A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details
Exploiting the XMM science archive

For X-ray galaxy cluster mass calibration and other astrophysical applications

David J. Turner

Submitted for the degree of Doctor of Philosophy
University of Sussex
August 2022
Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

David J. Turner
Abstract

The exploitation of the X-ray archive of XMM is essential to the work of the XMM Cluster Survey (XCS), which makes use of the XMM archive to identify and quantify galaxy clusters. To that end, this thesis provides a detailed account of the design and capabilities of the new open-source Python module X-ray: Generate and Analyse (XGA). This software allows for the simple generation and analysis of XMM data products through its ‘source-based’ paradigm, where the user need only input a coordinate to select all observations relevant to that source. The application of XGA to the measurement of galaxy cluster hydrostatic masses is then presented, including measurement comparisons between XGA and published results. Once the veracity of XGA measurements has been demonstrated, mass measurements of mass for the SDSSRM-XCS sample are given. Following this, the assembly of the new ACTDR5-XMM sample is detailed and (along with clusters from the DESY3RM-XCS sample) new hydrostatic mass measurements are presented; this results in a combined sample of 334 unique clusters with hydrostatic masses.

Next the first independent verification of the veracity of eROSITA galaxy cluster analyses is presented, making use of the eFEDS catalogues. XGA is used to construct a sample of 37 eFEDS galaxy clusters with XMM and HSC confirmation. Several failure modes of the eROSITA cluster finder are identified. X-ray luminosities measured using XMM and eROSITA data were found to be in excellent agreement, though XMM measured temperature were found to be (on average) 25% hotter than eROSITA. Finally, short summaries of other research projects undertaken during the course of this PhD are presented. This is followed by a final discussion of the contents of this thesis, alongside overviews of planned future work that builds on the research performed during the PhD.
Acknowledgements

Undertaking this PhD has been one of the most difficult things I’ve ever done, not exactly helped by a global pandemic happening in the middle of it, and there were points where I thought that I would not be able to finish it. I would not have gotten to this point without the support of my family and my friends, an incredible group of people who mean the world to me.

I have to start by thanking my Mum (Pauline), my Dad (Duncan), and my brother (Dan) for not only giving me their unwavering support, but for helping to make me the person I am today; for nurturing my curiosity and interests as I grew up, and for encouraging me as I took that curiosity all the way to doing a PhD. The comfort and strength I take from being able to talk to them about anything is immeasurable, and I am lucky to have such fundamentally good people as my family and role models.

Without Aswin and Lucas, the time of the pandemic and the lockdowns would have been intolerable. While we may not have known each other very well when we moved in together, I was so lucky to end up living with such fantastic people. You made every day brighter (especially Fridays) and I’ve missed our talks and hangouts immensely since you moved away. Aswin’s culinary experiments and Lucas’ constant ‘It’s always sunny in Philadelphia’ references will be remembered for a very long time.

Next, my dear friend Tim, who is one of the most intelligent, driven, and imaginative people that I’ve ever met. We have a friendship that I treasure, and knowing that he always has my back (and vice versa) is such a special feeling. He’s put up with the poor communication abilities that coincide with the periods of extreme stress and anxiety that have occurred during this PhD, and he’s always ready to have fun playing games with me. I would be much worse off without him in my life.

My wonderful PhD colleagues also deserve a huge amount of thanks, people that I’ve shared offices with, lunches with, and games nights and pubs nights with. Maria, Itzi, Will, Stephen,
Cameron, Will, Jussi, Edward, Luke, Kareem, and everyone else who has helped make Sussex such a fantastic place to be. Some of you I have only known during my PhD, some I’ve known quite a bit longer, and you’ve all helped make my time here so much better.

There have been so many amazing people that I’ve met in the time that I’ve lived in Brighton, people I met through University (many of whom I lived with). People that I met through my love of music, and who helped grow that into a love of performing music, learning (and buying) many instruments along the way. James, Emma, Shane, Cat, Suze, Beth, Anna, Tom, George, Louis, and so many others. You are all special, and I am going to miss you during my time in the US.

Finally, I have to thank the research group that I’ve been lucky to be a part of (though I will forever root for Edward to beat Paul at darts); but I especially owe a huge debt of gratitude to Kathy Romer. If Kathy had not ‘collected’ me when I over-complicated a solution (that’s become a running theme of my work) to a Python problem nearly eight years ago, I would not be where I am today. Her mentorship, her guidance, and her support are things I have come to value more than nearly anything else; anyone that finds themselves working with her should consider themselves extremely lucky, because she is one of a kind. She helped nurture my scientific instinct, letting me explore things by myself and helping when I needed it, and taught me the values of openness and collaboration that I think should embody all scientific exploits. It is time now for me to finally move away, to work with some new people and start the next phase of my scientific career, but if I ever find myself in the same position as her, supervising and nurturing new scientists, I can only hope that I’m even half as good to them as she was to me.
“Living either never has any point, or is always its own point; being a naturally cheery soul, I lean towards the latter.”

— Iain M. Banks, The Hydrogen Sonata
## Contents

### List of Tables

### List of Figures

### 1 Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Motivation and aims</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Galaxy Clusters</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Formation of Galaxy Clusters</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Composition of Galaxy Clusters</td>
<td>4</td>
</tr>
<tr>
<td>1.2.3 Galaxy Cluster Scaling Relations</td>
<td>10</td>
</tr>
<tr>
<td>1.2.4 Cosmology with Galaxy Clusters</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Observations of Galaxy Clusters</td>
<td>16</td>
</tr>
<tr>
<td>1.3.1 X-ray</td>
<td>16</td>
</tr>
<tr>
<td>1.3.2 Optical and Near-Infrared</td>
<td>23</td>
</tr>
<tr>
<td>1.3.3 Radio and Millimetre Wave</td>
<td>25</td>
</tr>
<tr>
<td>1.4 The XMM Cluster Survey (XCS)</td>
<td>27</td>
</tr>
<tr>
<td>1.4.1 XCS Science Goals and Achievements</td>
<td>28</td>
</tr>
<tr>
<td>1.4.2 Data preparation</td>
<td>29</td>
</tr>
<tr>
<td>1.4.3 Existing software tools</td>
<td>31</td>
</tr>
<tr>
<td>1.4.4 Master Source List (MSL)</td>
<td>32</td>
</tr>
<tr>
<td>1.5 Relevant Software</td>
<td>32</td>
</tr>
<tr>
<td>1.5.1 The XMM Science Analysis System (SAS)</td>
<td>33</td>
</tr>
<tr>
<td>1.5.2 The XSPEC Spectral Fitting Code</td>
<td>34</td>
</tr>
<tr>
<td>1.5.3 Astropy</td>
<td>35</td>
</tr>
<tr>
<td>1.6 Samples of Galaxy Clusters</td>
<td>36</td>
</tr>
<tr>
<td>1.6.1 SDSS redMaPPer (Optical/NIR selected)</td>
<td>36</td>
</tr>
<tr>
<td>1.6.2 DESY3 redMaPPer (Optical/NIR selected)</td>
<td>39</td>
</tr>
<tr>
<td>1.6.3</td>
<td>Ultimate XMM eXtragalactic (XXL)-100-GC (X-ray selected)</td>
</tr>
<tr>
<td>1.6.4</td>
<td>Local Cluster Substructure Survey (LoCuSS) High-$L_X$ (X-ray selected)</td>
</tr>
<tr>
<td>1.6.5</td>
<td>eROSITA Final Equatorial-Depth Survey (eFEDS; X-ray selected)</td>
</tr>
<tr>
<td>1.6.6</td>
<td>ACT-DR5 (SZ selected)</td>
</tr>
<tr>
<td>1.7</td>
<td>Thesis Overview</td>
</tr>
</tbody>
</table>

### 2 X-ray: Generate and Analyse (XGA)

| 2.1 | XGA Paper | 44 |
| 2.1.1 | Summary | 44 |
| 2.1.2 | Statement of need | 45 |
| 2.1.3 | Features | 46 |
| 2.1.4 | Existing software packages | 47 |
| 2.1.5 | Research projects using XGA | 49 |
| 2.1.6 | Future Work | 49 |

| 2.2 | The overall design of XGA | 49 |

| 2.3 | XGA Sources and Samples | 50 |
| 2.3.1 | What are sources and samples? | 50 |
| 2.3.2 | Types of source | 51 |
| 2.3.3 | Types of sample | 53 |
| 2.3.4 | Fetching relevant XMM data | 54 |
| 2.3.5 | Matching to regions | 55 |

| 2.4 | XGA Products | 58 |
| 2.4.1 | What are products? | 58 |
| 2.4.2 | The four base types of XGA product | 59 |
| 2.4.3 | How does XGA generate XMM data products using SAS? | 61 |
| 2.4.4 | EventList | 62 |
| 2.4.5 | Image | 62 |
| 2.4.6 | ExpMap | 63 |
| 2.4.7 | RateMap | 63 |
| 2.4.8 | PSF and PSFGrid | 72 |
| 2.4.9 | Spectrum | 77 |
| 2.4.10 | AnnularSpectra | 80 |
| 2.4.11 | XGA profile classes | 85 |

| 2.5 | Future development plans for XGA | 92 |
| 2.5.1 | Support for multi-mission analyses | 92 |
2.5.2 Simple analysis of simulated X-ray data ........................................ 93
2.5.3 Measurement of 2D property maps and development of 2D models ...... 93
2.5.4 Support for time-domain analyses .................................................. 94
2.5.5 Deployment on the ESA DataLabs platform .................................... 94

3 Tests of XGA measurements and new masses for the SDSSRM-XCS, DESY3RM-XCS, and ACTDR5-XCS samples 96
3.1 Introduction ................................................................. 96
3.2 XCS Data and Cluster Samples ................................................ 99
  3.2.1 XCS cleaned data and source regions ....................................... 99
  3.2.2 SDSSRM-XCST_{X}\text{,vol} .............................................. 100
  3.2.3 XXL-100-GC .......................................................... 100
  3.2.4 LoCuSS High-\(L_{X}\) ..................................................... 100
  3.2.5 Comparison of samples ....................................................... 101
3.3 The X-ray: Generate and Analyse (XGA) software package ................. 101
  3.3.1 Summary of XGA ......................................................... 101
  3.3.2 Selecting XMM data ...................................................... 102
  3.3.3 Generating and interacting with photometric data ......................... 104
  3.3.4 Identifying and removing contaminating sources ......................... 104
  3.3.5 Manual adjustment of detection regions and inspection of data .......... 105
  3.3.6 Generating and fitting spectra .......................................... 105
  3.3.7 Comparing XGA \(T_{X}\) and \(L_{X}\) measurements to literature .......... 107
3.4 Measuring gas density, temperature, and mass profiles ....................... 115
  3.4.1 Fitting models to radial profiles in XGA ................................ 115
  3.4.2 PSF correction of XMM images ......................................... 117
  3.4.3 Gas density profiles .................................................... 119
  3.4.4 Comparisons of XGA gas mass measurements to literature ............ 122
  3.4.5 Generating and fitting annular spectra with XGA ......................... 125
  3.4.6 Three-dimensional gas temperature profiles ............................... 127
  3.4.7 Hydrostatic mass profiles ............................................. 131
  3.4.8 Comparison of XGA hydrostatic mass measurements to the LoCuSS High-
             \(L_{X}\) sample ......................................................... 131
  3.4.9 Inspection of data and measurements ..................................... 132
3.5 Mass analysis and scaling relations for the SDSSRM-XCS sample ............ 133
  3.5.1 Masses for a subset of the SDSSRM-XCS sample ......................... 133
3.5.2 Construction of mass-observable relations .................................................. 134
3.6 Discussion and next steps ................................................................................. 139
  3.6.1 Discussion .................................................................................................... 140
  3.6.2 Future Work ................................................................................................. 141
3.7 Constructing the ACTDR5-XMM sample ......................................................... 142
  3.7.1 ACT DR5 clusters with XMM data ................................................................. 143
  3.7.2 Properties of the ACTDR5-XMM sample ................................................... 144
3.8 Properties of the DESY3RM-XMM sample ....................................................... 146
3.9 Measuring masses for the ACTDR5-XCS and DESY3RM-XCS samples ........ 147
3.10 The largest combined sample of hydrostatic masses ......................................... 151
  3.10.1 Identifying catalogue overlaps between SDSSRM-XCS, DESY3RM-
      XCS, and ACTDR5-XMM .............................................................................. 151
  3.10.2 Mass observable relations for ACT DR5 and DESY3RM clusters .............. 155
3.11 What is there left to do? .................................................................................... 155
  3.11.1 Inspection of data ........................................................................................ 156
  3.11.2 Measure new $R_\Delta$ values ..................................................................... 156
  3.11.3 Joint fitting density and temperature profiles to enforce physicality ............ 157

4 Comparison of eROSITA and XMM cluster measurements .................................. 160
  4.1 Introduction ...................................................................................................... 160
  4.2 Comparison of the eFEDS Optically Confirmed and XXL-100-GC catalogues .... 163
  4.3 Understanding the eFEDS catalogue contamination fraction ......................... 165
    4.3.1 eFEDS Cluster Candidates in the XMM Footprint .................................... 166
    4.3.2 Constructing the eFEDS-XCS Sample ...................................................... 171
    4.3.3 Categories of contaminating objects in the eFEDS X-ray cluster candidate
      catalogue ...................................................................................................... 172
    4.3.4 Clusters with contaminated X-ray emission .............................................. 173
    4.3.5 The eFEDS contamination fraction ......................................................... 174
  4.4 Comparisons of cluster properties measured by eFEDS and XCS .................... 175
    4.4.1 Fitting Procedure ..................................................................................... 175
    4.4.2 Luminosity Comparison .......................................................................... 178
    4.4.3 Temperature Comparison ........................................................................ 179
    4.4.4 Temperature Calibration .......................................................................... 179
  4.5 Discussion ........................................................................................................ 182
    4.5.1 Comparison of eROSITA and XMM X-ray scaling relations ...................... 182
4.5.2 Future work to improve the calibration of the XMM to eROSITA temperature offset ................................................................. 187
4.6 Summary ................................................................................. 188

5 Conclusions and future work .................................................. 191
5.1 Other Projects ........................................................................ 191
  5.1.1 Artificial XMM observations of simulated galaxy clusters .... 191
  5.1.2 Building samples of Pea galaxies with machine learning ..... 198
  5.1.3 XMM properties of LoTSS DR2 radio sources ................. 208
5.2 Conclusions and discussion .................................................... 212
5.3 Future work ........................................................................... 214

Bibliography ............................................................................. 215

A Contributions to other work ................................................... 256
  A.1 X-ray confirmation of low-mass AGN candidates ............... 256
  A.2 Dynamical state of a cluster using MeerKAT and XMM ...... 256
  A.3 Upper limit X-ray luminosities of galaxy clusters .......... 257

B eFEDS-XMM cluster properties .............................................. 258
  B.1 Excluded Cluster Candidates ............................................. 258
  B.2 eFEDS Candidate 1023 ..................................................... 258
  B.3 eFEDS-XCS Data and Measurements .............................. 262
List of Tables

1 The types of region output by XCS’ XAPA source finder (apart from white regions), including both their colour and their description. These are also the region colours supported by XGA. ................................................................. 56

1 Summary of galaxy cluster samples used in this work.
† The full XXL-100-GC sample contains 100 galaxy clusters, but we remove XLSSC504 due to known issues with its XXL-measured X-ray properties.
‡ The full LoCuSS High-$L_X$ sample contains 50 clusters, 32 have published $XMM$ properties. We analyse clusters that have had $XMM$ observations since then. . . . 98

2 Summaries of the $XGA$ 1D radial models used in this work (though others are implemented and available). Model names and descriptions of their use are included. The units of model parameters are given, as well as the start parameters used for initial fits. Finally details of the parameter priors used for the full MCMC fit are given, where $\mathcal{U}[A, B]$ indicates a uniform distribution with limits $A$ and $B$. ................................................................. 116

1 Summary of the samples defined in this work. $N_{cl}$ is the number of clusters, $N_{T_eFEDS,500kpc}$ the number with eFEDS $T_{500kpc}$ values, and $N_{T_XCS,500kpc}$ the number with XCS $T_{500kpc}$ values. ................................................................. 164

2 The normalisation, slope, and intrinsic scatter values of the fitted temperature calibration models for 500 kpc apertures. $A_{TT}$ is normalisation, $B_{TT}$ is slope, and $\sigma_{T_{ROSITA}/T_{XMM}}$ the intrinsic scatter. ................................................................. 180

3 The normalisation, slope, and residual scatter values from the LIRA fits of the different datasets, for the $L_{X,0.5-2.0}^{500kpc}$ scaling relation. ................................................................. 182
1. eFEDS-XMM galaxy cluster candidates excluded from further analysis due to one or more XMM-Newton data quality issues. The EL and DL columns correspond to the extent likelihood (EXT_LIKE) and detection likelihood (DET_LIKE) columns in the eFEDS catalogue.

† indicates that the candidate was present in the optically confirmed sample from Klein et al. (2021).

2. eFEDS-XMM galaxy cluster candidates classed as contaminants during our visual inspection of XMM, eROSITA, and SDSS images. The EL and DL columns correspond to the extent likelihood (EXT_LIKE) and detection likelihood (DET_LIKE) columns in the eFEDS catalogue.

† indicates that the candidate was present in the optically confirmed sample from Klein et al. (2021).

3. eFEDS-XMM galaxy cluster candidates which appear to be galaxy clusters whose X-ray emission is significantly contaminated by another source. The EL and DL columns correspond to extent (EXT_LIKE) and detection likelihood (DET_LIKE) columns in the eFEDS catalogue.

† indicates that the candidate was present in the optically confirmed sample from Klein et al. (2021).

4. The XMM data used in the analysis of the eFEDS-XCS sample, individual clusters denoted by their unique eFEDS ID. ObsID contains the unique identifier(s) of the XMM observation(s) used. T denotes true, F denotes false, - denotes that either no successful spectral fit was performed, or the data for that cameras was not available. Columns with a subscript A (e.g. PN\_A) indicate whether that instrument is available for an ObsID. Columns with a subscript radius (e.g. PN\_500kpc) indicate whether that instrument’s data contributed to the final XSPEC fit from which we extract temperature and luminosity information.

5. eFEDS-XCS galaxy cluster XGA measured values, RA, Dec, and redshift are taken from the eFEDS X-ray cluster candidate catalogue. $T_{x,500kpc}^{XGA}$ are temperatures within 500 kpc apertures, given in keV. $L_{x,500kpc}^{XGA,52}$ and $L_{x,500kpc}^{XGA,bol}$ are 0.5-2.0keV and bolometric luminosities within a 500 kpc apertures, in units of $10^{43}$ erg s$^{-1}$. All uncertainties calculated from 68% confidence limits, equivalent to $1\sigma$. 

259

260

261

263

265
List of Figures

1.1 A visualisation of dark matter filamentary structure created from an N-body dark matter only simulation by Sousbie et al. (2011). This is a 250×250×20 $h^{-1}$ Mpc slice of a 5123 particles of a 250 $h^{-3}$ Mpc$^3$ large simulation. It highlights the cosmic web structure formed through hierarchical collapse of the primordial density field. ................................................................. 5

1.2 A figure containing two halo mass functions generated using the same cosmological parameters ($\Omega_m = 0.318$ and $\sigma_8 = 0.803$), but at two different redshifts ($z = 0$ - top red line, $z = 1$ - bottom black line). The halo mass functions were generated using CosmoSIS (Zuntz et al., 2015). This demonstrates that the same cosmology creates halo mass functions with significant differences at different redshifts. ................................................................. 13

1.3 A figure containing two halo mass functions, both generated at $z = 1$, with $\Omega_m = 0.318$, and $\sigma_8 = 0.835$, but with different values of $w$, the dark energy equation-of-state parameter. The bottom, black, line is for $w = -1$ (the concordance value) and the top, red, line is for $w = -0.5$. ................................................................. 15

1.4 On-axis effective area curves for the eROSITA telescope modules, original Chandra ACIS-I calibration, 2020 Chandra ACIS-I calibration, Chandra HRC-I, combined XMM instruments (EPN, EMOS1, EMOS2), and the ROSAT PSPC (Predehl et al., 2021). On-axis means that these effective areas are for the central part of the instruments, which are the most sensitive. The curves illustrate how the sensitivity of each instrument depends on the energy of the photon. ................................................... 17

1.5 The chip configuration and position of the three EPIC cameras on the XMM-Newton X-ray telescope, for observation 0863401401. This is as of 17/01/2021, as MOS1 has sustained significant damage to two of its chips, in two separate micro-meteoroid impacts; at launch the middle and right figures show identical chip configurations. ................................................................. 19
1.6 Average field-of-view (FoV) grasp for the *eROSITA* telescope modules, original *Chandra* ACIS-I calibration, 2020 *Chandra* ACIS-I calibration, *Chandra* HRC-I, combined *XMM* instruments (EPN, EMOS1, EMOS2), and the *ROSAT* PSPC (Predehl et al., 2021). Grasp is defined as the effective area multiplied by the FoV area, and is a measure of the combination of how much of the sky the FoV covers and also how sensitive the instrument is. .................................................. 20

1.7 The sky coverage of XCS, DES, LSST, and German eRASS surveys. Each survey, as well as intersections between different surveys, are highlighted by a different colour in the legend at the bottom of the figure. This figure was created by Reese Wilkinson. ................................................................. 29

1.8 A demonstration of the effect of periods of high flare activity on hard-band (12-15 keV) EPIC-PN light curves. The right hand side figure shows the full light curve, with several severe peaks in the number of high energy events being registered. The left hand side figure shows the light curve once the XCS cleaning procedure has been applied; the periods of flaring have been removed, leaving a much more stable lightcurve. .................................................. 30

1.9 Locations of galaxy cluster candidates located by the redMaPPer algorithm. Red points indicate a candidate in the SDSS DR8 redMaPPer public catalogue. Blue points indicate candidates in the DESY3 redMaPPer catalogue. Part of the SDSS catalogue is obscured by part of the DES catalogue. .............................. 37

2.1 Demonstration of the view methods of the RateMap and Spectrum classes, when applied to the Abell 907 galaxy cluster. Data from the *XMM* EPIC-PN instrument of 0404910601 is used. *Left*: A count-rate map with a mask that removes contaminant sources (using XCS region information) and applies an $R_{500}$ aperture. *Right*: A spectrum generated for the $R_{500}$ region with contaminants removed, and fit with an absorbed plasma emission model using XSPEC. ......................... 46

2.2 A flowchart giving a brief overview of the XGA workflow. ......................... 48

2.3 An XGA generated RateMap visualisation of the Abell 478 galaxy cluster. A mask has been applied to remove all contaminating sources (any region that has not been matched to the source being analysed). The mask also applies a circular analysis aperture. The cross-hairs indicate the user supplied coordinates of the cluster. A large part of the cluster’s emission near the centre of the image has been spuriously removed, something that is corrected by checking for extended source regions that intersect with initially matched source regions. .............................. 57
2.4 A masking array generated by an XGA RateMap generated for the EPN instrument of the XMM observation 0201903501. Black parts of the mask indicate areas set to zero (which means those pixels will be masked when this is applied to image data), white parts indicate areas set to one (meaning nothing will change when this is applied to image data).

2.5 A RateMap view (zoomed-in) before (on the left hand side) and after (on the right hand side) the edge mask was applied. The visualisations are centered on a ‘hot pixel’, with the white dashed circle highlighting its location. Comparing the two visualisations demonstrates that the edge mask has removed the hot pixel.

2.6 Demonstration of the simple peak finding method, run without contaminating source masking for demonstrative purposes. The left hand image shows a full count-rate map of the EPN data of observation 0201901401, for context. The middle image shows the hot pixel highlighted in Figure 2.5, and demonstrates that the simple peak finder run without edge masking selects it as the brightest pixel. The right hand image shows that once the edge mask is applied, the simple peak finder selects a more obviously physical source as the X-ray peak, a bright point source.

2.7 An example of the point clusters found by RateMap’s hierarchical clustering peak finder algorithm. The black crosses indicate the largest point cluster constructed via hierarchical clustering, the blue crosses indicate a secondary point cluster coinciding with the bright point source selected by the simple peak finder. White cross-hair indicates the X-ray peak selected by the algorithm.

2.8 An example grid of XMM PSFs generated and stored by XGA. This demonstrates the complex morphology and spatial variation of the XMM PSF effects. This particular grid of PSFs is for 0.5-2.0 keV band for the PN instrument. PSFs near the centre of this image are taken from close to the aimpoint of the observation.

2.9 An XMM observation with a grid overlaid indicating the PSF bins that the observation was divided up into by XGA. The diamonds indicate the central coordinates of the PSF bins (which is where a realisation of the PSF model is generated). This is a companion to Figure 2.8, and is the PN data for observation 0201903501.
2.10 Two demonstrations of XGA’s image PSF correction. The top left figure shows an observation of the Castor sextuple star system, which is such a bright point source that the emission has been spread over a wide area, as well as into characteristic spokes that can be seen in Figure 2.8 as a feature of XMM’s PSF. The top right figure is the same observation after it has been PSF corrected; most of the emission now lies where we would naively expect it to. The spreading and spoking effects have been almost entirely eliminated. The bottom left figure shows the Abell 907 galaxy cluster, with contaminating sources masked (the mask was applied after PSF correction). The bottom right figure shows the image after PSF correction, with the core of the cluster more concentrated and the point source on the left more completely removed after masking.

2.11 An XGA visualisation of a spectrum produced and fit by the XGA SAS and XSPEC interfaces respectively. This is an example of a galaxy cluster spectrum, and is the PN data of observation 0693010301. A model with two main components, a plasma emission model and a hydrogen column absorption model, is fit to the data, with the best fit line plotted in blue.

2.12 An example visualisation of a 3D annular spectrum visualisation produced by an XGA AnnularSpectra instance. The data are from one specific ObsID-Instrument combination, the PN observation of 0652010401. The spectra have been fit with an absorbed plasma emission model, and the fit lines added to the figure.

3.1 Histograms of properties of samples used in this work, taken from literature. X-ray properties presented in this figure have been measured in different regions, indicated in the legend of each plot. a) Intra-cluster medium average temperatures, b) cluster redshifts from spectroscopic follow-up (XXL-100-GC), from the RASS catalogues the sample was selected from (LoCuSS High-$L_X$), and from the redMaPPer cluster finder (SDSSRM-XCS), c) bolometric X-ray luminosities taken from literature, measured within $R_{500}$ (LoCuSS High-$L_X$ and SDSSRM-XCS) and 300 kpc (XXL-100-GC).

3.7 A comparison of global, core-excised, X-ray temperatures and luminosities measured by XGA and LoCuSS analyses, for a subset of the LoCuSS High-$L_X$ sample. a) Shows the temperature values measured within the 0.15-$R_{500}$ region, and b) shows the comparison for full $R_{500}$ bolometric luminosity measurements.
3.8 A figure demonstrating the image PSF correction capabilities of XGA. The left hand side shows a zoomed view of the original combined RateMap (stacked images from multiple observations and instruments divided by the equivalent stacked exposure maps). The right hand side shows the same zoomed view of the PSF-corrected stacked RateMap, where the individual images were all corrected separately and then stacked, then divided by the same combined exposure maps.

3.9 This figure shows visualisations of profiles generated by XGA for a galaxy cluster from the SDSSRM-XCS sample, SDSSRMXCS-134. The left hand side shows a surface brightness profile, generated from a combined image in the 0.5-2.0 keV energy band. It has been fitted with a double beta model. The right hand side shows the density profile generated from the surface brightness profile, with a simplified Vikhlinin density model fitted.

3.10 A one-to-one comparison of gas masses measured for the XXL-100-GC cluster sample, within $R_{500}$, by Eckert et al. (2016) and an XGA reanalysis. $S_B$ profiles were fitted with Beta profiles, density profiles with King profiles. Contains measurements for 91 of 96 XXL-100-GC clusters analysed with XGA.

3.11 A one-to-one comparison of gas masses measured for the LoCuSS High-$L_X$ cluster sample, within $R_{2500}$ and $R_{500}$, by Martino et al. (2014) and an XGA reanalysis.

3.12 Set of annular spectra fitted folded models for SDSSRMXCS-134, performed by XSPEC, are shown as solid lines. Spectral data points are not included for clarity. The radius axis indicates the average of the inner and out radii of the annulus each spectra was measured from, with the zero point being the coordinate defined as the centre of the cluster.

3.13 An example fitted three-dimensional temperature profile generated by XGA for a galaxy cluster in the SDSSRM-XCS sample, SDSSRMXCS-134. This visualisation was created using the view() method of the temperature profile. Data points are shown in black, and the simplified Vikhlinin temperature profile model is shown in green. This temperature profile was generated by deprojecting temperatures measured by fitting the annular spectra models shown in Figure 3.12.

3.14 An example hydrostatic mass profile generated for a galaxy cluster in the SDSSRM-XCS sample, SDSSRMXCS-134. This is profile was generated using the model fits to the density profile shown in the right hand side of Figure 3.9, and the temperature profile shown in Figure 3.13.
3.15 A one-to-one comparison of hydrostatic masses measured for the LoCuSS High-$L_X$ cluster sample, within LoCuSS measured $R_{2000}$ and $R_{500}$ values, by Martino et al. (2014) and an XGA reanalysis. ........................................ 132

3.16 A hydrostatic mass to temperature relation constructed using masses measured from 102 clusters from the SDSSRM-XCS sample, shown in blue. A re-fitted scaling relation using the Arnaud et al. (2005) data is shown in orange for comparison. SDSSRM-XCS masses and temperatures are measured within $R_{500}$. Arnaud et al. (2005) temperatures are measured within [0.1-0.5]$R_{200}$. ........................................ 135

3.17 A hydrostatic mass to luminosity relation constructed using masses measured from 102 clusters from the SDSSRM-XCS sample. Masses and luminosities are measured within $R_{500}$. Luminosities are unabsorbed and measured within the soft band (0.5-2.0 keV). ........................................ 136

3.18 SDSSRM-XCS hydrostatic mass to richness relation (in purple) constructed using masses measured from 102 clusters from the SDSSRM-XCS sample, with masses measured within $R_{500}$ and richnesses are taken from the SDSS redMaPPer catalogue. A comparison relation re-fitted from masses and richnesses presented by Andreon and Congdon (2014); these masses are derived from weak lensing and both masses and richnesses are measured within 0.5 Mpc apertures. ................................. 138

3.19 A figure showing the distribution of ACT DR5 SZ selected galaxy clusters on the sky. Each cluster is represented by a point, with black points indicating clusters that are in the ACTDR5-XMM sample, and red indicating clusters that have no XMM data. ........................................ 144

3.20 A figure showing two mass-temperature relations, one generated from ACT DR5 (purple) clusters, and one from DESY3RM (green) clusters. The DESY3RM galaxy clusters used to fit the green scaling relations are unique, in that they do not appear in either the SDSSRM-XCS or ACTDR5-XMM cluster samples. The shaded region of each relation indicates the 90% confidence limits. ................................. 148

3.21 A companion to Figure 3.20, this figure shows the parameter posterior distributions of the scaling relation model that was used to fit the mass-temperature relation. ........................................ 149

3.22 A mass-richness relation for galaxy clusters in the DESY3RM-XCS sample what do not appear in either the SDSSRM-XCS or ACTDR5-XMM samples. ................................. 150
3.23 A preliminary scaling relation between X-ray hydrostatic mass (as measured by XGA using XMM data), and ACT DR5 SZ signal to noise, as published in the DR5 catalogue. No uncertainties for the signal to noise are available, so the scaling relation model was fit using only mass uncertainties.

3.24 A figure showing the 334 unique galaxy cluster masses for clusters from the SDSSRM-XCS, ACTDR5-XMM, and DESY3RM-XCS samples, and how they relate to the X-ray temperature measured within the $R_{500}$ region. No model is fit because selection effects for the three samples will be different enough that it would not be valid (particularly between selection from optical/NIR for the SDSS and DES samples, and SZ for the ACT sample).

3.25 A figure showing the 334 unique galaxy cluster masses for clusters from the SDSSRM-XCS, ACTDR5-XMM, and DESY3RM-XCS samples, and how they relate to the X-ray soft-band (0.5-2.0 keV) luminosity measured within the $R_{500}$ region. No model is fitted because selection effects for the three samples will be different enough that it would not be valid (particularly between selection from optical/NIR for the SDSS and DES samples, and SZ for the ACT sample).

4.1 Redshift, temperature, and fractional temperature error distributions of the eFEDS and XXL-100-GC samples. Redshifts from both samples come from a variety of sources, and temperatures are measured within 300 kpc apertures centered on clusters. There are no clusters in common between the two samples.

4.2 Footprint of eFEDS, given by the black solid line. Cluster candidates present in the eFEDS X-ray catalogue are highlighted by red diamonds. The grey circles highlight XMM observations, with a radius of 15′ (the approximate radius of the XMM FoV).

4.3 Distribution of exposure times for eFEDS-XMM cluster candidates, measured at the eFEDS coordinates. Exposures taken from 0.5-2.0 keV exposure maps, corrected for flaring and vignetting. Dashed line indicates the average vignetting corrected exposure of the eFEDS field reported by Liu et al. (2021a).

4.4 Comparison of eFEDS and XAPA central coordinates, for the subset of the eFEDS-XCS sample that have been detected by XAPA.
4.5 eFEDS-XMM cluster candidate (eFEDS ID 1644) identified as a pair interacting galaxies with ongoing AGN activity (see Section 4.3.3.1). The cross-hair indicates the eFEDS position. Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0822470101), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 750 kpc, HSC image has a half-side-length of 750 kpc (at the redshift provided by eFEDS).

4.6 eFEDS-XMM cluster candidate (eFEDS ID 3334) without an obvious corresponding source of emission (see Section 4.3.3.2). The cross-hair indicates the eFEDS position. Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0822470101), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 500 kpc, HSC image has a smaller half-side-length of 250 kpc (at the redshift provided by eFEDS).

4.7 Two eFEDS-XMM cluster candidates in the outskirts of a low redshift foreground AGN. A spurious eFEDS-XMM cluster candidate (eFEDS ID 8922) is indicated by the cross-hair (see Section 4.3.3.2). An eFEDS-XMM cluster candidate (eFEDS ID 16370) is indicated by the dashed circle. Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0655340133), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 1000 kpc, HSC image has a half-side-length of 1000 kpc (at the redshift for eFEDS ID 8922 provided by eFEDS).

4.8 eFEDS galaxy cluster split into two candidates by the source finder (see Section 4.3.3.3). Cross-hairs indicate one candidate (eFEDS ID 8602), and the white diamond indicates the other (eFEDS ID 1023). Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0761730501), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 800 kpc, HSC image has a half-side-length of 800 kpc (at the redshift provided by eFEDS, which is the same for 8602 and 1023).

4.9 A low redshift eFEDS-XMM cluster candidate (eFEDS ID 150) whose X-ray emission is dominated by an X-ray bright elliptical galaxy (see Section 4.3.4). Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0673180201), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 100 kpc, HSC image has a half-side-length of 100 kpc (at the redshift provided by eFEDS).
4.10 Comparison of unabsorbed cluster luminosities within a 500 kpc aperture, in the 0.5-2.0 keV energy band, centered on eFEDS coordinates. Pale blue line indicates best fit power-law, with 68% confidence levels given by shaded region. Grey line indicates a power-law fit with slope set to 1 (with 68% confidence levels given by grey shaded region). Cyan diamond is for the split cluster discussed in Appendix B.176

4.11 Temperature and fractional temperature error distributions of the eFEDS (red) and eFEDS-XCS (pale blue) samples, for measurements made within 500 kpc apertures, centered on eFEDS coordinates. The eFEDS sample plotted in red contains 95 eROSITA temperature measurements, and the eFEDS-XCS sample plotted in pale blue contains 28 XMM temperature measurements. 177

4.12 Comparison of eFEDS and XCS cluster temperatures within 500 kpc, centered on eFEDS coordinates. Pale blue line indicates best fit power-law (slope free to vary), with 68% confidence levels given by shaded region. Grey line indicates a power-law fit with fixed slope of unity (with 68% confidence levels given by the grey shaded region). Cyan diamond is for the split cluster discussed in Appendix B.180

4.13 Soft-band (0.5-2.0 keV) luminosity-temperature relations for the eFEDS and eFEDS-XCS samples. Properties measured within a 500 kpc fixed aperture centered on the eFEDS positions. eFEDS data points are green crosses and the model fit is green, eFEDS-XCS data points are black diamonds with a grey model fit. 183

4.14 Corner plot of the 1σ and 2σ confidence contours of the \( L_{500kpc}^{x,0.5-2.0} - T_{500kpc} \) relation parameters, for the eFEDS (green contours), eFEDS-XCS (grey contours) and eFEDS-XCS calibrated (blue contours) samples. The diagonal shows the posterior densities of each parameter. 184

4.15 Soft-band luminosity-temperature relations for eFEDS, eFEDS-XCS, and calibrated eFEDS-XCS. Properties measured within a 500 kpc fixed aperture centered on the eFEDS positions. 185

4.16 Distribution of XMM observations projected over the sky, indicated by grey points. The eRASS-DE half of the sky is highlighted by the red region. The background is the Planck-DustPol (Padovani et al., 2012) map, available for download on NASA’s Lambda service. 186
5.1 Sets of polygons that map boundaries of XMM detectors for observation 0201903501. The maps are in XMM detector coordinates, one of the internal XMM coordinate systems. Left hand side shows the detector boundary map for MOS2, the right hand side shows it for PN.

5.2 Combined XMM images of an observation where an artificial observation of a simulated galaxy cluster has been added. The left hand side shows the original image, and the right hand side shows the image generated from the event list which has had artificial photons added to it.

5.3 A galaxy cluster luminosity-temperature relation for a sample of real galaxy clusters analysed by XCS (in black, with a purple fit line), and properties measured from artificial XMM observations of a sample of simulated clusters taken from IllustrisTNG. The simulated cluster measurements are shown in blue, with a red fit line.

5.4 A galaxy cluster luminosity-temperature relation for a sample of real galaxy clusters analysed by XCS (in black, with a purple fit line), and properties measured from artificial XMM observations of a sample of simulated clusters taken from the 300 project. The simulated cluster measurements are shown in blue, with a red fit line.

5.5 Example g,r,i composite colour 50′′ x 50′′ SDSS images classified by Pea hunters, with the r-band representing green light. The distinctly green colour and compact morphology makes the Peas (left 3 images) easily distinguishable from the classical elliptical (right image). The elliptical galaxy is clearly red and has a smooth profile, while the Peas are r-band dominated and unresolved in these images, appearing like stellar point sources. All objects shown here are at $z \sim 0.2$, and this figure is taken from Cardamone et al. (2009).

5.6 The top left figure shows an image of a Green Pea galaxy, taken by the SDSS photometric survey, and the top right figure is the corresponding spectrum for that Green Pea galaxy, taken with the BOSS spectrograph. The bottom left figure shows an image of a Purple Pea galaxy (at lower redshift than the Green Pea), taken by the SDSS photometric survey, and the bottom right figure is the corresponding spectrum for that Purple Pea galaxy, taken with the SDSS spectrograph. Both spectra have filter curves overlaid to indicate which filter then emission lines fall within.
Comparisons of photometric redshifts and spectroscopic redshifts measured by SDSS. The left hand side shows the comparison for a sample of Green Pea galaxies selected from SDSS, and the right hand side shows the comparison for a sample of randomly selected galaxies in the same redshift range as the Green Peas.

5.8 Colour-colour plots for randomly drawn galaxies and QSOs, as well as the original sample of Green Peas presented by (Cardamone et al., 2009), this plot recreates Figure 2 of their paper. The left hand side figure shows the $g - r - r - i$ colour space, with the black line indicating the boundary chosen by the original paper, and the right hand side figure shows the same but for the $u - r - r - z$ colour space.

