-0.52}	1.40 ^{+0.95+1.40} _{-1.56-3.12}
4	44.57 ^{+0.04+0.08} _{-0.05-0.09}	3.09 ^{+0.09+0.17} _{-0.08-0.16}	0.32 ^{+0.03+0.07} _{-0.03-0.05}	0.55 ^{+0.61+1.14} _{-0.49-0.96}	-0.42 ^{+0.53+1.06} _{-0.56-1.15}	—	—
5	44.56 ^{+0.05+0.09} _{-0.05-0.09}	3.26 ^{+0.15+0.31} _{-0.12-0.22}	0.30 ^{+0.02+0.03} _{-0.01-0.03}	0.76 ^{+0.61+1.20} _{-0.55-1.06}	0	1.79 ^{+0.66+1.18} _{-0.42-0.64}	1.50 ^{+0.79+1.24} _{-1.53-3.02}
6	44.61 ^{+0.02+0.04} _{-0.02-0.05}	3.24 ^{+0.14+0.31} _{-0.11-0.21}	0.31 ^{+0.03+0.07} _{-0.03-0.05}	0	-0.27 ^{+0.52+1.06} _{-0.55-1.18}	1.58 ^{+0.80+1.48} _{-0.27-0.50}	1.15 ^{+1.17+1.68} _{-1.37-3.00}
7	44.57 ^{+0.04+0.08} _{-0.04-0.09}	3.30 ^{+0.14+0.29} _{-0.13-0.23}	0.32 ^{+0.03+0.07} _{-0.03-0.05}	0.73 ^{+0.48+0.97} _{-0.45-0.87}	-0.64 ^{+0.50+0.98} _{-0.54-1.11}	2.29 ^{+0.47+1.12} _{-0.82-1.08}	2.03 ^{+0.44+0.78} _{-1.49-3.07}

In Table 3, we provide the expected value (i.e. mean), standard deviation and skewness for the same marginal posterior distributions. As can be seen, some of the marginal posterior distributions deviate significantly from a Gaussian distribution, for which the skewness is zero. For these distributions the median and mean do not coincide, which is why we decided to include more information about the marginal posterior distributions than usual. The first three (raw) cumulants of a given probability distribution coincide, respectively, with the mean, the variance (i.e. the square of the standard deviation), and the product of the skewness with the cube of the standard deviation. The values of these cumulants can be used to approximate each marginal posterior distribution through an Edgeworth expansion around a Gaussian distribution with the same mean and variance, for example using the R package PDQutils available both at CRAN⁶ and GitHub⁷.

It is clear from both Tables 2 and 3, that the introduction of a Gaussian (mixture) prior mainly affects the expected value for β . This can be understood by first noticing that multiplying the likelihood by a Gaussian effectively leads to a re-weighting of the sample data, where the data points with values for $\log(T/5)$ closer to the Gaussian expected value have their relative importance downweighted with respect to the others. Further, for all models considered, the Gaussian prior expected value is always around -0.26

⁶ <https://cran.r-project.org/web/packages/PDQutils/>

⁷ <https://github.com/shabbychef/PDQutils/>

(2.75 keV). Although that is close to the mode ($-0.325, 2.37$ keV) of the sample distribution for $\log(T/5)$, this has a significant positive skew, as can be seen in Figure 8, with median and mean equal to -0.17 (3.38 keV). Therefore, the introduction of such a Gaussian (mixture) prior effectively leads to a re-weighting of the data points in favour of those located in the high temperature tail of the sample distribution. Because these prefer a higher value for β , as can be seen by what happens to the marginal posterior distribution for this parameter in the models where a change-point is introduced in the $L_X - T$ relation, the re-weighting in their favour leads to the observed increase in the expected value for β when a Gaussian (mixture) prior is introduced (in model 0 vs 0u, and in model 1 vs 1u).

Table 3. The mean, standard deviation and skewness for the marginal posterior distributions of the interesting hyper-parameters associated with each model.

Model	α	β	σ	γ	γ_σ	T_{break}	β_{break}
0u	44.61, 0.02, 0.03	3.00, 0.08, 0.04	0.30, 0.01, 0.17	0	0	—	—
0	44.61, 0.02, -0.02	3.12, 0.08, 0.06	0.30, 0.01, 0.18	0	0	—	—
1u	44.56, 0.04, -0.14	2.99, 0.08, -0.03	0.30, 0.01, 0.15	0.75, 0.54, 0.22	0	—	—
1	44.57, 0.04, -0.12	3.11, 0.08, -0.01	0.30, 0.02, 0.21	0.56, 0.51, 0.20	0	—	—
2	44.61, 0.02, -0.04	3.11, 0.08, 0.01	0.32, 0.03, 0.51	0	$-0.42, 0.57, -0.18$	—	—
3	44.61, 0.02, -0.13	3.27, 0.12, 0.44	0.30, 0.01, 0.13	0	0	1.83, 0.56, 1.56	1.10, 1.31, -1.44
4	44.57, 0.04, -0.11	3.09, 0.08, 0.03	0.32, 0.03, 0.41	0.58, 0.54, 0.17	$-0.43, 0.55, -0.06$	—	—
5	44.56, 0.05, -0.12	3.27, 0.14, 0.41	0.30, 0.01, 0.23	0.79, 0.57, 0.16	0	1.90, 0.55, 1.64	1.19, 1.17, -0.91
6	44.61, 0.03, -5.36	3.26, 0.14, 2.00	0.31, 0.03, 0.45	0	$-0.28, 0.55, -0.14$	1.81, 0.70, 4.57	1.00, 1.41, -2.41
7	44.56, 0.05, -0.20	3.26, 0.13, 0.38	0.31, 0.03, 0.47	0.80, 0.57, 0.22	$-0.31, 0.55, -0.16$	1.90, 0.52, 0.55	1.20, 1.23, -2.00

The consequence of completely removing the data points associated with lower temperatures from the estimation of β can be seen when a change-point is introduced, in model 3 vs 0, in model 5 vs 1, model 6 vs 2 and in model 7 vs 4. In every case the expected value for β increases, by a similar margin as a result of introducing a Gaussian (mixture) prior. That happens because, as previously mentioned, the data points associated with lower temperatures prefer a smaller value for β , as can be seen from the expected values for β_{break} .

Finally, the introduction of the time-evolution parameters, γ and γ_σ , affects most importantly the expected values for the base parameters to which they are associated, respectively α and σ , as can be seen by comparing the expected values for the models 1u vs 0u, 1 vs 0, 4 vs 2, 5 vs 3 and 7 vs 6 (with respect to γ), and models 2 vs 0, 4 vs 1, 6 vs 3 and 7 vs 5 (with respect to γ_σ), and taking into account the relative uncertainty associated with those parameters. Although not as strongly, the values for the change-

point associated parameters, T_{break} and β_{break} , especially the former, are also affected by the introduction of γ and/or γ_σ .

In Table 4, we provide the estimates for WBIC and WAIC, as well as the associated sample standard deviations, with respect to each model considered.

Table 4. Estimates for WBIC, its sample standard deviation, σ_{WBIC} , WAIC, and its sample standard deviation, σ_{WAIC} , with respect to each model.

Model	WBIC	σ_{WBIC}	WAIC	σ_{WAIC}
0u	2052	38	6.51	0.11
0	-1278	37	-2.88	0.11
1u	2054	38	6.52	0.11
1	-1277	37	-2.88	0.11
2	-1277	37	-2.88	0.11
3	-1277	37	-2.85	0.11
4	-1282	37	-2.89	0.11
5	-1280	37	-2.85	0.11
6	-1278	37	-2.85	0.11
7	-1281	37	-2.85	0.11

The results clearly indicate strong support for the introduction of a Gaussian (mixture) prior to model the joint effect of the population-level temperature function and any sample selection effects not explicitly taken into consideration (like those associated with SDSS-DR8-redMaPPer catalogue assembly, in our case). However, the data does not seem to hold enough information to differentiate between all favoured models, 0 to 7, in that respect. All these models have associated WBIC and WAIC estimates which overlap well within the respective sample standard deviations. This suggests that another, statistically equivalent, sample taken from the same underlying population could easily lead to a different ranking of the models 0 to 7, based on their associated WBIC and WAIC estimates for such sample. Interestingly, model 4 is that which gathers more support from the data, according to both the WBIC and WAIC model comparison methods. But, these two methods differ with respect to which individual feature, among those that make models 1, 2 and 3 distinct from model 0, is better supported by the data: WBIC suggests it is time-evolution of the intrinsic scatter about the $L_X - T$ relation, parametrised by γ_σ ; while WAIC prefers time-evolution of the normalization of the $L_X - T$ relation beyond the self-similar expectation, parametrised by γ . However, both methods agree that the individual feature with least support in the data is an abrupt transition or change-point in the $L_X - T$ relation, parametrised through T_{break} and β_{break} . Assuming a priori any of models 0 to 7 to be true, the data consistently suggests a positive deviation with respect to self-similar evolution for the $L_X - T$ relation (i.e. at fixed temperature the X-ray

luminosity increases with redshift), a negative redshift evolution in the intrinsic scatter about the $L_X - T$ relation (i.e. the scatter decreases with redshift), as well as a change-point location just below 2 keV and an increase in the exponent for the $L_X - T$ relation when moving from the group to the cluster regime.

In Figure 24, we compare, for all 353 systems in our sample, the expected values and associated standard deviations associated with: the observed temperatures and X-ray luminosities, T_{obs} and L_{obs} ; the temperatures and X-ray luminosities, T_{model} and L_{model} , obtained through LIRA, assuming a priori the $L_X - T$ relation to be described by model 4. In the former case, expected values and standard deviations were estimated by approximating the probability distributions for the temperatures and X-ray luminosities that result from our X-ray analysis pipeline as univariate split Gaussian distributions (Villani et al. 2006), and using the modes and 68% confidence intervals presented in Table 1. Further, the hierarchical modelling framework, implemented through LIRA, actually yields expected values and standard deviations for the (base 10) logarithms of T_{model} and L_{model} . However, the posterior probability distributions for these quantities are nearly Gaussian, thus implying that the posterior probability distributions for T_{model} and L_{model} can be well approximated by (base 10) log-normal distributions. This means that the expected values, μ_{model} , and associated standard deviations, σ_{model} , for T_{model} and L_{model} can be estimated from the respective quantities, $\mu_{model,log}$ and $\sigma_{model,log}$, for $\log T_{model}$ and $\log L_{model}$, through

$$\mu_{model} = 10^{(\mu_{model,log} + \frac{\ln 10}{2} \sigma_{model,log}^2)} \quad (54)$$

and

$$\sigma_{model} = \mu_{model} \sqrt{10^{\ln(10) \sigma_{model,log}^2} - 1} \quad (55)$$

where, as usual, $\ln$ stands for the natural logarithm.

In Figure 25, we compare the sample distributions of most probable values for the observed temperatures and X-ray luminosities, with those obtained through LIRA, assuming a priori the $L_X - T$ relation to be described by model 4. In the later case, the most probable values (i.e. modes) are given by

$$Mo_{model} = 10^{(\mu_{model,log} - \ln 10 \sigma_{model,log}^2)} \quad (56)$$

and are listed in a supplementary table available with the online edition of this article, together with the associated expected values and standard deviations, with respect to

the temperature and X-ray luminosity posterior distributions for all systems in our sample. The first 10 rows are shown in Table 5.

Table 5. The first 10 rows of a supplementary table available with the online edition of this article, providing the XCS ID, most probable values for the temperature, $Mo_{T,model}$ (in units of keV), and X-ray bolometric luminosity within R_{500} , $Mo_{L_x,model}$ (in units of $10^{44} \text{ erg s}^{-1}$), as well as the associated expected values, $\mu_{T,model}$ and $\mu_{L_x,model}$, and standard deviations, $\sigma_{T,model}$ and $\sigma_{L_x,model}$, obtained through LIRA, assuming a priori the $L_X - T$ relation to be described by model 4, for all 353 sample systems.

XCS ID	$Mo_{T,model}$	$\mu_{T,model}$	$\sigma_{T,model}$	$Mo_{L,model}$	$\mu_{L,model}$	$\sigma_{L,model}$
XMMXCS J140101.9 + 025238.3	6.64	6.64	0.13	55.39	55.39	0.10
XMMXCS J142601.0 + 374937.0	8.65	8.65	0.53	31.88	31.88	0.22
XMMXCS J172009.8 + 263725.8	6.15	6.15	0.21	21.50	21.50	0.09
XMMXCS J224321.4 – 093550.2	7.24	7.24	0.40	33.11	33.11	0.11
XMMXCS J212939.7 + 000516.9	5.29	5.30	0.27	22.75	22.75	0.13
XMMXCS J075124.2 + 173057.7	3.29	3.29	0.26	2.86	2.86	0.02
XMMXCS J172227.0 + 320758.0	7.21	7.21	0.43	27.57	27.57	0.19
XMMXCS J104044.2 + 395711.1	3.74	3.74	0.18	7.39	7.39	0.04
XMMXCS J111253.4 + 132640.2	4.89	4.89	0.29	6.12	6.12	0.04
XMMXCS J024803.3 – 033143.4	3.77	3.77	0.19	9.64	9.64	0.06

As expected, assuming a priori a specific model for the $L_X - T$ relation induces regression towards the mean, leading to a decrease in the variance of the temperature and X-ray luminosity sample distributions. This effect is significantly stronger for the former with respect to the later, as a result of the higher uncertainty level associated with the data-only based estimates for the temperatures. Substituting model 4 by any other model between 0 and 7 yields very similar results. On an individual system basis, the higher the overall uncertainty in data-only based estimates, the larger can be the difference between these and estimates that assume a priori a specific model for the $L_X - T$ relation. This is well illustrated in the case of cluster XMMXCS J222855.3+203920.1, which stands out as an outlier in the L_{obs} versus L_{model} plot in Figure 24. The significant difference between these two estimates stems from the large uncertainty in both the temperature and X-ray luminosity data-only based estimates for this system, together with a well below-average value for the later given the former, with respect to the sample distribution of such estimates.

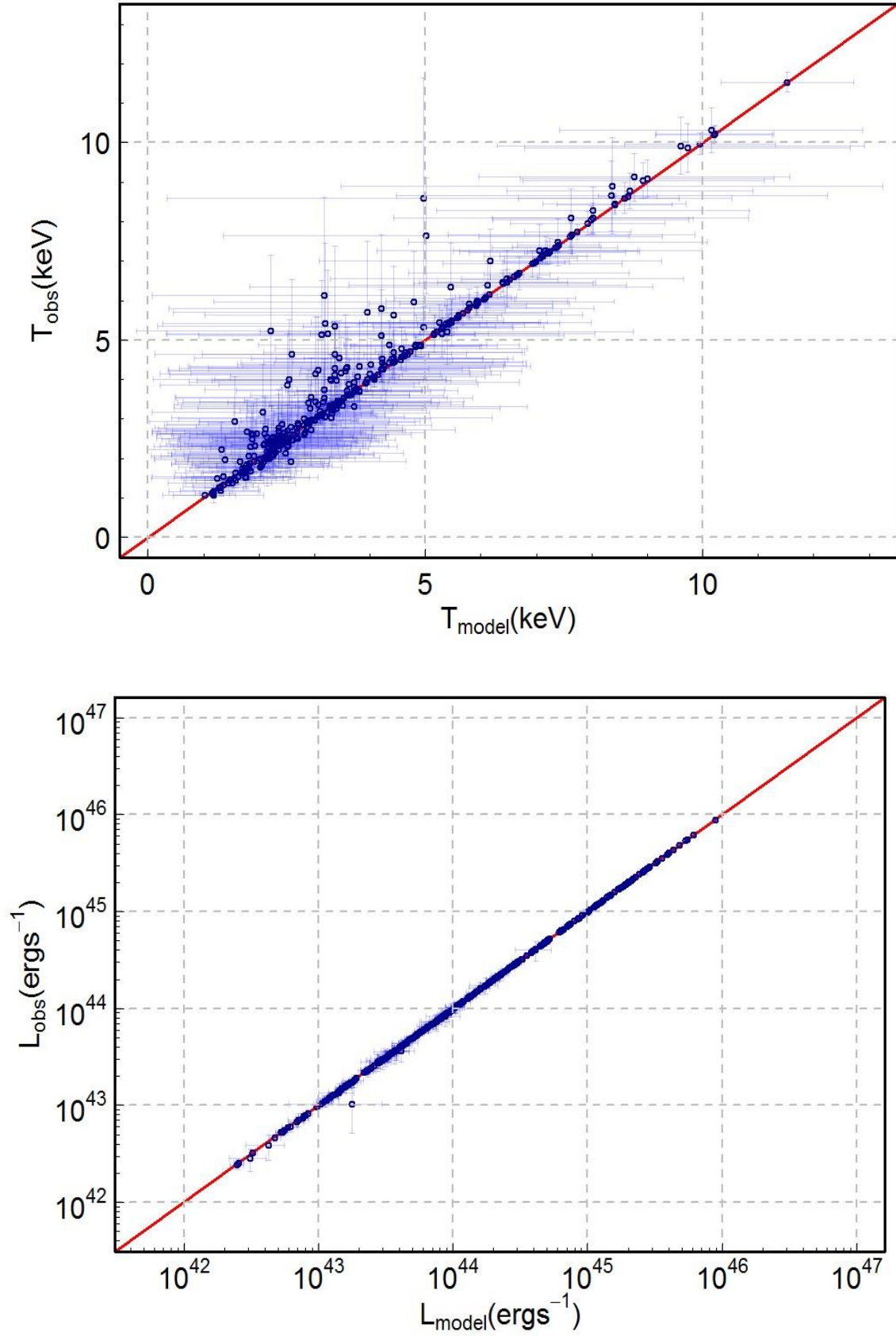


Figure 24. Comparison, for all 353 systems in our sample, of the expected values and associated standard deviations associated with: the observed temperatures and X-ray luminosities, T_{obs} and L_{obs} ; the temperatures and X-ray luminosities, T_{model} and L_{model} , obtained assuming a priori the $L_X - T$ relation to be described by model 4.

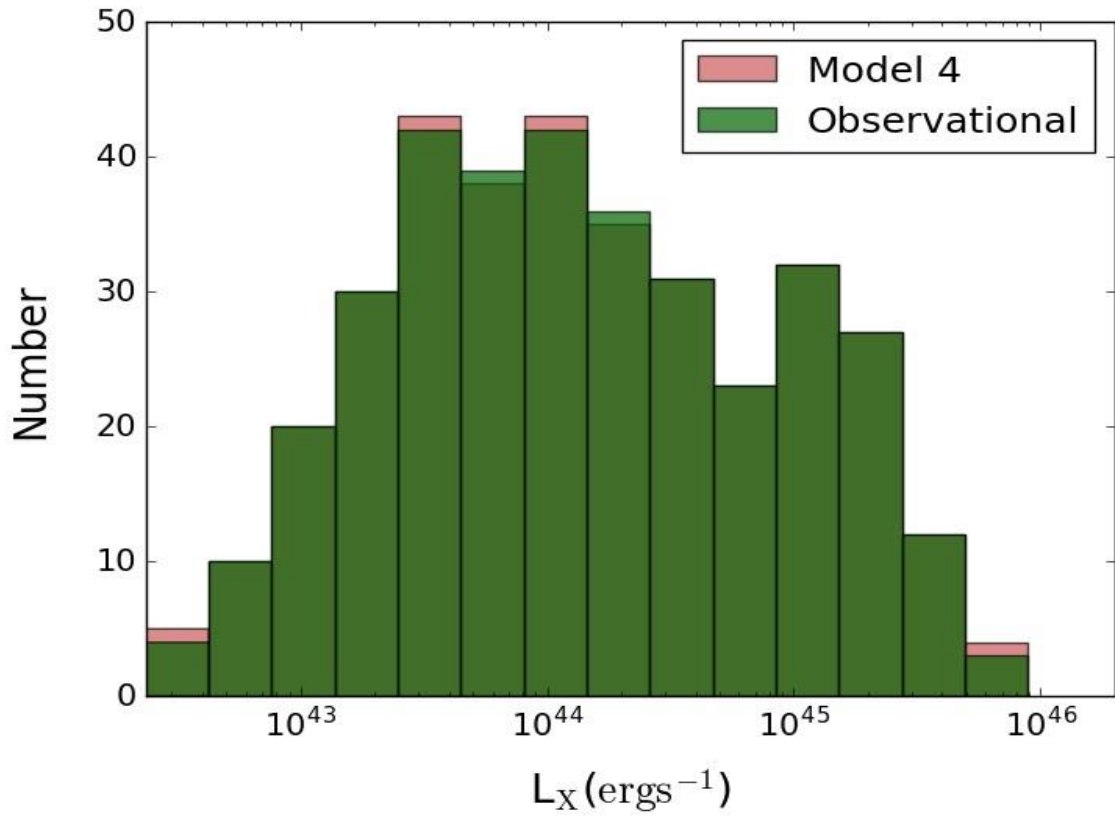
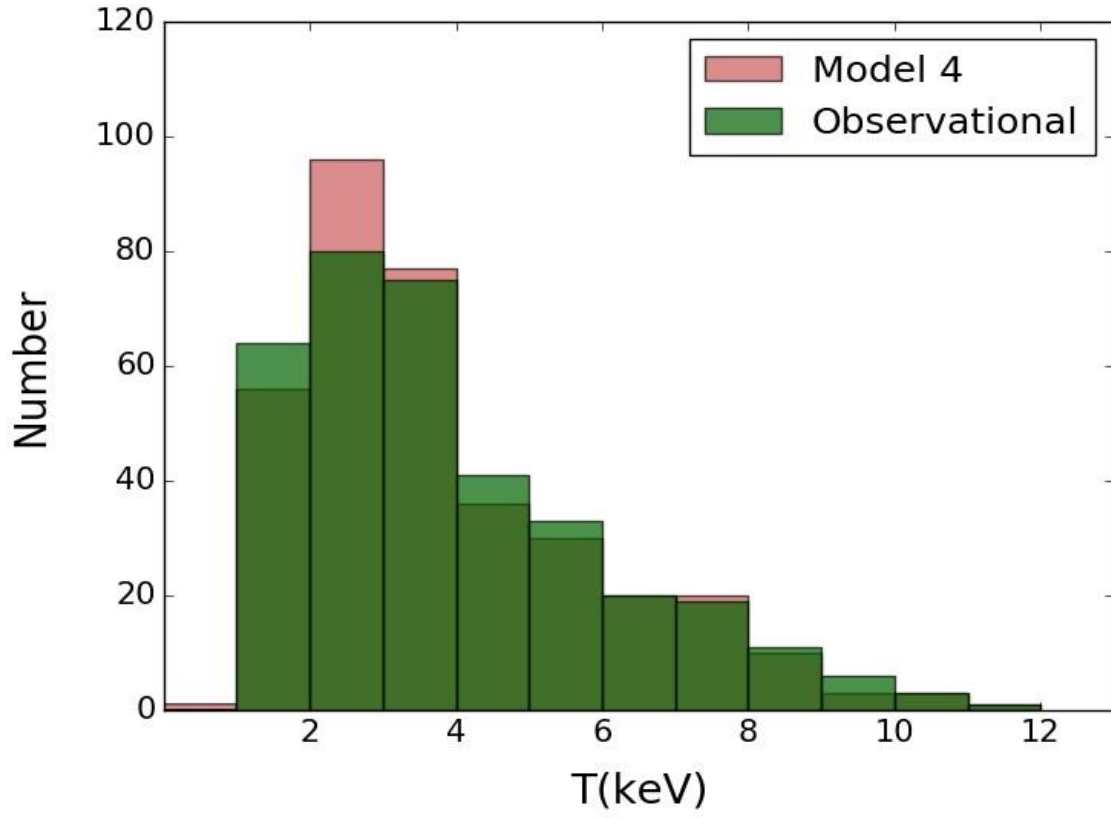


Figure 25. Comparison of the sample distributions of most probable values for the observed temperatures and X-ray luminosities, with those obtained through LIRA, assuming a priori the $L_X - T$ relation to be described by model 4.

3.4.2 Taking into account sample selection effects

The sample selection effects should have a bigger impact towards the lower temperature and luminosity end of the underlying distribution of the objects we are considering, i.e. towards the group regime. This means that, in principle, those models which will be most affected, both in terms of their associated WBIC and WAIC as well as the location and shape of the joint posterior probability distribution for the associated parameters, will be those which allow for a change-point.

Unfortunately, after computing the IS ratios for the models considered, the diagnostics showed a stable solution for the target distribution could not be achieved except for model 0. As expected, because this model does not include a change-point, its associated WBIC and WAIC as well as the location and variance of the marginal posterior probability distributions for the associated parameters did not change significantly. For example, the mean value and its associated standard deviation for α , β and σ is 44.61 ± 0.02 , 3.16 ± 0.10 and 0.30 ± 0.01 respectively when IS ratios are computed for model 0. It demonstrates that the results have not changed significantly comparing to when sample selection effects, thus IS ratios, are not taken into account (see Table 3 for model 0).

The second approach for taking into account selection effects, using a Gaussian mixture to approximate the joint posterior probability distribution sampled by the LIRA MCMC output, is currently still being implemented (Ebrahimpour et al. 2019).

3.5 Discussion

We have shown that the data considered, generally, does not support the existence of deviations from the standard modelling of the $L_X - T$ cluster and group scaling relation, as a single power-law, with a time-evolving normalization that follows the self-similar expectation and a non-evolving scatter. In particular, among the extra features we considered could be added to this standard model, the existence of a change-point in the $L_X - T$ relation, marking the transition between what would be the group and cluster regimes, is that which seems less justified given the data.

It has been shown by Maughan 2014 that the exponent of the $L_X - T$ scaling relation, β , can be written as the ratio of the exponents of the $L_X - M$ and $T - M$ scaling relations, where M stands for mass, respectively β_{LM} and β_{TM} . These can be thought as more fundamental, given that the presence of mass is the most determining factor in the assembly of both groups and clusters, as the result of gravitational instability. In fact,

under the self-similar assumption (Kaiser 1986), groups and clusters are assumed to be structures that virialize under the influence of gravity alone. This leads to a self-similar expectation of $\beta_{TM} = 2/3$. The derivation of the self-similar expectation for β_{LM} is more complicated, and (usually) implies assuming that the bolometric luminosity is the result of bremsstrahlung emission only, as well as that both the total amount of gas and the way it is distributed do not depend on the mass of the structures being considered. These assumptions then lead to the self-similar expectation $\beta_{LM} = 4/3$, and hence $\beta = 2$. However, the contribution of line emission to the bolometric luminosity in structures with temperatures below about 2 keV cannot be neglected, and thus the self-similar expectation for both β_{LM} and β should be different in this temperature regime. Indeed, we have confirmed that the ratio between bremsstrahlung and total emission increases roughly with the square root of the temperature for all systems with temperatures below 2 keV in our sample, as proposed by Zou et al. 2016. Hence, the total bolometric luminosity is in fact proportional to its bremsstrahlung component divided by the square root of the plasma temperature. The self-similar expectations for β_{LM} and β then become, respectively, 1 and 3/2, for structures with temperatures below about 2 keV (see also Sun 2012). The self-similar expectation is thus of a steepening of the $L_X - T$ relation when moving from the group to the cluster regime.

However, most analysis suggest the values for β_{LM} , β_{TM} and β predicted under the self-similar assumption are disfavoured given the available cluster data (e.g. see review by Giodini et al. 2013). This is most probably the result of radiative cooling and AGN feedback acting alongside gravity during the assembly and subsequent evolution of groups and clusters, as shown by simulations (e.g. Short et al. 2013; Planelles et al. 2014; Barnes et al. 2017; Le Brun et al. 2017; Truong et al. 2018). Both effects seem to lead to an increase in both β_{LM} and β , relatively to the self-similar expectation, with β_{TM} decreasing proportionally less. If these effects were more important, with respect to those induced by gravity, in groups than in clusters, they would contribute to erasing some of the difference between the group and cluster self-similar expectations for β_{LM} and β . Nevertheless, it would be somewhat surprising that, by coincidence, the stronger effects of radiative cooling and AGN feedback in groups versus clusters, would just compensate a completely unrelated effect due to the temperature dependence of line emission. Therefore, it is expectable a priori the existence of a change in the exponent of the $L_X - T$ scaling relation, when moving from the group to the cluster regime, albeit possibly weaker than the selfsimilar expectation.

The fact that we do not find statistical support for the existence of a change in the exponent of the $L_X - T$ scaling relation, could be due to its real absence or to a lack of sufficient information in the data to allow for its unambiguous identification. Nevertheless, if we indeed assume a priori a change-point in the $L_X - T$ relation to exist, as in models 3, 5, 6 and 7, we do find concordance with the self-similar expectation of an increase in the expected value for the exponent of the $L_X - T$ relation when moving from the group to the cluster regime. In the context of those models, the expected value for the $L_X - T$ relation exponent in the group regime, around 1.1, is even smaller than the self-similar expectation, although the difference is not statistically significant. On the other hand, the large positive difference between the expected value found here for the exponent in the cluster regime, close to 3.3, and the self-similar expectation, is statistically robust and in agreement with most previous results (e.g. Mantz et al. 2010a; Mittal et al. 2011; Maughan et al. 2012; Takey et al. 2013; Maughan 2014; Giles et al. 2016; Mantz et al. 2016). It seems to be essentially the result of a steeper than self-similar $L_X - M$ relation (Maughan 2014; Mantz et al. 2016), although there could be also some contribution from a shallower than self-similar $T - M$ relation (Maughan 2014; Lieu et al. 2016; Mantz et al. 2016). A smaller value for the exponent in the group regime, with respect to the self-similar expectation, together with a higher value in the cluster regime, could indicate that either the $L_X - M$ relation is shallower, or the $T - M$ relation is steeper, in the case of groups and again with respect to the self-similar expectation. For example, assuming the $T - M$ relation to be similar for both groups and clusters would imply that the exponent of the $L_X - M$ relation for groups would be close to the ratio between the expected values for the exponents of the $L_X - T$ relation for groups (1.1) and clusters (3.3) times the exponent of the $L_X - M$ relation for clusters. Assuming a exponent around 1.6 for the cluster $L_X - M$ relation (Maughan 2014), which is 20% higher than the self-similar expectation, would then imply the expected value for the exponent of the group $L_X - M$ relation would be 0.5, clearly shallower than the self-similar expectation, although compatible within the statistical uncertainties.

The expected values we obtain for the exponent of the $L_X - T$ relation in the group regime, in the context of models 3, 5, 6 and 7, are significantly smaller than what was found by Bharadwaj et al. 2015, who updated the analysis performed by Eckmiller et al. 2011, and Zou et al. 2016, but compatible within the statistical uncertainties. In Table 6 we compare the expected values found for the exponent of the $L_X - T$ relation by Bharadwaj et al. 2015 and Zou et al. 2016 with those obtained here assuming models 3, 5, 6 and 7. We also present the expected values obtained when models 0u, 0, 1u, 1, 2 and 4 are applied to each of the sub-samples that result from splitting the objects in our

full sample according to whether their observed (most probable) temperature is lower or higher than either 2 or 3 keV . In the case of the former there are 55 systems below the threshold, with 298 above, while for the later there are 142 systems below the threshold, and 211 above. We chose the 2 keV threshold because it is close to the expected value of T_{break} found for most of the models considered, but they cannot be directly compared. A hard selection threshold in observed temperature translates into a probabilistic boundary with respect to the model-inferred expected temperatures (and vice-versa), in the sense that some systems that end up in the e.g. lower observed temperature sub-sample actually have model-inferred expected temperatures higher than the value associated with the threshold (and vice-versa). The alternative 3 keV threshold was chosen because it leads to an observed temperature range, for the lower temperature sub-sample, similar to that associated with the group samples considered in Bharadwaj et al. 2015 and Zou et al. 2016. In these analysis, a model for the $L_X - T$ relation that is equivalent to our model 0u was assumed, and the expected values for the parameters of such model, θ_0 , were then determined, both with and without trying to take sample selection effects into account.