5.9 A corner plot showing the feature space used to train a Green Pea classifying SVM. The blue points and histograms represent the training set of Green Peas, whereas the orange points and histograms represent random galaxies drawn from the same redshift range.

5.10 This figure shows the hyperplane in Figure 5.9 projected into two dimensions using the UMAP (McInnes et al., 2018) dimensionality reduction technique. Once again the blue points represent Green Peas, and the orange points represent randomly selected galaxies from the same redshift range.

5.11 The confusion matrix for a two-class SVM classifier trained to separate Green Peas and random galaxies drawn from SDSS. This confusion matrix was generated from the testing set, and as such the data were not used during training.

5.12 The confusion matrix for a multi-Pea SVM classifier trained to separate Pea galaxies from other galaxies drawn randomly from the same redshift range as each Pea type. The classifier is ignorant of the redshifts of the galaxies passed into it, though redshift is used in the construction of the training samples. This confusion matrix was generated from the testing set, and as such the data were not used during training.

5.13 A figure showing the number density of LoTSS DR2 radio sources per square degree. The LoTSS DR2 catalogue was constructed from two separate observation fields, which are evident in this Mollweide projection sky plot. LoTSS DR2 source positions were binned using HEALPix to calculate source density on the sky, with the number of sides set to 64. Dark blue areas indicate that no LoTSS DR2 sources are located in that part of the sky.
5.14 A figure showing the locations of LoTSS DR2 radio sources that have been found to fall on at least one XMM observation. They do not necessarily have corresponding detections in the XCS master source list.
Preface

Chapter 1 is made up of background information that is entirely review material, not original work.

Chapter 2 is entirely original work, with Section 2.1 uploaded to the arXiv as a software paper (Turner et al., 2022). Other parts of Chapter 2 are based upon software documentation that is available online, and again is entirely original work.

Chapter 3 is also original work, with the first part to be submitted as a paper both to arXiv and the Monthly Notices of the Royal Astronomical Society Turner et al. (prepa). Comments and proof-reading were provided by co-authors, but the paper was written by me. The second part of this chapter will eventually form part of a publication, but not in the immediate future, and is entirely my own work.

Chapter 4 has been submitted to the Monthly Notices of the Royal Astronomical Society as a paper, and is also released on the arXiv (Turner et al., 2021). Again, this is entirely my own work, though comments and proof-reading were provided by co-authors.

Finally, Chapter 5 presents brief discussions of other projects I’ve undertaken that are not yet near publication, and ends with a conclusions and a discussion of the work in this thesis; all original work.
Chapter 1: Introduction

1.1 Motivation and aims

The main aim of this PhD was to measure masses of clusters of galaxies in order to construct scaling relations, calibrating the relation normalisation and the scatter of mass with observable. These masses were to be derived from archival XMM-Newton X-ray telescope data, and provide independent mass calibrations for large cluster surveys such as the Dark Energy Survey (DES) and the Legacy Survey of Space and Time (LSST). Independent mass measurements are particularly valuable to large optical/near-infrared surveys, as the method of mass measurement most often used (stacked weak-lensing) cannot provide information on the scatter of cluster masses with different observables.

A secondary goal of this work was to be sustainable and useful in the long term. As observations continue to be taken, the number of masses that can be measured using XMM will grow, thus any tools developed to enable this research had to be maintainable and usable past the end of this PhD.

The final aim was to perform preliminary work in preparation for the full release of the eROSITA All-Sky Survey, an X-ray survey of the full sky with the recently launched eROSITA telescope. This survey will provide the largest X-ray selected galaxy cluster (and indeed any X-ray source) catalogues ever produced, and will allow for cutting-edge X-ray cluster cosmology analyses. As it is being performed with a brand new telescope however, external validation is required to fully understand the limitations of the survey.

1.2 Galaxy Clusters

Galaxy clusters are the largest gravitationally bound structures in the Universe, and are made up of three main components; a host dark matter halo, the hot, diffuse, gas of the intra-cluster medium,
and the component galaxies. This makes them uniquely powerful laboratories for various different astrophysical studies.

### 1.2.1 Formation of Galaxy Clusters

Galaxy clusters form at the nodes of dark matter filamentary structure due to the collapse of peaks in the primordial density field, making them important parts of the local large scale structure. The formation process begins at very high redshift, with the (commonly assumed to be Gaussian) primordial density field initially undergoing a linear evolution where all scales of dark matter structure evolve independently and at the same rate, described by the linear growth factor. This theoretical approach to the growth of structure in the Universe is relatively simple, and does neglect some of the effects that neutrinos, for instance, can have on structure formation. Linear evolution is not thought to hold for the entire age of the Universe however, and theoretical models for non-linear structure evolution are still an area of intense study. The linear assumption breaks down once the general amplitude of fluctuations in the density field reaches a point where a small region of the Universe can be substantially denser than the average. Those slightly denser points in the initial density field will always become denser in contrast to the average, as they will expand slower than the less dense areas. At this point the process of forming objects has truly begun, with overdensities becoming dark matter halos that may later play host to individual galaxies, or even to a collection of galaxies, a galaxy cluster.

A simple, commonly used, theoretical framework for this stage of structure formation is the ‘spherical collapse model’, which assumes that a density peak can be described as a constant overdensity spherical perturbation with some some characteristic radius $R$. In the spherical collapse model $R$ will initially increase as the Universe expands, as the initial density peak is small, but the radius growth will slow and eventually reach a turn-around point where it starts to collapse due to the excess mass enclosed by the radius. This region is overly dense compared to the rest of the Universe. At this point the overdensity has decoupled from the Hubble flow. Non-linear evolution of structure is much faster than linear evolution, highlighted by the prediction that, if the density contrast would linearly grow to 1.69, the region would collapse to a virialised object with an over-density of 178 in the spherical collapse model (Mead, 2017). Any regions less over-dense than this will not collapse under this model.
1.2.2 Composition of Galaxy Clusters

Galaxy clusters are complex objects, with several different components, and often complicated interactions between them. This section will briefly overview what makes up a galaxy cluster, as well as the mechanisms of photon emission associated with it (if any).

1.2.2.1 Dark Matter

Dark matter is an inferred type of matter that only interacts through gravity, not through electromagnetic waves, and is currently thought to make up $\sim 87\%$ of the mass of a typical galaxy cluster (Gonzalez et al., 2007). The concordance cosmology model of $\Lambda$-CDM assumes that dark matter velocities are much slower than the speed of light, that it is collisionless, and that it does not radiate photons of any kind. As such, dark matter cannot be directly observed by studying the electromagnetic spectrum, rather its presence has been inferred through observations of objects that do not fit with predictions based on our current understanding of the Universe.

In a Universe with cold dark matter the process of structure formation briefly described in Section 1.2.1 is expected to result in a ‘cosmic web’ where the dark matter comes to exist in a system of filaments and nodes. The nodes of the cosmic web are where the dark matter halos that host galaxy clusters form, which in turn contain separate overdense regions called sub-halos, where the member galaxies of the cluster form. A visualisation of a cosmic web derived from simulations is shown in Figure 1.1.

The dark matter component of clusters of galaxies can be inferred from the effect of gravity on the emission of background galaxies, a process called ‘gravitational lensing’. This is where the mass distribution of the galaxy cluster causes a shearing, or change of shape, of galaxies that are at a higher redshift but when projected on the sky appear to be near the cluster in question; the gravitational potential of the cluster having changed the path that the photons emitted from those galaxies follow. Gravitational lensing is a prediction of the general relativity theory of gravity, and its dependence on the mass of the cluster shows us that there is far more mass in a galaxy cluster than can be accounted for by the observable baryonic components. A requirement for more mass than can be directly observed is also evident in the velocity dispersion of galaxy clusters; indeed arguably the first evidence for dark matter (Zwicky, 1933) was the orbits of galaxies within the Coma cluster. Studies of the rotation curves of gas and stars within galaxies found discrepancies that also hinted at there being more mass than is directly observable (Rubin et al., 1980).
Figure 1.1: A visualisation of dark matter filamentary structure created from an N-body dark matter only simulation by Sousbie et al. (2011). This is a 250×250×20 $h^{-1}$ Mpc slice of a 5123 particles of a 250 $h^{-3}$ Mpc$^3$ large simulation. It highlights the cosmic web structure formed through hierarchical collapse of the primordial density field.
The nature of cold dark matter is still a mystery, with theories ranging from extensions to the standard model of particle physics that introduce weakly interacting massive particles, to theoretical axion-like particles; Arbey and Mahmoudi (2021) give an overview of many current theories. Progress continues to be made in ruling out possible theories of dark matter, with recent work by Bhargava et al. (2020) effectively discarding the theory of sterile neutrino dark matter, as the predicted X-ray emission line at 3.5 keV was found to be absent in the vast majority of galaxy clusters.

Though it is the currently accepted model, cold dark matter is not the only flavour of dark matter being studied. One alternative theory is ‘hot dark matter’ (Primack and Gross, 2001), where dark matter particles move at velocities very close to the speed of light, with those particles having considerably lower mass than posited cold dark matter particles (though this theory fails to allow small scale structure formation). Another possibility is warm dark matter (Viel et al., 2013), which has properties in between the cold and hot dark matters (one of the particle candidates are sterile neutrinos, which have been effectively ruled out). Modifications to cold dark matter also exist, with ‘fuzzy cold dark matter’ (Mocz et al., 2019) theorised to consist of very light particles; this type of dark matter would display wave like behaviours on the scale of galaxies, causing interference patterns in large scale structure.

A dark matter like particle which only interacts gravitationally is not universally accepted however, and there is much research into possible alternatives to an invisible mass component of the Universe. These tend to take the form of modifications to the theory of general relativity; rather than inferring that there is more mass than is visible from our observations, these works take the approach that our understanding of gravity itself is flawed. Inevitably these theories have to contend with the fact that general relativity provides an accurate description of gravity in almost every circumstance, and is probably one of the most thoroughly tested theories of modern physics. Just as with dark matter, there are different flavours of theory that try to explain our observations. These include chameleon gravity (Wilcox et al., 2015; Tamosiunas et al., 2021; Tamosiunas, 2020), which introduces a scalar field coupled to the matter components of the Universe, giving rise to a fifth force; this fifth force is ‘screened’ in deep potential wells, allowing gravity to behave as predicted by general relativity. Another theory is that of ‘emergent gravity’ (Jacobson, 1995), which demonstrates that general relativity resembles thermodynamics, and as such might mean that space-time is made up of small elements, and that their collective motion gives rise to gravity. This would make gravity an ‘emergent’ property, not a fundamental force. There is also a theoretical prediction that a elements of space-time would push inwards on matter, accounting for the
additional mass that we currently take to be dark matter (Verlinde, 2017).

1.2.2.2 The Intra-Cluster Medium (ICM)

The intra-cluster medium (ICM) is a hot, diffuse, plasma that fills the space between the member galaxies of the galaxy cluster, making up ∼7% of the mass of a typical cluster (Gonzalez et al., 2007). The gas that makes up the content of the ICM is drawn into the galaxy cluster potential well as the cluster forms, heating up to typical temperatures of between ∼1–14 × 10^7 K (or ∼1–12 keV, see Giles et al. (2022b) for a large set of recent temperature measurements; it is more typical to refer to the temperature of the ICM in terms of its energy $k_B T$). The average baryon density of the intra-cluster medium is of order ∼10^{-6} M_⊙kpc^{-3} (∼10^{-35} kg m^{-3}; calculated from a gas mass of ∼1 × 10^{13} M_⊙ within a radius of ∼1200 kpc; Mantz et al., 2022).

The hot, ionised, state of the ICM causes it to emit X-ray photons, with two main mechanisms of emission. The first mechanism is ‘thermal bremsstrahlung’ (braking radiation), which results in a smooth continuum emission from the ICM. When references are made to measuring the X-ray temperature (commonly styled $T_X$) of a galaxy cluster, the plasma temperature (whether it be a global average across the cluster or a more localised temperature) that feeds thermal bremsstrahlung is what is being measured (or rather a density-weighted temperature). Thermal bremsstrahlung is also called ‘free-free emission’, with this process occurring when a free electron deflects off of a charged ion and decelerates, with the energy of the interaction conserved by the emission of a high-energy photon; the ‘free-free’ name for this process refers to the fact that the electron is not captured by the ion, and remains free after the interaction. When this process occurs in a system of electrons and ions that are in local thermodynamic equilibrium, the volume emissivity (emitted energy per time per volume, at frequency $\nu$) of an ion with atomic number (charge) $Z$ in a plasma with temperature $T$ is defined as

$$\epsilon_{\nu} = \frac{2^5 \pi e^6}{3 m_e c^3} \left( \frac{2 \pi}{3 m_e k_B} \right)^{1/2} Z^2 n_e n_i g_{ff}(Z, T, \nu) T^{1/2} e^{-\hbar \nu/k_B T},$$

(1.1)

where $e$ is the fundamental charge of an electron, $m_e$ is the mass of an electron, $k_B$ is the Boltzmann constant, $Z$ is the ion atomic number, $n_e$ is the electron number density, $n_i$ is the number density of the ion, $g_{ff}$ is the Gaunt factor, $T$ is the plasma temperature, $h$ is the Planck constant, and $\nu$ is the frequency of emission (Sarazin, 1986). The thermal bremsstrahlung emission is thus strongly dependant upon the electron and ion densities, and has a weaker dependence on the temperature of the plasma. The composition of the plasma (in terms of densities of different
ions) is obviously also important to the overall emission observed from a complex plasma, though in that regard the intra-cluster medium is often assumed to only consist of electrons and ionised hydrogen. Metals are present in the ICM however, as evidenced by the second main emission mechanism of the ICM, transition lines of different elements present in the plasma of the ICM. The line emission is contributed to by various processes, chief amongst which are collisional excitation and radiative recombination. The emissivity of line emission due to collisional excitation can be described as

\[
\int e_\nu^\text{line} d\nu = n(X^i)n_e \frac{h^3 \nu \Omega(T) B}{4 \omega_{gs}(X^i)} \left( \frac{2}{\pi^3 m_e^3 k_B T} \right)^{1/2} e^{\Delta E/k_B T},
\]

where \( h \nu \) is the transition energy, \( \Delta E \) is the excitation energy above the ground state of the excited level, \( B \) is the probability that the upper state decays through this transition, and \( \Omega \) is the collisional strength. Databases of plasma transition lines for X-ray spectroscopy are available, with the most commonly used being AtomDB (Foster and Heuer, 2020).

The intra-cluster medium does not exclusively emit in the X-ray part of the electromagnetic spectrum, radio emission has been observed from galaxy clusters for decades. The emission is not simply emanating from radio-bright member galaxies (though of course this can happen), but from the intra-cluster medium itself. There are several mechanisms for radio emission from a galaxy cluster, but the most prominent involves interaction between cosmic-ray and magnetic field components of the ICM plasma. High-energy (of order GeV) cosmic ray electrons can interact with the magnetic field of the ICM, emitting synchrotron radiation, though often a triggering event needs to occur to accelerate the cosmic rays to high enough velocities; these events come in the form of ‘shocks’ to the ICM, commonly caused by mergers between galaxy clusters (van Weeren et al., 2019). As such, galaxy clusters that have undergone recent mergers tend to exhibit an accompanying extended radio halo (Cassano et al., 2010).

The temperature and density of the intra-cluster medium are not uniform throughout a galaxy cluster, with both dropping as you move from the centre of the potential well radially outward. For some clusters (known as ‘cool cores’), the temperature can also be lower in the centre of the halo, resulting in a temperature profile that rises and then falls as radius increases. This drop off in density and temperature in the outskirts of clusters is intuitive, as matter is gravitationally drawn into the cluster it will gain kinetic energy, which will become thermal energy when collisions with other elements of gas occur. The drop off in density with radius is a result of the hierarchical structure formation of the Universe. Of course, galaxy clusters do not form in a specific thermodynamic
state and then stay that way forever, the dynamics of the intra-cluster medium are complex and not completely understood. The ICM can reduce its temperature through cooling mechanisms, primarily radiative cooling. In some cases the cooling time at the centre of the cluster is much less than the Hubble time, and as such the gas in the core has had time to cool significantly; these cool core clusters are also correspondingly denser in the core, though the temperatures are still higher than expected from radiative cooling predictions, meaning that some form of heating must be present.

1.2.2.3 Galaxies

Obviously, one of the key components of a galaxy cluster are the component galaxies themselves; for a typical cluster they make up $\sim 3\%$ of the mass. Galaxies in a cluster are a distinct population from galaxies that are not (‘field’ galaxies), as clusters tend to play host to older elliptical galaxies, rather than younger galaxies with more complex morphologies. In fact this gives rise to one method of locating galaxy clusters in optical and near-infrared survey, searching for the so called ‘red sequence’ of galaxies (see Section 1.6.1). Galaxy clusters also typically have a massive elliptical galaxy close to the centre of the potential well, the Brightest Cluster Galaxy (BCG). These galaxies tend to form early in the cluster history through the mergers of various massive elliptical galaxies, and are some of the most massive galaxies in the Universe, though typically are not actively forming stars. They also often have a surrounding stellar population, a significant number of stars that have been removed from their original host galaxies. This can cause detectable optical emission from around the BCG called intra-cluster light (ICL), and which has been shown to be an excellent proxy for the overall mass of the cluster (Sampaio-Santos et al., 2021).

Galaxy clusters can gather more galaxies than they were formed with through the hierarchical evolution of structure, by accreting them through gravitational attraction. Through the use of N-body simulations of dark matter halos (Berrier et al., 2009, and associating dark matter sub-halos with galaxies) demonstrated that the galaxy assembly of clusters is dominated by the infalling of isolated galaxies and groups of galaxies. Such infalling galaxies will orbit within the cluster, with their orbits depending on the mass and redshift of the host halo (Wetzel, 2011). This process of accreting galaxies also gives rise to a ‘splashback’ feature that presents an alternative characterisation of cluster radius versus the traditional overdensity radius definition which comes from non-linear evolution of structure. The infalling galaxies that are trapped by the cluster populate a shell at the apocenter (the furthest point from the centre of the potential well) of their first orbit. This will cause a spike in density at that characteristic radius, which can be searched for (Chang
et al., 2018; Bianconi et al., 2021). The splashback radius is also thought to be a possible probe for modified gravity (Adhikari et al., 2018), as transitions from screened to non-screened gravity are predicted to occur in the cluster outskirts.

The galaxies and the overall cluster have a reciprocal relationship, with the cluster affecting the development and evolution of the galaxies, and certain galaxies having a measurable effect on the galaxy cluster itself (particularly the intra-cluster medium, see Section 1.2.2.2). For instance, the interaction of the intra-cluster medium with infalling galaxies can cause an effect known as ‘ram pressure stripping’ (Gunn and Gott, 1972), which was originally proposed to explain an absence of gas-rich galaxies in galaxy clusters. The interaction of the ICM with the gas in the galaxy can overcome the gravitational attraction keeping the gas in the galaxy, thus stripping it and resulting in a gas-poor galaxy. This makes it a significant evolutionary process for infalling galaxies (Hester, 2006). Gas removed from galaxies in this fashion is also a possible source for the metals that are observed to be present in the ICM (Domainko et al., 2004). Ram pressure stripping is not the only mechanism by which gas can be removed from a galaxy by a galaxy cluster; gas from a member galaxy can be gravitationally drawn out by the potential of its host cluster, a process called tidal stripping (Merritt, 1983). Some galaxies can also have quite extreme effects on their host cluster, with feedback from active galactic nuclei playing a significant role in the evolution of the intra-cluster medium within a cluster. AGN can inject large amounts of energy into the intra-cluster medium, offsetting cooling processes and altering the evolution of the ICM (Gitti et al., 2012).

### 1.2.3 Galaxy Cluster Scaling Relations

Some properties of galaxy clusters are closely correlated with one another, which is both predicted from theoretical work on the physics of the ICM and shown empirically through observations. Such correlations allow for the construction of useful ‘scaling relations’, which provide a predictive link between two properties of a galaxy cluster (for instance X-ray luminosity and temperature); such relations are useful both for testing the theoretical frameworks that predicted them, and as a method of predicting difficult-to-measure quantities (such as the mass of a galaxy cluster) from easier-to-measure quantities (such as the cluster’s X-ray luminosity). Mass-observable relations, for instance, are almost indispensable for the measurement of cosmological parameters from galaxy clusters using the halo mass function (see Section 1.2.4).

The key assumption for those scaling relations that can be derived from theory is that of ‘self-similarity’, first proposed by Kaiser (1986); the idea that galaxy clusters are functionally identical
objects when scaled by their mass (i.e. one galaxy cluster is just a larger or smaller version of another galaxy cluster). Whether clusters are truly self similar is still being studied, with some work showing that scaling relation slopes obey self-similarity (Sereno and Ettori, 2015), and others finding significant departures from self-similarity (Maughan et al., 2012; Giodini et al., 2013).

Finding significant departures from self-similarity could indicate that astrophysical processes such as AGN feedback, magnetic fields, and cosmic rays are contributing significantly to the overall energy budget of the cluster. There is evidence that such processes are more impactful in the low-mass (i.e. galaxy group) regime Lovisari et al. (2021). When constructing relations one also needs to be mindful of selection effects, where we preferentially locate particular types of galaxy clusters, or objects above a certain brightness. For instance Eckert et al. (2011) showed that X-ray selected samples will be biased towards selecting relaxed clusters with bright cores, and O’Sullivan et al. (2017) found that $\sim 20\%$ of X-ray bright groups in the local Universe may have been missed.

Self-similarity assumes that the ICM is adiabatic, which it is not, and that clusters form from a single gravitational collapse, which is also not true. However, the self-similar model allows for the derivation of simple power-law scaling relations which can be compared to power-laws fitted from observed properties. By using the average kinetic energy of the particles in the intracluster medium, under the assumptions necessary for the self-similar model, and combined with the virial theorem it is possible to derive that,

$$M \propto E(z)^{-1} T^{3/2}.$$  \hspace{1cm} (1.3)

Where $M$ is the total mass of the cluster, $T$ is the temperature of the intra-cluster medium, and $E(z)$ is the dimensionless Hubble parameter. The Hubble parameter is defined in terms of cosmological parameters,

$$E(z) = \frac{H(z)}{H_0} = \sqrt{\Omega_r (1+z)^4 + \Omega_m (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda},$$ \hspace{1cm} (1.4)

As such it is possible to measure both the masses and temperatures of galaxy clusters, construct scaling relations in this form by multiplying the masses of the clusters by the $E(z)$ calculated at each cluster’s redshift, and then compare the slope of the measure scaling relation with the theoretical prediction of $1.5$ from Equation 1.3.

A similar process can be used to derive a relation between the X-ray luminosity of a system, and its total mass; in this case the prediction has a different dependence on $E(z)$ and predicts a different
slope for the relation. In this case the relation

\[ L_X \propto E(z)^{4/3} M^{4/3}, \]  

(1.5)

and its inverse can be used to predict the mass of a cluster from its luminosity, a relatively simple X-ray quantity to measure. That makes this relation particularly important for cluster cosmology using X-ray data, as luminosity measurements are much easier to acquire at scale than direct mass measurements.

Finally, and in the same manner, a relation between the luminosity and the temperature of a cluster can be derived:

\[ L_X \propto E(z) T^2, \]  

(1.6)

providing yet another avenue for testing the assumption of self-similarity. This is a commonly measured scaling relation (e.g. Turner et al., 2021; Giles et al., 2022b, 2016), with some evidence of departure from self-similarity reported.

There also exist scaling relations that are not derived from self-similarity, but instead are empirically observed, or predicted through physical intuition. The relation between galaxy cluster mass and richness (a probabilistic measure of the number of member galaxies in a cluster) is one example. Logically the number of galaxies that are bound to a cluster should be related to its mass, and the mass-richness relation has become a key part of cluster cosmology efforts where the clusters are selected from optical/near-infrared surveys (e.g. the Sloan Digital Sky Survey and the Dark Energy Survey).

### 1.2.4 Cosmology with Galaxy Clusters

As well as acting as powerful astrophysical laboratories, galaxy clusters can also be used to investigate several aspects of cosmology. This includes the derivation of cosmological parameters such as \( \Omega_m \) and \( \sigma_8 \) (Vikhlinin et al., 2009; DES Collaboration et al., 2020; Mantz et al., 2022; Clerc and Finoguenov, 2022), measurement of the Hubble parameter (Wan et al., 2021), and the investigation of different types of modified gravity theory (Wilcox et al., 2015; Tamosiunas, 2020; Tamosiunas et al., 2021).
Figure 1.2: A figure containing two halo mass functions generated using the same cosmological parameters ($\Omega_m = 0.318$ and $\sigma_8 = 0.803$), but at two different redshifts ($z = 0$ - top red line, $z = 1$ - bottom black line). The halo mass functions were generated using CosmoSIS (Zuntz et al., 2015). This demonstrates that the same cosmology creates halo mass functions with significant differences at different redshifts.
Different methods of measuring cosmological parameters with galaxy clusters exist, with the construction of a galaxy cluster halo mass function (HMF) being a common approach. A halo mass function is a measure of the number density of ‘halos’ (i.e. clusters; see Section 1.2.1) for a given mass range, and often for a given redshift range (as the HMF evolves with redshift). As the number density of galaxy clusters of different masses is sensitive to the formation of large scale structure in the Universe, modelling theoretical halo mass functions and comparing them to measurements provides a way to derive cosmological parameters. An analytical expression for the HMF was derived by Press and Schechter (1974), and is known as the Press-Schecter formalism, however it significantly underestimates the proportion of particles which are ‘locked’ in halos, and so a correcting factor is added.

The Press-Schecter HMF is not often used, as methods of constructing numerical HMFs using simulations have emerged. For instance, Tinker et al. (2008) used a large set of dark matter only cosmological simulations (run using several simulation codes, and with slightly varying cosmological parameters) to construct an analytical HMF for dark matter halos. More recent work has focused on being able to generate halo mass functions for a range of cosmologies by creating ‘emulators’ from suites of dark matter only simulations (McClintock et al., 2019; Bocquet et al., 2020); these suites are sets of cosmological sized simulations that cover a wide cosmogical parameter space. The resulting emulator can then output a HMF for any combination of input parameters, with high precision. Figure 1.2 gives an example of two halo mass functions, generated with the same cosmological parameters, but at two different redshifts, demonstrating the evolution of the HMF; the mass range is appropriate for galaxy groups and clusters, between $10^{13} - 10^{15} M_\odot$. Halo mass functions in different redshift bins can be jointly analysed to put constraints on cosmological parameters.

This method of deriving cosmological parameters allows for constraints to be put on the dark energy equation-of-state parameter ($w$), with Figure 1.3 showing the difference in HMFs with two different values of $w$. The $z = 1$ halo mass function provides the greatest distinction between the two curves, which can inform targeting priorities for observing campaigns.

Another approach to the derivation of cosmological parameters uses measurements of galaxy cluster baryon mass fractions (the ratio of the total mass of the baryons in the ICM to the total mass of the galaxy cluster). As they directly probe the baryon content of a galaxy cluster, and can also be used to measure total masses (see Section 3.4), X-ray observations of galaxy clusters (see Section 1.3.1) are ideal for this sort of analysis (Ettori et al., 2009; Mantz et al., 2022). The ratio of the cosmological baryon density parameter ($\Omega_b$) and the cosmological mass density parameter
Figure 1.3: A figure containing two halo mass functions, both generated at $z = 1$, with $\Omega_m = 0.318$, and $\sigma_8 = 0.835$, but with different values of $w$, the dark energy equation-of-state parameter. The bottom, black, line is for $w = -1$ (the concordance value) and the top, red, line is for $w = -0.5$. 
(\(\Omega_m\)) is easily relatable to the galaxy cluster baryon mass fraction through the use of simulations, requiring a multiplicative ‘gas depletion’ expression. Modelling of hydrostatic mass biases and the gas depletion fraction allow for constraints to be placed on the \(\Omega_b/\Omega_m\) ratio, though not on other parameters such as \(h\) (the dimensionless Hubble constant; Croton, 2013). Marginalising over a realistic prior for \(h\), the evolution of the cluster gas fraction with redshift can be used to constrain \(\Omega_m\) independently of \(\Omega_b\), as well as to constrain \(w\) and \(\Omega_\Lambda\).

1.3 Observations of Galaxy Clusters

The signatures of galaxy clusters can be found in most parts of the electromagnetic spectrum, with each different way of observing telling you something different about the cluster. This section will introduce relevant instruments used for each wavelength, their capabilities, and what part of the galaxy cluster they observe.

1.3.1 X-ray

Observations of galaxy clusters in the X-ray band directly trace the processes of the intra-cluster medium, which produce high-energy photons that can be detected using X-ray telescopes. So although X-ray observation cannot tell us much about the component galaxies of the galaxy cluster, it can tell us a lot about their environment, about the properties of the ICM, and the global properties of the overall halo.

As the Earth’s atmosphere is essentially opaque to X-ray photons, orbital observatories must be constructed to study the X-ray emission of clusters of galaxies. There are several currently flying observatories that are relevant to this work, and their capabilities and strengths will be summarised in the next few sections.

1.3.1.1 XMM-Newton

_XMM-Newton_ (henceforth referred to as _XMM_) is the European Space Agency’s (ESA) X-ray telescope, and has been in service since the end of 1999; as such there is a scientific archive dating back over 20 years, with more data collected and processed every year.

_XMM’s_ X-ray instrumentation is comprised of three separate co-aligned (pointing at the same part of the sky) X-ray telescopes, with each telescope made up of 58 nested Wolter Type-1 mirrors,
Figure 1.4: On-axis effective area curves for the eROSITA telescope modules, original Chandra ACIS-I calibration, 2020 Chandra ACIS-I calibration, Chandra HRC-I, combined XMM instruments (EPN, EMOS1, EMOS2), and the ROSAT PSPC (Predehl et al., 2021). On-axis means that these effective areas are for the central part of the instruments, which are the most sensitive. The curves illustrate how the sensitivity of each instrument depends on the energy of the photon.
and a focal length of 7.5 m. They give XMM a field-of-view (FoV) with a radius of $\sim 15'$ for X-ray observations. One of the telescopes is dedicated entirely to the European Photon Imaging Camera (EPIC)-PN camera, whereas the other two telescopes each host an EPIC-MOS camera and a Reflection Grating Spectrometer (RGS). The RGS instruments allow for high-energy-resolution X-ray spectroscopy as the grating spectrographs disperse photons onto focal plane cameras, though this has the side effect of reducing the number of photons that reach the EPIC-MOS cameras by $\sim 50\%$.

The two main types of camera on XMM (EPIC-PN and EPIC-MOS) are different kinds of Charge Coupled Device (CCD); PN is a back illuminated design, where X-ray photons hit the silicon back of a PN chip, generating electrons and holes whose charge is detected and stored; MOS (which stands for Metal Oxide Semi-conductor) is a front-illuminated design with modifications to make it much more sensitive at lower energies (below 700 eV) than previous front illuminated CCDs. The three EPIC cameras provide an on-axis effective area of over 2000 cm$^2$ at $\sim 1.5$ keV when combined, with less drop-off at higher and lower energies than other X-ray telescopes (see Figure 1.4). Capabilities of the two CCD types are comparable, though the MOS cameras have a smaller ($1.1''$) pixel size than PN ($4.1''$), a slightly smaller point spread function (PSF; full width half maximum of $5''$) than PN ($6''$), and a slightly better energy resolution ($\sim 70$ eV) than PN ($\sim 80$ eV). The PSF width is largely dependant on the optics and support structure, but each camera has a slightly different central PSF shape. On the other hand, PN can be operated in fast timing modes that have significantly better temporal resolution (0.03 ms) to MOS (1.75 ms), and as PN does not have to share its telescope with RGS it is nearly twice as sensitive as MOS in the XMM configuration. Both the PN and MOS cameras are made up of multiple CCD chips, and their spatial position and configuration differs (see Figure 1.5), with PN being made up of 12 rectangular chips and MOS consisting of 7 chips.

XMM also has a fourth telescope called the Optical Monitor (OM), which is capable of observations in the optical and low-energy UV bands. It consists of a 30 cm Ritchey-Chreien telescope with a focal length of $\sim 3.8$ m, two redundant detectors and filter wheels, and total coverage of the 170-650 nm range within a $17'$ square FoV. The filter wheel(s) contain six broad filters (U, B, V, UVW1, UVM2, and UVW2) covering optical and UV wavelengths. They also contain two grisms to enable basic UV and optical spectroscopy, and a magnifier which effectively increases the focal length and gives a higher resolution on the sky.
Figure 1.5: The chip configuration and position of the three EPIC cameras on the XMM-Newton X-ray telescope, for observation 0863401401. This is as of 17/01/2021, as MOS1 has sustained significant damage to two of its chips, in two separate micro-meteoroid impacts; at launch the middle and right figures show identical chip configurations.

1.3.1.2 eROSITA

eROSITA is an X-ray instrument mounted on the joint Russian-German Spectrum-Roentgen-Gamma (SRG or Spektr-RG, Predehl et al., 2021) observatory. It covers a similar energy band (0.2-10 keV) to XMM-Newton (Section 1.3.1.1). A Russian instrument, the Astronomical Roentgen Telescope X-ray Concentrator (ART-XC), is also mounted on SRG but observes harder X-rays (4-30 keV). SRG was launched to the second Sun-Earth Lagrange point (L2) on the 13th of July 2019, using a Russian Proton rocket.

The primary science goal of eROSITA is to complete an all sky survey in the 0.2-10 keV band, the eROSITA All-Sky Survey (eRASS). This is intended as a successor to the ROSAT All-Sky Survey (RASS), which was performed using the Position Sensitive Proportional Counters (PSPC) instruments mounted on the ROSAT (short for Röntgensatellit) X-ray observatory. RASS detected ∼150000 X-ray sources, and eROSITA is considerably more sensitive (see Figure 1.4 for a comparison of effective areas) in a wider X-ray band (the PSPC instruments observed in the 0.1-2.4 keV band). eRASS is expected to detect ∼1 × 10^5 galaxy clusters, ∼3 × 10^6 active galactic nuclei (AGN), and ∼7 × 10^4 local stars.

The eROSITA instrument is made up of seven identical co-aligned ‘telescope modules’, providing an observing capability with a FoV of radius ∼1 deg, considerably larger than XMM. Each telescope module has a mirror assembly made up of 54 nested mirrors in a Wolter type-1 configuration, the same mirror geometry as XMM. The focal length of each mirror assembly is 1.6 m, with a PSF FWHM of ∼18'. The size of the PSF is larger than the PSF of XMM’s telescopes, but
Figure 1.6: Average field-of-view (FoV) grasp for the *eROSITA* telescope modules, original *Chandra* ACIS-I calibration, 2020 *Chandra* ACIS-I calibration, *Chandra* HRC-I, combined *XMM* instruments (EPN, EMOS1, EMOS2), and the *ROSAT* PSPC (*Predehl et al.*, 2021). Grasp is defined as the effective area multiplied by the FoV area, and is a measure of the combination of how much of the sky the FoV covers and also how sensitive the instrument is.
considerably better than the angular resolution of the PSPC instrument (∼1.8′) on ROSAT, the last X-ray telescope to perform an all sky survey.

Each telescope module is equipped with a camera based on the same principle as XMM’s EPIC-PN, though with just one chip rather than twelve. This eliminates the need to deal with chip gaps in your data, where no parts of the image have no data available. Improvements have been made to the design, in particular the addition of a ‘framestore’ device. The framestore reduces the number of out of time events that EPIC-PN observations suffer from, which occur when photons are recorded whilst the CCD charges are being read out. Predehl et al. (2021) report that the framestore almost entirely resolves this issue. The combined on-axis effective area of the seven telescope modules at ∼1.5 keV provides a sensitivity effectively equal to the combination of the three EPIC cameras on XMM (see Figure 1.4), though the combined XMM cameras are more sensitive at higher and lower energies. Comparisons of the grasp (effective area multiplied by FoV area) however, show that eROSITA has a significant advantage over other X-ray telescopes in the soft X-ray regime in terms of the speed at which large areas of sky can be well observed.

1.3.1.3 Chandra

The Chandra X-ray Observatory is the National Aeronautics and Space Administration’s (NASA) flagship X-ray telescope, and was launched in July of 1999. As such it has been in service for slightly longer than XMM-Newton, with the two telescopes complementing each other with their different strengths. Just as with XMM, this long service time means that there is a very large (and still growing) archive of observations that can be drawn upon for X-ray science.

Chandra’s consists of a single Wolter type-1 X-ray telescope, with 4 nested mirror shells and a focal length of 10 m. The telescope has significantly fewer nested mirrors than its ESA counterpart XMM, but the thick optical substrate used to construct the mirrors, the Iridium coating, and the extreme care taken when polishing and calibrating the mirrors means that it exceeds the performance of XMM’s telescopes. The telescope’s PSF is small enough to allow an angular resolution of 0.5″, which is approximately ten times better than that of XMM. Such fine spatial resolution allows observations that can resolve the sources of X-rays inside of galaxies, as well as probe substructure of the ICM of galaxy clusters. A downside of Chandra’s telescope design, and the fact that there is only one telescope, is that it has a smaller effective area than XMM’s telescopes, even before instrument response and number of telescopes is considered; Chandra’s telescope has an effective area of ∼800 cm² at 1 keV, whereas a single XMM telescope has an effective area of ∼1500 cm².
Several instruments are available for use on Chandra; the High Resolution Camera (HRC) detectors, the Advanced CCD Imaging Spectrometer (ACIS) detectors, and the High/Low Energy Grating Spectrometers (HETG and LETG; these are used in combination with an HRC or ACIS detector). HRC consists of two detectors, HRC-I (a 30x30′ detector optimised for imaging, with no chip gaps), and HRC-S (a 99x6′ detector optimised for grating spectroscopy). The HRC detectors are based on a technology called ‘micro-channel plates’, which are made up of lead-oxide glass tubes ∼10 μm in diameter. The tubes have a coating that emits electrons when struck by a high energy (X-ray in this case) photon, and that charge is amplified by a photo-multiplier tube until its detectable by a grid of wires (the design is slightly different for HRC-S, where one axis is wires and one is gold lines on a ceramic base).

ACIS is also made up of two detectors, ACIS-I (a 16x16′ detector optimised for wide field imaging), and ACIS-S (50x8′ detector optimised for imaging and grating spectroscopy). The ACIS detectors are made up of arrays of CCDs, they are not single piece detectors like HRC-I is. Both ACIS-I and ACIS-S can be used for imaging/imaging spectroscopy, but ACIS-S is also capable of high-resolution grating spectroscopy when used in conjunction with the HETG instrument. The technology of these instruments is more comparable with those mounted on XMM-Newton than HRC is, but there are also key design differences. Whilst XMM detectors use both front-illuminated and back-illuminated CCDs, they are not used in conjunction in a single detector; MOS cameras use front-illuminated and PN uses back-illuminated. On the other hand, the ACIS-S detector (an array of six CCDs) uses four front-illuminated and 2 back-illuminated CCDs. They provide different capabilities, with the two back-illuminated CCDs more capable at lower energies and the front-illuminated more sensitive at higher energies.