The expected values we obtain for the exponent of the $L_X - T$ relation in the case of model 0u, are only 0.12 ($T < 2 keV$) and 0.35 ($T < 3 keV$), for the lower temperature sub-samples, raising to 2.73 ($T > 2 keV$) and 2.61 ($T > 3 keV$), for the higher temperature sub-samples. The former are significantly smaller than the exponents obtained by Bharadwaj et al. 2015 and Zou et al. 2016, respectively 2.17 and 2.80, which suggests that the selection effects acting on the low temperature end of our sample are significantly different from those acting on their group samples. Such effects could be, for example, the result of a different ratio of strong cool core versus non-cool core (Bharadwaj et al. 2015; Giles et al. 2016; Zou et al. 2016), or relaxed vs. non-relaxed (e.g. Chon & Böhringer 2017), systems among the samples. Allowing for deviations from self-similar redshift evolution (model 1u) does not change significantly the expected values for the exponent of the $L_X - T$ relation, for both temperature sub-samples. If we now attempt to correct for sample selection effects, by assuming a Gaussian prior distribution for our covariate, $\log(T/5)$, we then find for the exponent expected values, in the case of model 0, 2.55 ($T < 2 keV$) and 2.05 ($T < 3 keV$), and in the case of model 1, 1.84 ($T < 2 keV$) and 1.68 ($T < 3 keV$), for the lower temperature sub-samples. These values rise to 3.20 ($T > 2 keV$) and 3.06 ($T > 3 keV$), in the case of model 0, 3.20 ($T > 2 keV$) and 3.08 ($T > 3 keV$), in the case of model 1, for the higher temperature sub-samples. Therefore, as expected given the nature of the selection effects acting on group and cluster samples, taking these into account always leads to a

significant increase in the expected value for the exponent of the $L_X - T$ relation, as also found by Bharadwaj et al. 2015 and Zou et al. 2016. This increase, as well as the larger associated uncertainty, is enough to bring our results for the $T < 2 \text{ keV}$ (but not for $T < 3 \text{ keV}$) sub-sample, in the context of model 0, to agreement with those obtained by Bharadwaj et al. 2015 and Zou et al. 2016. The same conclusions are reached in the context of models which allow for possible redshift evolution beyond the self-similar expectation (models 1 and 4) and/or in the scatter (models 2 and 4), although the expected values for the exponent of the $L_X - T$ relation become significantly smaller assuming models 1 and 4 than those inferred in the context of model 0. This difference seems to be the result of a preference, in the case of low temperature systems, for very strong positive redshift evolution beyond what is expected in the self-similar model.

The separate modelling of scaling relations for groups and clusters implicitly assumes that these two types of objects constitute easily distinguishable populations, with minimum overlap with respect to their properties and/or clearly different formation and evolution paths. However, what we know about the process of large-scale structure formation and evolution suggests the same physical phenomena were responsible for the assembly of groups and clusters, though with varying relative influence across what would then constitute a single population. Consequently, scaling relations that jointly describe the dependencies between group and cluster properties should not show discontinuities. The simplest joint scaling relation model which ensures this just includes a change-point. It allows for a possible transition, in the way the properties being considered relate to each other, when moving from the group to the cluster regime, i.e from low to high temperature in the case of the $L_X - T$ relation. Interestingly, when this joint modelling is performed, by applying models 3 and 6 to the full sample, in contrast to applying models, respectively, 0 and 2, independently to the two temperature sub-samples, a widening of the difference between the $L_X - T$ relation exponent in the group (low temperature) and cluster (high temperature) regimes is seen. All these models assume self-similar evolution for the normalisation of the $L_X - T$ relation. When this assumption is relaxed, the same trend is found, although weaker. This can be seen by comparing the results of applying models 5 and 7 to the full sample, versus the results of applying models, respectively, 1 and 4 independently to the two temperature sub-samples.

Table 6. The exponent of the $L_X - T$ relation estimated by Bharadwaj et al. 2015 and Zou et al. 2016 (usin $Y | X$ g regres-
sion, private communication), compared with the expected values for the exponent we obtain assuming different models
and samples.

Sample selection effects not considered	
Bharadwaj et al. (2015)	2.17 ± 0.26
Zou et al. (2016)	2.80 ± 0.22
Model $0u$ (for $T < 2$ keV)	0.12 ± 0.10
Model $0u$ (for $T > 2$ keV)	2.73 ± 0.11
Model $0u$ (for $T < 3$ keV)	0.35 ± 0.16
Model $0u$ (for $T > 3$ keV)	2.61 ± 0.14
Model $1u$ (for $T < 2$ keV)	0.05 ± 0.07
Model $1u$ (for $T > 2$ keV)	2.71 ± 0.11
Model $1u$ (for $T < 3$ keV)	0.26 ± 0.13
Model $1u$ (for $T > 3$ keV)	2.62 ± 0.14
Sample selection effects taken into account	
Bharadwaj et al. (2015)	3.20 ± 0.26
Zou et al. (2016)	3.30 ± 0.20
Model 0 (for $T < 2$ keV)	2.55 ± 0.55
Model 0 (for $T > 2$ keV)	3.20 ± 0.11
Model 0 (for $T < 3$ keV)	2.05 ± 0.29
Model 0 (for $T > 3$ keV)	3.06 ± 0.14
Model 3 (for all T , β_{break})	1.10 ± 1.31
Model 3 (for all T , β)	3.27 ± 0.12
Model 1 (for $T < 2$ keV)	1.84 ± 0.70
Model 1 (for $T > 2$ keV)	3.20 ± 0.11
Model 1 (for $T < 3$ keV)	1.68 ± 0.29
Model 1 (for $T > 3$ keV)	3.08 ± 0.14
Model 5 (for all T , β_{break})	1.19 ± 1.17
Model 5 (for all T , β)	3.27 ± 0.14
Model 2 (for $T < 2$ keV)	2.52 ± 0.60
Model 2 (for $T > 2$ keV)	3.19 ± 0.11
Model 2 (for $T < 3$ keV)	2.06 ± 0.30
Model 2 (for $T > 3$ keV)	3.06 ± 0.14
Model 6 (for all T , β_{break})	1.00 ± 1.41
Model 6 (for all T , β)	3.26 ± 0.14
Model 4 (for $T < 2$ keV)	1.97 ± 0.62
Model 4 (for $T > 2$ keV)	3.18 ± 0.11
Model 4 (for $T < 3$ keV)	1.69 ± 0.28
Model 4 (for $T > 3$ keV)	3.07 ± 0.15
Model 7 (for all T , β_{break})	1.20 ± 1.23
Model 7 (for all T , β)	3.26 ± 0.13

We also explored the possibility of the change in exponent of the $L_X - T$ relation being driven by our prior assumptions that the scatter about the $L_X - T$ relation does not change across the sample temperature range, and a possible change in exponent occurs abruptly. Allowing for the scatter to change abruptly at the same temperature as the exponent, thus introducing an extra free parameter with respect to model 7, does not lead to a significant change in the expected value for the exponent in the cluster regime. But it leaves the exponent in the group regime essentially unconstrained, although with an expected value even smaller than what was obtained assuming model 7. While allowing for a transition region with respect to both the exponent and scatter around the

changepoint, hence introducing yet another extra free parameter, the width of the transition region, we were unable to ascertain convergence for the MCMC produced by Gibbs sampling. This is probably due to a strong degeneracy between this parameter and one or more of the other parameters. Allowing for such a transition region should lead to a decrease in the expectation value for β , because systems with temperatures lower than that associated with the change-point would now also contribute to the estimation of the regression parameters for the higher temperature systems, and an increase in the expectation value for β_{break} , due to the reciprocal effect. More data will be necessary to enable constraints to be imposed on the parameters of such complex models.

The self-similar expectation for the redshift evolution of the $L_X - T$ relation is usually assumed to mean that its normalisation is proportional to $E(z)^{\gamma_{LT}}$ with $\gamma_{LT} = 1$. Note that given our definition of γ , we have $\gamma_{LT} = \gamma + 1$. However, Maughan 2014 has also shown that the self-similar model implies more generally that β_{LM}

$$\gamma_{LT} = \gamma_{LM} - \frac{\beta_{LM}}{\beta_{TM}} \gamma_{TM} \quad (57)$$

$$= \gamma_{LM} - \beta \gamma_{TM} \quad (58)$$

where $E(z)^{\gamma_{LM}}$ and $E(z)^{\gamma_{TM}}$ are assumed to describe the redshift evolution of the normalisation of the $L_X - M$ and $T - M$ scaling relations, respectively. The self-similar expectation for γ_{LT} is then recovered if γ_{LM} , β and γ_{TM} equal their self-similar values of 7/3, 2 and 2/3, respectively. However, as discussed before, this will only hold true in the cluster regime, given that in the group regime the self-similar expectation for β is 3/2, and hence for γ_{LT} is 4/3. Further, if β deviates from its self-similar expectation, as we have found seems to happen both in the group and cluster regimes, the same will happen for γ_{LT} even assuming self-similar evolution for the $L_X - M$ and $T - M$ relations. Given that our results suggest values for β in the group and cluster regimes of 1.1 ± 1.3 and 3.3 ± 0.1 , respectively, the self-similar expectation would then suggest values for γ_{LT} in the group and cluster regimes of 1.6 ± 0.9 and 0.1 ± 0.1 , respectively. But, if groups and clusters are assumed to belong to a single population of objects, distributed in the space of their properties in a continuous manner, then γ_{LT} cannot change abruptly for any value of those properties, namely temperature or X-ray luminosity, unless the location of the respective change-point also changes (continuously) with redshift. Unfortunately, LIRA does not allow for this last option. Consistency with Equation (58), then implies that modelling the $L_X - T$ relation using a change-point through LIRA in effect imposes different evolution rates with redshift for the normalisation of the $L_X - M$ and/or $T - M$

scaling relations in the group versus the cluster regimes, and thus the location of the change-point for the differentially evolving scaling relation(s) among these to change with redshift.

However, when models 1 and 4 are applied to the low temperature sub-samples the expected values for γ_{LT} that are obtained are, respectively, 4.7 ± 2.1 and 4.4 ± 2.1 for $T < 2 \text{ keV}$, as well as 6.7 ± 1.0 for both models in the case of $T < 3 \text{ keV}$. When the same models are applied to the high temperature sub-samples those values are just, respectively, 0.3 ± 0.5 and 0.4 ± 0.5 for $T > 2 \text{ keV}$, as well as -0.8 ± 0.6 and -0.9 ± 0.6 for $T > 3 \text{ keV}$. This indicates that if a change-point is introduced in the modelling of the $L_X - T$ relation, then its location should be allowed to change with redshift. In particular, a significantly faster evolving normalisation of the $L_X - T$ relation in the group regime suggests the value for the temperature associated with its change-point is increasing with redshift. The inclusion of an extra parameter in LIRA to account for a possible evolution with redshift of the location of the changepoint would thus be an interesting avenue to pursue, something we expect to do in future work.

The expected values obtained for γ_{LT} with respect to the low temperature sub-samples are significantly higher than the selfsimilar expectation according to Equation (58), i.e. 1.6 ± 0.9 . But statistically indistinguishable, with respect to 0.1 ± 0.1 , for the high temperature sub-samples. A larger positive deviation with respect to the expected self-similar evolution for the $L_X - T$ relation in the cluster regime, in accordance with Equation (58), was found by Giles et al. 2016. On the other hand, Reichert et al. 2011 and Clerc et al. 2014, as well as Hilton et al. 2012 using a distinct XCS sample, found negative evolution in the cluster $L_X - T$ relation with respect to self-similar. However, the results of Reichert et al. 2011 and Clerc et al. 2014 could be explained by the difficulty of fully taking into account the distinct selection biases acting on the different cluster samples, spanning different redshift intervals, they used to determine the evolution of cluster properties, as extensively discussed in Giles et al. 2016 with respect to the results obtained by Clerc et al. 2014. The reasons driving the differences between the results presented here and those described in Hilton et al. 2012 are discussed in the Appendix. The redshift evolution in the scatter of the $L_X - T$ relation has been much less studied, but the results of Mantz et al. 2016 are in agreement with the negative evolution suggested by the data we used.

The most recent numerical simulations, which include additional energy injection into the ICM/ICG from star formation and AGN feedback, as well as metal-dependent radiative cooling, are able to reproduce the observed $L_X - T$ relation for clusters at low redshift,

including a larger exponent than the self-similar expectation (e.g. Short et al. 2013; Planelles et al. 2014; Barnes et al. 2017; Le Brun et al. 2017; Truong et al. 2018). In agreement with our modelling assumption, Truong et al. 2018, Barnes et al. 2017 and Le Brun et al. 2017 also found that, up to a redshift of 0.6, the exponent remained constant within the uncertainties. Further, their results for the evolution with redshift of the amplitude and scatter of the $L_X - T$ relation, up to a redshift of 0.6, are in agreement with our results, supporting an increase in the amplitude above the self-similar expectation and a decrease in the scatter. Interestingly, Le Brun et al. 2017 also tried to model the $L_X - T$ relation using a broken power-law, inserting a fixed change-point at 2 keV. Contrary to our results they seem to find similar exponents for the $L_X - T$ relation in the group and cluster regimes, although in agreement with our findings they see faster evolution in its amplitude in the group versus the cluster regime.

Nevertheless all the previous discussion, one should bear in mind that it was based on results that do not take into account sample selection effects.

3.6 Conclusions

We characterized the X-ray luminosity–temperature ($L_X - T$) relation using a sample of 353 clusters and groups of galaxies with temperatures in excess of 1 keV and redshifts in the range $0.1 < z < 0.6$. We did not find strong statistical support for deviations from the usual modelling of the $L_X - T$ relation as a single power-law, where the normalisation evolves self-similarly and the scatter remains constant with time. Nevertheless, assuming a priori the existence of the type of deviations considered, then faster evolution than the self-similar expectation for the normalisation of the $L_X - T$ relation seems favoured, eventually signalling an increase with redshift in the average group/cluster X-ray luminosity at fixed temperature, as well as a decrease with redshift in the scatter about the $L_X - T$ relation. Further, the preferred location for a possible temperature dependent change-point in the exponent of the $L_X - T$ relation seems to be around 2 keV, possibly marking the transition between the group and cluster regimes. Our results also suggest an increase in the power-law exponent of the $L_X - T$ relation when moving from the group to the cluster regime, and faster evolution in the former with respect to the later, driving the temperature dependent changepoint towards higher values with redshift. We confirm past results that found an exponent for the $L_X - T$ relation in the cluster regime significantly steeper than the self-similar expectation. But find the data suggests a shallower exponent with respect to it in the group regime. This implies that

either the $L_X - M$ relation is shallower, or the $T - M$ relation is steeper, in the case of groups and again with respect to the self-similar expectation.

The absence of statistical support for the existence of deviations from the standard modelling of the $L_X - T$ cluster and group scaling relation as a single power-law, with a time-evolving normalization that follows the self-similar expectation and a nonevolving scatter, is somewhat of a surprise. In particular, because the self-similar approach actually predicts the $L_X - T$ relation to deviate from a single power-law towards the lower temperature regime. This deviation could be weakened by non-gravitational effects, namely radiative cooling and AGN feedback, acting more strongly in groups versus clusters. However, if the impact of nongravitational effects depends on the depth of the gravitational potential of the system they are acting on, then these effects should also leave an imprint on the redshift evolution of groups and clusters, because, at fixed mass, haloes are denser at higher redshifts. This stands in contrast with the absence of statistical support for the existence of deviations from the self-similar expectation for the redshift evolution of the $L_X - T$ relation.

However, these conclusions are based on results that do not take into account sample selection effects. Considering these may change some of the conclusions, particularly those with respect to the behaviour of the $L_X - T$ relation in the group regime.

Hopefully, with the expected increased availability of large group and cluster samples, like those to be obtained from the Dark Energy Survey (DES, Abbott et al. 2016) and eROSITA (Predehl et al. 2014), we will reach a stage where enough information will be available in the data to allow for the statistical detection of possible deviations from the standard modelling of the $L_X - T$ relation. In particular, it would be important to expand the current group and cluster samples towards lower temperatures and higher redshift, in order to gain more leverage with respect to the exponent and evolution of the normalisation of the $L_X - T$ relation in the group regime. In the meantime, the prior assumption of more complex models than the standard one does not seem warranted.

The work described in this chapter has been submitted for publication in Monthly Notices of the Royal Astronomical Society and is currently available in <https://arxiv.org/abs/1805.03465>.

Chapter 4

هر کجا هستم باشم،
آسمان مال من است...

سهراب سپهری

Wherever I am,
Let me be,
The sky is mine...

Sohrab Sepehri

Mass estimation of galaxy clusters

4.1 Introduction

Gaussian Processes (GPs), a non-parametric statistical method explained in section 2.4, are used to predict future data based on the current state of the observed data. Clusters at different epochs may not obey the same scaling relations, which may also change as a function of cluster properties. Therefore, using these to predict cluster mass may require the flexibility of a non-parametric model like GPs.

In this section, we will describe the results obtained by applying GPs on three catalogues used by Sereno & Ettori 2017 (hereafter CoMaLit V). We have estimated the mass of galaxy clusters within those catalogues using their optical, SZ or X-ray properties. In each catalogue, those clusters with available Weak Lensing (WL) mass were used as training data to estimate the mass of the remaining clusters. Then, the estimated masses were compared to the masses that were obtained in CoMaLit V. The mass forecast in CoMaLit V for three mentioned catalogues are available [here](#). The results in CoMaLit V were estimated through applying LIRA, described in section 3.3.1, in which the intrinsic distribution of the independent variable is approximated using a mixture of Gaussian distributions. Hence, in GPs we also included a Gaussian mixture in the likelihood , to

mimic a selection function and to be able to compare our results to CoMaLit V predictions.

This is the first time that GPs are used to predict the mass of galaxy clusters. Our goal is to determine whether there is evidence that the traditional power-law modelling of scaling relations should be substituted by a more flexible, non-parametric model, like GPs. We will achieve this by comparing our results to those presented in CoMaLit V.

4.2 The datasets

4.2.1 The Planck SZ catalogue

PSZ2, as the second Planck Catalogue of Sunyaev-Zel'dovich Sources, is the largest SZ catalogue and the deepest all-sky catalogue of galaxy clusters yet produced. From 1653 sources in this catalogue which have been detected with a signal to noise ratio higher than 4.5 during the mission of 29 months, 1203 sources have been confirmed as clusters by finding matches in optical or X-ray samples or by further follow-ups. (Planck Collaboration 2016).

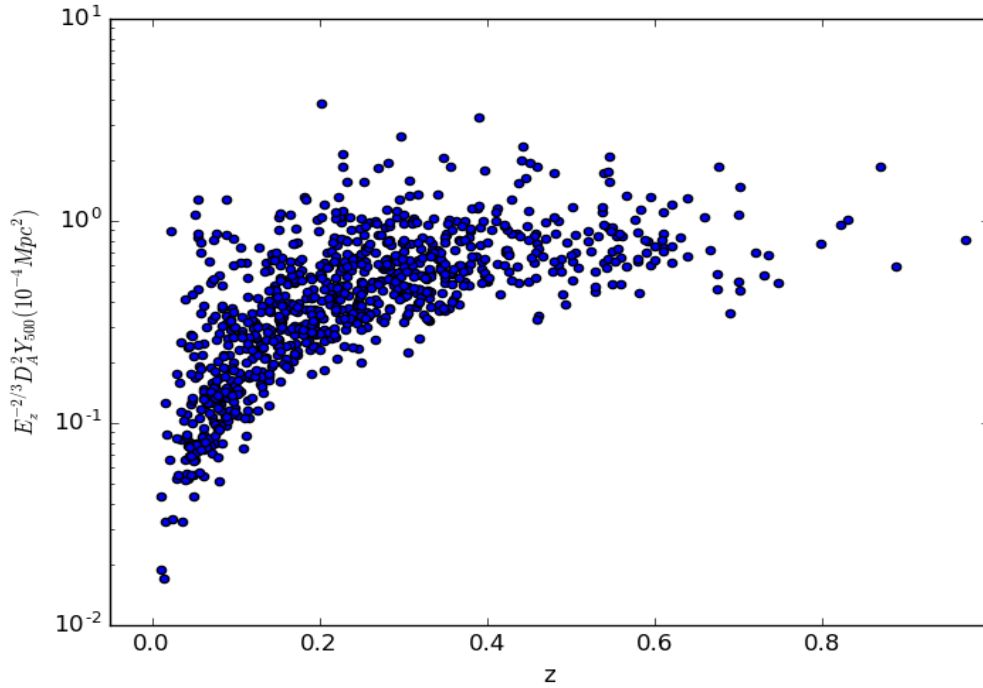


Figure 26. The self-similar scaled comptonisation parameter versus redshift for 926 clusters of PSZ2 with available redshift.

The clusters that we used are the subset of 1271 clusters identified with the Matched Multi-Filter method MMF3. Among these clusters, the redshift is available for 926 candidates. The parameter Y_{SZ} , the spherically integrated Comptonisation parameter, is

the observable for SZ surveys, which depends on the cluster gas thermal energy through its relation to the integrated electron pressure. Figure 26 shows the integrated Comptonisation parameter for those 926 clusters of PSZ2 with available redshift versus redshift. The clusters span the redshift range from $z \sim 0.01$ to 0.97 (median 0.22) and mass range from $M_{500} \sim 0.8$ to $16 \times 10^{14} M_{\odot}$ (median $\sim 5 \times 10^{14} M_{\odot}$). In PSZ2, the integrated Comptonisation parameter has been estimated within the radius $5R_{500}$, Y_{5R500} . The radius $5R_{500}$ has been selected because not only the bias for the total integrated Comptonisation estimated by it, is a few percent (almost unbiased) but also it is small enough that confusion effects from nearby structures can be neglected. In order to calculate Y_{500} we used Equation (24) of CoMaLit V. In this equation, M_{HE} is the hydrostatic equilibrium mass, which has been estimated and recorded in the PSZ2 catalogue. We calculated the self-similarity scaled integrated Comptonisation parameter as follows:

$$E_z^{-2/3} D_z^2 Y_{500} = 10^{\alpha} \left[\frac{M_{HE}}{M_{pivot}} \right]^{\beta} [10^{-4} Mpc^2], \quad (59)$$

where D_z is the angular diameter distance to the cluster, $M_{pivot} = 6 \times 10^{14} M_{\odot}$, $\alpha = -0.186 \pm 0.011$ and $\beta = 1.79 \pm 0.06$.

We also calculated the bias-corrected mass, M_{500} , using $M_{500} = M_{HE}/(1 - b)$. The mean bias $(1 - b)$ has been estimated by comparing the observed relation with predictions from several sets of available numerical simulations and the best guess for it has been considered to be 0.8 , without any clear dependency on mass or redshift (Planck Collaboration 2014). In section 4.2 the results are compared to both M_{HE} and the bias-corrected mass, M_{500} .

4.2.2 The SDSS redMaPPer catalogue

The SDSS redMaPPer (red-sequence Matched-filter Probabilistic Percolation) catalogue was compiled by applying a red-sequence cluster finder to $\sim 10000 \text{ deg}^2$ of DR8 release of the SDSS (Sloan Digital Sky Survey) (Rykoff et al. 2013). The result is a catalogue which contains ~ 26000 candidate clusters over the redshift range $0.08 < z < 0.55$. Redshifts are generally photometric but when available, spectroscopic redshifts are used. The redMaPPer optical richness λ of a cluster is the summation over the probabilities of finding galaxies close by a cluster to be members of the cluster. All the galaxies should be participated in the sum. In practice, a cutoff radius, R , which scales with richness, and a luminosity cut, should be defined to select the galaxies that are considered.

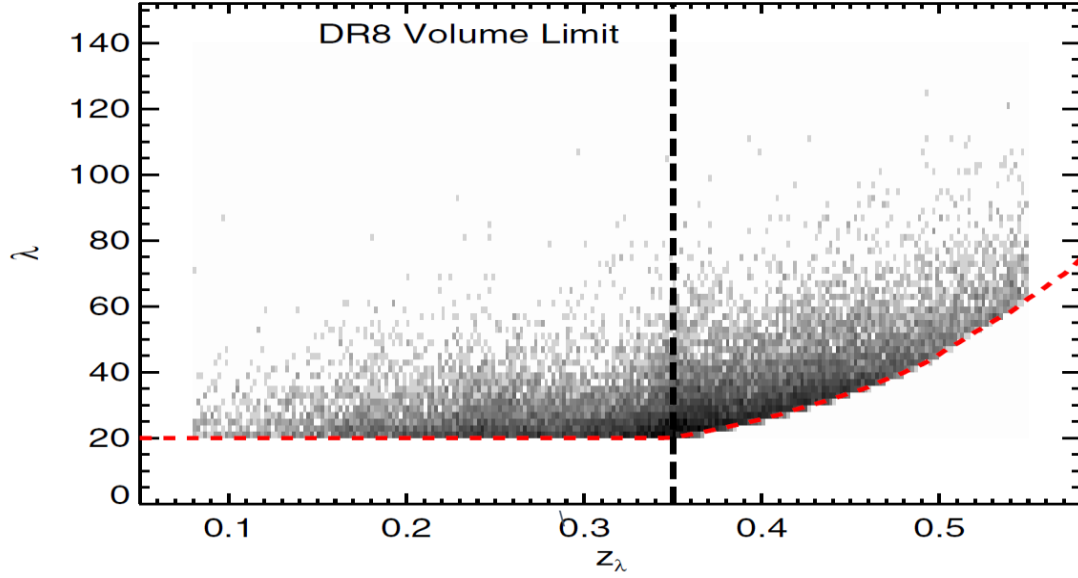


Figure 27. The richness versus redshift for clusters of the SDSS redMaPPer catalogue. Taken from Rykoff et al. 2013.

One of the parameters that is applied to construct the SDSS redMaPPer catalogue is the scale factor, $S(z)$, which is a function of the photometric redshift of the cluster. The definition of $S(z)$ requires the number of galaxies above the limiting magnitude of the survey for a cluster with richness λ to be $\lambda/S(z)$. If λ exceeds 20 times the scale factor, the clusters are included in the catalogue. It means that those clusters with at least 20 galaxy counts above the cut-off luminosity of the survey at the cluster redshift are selected (Rykoff et al. 2013). Figure 27 shows the richness λ versus redshift for the final redMaPPer catalogue. The red dashed line shows the redshift-dependent richness cut $\lambda > 20S(z)$.

4.2.3 Meta-Catalogue of X-ray detected Clusters of galaxies

The MCXC (Meta-Catalogue of X-ray detected Clusters of galaxies) gathers 1743 confirmed X-ray clusters from ROSAT All Sky Survey-based (RASS) and serendipitous cluster catalogues (Piffaretti et al. 2011). For each cluster the redshift, coordinates, the X-ray luminosity, L_{500} , which is systematically homogenized to the $[0.1-2.4] \text{ keV}$, the total mass, M_{500} , and radius, R_{500} , are available in the catalogue. Since the uncertainties are not provided, the statistical uncertainty that has been assumed on the luminosity is 10 per cent of the luminosity for each cluster as it is mentioned in CoMaLit V. As it is mentioned in Piffaretti et al. 2011 the luminosities within R_{500} and $0.1 - 2.4 \text{ keV}$ have been obtained by adopting the average gas density profile or the $L - M$ relation derived from the representative X-ray cluster sample REXCESS. The same $L - M$ was also used to compute M_{500} and R_{500} through the luminosities. Figure 28 shows the distribution of the MCXC clusters in the $L_{500} - z$ plane.

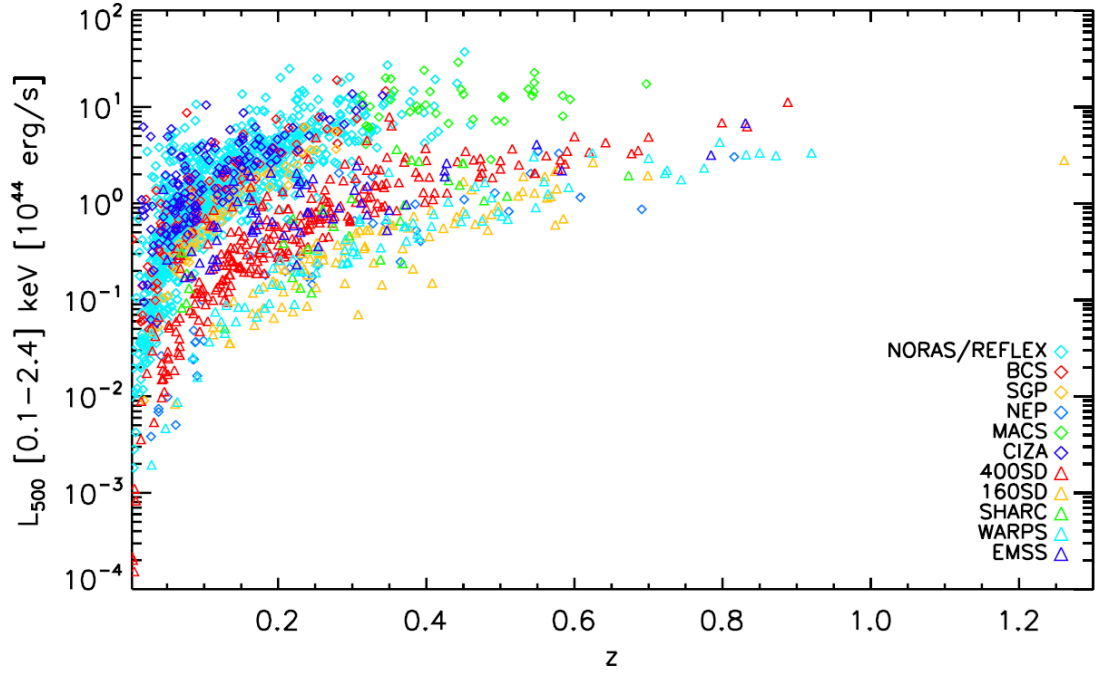


Figure 28. The luminosity versus redshift for clusters of MCXC catalogue. Taken from Piffaretti et al. 2011.

4.2.4 Weak lensing data

The WL mass of those clusters which were used as training data, came from the Literature Catalogs of Weak Lensing Clusters (LC2), presented in Sereno 2015 (hereafter CoMaLit III). WL masses of LC2 have been reported at over-densities of 2500, 500, 200, and at the virial one in the reference Λ CDM model. We used M_{500} as the mass of the clusters in the analysis.

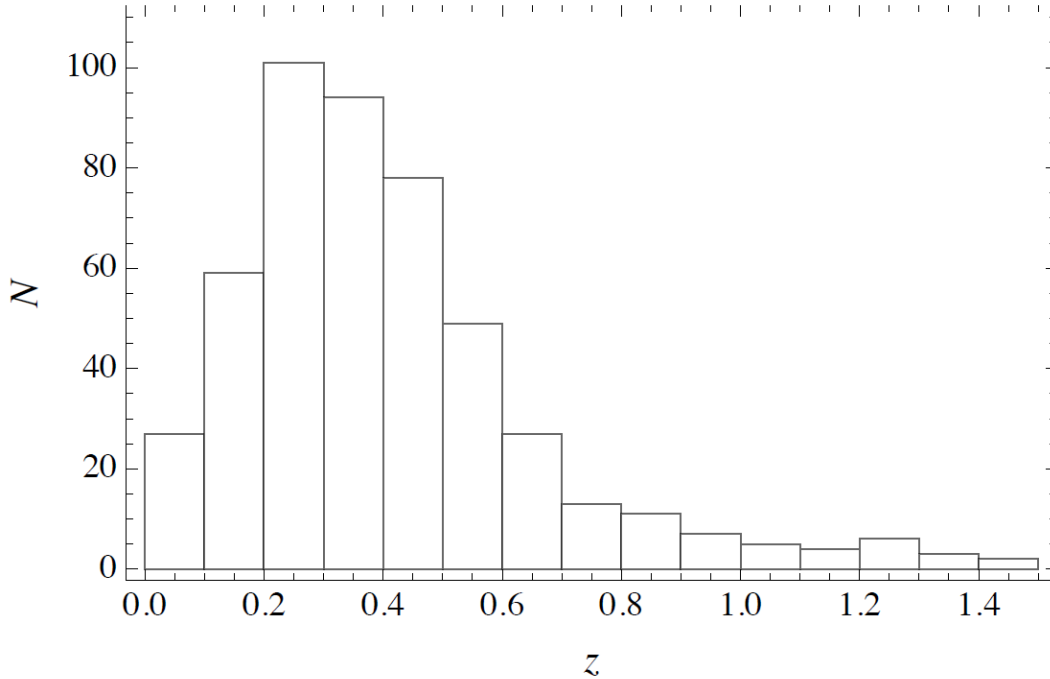


Figure 29. Redshift distribution of 485 WL clusters in the LC2-single catalogue. Taken from CoMaLit III.