Figure 1.4 provides a comparison of the effective area of the Chandra imaging optimised (HRC-I and ACIS-I) instruments to other X-ray telescopes. In the case of the ACIS-I detector it shows two effective area curves, one from the launch year of 1999 and one from 2020. This is because ACIS has undergone significant degradation during its operational lifetime, with a significant loss of low-energy sensitivity. We can see that both HRC-I and ACIS-I are significantly less sensitive than the combined XMM instruments, as well as the combined eROSITA telescope modules. Finally, although the HRC and ACIS instruments offer impressive capabilities, they have to share the focal plane of the singular telescope. As such they cannot be used simultaneously like PN, MOS1, and MOS2 on the XMM-Newton telescope.
1.3.2 Optical and Near-Infrared

Whereas X-ray observations probe the ionised gas between the member galaxies of a cluster, optical and near-infrared (NIR) observations allow us to investigate the galaxies themselves, something that is largely impossible with X-ray observations. Additionally, we can use the member galaxies photometric redshifts to measure an ensemble redshift for the cluster. This is a particularly valuable service in the context of X-ray analysis of galaxy clusters, as it is very difficult to measure galaxy cluster redshifts from X-ray observations; knowledge of the redshift is essential to most analyses, so optical/NIR surveys are very useful to X-ray astronomers.

1.3.2.1 Sloan Digital Sky Survey (SDSS)

The Sloan Digital Sky Survey (SDSS) is a pioneering optical/near-infrared photometric and spectroscopic survey primarily of the northern sky. SDSS began in 2000, and is performed using a 2.5 m optical telescope at the Apache Point observatory in the United States (with additional spectroscopic data coming from the Irénée du Pont Telescope at Las Campanas Observatory in Chile). Results and data are routinely released to the public so that everyone can benefit from the observations; the latest being DR17 (Abdurro’uf et al., 2022). All recent data releases have added spectroscopic data, with the final photometric datasets delivered in SDSS DR9 (Ahn et al., 2012); the SDSS imaging survey covers 14555 deg² (accounting for overlap) in the griz bands, with limiting magnitudes of 22, 22.2, 22.2, 21.3, and 20.5 respectively. The various spectroscopic observations performed by SDSS have measured a total of 5789200 spectra (as of DR17), which include galaxies, quasars, stars, and other objects. There was also an integral field unit spectroscopy program (MaNGA), which provided spatially resolved spectroscopy of galaxies. Survey operations for the fifth phase of SDSS (SDSS-V) have recently begun, and SDSS will spend the next five years collecting optical and infrared spectra of a variety of objects.

1.3.2.2 Dark Energy Survey (DES)

The Dark Energy Survey (DES) is a photometric survey in the grizY bands using the DECam instrument mounted on the 4 m Victor M. Blanco telescope at the Cerro Tololo Inter-American Observatory in Chile. Images are taken using the Dark Energy Camera (DECam), an array of 74 CCDs with a total resolution of 570 megapixels. The Dark Energy Survey was founded and planned with the express goal of investigating the magnitude and nature of dark energy through
the measurement of cosmological parameters. These measurements are undertaken in several different ways, including galaxy clustering, galaxy-galaxy lensing, cosmic shear, and galaxy cluster number counts. To this end, observations were undertaken for six years, from late-2012 to January of 2019. The second, and final, public data release (Abbott et al., 2021) included source catalogues derived from observations of $\sim 5000$ deg$^2$ of the southern sky; with the catalogues containing $\sim 691$ million unique objects. Figure 1.7 illustrates the DES sky coverage in comparison to other surveys, with the DES footprint shown as a blue ‘tank’. The median limiting magnitudes across the DES sky for the different filters are $g=24.7$, $r=24.4$, $i=23.8$, $z=23.1$, and $Y=21.7$. For the filters in common with SDSS, DES demonstrates considerably deeper limiting magnitudes, a result of the Blanco telescope’s larger aperture, and different survey design. Several parts of the DES footprint were imaged repeatedly (‘deep fields’ consisting of $\sim 27$ deg$^2$ of the DES footprint) in the $griz$ bands for a supernova survey with a cadence of $\sim 7$ days. This was in concert with a spectroscopic survey called OzDES, which used the Anglo-Australian Telescope to gather repeated spectroscopic observations of objects in the deep fields.

### 1.3.2.3 Hyper Suprime Cam (HSC) Subaru Strategic Program (SSP)

The Subaru Strategic Program (SSP) is a set of three surveys of the northern sky conducted using the Hyper Suprime Cam (HSC) mounted on the 8.2 m aperture Subaru telescope. The Subaru telescope is the flagship observatory of the National Observatory of Japan, is located on the summit of Mauna Kea in Hawaii, and has been in operation since January of 1999. The Hyper Suprime Cam is the camera that is currently mounted on the Subaru telescope (first light was achieved in 2013), and is an array of 104 CCDs with a combined resolution of 900 megapixels. This resolution, combined with the telescope aperture, was used to perform three very sensitive surveys in the optical and near-infrared ($grizy$) with extra observations taken using narrow band filters. The wide survey imaged $\sim 1800$ deg$^2$ and has limiting magnitudes of $g=26.5$, $r=26.1$, $i=25.9$, $z=25.1$, and $Y=24.4$; the deep survey covers $\sim 27$ deg$^2$ and has limiting magnitudes of $g=27.5$, $r=27.1$, $i=26.8$, $z=26.3$, and $Y=25.3$; the ultra-deep survey covers $\sim 3.5$ deg$^2$ and has limiting magnitudes of $g=28.1$, $r=27.7$, $i=27.4$, $z=26.8$, and $Y=26.3$. As such, all surveys are deeper than DES, and considerably deeper than SDSS, though the area of sky covered is much smaller than either.
1.3.2.4 Vera Rubin Observatory Legacy Survey of Space and Time (LSST)

The Vera Rubin Observatory (henceforth referred to as Rubin) is an observing facility with an 8.4 m aperture telescope that is currently under construction on the summit of Cerro Pachón. The LSST camera that will be mounted to the telescope is the largest digital camera ever produced, is made up of an array of 189 CCDs, and will provide wide, high resolution images with its total resolution of 3.2 gigapixels and 3.5 deg field of view. Rubin will be used to perform the Legacy Survey of Space and Time (LSST), a survey that will take observations in the $ugrizY$ bands of $\sim 18000 \text{ deg}^2$ of the southern sky. The survey is currently planned to go on for ten years, and will provide uniform coverage of the chosen filters, imaging the entire southern sky every few nights. Predicted single exposure (ten year depth) limiting magnitudes are $u=23.9(26.1)$, $g=25.0(27.4)$, $r=24.7(27.5)$, $i=24.0(26.8)$, $z=23.3(26.1)$, and $Y=22.1(24.9)$. Even the single exposure limiting magnitudes are comparable with DES full depth (while covering a much wider area), and significantly better than SDSS limiting magnitudes (while also being better resolution). The Rubin observatory’s Legacy Survey of Space and Time will provide very powerful datasets for many types of study, including the analyses of large scale structure and cosmology performed by DES, the investigation of transient astronomical phenomena such as supernovae, and investigations of objects within our own solar system. It is currently expected to detect $\sim 20 \times 10^9$ galaxies, $\sim 17 \times 10^9$ stars, and $\sim 6 \times 10^6$ orbits of objects in our solar system.

1.3.3 Radio and Millimetre Wave

Several aspects of cluster science (as well as member galaxy properties) can be probed using radio and millimetre wave observations. Radio emission can highlight the presence of magnetic fields and cosmic in rays in the ICM, as well as indicating shocks in the ICM where particles have been accelerated to relativistic speeds as the result of mergers (van Weeren et al., 2019). Galaxy cluster radio emission can be on the scale of the whole cluster (giant radio halos) or localised, indicating a shock in the ICM, a radio-bright galaxy, or cases where AGN plasma has been re-energised by the ICM.

The millimetre wave regime (high frequency radio, between $\sim 30 – 300 \text{ GHz}$) is also useful for the investigation of galaxy clusters. The Sunyaev-Zel’dovich (SZ) effect can be observed in this range and is an effective means of locating galaxy clusters. This is where the photons of the cosmic microwave background (CMB) are boosted in energy through inverse Compton scattering, where photons gain energy through an interaction with a charged particle. This affects the local
temperature of the CMB, leading to ‘cold spots’ when observed at frequencies below 220 GHz, and ‘hot spots’ when observed above 220 GHz (Hilton et al., 2021).

1.3.3.1 Atacama Cosmology Telescope (ACT)

The Atacama Cosmology Telescope is a 6 m aperture telescope located on Cerro Toco in the Atacama desert, making high-resolution/sensitivity observations at a variety of frequencies. It was designed to make surveys of the cosmic microwave background, with an aim of performing cosmological measurements and detecting large samples of galaxy clusters. First light was achieved in 2007, and there have been several upgrades to the instrumentation since then, with the current instrument being the Advanced ACTPol camera. This is the third generation of instrument, and represents a significant upgrade to the last instrument, ACTPol. It supports high-sensitivity, high-resolution (arcminute scale) observations in five frequency bands, the centre frequencies of which are 28 GHz, 41 GHz, 90 GHz, 150 GHz, and 230 GHz. As such it is most useful for observation of the SZ effect, identifying clusters and providing an estimate of their mass (which is strongly correlated with the signal-to-noise of SZ observations. ACT can also make measurements of the polarisation of the light it observes, providing another probe of the CMB. Observations with ACT cover \( \sim 18000 \) deg\(^2\) of the sky, intersecting with surveys in other wavelengths such as SDSS and DES, though several thousand square degrees of that are of the galactic plane where dust makes it very difficult to detect galaxy clusters (Hilton et al., 2021).

1.3.3.2 South Pole Telescope (SPT)

The South Pole Telescope is a 10 m aperture telescope built at the Amundsen-Scott South Pole station in Antarctica. It achieved first light in 2007, and was built with a very similar mission to ACT; finding and investigating galaxy clusters and constraining cosmological parameters. It has also had three generations of instruments mounted on it since it was built, with the ability to measure polarisation introduced in the second generation. The current instrument mounted on the telescope is called SPT-3G, and can perform observations in bands with centre frequencies of 90 GHz, 150 GHz, and 220 GHz. This is not as many frequency bands as ACT, and the frequency range does not extend as low as ACT so in some ways ACT is more capable, however SPT has a significantly larger aperture than ACT with the corresponding increase in light collection. SPT has been used for different surveys, including a 2500 deg\(^2\) SPT-SZ survey, a 500 deg\(^2\) SPT-POL (polarisation) survey, and a subsequent 1500 deg\(^2\) high-sensitivity polarisation survey with the
SPT-3G instrument.

1.3.3.3 LOw-Frequency ARray (LOFAR)

The LOw-Frequency ARray is a radio telescope made up of an array of small simple antenna, combined using interferometry to produce a telescope with a massive effective collecting area of $\sim 300000 \text{ m}^2$. The bulk of the telescope’s antenna are based in the Netherlands, but there are also LOFAR stations in Germany, Poland, France, Ireland, Latvia, Sweden, and the United Kingdom. LOFAR is made up of $\sim 20000$ antenna, and is considered a ‘software telescope’, as the signals from each system are sent to a central location and processed into a map of the sky. Two types of antenna are deployed, Low Band Antennas (LBA) that are sensitive in the 10-80 MHz band and High Band Antennas (HBA) that are sensitive in the 120-240 MHz band. This is key to its science goals, which are to map the Universe in low radio frequencies at a greater resolution and sensitivity than previous observatories. The 10-240 MHz observing capability of LOFAR also allows for the investigation of the 21 cm hydrogen line in emissions from the epoch of reionisation, a field of study that is of great current interest. With its large baseline, LOFAR is able to achieve sub-arcsecond spatial resolution over most of its frequency range (van Haarlem et al., 2013; Morabito et al., 2022). ACT and SPT both observe at considerably higher frequencies, which is what allows them to make observations of the SZ effect (something LOFAR cannot do). However, LOFAR’s lower frequency band allows it to probe cluster radio halos (such as from mergers) and emission from radio galaxies, physics that can be complementary to locating clusters through the SZ effect. They have also made use of the recent data release two (Shimwell et al., 2022b, covering 5635 deg$^2$ of the northern sky, containing 4396228 sources) to probe radio emission from clusters (Botteon et al., 2022) detected by the Planck CMB telescope using the SZ effect (Planck Collaboration et al., 2016).

1.4 The XMM Cluster Survey (XCS)

The XMM Cluster Survey (XCS, Romer et al., 1999) is a serendipitous survey of the entire XMM-Newton archive, with a particular focus on the location and analysis of galaxy clusters. All available XMM observations are downloaded, processed and cleaned (see Section 1.4.2), and then have a source detection algorithm run on them. Galaxy clusters are not the only type of object explored by XCS, as XCS source catalogues contain hundreds of thousands of point sources.
1.4.1 XCS Science Goals and Achievements

The primary aim of XCS is to construct premier samples of galaxy clusters, measure their properties, and use that information to perform and enable the derivation of cosmological parameters. Initially XCS was going to use its cluster samples and X-ray properties to perform ‘in-house’ (i.e. constructing an XCS halo mass function) measurements of cosmological parameters, and preparatory work was performed (Sahlén et al., 2009) to predict the cosmological constraints that could be expected.

Ultimately it was decided that the very limited XCS sky coverage (Sahlén et al., 2009, predicted that XCS would cover ∼500 deg$^2$) meant that it would be better for XCS to support other cluster surveys with much larger area coverage. This support would be provided in the form of scaling relations between cluster properties (including between cluster hydrostatic mass and easier-to-measure observables), as well as multi-wavelength exploration of cluster catalogue completeness (Upsdell et al., prep), and calibration of cluster mis-centering models (Zhang et al., 2019). Figure 1.7 shows the sky coverage of XCS using black points, and illustrates the stark contrast between the coverage of XMM pointings and other optical/NIR surveys such as DES and LSST. Likewise XCS’ sky coverage is far outstripped by that of the German half of the eRASS (see 1.6.5).

The measurement of cosmological parameters is not the only focus of XCS however; the cluster catalogues produced by XCS are world-leading and can be used for a variety of science. XCS has contributed to the study of dark matter (Bhargava et al., 2020), finding no evidence that a previously reported (Bulbul et al., 2014) 3.5 keV emission line is consistent with sterile neutrino dark matter. Additionally, XCS has investigated theories of modified gravity Wilcox et al. (2015) using XCS clusters and Canada-France-Hawaii-Telescope Lensing Survey (CFHTLens, Heymans et al., 2012) weak lensing measurements, putting constraints on chameleon gravity parameters.

Finally, the astrophysics of galaxies and galaxy clusters are also of interest to XCS. For instance, there have been investigations into the dynamics of individual clusters (Hilton et al., 2007; Pillay et al., 2021) using combinations of XMM and spectroscopic optical and XMM and MeerKAT (Jonas and MeerKAT Team, 2016) data respectively. XCS has also analysed the XMM data of clusters taken from the XCS-DR1 catalogue (Mehrtens et al., 2012) to see whether there is evolution of the $L_X$-$T_X$ scaling relation with redshift; no evolution was found, but the data were found to favour feedback models in which the majority of the energy injection occurs at high redshift. The individual galaxies that make up the galaxy clusters in XCS catalogues have also been the subject of
Figure 1.7: The sky coverage of XCS, DES, LSST, and German eRASS surveys. Each survey, as well as intersections between different surveys, are highlighted by a different colour in the legend at the bottom of the figure. This figure was created by Reese Wilkinson.

XCS work; the build-up of stellar mass in specifically the brightest cluster galaxies (BCG; Stott et al., 2010) was traced using J and K band photometry and the intracluster light component of 18 XCS galaxy clusters was measured using I-band data Furnell et al. (2021).

1.4.2 Data preparation

The XMM Cluster Survey is a serendipitous survey of the entire XMM archive, and as such has searched the entire XMM scientific archive (consisting of over two decades worth of data) for X-ray sources. This involves downloading and processing every applicable XMM observation into science-ready data. The final output of this process are X-ray ‘event lists’, tables that contain information about the photon events that were recorded during a particular observation. This includes their position on the detector (also mapped to a corresponding position on the using XMM attitude files), the time that they arrived, and the energy of the photon. Such files can then be binned in two spatial dimensions to form images, or binned in energy to make spectra.

The main difficulty in XCS’ X-ray data preparation lies in the treatment of background noise caused by flares of high energy particles and protons from the Sun. Such flares are a large part of XMM’s background (Lumb et al., 2002; Read and Ponman, 2003) and can cause additional events
Figure 1.8: A demonstration of the effect of periods of high flare activity on hard-band (12-15 keV) EPIC-PN light curves. The right hand side figure shows the full light curve, with several severe peaks in the number of high energy events being registered. The left hand side figure shows the light curve once the XCS cleaning procedure has been applied; the periods of flaring have been removed, leaving a much more stable lightcurve.

to be detected in the soft ($\lesssim 1$ keV) X-ray band, as well as in the harder X-ray band ($\sim 10 - 15$ keV). Such flaring can be highly temporally variable, and at certain points it is preferable to simply remove a period of the observation, in an attempt to maximise the overall signal-to-noise. The XCS data preparation procedures are fully described by Lloyd-Davies et al. (2011), who designed an automated method to identify and remove periods of high particle background. This process involves generating a hard-band light-curve (where the counts received in a certain energy band are plotted against the time they were recorded), and then using an algorithm to locate periods of much greater than average activity. Figure 1.8 shows a flared light curve, with significant spikes in the number of counts at several points, and then shows the light curve after the XCS algorithm was applied. Firstly a hard-band light curve, binned in intervals of 50 s, is generated, and a count-rate threshold is applied to remove extremely high rate bins. Then the mean and standard deviation of the light curve are calculated and 50 s bins further than $3\sigma$ from the mean are excluded. The mean and standard deviation are then re-calculated and the process repeated up to fifty times. The whole process is then performed again with a soft-band light curve, to account for the soft-proton component of flaring.

Once cleaned event lists are available, a set of standard images and exposure maps (which describe the effective exposure across the field of view of the telescope, it changes due to vignetting) are generated. They are generated in specific energy bands, 0.5-2.0 keV (which is most useful for finding objects like clusters which emit lower energy X-rays) and 2.0-10.0 keV, both for the individual EPIC cameras (PN, MOS1, MOS2; Section 1.3.1.1) and also for the combination of all
1.4.3 Existing software tools

There are two key pieces of existing software that enable XCS’ science efforts; the XAPA (Lloyd-Davies et al., 2011) source finding algorithm, and the XCS Luminosity-Temperature pipeline (Lloyd-Davies et al., 2011; Mehrtens et al., 2012; Giles et al., 2022b).

The XCS source finder, XAPA, is the linchpin of XCS science; it has been run on every XMM observation that is available for download (over 14000 pointings), and is capable of finding both extended and point-like X-ray sources. While XCS’ primary interest is in the location and analysis of galaxy clusters, not all extended sources detected by XAPA are the result of cluster emission; some will be the result of the energetic remnants of a supernova, or an X-ray bright local galaxy, for instance. One benefit of a multi-wavelength approach (as discussed in Section 1.4.1) is that the observations in other wavelengths help to define the nature of the object that is emitting X-rays. XAPA is based upon a source detection package called WAVDETECT (Freeman et al., 2002), though other source detection methods were considered; e.g. XMM SAS EWAVELET and SEXTRACTOR (Bertin and Arnouts, 1996). It functions by convolving a window function (on various scales) with the image, comparing the convolved image with a ‘threshold image’ and finding pixels where the convolved image has a significantly higher value than the threshold image. Such pixels are called ‘significant pixels’, and are assumed to be associated with astronomical sources rather than a diffuse X-ray background. In total nine different scales of window function are convolved, to locate sources of varying size, with bright sources being identified (and then masked) first, so as to not affect the detection of more extended sources. Once significant pixels have been located, they have to be grouped together into discrete sources, this is the second stage of XAPA, and it is performed separately at each wavelet scale. Those objects which are detected at multiple scales will be extended sources. The original implementation of WAVDETECT uses the size of the instrumental point spread function (PSF) to create the groups that become source regions, but this was found to have a negative impact on the detection of extended sources, as large objects could be broken up into multiple PSF-sized regions. As such a modified version was written that does not a priori assume the size of the PSF, and has been shown to be more capable at finding and fitting elliptical source regions (Lloyd-Davies et al., 2011). XAPA is run on the combined (PN+MOS1+MOS2) images, and so a single source file is produced for each XMM observation.

The XCS luminosity-temperature pipeline (XCS Post Processing Pipeline, or XCS3P) is an it-
erative pipeline that goes from an input catalogue of galaxy clusters to a set of reliable X-ray temperatures and luminosities within an estimate of $R_{500}$. The input catalogue is required to have positions and redshifts, but no $R_{500}$ (or any other overdensity radius) is required. XCS3P is originally described by Lloyd-Davies et al. (2011), though was recently updated by Giles et al. (2022b). It has been used for multiple XCS publications, such as the XCS-DR1 catalogue (Mehrtens et al., 2012), and the recent SDSSRM-XCS catalogue (Giles et al., 2022b), as well as for XCS supported work by the Dark Energy Survey (Farahi et al., 2019a). XCS3P works by generating spectra (and accompanying files necessary for analysis, including background spectra) within circular apertures centered on XAPA detections of X-ray emission from clusters. The initial aperture is based on the semi-major axis of the original XAPA detection of a cluster, and once the spectra have been generated a fit is performed to the data using an absorbed (with $\text{tbabs}$; Wilms et al., 2000) plasma emission model ($\text{APEC}$; Smith et al., 2001). The model fit can be used to retrieve the global temperature and luminosity of the cluster ICM, which in turn is used to estimate an $R_{500}$ from a radius-temperature scaling relation (Arnaud et al., 2005). This process iterates (with the new estimate of $R_{500}$ used as the analysis aperture) until the measured temperatures converge to within 10%.

1.4.4 Master Source List (MSL)

As discussed in Section 1.4.3, the XAPA source finder is run on every $XMM$ available to XCS, and a list of source regions is produced for each observation. The XCS master source list (MSL) is the compilation of regions for individual observations into a final catalogue with no duplicated sources. It is not uncommon for there to be overlapping $XMM$ observations (such as in the XXL fields, see Section 1.6.3), or multiple observations of the exact same object, so cross-matching between region lists is required for the construction of the master source list.

As of March 2022 (the last time the XCS list of processed observations was updated) the master source list contains 400225 entries. Of these, 357478 are point sources, and the remaining 42747 are considered to be extended sources.

1.5 Relevant Software

Some existing software developed by XCS has already been introduced in Section 1.4.3, but some software external to XCS is used throughout this thesis thus will be introduced here.
1.5.1 The XMM Science Analysis System (SAS)

The XMM Science Analysis System (SAS; Gabriel et al., 2004) is the name given to the tools and scripts provided by the XMM team for the purposes of reducing XMM data and readying it for analysis. They can be installed on any Unix based system and are generally used from the command line, rather than by using another programming language (such as Python) as an interface; however there are a very limited number of functions usable through Python, and new Python interfaces for many common SAS tools should be available with the next release.

There are many tools that make up the SAS software, covering a wide range of functions, so not all of them can be introduced here. The most relevant of the SAS tasks to XCS are:

- **evselect** - The most generally useful piece of software in the SAS toolkit, this is used for selecting events relevant to whatever source you’re analysing, and then processing them into a data product. This is the tool that is used to generate both XMM images and spectra, as they are largely the result of the same processes; the only difference is that the selected events are binned in two spatial dimensions to form an image, and in one energy (or rather detector channel) dimension to form a spectrum. An expression that determines which events are selected can be passed, imposing spatial (if generating a spectrum for instance) or event energy limits (if generating an image of the whole XMM observation within a specific energy band), amongst other things.

- **selectlib** - This part of SAS provides the capability to select specific events referred to in the evselect introduction. It is a suite of tools that allow filtering based on spatial position, temporal position, or energy. This includes implementing expressions that simplify the imposing of spatial masks of particular shapes; circular masks for instance can be defined with the central coordinates and radius, and it provides similar interfaces for more complex shapes such as ellipses, annuli, and polygons.

- **eexpmap** - This performs the calculations necessary to generate XMM exposure maps, taking into account the quantum efficiency of the detectors as it varies with energy, as well as the vignetting effects caused by the X-ray optics as you move closer to the edge of the field of view. This tool is essential for the creation of count-rate maps, as assuming a single exposure time would result in measured count rates diverging from true values as you moved closer to the edge of the field of view.

- **emosaic** - This tool takes an arbitrary number of images (or exposure maps) and combines
them into a single mosaic-ed image or exposure map. This is useful for cases where an object of interest has been observed multiple times and you wish to make use of all the data.

• arfgen - Essential for the analysis of spectra generated for XMM observations, this tool creates the Auxiliary Response Files (ARF) that describe the effective area (i.e. sensitivity) of the detector at different energies. This information allows emission models being fitted to be ‘folded’ with the response so they represent what the telescope is actually observing.

• rmfgen - Also essential for the analysis of XMM spectra, this tool generates the Response Matrix Files which essentially provide the mapping between detector channel and energy. As all XMM events initially report a detector channel, this is necessary to convert into physical units.

• backscale - A very simple tool that calculates the area covered by the spatial region inside which a spectrum has been generated. This is done for source and background spectra and allows background spectra to be scaled to account for difference in area.

• specgroup - Finally, this tool is used to take spectra generated using evselect and group channels to improve the signal to noise of the spectra.

The SAS software also contains a source finder, but as XCS has implemented their own source finder XAPA, it is not used in any capacity within this thesis. SAS tools are also used extensively by XCS during the data reduction and cleaning steps introduced in Section 1.4.2, but as the work in this thesis just makes use of the already prepared XCS data, those tools are not introduced here.

1.5.2 The XSPEC Spectral Fitting Code

This code was developed by Arnaud (1996), and allows the user to fit input X-ray spectra with a large number of X-ray emission and absorption models. Other tools for X-ray spectroscopy exist (ISIS for example; Houck and Denicola, 2000), but XSPEC is still the most widely used tool for this purpose. XCS have performed all their spectral analysis of galaxy clusters (and other sources) using XSPEC. The software can be used in interactive and script modes, making it useful for both exploratory analysis (loading in a spectrum and its supporting files, testing how well different models fit and visualising the results for instance), as well as for setting up fitting scripts for large samples of objects to be analysed in the same way. It supports scripting in the TCL programming language, so traditional programming tools such as loops, boolean statements, and different data types such as strings, integers, and arrays are all supported.
1.5.3 Astropy

Astropy (Astropy Collaboration et al., 2013, 2018) is an open-source Python module that provides a large range of functionality related to the study of astronomy, astrophysics, and cosmology, and the analysis of astronomical data. This ranges from providing interfaces to common cosmological calculations, with easy to access cosmology instances for cosmologies measured by specific works, like the Planck Collaboration et al. (2020) results, to providing tools that help to visualise astronomical data and interact with fits files.

This module is a large part of all analyses presented in this thesis, powering a many of the background processes that go into preparing and interacting with data. It is used extensively in the open-source X-ray analysis module that was developed during the course of this PhD, and is described in Chapter 2, and though there are too many functions to allow Astropy to be described in detail, some of the most important functionality is described here:

• **Units and Quantities** - One of the most useful aspects of Astropy is the ability to define ‘quantities’, which have both a value and a unit associated with them. Once defined they can be easily converted to other compatible units (metres to kiloparsecs for instance), and are protected from being converted to incompatible units. Astropy quantities can also be used in place of numpy arrays, meaning that multiple values with the same unit (supporting vectorisation) can be stored in one object. As such astropy quantities can also be used in calculations, with the result of maths operators having appropriate units (so multiplying a quantity with a unit of $\text{m}^{-1}$ would result in a dimensionless quantity for instance).

• **Reading FITS** - FITS files are ubiquitous in astronomy, and are used for anything from storing catalogues and spectra with headers in binary tables, to storing images in arrays. Interacting with them and extracting data is a common occurrence in most analyses, and astropy provides a very convenient set of tools to extract data and headers from them. The tools can extract data from any type of FITS file, whether it stores images or binary tables. It is also possible to setup objects to interact with World Coordinate System (WCS) information stored in image headers, which allow for relatively simple conversions between different coordinate systems that are embedded in an image, another common occurrence in analyses.

• **Image Visualisation** - It is often necessary to display visual representations of astronomy
images, whether for generating figures for publication or for visual inspection of a particular source. This commonly involves scaling images, applying specific colour maps, and smoothing the data. Astropy provides easy to use implementations of all of this functionality, with different stretches (such as log stretch and arcsin stretch) and different smoothing kernels available.

- **Cosmology** - Finally, as has already been mentioned, Astropy provides interfaces with different cosmologies (including the ability to set up your own cosmology with user defined properties). These interfaces can be used to calculate different distance measures, the value of the Hubble parameter at user-defined redshifts, and lookback times, amongst other things. They are much more convenient than implementing these calculations for yourself, for every analysis you wish to perform.

## 1.6 Samples of Galaxy Clusters

This subsection provides an overview of the different samples of galaxy clusters used throughout this work, including how each sample was constructed and what survey they were selected from.

### 1.6.1 SDSS redMaPPer (Optical/NIR selected)

This is a large catalogue of galaxy cluster candidates located using optical/near-infrared photometry from SDSS DR8 (Aihara et al., 2011), using the red-sequence matched-filter Probabilistic Percolation (redMaPPer, Rykoff et al., 2014, 2016) cluster finder. The redMaPPer algorithm, also referred to as RM in this work, was designed for use on large scale photometric surveys. This makes it very useful for the current generation of photometric surveys, such as SDSS (Section 1.3.2.1) and DES (Section 1.3.2.2), and a key part of the cluster science goals of the upcoming LSST (Section 1.3.2.4). The algorithm works by taking an input galaxy catalogue derived from a photometric survey and attempting to find red sequence galaxies; this technique was initially proposed for use on the SDSS photometric survey (Annis et al., 1999; Gladders and Yee, 2000), and has enjoyed considerable success. The red sequence is an empirical property of galaxy clusters, where the cluster tends to host a population of early-type elliptical galaxies with low star-formation rates, formed at high redshift, and lying along a linear colour-magnitude relation (Bower et al., 1992).

The algorithm is made up of two stages. Firstly, redMaPPer calibrates the red sequence as a
Figure 1.9: Locations of galaxy cluster candidates located by the redMaPPer algorithm. Red points indicate a candidate in the SDSS DR8 redMaPPer public catalogue. Blue points indicate candidates in the DESY3 redMaPPer catalogue. Part of the SDSS catalogue is obscured by part of the DES catalogue.
function of redshift by using a training catalogue of brightest cluster galaxy (BCG) spectroscopic redshifts. The creators of redMaPPer (Rykoff et al., 2014) believed this to be preferable to using photometric redshift estimates from galaxy catalogues because cluster galaxies are a ‘particular population’ that should be treated separately from other galaxies. Self-training with spectroscopic redshifts for cluster red sequence galaxies avoids this issue, and allows redMaPPer to measure photometric redshifts for the cluster member galaxies that it locates. The second stage involves using the trained model to identify red sequences on the sky, assumes that they represent an overdensity, and then obtains a richness ($\lambda$; where richness is a probabilistic measure of the number of galaxies in a cluster) for each system. The whole process is iterative; once a first rough set of clusters has been identified they can be used to recalibrate the initial red sequence model and refine the results.

The output of redMaPPer is a cluster candidate catalogue with positions, richness estimates, and estimates of the redshifts of the clusters. A redshift for a cluster is obtained by a simultaneous fit of all high probability cluster members to the red sequence model. The cluster positions output by redMaPPer correspond to the most likely candidates for the central galaxy, with the five most likely positions included for each cluster. The central massive galaxy of a cluster has been shown to be a useful estimate of the ‘true’ centre for optical/NIR observations, though other wavelengths (such as X-ray) provide better proxies for the cluster centre (Yan et al., 2020; Zhang et al., 2019).

There are 26111 entries in the catalogue (the updated version from Rykoff et al., 2016), with the lowest redshift candidate at $z = 0.08$, and the highest redshift candidate at $z = 0.60$. The lowest richness ($\lambda$) system in the published catalogue is $\lambda = 19.85$, and the highest is $\lambda = 299.46$. Figure 1.9 shows (in red) the distribution on the sky of the full SDSSRM catalogue.

Subsets of the SDSSRM cluster catalogue containing XCS confirmed galaxy clusters were created by Giles et al. (2022b), in order to measure a set of scaling relations between X-ray temperature, X-ray luminosity, and richness. This process involved determining which galaxy cluster candidates fell on at least one XMM observation, as well as which candidates could be matched to a corresponding XCS detection of an X-ray extended source. Following this, images of both the X-ray and SDSS observations were inspected, with extra information such as redMaPPer member galaxies and XCS sources overlaid on the images. The resulting sample of galaxy clusters with successful temperature measurements includes 382 objects. A volume limit was also imposed on the clusters by selecting only clusters that have a redshift $0.1 < z < 0.35$, and then selecting only clusters with fractional temperature uncertainties of less than 25%, leaving a sample of 150 clusters.
1.6.2 DESY3 redMaPPer (Optical/NIR selected)

This cluster catalogue is closely related to the previously discussed SDSS redMaPPer catalogue (1.6.1). The DESY3 redMaPPer catalogue is the result of running the redMaPPer (Rykoff et al., 2014, 2016) algorithm on the DES galaxy catalogue derived from the first three years of DES data. It supersedes both the DES science verification (SV; Rykoff et al., 2016) redMaPPer catalogue and the DESY1 redMaPPer catalogue (based on the first year of DES data Abbott et al., 2020).

The full DESY3 redMaPPer catalogue (v2.2.1) contains 53610 galaxy cluster candidates, see the blue points in Figure 1.9 for an illustration of their positions on the sky. The lowest redshift candidate is at $z = 0.10$, the highest redshift candidate is at $z = 0.95$; this maximum redshift immediately indicates that the Dark Energy Survey is sensitive to clusters at higher redshift than SDSS is, due to DES’ (Section 1.3.2.2) limiting magnitudes being better than SDSS’ (Section 1.3.2.1). The lowest richness cluster has $\lambda = 20$ and the highest has richness has $\lambda = 221.67$. The catalogue is artificially limited to a minimum richness of 20 as Farahi et al. (2016) demonstrated (using simulations) that 99% of redMaPPer clusters above this limit can be mapped to an individual dark matter halo.

The same visual inspection and classification process that was used to build the SDSSRM-XCS catalogue (Section 1.6.1) was used to create a DESY3RM-XCS sample. This sample contains 325 galaxy clusters, of which 293 have sufficiently high quality XMM data to measure a temperature. A fractional temperature error cut of 25% was then imposed, with 170 clusters remaining in the final sample.

1.6.3 Ultimate XMM eXtragaLactic (XXL)-100-GC (X-ray selected)

XXL-GC-100 is a sub-sample (Pacaud et al., 2016) of the galaxy clusters detected by the Ultimate XMM eXtragaLactic (XXL) survey (Pierre et al., 2016), the largest XMM observing program. The survey’s main goals are to provide constraints on the dark energy equation of state from the space-time distribution of clusters of galaxies and to serve as a pathfinder for future, wide-area X-ray missions such as eROSITA. The original 542 pointings used for XXL totalled approximately 7 Ms of observing time, covering two contiguous regions of the sky. These two fields, called XXL-North and XXL-South, total $\sim 50$ square degrees and contain $\sim 500$ galaxy clusters. A comprehensive follow-up program was then undertaken to obtain spectroscopic or photometric redshifts for the whole galaxy cluster candidate catalogue. The XXL fields appear in Figure 1.7, as the two large
areas of black at bottom and top of the Dark Energy Survey ‘tank’ region.

The XXL-GC-100 sub-sample was selected by setting a lower limit of $3 \times 10^{-14}$ erg s$^{-1}$ cm$^{-2}$ on the flux measured within a 1’ aperture. This flux limit was specifically selected to return the 100 brightest clusters from the full cluster sample. The goal was to select a homogeneous set of clusters with at least 100 X-ray counts in a high signal-to-noise area, to allow ‘statistical studies of cluster physics in the group-mass range at intermediate redshifts along with a preliminary cosmological analysis’ (Pacaud et al., 2016).

This sample of 100 clusters has been used in several XXL analyses, including their work on mass temperature relations (Lieu et al., 2016), and an exploration of the baryonic properties of the sample (Eckert et al., 2016). As such measurements of the X-ray temperature and X-ray luminosity are available (Giles et al., 2016). The total masses used for the XXL mass-temperature relations are not measured using X-ray data, however the gas masses of the clusters in this sample were measured during the exploration of their baryonic properties, and are also publicly available (Eckert et al., 2016).

The XXL-100-GC clusters cover a redshift range of 0.043 to 1.05, and a temperature range (as measured by XXL within an aperture of 300 kpc by Giles et al., 2016) of 0.64-13.8keV.

### 1.6.4 Local Cluster Substructure Survey (LoCuSS) High-$L_X$ (X-ray selected)

The LoCuSS sample is a sample of 50 galaxy clusters selected from the ROSAT all sky survey (RASS) catalogues (Ebeling et al., 1998; Böhringer et al., 2004), for which hydrostatic masses have been measured using a mix of XMM-Newton and Chandra observations (Martino et al., 2014). The full LoCuSS sample of 165 clusters was derived from the RASS catalogues by applying cuts to various values; $0.15 < z < 0.3$, $n_H < 7 \times 10^{20}$ cm$^{-2}$, and $-70^\circ < \delta < 70^\circ$. The High-$L_X$ sample was then selected from this larger sample by a harsher cut on the allowed declination values ($-25^\circ < \delta < 65^\circ$), and a cut on the cluster luminosities measured within the $0.1 - 2.4$keV band ($L_X E(z)^{-2.7} \geq 4.2 \times 10^{44}$ erg s$^{-1}$).

Many pieces of work have made use of this sample, and as such the LoCuSS High-$L_X$ clusters are very well studied, particularly in X-ray. Common X-ray global properties, such as X-ray temperature within $R_{500}$ and X-ray luminosity within $R_{500}$ (both with and without core excision), have been measured and published (Martino et al., 2014; Mulroy et al., 2019). More complex properties such as cluster gas and hydrostatic masses are also available (Martino et al., 2014).
There has also been optical follow-up with excellent weak lensing (Okabe and Smith, 2016) using the Hyper Suprime Cam (see Section 1.3.2.3).

At the time of the original LoCuSS hydrostatic mass work, 43 of the clusters had been observed by *Chandra* and 39 by *XMM*, with 27 of the clusters having data from both telescopes. A search through the *XMM* science archive now shows that 46 of the LoCuSS High-Lx clusters have been observed by *XMM*, though no such check was performed on the *Chandra* archive.

As this sample of galaxy clusters was selected from the brightest clusters detected in the RASS catalogues, it preferentially selects high mass systems. This is a consequence of the strong correlation between cluster mass and X-ray luminosity and temperature (see Chapter 3). As such, while this is an extremely useful sample for various aspects of cluster science (with a great many of the cluster properties well measured) it doesn’t tell us much about the lower mass cluster and group scale.

### 1.6.5 *eROSITA* Final Equatorial-Depth Survey (eFEDS; X-ray selected)

The *eROSITA* Final Equatorial-Depth Survey (eFEDS) is the commissioning survey for the final *eROSITA* All-Sky Survey (eRASS), and was one of the first sets of observations taken by *eROSITA*. It covers $\sim$140 deg$^2$ of the northern sky, intersecting with prominent optical/NIR surveys such as SDSS and HSC-SSP. The eFEDS observations were designed so that eFEDS’ depth slightly exceeded the full-depth eRASS survey that will be made up of eight passes of the sky. A full source catalogue as measured by the *eROSITA* source finder was a part of the original (Brunner et al., 2021) eFEDS data release. The full catalogue contains 27910 sources detected in the 0.2-2.3 keV energy band with a detection likelihood of above six, with a supplementary catalogue providing information on 4774 sources detected with a lower significance.