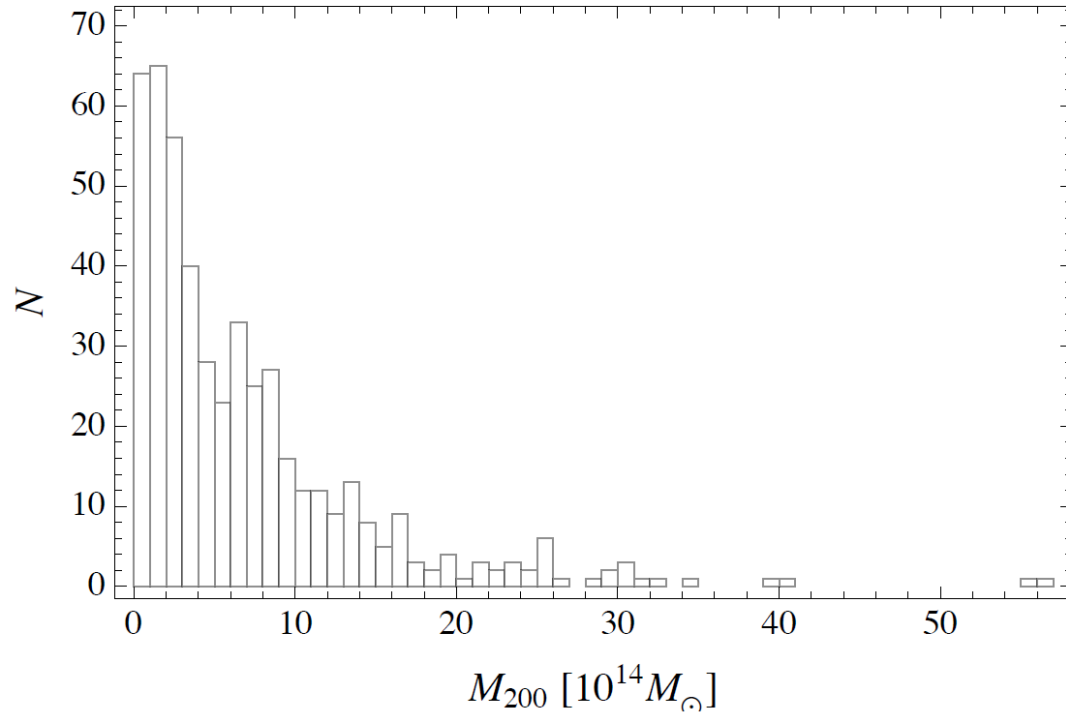


Figure 30. Mass distribution of 485 WL clusters in the LC2-single catalogue. Taken from CoMaLit III.

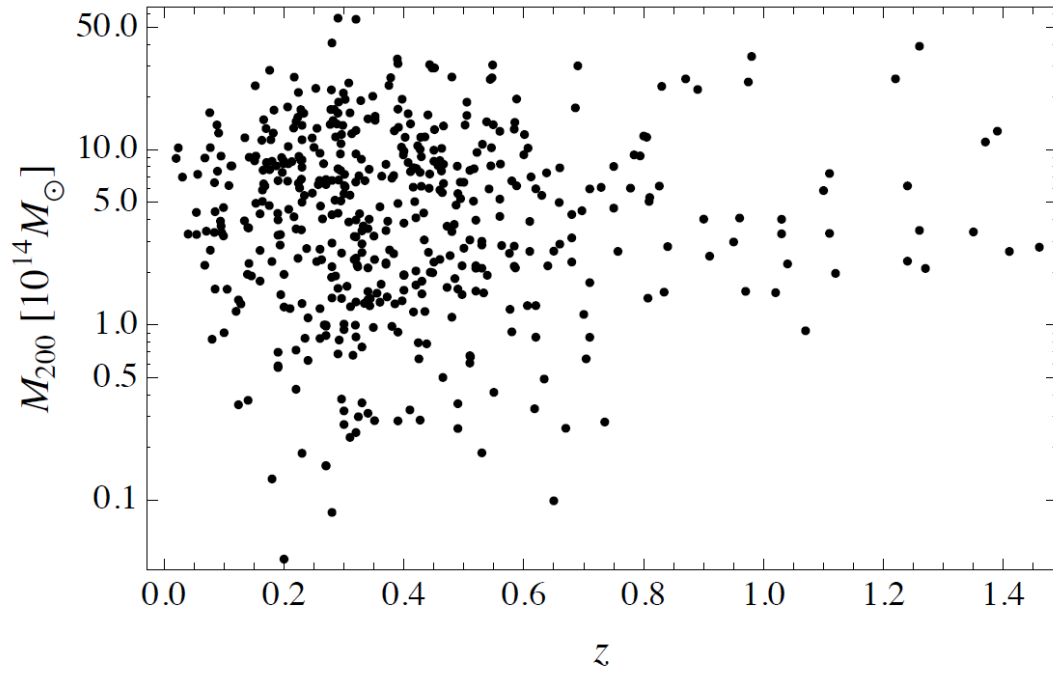


Figure 31. The distribution of 485 WL clusters in the LC2-single catalogue in a mass-redshift plane. The mass is in units of $10^{14} M_{\odot}$. Taken from CoMaLit III.

The catalogue we will be using, called LC2-single, contains 485 unique groups and clusters whose masses were determined with a single-halo modeling with no duplicate entries. The presented masses in this catalogue are the most sensible masses that can be used in order to compare to other cluster properties like the optical richness, the X-ray luminosity or the SZ flux (CoMaLit III). The redshift range for the groups and clusters in this catalogue is $0.02 - 1.46$ with a mode at $z \sim 0.35$ and 50 clusters with $z > 0.7$.

Figure 29 shows the redshift distribution of the clusters in the LC2-single catalogue. The mass distribution in the LC2-single catalogue is shown in Figure 30. The median for the mass distribution is $M_{200} \sim 4.5 \times 10^{14} M_{\odot}$ and extends to M_{200} larger than $5 \times 10^{15} M_{\odot}$. From Figure 31 it seems that there is no correlation between cluster masses with redshift and the reason is the heterogeneous nature of the sample (CoMaLit III).

4.3 Results and discussion

In section 2.4 we discussed Gaussian Processes (GPs), a non-parametric Bayesian statistics approach that enables us to predict the true values for a given quantity, mass in our case. We applied this algorithm to each of the three catalogues that we described about in sections 4.1.1 to 4.1.3, to predict the mass of the clusters therein, using the WL masses in section 4.1.4. We used the package `gptools` in Python to implement GPs, which is available both at PyPi⁸ and GitHub⁹.

Among the 926 candidates with available redshift in PSZ2, there are 135 clusters whose WL masses are available through LC2 (CoMaLit III). We used the integrated comptonisation parameters (Y_{500}), the WL masses and the associate uncertainties of these 135 clusters (as training points) to predict the mass of 676 clusters whose integrated comptonisation parameter are within the range of those of training data (as test points). Figure 32 shows the predicted mass for those 676 clusters of PSZ2 versus the integrated comptonisation parameter.

We compare our results (in blue) to what has obtained in CoMaLit V (in green) and in PSZ2 (in red). The full result of the comparison is available [here](#). The first 10 entries of the catalogue are presented in Table 7.

In general, it is clear that GPs and CoMaLit V predictions are in agreement with each other although it seems that in higher fluxes GPs tend to overestimate mass with respect to CoMaLit V masses. However, they both are higher than Planck predictions.

⁸ <https://pypi.python.org/pypi/gptools/>

⁹ <https://github.com/markchil/gptools/>

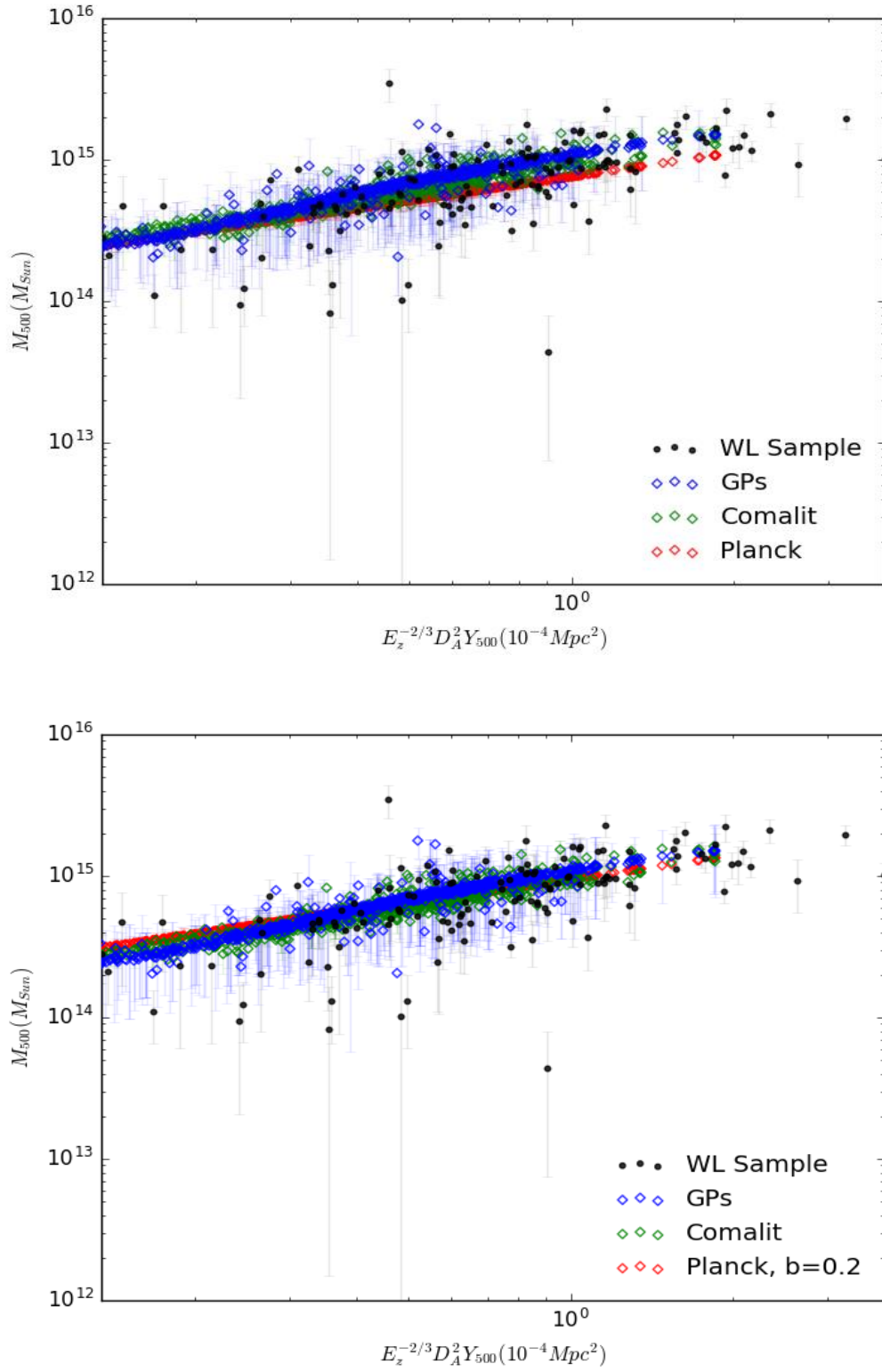


Figure 32. The predicted mass for 676 clusters of Planck-SZ2 (blue dots) using the integrated Compton parameter and WL mass of 135 clusters with available WL mass in LC2 (black dots) versus the integrated Compton parameter. The green points show the estimated mass in CoMaLit V and the red points show the estimated mass(upper panel) or estimated biased mass (lower panel) by Planck collaboration.

The upper panel in Figure 32 includes Planck mass estimates while the lower panel shows the bias-corrected mass estimated by Planck collaboration. Since the bias predicts higher values of mass, the bias-corrected mass is compatible with CoMaLit V results at higher SZ fluxes while overestimates mass for clusters with lower SZ flux.

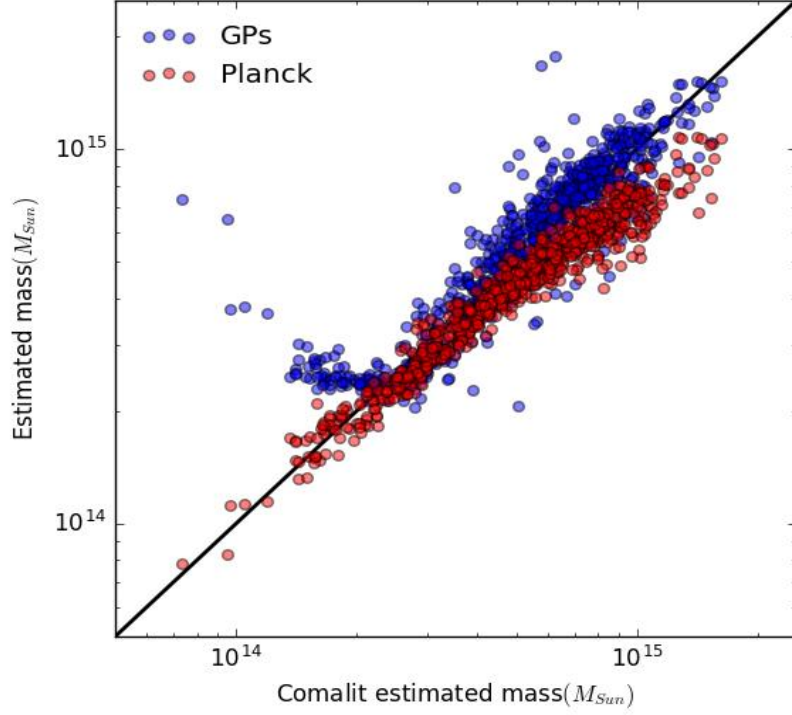


Figure 33. Comparison of the predicted masses for 676 clusters of Planck-SZ2 by GPs (blue circles) and Planck collaboration (red circles) to the masses estimated in CoMaLit V.

Table 7. The first 10 entries of the catalogue of the comparison of GPs masses to CoMaLit V and Planck masses, *HFI_PCCS_SZ-MMF3_R2.08.fits*. The forecast by CoMaLit V is available here. Col. 1: index. Col. 2: name of the cluster. Col. 3: GPs mass and associated error, Col. 4: CoMaLit V mass and associated error, Col. 5: Planck mass and associated error. Masses are within R_{500} and in units of $10^{14}M_{\odot}$.

Index	Name	M_{GPs}	M_{CoM}	M_{Pl}
1	PSZ2G000.04+45	6.106 ± 2.933	4.461 ± 0.523	$3.962^{+0.393}_{-0.370}$
2	PSZ2G000.13+78	6.957 ± 2.658	6.210 ± 0.480	$5.122^{+0.351}_{-0.323}$
3	PSZ2G000.77-35	10.397 ± 2.822	8.518 ± 0.875	$6.334^{+0.590}_{-0.611}$
4	PSZ2G002.04-22	2.167 ± 1.401	2.874 ± 0.738	$2.835^{+0.370}_{-0.463}$
5	PSZ2G002.82+39	4.345 ± 1.958	5.459 ± 0.491	$4.716^{+0.322}_{-0.370}$
6	PSZ2G003.21-76	4.044 ± 1.488	3.386 ± 0.867	$3.592^{+0.465}_{-0.500}$
7	PSZ2G003.91-42	7.875 ± 4.488	9.706 ± 0.973	$7.123^{+0.643}_{-0.679}$
8	PSZ2G003.93-59	14.568 ± 2.592	15.429 ± 1.588	$10.357^{+0.677}_{-0.724}$
9	PSZ2G004.04+42	5.528 ± 2.779	4.313 ± 0.799	$4.286^{+0.387}_{-0.417}$
10	PSZ2G004.13+56	16.772 ± 7.745	5.775 ± 1.193	$5.515^{+0.647}_{-0.715}$

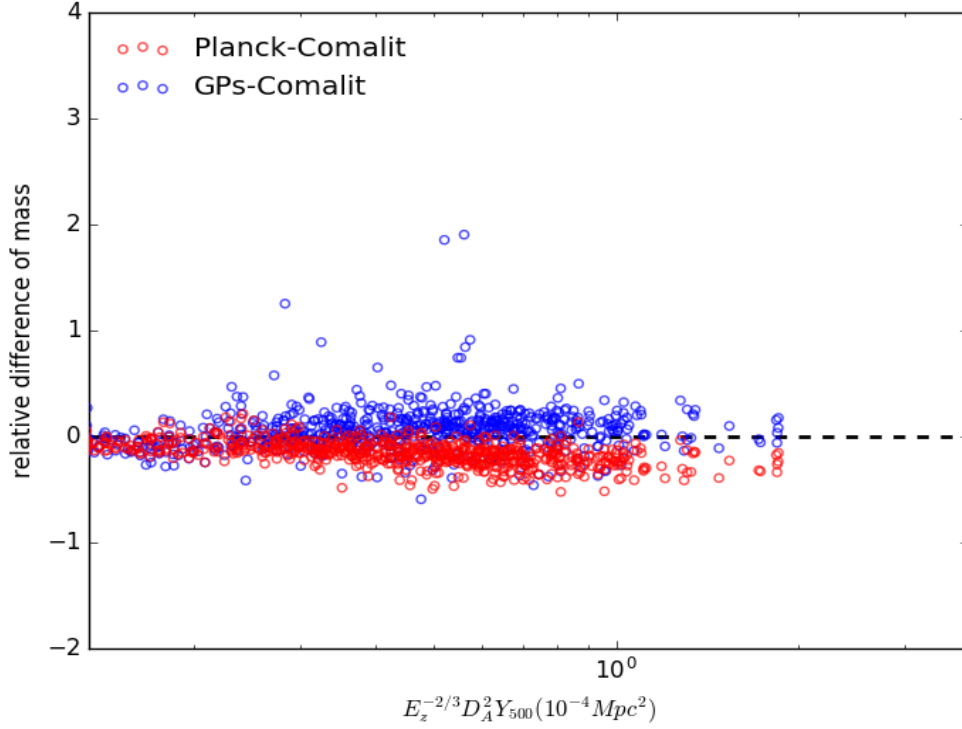


Figure 34. Relative difference of masses predicted by Planck collaboration (red circles) and GPS (blue circles) with respect to the mass estimates by CoMaLit V versus the integrated Compton parameter.