A subset of the full catalogue was also published, specifically containing eFEDS X-ray cluster candidates (Liu et al., 2021a), as well as a companion catalogue with optical counterpart information for those cluster candidates (Klein et al., 2021). These catalogues contain 542 X-ray cluster candidates, with 477 of those being considered ‘optically confirmed’ (Klein et al., 2021). Optical confirmation was performed using a multicomponent matched filter (MCMF) tool (Klein et al., 2018, 2019), which also provides redshift measurements for all 542 X-ray cluster candidates. MCMF, similarly to redMaPPer, makes use of the red sequence of cluster galaxies. In this case it uses the X-ray candidate position as a starting point, then calculates richnesses in 230 redshift bins out to $z = 1.3$, using HSC-SSP (Aihara et al., 2018) and Legacy survey photometric data (Dey...
et al., 2019). The minimum redshift candidate has \( z = 0.02 \), and the maximum redshift candidate has \( z = 1.30 \).

Attempts were made to measure X-ray luminosities and temperatures within fixed apertures (300 kpc and 500 kpc) for the 542 X-ray cluster candidates. Luminosity measurements were made using a photometric method, and 91% (493) of the X-ray cluster candidates have a viable measurement. As X-ray temperatures of galaxy clusters must be obtained through spectroscopic fitting, correspondingly fewer of the eFEDS X-ray candidates have successful measurements; 69 have successful temperature measurements within 300 kpc, and 95 have successful measurements within 500 kpc.

### 1.6.6 ACT-DR5 (SZ selected)

A Sunyaev-Zel’dovich (SZ) selected sample of galaxy clusters that were originally detected by the Atacama Cosmology Telescope (see Section 1.3.3.1); the work by Hilton et al. (2021) contains a full account of the data processing, detection methods, and optical confirmation. The sample contains 4195 optically confirmed galaxy clusters (with an SZ signal to noise greater than 4), from a redshift range of \( 0.04 < z < 1.91 \). The lowest cluster SZ signal to noise is 4.01, and the highest cluster signal to noise is 53.68.

As the optical data used by the ACT team to confirm the SZ detections comes from a variety of surveys, so too do the redshift measurements. Around 60% of the redshifts are photometric (some from cluster finders such as redMaPPer (Rykoff et al., 2014, 2016) and CAMIRA (Oguri, 2014), some from other methods using broadband photometry, such as zCluster (Hilton et al., 2018), with the remaining 40% coming from spectroscopic sources. Comparisons were performed between spectroscopic and photometric redshifts for 1168 clusters, and more than 98% are within \( \Delta z/(1 + z_s) < 0.05 \), so the two types of redshift measurement agree very well.

This is a very large, and high quality, sample covering a wide range of cluster masses (which are proportional to the SZ signal-to-noise measure, see Section 3.10.2) and redshifts (as the SZ effect is not redshift dependant clusters can be detected at high redshifts), it is a great sample to use for follow-up X-ray analyses.
1.7 Thesis Overview

Chapter 2 introduces and describes X-ray: Generate and Analyse (XGA), a generalised open-source Python module designed to make archival X-ray astronomy analysis quick and easy. This chapter first gives a general overview of the module, then goes on to provide details of the structure and features of the software.

Chapter 3 demonstrates the analysis of galaxy clusters using XGA, including presenting a new set of hydrostatic mass measurements for the SDSSRM-XCS sample presented in Giles et al. (2022b). The veracity of XGA measurements of ICM properties is illustrated by using XGA to re-analyse several samples of clusters from literature, then comparing the new results to the original work.

Chapter 4 presents the construction and analysis of a sample of galaxy clusters selected from the eFEDS X-ray cluster candidate catalogue. The newly constructed sub-sample was made up of eFEDS cluster candidates with at least one XMM observation, and was used to probe the contamination level of the eFEDS catalogue, as well as to compare their published $L_X$ and $T_X$ values with XMM measurements.

Chapter 5 first gives high-level overviews of other projects that have been undertaken during the course of this PhD. Then plans for future work and general conclusions are presented.
Chapter 2: X-ray: Generate and Analyse (XGA)

The first part of this chapter (Section 2.1) is made up of a short software paper that aims to give a very high level overview of the X-ray: Generate and Analyse Python module. It lays out the motivation behind the creation of this module, its general capabilities, and compares it to other relevant software packages. This paper is currently on arXiv (Turner et al., 2022), and will be submitted to the Journal of Open Source Software (JOSS), a journal specifically for software papers that relate to completely open source scientific packages. JOSS peer reviews both the paper and the code, with the idea being that the review process can help you to improve your software.

The second part of this chapter (starting Section 2.2) gives a detailed description of the design and capabilities of this software package. Some information from Section 2.1 is repeated, as the paper has been added to this chapter in the same form as it was made public, but there is a far greater amount of detail in the rest of the chapter.

2.1 XGA Paper

2.1.1 Summary

The XMM Cluster Survey (XCS, Romer et al., 2001) have developed a new Python module, X-ray: Generate and Analyse (hereafter referred to as XGA) to provide interactive and automated analyses of X-ray emitting sources observed by the XMM-Newton space telescope. XGA only requires that a set of cleaned, processed, event lists has been created, and (optionally) that a source detector has generated region lists for the observations. XGA is centered around the concept of making all available data easily accessible and analysable. The user provides information (e.g. RA, Dec, redshift) on the source they wish to investigate, and XGA will locate all relevant observations and generate all required data products. This approach means that the user can quickly and easily
complete common analyses without manually searching through large amounts of archival data for relevant observations, thus being left free to focus on extracting the maximum scientific gain. In the future, we will add support for X-ray telescopes other than XMM (e.g. Chandra, eROSITA), as well as the ability to perform multi-mission joint analyses. With the advent of new X-ray observatories such as eROSITA (Predehl et al., 2021), XRISM (XRISM Science Team, 2020), ATHENA (Nandra et al., 2013), and Lynx (Gaskin et al., 2019), it is the perfect time for a new, open-source, software package that is open for anyone to use and scrutinise.

2.1.2 Statement of need

X-ray telescopes allow for the investigation of some of the most extreme objects and processes in the Universe; this includes galaxy clusters, active galactic nuclei (AGN), and X-ray emitting stars. This makes the analysis of X-ray observations useful for a variety of fields in astrophysics and cosmology. Galaxy clusters, for instance, are useful as astrophysical laboratories, and provide insight into how the Universe has evolved during its lifetime.

Current generation X-ray telescopes have large archives of publicly available observations; XMM-Newton has been observing for over two decades, for instance. This allows for analysis of large amounts of archival data, but also introduces issues with respect to accessing and analysing all the relevant data for a particular source. XGA solves this problem by automatically identifying the relevant XMM observations then generating whatever data products the user requires; from images to sets of annular spectra. Once the user has supplied cleaned event lists (and optionally region files) an analysis region can be specified and spectra (along with any auxiliary files that are required) can be created.

Software to generate X-ray data products is supplied by the telescope teams, and most commands require significant setup and configuration. The complexity only increases when analysing multiple observations of a single source, as is often the case due to the large archive of data available. XGA provides the user with an easy way to generate XMM data products for large samples of objects (which will scale across multiple cores), while taking into account complex factors (such as removing interloper sources) that vary from source to source.
Figure 2.1: Demonstration of the view methods of the RateMap and Spectrum classes, when applied to the Abell 907 galaxy cluster. Data from the XMM EPIC-PN instrument of 0404910601 is used. Left: A count-rate map with a mask that removes contaminant sources (using XCS region information) and applies an $R_{500}$ aperture. Right: A spectrum generated for the $R_{500}$ region with contaminants removed, and fit with an absorbed plasma emission model using XSPEC.

2.1.3 Features

XGA is centered around source and sample classes. Different source classes, which represent different types of X-ray emitting astrophysical objects, all have different properties and methods. Some properties and methods are common to all sources, but some store quantities or perform measurements that are only relevant to a particular type of astronomical source.

XGA also contains product classes, which provide interfaces to X-ray data products, with built-in methods for analysis, manipulation, and visualisation. The RateMap (a count rate map of a particular observation) class for instance includes view methods (left hand side of Figure 2.1), methods for coordinate conversion, and for measuring the peak of the X-ray emission. We also provide classes for interacting with, analysing, and viewing spectra (see right hand side of Figure 2.1), both global and annular; as such we can use XGA to investigate both average properties and, in the case of extended sources, how these properties vary radially. Similar procedures for image based analysis are also available, where images (and merged images from all available data for a given source) can be easily generated en masse, then combined with masks automatically generated from supplied region files to perform photometric analyses.

We also include a set of profile classes, with built-in viewing methods, and a fitting method based around the emcee ensemble MCMC sampler (Foreman-Mackey et al., 2013). Profiles also support storing and interacting with fitted models; including integration and differentiation methods,
inverse abel transforms, and predictions from the model. An example of the utility of these profiles is the galaxy cluster hydrostatic mass measurement feature; this requires the measurement of 3D gas density profiles, 3D temperature profiles, gas mass, and total mass profiles.

To extract useful information from the generated spectra, we implemented a method for fitting models, creating an interface with XSPEC (Arnaud, 1996), a popular X-ray spectral fitting language. This interface includes the ability to fit XSPEC models (e.g. plasma emission and black-body) and simplifies interaction with the underlying software and data by automatically performing simultaneous fits with all available data.

2.1.4 Existing software packages

To the knowledge of the authors, no software package exists that provides features completely equivalent to XGA, particularly in the open source domain. That is not to say that there are no software tools similar to the module that we have constructed; several research groups including XCS (Lloyd-Davies et al., 2011), XXL (Giles et al., 2016), LoCuSS (Martino et al., 2014), and the cluster group at UC Santa Cruz (Hollowood et al., 2019) have developed pipelines to measure the luminosity and temperature of X-ray emitting galaxy clusters, though these have not been made public. It is also important to note that these pipelines are normally designed to measure a particular aspect of a particular type of X-ray source (galaxy clusters in these cases), and as such they lack the generality and flexibility of XGA. Our new software is also designed to be used interactively, as well as a basis for building pipelines such as these.

Some specific analyses built into XGA have comparable open source software packages available; for instance pyproffit (Eckert et al., 2020) is a recently released Python module that was designed for the measurement of gas density from X-ray surface brightness profiles of galaxy clusters. We do not believe that any existing X-ray analysis module has an equivalent to the source and sample based structure which XGA is built around, or to the product classes that have been written to interact with X-ray data products.

The XSPEC (Arnaud, 1996) interface we have developed for XGA is far less comprehensive than the full Python wrapping implemented in the PyXspec module, but scales with multiple cores for the analysis of multiple sources simultaneously much more easily.
Figure 2.2: A flowchart giving a brief overview of the XGA workflow.
2.1.5 Research projects using XGA

XGA is stable and appropriate for scientific use, and as such it has been used in several recent pieces of work; this has included an XMM analysis of the eFEDS cluster candidate catalogue (Turner et al., 2021), where we produced the first temperature calibration between XMM and eROSITA, a multi-wavelength analysis of an ACT selected galaxy cluster (Pillay et al., 2021), and XMM follow-up of Dark Energy Survey (DES) variability selected low-mass AGN candidates (Burke et al., 2021).

There are also several projects that use XGA nearing publication. The first of these is a hydrostatic and gas mass analysis of the redMaPPeR (Rykoff et al., 2014) SDSS selected XCS galaxy cluster sample (Giles et al., in prep.) and well as the ACTDR5 (Hilton et al., 2021) Sunyaev-Zel’dovich (SZ) selected XCS sample of galaxy clusters. This work also compares commonly measured X-ray properties of clusters (the X-ray luminosity $L_X$, and the temperature $T_X$ both to results from the existing XCS pipeline and from literature, confirming that XGA measurements are consistent with previous work. This process is repeated with XGA’s galaxy cluster gas and hydrostatic mass measurements, again showing they are consistent with previous work. XGA’s ability to stack and combine X-ray surface brightness profiles is currently being used, in combination with weak lensing information from DES, to look for signs of modified gravity in galaxy clusters.

2.1.6 Future Work

In the future we intend to introduce support for the analysis of X-ray telescopes other than XMM-Newton, first focusing on Chandra and eROSITA, and then possibly considering the addition of other X-ray instruments. This will include the same ability to find relevant data and generate data products as is already implemented for XMM, and will also involve the introduction of powerful multi-mission joint analyses to fully exploit the X-ray archives. We are also happy to work with others to introduce specific analysis features that aren’t already included in the module.

2.2 The overall design of XGA

From the outset, XGA was designed around a ‘source-based’ paradigm; that every XGA source instance (a Python term for a declared and setup class object) would represent a real astrophysical source. Such source class instances would have physical properties associated with them (location,
size, redshift, temperature etc.), and would also have any available, relevant, data associated with them as well. That way analyses could be run using the available data, and the physical properties inferred from those analyses could be stored back in the source class instance. Such a design was meant to make it easier for users to set up their analyses, as well as insulating them from steps like data collation and matching that can be automated.

Beyond sources and samples, XGA was designed to make interacting with and analysing XMM data products as easy as possible; as well as allowing such analyses to be performed in an automated or interactive fashion depending on the user’s choice. Each type of data product that could be generated for an XMM observation (e.g. images, spectra) would have an equivalent XGA product class. An instance of a product class would be used to store and interact with the data, and would itself be stored in the source instance it was generated for. Common techniques and functionality would be built into the product classes; e.g. images have a method that can easily use WCS information to convert from sky coordinates to detector coordinates, and a view method to generate publication-quality visualisations, spectra can also generate visualisations, as well as storing results from XSPEC fits performed using XGA’s XSPEC interface.

### 2.3 XGA Sources and Samples

This section will explain what XGA sources and samples are and what they can do. It also explains the mechanism that allows relevant XMM data for each source to be selected, and the source region matching criteria for different types of source.

#### 2.3.1 What are sources and samples?

XGA revolves around ‘source’ classes, instances of which are representative of real X-ray emitting astrophysical sources. Source classes are at the heart of any analysis performed using XGA, with each class representing a different type of object that emit X-rays. Distinctions are made between the different types of source due to the different information they can require for their analysis (galaxy clusters need overdensity radii for instance, whereas that isn’t a useful concept for an AGN). Different source classes also have some different procedures and methods built into them, as we often wish to measure different things for different types of source.

At their most basic, all that is required to define an XGA source is a position on the sky. The source will then find all relevant XMM data that is available for use on the system (see Section 2.3.4 for
more information). This approach means that the user doesn’t have to directly deal with data if they don’t want to, as XGA will fetch all available data by itself. When it comes to actually analysing and measuring quantities from the source, all the data will be used, not just single observations. If region files (generated by a source finding algorithm, detailing where sources were detected in a particular observation) are supplied to XGA along with the XMM data, an XGA source instance can match the input coordinates to a particular source region (see Section 2.3.5). In this case the XGA source instance will also be aware of where other detected sources are in the associated data, this allows it to define any contaminating sources that have to be excluded from spectrum generation and photometric analysis.

An XGA sample is a group of the same type of object that we wish to analyse as a population; for instance you might want to analyse multiple galaxy clusters and derive a scaling relation from them. There is a secondary benefit to using a sample object rather than multiple Source objects, a sample can be passed into any function that will accept a source, and the function will perform its job on every source in it. This not only makes writing analysis code easier and cleaner, but can also be more efficient; XGA will run many operations in parallel, rather than having to run them sequentially in a loop. When an instance of an XGA sample class is declared, it automatically discards any sources that do not fall on an XMM observation (providing a warning about such sources). This makes it very convenient for finding which sources in a large dataset have XMM data. Some sample classes have convenience methods included, which generate certain common output products for samples of that type of object; the ClusterSample class for instance can generate several common scaling relations by just calling a class method.

### 2.3.2 Types of source

This section introduces and summarises the source classes currently implemented in XGA. The base functionality of XGA sources is all implemented in the BaseSource class (fetching data, storing XGA products, providing basic information to the user), and as such new, specialised, source classes can quickly and easily be implemented if requested by a user. Indents in this bulleted-list indicate where a source class is a sub-class of another class.

- **BaseSource** - The superclass (parent class; everything implemented in BaseSource is inherited by its child classes) for all the other source classes, and the simplest of them all, there are very few circumstances where this class should be initialised by users. BaseSource only needs an RA-Dec position to be initialised, and if a name is not supplied by the user then
one will be generated from those coordinates (in the standard format).

- **ExtendedSource** - This is a general class for extended X-ray sources, it is also the superclass of the GalaxyCluster class. *XGA* will attempt to find a matching extended source from the supplied region files (Section 2.3.5), and if it does then that region will be used for any analysis. The user may also supply a custom circular region in which to analyse the object. Unless it is told not to, this *XGA* source will attempt to find the X-ray peak coordinates of this extended object, and then use that as the central coordinate for analyses.

  - **GalaxyCluster** - This class is specifically for the analysis of Galaxy Clusters, and is a subclass of ExtendedSource. Defining an instance of this class requires a redshift to be passed, as well as at least one overdensity radius (R\textsubscript{200}, R\textsubscript{500}, and R\textsubscript{2500} are supported). Also supports passing weak lensing mass and richness values, for use in multi-wavelength analyses. Point sources close to the centre of the cluster will not be removed, as they could be a misidentified cool core (see Section 2.3.5 for more information).

- **PointSource** - Similar to the ExtendedSource class in that this is a superclass for more specific point source classes. There are no methods in this class to produce radial plots for instance, as for point-like sources the ideal of a radial profile has very little meaning. When a PointSource is declared, an attempt will be made to match to a point source in region files, if they are supplied.

  - **Star** - An *XGA* class for the analysis of X-ray emission from stars within our galaxy. As such it does not accept a redshift argument, instead taking an optional distance measure. It will also accept either a proper motion magnitude, or a vector of proper motion in RA and Dec directions; this is in order to try and account for significant proper motion changing the location of the star in subsequent observations. Matching to region files also differs from the PointSource superclass, with point source regions within a matching radius being designated as matches - this is because the local nature of stars can throw up problems with the strict matching of RA-Dec within region that PointSource uses.

- **NullSource** - This class of source is an exception to *XGA’s* design philosophy that an *XGA* source represents a real X-ray emitting object. By default NullSource associates every ObsID available to *XGA* with itself (though you may specify which ObsIDs to associate with
it), and as such shouldn’t be used for astrophysical analysis. This class of source should only be used for bulk generation of products such as images and exposure maps.

### 2.3.3 Types of sample

The XGA sample classes mostly mirror the types of source available in XGA. Specific sample types can have properties or methods unique to that type of astrophysical object.

- **BaseSample** - The superclass for all the other sample classes, there are very few circumstances where this class should be initialised by users. All a BaseSample requires to be instantiated are two numpy arrays, containing RA and Dec values. Arrays of names and redshifts may also be supplied (though names supplied to sample definitions must be unique).

- **ExtendedSample** - For a population of some generic extended X-ray source. Supports setting custom analysis regions, either one aperture size for all or setting an individual aperture size for each source. This sample ensures that all sources have stacked images and exposure maps generated for them, which can then be used for finding the position of the X-ray peak (if selected by the user).

- **ClusterSample** - For a population of Galaxy Clusters. Just as with the GalaxyCluster source class, here we require that redshift and overdensity radius information be provided on declaration. Many convenient features have been added to this sample class, for instance you can retrieve temperatures of all clusters in the sample (if measured) using a ClusterSample method. You can also easily generate common scaling relations by calling methods of the ClusterSample class, using several different fitting methods.

- **PointSample** - For a population of some generic type of point source. Again only RA and Dec values have to be supplied, though redshift information can be provided if available. As this is a general point source class, no methods to generate scaling relations have been provided.

- **StarSample** - A simple extension of the Star class designed to allow easy analysis of multiple stars’ X-ray emission. The same arguments as the Star source class are taken, including proper motion and matching radius.
2.3.4 Fetching relevant XMM data

When an XGA source has been declared, it will attempt to find any available XMM data that are relevant to the object being analysed. As XGA does not download, reduce, and process XMM data itself, it has to be given access to an already processed set of data. In the case of the XMM Cluster Survey, this is the entire XMM archive, as new observations are routinely reduced and processed in preparation for serendipitous surveys.

When XGA is first launched, and anytime that it detects a new observation in the directory it has been told contains XMM data, creates an observation census. The census is a table containing a list of available (i.e. can be found in the directory supplied by the user) observation identifiers (ObsIDs), each ObsID’s pointing coordinate (RA and Dec), and columns to store whether each instrument’s (PN, MOS1, and MOS2) data should be used by XGA. Data are marked as unusable only when they were taken with the filter wheel in the closed position; these observations are identified using the FILTER keyword (either ‘CalClosed’ or ‘Closed’) in the headers of the event lists (see Section 1.4.2). Checks for new data occur when the module is first imported.

When a source is initially declared, it uses the user-provided RA-Dec coordinate to search for ObsIDs with pointing coordinates within 30′ (by default, the search distance is user-adjustable). This distance is approximately two times the XMM field of view radius, and is chosen to try and ensure that extended sources approaching the size of the field of view have any nearby XMM observations included as well. Relevant observations are found by calculating the Euclidean distance between the object coordinates and all ObsID pointing coordinates, then selecting only those which fall within the matching distance. XGA labels observations as on-axis if an ObsID’s pointing coordinate falls within 5′ of the object coordinate. This method to search the XGA census is implemented in a function separate from the source classes, so can be used without declaring XGA sources; it’s fully parallelised, and so can be used to quickly find nearby XMM observations for large samples of objects. A more capable, but slower, version of this function also checks whether the object coordinates fall on an XMM camera (though this is purely for independent use and it not used in XGA source declaration).

Selecting observations that are merely nearby an object of interest is indiscriminate, and can lead to selecting ObsIDs that contain very little (or no) emission from the object of interest. As such, most XGA source classes contain an additional (and optional, though turned on by default) cleaning step to identify and remove such observations. This cleaning step requires that a certain fraction of an analysis aperture is on an observation, otherwise it is ‘disassociated’ from the source and
discarded (each instrument of an observation is checked separately, due to damage to the MOS1 camera); both the fraction required and the analysis aperture are user configurable. Calculation of the coverage fraction of an analysis aperture follows these steps:

1. Exposure maps for the candidate observations are generated.
2. Exposure maps are converted into ‘detector maps’, where any part of the exposure map that is not zero is set to one. These indicate what is on and off of a camera.
3. Masks that exclude everything but the analysis aperture are generated for the candidate observations.
4. A mask is applied to its matching detector map, and the resulting masked detector map is summed, giving a coverage area in square pixels.
5. The coverage area is divided by the sum of the analysis aperture mask, giving a fractional area coverage of the analysis aperture.

2.3.5 Matching to regions

\textit{XGA} requires a pre-existing set of \textit{XMM} observations, and it also useful to have a pre-existing set of region files (see Section 1.4.3), as \textit{XGA} does not currently have source-finding capabilities. Ideally, every available ObsID should have a corresponding set of regions.

Once an \textit{XGA} source has finalised the set of observations that will be used in any potential analysis, it attempts to find a region from each of them that corresponds to the source. This involves searching the relevant lists of regions, finding any that are coincident with the user supplied source coordinates, and then deciding whether or not matching to that region is appropriate. For instance, matching a type of source (see Section 2.3.2 for a list of \textit{XGA} sources that correspond to astrophysical sources) that would be expected to have extended emission (e.g. a galaxy cluster) to a point-like emission that happens to be nearby would not make sense. As such any region lists supplied to \textit{XGA} are required to at least separate detections into two classes; point sources and extended sources. The distinction is made using the colours assigned to the sources, green indicates extended and red indicates point (see Table 1 for a full list of region types).

An advantage of having different \textit{XGA} classes analogous to astrophysical sources is that each type can have its own subtly different matching algorithm, depending on context. The first step for all the matching processes is to sort the regions by distance (of each regions central coordinate) from
Table 1: The types of region output by XCS’ XAPA source finder (apart from white regions), including both their colour and their description. These are also the region colours supported by XGA.

<table>
<thead>
<tr>
<th>Region Colour</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>Point source</td>
</tr>
<tr>
<td>Green</td>
<td>Extended source</td>
</tr>
<tr>
<td>Magenta</td>
<td>PSF-sized extended source</td>
</tr>
<tr>
<td>Blue</td>
<td>Extended source with significant point source contribution</td>
</tr>
<tr>
<td>Yellow</td>
<td>Extended source with less than 10 counts</td>
</tr>
<tr>
<td>White</td>
<td>Custom regions created through an XGA process</td>
</tr>
</tbody>
</table>

the source coordinate, this is done immediately upon the read-in of the region files. Initial matches, where the source coordinate falls within a source region, are then assigned for each observation associated with the XGA source.

The next steps differ based on the XGA source class that has been declared; here are examples for some currently implemented classes:

- **ExtendedSource** - The base class for specific extended source classes (such as GalaxyCluster), the only extension to the base matching process is that only extended source regions are considered as matches.

- **PointSource** - The base class for specific point source classes (such as Star), the only extension to the base matching process is that only point source regions are considered as matches.

- **GalaxyCluster** - The type of all initial matching regions for all ObsIDs is ascertained, only those that are extended sources are considered as matches. A check is performed for point sources within ~0.15R_{500} of the user coordinates, if any are found then they are classified as ‘alternative matches’. This is to account for the possibility that a galaxy cluster cool core has been falsely detected as an independent point source. Finally, an attempt is made to check for fragmentation of clusters by the source finder, which can cause issues where low redshift clusters are split up into multiple extended sources. Any non-matched sources which are extended and intersect with a source region that has been designated the actual source will be classified as alternative matches. This is also useful in making sure regions which are not consistent across observations do not remove chunks of the cluster. Figure 2.3
Figure 2.3: An XGA generated RateMap visualisation of the Abell 478 galaxy cluster. A mask has been applied to remove all contaminating sources (any region that has not been matched to the source being analysed). The mask also applies a circular analysis aperture. The cross-hairs indicate the user supplied coordinates of the cluster. A large part of the cluster’s emission near the centre of the image has been spuriously removed, something that is corrected by checking for extended source regions that intersect with initially matched source regions.
demonstrates what can happen without this last step.

• **Star** - The type of all initial matching regions for all ObsIDs is ascertained, only those that are point sources are considered as matches. Matching to region files differs from the PointSource superclass, with point source regions within a matching radius being designated as matches. This is because the local nature of stars can cause problems with the strict matching of RA-Dec falling within a region that PointSource uses.

It is allowable for there to be no matching region for some (or all) ObsIDs, in which case XGA classifies the source as undetected. All analyses can still be performed in the case of a non-detection, but with warnings that results are likely to be unreliable or impossible to obtain.

Any regions which have not been classed as either a match, or an alternative match, to the source being analysed are designated as ‘contaminating sources’. The lists of contaminating regions for each ObsID are used to build masks that remove their emission, to avoid biasing analyses. All regions from all ObsIDs are used to build each mask; a mask for a single observation generated by a source with multiple ObsIDs will have regions from all ObsIDs included. Additional masking can be included, Figure 2.3 demonstrates an image mask that has had an analysis aperture included along with the masks for contaminating sources. XGA sources can also generate masking commands in XMM detector coordinates, specifically for the generation of spectra using XMM SAS tools.

It is possible to run XGA without region files, and then an interactive feature of the Image product class (see Section 2.4.5) can be used to draw on regions.

### 2.4 XGA Products

This section will briefly introduce XGA’s product classes, as well as providing information on some of the most useful analysis methods that are built into them. It will also explain how the products are generated, whether from an XMM SAS routine, or by XGA functions.

#### 2.4.1 What are products?

XGA products are Python classes used to interface with various types of X-ray data, both wrapping files produced by XMM’s SAS routines, and storing/allowing access to various XGA generated
data. They have general functionality, such as making sure files exist and parsing SAS output for errors, but they also provide functionality specific to each type of product.

When using XGA sources and samples, instances of product classes are instantiated and setup by internal XGA functions; in other words the user never needs to set them up themselves. However, it is quite possible for users to setup their own products, if they wish to leverage the power of XGA’s functionality with data from an external source. Setting up instances of the XGA Image class, for instance, is used to create publication quality visualisations of eROSITA images in Chapter 4; eROSITA is not yet supported by the XGA source and sample system, but product classes can still be useful.

Products instances setup by an XGA source are immediately ‘associated’ with (stored in) the source object, though beyond storing the name of the source in the product’s internal structure, they have no knowledge or awareness of the XGA source and its properties. This was by design, to make sure that any functionality built into a product would work regardless of the type of XGA source, and even if it was defined independently of any source at all. Products can be easily retrieved from the source using a set of built-in methods if the user needs to directly interact with them.

2.4.2 The four base types of XGA product

While there are many different types of XGA product class (25 currently in the main branch), the majority of them are specialised sub-classes of the four product classes that are described in this section. This will be a high-level overview of the four main base classes, with later sections describing specific products and their functionality.

- **BaseProduct** - The superclass for XGA product classes that are used as wrappers for files generated by SAS. An instance of BaseProduct itself should never be declared, but this class contains the functionality that is common to all XGA products that act as wrappers (or interfaces) to files. This includes simple sanity checks such as making sure that the supplied path is valid, storing information such as the ObsID, instrument, and energy band the product is from, to a mechanism that parses SAS output during generation and searches for errors/warnings.

- **BaseAggregateProduct** - This is the superclass for XGA product classes that are made up of combinations of files generated by SAS. The ‘Aggregate’ part of the name refers to the combining of multiple other products into a single multi-part product. The most important
example of an aggregate product in XGA is the AnnularSpectra (see Section 2.4.10),
where sets of spectra generated in different annuli are brought together to form a single
object that can be used for investigating how spectral properties change with radius. The
BaseAggregateProduct class is built to store and index sets of individual products
that are sub-classes of the BaseProduct class. As such its features are largely centered
around sanity checking the component products it is made up from, and allowing them to
be retrieved in a way related to their generation; i.e. if the aggregate product is made up of
individual products that were generated at different spatial locations, those locations should
be able to be used to retrieve products.

- **BaseProfile1D** - The superclass for radial profiles generated by XGA, with the ‘1D’
in its name referring to the radial nature of the profile; it can still be used as a base for
3D radial profiles such as the three-dimensional mass distribution. This superclass con-
tains a significant amount of functionality, including storing information on the data points
and their uncertainties, identifiers for the ObsID and instrument used to generate the in-
formation, attributes which set-up the physical context of the profile (i.e. identifiers which
ensure you couldn’t accidentally use a temperature profile in place of a density profile). The
BaseProfile1D class also includes mechanisms for fitting XGA model instances (Sec-
tion 3.4.1) to profiles (with multiple choices of fitting algorithm), visualising the posterior
distributions and chains of model parameters if fit using emcee (Foreman-Mackey et al.,
2013), and making publication quality figure showing the data and any model fits.

- **ScalingRelation** - A special base class of XGA product, as ScalingRelation does
not act as a superclass for any other classes. It is designed to hold data points (including
uncertainties) and fits that have been used to build a scaling relation. Once declared a
ScalingRelation instance can be used to create publication quality visualisations of
the relation, view posterior distributions and chains if an MCMC based method was used to
fit the data, and to predict a value for Y given an X input.

Two other high-level XGA product classes have also been implemented,
BaseAggregateProfile1D and AggregateScalingRelation. Akin to the
BaseAggregateProduct, they are classes of product made up of multiple profile and
scaling relation instances respectively. They are not mentioned in the list above as they will never
be used as a superclass to another XGA product, and they have only one function; to help create
visualisations of multiple radial profiles and scaling relations respectively on a single figure. It is
a common analysis requirement to compare different scaling relations (for instance a new result
to a result from literature), or to plot multiple radial profiles at once for comparison, and these classes were designed to ensure that process was as simple as possible. The ‘aggregate’ versions of profile and scaling relation classes can be created by simply combining two (or more) profile or scaling relations with the Python ‘+’ operator; they are just added together.

2.4.3 How does XGA generate XMM data products using SAS?

The Science Analysis System (SAS; see Section 1.5.1) is the suite of software released and maintained by the XMM team, designed for the creation of XMM data products. The main downside to these tools is (at the time of writing) most of the frequently used tools can be run only from the command-line, and the various data and configuration arguments required mean that setting up commands can be complex.

As such an interface to SAS was designed and implemented in XGA, both to allow the user to run common (and useful) SAS commands from a Python interface (arguably the most commonly used programming language in Astronomy), and to perform the complex setup processes required to run SAS methods in a completely consistent manner. Each ‘task’ (such as generating an image, or a spectrum) has a separate method which generates the requisite commands, populating them both with information from the XGA source object that a product is being generated for and with user-supplied configuration information (such as the energy range in which to generate an image).

The bulk-generation of data products such as images and spectra is an embarrassingly parallel problem (where little effort is required to break a workload down into a set of parallel tasks), and as such the XGA SAS interface will run tasks in parallel, using the number of cores allotted to XGA in the configuration file. Each XGA source has a SAS queue, which is populated by SAS commands generated by various tasks until the queue is retrieved by XGA’s ‘SAS run’ function, at which time the commands are run in parallel, and the resulting files are wrapped by XGA product classes. Sometimes multiple files for a single task must be generated, in a specific order (such as generating a spectrum and its ancillary/response files), and consequently such XGA SAS tasks are organised into ‘stacks’. Such command stacks can be added to the XGA source queue, and they will be executed as a single task rather than in parallel. So if there are many cores available on the machine used for analysis, it is generally more efficient to be studying a sample of sources rather than just one.
2.4.4 EventList

A very simple product class which differs only slightly from the *BaseProduct* class, it is only used to store path and header information for the *XMM* event lists of observations associated with a source. This class was introduced so that information about event lists could be retrieved from *XGA* source instances (by *XGA* functions and by the user) in an analogous way to how all other *XGA* products can be accessed. *EventList* instances are set up for every ObsID-instrument combination associated with a given source, after the source has identified relevant observations per Section 2.3.4.

2.4.5 Image

An extremely useful *XGA* product with many extra features, it is used as a wrapper for fits images generated using the *XGA SAS* interface (or fits images from an external source, if the user is not using *XGA* source classes). A user can generate single ObsID-instrument images for all observations associated with an *XGA* source using the `evselect_image` function, with the primary user-configurable argument being the energy range the image is generated in; the default energy range is 0.5-2.0 keV. Additionally, if the user wants to automatically make stacked (or ‘combined’ as they are called in *XGA*) images from all data available for a source, then the `emosaic` function can be used. This will generate combined images or exposure maps within a user-specified energy range, storing them in *Image* class instances, but with information about all ObsIDs and instruments that went into the combined image stored in the product instance.

The data and header information are read into memory only when required (i.e. when a function or the user call a method of the *Image* class that requires either image data or header information), this saves using excessive amounts of RAM for large samples of objects. Once declared the user can use the ‘data’ property to get access to the image data in the form of a numpy array, making it easy to analyse the data.

One of the most convenient pieces of functionality implemented in the *Image* class is a method that uses World Coordinate System (WCS) information extracted from the image header to convert a coordinate (or a set of) between different coordinate systems. The user just needs to pass the input coordinate as an Astropy (see Section 1.5.3) quantity, the method will check that the quantity unit is allowed in this context (degrees, pixels, or *XMM* detector coordinate units), and then convert the coordinate to an allowed unit. *XMM* images typically contain two sets of WCS
information, one that converts between image pixels and sky coordinates (RA-Dec), and one that converts between image pixels and XMM detector coordinates (required for setting up SAS spectrum generation masks). The convenience function built into the Image class will automatically determine which WCS information is required for the requested transformation.

A view method has also been implemented, with comprehensive options for altering the appearance and quantity of data that is displayed and overlaid on top of the image. This includes the ability to add crosshair(s) to indicate certain positions on an image, applying a mask to the data and automatically zooming in so that areas of the image with no data are removed, overlaying source regions, and changing the colour-map and scaling. As such the view method is extremely versatile and can easily be used to make publication-quality visualisations.

An extension to the view method (called edit_regions) has been implemented in such a way that the visualisation can be directly interacted with in order to edit or add source regions. This can be extremely useful in the case where your source detector of choice has not detected a source, or in the case where no source finder has been run and the user wishes to manually add regions. This feature allows a user to quickly fine tune regions for large samples of objects, with all changes being automatically passed back into XGA analyses, rather than having to manually edit region files.

2.4.6 ExpMap

A subclass of the XGA Image class (Section 2.4.5), this is a very simple extension to Image that adds a method to easily retrieve an exposure time at a given angular or pixel coordinate. The get method for exposure will automatically convert RA-Dec and XMM detector coordinates to pixel coordinates, before fetching the exposure map. The entire exposure map data array can be accessed in the same way as data can from the Image class.

2.4.7 RateMap

The class that most photometric analyses are based around, and another subclass of the Image class. This class represents a ‘count rate map’, essentially the image divided by the exposure map; this helps to account for the varying effective area of the detector over the field of view due to vignetting. Setting up an instance of this XGA product class requires a matched pair of Image and ExpMap instances; among the first things that the RateMap class does upon initialisation is to
check that the passed `Image` and `ExpMap` have the same ObsID, instrument, and energy range.

As a subclass of the `Image` class, `RateMap` inherits all of its methods and properties, but some methods have to be overwritten to account for the fact that `RateMap` doesn’t represent a single file (like `Image` and `ExpMap` do). Unlike its superclass, `RateMap` can’t load data from a single file on demand (when the user or a function asks for access to the data), it has to load in both the image and exposure map data, then divide one by the other.

In addition to the functionality inherited from `Image` (such as easy coordinate conversion and a powerful set of viewing and interactive viewing methods) the `RateMap` class has new methods for tasks like X-ray peak finding, signal to noise/background subtracted counts calculation, and the creation of ‘detector maps’ as mentioned in Section 2.3.4.

In these sections some of the specialised capabilities of the `RateMap` class are introduced, with brief explanations of what they are used for.

### 2.4.7.1 Addressing hot pixels with `RateMap` edge masks

Count-rate maps created from `XMM` observations can exhibit ‘hot pixels’ on the edges of the detector chips, where the exposure map calculation in some spots has calculated an erroneously low value. As such when the image data is divided by the exposure map data to calculate the count-rate map, pixels with much erroneously low exposure map values inherit a boosted count-rate value, such that they are among the brightest pixels in the count-rate map. This can cause issues in several aspects of photometric analysis, but particularly when we attempt to find the brightest pixel associated with a particular source (known as the X-ray peak), a very useful quantity (particularly for the analysis of extended objects such as galaxy clusters). `XGA RateMap` instances can make use of their access to exposure map data to generate a mask that removes these pixels, by obscuring any pixel on the edge of a detector.

For a given ObsID-instrument combination, an edge finding algorithm was developed to map the detector and decide which pixels are on a detector edge. The algorithm is run on a ‘detector map’; a copy of the exposure map where every non-zero value is set to one, and as such the only values in the array are one and zero, where one indicates being on the detector and zero indicates being off the detector. The modified detector map array is differentiated in the X and Y directions, and then the two resultant arrays are added together to form an ‘edge map’ (edge maps show where the edges and nodes of the detector boundaries are, rather than acting as a way of masking them).
Figure 2.4: A masking array generated by an XGA RateMap generated for the EPN instrument of the XMM observation 0201903501. Black parts of the mask indicate areas set to zero (which means those pixels will be masked when this is applied to image data), white parts indicate areas set to one (meaning nothing will change when this is applied to image data).
Figure 2.5: A RateMap view (zoomed-in) before (on the left hand side) and after (on the right hand side) the edge mask was applied. The visualisations are centered on a ‘hot pixel’, with the white dashed circle highlighting its location. Comparing the two visualisations demonstrates that the edge mask has removed the hot pixel.

As such an edge map pixel $e_{i,j}$ is

$$e_{i,j} = (p_{i,j} - p_{i+1,j}) + (p_{i,j} - p_{i,j+1})$$

(2.1)

where $p_{i,j}$ is the pixel value of the modified exposure map at coordinates $i, j$.

This makes it easy to know when you’re going from ‘not on a chip’ to ‘on a chip’ and vice versa; $e_{i,j}$ is -1 when moving from ‘not on a chip’ to ‘on a chip’, and 1 when moving from ‘on a chip’ to ‘not on a chip’. The modified exposure map is differentiated in both X and Y directions to make sure we capture all the edges, but some will be detected in both directions, so when the arrays are added together there could be edge map pixels with a value of 2, which generally means they are a corner of a chip. As we wish to know where the edges are, any values of -1 are shifted over by 1, then set to 1, in the direction of the differentiation they belong to. The array is then inverted, so everywhere but the edges is 1, see Figure 2.4 for an example. The resulting mask can multiplied with the RateMap data to remove possible bright pixels, a step which every RateMap performs automatically when the data array is retrieved from it (though it is still possible to access the unedited data). See Figure 2.5 for an example of its an edge mask’s application to an XGA visualisation.
Figure 2.6: Demonstration of the simple peak finding method, run without contaminating source masking for demonstrative purposes. The left hand image shows a full count-rate map of the EPN data of observation 0201901401, for context. The middle image shows the hot pixel highlighted in Figure 2.5, and demonstrates that the simple peak finder run without edge masking selects it as the brightest pixel. The right hand image shows that once the edge mask is applied, the simple peak finder selects a more obviously physical source as the X-ray peak, a bright point source.

2.4.7.2 Finding X-ray peak coordinates using an XGA RateMap

This section explains and introduces the two peak finding methods XGA built into the RateMap class. Finding the coordinates of the peak of X-ray emission can be useful for the analysis of X-ray sources (extended sources in particular). For galaxy cluster analysis it can be difficult to define the ‘centre’ of a given cluster, but knowing where to centre your analyses is important, especially when we wish to produce radial profiles of cluster properties. Different methods are used by different analysis teams, and vary depending on the wavelength of the observation; in optical observations, for instance, the brightest cluster galaxy (BCG) is often used as a proxy for the centre. X-ray peaks are considered a good proxy for the true centre of a galaxy cluster (Yan et al., 2020), as X-ray photons trace the intra-cluster medium, and in effect the potential well of the halo. X-ray derived central coordinates have been used to quantify miscentering between redMaPPer coordinates and the true centre of the cluster by the Dark Energy Survey (Zhang et al., 2019).

The Simple Peak Finder

The first peak finder implemented in RateMap simply searches for the brightest pixel, converts the pixel coordinates to RA-Dec using the coordinate conversion function that RateMap inherits from Image (see Section 2.4.5). The implementation is efficient as it takes advantage of the speed
of numpy array operations.

The data array in which the peak is searched for will automatically have an edge mask (Section 2.4.7.1) applied to remove any hot pixels, as they would be very likely to be selected by this algorithm. The simple_peak method supports a mask being passed, which in real-world use cases will remove any known contaminating sources and restrict the search area to a given aperture (see Section 2.4.7.2 for a information on how peak finding is called by XGA sources).

To demonstrate the simple peak finding method, it is applied to the EPN data of observation 0201901401. This is a targeted observation of the galaxy cluster Abell 907, though in this case no mask to remove contaminating sources is applied, for demonstrative purposes. First the simple peak finder is applied with edge masking turned off, to demonstrate the issue that can arise if hot pixels are not dealt with, see the middle image of Figure 2.6. You can quite clearly see that the pixel that has been selected is a) on a chip edge, and b) much brighter than anything else in the ratemap. Ideally this would have selected the cluster that we can see nearby, but evidently we must deal with these spuriously bright pixels that can appear on chip edges.

Then edge masking is turned back on and the simple peak finder is applied again, and we see that it selects a bright point source as the X-ray peak position. Intuitively this makes sense, as it is obvious from the colour map of the visualisation that the point source is very bright. Of course in a realistic situation the point source would be masked out as a contaminating source, but even then there is no guarantee that all the point source emission would be removed. This is actually the case for this observation, and a bright pixel on the outskirt of the same point source is still selected as the brightest pixel, whereas we wish to know the peak of the cluster emission. This is the motivation for the new peak finding method developed for XGA.

Hierarchical Clustering Peak Finder

To account for any point source remnants in the RateMap, a new method was developed that uses a hierarchical clustering algorithm to choose the pixels most likely to belong to the (in this case) cluster for which a peak coordinate needs to be measured.

Hierarchical clustering is a type of unsupervised machine learning algorithm that can be used to cluster data points in some arbitrary feature space until a completion criteria has been met. In this case clustering is performed on X and Y spatial coordinates, so intuitively it is quite easy to understand, but the algorithm can cluster any kind of information. It starts off by assuming
Figure 2.7: An example of the point clusters found by RateMap’s hierarchical clustering peak finder algorithm. The black crosses indicate the largest point cluster constructed via hierarchical clustering, the blue crosses indicate a secondary point cluster coinciding with the bright point source selected by the simple peak finder. White cross-hair indicates the X-ray peak selected by the algorithm.
that each point is a separate cluster, then repeatedly identifies the clusters that are closest together and combines them. This XGA implementation uses the scipy `fclusterdata` function, with a distance criterion of five pixels by default, though the distance can be changed by setting the `clustering_peak()` method’s `max_dist` argument.

The spatial coordinates of the pixels with the top 5% (by default) of values in the `RateMap` are selected, then the hierarchical clustering algorithm is run on them. Once it’s complete the largest cluster of points is selected and assumed to be the relevant source. This excludes any small patches of emission that might be left over from point sources, and so you can just select the pixel with the maximum value in the chosen point cluster.

A visualisation of the hierarchical clustering peak finder’s point clusters can be constructed using the `view()` method of the `RateMap` class. Figure 2.7 shows the same observation as the left hand image of Figure 2.6 (though zoomed into the region of interest), with two different point clusters overlaid. In blue (to the left side of the visualisation) is the point source that was previously selected as containing the X-ray peak, and in the middle of the visualisation is the black point cluster representing the galaxy cluster for which a peak is to be found. We can see that the clustering peak finder has noticed the bright point source to the left, but has decided that because the galaxy cluster point cluster is larger, it will take the peak from there. As with the demonstration of the simple peak finder, this demonstration does not include any of the masking of contaminating sources that would occur in a real use case; this is a worst case scenario, and the correct peak was still selected.

**How are peak finders called in XGA sources?**

The implementation of this process in the XGA source classes is iterative. Firstly, an aperture with a radius of 500 kpc (or 5′) is placed at the user supplied source coordinates, with interloper sources being masked out. The radius is only set to 5′ if the user has defined an extended source with no redshift information (which is allowed), if the source is a `GalaxyCluster` object, redshift information is required so it will always use 500 kpc.

Then the algorithm follows these steps (for up to 20 iterations before it throws an error):

1. Runs the hierarchical peak finder.
2. Re-centres the search aperture at the new coordinates.
3. Checks to see if the new peak is within 15 kpc (or 0.15′) of the last central coordinate.

4. If it is, consider the peak converged and exit, if it isn’t (or if the algorithm is on the first iteration), go back to 1.

Initially run on the combined ratemap, then on the individual ratemaps of the different observations/instruments. If the combined ratemap peak won’t converge, a hard error will be thrown (as this is what XGA analyses uses as the centre), the individual peaks are allowed to not converge.

The user can also choose to use their initial coordinates as the cluster centre, if they don’t want to find and use the peak.

### 2.4.7.3 Calculating SNR and background subtracted counts with a RateMap

The calculation of background subtracted counts or signal to noise (SNR) within a particular analysis region is a common task in X-ray astronomy, particularly in assessing the quality of a particular observation. As only photometric data (and information on the analysis aperture) are required for such calculations, they were added to the RateMap class as methods.

Background subtracted counts are calculated using

\[
C_{BS} = C_S - C_B \times \frac{A_S}{A_B},
\]

(2.2)

where \(C_{BS}\) are the final background subtracted counts, \(C_S\) are the number of counts in the source analysis aperture, \(C_B\) are the number of counts in the background analysis annulus, \(A_S\) is the area of the source analysis aperture, and \(A_B\) is the area of the background analysis annulus.

The uncertainty on the background-subtracted counts (\(C_{BS}\)), is calculated assuming Poisson statistics,

\[
\sigma_{C_{BS}} = \sqrt{C_S + C_B \times \left(\frac{A_S}{A_B}\right)^2},
\]

(2.3)

and can then be used to calculate the signal-to-noise of the particular source. The signal to noise is simply calculated using

\[
SNR = \frac{C_{BS}}{\sigma_{C_{BS}}},
\]

(2.4)
Every source automatically calculates the combined signal to noise (i.e. using a stacked RateMap) within whatever analysis apertures it has been assigned; a GalaxyCluster might calculate it within $R_{500}$ and $R_{200}$ for instance, depending on which overdensity radii it has been supplied with on declaration. These combined SNRs are displayed in the `info()` method of each source, which additionally provides a complete set of summary information for the source.

Each XGA source also has methods that allow it to calculate the SNR and background subtracted counts for a given aperture for every ObsID-Instrument combination that is associated with it. The calculated SNR or counts are then returned, along with an array detailing which result belongs to which ObsID-Instrument combination. Another pair of methods will use the bulk SNR and count calculation methods implemented in each source to rank the ObsID-Instrument combinations in order of ascending SNR or counts; this is useful in other XGA functions when it comes to deciding which data to use for particular analysis choices (such as deciding on annular radii for the generation of annular spectra, see Section 2.4.10.1).

The PSF class is a subclass of XGA’s Image class, and is used as an interface to two-dimensional PSF images that describe the XMM PSF at a given position on a given camera. As a sub-class of Image, this class inherits all of the superclass’s methods and properties. Few methods have been added, and it is unlikely a user will ever need to interact directly with this. The added functionality here is the ability to re-sample the PSF at a scale (pixel scale in degrees, psfgen does not generate realisations of models at the same scale as our images) provided by a passed in Image object.

### 2.4.8 PSF and PSFGrid

PSF images are generated by the psfgen SAS routine. This routine allows the user to choose from several different XMM PSF models, and has a corresponding XGA interface function that allows XGA users to generate realisations of the PSF models as a 2D image. The default PSF model is ELLBETA, which was developed by Read et al. (2011), and accounts for the complex morphology of the XMM PSFs, as well as the change of PSF with off-axis position.

An analogous aggregate product has also been implemented, the PSFGrid class. This is designed to store and provide access to a grid of PSF images generated for different points on an XMM detector. All component PSFs are stored with a key based upon their central position in Image coordinates, where the image in question is the energy and ObsID-Instrument combination that the PSFs were generated for. The XGA SAS interface to the psfgen routine is designed to take a source, and for a given number of ‘psf bins’ $N$, generate an $N \times N$ grid of PSFs, so as to sample
Figure 2.8: An example grid of XMM PSFs generated and stored by XGA. This demonstrates the complex morphology and spatial variation of the XMM PSF effects. This particular grid of PSFs is for 0.5-2.0 keV band for the PN instrument. PSFs near the centre of this image are taken from close to the aimpoint of the observation.
the changing PSF properties over the field of view of an instrument.

### 2.4.8.1 PSF correcting XMM images

The three XMM cameras (PN, MOS1, and MOS2) all have quite significant (but slightly different) PSF effects, that change shape and size with off-axis position. The full-width half maximum of the MOS on-axis PSF is 5″, and the full-width half maximum of the PN on-axis PSF is (6″). As such photons can be spread a considerable distance from where they ‘should’ have landed on the detector, meaning that it is advisable to address the PSF in some analyses.

Specific analyses often have different ways of dealing with the PSF; for instance, the generation of 3D X-ray emissivity profiles for galaxy clusters have previously used a 1D characterisation of the PSF effect to account for photons being displaced from their ‘true’ position. As XGA is a generalised Python package, capable of analysing any kind of X-ray emitting source, a generalised method applicable to any source is preferable.

As such a way of correcting PSF effects in whole XMM images was developed, and as it is just applied to the images it is applicable to any kind of source (no source information is required for correction). The Richardson-Lucy (Richardson, 1972; Lucy, 1974) algorithm is an iterative method of recovering a source image after its been blurred by a PSF. As the PSF effect cannot be described by a single function (see Figure 2.8 for a demonstration of how an EPN PSF changes over the field of view), the source image is split up into a grid and a PSF realisation is generated for the central coordinate of each PSF bin (see Figure 2.9 for a visualisation of the gridding process). Following this the \( N^2 \) copies of the image are made (where \( N \) is the number of PSF bins per side), and each copy is iteratively convolved per the Richardson-Lucy algorithm. Finally the PSF-corrected image is assembled by stitching together the PSF bin regions from each separately convolved image, forming a whole image where regions have been PSF corrected with a local PSF realisation.

As previously stated, PSF effects vary between XMM instruments and with off-axis position. As such there would be no valid way of PSF correcting a combined (or stacked image). As such, when the user calls the `rl_psf` function to generate PSF corrected images, each ObsID-Instrument combination are PSF corrected separately, and then stacked at the end to create a final combined image. Figure 2.10 shows zoomed-in visualisations of a galaxy cluster and the Castor star system, providing an before and after the application of the PSF correction method.
Figure 2.9: An *XMM* observation with a grid overlaid indicating the PSF bins that the observation was divided up into by *XGA*. The diamonds indicate the central coordinates of the PSF bins (which is where a realisation of the PSF model is generated). This is a companion to Figure 2.8, and is the PN data for observation 0201903501.
Figure 2.10: Two demonstrations of XGA’s image PSF correction. The top left figure shows an observation of the Castor sextuple star system, which is such a bright point source that the emission has been spread over a wide area, as well as into characteristic spokes that can be seen in Figure 2.8 as a feature of XMM’s PSF. The top right figure is the same observation after it has been PSF corrected; most of the emission now lies where we would naively expect it to. The spreading and spoking effects have been almost entirely eliminated. The bottom left figure shows the Abell 907 galaxy cluster, with contaminating sources masked (the mask was applied after PSF correction). The bottom right figure shows the image after PSF correction, with the core of the cluster more concentrated and the point source on the left more completely removed after masking.
2.4.9 Spectrum

A complicated sub-class of BaseProduct, the Spectrum class is designed to wrap and provide a Python interface to spectra generated by SAS, along with all the ancillary files required for their analysis (redistribution matrix files, auxiliary response files, and background spectra). Similarly to the Image class (Section 2.4.5), instances of this class, while normally generated and setup by internal XGA functions, can quite easily be setup by a user from an external data source.

The Spectrum class has some added functionality over its superclass, including a view method designed to visualise the spectrum and create publication quality figures, extra properties to store and retrieve paths to ancillary files, and the infrastructure required to store and retrieve the results of XSPEC fits to spectra. There are also properties that store how the spectrum was generated, which can be important as there are plenty of user-configurable options when generating spectra, such as whether any grouping has been applied to the channels, and what metric was used to group on.

To generate spectra using XGA, the user can call the evselect_spectrum function. It will take an XGA source (or sample) object, an analysis aperture that you wish to generate spectra within, and configuration arguments on how the spectra should be generated (whether to grouping, grouping metric, number of CPU cores to use for generation). Both an inner and outer radius can be set, to generate a single annular spectrum, but by default the inner radius is set to zero so as to generate a circular spectrum. The function then cycles through the sources (if a sample was passed), and through each source’s associated observations, setting up the commands necessary to generate valid spectra.

The first step is to use the XGA source’s knowledge of contaminating source regions to construct a mask in the XMM detector coordinates required by the SAS evselect tool. Each instance of an XGA source has a method called regions_within_radii which, given inner and outer radii, will return a list of regions that a) have not been matched to the source being analysed, and b) have some part of their boundary that falls within the specified radial bounds. This method call is repeated for the background spectrum inner and outer radii (by default they are $1.05R_{outer}$ and $1.5R_{outer}$, though the user can change that choice and those boundaries aren’t necessarily valid for all spectra. Once the list of relevant contaminating source regions for the main spectrum and the background spectrum have been assembled, they are converted to XMM detector coordinates by using the coordinate conversion method of the XGA Image of the same ObsID-Instrument combo. They are then assembled into two strings that explicitly exclude those regions, and include a
region describing the spectrum being generated. They are then passed to the ‘expression’ keyword of the SAS evselect tool used to generate the main and background spectra. This is how evselect is told which events from the event list to select based on their spatial coordinates.

The XGA evselect_spectrum function then sets up the SAS commands to generate all files necessary for spectral analysis:

1. Sets up the previously discussed SAS evselect commands for the source and background spectrum. Events in the event list are only selected if they have a ‘pattern’ of less than four (for PN) or less than twelve (for MOS), which are an indication of how many pixels a particular event has been detected in (charge from a particular pixel of the detector can spill over into adjacent pixels)\(^1\). The choice of pattern selection was informed by XMM calibration team recommendations. Events must also have a ‘FLAG’ value of zero to be selected, which excludes events that land on/near known bad pixels and chip edges.

2. The Redistribution Matrix File (RMF; defines mapping between energy and detector channel) command is generated, taking the path to the spectrum file as one of its arguments; as such RMF must be generated after the spectrum file. XGA addresses this problem by creating a ‘command stack’, as explained in Section 2.4.3, where commands in the stack will not be executed in parallel, but sequentially. The rmfgen command also requires context as to whether the source is extended or not, which can be extracted from the type of XGA source passed into evselect_spectrum. If the source is extended a detector map (an image of the observation in detector coordinates) is used as part of the RMF calculation, giving information on the emission distribution.

3. Accompanying Auxiliary Response File (ARF) command is generated. The ARF describes the effective area of the detector as a function of energy at the generation position, and is used in the fitting process to account for how good XMM instruments are at detecting different energies of photon. The arfgen command requires the same information as rmfgen about whether the source is extended or not, and uses the same detector map (if extended). It also takes the previously generated RMF as an argument.

4. The areas of the source and background regions are calculated, accounting for any chip gaps and removed regions. This is performed using the SAS backscale tool, and the resulting areas are written to the headers of the spectral files. Knowing the areas allows for area-based scaling of background counts during the XSPEC fitting process.

\(^1\)XMM pattern definitions
5. Finally, the SAS routine specgroup is used to group the channels in the source spectrum, so as to increase the signal to noise of the spectrum. The user can choose to group on minimum counts in a bin, minimum signal to noise, or minimum source to background ratio. The default for XGA spectra is to group in order to achieve a minimum of 5 counts per bin.

Following generation and Spectrum instance set-up, the spectrum is stored in the XGA source it was generated for. Such spectra can be used with the XGA XSPEC interface to perform and retrieve parameters from spectral fits.

2.4.9.1 Fitting models to spectra with the XGA XSPEC interface

XSPEC is a tool used for the spectral fitting of models to data for X-ray emission (see Section 1.5.2). It is used by XGA to fit emission models to the spectra generated by the SAS interface and stored in XGA Spectrum. A Python interface to XSPEC (called PYXSPEC does exist, but has limitations that made it unsuitable for use in XGA, including difficulties in installation and lack of support for running many XSPEC fits in parallel.

As such XGA generates XSPEC scripts and then uses the same method as the SAS interface to execute them through the command line, executing as many fits in parallel as there are CPU cores that have been assigned to XGA. As scripts for XSPEC fitting can be written in TCL, a general fit script template was written for XGA; it is completely configurable and in theory should support any combination of XSPEC models, though they have not all been tested. While some very specialised spectral analyses will not be able to be executed through this interface, it is suitable for the majority, and further XSPEC functionality can always be added to XGA on request.

Various XGA functions representing different model combinations have been implemented to properly populate the template scripts with parameter values. The models currently represented include tbabs*apec (a hydrogen column absorption model combined with a plasma emission model), tbabs*powerlaw (a hydrogen column absorption model combined with powerlaw), and other absorbed powerlaws with a gaussian absorption feature. The functions take an XGA source or sample, an analysis aperture (including an annulus with a non-zero inner radius), and use source object methods to retrieve matching spectra (or call the evselect_spectrum function to generate them). They then use the spectra and source information, alongside model information and various other user-configurable parameters to populate a template XSPEC script. The user can set whether parameters are to be frozen during fitting through arguments passed to the Python
function, as well as the start values for model parameters, and other details such as which solar abundance table should be used. The \textit{XGA} source is used to retrieve information such as redshift and cosmology (if available).

By default \textit{XGA} sets up \textit{XSPEC} fits so that all available data are fitted simultaneously, contributing to a single final model from which fit properties are extracted. Individual \textit{XSPEC} models are set up for the available spectra (retrieved from the \textit{XGA} source object during the population of the \textit{XSPEC} script), with parameters ‘linked’ between different models so that they all change during fitting. \textit{XGA} also adds an extra multiplicative constant to each model, so when the \textit{XGA} function that represents $\text{tbabs} \times \text{apec}$ is called it actually sets up the models as $\text{constant} \times \text{tbabs} \times \text{apec}$; this is so that the constant can be left unlinked and free to vary so that each model can take the differing \textit{XMM} instrument effective area for each spectra into account. An optional ‘cleaning’ step is also implemented, where every spectrum is first fit individually and has to pass certain quality checks before it is allowed to contribute to the final, simultaneous, fit. For the $\text{tbabs} \times \text{apec}$ model for instance, the temperature measurements extracted have to be above 0.01 keV, below 20 keV, and the temperature uncertainty less than 15 keV. The quality checks are all configurable when setting up functions to fill in the template script, so new convenience functions for fitting different models can easily be implemented.

When complete, the fit results are saved in files and then read back into \textit{XGA}, which proceeds to match the results files to the \textit{XGA} sources (if multiple sources have had \textit{XSPEC} fits performed) and store the fit parameters in the spectra they were derived from. Each \textit{XGA} source (and sample) class then has convenient methods to quickly retrieve fit results for a given model, within a given analysis aperture. There are also methods to retrieve X-ray luminosities from sources and samples in the same way, as every \textit{XGA} \textit{XSPEC} fit also measures an unabsorbed luminosity. Plotting data for the fitted model is also saved in files and read back in, and can then be used by the \textit{Spectrum} view method to plot the model on the spectral data, as seen in Figure 2.11.

### 2.4.10 AnnularSpectra

The \texttt{AnnularSpectra} class is another example of an aggregate \textit{XGA} product, in this case it stores sets of \textit{XGA Spectrum} objects, that have been generated in concentric annuli around a central point. Additionally, the \texttt{AnnularSpectra} class is designed to hold annular spectra from multiple observations and instruments, making it easy and convenient to access a spectrum from a) a specific annulus, b) a specific observation, and c) a specific instrument. The spectra are generated
Figure 2.11: An XGA visualisation of a spectrum produced and fit by the XGA SAS and XSPEC interfaces respectively. This is an example of a galaxy cluster spectrum, and is the PN data of observation 0693010301. A model with two main components, a plasma emission model and a hydrogen column absorption model, is fit to the data, with the best fit line plotted in blue.
Figure 2.12: An example visualisation of a 3D annular spectrum visualisation produced by an XGA AnnularSpectra instance. The data are from one specific ObsID-Instrument combination, the PN observation of 0652010401. The spectra have been fit with an absorbed plasma emission model, and the fit lines added to the figure.

and fitted with the exact same procedures as Spectrum objects, but with extra flags that tell XGA they should be stored in an AnnularSpectra instance as they belong to a set of spectra. The background spectrum assigned to every annulus is generated outside the final annulus, with the default settings placing it between $1.05R_{\text{outer}}$ and $1.5R_{\text{outer}}$. Each individual annular spectrum has its own ARF, which helps to account for the vignetting effects which decrease the effective area of the telescope towards the edge of the field of view.

Spectral fits are performed on all annuli separately (and thus in parallel), using all available data, though with the possibility of quality checks being used to exclude some spectra if a particular model has that feature set-up (as was described in Section 2.4.9.1). The spectral fit results can be used to setup radial profiles, which are very useful in investigating how the properties of extended X-ray emitting objects vary with spatial position. Currently only the $\text{tbabs}$*apec model has an XGA fitting method for profiles, and if a fit has been successfully completed (i.e. every annulus has a fit result with no failure indicated by XSPEC) then projected temperature and apec normalisation profiles (see Section 2.4.11) are setup automatically and stored in the XGA source object that generated them. If the metallicity in the fit was allowed to vary (by default metallicity is frozen at
a value of 0.3 $Z_{\odot}$) then a metallicity radial profile will also be constructed automatically.

This class has additional methods implemented to retrieve spectra, the radial boundaries of the annuli the set of spectra were generated with, and other properties of the spectra. Also added were methods for visualising the spectra of multiple annuli at the same time, they create a 3D figure where the added third axis represents distance from the central coordinate of the set of spectra. Figure 2.12 demonstrated one such method, where the annular spectra (and their model fits) of a particular ObsID-Instrument combination have been displayed on a single figure.

### 2.4.10.1 Determining what annular boundaries to use

As discussed in Section 2.4.9 and Section 2.4.10, the XGA evselect_spectrum function can be used to generate sets of annular spectra within a given set of annular radii, the inner and outer boundaries of the annuli. Such annular radii can be decided upon manually and passed into XGA’s spectrum generating and fitting functions, but a method for automatic generation of annuli was implemented in order to decrease the need for human interaction while still leaving manual annuli as an option.

A method was designed and implemented based on two considerations; a) that the width of the annular bins should be at least equal to a minimum annulus width provided by the user, and b) that each annular bin should contain at least a minimum number of X-ray counts as specified by the user. The minimum width requirement is because while we can apply PSF correction to images (see Section 2.4.8.1) we cannot apply this same technique to the spectra, instead we take an annulus of a minimum width to try and minimise PSF effects moving photons from one annulus to another. The minimum count requirement is to try and ensure that there are enough photons in a spectrum to perform a useful model fit. For instance, previous XCS work (Lloyd-Davies et al., 2011) found that a minimum of 1000 soft band (0.5-2.0 keV) was required to attain a 10% fractional error for a temperature measurement of 1.5 keV, or 30% for a 8 keV measurement; 2000 soft band counts should attain below 20% fractional temperature error for all temperatures.

This method is implemented as part of the min_cnt_proj_temp_prof function, which is designed to take a GalaxyCluster (or ClusterSample) input and go all the way from deciding which annular boundaries, to generating annular spectra, to fitting them and returning projected temperature profiles. There are many user configurable options; setting the minimum width and counts, which image should be used for the minimum count determination (can either use the combined image from all available data, or the median image in terms of counts within the
annulus region), what energy range of image should be used for minimum count determination, and parameters that are passed through to spectrum generation and fitting. The user is also required to choose an outer radius for the annular spectra to be generated out to; a separate radius can be passed for each source if an XGA sample is passed. The method is iterative, with the following steps:

1. Initial annuli are laid down, taking into account the user-supplied minimum width. As a consequence of this the outermost radius is likely to change slightly from what the user specified, as it must be an integer multiple of the minimum width. The initial inner and outer radii of the annuli are stored in ‘current radii’ variables. A set of masks corresponding to the annuli is generated using the XGA source object, accounting for contaminating sources by multiplying the annular masks by the contaminating source mask that a source can produce. The set of masks comes in the form of an $X \times Y \times N$ array, where $X$ and $Y$ are the dimensions of the image, and $N$ is the number of initial annuli.

2. A check on the current number of annuli is made. If it is less than or equal to four then the annuli are returned to the user as they, and a warning issued that the minimum number of annulus has been reached. The method to calculate background subtracted counts given a source mask and a background mask (discussed in Section 2.4.7.3) is used to measure background subtracted counts for each annulus.

3. The background subtracted counts in each annulus are compared to the minimum counts specified by the user, and any annuli that don’t meet the requirements are flagged. Working from the outermost ‘bad annulus’ inwards, any annulus that doesn’t meet requirements is merged with the annulus to its right by the expedient method of adding their masks together, then calculating and updating the current annular radii variables. If the outermost annulus has below the required number of counts then it is merged to the left.

4. Repeat from step two until the requirements are met, or the minimum number of annuli has been reached.

5. The final annuli are returned to the calling function and used to generate and fit sets of annular spectra. The final annuli are also returned directly to the user, in case they are required for any external analysis.
2.4.11 XGA profile classes

This section summarises the various specialised sub-classes of the BaseProfile1D class discussed in Section 2.4.2, including explanation of any extra methods/functionality implemented for the subclass.

- **SurfaceBrightness1D** - Mostly meant for galaxy clusters, this class will store a 1D surface brightness profile, and enable the fitting of valid models such as a beta profile, or double beta profile.

- **GasDensity1D** - This is meant to store a gas density profile as calculated by XGA, and includes methods to calculate a total gas mass within a given radius, as well as to generate a cumulative gas mass profile.

- **GasMass1D** - A class for storing gas mass 1D profiles, it currently has no extra functionality over BaseProfile1D and is generated by a method of the GasDensity1D class.

- **ProjectedGasTemperature1D** - A class for the projected temperature profiles which are directly measured by fitting plasma emission models to annular spectra. They are ‘projected’ because they are a combination of temperatures of the 3D shells which are intersected along the line of sight by the annulus.

- **APECNormalisation1D** - A class for storing profiles of the normalisation of the APEC plasma emission model, which is extracted from the same fitting process (run on an AnnularSpectra) that produces the projected temperature profiles. This profile can be used to measure the 3D density profile, and when converted to an emission measure profile and combined with knowledge of the projected temperature profile allows us to infer the 3D temperature profile.

- **EmissionMeasure1D** - Calculated from an APECNormalisation1D, and knowledge of the cosmology and the redshift of the source. The emission measure profile can be used to help infer the 3D temperature profile of a galaxy cluster, when combined with the projected temperature profile and assumptions about the source geometry.

- **ProjectedGasMetallicity1D** - Another profile that *can* be measured from the fitting of AnnularSpectra, though only if metallicity is allowed to vary as a free parameter. Again it is ‘projected’ because the metallicities are a combination of the metallicities of the 3D shells intersected along the line of sight by the annuli.
• **GasTemperature3D** - A three-dimensional radial map of the plasma temperature of the intra-cluster medium of a galaxy cluster. This can be used, in combination with knowledge of the 3D gas density, to measure a mass profile for a galaxy cluster.

• **HydrostaticMass** - Defined with a gas density profile and a 3D temperature profile, this type of profile describes the change of the total mass contained within a radius, and has methods to measure a mass at whatever radius the user wants to. This profile class has significant extra functionality, including methods to view the mass distribution measured at a particular radius, the ability to measure overdensity radii using an iterative process, and a function to leverage this classes access to both hydrostatic and gas density profiles to calculate baryon fraction profiles. This class is used extensively in Chapter 3.

• **BaryonFraction** - Can be generated by a HydrostaticMass profile, this shows the change in total baryon fraction within a radius, with radius. Again the value at a specific point can be calculated using a method implemented in this class.

### 2.4.11.1 XGA radial models

Fitting models to radial profiles to attain a parametrised representation of radial properties is a common step in many X-ray analyses, and to make the process as simple as possible for the user (and to provide powerful and convenient features) an XGA BaseModel1D class was implemented. This is not counted as one of XGA’s product classes as it is simply used to configure and store results from fits to profile products, it is not a product in its own right.

The BaseModel1D class has a significant amount of functionality. Its primary function is to act as storage for best fit model parameters and parameter chains (if an MCMC method is used to fit the model, see Section 2.4.11.2), as well as to use those best fit parameters and parameter posterior distributions to provide predictions based on an radius value passed into a class instance. Every part of the model parameter storage uses Astropy (see Section 1.5.3) quantity objects, so that the unit of every parameter is consistent, the parameters can be used in the model equation to retrieve an answer of the expected unit, and the parameters themselves can be easily converted to other (compatible) units with a simple method call. XGA models also store the start values and priors for the fitting process, providing them to the fit functions when required; to make the fitting solution as flexible as possible the parameter start values and priors are user-configurable when an XGA model instance is setup. Some mathematical operations have been built directly into the BaseModel1D class; these include the first derivative (for certain sub-classes whose equation has
an analytical solution this is implemented in an analytical form), n\textsuperscript{th} derivative (using the scipy implementation of numerical differentiation), the volume integral, and inverse Abel transforms using various methods implemented in the PyAbel Python module. Finally, there are visualisation methods built into the BaseModel1D class; the first will display the model curve as it is currently described by the parameter distributions (though without any data, as XGA models do not store any data points), and the second will display histograms of the current posterior distributions of parameters (if they are available).

At no point is the BaseModel1D class directly utilised by the user, as the base class does not have an equation associated with it, nor does it have any physical context (each model class has a label that defines what it describes, to ensure that the user cannot use unphysical models with a particular profile); instead several sub-classes of the BaseModel1D superclass have been setup in XGA, with an equation and default parameter start values/priors associated with them. The following model classes are currently implemented in XGA:

- **Surface Brightness Models:**

  - **BetaProfile1D** - The simplest model of galaxy cluster X-ray surface brightness, a projected isothermal King profile. Described by a simple power law,

    \[
    S_X(R) = S_0 \left( 1 + \left( \frac{R}{R_c} \right)^2 \right)^{-3\beta + 0.5},
    \]

    where \( S_0 \) is the normalisation of the model in units of \( ct \ s^{-1} \ arcmin^{-2} \), \( R_c \) the ‘core radius’ in distance units dependant on the profile radial unit, and \( \beta \) the powerlaw index.

  Using this model makes the implicit assumption that the cluster surface brightness profile is a smooth curve, which is not necessarily valid if the cluster X-ray emission has significant substructure.

  - **DoubleBetaProfile1D** - A simple extension to the Beta profile model that can account for surface brightness profiles that display more structure than a simple curve can account for. Described by the sum of two Beta models,

    \[
    S_X(R) = S_{01} \left( 1 + \left( \frac{R}{R_{c1}} \right)^2 \right)^{-3\beta_1 + 0.5} + S_{02} \left( 1 + \left( \frac{R}{R_{c2}} \right)^2 \right)^{-3\beta_2 + 0.5},
    \]

    where the parameters are equivalent to those of the BetaProfile1D model, but there are two independent sets for the two summed beta models.

- **Gas Density Models:**
- **KingProfile1D** - The simplest model of galaxy cluster gas density, an isothermal-sphere King profile. Described by a simple power law,

\[
\rho_g(R) = \rho_0 \left( 1 + \left( \frac{R}{R_c} \right)^2 \right)^{-3\beta},
\]

where \( \rho_0 \) is the normalisation of the model in units of \( M_\odot \text{Mpc}^{-3} \), \( R_c \) the ‘core radius’ in distance units dependant on the profile radial unit, and \( \beta \) the power law index. This model has a simple analytical solution to the Abel transform (a projection assuming spherical symmetry), the Beta profile (see surface brightness models).

- **DoubleKingProfile1D** - A simple extension to the King profile model that can account for slightly more complex density profile structures. Described by the sum of two King models,

\[
\rho_g(R) = \rho_{01} \left( 1 + \left( \frac{R}{R_{c1}} \right)^2 \right)^{-3\beta_1} + \rho_{02} \left( 1 + \left( \frac{R}{R_{c2}} \right)^2 \right)^{-3\beta_2},
\]

where the parameters are equivalent to those of the KingProfile1D model, but there are two independent sets for the two summed King models. This can be projected to become the double beta model.

- **VikhlininDensity1D** - A commonly used model developed by Vikhlinin et al. (2006) that can represent quite complex galaxy cluster gas density profiles. It was constructed empirically to account for various features that are seen in density profiles and is described by,

\[
\rho_g(R) = \sqrt{\rho_{01}^2 \left( \frac{R}{R_{c1}} \right)^{-\alpha} \left( 1 + \left( \frac{R}{R_{c1}} \right)^2 \right)^{(3\beta_1 - \alpha)/2} \left( 1 + \left( \frac{R}{R_s} \right)^\gamma \epsilon/\gamma \right)} + \rho_{02}^2 \left( 1 + \left( \frac{R}{R_{c2}} \right)^2 \right)^{3\beta_2},
\]

where \( \rho_{01} \) is the normalisation of the first term of the model (a broken power law designed to account for steepening of density profiles near the outskirts), \( R_c \) is the core radius of the first term, \( \alpha \) is a power law index, \( \beta_1 \) is another part of the first power law index, \( R_s \) is a transition radius between two power laws, \( \gamma \) controls the width of the transition region \( \epsilon \) is the change of slope between power laws, \( \rho_2 \) is the normalisation of the second term of the model (which increases modelling freedom near the core), \( R_{c2} \) is the core radius of the second term powerlaw, and \( \beta_2 \) is the index of the second term powerlaw.
- **SimpleVikhlininDensity1D** - A simplified version of the original Vikhlinin density model, with fewer parameters making it easier to fit. The model is represented by,

\[
\rho_g(R) = \rho_0 \sqrt{\frac{R}{R_c}}^{-\alpha} \left(1 + \left(\frac{R}{R_c}\right)^2\right)^{-\beta} \left(1 + \left(\frac{R}{R_s}\right)^3\right)^{\epsilon/3}, \tag{2.10}
\]

where all parameters have the same definitions as in the original model, though the second term and its associated parameters have been removed.

- **Gas Temperature Models:**

  - **VikhlininTemperature1D** - A model for the radial temperature profile of galaxy clusters, created by Vikhlinin et al. (2006) to fit the observed profiles of clusters. The model is described by,

\[
T(R) = T_0 + \frac{T_{\text{min}}}{T_0} \left(\frac{R}{R_{\text{cool}}}\right)^{a_{\text{cool}}} \left(\frac{R}{R_T}\right)^{-\alpha} \left(1 + \left(\frac{R}{R_{\text{cool}}}\right)^{b} \left(1 + \left(\frac{R}{R_T}\right)^{c} \right)^{\frac{1}{2}}\right), \tag{2.11}
\]

where \(T_0\) is the normalisation of the model, \(T_{\text{min}}\) should represent the minimum temperature of the profile, \(R_{\text{cool}}\) is the turnover radius for a possible cool-core region, \(a_{\text{cool}}\) is the slope of a cool core region powerlaw, \(R_T\) is the transition radius for a broken powerlaw modelling the outskirts of the cluster, \(a\), \(b\), and \(c\) are all indices of the broken powerlaw.

  - **SimpleVikhlininTemperature1D** - A simplified version of the original Vikhlinin temperature profile model, with fewer parameters, making it easier to fit. The model is described by,

\[
T(R) = \frac{T_{\text{min}}}{T_0} \left(\frac{R}{R_{\text{cool}}}\right)^{a_{\text{cool}}} \left(1 + \left(\frac{R}{R_{\text{cool}}}\right)^{2 \frac{1}{2}} \left(1 + \left(\frac{R}{R_T}\right)^{2} \right)^{\frac{1}{2}}\right), \tag{2.12}
\]

where each parameter has the same meaning as in the original Vikhlinin temperature model.
2.4.11.2 Fitting models to radial profiles

The XGA class for radial profile models has been introduced (Section 2.4.11.1), and this section will discuss how XGA the radial profile classes (see Section 2.4.2 and Section 2.4.11 fit models to their data. Several fitting methods have been implemented in the BaseModel1D class, a simple non-linear least squares fitting method, an orthogonal distance regression method, and an MCMC method based on the emcee (Foreman-Mackey et al., 2013).

Once a specific model class (Section 2.4.11) has been setup, the user can choose to fit appropriate models to the data (as setting up a profile entails passing the radii and values, as well as their uncertainties). This is done by calling the fit method of the profile class and passing a model that you wish to fit to the data; this can be either the name of the model as a string, or an already setup model object. Setting up the model object beforehand allows for the user to have fine grained control over the parameter start values and prior distributions. If the user passed a model name then the profile will check that it matches a model implemented in XGA, and then that it is physical to fit the model to this particular profile (i.e. you cannot fit a temperature model to a density profile). The physicality check is also performed if the user passes an instantiated model.