From Figures 32 and 33 one can see that GPs predictions differ for low and high SZ flux clusters. One reason could be WL data which is more scarce for lower SZ fluxes.

In Figure 33, the GPs and Planck masses are compared to what has been estimated in CoMaLit V. In this Figure, it is more obvious that while in general GPs overestimates mass with respect to CoMaLit V estimations for higher SZ flux clusters, Planck results show an underestimation of the mass with respect to CoMaLit V results for the same region. We also calculated the relative difference of mass predicted by Planck collaboration and GPs with respect to the CoMaLit V mass estimates using $relative\ difference = \frac{M_X - M_{Com}}{M_{Com}}$. Here X corresponds to Planck or GPs. The results are shown in Figure 34.

Among the clusters in the SDSS redMaPPer catalogue, the WL mass is available for 144 clusters, found by cross matching with LC2. The match has been done by finding pairs with a redshift difference less than $\Delta z = 0.05$ and a maximum angular separation of 6 arcminutes in the sky. Since the number of training data (144 clusters) whose WL mass is going to be used to estimate the mass of the remaining clusters in the catalogue is very low in comparison with the total number of the clusters (~ 26000), we first randomly

selected about 10% of clusters out of those whose richness are within the range of richness for training data in the main catalogue as test data and then used the mass and richness of those 144 clusters as training data to estimate the mass of test data given their richness.

Figure 35 shows the predicted mass for those 2496 clusters of SDSS-redMaPPer whose richnesses are within the interval of training data. We compared the estimated mass to what has been obtained in CoMaLit V. The full results of the comparison is available [here](#). The first 10 entries of the catalogue are presented in Table 8. From Figure 35 it is clear that the GPs predicted masses tend to be higher than those estimated in CoMaLit V. Figure 36 shows the comparison between GPs and CoMaLit V estimated masses. There are regions with huge uncertainties in GPs predictions, which can be because of higher uncertainties in richness of those clusters or as a result of lack of training data in those regions. This implies that the more clusters with known WL mass and the less uncertainties in observational measurements we have, the more accurate and precise is the GPs mass we predict. Figure 37 shows the relative difference of GPs predicted mass with respect to the mass estimates by CoMaLit V versus the richness.

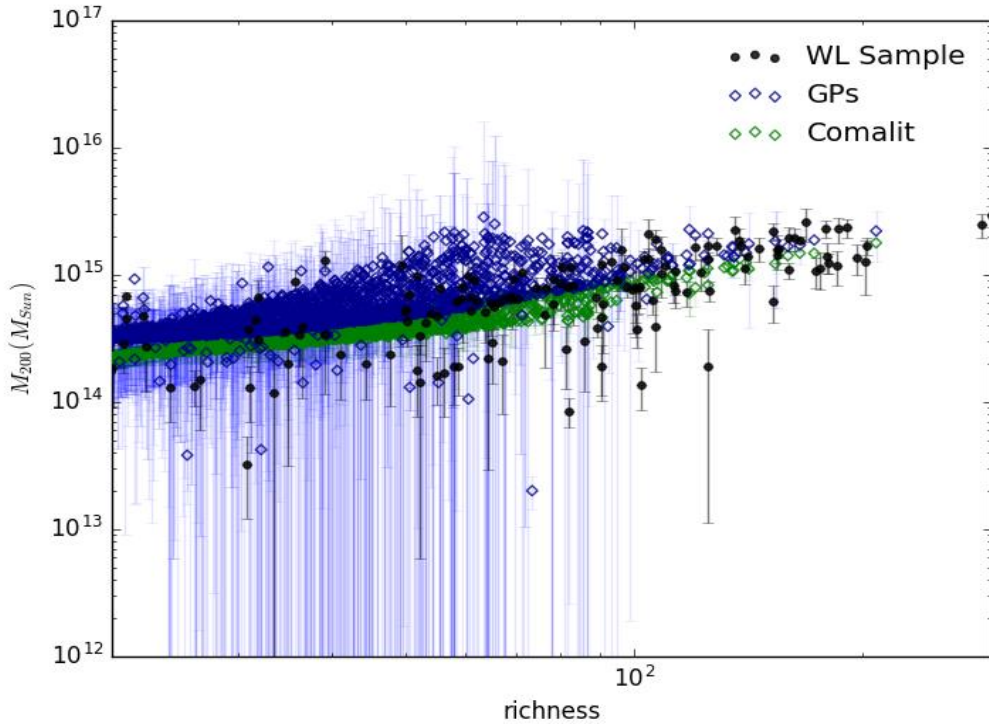


Figure 35. The predicted mass for 2496 clusters of SDSS redMaPPer (blue diamonds) using the richness and WL mass of 144 clusters with available WL mass in LC2 (black dots). The green diamonds show the estimated mass in CoMaLit V.

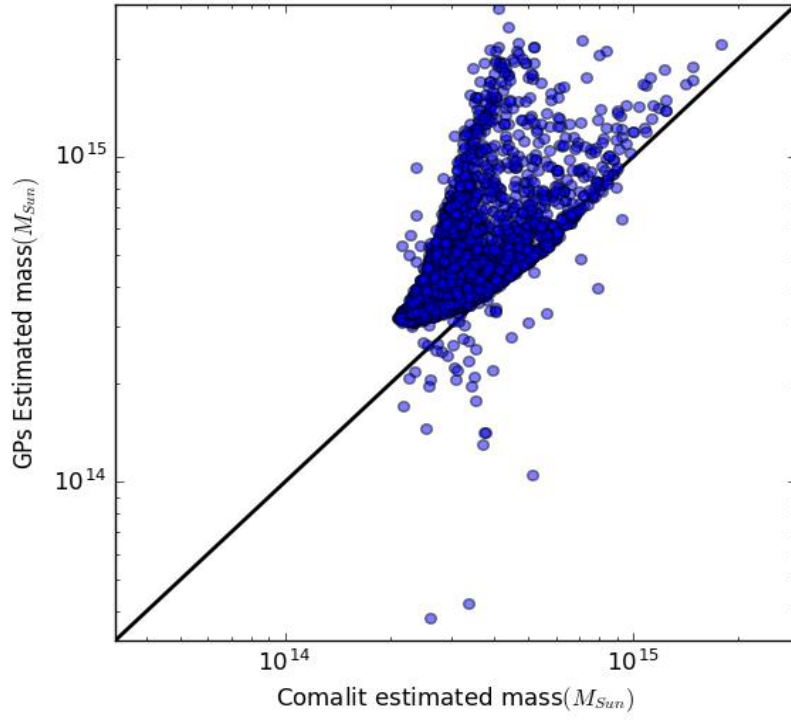


Figure 36. Comparison of the predicted masses for 2496 clusters of SDSS redMaPPer by GPs to the masses estimated in CoMaLit V.

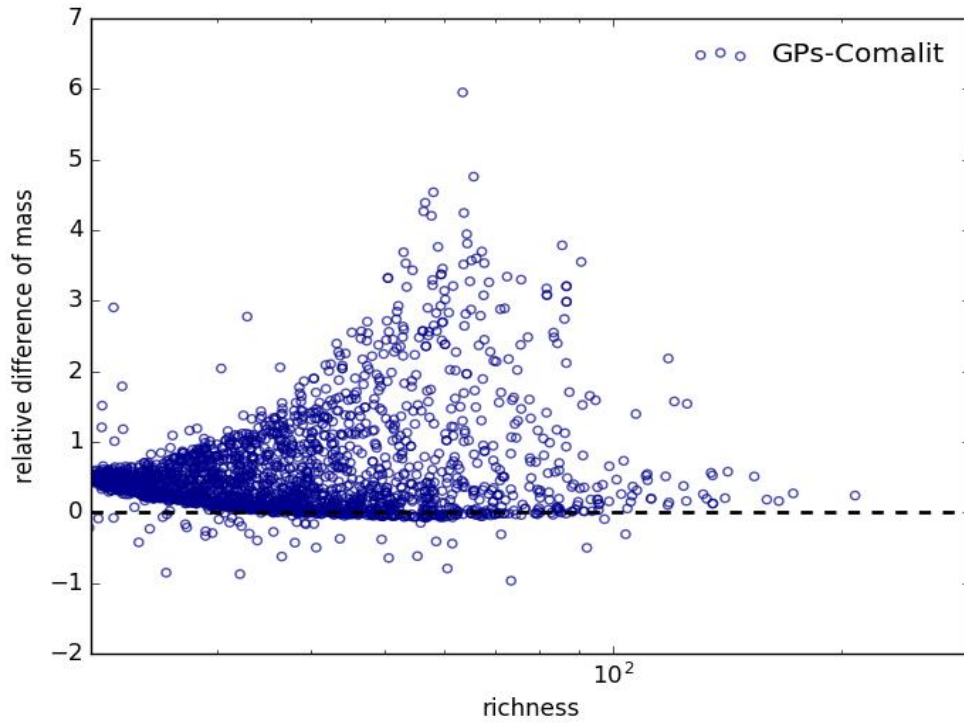


Figure 37. Relative difference of GPs masses with respect to the mass estimates by CoMaLit V versus the richness.

Table 8. The first 10 entries of the catalogue of the comparison of GPs masses to CoMaLit V masses for a subsample of SDSS DR8 redMaPPer clusters, [redmapper_dr8_public_v6.3_catalog.fits](#). The forecast by CoMaLit V is available [here](#). Col. 1: index. Col. 2: name of the cluster. Col. 3: GPs mass and associated error, Col. 4: CoMaLit V mass and associated error. Masses are within R_{500} and in units of $10^{14}M_{\odot}$.

Index	Name	M_{GPs}	M_{CoM}
1	RMJ010739.8+084	3.311 ± 0.985	5.622 ± 0.599
2	RMJ222714.4-023	5.736 ± 2.594	2.282 ± 0.366
3	RMJ145632.6+184	0.382 ± 0.183	2.607 ± 0.503
4	RMJ112124.1+515	2.686 ± 1.346	2.501 ± 0.389
5	RMJ130629.0+135	1.057 ± 0.485	5.137 ± 1.187
6	RMJ022859.3+041	2.668 ± 1.473	2.742 ± 0.400
7	RMJ112716.1+204	4.867 ± 2.097	4.422 ± 0.435
8	RMJ100417.8-011	4.779 ± 2.110	2.378 ± 0.398
9	RMJ104252.5+341	4.996 ± 2.614	2.265 ± 0.360
10	RMJ002148.9+065	3.357 ± 2.058	3.047 ± 0.501

Finally, we applied the GPs algorithm on MCXC data. There are 196 clusters within MCXC catalogue whose WL mass are available through LC2 and these constitute our training data. These matching clusters are those with a redshift difference less than $\Delta z = 0.05$ and a maximum angular separation of 2 arcminutes in the sky. We used the mass and luminosities of these clusters to estimate the mass of the remaining clusters given their luminosities.

Table 9. The first 10 entries of the catalogue of the comparison of GPs masses to CoMaLit V and MCXC masses, [J/A+A/534/A109/mcxc.dat](#). The forecast by CoMaLit V is available [here](#). Col. 1: index. Col. 2: name of the cluster. Col. 3: GPs mass and associated error, Col. 4: CoMaLit V mass and associated error, Col. 5: MCXC mass. Masses are within R_{500} and in units of $10^{14}M_{\odot}$.

Index	Name	M_{GPs}	M_{CoM}	M_{xc}
1	J0001.6-1540	2.843 ± 2.136	2.067 ± 0.236	1.656
2	J0001.9+1204	3.318 ± 2.232	3.437 ± 0.297	2.693
3	J0003.2-3555	2.758 ± 4.076	1.452 ± 0.200	1.202
4	J0003.8+0203	2.072 ± 1.726	2.109 ± 0.240	1.734
5	J0004.9+1142	2.190 ± 2.145	1.594 ± 0.209	1.301
6	J0005.3+1612	3.111 ± 2.720	3.010 ± 0.287	2.493
7	J0006.0-3443	3.901 ± 2.884	3.246 ± 0.303	2.712
8	J0006.3+1052	4.301 ± 2.882	3.701 ± 0.316	2.994
9	J0008.9+4110	3.343 ± 2.315	3.550 ± 0.309	2.896
10	J0009.7-3516	3.082 ± 4.928	1.080 ± 0.168	0.873

Figure 38 shows the predicted mass for 1382 MCXC clusters with luminosities within the interval of training data given their X-ray luminosity and the WL mass of 196 clusters in LC2. The result of the comparison between GPs, CoMaLit V and MCXC masses is available in [this](#) catalogue. The first 10 entries of the catalogue are also presented in Table 9.

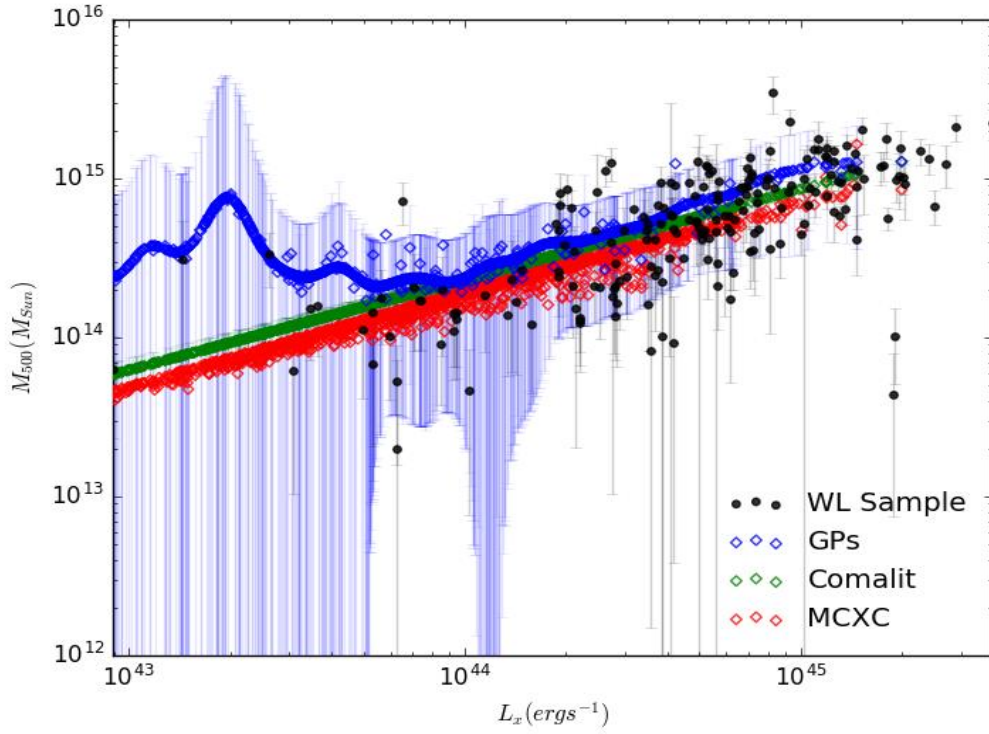


Figure 38. The predicted mass for 1382 clusters of MCXC (blue diamonds) using the X-ray luminosity and WL mass of 196 clusters with available WL mass in LC2 (black points). The green diamonds show the estimated mass in CoMaLit V and the red diamonds show the MCXC masses.