When the user calls the fit method they can also choose the type of fitting that they want to use, along with passing any extra arguments that type of fitting might support; the MCMC fitting for instance allows the user to define the number of steps and the number of walkers. Once a model fit is complete, and the parameter best fit values and parameter posterior distributions are stored in the model object, the model in turn is stored within the radial profile object. Each profile can store multiple instances of the same model if they were fit using different methods. Storing the model in the profile allows the user easy access to the model (using the get_model_fit function to retrieve it), allows the view method of the radial profile object to incorporate any model fits into its visualisation, and also allows the user to generate additional visualisations related to the model fit, namely parameter chain and parameter posterior distribution corner plot figures.

Non-Linear Least Squares

This is the simplest fitting method offered by the profile classes, and makes use of the scipy curve_fit function to perform the fit. Non-linear Least Squares (NLSS) is used for non-linear regression, and is suitably generalised that it will support any model function that can be implemented in XGA. It is computationally inexpensive (especially compared to the emcee fitting
process), and will quickly provide an acceptable set of best fit parameters. It makes use of the parameter start values setup in the model instance, as well as the parameter priors (if they are uniform distributions). The NLLS implementation from scipy will take value uncertainties into account, but not uncertainties on the radial position of each value. The `curve_fit` function returns a set of best fit parameter values and a covariance matrix, so the parameter uncertainties are calculated from the variance values and used to sample a posterior distribution (assuming a Gaussian distribution).

**Orthogonal Distance Regression**

A slightly more complex fitting method that does not treat the radii as fixed known quantities, but rather takes into account that they also have uncertainties. These uncertainties stem from the methods used to generate radial profiles, which more often than not involve placing down annuli of some width; any quantity measured within an annulus is inevitably derived from the combination of emission within that annulus, and thus cannot be attributed to a single point with no uncertainty.

`XGA` uses the scipy implementation of Orthogonal Distance Regression (ODR), which in turn is a Python wrapper for the ODRPACK Fortran-77 library. Parameter start values are again extracted from the `XGA` model object, and once fitting is complete the parameter uncertainties are used to generate Gaussian posterior distributions (just as in the Non-Linear Least Squares case). This fitting method has not been used in any publication using `XGA`, but tends to produce better results than the Non-Linear Least Squares method, and it is similarly computationally inexpensive when compared to `emcee`.

**emcee**

Finally, `XGA` can use the `emcee` ensemble MCMC sampler (along with a Gaussian likelihood function) to fit profile models to data. The `emcee` module has been widely used in the astronomy community (Foreman-Mackey et al., 2013). The `emcee` sampler runs multiple interconnected MCMC chains to properly explore the parameter space. The number of separate instances of chains is called the number of walkers, and each walker takes the number of steps specified by the user on the calling of the `fit` function.

The NLLS (Section 2.4.11.2) fitting method is used to help generate parameter start values for the `emcee` fitting method. The specified model is first fit using NLSS using the starting values
taken from the model instance. The orders of magnitude of each best-fit model parameter are then calculated and perturbed by drawing an $N_{\text{walker}} \times N_{\text{par}}$ matrix of random values from a $\mathcal{N}(0,1)$ distribution, multiplying each random value by the order of magnitude of the respective parameter, and then adding this perturbation to the original NLLS best fit value; $N_{\text{walker}}$ is the number of walkers specified by the user, and $N_{\text{par}}$ is the number of parameters in the model. Before the sampler is started, a check is performed to ensure that the start positions all fall within the allowed boundaries of the parameter priors (as defined by the model instance), and if not then they are re-drawn until they do. On the rare occasions that the `curve_fit` run fails, the `emcee` fit method falls back on a maximum likelihood method where it attempts to minimise the negative log likelihood to find best fit parameters, then those parameters are perturbed in the same way to generate start positions for the walkers.

Once the sampler run is complete, integrated autocorrelation times are calculated for the model parameters, if all autocorrelation times are more than 400 times smaller than the number of steps taken by a walker, then the mean autocorrelation time for all parameters is calculated, rounded up to the nearest 100, and doubled to find a good number of steps to remove from each walker’s chains as a burn-in period. If not all parameters had a autocorrelation time that fit this criteria, then a brute force approach is taken and the first 30% of steps in each walker’s chains is removed. Finally, $N_{\text{samp}}$ random points is drawn from the combined parameter chains of all walkers to create the final posterior distributions; where $N_{\text{samp}}$ is the number of samples and can be set by the user on calling the `fit` method. The medians of the parameter posterior distributions are also measured to act as ‘best-fit’ values.

### 2.5 Future development plans for XGA

XGA is a large and fully featured X-ray analysis module, but its development is still ongoing and there are many planned new features that will make it even more capable. This section discusses some of the functionality that is planned to be implemented, but it is also likely that other features will be implemented on request of current and future collaborators.

#### 2.5.1 Support for multi-mission analyses

As XGA was designed for the XMM Cluster Survey, it currently only supports analysis of objects that have been observed by XMM. The work that XGA has been used for, and the ease and speed
with which it can be used to analyse a new sample of objects, has confirmed that its ‘source based’ paradigm is successful and worth developing further. X-ray astronomers have access to large sets of data thanks to several very successful telescopes over the last two decades, with new telescopes either online (eROSITA) or due to launch soon (such as the Japan Aerospace Exploration Agency X-ray telescope XRISM); furthermore there are powerful X-ray telescopes planned to launch within the next decade (e.g. Athena and Lynx. As such the current science archive is large, and its growth will accelerate over the next few years. A downside of this is that science output can become limited by the difficulty of gathering and analysing all data for a source. XGA has solved this problem for XCS, and can do the same for Chandra and eROSITA, as well as other future telescopes.

Not only will this simplify access to data from other X-ray telescopes, it will allow for the creation of powerful joint analyses, where the telescope observations are combined to take full advantage of their individual strengths. The implemented analyses should all generalise to being performed with any individual X-ray telescope, and then new joint methods can be devised.

### 2.5.2 Simple analysis of simulated X-ray data

Once XGA has been generalised to support data from other X-ray telescopes (as discussed in Section 2.5.1), it will be possible to add greater support for simulated observations from current and future telescopes. XGA contains many powerful analyses that have been used in observational data analyses, which makes it an ideal tool for taking measurements from simulated observations for purposes such as comparing simulation and observed properties. In some cases, such as XGA’s XCSim (Turner et al., prepb) tool, it is as simple as pointing XGA to a different set of event lists, as XCSim produces data products that are identical to real XMM observations. In other cases (such as for telescopes that haven’t launched yet), interfaces with simulated observation codes will have to be developed so as to convert the output data into a format that XGA can use.

### 2.5.3 Measurement of 2D property maps and development of 2D models

Currently, the only measure of spatially varying properties measured from spectra that is supported by XGA are radial profiles of XSPEC derived parameters. It is also possible to measure 2D maps of projected spectral quantities, and it has been used in several instances in the literature (Schmidt et al., 2002; Sanders, 2006; Alden et al., 2019) to explore things like the temperature structure of a galaxy cluster. Setting up and measuring the maps is very complex, requiring the generation
and fitting of hundreds of spectra. Implementing that functionality into XGA will make it far more convenient, and designing a PropertyMap class that will provide the same sort of ease of use as XGA radial profiles will allow for many more analyses to use property maps. To complement this new functionality, new two-dimensional models of temperature and density will be developed and integrated into XGA, so spatially varying properties can be parametrised as they are for radial profiles.

2.5.4 Support for time-domain analyses

Some X-ray sources (such as AGN, Quasars, X-ray binaries) are known to be variable, with emission that changes with time. Some X-ray emission, such as that resulting from Supernovae and Neutron Star-Neutron Star collisions, is entirely transient, and can be used to investigate various properties of the emitting objects or events. As such XGA will be expanded to support time-domain analyses, including the creation and fitting of light curves, which describe how the emission of an object changes with time (collaboration with X-ray variable source experts will be sought for these features). As XGA already fetches all available XMM data for a given source, it will be easy to use each observations temporal information to create cross-observation light curves, again maximising the use of the XMM science archive. Additionally, once the work detailed in Section 2.5.1 is complete, this feature will become even more powerful as other telescope archives will be able to be used to help fill in the gaps. Finally a new X-ray product class to store and interact with these light-curves will be implemented, to make those analyses as easy as the currently implemented analyses are.

2.5.5 Deployment on the ESA DataLabs platform

In 2021 a talk about XGA was given to members of the European Space Agency (ESA) science teams, and the XMM mission director expressed an interest in XGA being made a part of the a new ESA platform that will enable easy access to and analysis of the archives of all ESA missions. An online interface will allow the user to select the exact calibration and software stack that they wish to use for their analysis, and then use Jupyter Notebooks to run code and save their results. This represents a near-perfect use case for XGA, as there will be access to a cleaned set of XMM data with region files, and the user will not have to perform any of the XGA setup (which is more involved than a typical Python module due to the need to point XGA at your processed data). This work will ideally begin when the DataLabs platform is in a closed-beta stage, and will mostly
involve setting up and testing on the DataLabs platform; however it may also entail replacing some of XGA’s SAS interfaces with new Python interfaces that are currently being created by the XMM team.
Chapter 3: Tests of XGA measurements and new masses for the SDSSRM-XCS, DESY3RM-XCS, and ACTDR5-XCS samples

This chapter contains work related to the measurement of hydrostatic masses of samples of galaxy clusters. The sections prior to Section 3.7 make up a paper that is due to be submitted to MNRAS, which lays out the XGA approach to measuring hydrostatic masses and other cluster properties, and makes comparisons with published results from XCS, XXL, and LoCuSS. As the paper has been included in this chapter in the form it will be submitted to the journal in, there may be some overlap between the content of this and other chapters. The final sections relate to the construction of an ACTDR5 selected XMM sample, and the measurement of masses for the ACTDR5 and DESY3RM selected samples.

3.1 Introduction

Galaxy clusters are the most massive gravitationally bound structures in the Universe. They formed through the collapse of the primordial density field, and as such are a useful way to investigate the evolution of the Universe through the measurement of cosmological parameters. The mass of a galaxy cluster is split into three main components; the dark matter halo (87%), the intra-cluster medium (7%), and the component galaxies (3%) (Gonzalez et al., 2007); where the intra-cluster medium (ICM) is a high-temperature, low density plasma largely made up of ionised hydrogen. Just as the formation of clusters makes them useful for investigating cosmology, the nature of the ICM makes them ideal astrophysical laboratories.

Cosmological parameters can be derived from measurements of the galaxy cluster mass func-
Figure 3.1: Histograms of properties of samples used in this work, taken from literature. X-ray properties presented in this figure have been measured in different regions, indicated in the legend of each plot. a) Intra-cluster medium average temperatures, b) cluster redshifts from spectroscopic follow-up (XXL-100-GC), from the RASS catalogues the sample was selected from (LoCuSS High-$L_X$), and from the redMaPPer cluster finder (SDSSRM-XCS), c) bolometric X-ray luminosities taken from literature, measured within $R_{500}$ (LoCuSS High-$L_X$ and SDSSRM-XCS) and 300 kpc (XXL-100-GC).

The hydrostatic equilibrium mass method uses X-ray data to infer the total mass of the cluster from the temperature and gas density profiles. The XMM telescope’s field of view (FoV), effective area, and large public archive of observations makes it the best available X-ray telescope for these measurements. Hydrostatic masses can (and have been) measured using other instruments (Vikhlinin et al., 2006; Sun et al., 2009; Giles et al., 2017; Sanders et al., 2021a), but have disadvantages compared to XMM. Chandra’s smaller FoV and lower sensitivity make it a less effective instrument for probing hydrostatic masses, and though eROSITA has a very large FoV and similar sensitivity to XMM, its current use as a survey telescope precludes the possibility of gathering enough photons
Table 1: Summary of galaxy cluster samples used in this work.

† The full XXL-100-GC sample contains 100 galaxy clusters, but we remove XLSSC504 due to known issues with its XXL-measured X-ray properties.

‡ The full LoCuSS High-$L_X$ sample contains 50 clusters, 32 have published $XMM$ properties. We analyse clusters that have had $XMM$ observations since then.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>$N_{CL}$</th>
<th>Brief description</th>
<th>Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDSSRM-XCS</td>
<td>150</td>
<td>Volume-limited sample ($0.1&lt;z&lt;0.35$) selected from SDSS redMaPPer catalogue, with $\Delta T_X&lt;25%$</td>
<td>Ensuring $XGA : L_X$ and $T_X$ measurements are consistent with existing XCS results. New measurements of masses of this sample are presented.</td>
</tr>
<tr>
<td>XXL-100-GC</td>
<td>99†</td>
<td>The 100 brightest clusters in the XXL survey regions</td>
<td>Testing whether $XGA : L_X$, $T_X$, and $M_{gas}$ measurements are consistent with external results.</td>
</tr>
<tr>
<td>LoCuSS High-$L_X$</td>
<td>45‡</td>
<td>A sample of high-$L_X$ clusters selected from the RASS and followed up by $XMM$ or $Chandra$</td>
<td>Testing whether $XGA : L_X$ and $T_X$ (global and core-excised) and $M_{gas}$ and $M_{hy}$ measurements are consistent with external results.</td>
</tr>
</tbody>
</table>

to determine temperature profiles. As such, while $eROSITA$ is predicted to detect $\sim 100000$ (Pillepich et al., 2012) galaxy clusters over the full eRASS survey, older X-ray telescopes (especially $XMM$) have an important part to play in the future of X-ray cluster mass measurements. Enough X-ray data exists in the telescope archives that we will be able to measure hundreds of masses at once, but a new, self-consistent, method is required.

will focus on describing and verifying our methodology, and then presenting a sample of masses measured for the SDSSRM-XCS sample, as a companion to Giles et al. (2022b).

The outline of the paper is as follows, §3.2 introduces the XCS cleaned data products used for this work, as well as the three galaxy cluster samples. §3.3 describes $XGA$, a new piece of software created by XCS, as well as demonstrating that its measurements of global properties are consistent
with literature measurements and our previous work. §3.4 demonstrates features of XGA that allow the generation and fitting of radial profiles of gas density, temperature and mass. We then introduce tests by comparing XGA gas and hydrostatic mass measurements to literature values. In §3.5 we present new measurements of hydrostatic mass for a subset of the SDSSRM-XCS sample, including scaling relations with richness, temperature, and luminosity. Finally, in §3.6 we present our conclusions and a discussion of the next steps of this work.

The analysis code, samples, and outputs are available in a GitHub repository¹. As we present comparisons to several sets of measurements from literature, we adopt the cosmology used by each of the original analyses; a concordance ΛCDM cosmology where $\Omega_M=0.3$, $\Omega_\Lambda=0.7$, and $H_0=70$ km s$^{-1}$ Mpc$^{-1}$ for the XCS and LoCuSS samples, and the WMAP9 cosmology where $\Omega_M=0.282$, $\Omega_\Lambda=0.719$, and $H_0=69.7$ km s$^{-1}$ Mpc$^{-1}$ (Hinshaw et al., 2013) for the XXL sample.

### 3.2 XCS Data and Cluster Samples

#### 3.2.1 XCS cleaned data and source regions

All XCS X-ray measurements presented in this work are based on cleaned event lists produced during the XCS analysis of the XMM-Newton public archive, which is fully described in Lloyd-Davies et al. (2011). The initial EPIC data were processed with v14.0.0 of the XMM Science Analysis Software (SAS), using EMCHAIN and EPCHAIN to generate event lists. Following this, the event lists were screened for periods of high background and particle noise by generating light curves in 50s time bins for the soft (0.1–1.0 keV) and hard (12–15 keV) bands. An iterative 3$\sigma$ clipping process was then performed on the light curves; time bins falling outside this range were excluded. We do not currently make use of XMM Reflection Grating Spectrometer (RGS) or Optical Monitor (OM) data.

Source detection is performed using a custom version of WAVDETECT (Freeman et al., 2002) called XCS Automated Pipeline Algorithm (XAPA). XAPA is run on merged EPIC (PN+MOS1+MOS2) images, with a pixel size of 4.35\textquoteright, that have been generated using events within an energy range of 0.5-2.0 keV. Once sources in an image have been located, XAPA then classifies them as either point or extended sources. Region lists are created for each individual XMM observation. Note that although the cleaned events lists were produced with SAS v14.0.0,

¹Analysis code and samples
as were the images upon which XAPA is run, the analyses described in this paper are based on XMM products generated using SAS v18.0.0.

### 3.2.2 SDSSRM-XCS$_{T_X,\text{vol}}$

The first sample used in this work is a set of optically selected galaxy clusters with available XMM data, recently described in Giles et al. (2022b). The sample was constructed by selecting from the red-sequence Matched-filter Probabilistic Percolation (redMaPPer) Sloan Digital Sky Survey (SDSS) catalogue (Rykoff et al., 2014). The final sample is made up of clusters withing $0.1 < z < 0.35$, and a fractional X-ray temperature error of less than 25%. This results in a sample of 150 galaxy clusters. The sample we use in this work is referred to as SDSSRM-XCS$_{T_X,\text{vol}}$ in the original Giles et al. (2022b) analysis, but we shall refer to it as the SDSSRM-XCS sample as it is the only sample from the original work that we use.

### 3.2.3 XXL-100-GC

The second set of clusters we utilise is the XXL-100-GC sample (Pacaud et al., 2016), a sub-sample of the galaxy clusters detected by the XXL survey (Pierre et al., 2016), X-ray selected by setting a lower limit of $3 \times 10^{-14}$ erg s$^{-1}$ cm$^{-2}$ on the flux measured within a $1'$ aperture. The XXL survey is a large-scale XMM survey that observed two contiguous regions of the sky, totalling $\sim$50 square degrees, made up of 542 XMM observations. This sample of 100 clusters has been used in several XXL analyses, including their work on mass temperature relations (Lieu et al., 2016), and measurements of gas mass (Eckert et al., 2016).

### 3.2.4 LoCuSS High-$L_X$

The final sample we use is Local Cluster Substructure Survey (LoCuSS) High-$L_X$, 50 X-ray selected galaxy clusters from the ROSAT all sky survey (RASS) catalogues; ICM temperatures ($T_X$), ICM luminosities ($L_X$), total gas masses ($M_{\text{gas}}$), and total hydrostatic masses ($M_{\text{hy}}$) have been measured using a mix of XMM-Newton and Chandra observations (Martino et al., 2014). At the time of the original LoCuSS hydrostatic mass work, 43 of the clusters had been observed by Chandra and 39 by XMM, with 27 of the clusters having data from both telescopes. This set of galaxy clusters does not extend to the low temperatures explored by the other samples we intend to analyse, as can be seen in Figure 3.1a. As they were specifically selected for their high X-ray
luminosities this is to be expected, but means that our comparison of masses does not extend to the low-mass regime.

### 3.2.5 Comparison of samples

Key information about these samples is summarised in Table 1, and distributions of three key values are presented in Figure 3.1. Figure 3.1a shows the temperature distributions; XXL-100-GC mostly contains clusters below 6 keV, SDSSRM-XCS peaks at a similar temperature to XXL-100-GC but extends to higher temperatures due to selection from optical observations, and the LoCuSS High-$L_X$ sample contains higher temperature clusters (similar to the tail of the SDSSRM-XCS distribution) to its selection from highly luminous systems. Figure 3.1b shows the redshift distribution. The XXL-100-GC sample contains higher redshift clusters than the other samples, however, the SDSSRM-XCS and LoCuSS High-$L_X$ samples includes redshift cuts, introducing artificial limits. Finally, Figure 3.1c shows luminosity distributions; XXL-100-GC is similar to the lower end of the SDSSRM-XCS distribution, whereas the LoCuSS High-$L_X$ sample is similar to the high luminosity tail of SDSSRM-XCS.

### 3.3 The X-ray: Generate and Analyse (XGA) software package

In this section we provide a brief overview of the basics of X-ray Generate and Analysis (XGA; Turner et al., 2022), including source matching, generation of images, generation of masks to remove contaminating sources, and the measurement of galaxy cluster temperatures and luminosities. Using the samples introduced in Section 3.1, we show that XGA produces $T_X$ and $L_X$ values consistent with previous measurements, and highlight that XGA can measure values using different cosmologies and energy bands. All analyses performed in this work used XGA v0.4.1, SAS v18.0.0, XSPEC v12.11.

#### 3.3.1 Summary of XGA

X-ray: Generate and Analyse (XGA)\(^2\), is a new, open-source, Python module developed by XCS, and recently described in Turner et al. (2022). It is a generalised X-ray analysis tool, capable of in-

\(^2\)X-ray: Generate and Analyse GitHub
\(^3\)X-ray: Generate and Analyse Documentation
vestigating any X-ray source that has been observed by XMM, with different analyses implemented
for specific types of astrophysical source.

XGA is based around the declaration of source objects, which are software proxies for actual astrophysical objects; they have physical properties associated with them, such as position, redshift, nH, luminosity, and more, dependent on the type of object they are. There are general source types for generic point and extended sources, and more specific classes for sources like galaxy clusters or stars. The type of source declared will affect what analyses can be performed, for instance attempting to measure the hydrostatic mass of an AGN source would not be allowed.

Almost every aspect of the generation and analysis routines implemented in XGA are configurable by the user. This ranges from being able to define the cosmology that is used for analysis, to deciding how spectra should be binned on generation, to setting the energy range that a spectral fit should be performed in. This allows us to recreate many of the analysis choices employed by previous work, simply by changing arguments passed to XGA functions.

### 3.3.2 Selecting XMM data

Once a galaxy cluster source has been declared, valid XMM observations with an aimpoint within 30′ of the cluster coordinates are located; an observation is discarded if less than 70% of the initial $R_{500}$ region falls on the detectors. If an observation aimpoint is within 5′ of the cluster coordinates, it is considered ‘on-axis’. The SDSSRM-XCS sample contains 99 (66%) objects with at least one on-axis observation, the XXL-100-GC sample contains 33 (34%; lower because of the contiguous survey nature of XXL), and LoCuSS High-$L_X$ 43 (96%). Our final analysis of the three samples makes use of ~400 unique XMM observations, not including observations that were removed due to data quality issues. All observations were inspected for flaring that could not be removed through cleaning, as well as for other data problems; we found and rejected 47 such observations.

During the source setup process an nH value for the cluster coordinates is found from the full-sky HI survey by the HI4PI Collaboration et al. (2016) using the HEASoft nh command, the chosen cosmology is used to make sure analysis radii (such as $R_{500}$) are available in kpc and degrees, then image World Coordinate System (WCS) information is used to make radii available in pixel units.
Figure 3.2: An *XMM* 0.5-2.0 keV stacked ratemap of a galaxy cluster in the SDSSRM-XCS sample, generated by XGA. A mask to remove contaminating sources has been applied. The solid line cross-hair indicates the X-ray peak measured by XGA, the dashed line cross-hair indicates the original input coordinate. Solid circle indicates initial $R_{500}$ from Giles et al. (2022b), dashed annulus indicates background region.
3.3.3 Generating and interacting with photometric data

We often need to examine or perform analyses on X-ray images or ratemaps (where the image is divided by the exposure map of the observation to produce a map of count-rate). As such, XGA has been given interfaces to the SAS evselect (used to generate images) and eexpmap (used to generate exposure maps) tools. Once a source (or a sample of sources) has been defined, and the relevant XMM observations have been located, images, exposure maps, and ratemaps in any energy range can be generated in bulk (and in parallel). By default, XGA uses the 0.5-2.0 keV energy band. Photometric products are loaded into instances of specific XGA product classes; Image, ExpMap, and RateMap. XGA’s photometric classes provide convenience functions for viewing (e.g., Figure 3.2), conversion of coordinates using WCS and extracting values at specific coordinates. In addition, XGA also supports creating ‘combined’ images, exposure maps, and ratemaps, using the SAS tool emosaic.

3.3.4 Identifying and removing contaminating sources

To identify which sources are contaminants and need to be removed from an observation, we must determine which source corresponds to the cluster of interest. During the setup stage of a galaxy cluster object, XGA searches the XCS region files for each observation to find source regions that match the cluster coordinates.

If the input cluster coordinates fall within an extended source region, that is considered a match. A cluster is considered detected if there is a match in at least one observation’s region list. We then search for point regions within 0.15R$_{500}$ of the cluster coordinates, as they could be an indication of a bright cool core, and make sure that they are not removed as contaminating sources. We also check for intersections between matched regions and other extended regions (from all observations) within R$_{500}$. This helps to avoid cases where a cluster is detected in multiple observations, but the supplied coordinates do not match to regions in all of them, as XAPA has placed a region on a different part of the cluster. Finally, all other regions (from all observations) are designated as contaminants, and will be removed by masks applied to images, as well as excluded in the event selection criteria used when generating spectra. Figure 3.2 shows a stacked count-rate map generated by XGA, with contaminating sources masked, and XAPA regions overlaid on top.
3.3.5 Manual adjustment of detection regions and inspection of data

In some instances we need to make manual adjustments to the source regions produced by XAPA, most commonly to increase the size of a region so that contaminating emission is better removed. Poorly removed contaminating point source emission can have a significant effect on measurements, particularly when the source is in the background region. As background emission is subtracted from the surface brightness profiles generated by XGA, and from the spectra during XSPEC fitting, extra emission can remove bias the observed emission level of the cluster significantly.

We address this by visually inspecting combined ratemaps of our galaxy cluster samples using a method in XGA (edit_regions) that allows us to easily interact with and modify existing regions, add new regions, adjust the scaling of the image and apply smoothing to check for sources that have been missed by XAPA. This is done in a graphical user interface, so the result of a new or modified region can be seen immediately.

We overlay analysis regions on the clusters (centered on their X-ray coordinates); a circle indicating where the $R_{500}$ region lies, and an outer annulus indicating where its corresponding background region lies. Some point source regions were adjusted so that more of their emission is removed from analysis. In extreme cases, bright point sources introduce artefacts in the image from XMM’s radial PSF spokes. Extended elliptical regions were added to remove this emission, with Figure 3.3 showing an example of this. Additionally, nearby cluster’s (e.g., mergers or foreground/background clusters) to the one in question had existing extended regions modified to remove more of the contaminating cluster’s X-ray emission.

We modified region files for 60 (40%) of the SDSSRM-XCS sample, 44 (44%) of the XXL-100-GC sample, and 26 (56%) of the LoCuSS High-$L_X$ sample; every cluster was inspected. The process took approximately three hours for all clusters.

3.3.6 Generating and fitting spectra

We use XGA to measure global properties ($T_X$ and $L_X$) of the galaxy clusters in the samples introduced in Section 3.1. Spectra are extracted within analysis regions specific to each sample, as the original works measured properties within different apertures; apertures are defined in the relevant subsection (Sections 3.3.7.1, 3.3.7.2, 3.3.7.3). Background spectra are extracted in annuli around the outside of the analysis region, with the inner and outer radii of a background spectrum dependant on the analysis aperture. We also define background regions in specific analysis subsections.
Figure 3.3: An example of an XGA region editor window, taken from an animated example of the manual adjustment process for SDSSRM-103. A manual region (in white) is being added to account for a bright poorly removed point source which is exhibiting strong PSF spoking.
For example, the solid white circle centered on the cross-hairs in Figure 3.2 shows the $R_{500}$ measured by Giles et al. (2022b), and the dashed white circles indicate the corresponding background annulus. All spectra are centered on their literature X-ray coordinates. Spectra for each camera of each observation of a given galaxy cluster are generated. XGA uses the SAS `evselect` tool to create the initial spectra, and then `specgroup` to re-bin each spectrum so that there are at least five counts per channel.

Once XGA has generated a set of spectra corresponding to the XMM observations (and each observation’s instruments) of a particular galaxy cluster, we use the XGA XSPEC (Arnaud, 1996) interface to fit an emission model, allowing us to derive cluster properties such as temperature and luminosity. We fit absorbed (with `tbabs`, Wilms et al., 2000) plasma emission models (APEC, Smith et al., 2001) to the spectra.

There are multiple stages to the fitting process. Initially, we reject any spectra that have fewer than 10 noticed channels. Next, each spectrum is fitted individually following the procedure above. If the estimated temperature for a spectrum is $0.01 \text{ keV} \leq T_X < 20 \text{ keV}$, or if the temperature uncertainty is greater than 15 keV, that spectrum is rejected from further analysis. Finally, we simultaneously fit the remaining spectra using a constant*`tbabs`*apec model. The $T_X$ and normalisation parameters of apec are tied across all models, the redshift is frozen at the input value, metallicity is frozen at $0.3 \ Z_\odot$, nH is frozen at the value determined by XGA, and the constant factor for each spectrum is allowed to vary to account for differences in normalisation due to effective area. Temperatures and unabsorbed luminosities are determined from this simultaneous fit. For all fits, $T_X$ is initially set to 3 keV, the normalisation is initially set to $1 \cm^{-5}$, and the constant is initially set to 1. All fits use the angr abundance table (Anders and Grevesse, 1989).

The fits are performed using the $c$-statistic (Cash, 1979). Each XSPEC operation (i.e. a fit to global spectra for a cluster) will time out if it has not been successful within 4 hours.

### 3.3.7 Comparing XGA $T_X$ and $L_X$ measurements to literature

Here we directly compare XGA measurements of galaxy cluster temperature and luminosity to measurements from literature. We attempt to minimise the differences in analysis between XGA and each literature dataset; we use the same energy bands during XSPEC spectral fitting as the original analyses, equivalent analysis regions (e.g. 300 kpc aperture for XXL), and the same cosmologies.
Figure 3.4: Comparisons between results measured using the existing XCS pipeline, and those measured using the new XGA package. a) The XGA APEC temperatures versus the XCS3P APEC temperatures within $R_{500}$ for the SDSSRM-XCS sample, a clear 1:1 relation with a small amount of scatter is evident, as would be expected in this case. b) Shows the equivalent comparison for the 0.5-2.0 keV X-ray luminosity measured within $R_{500}$, with a similarly tight relation between the two measurements. c) Shows the comparison for the bolometric X-ray luminosity measured within $R_{500}$, and is very similar to the soft-band comparison.

### 3.3.7.1 SDSSRM-XCS properties measured by XCS3P

We first compare XGA SDSSRM-XCS properties to the previous XCS analysis of Giles et al. (2022b), who measured temperatures and luminosities (0.5-2.0 keV and bolometric) within $R_{500}$ and $R_{2500}$ regions. The notebook we constructed to measure these quantities is accessible here, and fully explains all analysis steps.

Published measurements were made using the XCS3P pipeline (fully described in Lloyd-Davies et al., 2011; Giles et al., 2022b) that measures properties within the overdensity radii, found through an iterative process. An initial temperature and luminosity are measured within the XAPA defined source radius, then uses the temperature to estimate $R_{500}$ (or $R_{2500}$) using the $R_\Delta - T_X$ relation of Arnaud et al. (2005), and repeats until the radius converges. To account for the background, local background annuli were constructed between 1.05-1.5 $R_{500}$ or between 2-3 $R_{2500}$. XGA does not re-use any code from XCS3P, but the XGA analysis matches as close as possible that performed in Giles et al. (2022b) (e.g., the same source and background regions).

We find strong agreement between SDSSRM-XCS $T_X$, $L_X^{0.5-2.0}$, and $L_X^{bol}$ values measured by XGA and the XCS3P pipeline. Figure 3.4 demonstrates a 1:1 relationship for all measured properties within $R_{500}$. Figure 3.5 provides the same comparison, but for measurements made within $R_{2500}$. Again, there is a very tight agreement between the two sets of measurements.
Figure 3.5: Comparisons between results measured using the existing XCS pipeline, and those measured using the new XGA package. a) The XGA APEC temperatures versus the XCS3P APEC temperatures within \( R_{2500} \) for the SDSSRM-XCS sample; they agree very well with less scatter than the \( R_{500} \) comparison. b) Shows the equivalent comparison for the 0.5-2.0 keV X-ray luminosity measured within \( R_{2500} \), with a similarly tight relation between the two measurements. c) Shows the comparison for the bolometric X-ray luminosity measured within \( R_{2500} \) which is again very similar to the soft-band.

Several outliers can be observed in the temperature and luminosity comparisons, though the majority of measurements agree very well with past XCS work. These outliers can be attributed to a combination of two factors; the exclusion regions used by each analysis, and the use of XMM sub-exposures by the original XCS analysis. Some XMM observations contain multiple sub-exposures by the same instruments, either due to the request of the proposer, or to maximise the usage of the telescope. The analyses performed by XGA only make use of the longest individual sub-exposure for a particular instrument of a particular observation, whereas previous XCS work make use of all of the sub-exposures. The differences in the regions used for excluding contaminating sources are minor, but some individual clusters require manual intervention to modify the regions to better represent the emission in an image. This modification was performed seperately for the original XCS analysis and this XGA reanalysis, contributing to the outliers observed in Figure 3.4. XGA also contains a source region matching step that will ensure any point regions within 0.15 \( R_{500} \) of the cluster centre are not treated as contaminants, to avoid accidentally removing cool cores detected as a separate point source; this step is not present in the original XCS analysis.

It is interesting to note that, for three of the four high-XGA-luminosity outliers in Figure 3.4b, the XGA results reduce the scatter of luminosity with temperature compared to the original XCS results. This may indicate that these re-analysis measurements are more consistent with the actual physical properties of the clusters. The fourth high-luminosity outlier does not exhibit the same
Figure 3.6: A comparison of global X-ray values measured by XGA and XXL analyses, for the XXL-100-GC sample. a) Shows the temperature values measured within the 300kpc, and b) shows the equivalent for the 0.5-2.0keV luminosity measurements.

dramatic improvement in consistency with the luminosity-temperature relation, as it was already a low scatter point. This fourth cluster in particular is right on the edge of the field-of-view, making reliable measurements of properties more challenging. All four of these luminosity outliers correspond to temperature outliers in Figure 3.4a, as their constrained temperatures are also different to the original XCS measurements.

This demonstrates that XGA produces results that are consistent with previous XCS work, and the small amount of scatter can be attributed to differences in software versions, rounding, and how XGA removes contaminating sources.

3.3.7.2 XXL-100-GC

Next, we compare XGA luminosity and temperature measurements to the XXL-100-GC\(^4\) (Pacaud et al., 2016) sample. The notebook we constructed to measure these quantities is accessible here, and fully explains all analysis steps.

We attempt to minimise the difference between our XGA analysis and the XXL-100-GC analysis (as presented in Giles et al., 2016), but some variables are outside our control. The measurements of Giles et al. (2016) are based on data products that have gone through a different cleaning process\(^4\) XXL-100-GC baryonic properties VizieR table
to XCS data (Pierre et al., 2016). Not only that, a different source finder was used to identify and remove contaminating sources, and we do not replicate any potential manual exclusion of sources. Our measurements of the XXL-100-GC cluster sample $T_X$ and $L_X$ are centred on the coordinates reported by Pacaud et al. (2016) (derived from X-ray centroids), use the WMAP9 cosmological results, and use a 300 kpc aperture to be consistent with Giles et al. (2016). Attempting to make our analysis as consistent as possible with Giles et al. (2016), we:

- Perform XSPEC fits within an energy range of 0.4-7.0 keV.
- Use the same cosmology as the original analysis.
- Use a fixed metallicity of 0.3 $Z_{\odot}$ and the Anders and Grevesse (1989) abundance tables.
- Fit an absorbed APEC model to the spectra.

However, some differences between the analyses remain:

- Original work by Giles et al. (2016) used Kalberla et al. (2005) nH measurements, XGA uses HI4PI Collaboration et al. (2016) values; they have been shown to be consistent.

- Further XMM observations of XXL clusters may have been taken since the original analysis.
- XXL will have different cleaned event lists and region files.
- The background measurement technique used in the XXL analysis is significantly different to the method currently used by XGA. A background annulus centered on the aimpoint of the XMM observation is used, with a width determined by the diameter of the analysis region of the galaxy cluster (see Figure 1 of Giles et al., 2016). We use a background annulus local background with width 1000-1500 kpc centered on the galaxy cluster.

From these spectral fits we measure $T_{X,300} \text{ kpc}$ and unabsorbed $L_{X,300}^{0.5-2.0 \text{ keV}}$. In line with Giles et al. (2016) we do not include XLSSC-504 in our analysis due its unconstrained temperature errors. Two (XLSSC-61 and XLSSC-527) of the XXL-100-GC clusters do not pass the XGA cleaning step where at least 70% of the $R_{500}$ region must be present on an observation, and as such we do not measure a temperature or luminosity for those clusters. Likewise we do not measure a temperature or luminosity for XLSSC-11, as due to its very low redshift there are difficulties generating the required spectral products.

$^5$Comparisons of nH from different analyses
The comparison in Figure 3.6 demonstrates good agreement between our XGA based analysis, and the original XXL measurements. The temperature measurements are generally consistent with a 1:1 relation, though many have large uncertainty values. As the cluster observations in the XXL dataset are low signal to noise, the difference in background techniques could have a more substantial effect on the measured temperatures, increasing differences in some cases. The low signal to noise of these observations is likely to have amplified the effect of any differing analysis decisions. The XGA spectrum fitting technique also includes extra cleaning steps that are not included in XXL analyses, where the individual spectra are fit and checked for the quality of their results, to ensure that only spectra of sufficient quality are included in the final, simultaneous, fit. That could remove some data from analyses which were included by XXL. The luminosity measurements of our analysis are in agreement with the XXL work, though there appears to be a slight divergence at higher luminosities. One plausible explanation is the difference in our background methods. Since we are using a fixed background annulus surrounding the cluster there will be more cluster emission present in the background region for larger clusters, hence, more cluster emission will be removed for those systems. This explains the slight slope that is evident in the comparison, as larger (i.e. high luminosity) clusters would be more strongly affected. Luminosity depends mainly on the normalisation of the spectral fit models, and as such exhibits a stronger dependence on choice of background. This reanalysis of the XXL-100-GC cluster sample, and its good agreement with the original results published by Giles et al. (2016), shows that XGA is a reliable tool that produces measurements of global cluster quantities consistent with existing results.

3.3.7.3 LoCuSS High-$L_X$

Finally, we perform a reanalysis of the LoCuSS High-$L_X$ and again directly compare XGA measured $T_X$ and $L_X$ values to those from the original analysis. The notebook we constructed to measure these quantities is accessible here, and fully explains all analysis steps.

The temperatures which we compare to for these clusters are core-excised (measured within 0.15-$R_{500}$), though the luminosities are not. The $R_{500}$ values we use are those presented in Martino et al. (2014). Comparing to core-excised temperatures allows us to demonstrate and test XGA’s ability to measure core-excised galaxy cluster properties. We attempt to minimise differences between analyses, but the same caveats with respect to event cleaning and source regions apply as in Section 3.3.7.2. The cluster spectra are centered on RASS positions reported in Martino et al. (2014).
Figure 3.7: A comparison of global, core-excised, X-ray temperatures and luminosities measured by XGA and LoCuSS analyses, for a subset of the LoCuSS High-$L_X$ sample. a) Shows the temperature values measured within the 0.15-1$R_{500}$ region, and b) shows the comparison for full $R_{500}$ bolometric luminosity measurements.

We attempt to match LoCuSS analysis choices by adjusting the settings with which we run XGA:

- We perform XSPEC fits within an energy range of 0.7-10.0 keV.
- The same cosmology as the original analysis is used.
- Metallicity is left free to vary during XSPEC fits.

Some differences between our analysis and the original remain:

- Original work by Martino et al. (2014) used Dickey and Lockman (1990) nH measurements, XGA uses HI4PI Collaboration et al. (2016) values; they have been shown to be consistent$^5$.

- The original analysis left nH free in XSPEC fits, we do not.

- Further XMM observations of LoCuSS clusters may have been taken since the original analysis.

- LoCuSS will have different cleaned event lists and region files.

- Martino et al. (2014) use older models (wabs and mekal). Though we have found that temperature and luminosity results are consistent with tbabs*apec.

At the time of the original analysis, 39 of the LoCuSS High-$L_X$ clusters had an XMM observation.
During our analysis we found that 45 of the clusters now have XMM observations, though we cannot use the new observations to provide direct comparisons to literature.

Figures 3.7 a) and b) highlight the comparison in temperatures and luminosities respectively, between XGA and LoCuSS. The comparison of core-excised temperatures demonstrates a good 1:1 agreement. The scatter is attributed to the differences in analysis techniques, as well as the cleaning of the data, and the different analysis/exclusion regions used. Some clusters in the LoCuSS High-$L_X$ sample have also had follow-up observations since the initial work by Martino et al. (2014), which could have an effect on our results. The luminosity comparison has a similar level of scatter, though we observe a systematic offset between original and XGA-measured values, with XGA measurements slightly lower than original LoCuSS measurements. The LoCuSS background treatment for spectral analysis is more sophisticated than the current XGA technique of background spectrum generation and subtraction; LoCuSS spectral analysis models the known components of the background, likely providing a slightly different level of background subtraction. As the luminosity measurement stems primarily from the normalisation of the fitted spectral model, small changes to the level of background attributed to the source can significantly change the measured luminosity. Recent comparisons of cluster luminosity measurements by Giles et al. (2022b) have shown that differences in luminosity results between samples are common. We are also unsure of the exact origin of the luminosity measurement outer radius, as the original publication (Mulroy et al., 2019) indicates that the measurements were performed prior to the paper, but doesn’t specify the source of the overdensity radii; some LoCuSS publications use weak-lensing derived radii, some use X-ray scaling relation derived radii. Some uncertainty as to which region was used remains here, though as XGA luminosity measurements have been demonstrated to be adequate for XCS and XXL measurements, this does not overly concern us.

This comparison has further demonstrated XGA’s flexibility and the veracity of its global (and core-excised) luminosity and temperature measurements. The analyses and comparisons presented in Section 3.3.7.1, Section 3.3.7.2, and in this section have shown that XGA is consistent with previous work by XCS, capable of accurate fixed aperture measurements akin to those by XXL, and consistent with core-excised values as measured by LoCuSS. Additionally, we have illustrated that we can easily analyse large samples of galaxy clusters with XGA.
3.4 Measuring gas density, temperature, and mass profiles

We wish to measure hydrostatic masses of galaxy clusters from our XMM observations. The hydrostatic mass equation allows us to derive hydrostatic masses from knowledge of the 3D temperature and gas density profiles,

\[
M(<r) = -\frac{k_B r^2}{\rho_g(r) \mu m_u G} \left[ \rho_g(r) \frac{dT(r)}{dr} + T(r) \frac{d\rho_g(r)}{dr} \right],
\]

where \( r \) is the radius within which the mass is being measured, \( \rho_g(r) \) is the baryon mass density profile, \( T(r) \) is the temperature profile, \( \mu = 0.61 \) is the mean molecular weight, \( m_u \) is the atomic mass unit, \( G \) is the gravitational constant, and \( k_B \) is the Boltzmann constant. This is under the assumption that the cluster is spherically symmetric and in hydrostatic equilibrium.

In this section we describe how we use XGA to measure gas density and temperature profiles, fit models to them, and then derive hydrostatic mass profiles from which we can obtain masses within \( R_{500} \). As in Section 3.3 we present comparisons to measurements from literature, both of gas masses and hydrostatic masses.

XGA can measure various different types of radial profile for extended X-ray sources, both from X-ray images and spectra, and as such we implemented a set of classes to provide easy access to the data, model fitting, and visualisation. Specific profile classes also include other features, such as the hydrostatic mass profile’s ability to measure a mass or a gas fraction. Such profiles can either be generated by XGA or set up from existing data, though we do not compare profiles from literature to XGA profiles in this work.

3.4.1 Fitting models to radial profiles in XGA

Various aspects of this work require models to be fitted to radial data profiles. We have implemented several fitting methods, but for this work we make use of the emcee ensemble MCMC sampler which has been widely used in the astronomy community (Foreman-Mackey et al., 2013). This sampler runs multiple interconnected MCMC chains to explore the parameter space. We choose to use a simple Gaussian likelihood function and uninformative uniform priors. The priors are dependant upon the particular profile being fitted, as well as the model choice, and further information can be found in Table 2; we shall refer back to this table in the specific method sections where profile fitting is required.
Table 2: Summaries of the XGA 1D radial models used in this work (though others are implemented and available). Model names and descriptions of their use are included. The units of model parameters are given, as well as the start parameters used for initial fits. Finally details of the parameter priors used for the full MCMC fit are given, where \( \mathcal{U}[A, B] \) indicates a uniform distribution with limits A and B.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Brief description</th>
<th>Parameters</th>
<th>Start Values</th>
<th>Priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double Beta</td>
<td>A simple model for the surface brightness profile, the sum of two ( \beta ) models.</td>
<td>( \beta_1 )</td>
<td>1</td>
<td>( \mathcal{U}[0, 3] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( R_{\text{core}, 1}) [kpc]</td>
<td>100</td>
<td>( \mathcal{U}[1, 2000] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( S_{01} ) [cts(^{-1}) arcmin(^{-2})]</td>
<td>1</td>
<td>( \mathcal{U}[0, 3] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_2 )</td>
<td>1</td>
<td>( \mathcal{U}[0, 3] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( R_{\text{core}, 2}) [kpc]</td>
<td>100</td>
<td>( \mathcal{U}[1, 2000] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( S_{02} ) [cts(^{-1}) arcmin(^{-2})]</td>
<td>1</td>
<td>( \mathcal{U}[0, 3] )</td>
</tr>
<tr>
<td>Simplified Vikhlinin Density</td>
<td>A simplified version of the Vikhlinin et al. (2006) density profile, as used by Ghirardini et al. (2019).</td>
<td>( \beta )</td>
<td>1</td>
<td>( \mathcal{U}[0, 3] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( R_{\text{core}}) [kpc]</td>
<td>100</td>
<td>( \mathcal{U}[0, 2000] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \alpha )</td>
<td>1</td>
<td>( \mathcal{U}[0, 3] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( R_s) [kpc]</td>
<td>300</td>
<td>( \mathcal{U}[0, 2000] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \epsilon )</td>
<td>2</td>
<td>( \mathcal{U}[0, 5] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( N_0 ) [M(_{\odot}) Mpc(^{-1})]</td>
<td>( 1 \times 10^{13} )</td>
<td>( \mathcal{U}[1, 10000] \times 10^{12} )</td>
</tr>
<tr>
<td>Simplified Vikhlinin Temperature</td>
<td>A simplified version of the Vikhlinin et al. (2006) temperature profile, as used by Ghirardini et al. (2019).</td>
<td>( R_{\text{cool}}) [kpc]</td>
<td>50</td>
<td>( \mathcal{U}[10, 500] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( a_{\text{cool}} )</td>
<td>1</td>
<td>( \mathcal{U}[0, 5] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( T_{\text{min}}) [keV]</td>
<td>3</td>
<td>( \mathcal{U}[0.1, 6] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( T_{1}) [keV]</td>
<td>6</td>
<td>( \mathcal{U}[0.5, 15] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( R_F) [kpc]</td>
<td>200</td>
<td>( \mathcal{U}[100, 500] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \epsilon )</td>
<td>1</td>
<td>( \mathcal{U}[0, 5] )</td>
</tr>
</tbody>
</table>
All profile fitting performed for this work used an emcee sampler instance with 20 walkers, each taking 30000 steps. We chose to generate sets of start positions for the walkers by adapting a simple method suggested in the emcee documentation. We use the SciPy (Virtanen et al., 2020) implementation of the non-linear least squares (NLLS) fitting method (curve_fit) to fit the model to the profile. The NLLS fits use starting values for model parameters as defined in Table 2. We then find the order of magnitude of each fit parameter. These are then perturbed by drawing an $N_{\text{walker}} \times N_{\text{par}}$ matrix of random values from a $\mathcal{N}(0,1)$ distribution, multiplying each random value by the order of magnitude of the respective parameter, and then adding this perturbation to the original NLLS best fit value. Before starting, the sampler checks to ensure that the start positions all fall within the allowed boundaries of the priors, and if not then they are re-drawn until they do.

Once the sampler run is complete, we calculate integrated autocorrelation times for the model parameters, if all autocorrelation times are more than 400 times smaller than the number of steps taken by a walker, then we find the mean autocorrelation time for all the parameters, round it up to the nearest 100, and double it to find a good number of steps to remove from each walker’s chains as a burn-in period. If any of the parameters had an autocorrelation time that did not fit this criteria, then we take a brute force approach and remove the first 30% of steps in each walker’s chains. We draw 10000 random points from combined parameter chains from all walkers to create the final posterior distributions. We also measure medians of the parameter posterior distributions, as well as finding the $1\sigma$ regions.

These fits are performed using a set of radial model classes that we have implemented in XGA. They provide easy access to model posterior distributions, visualisations, and predictions. XGA models also have implementations of useful mathematical operations such as differentiation, spherical volume integration, and inverse Abel transformation of the models (using PyAbel; Hickstein et al., 2019; Gibson et al., 2022), making use of Astropy (Astropy Collaboration et al., 2013, 2018) to provide correct units for any calculated quantities. Model instances also provide information on model parameters, support custom start parameters, and also custom priors.

### 3.4.2 PSF correction of XMM images

To measure a valid density profile from the surface brightness profile of a galaxy cluster (presented in Section 3.4.3), the effect of the instrument point spread function (PSF) must be accounted for. The PSF will have caused detected photons to be deflected from the position on the detector
Figure 3.8: A figure demonstrating the image PSF correction capabilities of XGA. The left hand side shows a zoomed view of the original combined RateMap (stacked images from multiple observations and instruments divided by the equivalent stacked exposure maps). The right hand side shows the same zoomed view of the PSF-corrected stacked RateMap, where the individual images were all corrected separately and then stacked, then divided by the same combined exposure maps.

that corresponds to their true emission point on the cluster, changing the shape of the eventually measured surface brightness profile.

The XMM EPIC-PN on-axis PSF has a full width half maximum of $\sim12.5''$, and the EPIC-MOS1 and MOS2 camera on-axis PSFs have a FWHM of $\sim4.3''$. The PSF of all three EPIC cameras changes size and shape depending on the position on the detector, with the PSF causing stretching of bright sources along the azimuthal direction. PSF effects are particularly noticeable for very bright point sources (causing spoking and smearing of the source), as the high photon count-rate means that enough photons arrive to trace the structure of the PSF effect, rather than the structure of the source. The XMM PSFs also have a non-zero effect on observations of extended sources, especially brighter ones. Previous analyses (e.g., Croston et al., 2006; Eckert et al., 2016) have accounted for this by constructing a mixing matrix that describes the shifting of photons between different radial bins by a 1D model of the PSF.

We have taken a different, generalised, approach to correcting for PSF effects. Rather than apply PSF corrections in 1D after the surface brightness profile has been created, we opt to correct the image itself, then create a profile from that. This approach has several advantages when compared to the mixing matrix method. The primary benefit is the ability to correct in two dimensions, especially important for off-axis clusters (where the XMM PSF becomes highly non-symmetric).
It also reduces the spread of any point sources that may be contaminating the cluster of interest, making them easier to fully remove from the observation with the XGA generated mask. Also, as images are PSF corrected without reference to a particular source, they can be used in many different types of analysis.

We applied the widely used Richardson-Lucy algorithm, presented in both Richardson (1972) and Lucy (1974), to individual XMM images to correct for PSF effects. Instances of the ELLBETA PSF model developed by Read et al. (2011) are generated using the SAS psfgen tool. An added complication is the variation of the PSF over the field of view; the Richardson-Lucy algorithm involves convolving the PSF function with the image, and when the PSF varies dependant on position there is no one to use. As such, each observation is divided into a grid of chunks, with a PSF being generated for the centre of each chunk, and each chunk being of equal size. The number of bins per side of the image is user configurable, with all corrections in this work using four bins per side, making sixteen chunks. Then, for each chunk, the whole image is convolved with that PSF (this is to avoid edge effects caused by Fast Fourier Transform convolution), and then the chunk is cut out of the convolved image and stitched into a final, overall image. This convolution process is then iterated, until a final output image for each ObsID-Instrument combination associated with a source is generated (all correction in this work uses 15 iterations). Once individual PSF-corrected images are created, stacked PSF corrected images are created in the same way uncorrected stacked images are made (Section 3.3.3). They are then divided by the existing stacked exposure maps to create PSF-corrected stacked count-rate maps. Figure 3.8 shows an example comparison of a pre and post correction combined count-rate map.

3.4.3 Gas density profiles

In XGA we have implemented functionality to measure 3D density profiles from radial X-ray surface brightness profiles, \( S_B(r) \). This method implicitly assumes that the cluster is spherically symmetric. Furthermore, the analysis assumes the observed X-ray emission can be described by an APEC (Smith et al., 2001) plasma model with galactic hydrogen column absorption provided by tbabs (Wilms et al., 2000), and that emission from contaminating sources (and the X-ray background) has been properly removed.

We use XGA to construct radial surface brightness profiles from combined (Section 3.3.3) PSF corrected (Section 3.4.2) ratemaps. For the XXL-100-GC and LoCuSS High-\( L_X \) samples the profiles are centred on the X-ray derived coordinates from literature, and for the SDSSRM-XCS sample we
Figure 3.9: This figure shows visualisations of profiles generated by XGA for a galaxy cluster from the SDSSRM-XCS sample, SDSSRMXCS-134. The left hand side shows a surface brightness profile, generated from a combined image in the 0.5-2.0 keV energy band. It has been fitted with a double beta model. The right hand side shows the density profile generated from the surface brightness profile, with a simplified Vikhlinin density model fitted.

centre on the X-ray centroid positions from the XCS source finder XAPA. Galaxy cluster surface brightness profiles are generated out to $R_{500}$ (taken from literature), and an annulus of between 1.05-1.5$R_{500}$ is used to measure a background level. Contaminating sources are removed prior to profile generation by applying an XGA generated mask to the combined image used for generation. Surface brightness profiles are generated by the radial_brightness function, with radial bins of width 1 pixel. XGA surface brightness profiles are measured in the 0.5-2.0 keV energy range by default. We then fit the double beta model (Table 2) to the surface brightness profile data (fit method information in Section 3.4.1, model information in Table 2). Once a profile has been parametrised, we use an inverse-Abel transform to deproject it from a 2D into a 3D profile, assuming spherical symmetry (this has an analytical solution for the double beta model). This results in a 3D radial volume emissivity profile $\epsilon(R)$, in units of photons per second per volume (volume can be in various units depending on the user).

Next, we need to make use of the definition of the APEC emission model normalisation,

$$N_{APEC} = \frac{10^{-14}}{4\pi(D_A(1+z))^2} \int n_e n_H dV,$$

(3.2)

where $N_{APEC}$ is the normalisation of the APEC plasma emission model (in units of cm$^{-5}$), $D_A$ is the angular diameter distance to the cluster (in units of cm), $z$ is the redshift of the cluster, $n_e$
and \( n_p \) are the electron and hydrogen number densities (in units of \( \text{cm}^{-3} \)). This will help infer the gas density profile by enabling a conversion between the volume emissivity profile \( \epsilon(R) \), and the density profile.

We calculate a conversion factor, \( K_N \), between APEC normalisation \( \mathcal{N}_{\text{APEC}} \) and count-rate \( C_r \), where \( \mathcal{N}_{\text{APEC}} = K_N C_r \). To calculate the conversion factor, \( K_N \), we implement an XGA interface to the XSPEC tool FakeIt, which can be used to simulate spectra given an emission model and an instrument response.

This is used to generate simulated spectra for every instrument of every observation of a particular cluster (using ARFs and RMFs generated at the cluster coordinates during the generation of spectra, see Section 3.3.6). The simulations are performed using an APEC model absorbed with tbabs. The nH value for tbabs is set to the HI4PI Collaboration et al. (2016) value for that cluster. The spectra are simulated with a set normalisation of 1, a set temperature measured for that cluster (within \( R_{500} \) for LoCuSS High-L_X/SDSSRM-XCS, within 300 kpc for XXL-100-GC), a metallicity of 0.3 \( Z_{\odot} \), and the input redshift of the cluster. The simulated spectra are all ‘observed’ for a set exposure time of 10 ks, and are used to measure count-rates in the 0.5-2.0 keV band, which corresponds to the energy range of the ratemaps used to generate surface brightness profiles. The count-rates for each instrument of each observation are then weighted by the average effective area between 0.5-2.0 keV (drawn from the corresponding ARF) and combined into a single conversion factor \( K_N \) by the norm_conv_factor method of the GalaxyCluster class. This conversion factor is suitable for use with emissivity profiles generated from combined ratemaps.

This conversion factor, \( K_N \), allows Equation 3.2 to be written in terms of the 3D emissivity \( \epsilon(R) = \frac{C_r(R)}{V} \) (which we can measure), and the product of the electron and hydrogen number densities \( n_e n_H \) (which we can use to calculate the total gas density),

\[
    n_e(R)n_H(R) = \frac{4\pi K_N (D_A (1 + z))^2 C_r(R)}{10^{-14} V}.
\]  

(3.3)

At this point we must assume the ratio of electrons to hydrogen in the intra-cluster medium, so as to calculate a total gas density. We choose to use the solar abundances presented in Anders and Grevesse (1989) to calculate this ratio, \( R_{eH} = 1.199 \). The total gas number density \( n_g = n_e + n_H = (1 + R_{eH}) n_H \), is thus calculated by finding an expression for \( n_H \) from Equation 3.3, then using \( R_{eH} \) again to find \( n_e \) and sum the two number densities. At this point the mean molecular weight \( \mu = 0.61 \) and the atomic mass unit are used to convert number density to mass density.
Thus, we calculate gas density from emissivity with quantities that we can measure, or already know;

$$\rho_{\text{gas}}(R) = \mu m_u (1 + R_{eH}) \sqrt{\frac{4\pi K_N R_{eH} \epsilon(R)(D_A(1+z))^2}{10^{-14}}}.$$  \hspace{1cm} (3.4)

Once gas density is measured using Equation 3.4, it is stored in a `GasDensity3D` instance. We then use the fitting functionality discussed in Section 3.4.1 to fit simplified Vikhlinin density models (see Table 2) to the profiles.

Once a gas density profile has been measured (and had a model fitted to it) we perform a spherical volume integral,

$$M_{\text{gas}}(< R_\Delta) = 4\pi \int_0^{R_\Delta} \rho_{\text{gas}}(R) R^2 dR,$$  \hspace{1cm} (3.5)

to measure a gas mass within a given radius. `XGA` can also perform this integral at a set of radii to create a cumulative gas mass profile (stored in `GasMass1D`).

### 3.4.4 Comparisons of XGA gas mass measurements to literature

We compare `XGA` measurements to previous measurements of the XXL-100-GC and LoCuSS High-$L_X$ samples, similar to the comparisons we made in Section 3.3.7.

#### 3.4.4.1 XXL-100-GC gas masses

We generate baryon density profiles for 91 clusters in the XXL-100-GC sample, with outer radii of $R_{500}$ (overdensity radii measured by Eckert et al., 2016). To maximise consistency we use the XXL measured $T_{300kpc}$ values in the calculation of the conversion factors ($K_N$) between APEC normalisation ($N_{\text{APEC}}$) and count rate ($C_T$). Each cluster’s neutral hydrogen column density is used in the conversion factor calculation, as is its redshift, and we assume a metallicity of 0.3 $Z_\odot$.

Figure 3.10 shows the comparison between `XGA` measured and original XXL gas masses. The $R_{500}$s within which XXL gas masses are measured have been published with uncertainty values (some with a significant fractional uncertainty), and the gas mass measurements presented in Eckert et al. (2016) account for those uncertainties. As such we implement a method of measuring gas mass that can include information on the radius uncertainty in the measurement. The posterior distribution of radius for each XXL cluster is assumed to be Gaussian, with the mean being the
Figure 3.10: A one-to-one comparison of gas masses measured for the XXL-100-GC cluster sample, within $R_{500}$, by Eckert et al. (2016) and an XGA reanalysis. $S_B$ profiles were fitted with Beta profiles, density profiles with King profiles. Contains measurements for 91 of 96 XXL-100-GC clusters analysed with XGA.
Figure 3.11: A one-to-one comparison of gas masses measured for the LoCuSS High-$L_X$ cluster sample, within $R_{2500}$ and $R_{500}$, by Martino et al. (2014) and an XGA reanalysis.

There is broad agreement between the two sets of measurements. This is encouraging considering the significant differences in density measurement between the two samples; including the PSF correction method, fitting method, and other differences mentioned in Section 3.3.7.2. We observe that the comparison isn’t a perfect 1:1 relation, and that there is an offset between the two measurement sets that evolves with gas mass. An evolving offset between XGA and XXL $L_X$ measurements was also observed in Figure 3.6, with the explanation for both likely being different background treatments.

3.4.4.2 LoCuSS High-$L_X$ gas masses

Gas masses are also available for the LoCuSS High-$L_X$ cluster sample, and we perform a comparison between the original values measured by Martino et al. (2014) and our reanalysis of the sample. LoCuSS provides an extra point of comparison, especially as the clusters also have gas masses measured within $R_{2500}$. Observations of LoCuSS High-$L_X$ clusters also tend to be more massive and better observed than XXL-100-GC clusters (having dedicated pointed observations rather than being part of a survey). We estimate gas masses within the overdensity radii provided
in Martino et al. (2014) in order to minimise analysis differences.

Comparisons for $M_{\text{gas}}^{500}$ and $M_{\text{gas}}^{2500}$ are presented in Figure 3.11. We measure density profiles out to a radius of $R_{500}$. We see that the XGA measurements of $M_{\text{gas}}^{2500}$ are consistently higher than their LoCuSS counterparts; we can infer that the LoCuSS analyses measure fewer baryons in the core regions of the clusters than the XGA does. There are differences between the original LoVoCCS method and our reanalysis that likely account for the offset in gas mass measurements. The most significant are a different approach to PSF correction, where the entire image is corrected for two dimensional PSF effects, from which a surface brightness profile is constructed, rather than a brightness profile being constructed and then corrected for one dimensional PSF effects. The greatest effect of PSF correction is in the inner regions of the clusters, as the brightest part of the objects, which is where the greatest difference in gas mass is apparent. In addition to the differing methods of dealing with the XMM PSF, the differing treatment of background emission in the two analyses will contribute to the different gas mass values. The LoCuSS analysis has a more sophisticated approach involving the separate modelling of different background components. This may cause the background subtracted brightness profiles produced for LoCuSS may differ from the XGA profiles, possibly even affecting the slope of the profile - this would contribute to the gas masses being more similar within $R_{2500}$ than within $R_{500}$.

The uncertainties on the gas masses presented in Figure 3.11 are notably smaller than those shown for the XXL comparison in Figure 3.10. This is due largely to the large radius uncertainties of the XXL sample, which were derived from low signal to noise X-ray observations. The LoCuSS sample do not have published radius uncertainties, and as such they are not accounted for in the gas mass measurements.

### 3.4.5 Generating and fitting annular spectra with XGA

We wish to explore how certain spectral properties of a galaxy cluster vary with radius. As such, we implement methods to generate sets of annular spectra within radial bounds, as well as methods that attempt to find optimal bounds for the annuli. It is common practice to generate annuli with specific data quality; e.g. a minimum signal to noise or number of X-ray counts per annulus. This is to ensure that any spectra that are generated are of a sufficient quality for spectral fitting. We implement a function in XGA where annuli are constructed so that they have a minimum (user defined) background subtracted counts per annulus based on the combined observation of the cluster in question. We also impose a requirement of a minimum annulus width of 20ʺ (also
Figure 3.12: Set of annular spectra fitted folded models for SDSSRMXCS-134, performed by XSPEC, are shown as solid lines. Spectral data points are not included for clarity. The radius axis indicates the average of the inner and outer radii of the annulus each spectra was measured from, with the zero point being the coordinate defined as the centre of the cluster.
user configurable), to counter PSF effects of the extracted spectrum. Constructing the annuli is an iterative process, where initial annuli of the minimum allowed width are constructed between the cluster centre (the first annulus is a circle) and the outer radius specified for the set of annular spectra (though as the outermost radius must be an integer multiple of 20″ the true outer radius will likely be slightly larger than what was initially specified for the cluster). These initial annuli are combined until the minimum signal-to-noise is achieved in all annuli, or until only four annuli remain.

Annular spectra are generated using the same default settings as the global spectra detailed in Section 3.3.6. The set of annular spectra for a particular observation-instrument combination will (by default) have a background spectrum generated in the 1.05-1.5 $R_{\text{outer}}$ region, where $R_{\text{outer}}$ is the outer radius of the $N^{th}$ annulus. To measure a temperature profile, we employ the same fitting technique that we applied to global spectra in Section 3.3.6; each annulus has simultaneous fits performed on all available data using the constant*tbabs*apec model. The same parameter start values and spectrum quality verification steps are used.

The generated spectra are stored in instances of an XGA product class, AnnularSpectra, as are the fit results. Figure 3.12 shows a visualisation of an example set of annular spectra for SDSSRMXCS-134 as a function of radius. It clearly illustrates the decreasing emission moving from the core of the cluster to the outskirts, as well as the flattening of the slope between $\sim$1-2 keV that indicates a lessening temperature.

### 3.4.6 Three-dimensional gas temperature profiles

The second property we must measure to calculate a cluster’s hydrostatic mass is the three-dimensional temperature profile. First, we measure projected temperature profiles by generating spectra in annuli centered on the cluster coordinates, then performing spectral fits to each annulus to determine a temperature. Previous XCS work (Lloyd-Davies et al., 2011) has indicated that around 1000 counts are required to measure well constrained temperatures (with percentage uncertainties of 10-30%, depending on the temperature of the cluster). That means that all annuli need to have that many counts to achieve a well constrained projected temperature profile. We make use of the method described in Section 3.4.5 to generate annular radii, centered on the supplied cluster coordinates, that have a minimum number of counts defined by the user (3000 for the LoCuSS analysis, 1500 for the SDSSRM-XCS analysis).

Once the spectra have been generated, stored in an AnnularSpectra instance, and had XSPEC
fits performed, the resulting projected temperature measurements (as a function of radius) are stored in an `ProjectedGasTemperature1D` instance. Each projected temperature is a weighted combination of the temperatures in the three-dimensional shells of the cluster that the annulus intersects with. We opt to use the ‘onion-peeling’ method (see descriptions in e.g., Ettori et al., 2002; Ghirardini et al., 2018). First, the volume intersections between the analysis annuli (projected back into the sky) and the spherical shells which we separate each cluster into (defined as having the same radii as the annuli) must be calculated. The volume intersections can be calculated (see Appendix A of McLaughlin, 1999) as,

$$V_{\text{int}} = \frac{4\pi}{3} \left[ \max\{0, (R_{\text{so}}^2 - R_{\text{ao}}^2)^{\frac{3}{2}}\} - \max\{0, (R_{\text{so}}^2 - R_{\text{ao}}^2)^{\frac{3}{2}}\} + \max\{0, (R_{\text{si}}^2 - R_{\text{ao}}^2)^{\frac{3}{2}}\} - \max\{0, (R_{\text{si}}^2 - R_{\text{ai}}^2)^{\frac{3}{2}}\} \right],$$

where $R_{\text{so}}$ is the matrix of shell outer radii, $R_{\text{ao}}$ is the matrix of annulus outer radii, $R_{\text{si}}$ is the matrix of shell inner radii, and $R_{\text{ai}}$ is the matrix of annulus inner radii. The calculation is implemented in the XGA function `shell_ann_vol_intersect`. The $V_{\text{int}}$ matrix is a two-dimensional matrix describing volume intersections between all combinations of annuli and three-dimensional shells.

Second, an emission measure profile (EmissionMeasure1D) is calculated using the `emission_measure_profile` method of the APECNormalisation1D profile produced during the spectral fitting of the set of annular spectra. This calculation uses Equation 3.2 and rearranges to solve for the integral. As the projected temperature measured within a given annulus is a weighted combination of the shell temperatures intersected by that annulus, we can infer the temperature of a shell (for past usage see Ghirardini et al., 2018) with

$$T_{\text{shell}} = \left( V_{\text{int}} T \right)^{-1} \times T_{\text{annulus}} \times \text{EM} \times \left( V_{\text{int}} T \right)^{-1} \times \text{EM}. \quad (3.7)$$

$T_{\text{shell}}$ is the matrix of spherical shell temperatures that we aim to calculate, $V_{\text{int}}$ is the matrix of volume intersections between all combinations of annuli and shells (see Equation 3.6), $\text{EM}$ is the emission measure matrix, and $T_{\text{annulus}}$ is the projected temperature matrix; $\times$ represents the matrix product. Once calculated, the $T_{\text{shell}}$ values and uncertainties are stored in a `GasTemperature3D` instance.

Once three-dimensional, de-projected, gas temperature profiles have been measured, we use the methods discussed in Section 3.4.1 to model the profiles with the simplified Vikhlinin temperature model (see Table 2 for more information).
Figure 3.13: An example fitted three-dimensional temperature profile generated by XGA for a galaxy cluster in the SDSSRM-XCS sample, SDSSRMXCS-134. This visualisation was created using the view() method of the temperature profile. Data points are shown in black, and the simplified Vikhlinin temperature profile model is shown in green. This temperature profile was generated by deprojecting temperatures measured by fitting the annular spectra models shown in Figure 3.12.
Figure 3.14: An example hydrostatic mass profile generated for a galaxy cluster in the SDSSRMXCS sample, SDSSRMXCS-134. This profile was generated using the model fits to the density profile shown in the right hand side of Figure 3.9, and the temperature profile shown in Figure 3.13.
3.4.7 Hydrostatic mass profiles

Once we have measured a 3D radial density profile (details in Section 3.4.3) and a 3D radial temperature profile (details in Section 3.4.6), we can use Equation 3.1 to measure hydrostatic masses. As such, XGA creates hydrostatic mass profiles as a function of radius (e.g. Figure 3.14), which are stored in HydrostaticMass class instances. The density and temperature profiles must have models fitted prior to the construction of a hydrostatic mass profile, with the fitting process described in Section 3.4.1 and models described in Table 2. The hydrostatic mass equation not only involves the temperature and density profile values at each radius for which a enclosed mass is measured, but also the derivatives of those profiles with respect to radius. The simplified Vikhlinin density and temperature profiles both have analytical first-derivatives, and as such these are used to calculate the slope at a given radius, rather than a numerical approximation. XGA models can return posterior distributions of the value of the model (or model slope) at a given radius, rather than a single value. This allows for uncertainties to be propagated from the density and temperature profile through to the mass profile.

3.4.8 Comparison of XGA hydrostatic mass measurements to the LoCuSS High-\(L_X\) sample

As a consistency check, we use XGA to measure hydrostatic masses for the LoCuSS High-\(L_X\) sample.

To minimise differences with the original LoCuSS High-\(L_X\) hydrostatic mass analysis, we leave the metallicity of each annulus free to vary during the spectral fits. Density and temperature profiles are generated with an outer radius of \(R_{500}\), fitted with models as described in Section 3.4.3 and Section 3.4.6. Mass profiles are constructed (see Section 3.4.7) and used to measure \(M_{500}^{hy}\) and \(M_{2500}^{hy}\). The temperature profiles are constructed with a minimum annular width of 20'', and a minimum number of background subtracted counts per annulus of 3000 (an attempt to maximise the similarity of this reanalysis with the original work). The background regions are taken from the 1.05-1.5\(R_{\text{outer}}\) annulus, where \(R_{\text{outer}}\) is the outer radius of the \(N^{\text{th}}\) annulus, which will likely be very slightly larger than \(R_{500}\) (as described in Section 3.4.5, the outer radius must be an integer multiple of the minimum annulus width).

Figure 3.15 shows a comparison between the XGA and LoCuSS hydrostatic masses for \(M_{2500}^{hy}\) (left-hand-side) and \(M_{500}^{hy}\) (right-hand-side). We successfully measure mass profiles for 41 galaxy
Figure 3.15: A one-to-one comparison of hydrostatic masses measured for the LoCuSS High-$L_X$ cluster sample, within LoCuSS measured $R_{2500}$ and $R_{500}$ values, by Martino et al. (2014) and an XGA reanalysis.

clusters in the LoCuSS High-$L_X$ sample, 29 of which have XMM hydrostatic masses measured by LoCuSS. When plotted directly against the published LoCuSS masses (specifically those measured by XMM) we see a relatively good agreement for both overdensities. The agreement between the measured masses is not perfectly one-to-one; this is expected, since the number of components involved in measuring hydrostatic mass make it difficult to exactly match measurement methods between analyses. For example, XGA and LoCuSS take different approaches to measuring density profiles. Furthermore, additional XMM data is available for these clusters than there was at the time of the original Martino et al. (2014) analysis, which may have an effect on the constraints.

### 3.4.9 Inspection of data and measurements

To ensure that the masses measured are valid, visual inspection of key data products is performed. This expands upon the visual inspection and region adjustment that was described in Section 3.3.5. All profile-based products that contribute to the measurement of the hydrostatic mass profiles are inspected; surface brightness, density, temperature, and finally the mass profiles.

The inspections are to look for obviously invalid or unphysical data. For instance, more complex than usual surface brightness profiles that have been poorly described by the double beta model fit to them. Temperature profiles with poor fits, very poorly constrained data points, or unphysical
attributes (such as temperature increasing in the outskirts of a cluster) are flagged. Finally, mass profiles that do not appear to be strictly monotonically increasing are flagged, as each point in the mass profile represents an enclosed mass, so the mass measured at annulus $R_{j+1}$ must always be greater than at $R_j$. The HydrostaticMass class provides a method (diagnostic_view) to generate a set of visualisations (a combination of Figures 3.9, 3.13, and 3.14) for each cluster.

3.5 Mass analysis and scaling relations for the SDSSRM-XCS sample

Here we present hydrostatic masses for a subset of galaxy clusters in the SDSSRM-XCS sample presented by Giles et al. (2022b). With these masses, we present a series of mass-observable relations.

3.5.1 Masses for a subset of the SDSSRM-XCS sample

We attempt to measure masses for all 150 galaxy clusters in the SDSSRM-XCS sample. Measurements presented in this section are performed using XGA, though we use $R_{500}$ values measured by Giles et al. (2022b). In Section 3.3.7.1 we demonstrated that XGA $T_X$ and $L_X$ measurements are almost completely consistent with the original XCS measurements when measured within identical apertures.

Temperature and density profiles were initially generated out to $R_{500}$, where the $R_{500}$ used was measured by Giles et al. (2022b) using a radius-temperature relation. Density profiles were generated as described in Section 3.4.3. The temperature profile annuli are chosen so that each annuli has a minimum width of 20" and a minimum number of background subtracted counts per annulus of 1500. The minimum counts check is performed on the combined data for all relevant observations of a cluster.

We are able to successfully measure hydrostatic masses for 102 of the 150 galaxy clusters in the SDSSRM-XCS sample. The temperature profiles used for these measurements have between 4-26 annuli. We perform a full examination of relevant data products for each cluster, as described in Section 3.4.9. This process flags any suspect or blatantly unphysical results, and allows us to perform a deeper inspection and determine why a particular result has been measured.

This set of hydrostatic masses adds significant value to the SDSSRM-XCS sample, as well as
representing one of the largest samples of cluster hydrostatic masses available. As each mass is measured individually, rather than derived through the stacking of different clusters (as is usually the case for weak lensing cluster masses), there are many possible uses for the sample. Also, as SDSSRM-XCS clusters were selected from optical/NIR via the SDSS redMaPPer catalogue (Rykoff et al., 2014), this sample allows for powerful multi-wavelength studies of how cluster mass is related to optical properties.

### 3.5.2 Construction of mass-observable relations

We use the masses measured for the SDSSRM-XCS sample to construct new mass-observable scaling relations. These are compared to examples taken from literature. In all cases we perform the fits in log space using the R package LInear Regression in Astronomy (LIRA, Sereno, 2016b), fully described in Sereno (2016a).

We fit a relation of the form,

\[
\log \left( \frac{M_h^{\Delta}}{E(z)^y M_0} \right) = \log(A_{MO}) + B_{MO} \log \left( \frac{O}{O_0} \right) \pm \sigma_{M/O},
\]

(3.8)

where \(M_h^{\Delta}\) is the hydrostatic mass within an overdensity radius \(R_\Delta\), \(E(z)\) is the Hubble parameter, \(y\) is an exponent applied to the Hubble parameter, \(M_0\) is the mass normalisation, \(O\) is the observable parameter, \(O_0\) is the observable parameter normalisation. Finally, \(A_{MO}\) is the normalisation, \(B_{MO}\) the slope, and \(\sigma_{M/O}\) the intrinsic scatter of the relation. We construct scaling relations based upon multiple observables, both measured from X-ray observations and from the optical/NIR SDSS observations.

#### 3.5.2.1 \(M_h^{\Delta}\)-\(T_{X,\Delta}\) scaling relation

In Figure 3.16 we present the scaling relation between hydrostatic mass and temperature measured within literature values of \(R_{500}\) presented by Giles et al. (2022b), as well as a scaling relation from the masses and temperatures published by (Arnaud et al., 2005, the relation is re-fit with the same methodology as our new relation). The new SDSSRM-XCS mass-temperature relation demonstrates a strong correlation between the hydrostatic mass measurement, and measurements of temperature, as is predicted and has been found by other works (e.g. Arnaud et al., 2005; Lovisari et al., 2020). The slope of the new relation is lower than that of the comparison relation.
Figure 3.16: A hydrostatic mass to temperature relation constructed using masses measured from 102 clusters from the SDSSRM-XCS sample, shown in blue. A re-fitted scaling relation using the Arnaud et al. (2005) data is shown in orange for comparison. SDSSRM-XCS masses and temperatures are measured within $R_{500}$, Arnaud et al. (2005) temperatures are measured within [0.1-0.5]$R_{200}$. 
Figure 3.17: A hydrostatic mass to luminosity relation constructed using masses measured from 102 clusters from the SDSSRM-XCS sample. Masses and luminosities are measured within $R_{500}$. Luminosities are unabsorbed and measured within the soft band (0.5-2.0 keV).
though consistent within $2\sigma$; the slope is also consistent with the prediction for this relation from self-similarity (1.5). The greater number of mass measurements available for our new relation also allows for a more reliable estimate of the scatter of mass with temperature, and we find a higher median value (0.12) than the older relation (0.04). The normalisation of the SDSSRM-XCS mass-temperature relation is higher than that of the Arnaud et al. (2005) scaling relation, indicating that the masses measured in this work are larger when compared to the measured temperatures. This may be because the Arnaud et al. (2005) relation uses temperatures measured within a within $[0.1-0.5]R_{200}$ annulus, making for a slightly inconsistent comparison.

Figure 3.16 also demonstrates that the SDSSRM-XCS sample covers a wide range of temperature and mass values, with clusters and groups of every size represented. This is partially due to the selection of galaxy clusters from optical/NIR data, as selection from SZ and X-ray can have stronger mass dependencies on the signal to noise. This relation clearly demonstrates that X-ray temperatures are an excellent proxy observable for cluster mass, and should be the observable of choice for X-ray astronomers (among the relations presented in this work).

### 3.5.2.2 $M_{\Delta}^{\text{hy}} - L_{X,\Delta}$ scaling relations

The second scaling relation measured from the SDSSRM-XCS sample makes use of galaxy cluster X-ray luminosity as the observable. Figure 3.17 shows the fit performed to the data with LIRA, with a clear relationship between the mass of a cluster and its soft-band (0.5-2.0 keV) unabsorbed X-ray luminosity. More scatter in the mass with observable is evident here than in the mass-temperature relation presented in Figure 3.16; this is again in line with previous work on the measurement of mass luminosity scaling relations.