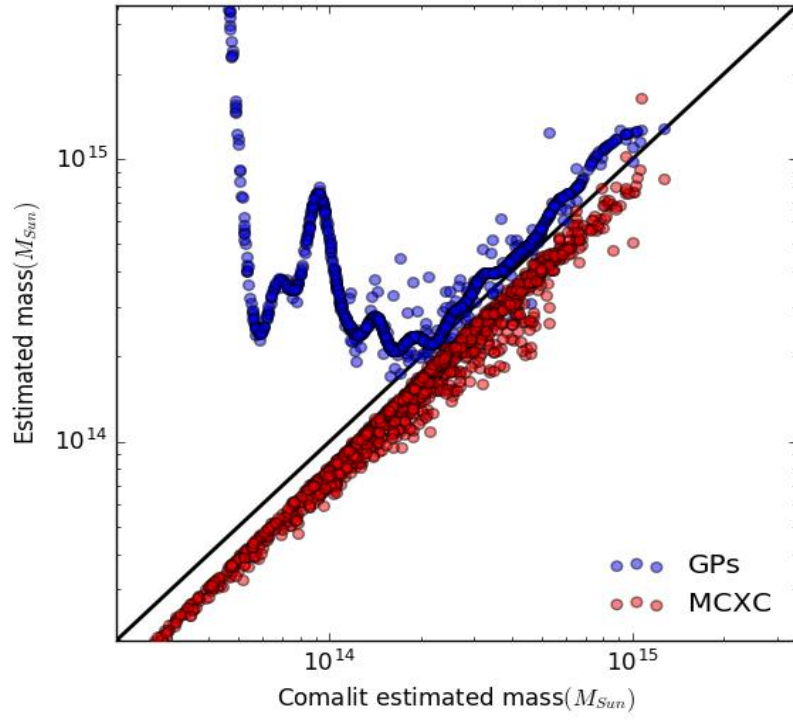


Figure 39. Comparison of the predicted masses for 1382 clusters of MCXC by GPs (blue circles) and MCXC (red circles) to the masses estimated in CoMaLit V.

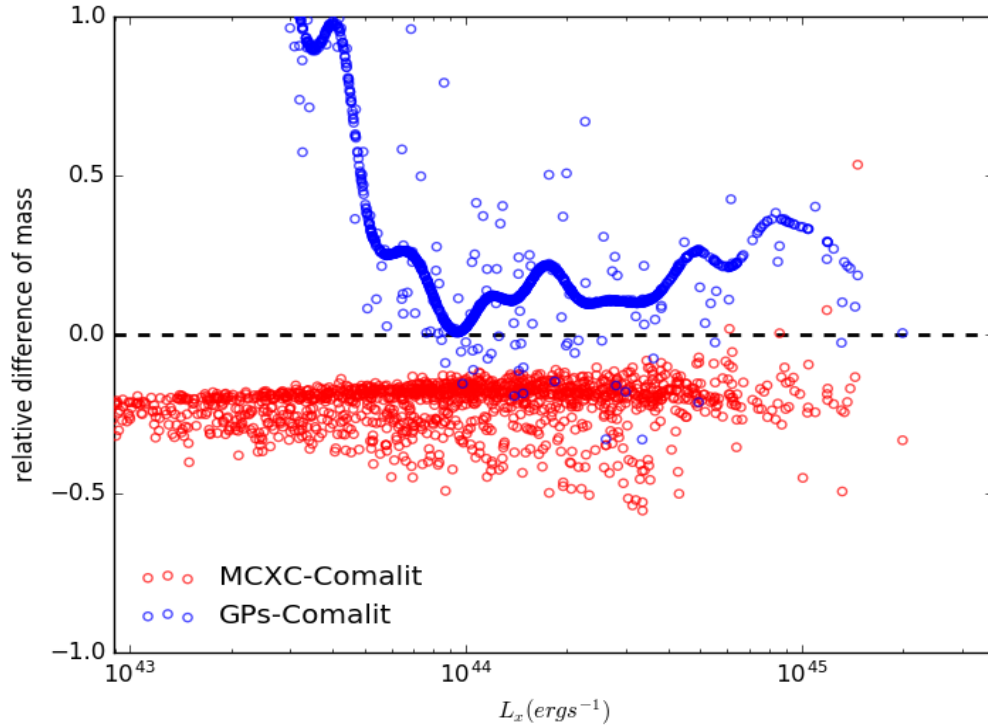


Figure 40. Relative difference of masses predicted by MCXC (red circles) and GPs (blue circles) with respect to the mass estimates by CoMaLit V versus the X-ray luminosity.

Figure 39 shows the comparison between the masses predicted by GPs and MCXC to CoMaLit V results. It is obvious that as long as we have enough training data, which are mostly found in higher mass ranges, GPs predictions overestimates mass with respect to CoMaLit V results, while towards lower masses, where there are few clusters with known WL mass, the GPs estimates are not reliable and the uncertainty in GPs predictions increases. It also shows that MCXC predictions are lower than what is estimated in CoMaLit V. In Figure 40, the predictions by MCXC and GPs are compared through their relative difference with respect to CoMaLit V predictions. The results confirm the importance of the existence of enough clusters with known WL mass to predict the mass of remaining clusters as precise as possible.

All the results highlight the power of GPs as a tool to predict the mass of galaxy clusters as long as there are enough clusters with known WL mass.

Chapter 5

خوش باش دمی که زندگانی نیست

عمر خیام نیشابوری

Be happy for this moment. This moment is your life

Omar Khayyam Neishabouri

Conclusions and future work

5.1 Conclusions

The objective of the work presented in this thesis was to develop and apply statistical tools to infer the properties of individual groups/clusters of galaxies in different wavelengths. This project was divided into two main projects, which can be summarized as follows:

- In the first part, we characterized the evolution of the X-ray $L_X - T$ relation through a hierarchical Bayesian statistical framework using a sample of 353 clusters and groups of galaxies from XCS. Based on the results that do not take into account the selection effects, we did not find strong statistical support for any type of the deviations considered from the usual modelling of the $L_X - T$ relation as a single power-law, where the normalisation evolves self-similarly and the scatter remains constant with time. Nevertheless, assuming a priori the existence of the type of deviations considered, then faster evolution than the self-similar expectation for the normalisation of the $L_X - T$ relation seems favoured, eventually signalling an

increase with redshift in the average group/cluster X-ray luminosity at fixed temperature, as well as a decrease with redshift in the scatter about the $L_X - T$ relation. In addition, although we did not find enough statistical evidence indicating that clusters and groups of galaxies obey different scaling relations, we did find some preliminary indications that might be the case. Further data is needed to draw more definitive conclusions. By considering sample selection effects, some of the conclusions, particularly those with respect to the behaviour of the $L_X - T$ relation in the group regime may change.

- In the second part, we developed a non-parametric Bayesian framework, based on Gaussian Processes and applied it to three samples of Planck SZ2, SDSS and MCXC clusters in order to forecast their mass given their SZ flux, optical richness or X-ray luminosity. The results demonstrated the power of GPs to forecast the mass of clusters as long as there are enough clusters with known WL mass to be used as training data in the analysis. The estimated mass for the clusters in three catalogues mentioned above were presented and compared to the masses obtained in CoMaLit V and through SZ flux/X-ray proxies. Our results show that while Planck and MCXC masses are lower than those estimated in CoMaLit V, GPs tend to an over estimation of the mass with respect to CoMaLit V predictions for Higher SZ flux clusters in PSZ2 and for all MCXC and SDSS clusters as long as there are enough clusters with available WL mass.

5.2 Future work

The methods that we used in this project can be improved and be used for other purposes. The issues that we would like to address in the future follows.

- We intend to consider the possibility of assuming an arbitrary function, other than a Gaussian mixture, as a prior distribution for the independent parameter through developing a code or adding it to LIRA. This possibility, together with considering the survey selection function, when available, leads to a more realistic estimation of the parameters of mass-based scaling relations. e.g. $L - M$ relation, by inserting the mass function as the prior distribution for the cluster mass. This may also lead to constrain cosmological parameters which are inserted in mass function by setting them as free parameters. However, the convergence of MCMC chains demands more observational data when a higher number of free parameters are assumed. These data will be available through existing

catalogues which are updating (e.g. XCS-DR2) or through future surveys (i.e. Euclid).

- Another work that can be done is to use GPs to predict the mass of clusters in other surveys like XCS DR2. The existing selection function of XCS could also be included in our analysis. In order to do that, first we should digest which clusters in XCS-DR2 have available WL mass in the literature. This can be done through cross matching XCS-DR2 clusters to LC2
- We intend to use more dimensional input data in our GPs analysis. In the first step, the redshift of the cluster can also be inserted as the input data in GPs in addition to the survey observable data, which is L_X , SZ flux or richness. This would improve GPs results in our future analysis. Then, we can make benefit of the multi-component nature of galaxy clusters in our GPs analysis. We aim to use as many as dimensions possible for the input data to predict the mass of clusters. For this goal, we aim to use the sample that we applied for the first part of this thesis project, $L_X - T$ relation. The richness, L_X and redshift of the X-ray clusters of this sample, which also appear in SDSS DR8, can be used to predict their mass through GPs analysis.

Appendix

Appendix

چشم ہارا باید شست، جور دیگر باید دید...

سہراب سہری

One must wash eyes, look differently to things...

Sohrab Sepehri

In this Appendix, we use the sample considered in Hilton et al. 2012 to illustrate the impact on the estimation of X-ray luminosities and temperatures of the changes described in Rooney 2015, Bermeo-Hernandez 2017 and Manolopoulou et al. 2019, as well as on the characterization of the $L_X - T$ relation.

In Figure 41, we compare the expected values and associated standard deviations for the X-ray temperatures and luminosities for 204 out of the 211 groups and clusters considered in Hilton et al. 2012, as then estimated using the X-ray analysis methodology described in Lloyd-Davies et al. 2011 and using the current X-ray analysis methodology (Rooney 2015; Bermeo-Hernandez 2017; Manolopoulou et al. 2019). We omitted seven systems on the original sample from the comparison, because the current X-ray analysis methodology does not yield reliable values for the temperature and/or luminosity. The omission of these systems from the analysis should not have much of an impact, and thus henceforth we will often distinguish between the the original sample with 211 systems and its sub-sample of 204 systems. With respect to the X-ray luminosities, the current values for these systems are slightly higher on average than the previous ones, with the mean increasing from 1.40 to $1.46 \times 10^{44} \text{ erg s}^{-1}$, and have significantly smaller associated uncertainties. For the X-ray temperatures, on the contrary, the current

values are on average lower than the previous ones, with the mean decreasing from 3.24 to 3.06 *keV*.

Although the model adopted for the $L_X - T$ relation in Hilton et al. 2012 is similar to what we here labelled as model 1u, we differ in the way the measurement uncertainties associated with the values of the sample X-ray temperatures and luminosities are propagated, and in how the likelihood is specified. With regards to the later, while Hilton et al. 2012 penalized deviations between observed values and model predictions, for the X-ray temperatures and luminosities, using orthogonal and bisector metrics, here we focused only on such deviations with respect to the X-ray luminosity, at fixed X-ray temperature, in a manner analogous to what is commonly known as $Y | X$ regression. If sample selection effects are not taken properly into account, these different approaches to likelihood specification can easily lead to significantly different results. Applying our model 1u to the sample in Hilton et al. 2012, assuming the expected values and associated standard deviations for the observed X-ray temperatures and luminosities as obtained with the previous X-ray analysis methodology, we obtain for the expected values and standard deviations: $\alpha = 44.22 \pm 0.07$, $\beta = 2.21 \pm 0.13$, $\sigma = 0.28 \pm 0.02$ and $\gamma = 0.11 \pm 0.46$. For comparison, the bisector method yielded $\alpha = 44.41 \pm 0.04$, $\beta = 2.65 \pm 0.09$, $\sigma = 0.34 \pm 0.02$ and $\gamma = -0.7 \pm 0.3$, while applying the orthogonal method resulted in $\alpha = 44.63 \pm 0.07$, $\beta = 3.02 \pm 0.15$, $\sigma = 0.27 \pm 0.03$ and $\gamma = -1.9 \pm 0.5$ (Table 3, Hilton et al. 2012). Therefore, as was somewhat expected given the different methodologies used, our analysis of the same data considered in Hilton et al. 2012 leads to markedly different results. However, it should be noted that our results differ as much from those yielded by the bisector method, as these differ from those that arise from applying the orthogonal method.

We have also applied models 1u and 1 to the sample in Hilton et al. 2012, but now assuming the expected values and associated standard deviations for the observed X-ray temperatures and luminosities as obtained with the current X-ray analysis methodology. Note that, in the case of model 1, as before we take into account the XCS selection function, but now we do not assume a priori that a lower threshold in temperature as been imposed in the process of sample selection. In Table 10 we present the results, and in Figure 42 we plot the distribution of the data points associated with the Hilton et al. 2012 sample, as well as the regression lines obtained assuming the expected (i.e. mean) values for the relevant model hyper-parameters, either assuming model 1u or 1. As happened when the same models were applied to the XCS-SDSS

data, we find an increase in the expected values for α and β , as well as a decrease in the expected value for γ , when sample selection effects are taken into account.

Table 10. The mean and standard deviation for the marginal posterior distributions of the interesting hyper-parameters associated with each model, given 204 groups and clusters in the Hilton et al. 2012 sample. The expected values and associated standard deviations for the observed X-ray temperatures and luminosities were obtained with the current X-ray analysis methodology.

Model	α	β	σ	γ
1u	44.02, 0.07	2.03, 0.14	0.33, 0.02	2.40, 0.54
1	44.23, 0.07	2.53, 0.15	0.31, 0.02	1.22, 0.52

The results obtained by applying model 1u to the sample in Hilton et al. 2012, assuming either the expected values and associated standard deviations for the observed X-ray temperatures and luminosities as obtained with the previous X-ray analysis methodology, or those obtained with the current X-ray analysis methodology, are compatible within the statistical uncertainties, except for the evolution parameter γ . The previous X-ray data suggested much weaker evolution with redshift for the normalisation of the $L_X - T$ relation than the current X-ray data. On the other hand, the expected values for the model parameters α , β and γ , are significantly different from what was obtained assuming the XCS-SDSS data, for both models 1u and 1. This would be expected in the former case, given the absence of correction for the different selection procedures behind the assembly of both samples. But it is somewhat surprising in the later case, although the discrepancy becomes smaller. Possibly, it reflects the greater difficulty of modelling the more complex nature of the selection function associated with the assembly of the sample used in Hilton et al. 2012.

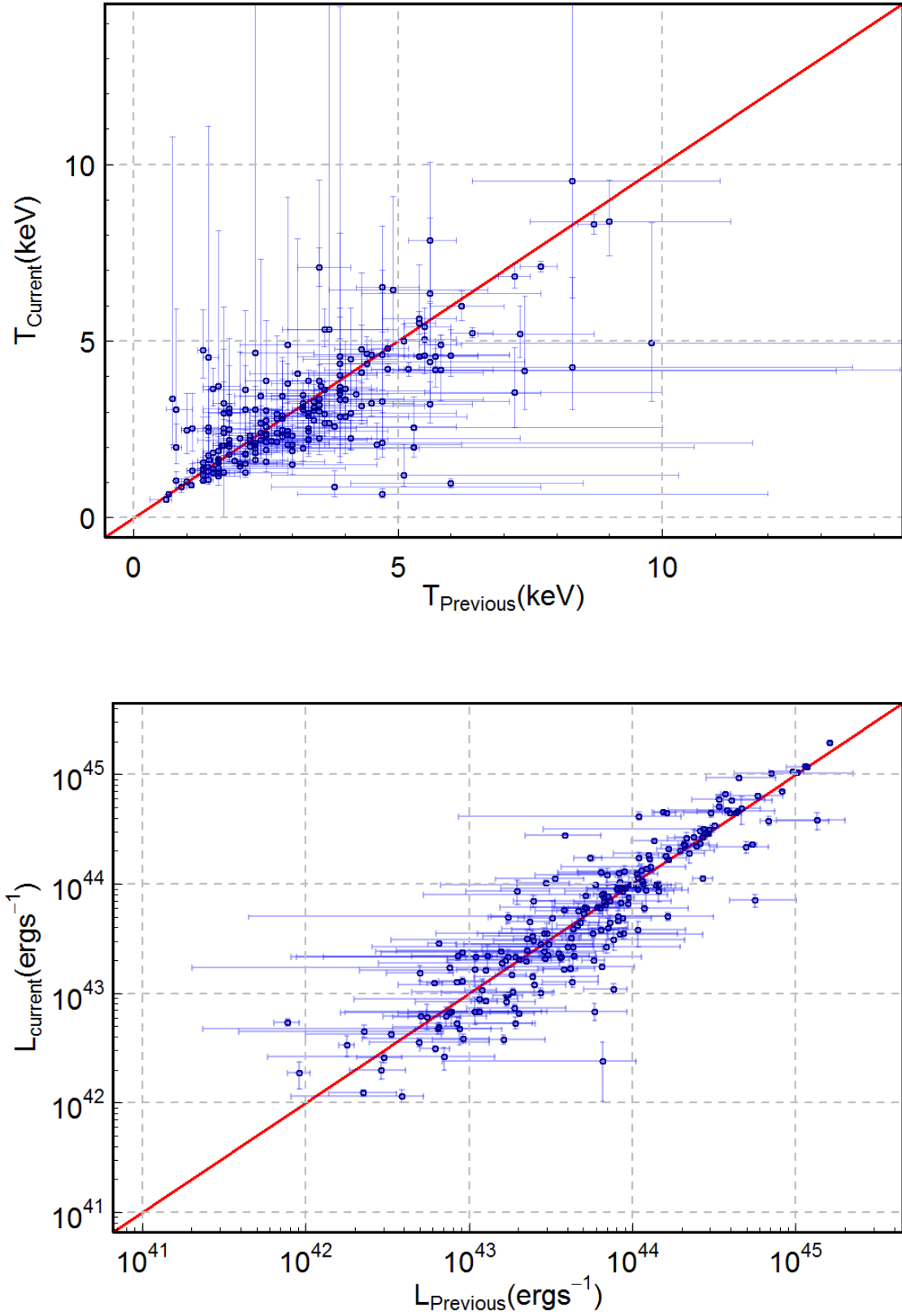


Figure 41. Comparison of the expected values and associated standard deviations for the X-ray temperatures and luminosities of 204 groups and clusters considered in Hilton et al. 2012, using the previous and current X-ray analysis methodology.

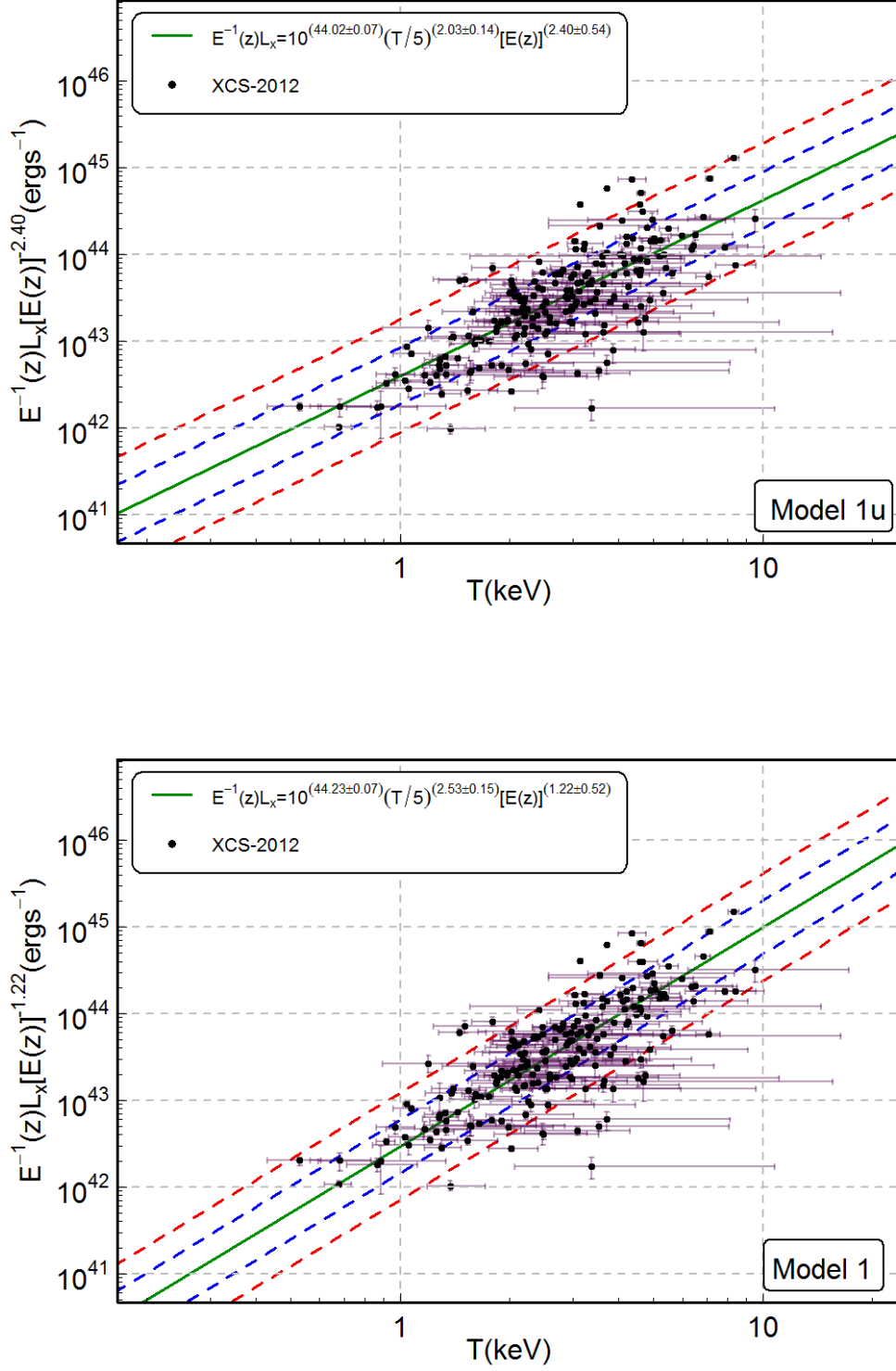


Figure 42. The $L_X - T$ relation assuming model 1, without (top, model 1u) and with (down, model 1) sample selection effects taken into account. Each point indicates the temperature, T , in keV , and re-scaled bolometric X-ray luminosity, L_X , expected for each of the 204 systems in the Hilton et al. 2012 sample considered here, given the X-ray data. The re-scaling of L_X with the cluster redshift, z , is performed dividing L_X by $E(z)^{1+\gamma}$, where γ is the expected (i.e. mean) value for γ . The error bars associated with each point identify the 1σ uncertainty intervals. The solid regression line was determined by fixing the model hyper-parameters to their expected (i.e. mean) values. The amplitude of the intrinsic vertical scatter about the regression line is indicated by the dashed lines (1σ , inner blue; 2σ , outer red).

Bibliography

- Akritas M. G. & Bershadsky M. A., 1996, *AJ*, 470, 706
- Allen S. W. et al., 2002, *MNRAS*, 334, L11
- Allen et al., 2011, *ARA&A*, 49, 409
- Andreon S. & Weaver S., 2015, *Bayesian Methods for the Physical Sciences Learning from Examples in Astronomy and Physics*, Springer Series in Astrostatistics
- Arnaud M. et al., 2010, *A&A*, 517, A92
- Aschenbach B., 1988, *Appl. Opt.*, 27, 1404
- Bahcall N. A. et al., 2000, *ApJ*, 541, 1
- Barnes D. J. et al., 2017, *MNRAS*, 465, 213
- Bartlett J.G., 1997, *Astronomical Society of the Pacific Conference Series*, 126, 365
- Bermeo-Hernandez A., 2017, *PhD Thesis*, University of Sussex
- Bharadwaj V. et al., 2015, *A&A*, 573, A75
- Böhringer H. et al., 2014, *A&A*, 570, A31
- Borgani S. & Guzzo L., 2001, *Nature*, 409, 39
- Borgani S., 2003, *Proceedings of The Riddle of Cooling Flows in Galaxies and Clusters of Galaxies*
- Borgani S., 2008, *Lecture Notes in Physics*, 740, Springer, Dordrecht
- Carlberg R. G. et al., 1996, *ApJ*, 462, 32
- Cavaliere A. & Lapi A., 2014, *A&A*, 571, A84
- Clerc N. et al., 2014, *MNRAS*, 444, 2723
- Czacon N. G. et al., 2015, *ApJ*, 806, 18
- D'Agostini G. & Raso M., 2000, *CERN-EP/2000-026 internal document* (hepex/0002056)
- de Haan T. et al., 2016, *ApJ*, 832, 95
- Dallaire P. et al., 2011, *Neurocomputing*, 74, 1945
- Ebrahimpour L. et al., 2019, submitted to *MNRAS*
- Eckmiller H. J. et al., 2011, *A&A*, 535, A105
- Eisenstein D. J. et al., 2001, *AJ*, 122, 2267
- Ettori S. et al., 2003, *A&A* 398, 879
- Euclid Red Book, 2011, URL:<http://sci.esa.int/jump.cfm?oid=48983>

Fabian A. C., 1991, MNRAS, 253, 29P

Feigelson E. D. & Babu G. J., 2012, <http://arXiv:1205.2064>

Friel N. et al., 2017, Stat. Comput., 27, 833

Gelfand A. E. et al., 1990, J. AM. STAT. ASSOC., 85, 1990

Gelfand A. E. & Smith A. F. M., 1990, J. Am. Stat. Assoc., 85, 398

Gelman A. & Rubin D. B., 1992, Stat. Sci., 7, 457

Gelman A. et al., 2008, Ann. Appl. Stat., 2, 1360

Gelman A. et al., 2014, Stat. Comput., 24, 997

Ghosh J. et al., 2017, <https://arxiv.org/abs/1507.07170v2>

Giles P. A. et al., 2016, A&A, 592, A3

Giodini S. et al., 2013, Space Sci Rev, 177, 247

Girardi M. et al., 2000, ApJ, 530, 62

Gitti M. et al., 2012, Advances in Astronomy, article ID 950641

Gregori Ph., 2005, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press

Gunn J. E. et al., 1998, AJ, 116, 3040

Gunn J. E. et al., 2006, AJ, 131, 2332

Hilton M. et al., 2012, MNRAS, 424, 2086

Hitchcock D. B., 2003, Am. Stat., 57, 254

Hogg D. W. et al., 2010, <http://arXiv:1008.4686>

Isobe T. et al., 1990, AJ, 364, 104

Jansen F. et al., 2001, A&A 365, L1

Kaiser N., 1986, MNRAS, 222, 323

Kelly B., 2007, AJ, 665, 1489

Lau T. et al., 2015, ApJ, 806, 68

Le Brun A. M. C. et al., 2017, MNRAS, 466, 4442

Lieu M. et al., 2016, A&A, 592, A4

Lloyd-Davies E. J. et al., 2011, MNRAS, 418, 14

Lovisari L. et al., 2015, A&A, 573, A118

Lombriser L. et al., 2012, Phys. Rev. D., 85, 124038

Manolopoulou M. et al., 2019, to be submitted for publication

Mantz A. et al., 2014, MNRAS, 440, 2077

Mantz A. 2016, MNRAS, 457, 1279

Mantz A. et al., 2016, MNRAS, 463, 3582

Mantz A. et al., 2010a, MNRAS, 406, 1773

Mantz A. et al., 2010b, MNRAS, 406, 1759

Maughan B. J., 2014, MNRAS, 437, 1171
 Maughan B. J. et al., 2012, MNRAS, 421, 1583
 Mehrtens N. et al., 2012, MNRAS, 423, 1024
 Metropolis N. et al., 1953, J. Chem. Phys., 21, 1087
 Mittal R. et al., 2011, A&A, 532, A133
 Mohr J.J., 2005, *Astronomical Society of the Pacific Conference Series*, 339
 Moscardini L. et al., 2001, MNRAS, 327, 422
 Padmanabhan N. et al., 2008, AJ, 674, 1217
 Pier J. R. et al., 2003, AJ, 125, 1559
 Piffaretti R. et al., 2011, A&A, 534, A109
 Piironen J. & Vehtari A., 2017, Stat. Comput., 27, 711
 Planck Collaboration, 2014, A&A, 571, A20
 Planck Collaboration, 2014, A&A, 571, A29
 Planck Collaboration, 2016, A&A, 594, A24
 Planck Collaboration, 2016, A&A, 594, A27
 Planelles S. et al., 2014, MNRAS, 438, 195
 Ponman T.J. et al., 1999, Nature, 397, 135
 Pratt G. W. et al., 2009, A&A, 498, 361
 Rasmussen C. E. & Williams C. K. I., 2006, *Gaussian Processes for Machine Learning*, the MIT Press
 Reichert A. et al., 2011, A&A, 535, A4
 Richards G. T. et al., 2002b, AJ, 123, 2945
 Robotham D. et al., 2015, Alzheimer's Res. Ther.
 Romer A. K. et al., 2001, ApJ, 547, 594
 Rozo E. et al., 2010, ApJ, 708, 645
 Rozo E. et al., 2010, MNRAS, 450, 592
 Rozo E. & Rykoff E. S., 2014, ApJ, 783, 80
 Rooney P. J., 2015, *PhD Thesis*, University of Sussex
 Rykoff E. S. et al., 2013, ApJ, 785, 104
 Rykoff E. S. et al., 2016, ApJS, 224, 1
 Sahlén M. et al., 2009, MNRAS, 397, 577
 Sánchez E. & the Des collaboration, 2010, J. Phys.: Conf. Ser, 259
 Sarazin Craig L., 1988, *Proceedings of the International School of Physics "Enrico Fermi"*
 Seikel M. et al., 2012, JCAP, 2012
 Sereno M., 2015, MNRAS, 450, 3665

Sereno M., 2016, MNRAS, 455, 2149

Sereno M. & Ettori S., 2015, MNRAS, 450, 3675

Sereno M. & Ettori S., 2017, MNRAS, 468, 3322

Sharma S., 2017, Annu. Rev. Astron. Astrophys., 55, 213

Short C. J. et al., 2013, 428, 1225

Stanek R. et al., 2010, ApJ, 715, 1508

Strauss M. A. et al., 2002, AJ, 124, 1810

Sun M., 2012, New J. Phys., 14, 045004

Takey A. et al., 2013, A&A, 558, A75

Tanner M. A. & Wong W. H., 1987, J. Am. Stat. Assoc., 82, 528

Trumper J., 1983, Adv. Space Res., 2, 241

Trumper J. et al., 1991, Nature, 349, 579

Truong N. et al., 2018, MNRAS, 474, 4089

Tucker D. L. et al. 2006, AN 327, No. 9, 821

Vehtari A. & Ojanen J., 2012, Stat. Surv., 6, 142

Vehtari A. et al., 2012, Stat. Compu., 27, 1413

Vehtari A. et al., 2017, <https://arxiv.org/abs/1507.02646>

Vikhlinin A. et al., 2009, ApJ, 692, 1060

Villani M. & Larsson R., 2006, Commun. Stat. Theory Methods, 35, 1123

Voges W. et al., 1999, A&A, 349, 389

Voit G. Mark, 2005, Rev. Mod. Phys., 77, 207

Watanabe S., 2010, J. Mach. Learn. Res., 11, 3571

Watanabe S., 2013, J. Mach. Learn. Res., 14, 867

White S. D. M. et al., 1993, Nature, 366, 429

Yanny B. et al., 2009, AJ, 137, 4377

York D. G. et al., 2000, AJ, 120, 1579

Zou S. et al., 2016, MNRAS, 463, 820