The closest comparison we can make is to the work performed by Lovisari et al. (2020), which measured hydrostatic masses and luminosities for a sample of Planck-selected galaxy clusters. Their relation between mass and luminosity exhibits a steeper slope than SDSSRM-XCS relation, though several factors complicate this comparison. The Lovisari et al. (2020) luminosities are measured within a different energy band (0.1-2.4 KeV), and the luminosity limit of the sample is much higher than the SDSSRM-XCS sample (this can alter the slope of the relation). Neither relation’s slope is wholly consistent with the self-similar prediction. More scatter is evident in the SDSSRM-XCS relation, particularly in the lower-luminosity regime that is not explored by the Planck-selected sample.

Mass-luminosity scaling relations will be of great importance to the ongoing eROSITA All-Sky
Figure 3.18: SDSSRM-XCS hydrostatic mass to richness relation (in purple) constructed using masses measured from 102 clusters from the SDSSRM-XCS sample, with masses measured within $R_{500}$ and richnesses are taken from the SDSS redMaPPer catalogue. A comparison relation re-fitted from masses and richnesses presented by Andreon and Congdon (2014); these masses are derived from weak lensing and both masses and richnesses are measured within 0.5 Mpc apertures.
Survey cosmology efforts, as luminosity is a much easier X-ray observable to measure than temperature in terms of the required data quality. As such many more luminosity than temperature measurements will be acquired, and thus many more inferred masses will be able to be included in their halo mass function and thus cosmological analysis.

### 3.5.2.3 $M_{\Delta}^{\text{hy}} - \lambda$ scaling relation

Finally, we construct multi-wavelength scaling relations between hydrostatic mass and optical richness. Richness ($\lambda$) is a probabilistic measure of the number of galaxies in a galaxy cluster, and cannot be measured using X-ray observations. It is one of the outputs of the redMaPPer galaxy cluster finder, which was run on SDSS photometry (Rykoff et al., 2014) and was used to construct the SDSSRM-XCS sample (Giles et al., 2022b). Figure 3.18 shows the first scaling relation constructed between X-ray hydrostatic masses and redMaPPer optical richnesses. Like the mass-luminosity relation this scaling relations exhibits more scatter than is seen in the mass-temperature relation, but this scaling relation puts some of the first direct constraints on mass normalisation with respect to richness (from a source independent of the observable measurements). Importantly this relation can also be used to constrain the scatter of mass with observable, which has to be factored in for large scale mass inference using scaling relations.

Also present in Figure 3.18 is a scaling relation (fit with the same methodology as the SDSSRM-XCS relation) between weak-lensing masses and non-redMaPPer richnesses presented by Andreon and Congdon (2014). The slopes of the two relations are completely consistent with one another, though the normalisations are not. This is due to the different origin of the mass measurements in the Andreon and Congdon (2014) relation; weak-lensing masses are known to be larger than hydrostatic masses measured by analysis of X-ray observations (Smith et al., 2016).

### 3.6 Discussion and next steps

Finally, in this section we prevent summaries of the work in this paper, discussions of the implications and results, and goals for related future work.
3.6.1 Discussion

We make use of three samples of galaxy clusters, SDSSRM-XCS, XXL-100-GC, and LoCuSS High-$L_X$ to explore the efficacy of measurements of galaxy cluster properties produced by our new software tool, \textit{XGA}. Once we have demonstrated that \textit{XGA} is reliable, we find and present new measurements of hydrostatic mass for the SDSSRM-XCS sample, using the values to construct new scaling relations. In summary:

- Comparisons of $T_X$ and $L_X$ measurements of the SDSSRM-XCS sample from an existing XCS analysis (Giles et al., 2022b) to those measured by \textit{XGA} demonstrate close agreement. This holds true for measurements within both $R_{500}$ and $R_{2500}$, with $R_{2500}$ measurements appearing to show less scatter. Close agreement of these measurements is expected because they use the same event lists and region files, as well as very similar analysis techniques.

- Similar comparisons of $T_X$ and $L_X$ to literature values measured for the XXL-100-GC and LoCuSS High-$L_X$ samples showed good agreement, and also demonstrated \textit{XGA}'s ability to perform measurements with different energy limits, with metallicity left free to vary, and different cosmologies.

- This work has highlighted that small analysis choices (such as the energy limits used for spectral fits) can have significant effects on global measurements made for a sample of galaxy clusters. This indicates that samples from literature should not be naively combined, but rather that a reanalysis should be performed with consistent analysis choices; \textit{XGA} makes that a realistic possibility.

- The next section of the work described and tested our approach to radial profiles, including the measurement of density and 3D temperature profiles for sets of galaxy clusters. We also demonstrate and explain \textit{XGA}'s PSF correction feature, which takes into account the spatially varying nature of \textit{XMM} PSFs, and is relevant to the measurement of density profiles. Comparisons of gas mass values measured for the XXL-100-GC and LoCuSS High-$L_X$ samples, and reanalyses using \textit{XGA}, demonstrated that \textit{XGA} produces gas mass measurements consistent with past work. The XXL comparison indicated an evolving offset between the original and reanalysis values, likely due to different background approaches. The LoCuSS comparison demonstrated a 1:1 relation (with scatter) between gas masses measured within LoCuSS $R_{500}$ values, but that $R_{2500}$ gas masses are slightly systematically larger when measured by \textit{XGA} versus the original work.
Our final comparison makes use of the hydrostatic masses measured within $R_{2500}$ and $R_{500}$ for the LoCuSS High-$L_X$ sample by Martino et al. (2014). This probes our measurement of temperature profiles. The $M_{2500}^{hy}$ comparison lies close to the 1:1 line, with minimal scatter. The $M_{500}^{hy}$ comparison is also broadly compatible with the 1:1 line, though the scatter and uncertainties are increased when compared to the $M_{2500}^{hy}$ plot. From this comparison, and the others presented in this work, we conclude that XGA mass measurements are consistent with previous work and can be relied on.

Finally, we present new measurements of hydrostatic mass for the SDSSRM-XCS galaxy cluster sample. They are used to construct mass-observable relations, with both X-ray and optical observables.

### 3.6.2 Future Work

Here we detail some of the work that we have planned for our new implementation of hydrostatic mass measurement, as well as for the samples of masses we have (and will) measure.

#### 3.6.2.1 Other Samples

We will produce mass measurements and scaling relations for other samples of galaxy clusters, selected from optical/NIR surveys such as DES, and from the ACT-DR5 cluster catalogue. This shall not only add a significant number of cluster masses to our sample, but will also open the possibility of constructing scaling relations between X-ray mass and Sunyaev–Zeldovich properties.

#### 3.6.2.2 Multi-mission X-ray Analyses

Planned additions to the XGA software package will enable support for X-ray telescopes other than XMM; e.g. Chandra, eROSITA, and XRISM. This will allow a user to draw on multiple archives of X-ray observations for analysis of samples, increasing the likelihood that samples selected from other wavelengths will have a serendipitous X-ray observation.

Multi-mission support will also allow us to design joint analyses that take advantage of the unique capabilities of each telescope. Joint analyses with all available X-ray data should become routine, rather than requiring special effort, and XGA can provide that capability.
We will also include support for simulated telescopes to enable preparatory work for future missions; e.g. *Athena, Lynx,* and *AXIS.* As such we will produce catalogues of hydrostatic masses measured with multiple current telescopes (*XMM* and *Chandra*), as well as explore how we may exploit the capabilities of planned telescopes. Work on simulated clusters will also include exploration of the hydrostatic mass bias from true mass.

### 3.6.2.3 Accounting for non-thermal pressure support

Hydrostatic masses are biased from true masses due to assumptions made during the derivation of Equation 3.1. The assumption of hydrostatic equilibrium implies that all pressure support is provided by the thermal gradient of the ICM, which is not the case. Various processes also help to balance gravitational collapse, and together are often referred to as non-thermal pressure (NTP) support. Methods to measure a total mass from X-ray observations exist (e.g. Eckert et al., 2019a), and we aim to apply them to large samples of clusters such as SDSSRM-XCS. By taking NTP support into account hydrostatic masses can become even more competitive with other direct mass measurement methods (such as individual weak lensing).

### 3.6.2.4 Comparison of density and temperature profiles

While we have compared values derived from radial profiles of density and temperature to the literature, it would be beneficial to compare *XGA* profiles directly to existing profiles. We will use the profiles measured by Lovisari et al. (2020) for comparison, as well as making comparisons to radial profiles from hydrodynamical simulations such as Illustris-TNG and the 300 project. This will give us an idea of how comparable our profiles are to existing observed profiles, and how the observed profiles reflect the true temperature and density structure of the galaxy clusters.

### 3.7 Constructing the ACTDR5-XCS sample

This section details the construction of a sample of galaxy clusters selected from the ACT DR5 SZ selected cluster catalogue (see Section 1.6.6). This sample is created in preparation for a follow-up paper to the work presented in the earlier parts of this chapter, both in order to increase the number of galaxy cluster properties and masses available, but also so that relations between X-ray and SZ properties measured by *XMM* and ACT can be explored.
3.7.1 ACT DR5 clusters with XMM data

The ACT DR5 sample contains 4195 optically confirmed galaxy clusters with an SZ signal to noise ratio greater than 4, detected in 13211 deg$^2$, all with redshift and mass estimates. This not only provides a large dataset to perform follow-up XMM analyses on, but makes it possible to fit models to the spectra to derive properties, as all of the clusters have redshift estimates. Also, thanks to the due diligence of the ACT team confirming that these objects are galaxy clusters, it is not necessary to perform the exhaustive visual inspection process that is normally undertaken during the creation of XCS samples.

As a pre-processing step, it is necessary to calculate an $R_{500}$ estimate for each of the ACT clusters, as it will be essential to use as an analysis aperture later on. In this case a mass estimate for each of the galaxy clusters was used to infer the corresponding overdensity radius, $R_{500}$, by making use of the definition of an overdensity radius;

$$R_{\Delta} = \left( \frac{3M_{\Delta}}{4\pi \Delta \rho_c(z)} \right)^{\frac{1}{3}},$$  

(3.9)

where $R_{\Delta}$ is the overdensity radius, $M_{\Delta}$ is the mass measured within the overdensity radius, $\rho_c(z)$ is the critical density of the Universe at redshift $z$, and $\Delta$ is the overdensity (set to 500 in this case). The ACT DR5 catalogue provides several estimates of mass and the ‘M500cCal’ entry was used for this calculation; this is a mass derived from a scaling relation between SZ signal to noise and mass, then scaled using weak lensing estimates.

To create the ACTDR5-XCS sample, it was first necessary to determine which of the clusters fell on or near an XMM observation. An initial search was performed using XGA’s simple xmm_match function, which simply locates XMM observations that have an aimpoint within a certain distance of the source coordinate; the match distance is set by the user, and in this case was set to 17', a distance that should ensure all relevant observations for each cluster are found (though a wider search radius might be necessary for some low-redshift systems). The ACT SZ detection coordinates were used for this search, rather than optically determined coordinates.

This search finds that 475 (11.32%) clusters fall on or near an XMM observation. The vast majority (309) fall on or near only one XMM observation, with 100 falling on or near two, and the rest falling on or near a greater number of observations (with the maximum being 9). Figure 3.19 shows a representation of the ACT DR5 cluster locations on the sky, with those marked in black falling on or near an XMM observation.
Figure 3.19: A figure showing the distribution of ACT DR5 SZ selected galaxy clusters on the sky. Each cluster is represented by a point, with black points indicating clusters that are in the ACTDR5-XMM sample, and red indicating clusters that have no XMM data.

The next step was to determine which of the clusters had large enough coverage by XMM to be of use for the measurement of temperatures, luminosities, and possibly hydrostatic masses. This was achieved by declaring an XGA ClusterSample instance (2.3.3), feeding in the sub-sample of ACT DR5 clusters on or near an XMM observation that was constructed in the previous step. The ACT RA-Dec positions were again used for this, with X-ray peak finding turned off so that XGA would not calculate an updated central coordinate using a peak finder routine (Section 2.4.7.2), and the sample was configured to discard any observations that did not cover at least 70% of the $R_{500}$ region. This led to a ‘cleaned’ sample of 431 galaxy clusters that will be referred to as the ACTDR5-XMM sample; of those 431 clusters, XGA found that 390 (90%) had at least one match to an XCS extended source.

### 3.7.2 Properties of the ACTDR5-XMM sample

Once a sample was assembled some understanding of the X-ray properties of the clusters was required. In this case that involved measuring ‘global’ X-ray temperature and luminosity values, as they provide good insight into galaxy cluster mass before performing more complex analyses. They are also used to create scaling relations, between X-ray properties, SZ properties, and even-
tually mass measurements. Measuring these values also brings this sample in line with other samples created and published by XCS (such as the SDSSRM-XCS sample; Giles et al., 2022b).

Measurements of temperature and luminosity were performed using the exact same methodology as was described in Section 3.3.6. Spectra were generated centered on the ACT SZ detection RA-Dec position, within the $R_{500}$ aperture estimated from mass in Section 3.7.1. Successful $T_X$ and $L_X$ measurements were achieved for 409 clusters in the ACTDR5-XMM sample. Once measured, the temperature constraints were used to create a sub-sample of ACTDR5-XMM with well constrained temperatures, just as Giles et al. (2022b) did for the SDSSRM-XCS sample. Clusters with a percentage temperature error of less than 25% were included in this sample; 345 clusters passed this check, and are referred to as the ACTDR5-XMM$_{\Delta T_X < 25}$ sample. The XMM suffix is used here rather than ‘XCS’ as the clusters in this sample do not necessarily have an XCS detection.

As the clusters presented in the ACT DR5 catalogue were optically confirmed, many of them have a measure of richness associated with them. These richness measurements come from varying data sources and cluster finders, but do provide insight into the galaxy population of each cluster. The richnesses are taken from the measurements performed on SDSS and DESY3 with the redMaP-Per algorithm (Rykoff et al., 2014), and measurements performed on HSC-SSP with the Camira algorithm (Oguri, 2014). Of the clusters in the ACTDR5-XMM sample, 285 (66%) had some measure of the richness from one of the three sources. The main SZ parameters provided in the ACT DR5 catalogue are SZ signal-to-noise and measures of the Compton-y parameter, which can also be used for creating scaling relations just like X-ray properties.

In summary, this newly constructed ACT DR5 selected sample with XMM data is a powerful and high quality dataset. The XGA-measured luminosities and temperatures represent one of the largest consistently selected samples of galaxy clusters with measured X-ray properties. It also serves as another demonstration of the usefulness of the XGA software, as it allowed for this sample to be put together with a minimum of effort, and for temperature and luminosity measurements to be retrieved within a few hours (running on 100 cores). The clusters in this sample are at redshifts between $0.05 \leq z \leq 1.91$, with the lack of redshift dependence of the SZ cluster signals allowing for the selection of some very high redshift clusters. They cover a mass range of between $1.53 \times 10^{14} M_\odot$ and $1.90 \times 10^{15} M_\odot$ (with masses drawn from the ‘M500cCal’ entry in the catalogue), a temperature range of between 0.18 and 14.08 keV, and a soft band luminosity range of between $2.17 \times 10^{43}$ and $4.18 \times 10^{45}$ erg s$^{-1}$. 
3.8 Properties of the DESY3RM-XCS sample

Unlike the ACTDR5-XMM sample, the DESY3RM-XCS sample has already been constructed and curated, as part of the continued XCS work in support of the Dark Energy Survey cluster science working group. The sample will be published in an upcoming DES paper on the X-ray properties of the DESY3 redMaPPer catalogue by Jeltema et al. (prep). The redMaPPer DES Y3 catalogue was cross-matched to the XMM Cluster Survey master source list, with a search performed for XCS extended source regions within 2 Mpc. This provides an initial list of cluster candidates with XMM observations for further investigation.

The next step of the process makes use of an XCS software tool called Object Classification Tools for Astronomy VisUaliSation (OCTAVIUS), which is a versatile, robust, platform that makes the ‘eyeballing’ process (visual inspection of candidates) as simple as possible. OCTAVIUS has been built to an industry standard by a member of the XCS Sussex team, is able to be deployed on any HPC and served through a comprehensive web portal to the user. Images from any wavelength of observation can be added to the visualisation pages (for the DESY3 work it was XMM and DES images), with options for zooming in, changing contrast, viewing with and without overlaid XCS regions or possible member galaxies as identified by redMaPPer. Then the user can classify the object with one of any number of flags that were configured when the classification task was setup, including flags for data problems so that problematic observations can easily be removed from future analyses. After this process a sample of 325 galaxy clusters remained.

Once a pure sample of galaxy clusters had been established, the XCS pipeline for measuring temperatures and luminosities (XCS Post Processing Pipeline or XCS3P, see Section 1.4.3) is applied. This both measures temperatures and luminosities for the DESY3 clusters, and provides an estimate of their $R_{500}$ derived during the iteration process from a radius-temperature scaling relation Arnaud et al. (2005). This process results in measurements of temperature and luminosity, if the data were good enough to get a measurement; this was the case for 293 clusters out of the original 325. Again a cut was made on the temperature percentage error (selecting only those with a percentage error of less than 25%) to produce a final sample of the highest quality, resulting in a final sample of 170 clusters. This sample is referred to as DESY3RM-XCS$\Delta T_{X<25}$.

As part of this work, temperatures and luminosities were re-measured for the DESY3RM-XCS$\Delta T_{X<25}$ sample using XGA. The reason for this is two-fold; a) while temperatures and luminosities measured by XCS3P and XGA within the same analysis aperture demonstrate excellent
agreement (see Figure 3.4), they are not identical, and b) it is necessary for XGA to be aware of the temperature within an aperture for the calculation of the conversion factor between emissivity and density, and measuring new values takes very little time.

In summary, the 170 clusters in the DESY3RM-XCS\(_{ΔT_x<25}\) sample cover a redshift range of 0.10 \(\leq z \leq 0.89\), have redMaPPer richnesses between 20.61 and 221.67, a temperature range of between 1.09 and 12.66 keV, and a soft band luminosity range of between \(4.02 \times 10^{42}\) and \(3.30 \times 10^{45}\) erg s\(^{-1}\).

### 3.9 Measuring masses for the ACTDR5-XCS and DESY3RM-XCS samples

Once global properties for both the ACTDR5-XMM\(_{ΔT_x<25}\) and DESY3RM-XCS\(_{ΔT_x<25}\) have been measured, the next step is to measure temperature and density profiles, and then derive hydrostatic mass profiles (as described in Section 3.4). The same settings, techniques, and models are used to generate the mass profiles. In summary PSF-corrected surface brightness profiles are generated and fit with double-beta profiles, which in turn are deprojected and converted into density profiles to be fit by a simplified Vikhlinin density profile. Then annular spectra are generated out to the supplied \(R_{500}\), with annular binning set to provide a minimum width of 30\('\) and at least 1500 ct per annulus, and the spectra are then fit with absorbed \texttt{apec} profiles to generate projected temperature profiles. These projected temperature profiles are deprojected to recover a three-dimensional temperature profile, which are then fit with simplified Vikhlinin temperature profiles. The density and temperature profiles for each galaxy cluster are used to set up a hydrostatic mass profile, and then masses are measured within the passed in \(R_{500}\).

Once complete this process results in hundreds of new galaxy cluster masses. Out of the 345 galaxy clusters in the ACTDR5-XMM\(_{ΔT_x<25}\) sample, 245 have a successfully measured hydrostatic mass. This represents the largest sample of galaxy cluster hydrostatic masses ever measured, over twice the size of a sample selected from Planck by Lovisari et al. (2020). Out of the 170 galaxy clusters in the DESY3RM-XCS\(_{ΔT_x<25}\), 87 have a successfully measured mass.
Figure 3.20: A figure showing two mass-temperature relations, one generated from ACT DR5 (purple) clusters, and one from DESY3RM (green) clusters. The DESY3RM galaxy clusters used to fit the green scaling relations are unique, in that they do not appear in either the SDSSRM-XCS or ACTDR5-XMM cluster samples. The shaded region of each relation indicates the 90% confidence limits.
Figure 3.21: A companion to Figure 3.20, this figure shows the parameter posterior distributions of the scaling relation model that was used to fit the mass-temperature relation.
Figure 3.22: A mass-richness relation for galaxy clusters in the DESY3RM-XCS sample what do not appear in either the SDSSRM-XCS or ACTDR5-XMM samples.
3.10 The largest combined sample of hydrostatic masses

This section briefly discusses the combination of the three galaxy cluster samples mentioned in this chapter for which new masses have measured, as well as how the masses for the ACT DR5 and DESY3RM sub-samples relate to other measured galaxy cluster observables.

3.10.1 Identifying catalogue overlaps between SDSSRM-XCS, DESY3RM-XCS, and ACTDR5-XMM

Combining the masses measured in Section 3.9 with the galaxy cluster hydrostatic masses measured for the SDSSRM-XCS sample in the earlier part of this chapter (see Section 3.5.1), a very large sample of directly measured masses is available to us. If the number of masses measured for the three separate samples are naively combined then the total number of masses is 442. However, due to the different surveys that the galaxy clusters were selected from there is some overlap between the samples. The DES and SDSS footprints overlap in the Stripe-82 region (see Figure 1.9 for an illustration of this cross-over), and the fact that some of the galaxy clusters in the ACT DR5 catalogue have richnesses from SDSS redMaPPer and/or DESY3 redMaPPer make it easy to infer that the ACT DR5 crosses over with the other samples.

A simple procedure was adopted to find which entries in the three samples are referring to the same galaxy clusters. This involved setting up a three-dimensional matching function, to find the catalogue entries that a) are coincident on the sky, and b) are within a certain matching distance (calculated from redshift) of one another in the direction of the line of sight. Both the DESY3RM-XCS and SDSSRM-XCS samples are matched to the ACT DR5 sample, with a matching sky separation of 500 kpc (converted to angular distance using the cluster redshift and a concordance cosmology), and a matching redshift separation of 0.05. This procedure resulted in a final sample of 334 unique, directly-measured, cluster masses; 245 from ACT DR5, 65 from SDSSRM-XCS, and 24 from DESY3RM-XCS accounting for overlap. The DES sample is considerably smaller simply because for a cluster to be included in the DES unique mass count, that cluster had to not be in the SDSS or ACT samples, whereas the SDSS clusters had to only not be in the ACT sample, and the ACT sample was unaltered. This is the largest sample of directly measured galaxy cluster masses for individual galaxy clusters (unlike weak lensing where stacking of the shear signal is required to attain high enough signal to noise for a mass measurement), and will be a powerful tool in understanding the evolution of galaxy cluster mass with other properties.
Figure 3.23: A preliminary scaling relation between X-ray hydrostatic mass (as measured by XGA using XMM data), and ACT DR5 SZ signal to noise, as published in the DR5 catalogue. No uncertainties for the signal to noise are available, so the scaling relation model was fit using only mass uncertainties.
Figure 3.24: A figure showing the 334 unique galaxy cluster masses for clusters from the SDSSRM-XCS, ACTDR5-XMM, and DESY3RM-XCS samples, and how they relate to the X-ray temperature measured within the $R_{500}$ region. No model is fitted because selection effects for the three samples will be different enough that it would not be valid (particularly between selection from optical/NIR for the SDSS and DES samples, and SZ for the ACT sample).
Figure 3.25: A figure showing the 334 unique galaxy cluster masses for clusters from the SDSSRM-XCS, ACTDR5-XMM, and DESY3RM-XCS samples, and how they relate to the X-ray soft-band (0.5-2.0 keV) luminosity measured within the $R_{500}$ region. No model is fitted because selection effects for the three samples will be different enough that it would not be valid (particularly between selection from optical/NIR for the SDSS and DES samples, and SZ for the ACT sample).
3.10.2 Mass observable relations for ACT DR5 and DESY3RM clusters

While a combined number of unique galaxy cluster masses has been discussed, the samples are not used to measure combined scaling relations due to the differing selection functions of the three samples; they would make naively combining and fitting overall scaling relations invalid. Scaling relations between mass and common observable can still be created for the individual samples however, as seen in Section 3.5.1 for the SDSSRM-XCS sample. All scaling relations in this section are fitted using the same LIRA setup as in Section 3.5.2. In this case the commonly used relations between mass and X-ray temperature/luminosity are created for the ACT DR5 sample and the unique cluster subset of the DESY3RM sample (when scaling relations are created for the Dark Energy Survey they will use every cluster in the sample, but here only clusters that do not appear in the SDSSRM or ACT DR5 samples are used). Figure 3.20 shows two mass temperature scaling relations plotted on the same axis with acceptable agreement, though as the unique DESY3RM subsample contains so few galaxy clusters the uncertainties on the corresponding scaling relation are very large. Figure 3.21 shows a companion corner plot to Figure 3.20, showing the posterior distributions of the scaling relation model; the contours for the ACTDR5-XMM samples are much better constrained than those of the DESY3RM-XCS sample due to the significant disparity in sample sizes between ACTDR5-XMM and the DESY3RM-XCS unique subsample. Figure 3.22 shows the relation between DESY3 masses and richness for the DESY3RM-XCS unique subsample. As far as we are aware this is only the second time a relation between X-ray hydrostatic mass and optical/near-infrared richness has been constructed, with the first time occurring earlier in the chapter for the SDSSRM-XCS sample. These relations between X-ray derived directly measured masses and the richnesses of the systems have the potential to be very valuable to current and upcoming cluster analyses, as with individual masses both the normalisation, slope, and scatter of the mass with observable can be probed. A scaling relation between hydrostatic masses measured for the ACTDR5-XMM clusters and their signal to noise is also created, shown in Figure 3.23. Finally, the entire unique set of 334 galaxy cluster hydrostatic masses are shown on the same axes (though without fitting a scaling relation model), showing both the mass-temperature plane (Section 3.24) and the mass-luminosity plane (Section 3.25).

3.11 What is there left to do?

While measurements of mass have been made for the ACT and DES selected samples, this work cannot yet be considered complete. This is because simply repeating the analysis performed in the
paper (starting at Section 3.1) that makes up the first part of this chapter will not provide as many new insights as expanding and improving upon the analyses. These samples can provide critical insights as to whether hydrostatic masses will be truly useful for future cluster studies, and what that their limits are. The plans in this section are in addition to the planned future work mentioned in Section 3.6.2 of the paper that makes up the first part of this chapter. Many of the plans in that section will be carried out using the ACT DR5 and DESY3RM samples before they are published.

### 3.11.1 Inspection of data

The masses presented in this part of the chapter are not yet part of a publication, and as such are preliminary. One of the major analysis steps that remains for these samples are the visual inspection of key data outputs, to check for generation and fit quality. The inspected products will include annular spectra, as well as surface brightness, density, temperature, and mass profiles; these will help ensure that the (eventually) published masses are valid and not the result of faulty measurements, without risking bias by only inspecting clusters that appear to be abnormal. This also includes a process of inspecting images; a) to identify flared or otherwise faulty observations, and b) to use visualisations with overlaid region information to ensure that all contaminant sources are being properly masked, including any infalling clusters that are not been properly removed by the XGA masks generated from XAPA region lists.

### 3.11.2 Measure new $R_\Delta$ values

Masses are commonly measured within a circular overdensity radius ($R_\Delta$). Overdensity radii are defined as the radius at which the density is $\Delta$ times the critical density of the Universe at the cluster redshift. All properties presented in this chapter have been measured within $R_\Delta$ estimates from an external source, that is to say that the published properties of each source have either directly provided $R_\Delta$ (e.g. SDSSRM-XCS, XXL-100-GC, LoCuSS High-$L_X$, and DESY3RM-XCS), or have been used to infer an overdensity radius (ACT DR5).

The mass profiles generated byXGA can be used to calculate the radius corresponding to any desired overdensity, when given the analysis cosmology and the redshift of the cluster. As there is no analytical solution to solving the hydrostatic mass density profile equation for $R$ we must minimise
\[ \delta \rho_{\Delta, \text{hy}} = \frac{3M_{\text{hy}}(R_{\Delta})}{4\pi R_{\Delta}^3} - \Delta \rho_c(z), \]  

where \( R_{\Delta} \) is the overdensity radius being calculated, \( M_{\text{hy}}(R_{\Delta}) \) is the hydrostatic mass within radius \( (R_{\Delta}) \), \( \Delta \) is the desired overdensity (2500, 500, and 200 are common choices), and \( \rho_c(z) \) is the critical density of the Universe at the cluster’s redshift. When an overdensity radius is being calculated, XGA initially measures \( \delta \rho_{\Delta, \text{hy}} \) for a range of radii between 100-3500 kpc in steps of 100 kpc. It can then find the two radii that bracket \( \delta \rho_{\Delta, \text{hy}} = 0 \), and between them test another range of radii (in steps of 1 kpc this time), finding the radius that corresponds to the minimum \( \delta \rho_{\Delta, \text{hy}} \) value.

### 3.11.3 Joint fitting density and temperature profiles to enforce physicality

While it is valid to separately generate and fit the density and temperature profiles necessary to measure a hydrostatic mass profile, there can be a distinct disadvantage. As the density and temperature profiles are treated completely separately, the validity of the hydrostatic mass profile that results from them is not known until after the entire process is complete. It is possible to measure a hydrostatic mass profile that is not monotonically increasing for instance, which considering the hydrostatic mass profile describes the amount of mass within each given radius, does not make physical sense; a mass at radius \( R_{j+1} \) cannot be larger than at \( R_j \). This can occur if one of the profiles has not been constructed entirely successfully or the profile fit has not worked well. Several methods are being developed for XGA to both flag suspicious profiles for human inspection, and to use information from both profiles simultaneously to inform the fitting of profile models.

Firstly, an automated method of flagging unusual temperature profiles is being developed to take advantage of XGA’s model classes (see Section 2.4.11.1). Specifically, the built in derivative calculation method for those profile models. In cases where there is an analytical solution to the first order derivative, then the model will have been implemented with that solution, but in all other cases a numerical differentiation method is used. The simplified Vikhlinin temperature model for instance, does have an analytical first derivative,
\[
\frac{dT(R)}{dR} = \left[ \frac{R^2 + R_T^2}{R_T^2} \right]^{-\frac{7}{2}} \left( a_{cool} T_0 \right) \left( \frac{R}{R_{cool}} \right)^{a_{cool}} \left( \frac{R^2 + R_T^2}{R_T^2} \right) \left( \left( \frac{R}{R_{cool}} \right)^{a_{cool}} + 1 \right) - a_{cool} \\
\left( \frac{R}{R_{cool}} \right)^{a_{cool}} \left( R^2 + R_T^2 \right) \left( T_0 \left( \frac{R}{R_{cool}} \right)^{a_{cool}} + T_{min} \right) - cR^2 \left( \left( \frac{R}{R_{cool}} \right)^{a_{cool}} + 1 \right) \\
\left( T_0 \left( \frac{R}{R_{cool}} \right)^{a_{cool}} + T_{min} \right) \right] \div \left[ \left( \frac{R}{R_{cool}} \right)^{a_{cool}} + 1 \right]^2,
\]

(3.11)

where all parameter values have meanings as defined in Section 2.4.11.1. This flagging method will take a fitted profile model (so is only useful after a fit has been performed) and perform derivatives at several radii that should be in the outskirts of the cluster. If the first derivative is positive in those locations (i.e., the temperature is getting hotter) then it is possible to infer that there is either a problem with the generation/fitting of that profile, or that perhaps there is an infalling system causing a temperature increase in measurements of the outskirts.

A similar logic can also be applied to the gas density profile, and the simplified Vikhlinin density profile also has an analytical first derivative,

\[
\frac{\rho_g(R)}{dR} = -\rho_0 \sqrt{\left( \frac{R}{R_c} \right)^{-\alpha} \left( \frac{R^2 + R^3_c}{R^2} \right)^{\frac{3}{2} - \beta} \left( \frac{R^3 + R^3_c}{R^2} \right)^{-\frac{5}{2}} \left( \alpha R^3 R^2_c + \alpha R^2 R^3_s + 6 \beta R^5 + 6 \beta R^2 R^3_s + \epsilon R^5 + \epsilon R^3 R^3_c \right)} \\
2R \left( R^5 + R^3 R^2_c + R^2 R^3_s + R^2 R^3_s \right)
\]

(3.12)

with the parameters again having the meanings defined for this model in Section 2.4.11.1.

This leads into a method under development that would not be used to flag problematic profiles after the fact, but actually enforce certain physical requirements on the final output mass profile during the fitting process of the input density and temperature profiles. The density and temperature profiles will be fit jointly, with each step in the MCMC chain used to not only calculate updated model instances and find the difference between them and the data, but also to quickly calculate a mass profile (and mass profile derivatives); these can then be used to ensure that the mass profile constructed from the current temperature and density profile models is strictly monotonically increasing, and isn’t that the masses calculated are within broad physical bounds. If the profile does not satisfy these conditions then that MCMC step is rejected. Though this would be possible without an analytical expression for the mass profile, this idea was inspired by the fact that there are analytical derivatives for the simplified Vikhlinin temperature (Equation 3.11) and density profiles (Equation 3.12), which in turn means that an analytical expression for the hydrostatic mass profile can be derived,
where this model is denoted the simplified Vikhlinin hydrostatic mass profile, and the parameters are as described for the simplified Vikhlinin density and temperature models in Section 2.4.11.1. It is also possible to derive an analytical first derivative of this mass model, but it is too large an expression to present here; the first derivative will be useful in ensuring that the mass model is strictly monotonically increasing. These analytical expressions will make it more convenient to construct mass profiles in XGA, resulting in a smooth mass distribution rather than the samples which are currently drawn at certain radii to create the data points of the HydrostaticMass profile. There is also the potential for setting up a model of the dark matter mass density profile for a cluster, given that we probe both the ‘total’ mass and the baryonic mass of the cluster with this technique.
Chapter 4: Comparison of eROSITA and XMM cluster measurements

This chapter consists of a paper that is on arXiv (Turner et al., 2021), and has been submitted to MNRAS (now pending the second set of reviewer comments). As this paper has been included in the same format it has been made public in, there may be some overlap in content between this and other chapters.

4.1 Introduction

X-ray observations of clusters of galaxies provide insights into various aspects of astrophysics (e.g., Hitomi Collaboration et al., 2016; Bhargava et al., 2020; Sanders et al., 2021b) and cosmology (e.g., Vikhlinin et al., 2009; Schellenberger and Reiprich, 2017). Clusters are among the largest gravitationally bound structures in the Universe and consist of a dark matter halo, the intra-cluster medium (ICM), and the component galaxies. The ICM is a high-temperature, low-density plasma that emits strongly in the X-ray band, with both continuum and emission line components.

The eROSITA instrument mounted on the joint Russian-German Spectrum-Roentgen-Gamma (SRG, Predehl et al., 2021) mission will contribute significantly to X-ray cluster astrophysics and cosmology. Its large field of view (~1 deg), sensitivity, and energy resolution combine to make it a revolutionary new instrument. The final eROSITA All-Sky Survey (eRASS) is predicted to detect approximately 100,000 galaxy clusters above a mass of $5 \times 10^{13} h^{-1} M_\odot$ (Pillepich et al., 2012).

The data sharing agreement between the German and Russian consortions that funded eROSITA involves sharing the sky equally. Most of these clusters will be accompanied by an X-ray luminosity ($L_X$) measurement, and roughly 20% (Liu et al., 2021a) of the observations will yield an X-ray temperature ($T_X$) measurement. However, apart from a handful of the highest flux clusters, it will not be possible to measure masses via the hydrostatic technique directly from eRASS data.
fore, until the all sky survey is complete, it will be necessary to supplement the eRASS cluster catalogue with mass measurements from the current generation of X-ray telescopes (i.e. XMM, Chandra) in order to maximise the scientific yield. After the all sky survey is complete eROSITA pointed observations of clusters, will produce some hydrostatic mass estimates, as demonstrated Sanders et al. (2021a).

The aim of this paper is to explore potential synergies between eRASS cluster catalogues and the data in the XMM-Newton public archive, and to probe calibration considerations required for such analyses. The eROSITA and XMM telescopes have different characteristics that allow them to complement one another, some of which (such as the effective area at different energies, and the background level) were explored by Predehl et al. (2021). Comparisons between the on-axis effective areas of the combined XMM cameras (PN, MOS1, and MOS2) and the combined eROSITA telescope modules show that the effective areas are effectively equal between $\sim 0.5$-$2.0$ keV, though outside this range XMM has an advantage. Here XMM complements eROSITA in that it will observe more source emission at higher energies, which could improve constraints on spectroscopic X-ray measurements of temperature and luminosity. The larger field of view of eROSITA ensures that its grasp (the product of effective area and observing area) is significantly greater than XMM’s below $\sim 3.5$ keV, though above that energy XMM’s grasp is greater. Comparisons between the background levels of a subset of eROSITA’s telescope modules and the XMM-Newton cameras using a simultaneous observation of NLS1 1H0707-495 (Boller et al., 2021) revealed that, although the eROSITA background is higher than pre-launch predictions, it is generally lower than XMM and more temporally stable. Soft-proton flaring does not significantly impact eROSITA, giving it an advantage over XMM in this regard, and possibly allowing it to locate more low-flux sources (such as low surface-brightness galaxy clusters).

In this work we will make use of the recent release of the eROSITA Final Equatorial-Depth Survey (eFEDS, Brunner et al., 2021). The eFEDS field covers approximately 140 square degrees of the equatorial ($-2.5^\circ < \delta < 6.0^\circ$) sky. It intersects with several optical/near-IR photometric and/or spectroscopic surveys, including the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP, Aihara et al., 2018), the Galaxy and Mass Assembly survey (GAMA, Driver et al., 2011), and the Sloan Digital Sky Survey (SDSS, Blanton et al., 2017). We make an indirect comparison to a similar X-ray survey using the XXL-100-GC catalogue (Pacaud et al., 2016), then make direct comparisons to XCS measurements.

We wish to ascertain the level of sample contamination in the eFEDS cluster catalogue, compare the central coordinates of the detected clusters to those measured by the XMM Cluster Survey
(XCS, Romer et al., 1999), and verify the accuracy of $L_X$ and $T_X$ measurements. As eFEDS is the same depth as the final eRASS, the accuracy of these measurements have implications for cosmological studies based on eRASS cluster detection (using weak lensing masses and X-ray luminosities for the mass observable relation). They will also impact studies based on optical or near-infrared detection (as luminosities can be used to explore scatter in the mass observable relations), and astrophysical studies of cluster luminosity-temperature relations to study the evolution of the intra-cluster medium.

There is a known difference between the galaxy cluster temperatures measured by XMM and Chandra. This difference has been quantified with functions to calibrate the temperatures of one telescope to another; Schellenberger et al. (2015) showed that the difference increases with temperature, with XMM EPIC temperatures being on average 7% and 23% lower than Chandra ACIS temperatures for 2 keV and 10 keV clusters respectively. Possible mechanisms for the discrepancy include instrument specific calibration errors (i.e. uncertainty on calculated effective areas), or fitting single temperature plasma models to multi-temperature plasma emission. Schellenberger et al. (2015) demonstrated that multi-phase ICM with extreme temperature differences can cause an overall temperature to be dependant on the instrument response. However, they concluded that effective area calibration uncertainties in the soft energy band (0.7–2) keV caused the observed differences in temperature between XMM and Chandra. Any analysis which uses both XMM and Chandra temperatures typically accounts for this (e.g., Farahi et al., 2019b; Migkas et al., 2020). An understanding of whether there is a similar difference in eROSITA and XMM temperatures will be necessary before any joint analyses with data from the two telescopes are undertaken, and before scaling relations from one telescope can be safely used by another.

In §4.2 we explore the general properties of the eFEDS cluster catalogue and provide comparisons to a catalogue with similar properties. In §4.3 we construct a cluster sample from the eFEDS catalogue with corresponding XMM observations, which includes a visual inspection of the X-ray data and SDSS/HSC images. We also compare eFEDS and XCS exposure times and central positions. In §4.4 we compare luminosities and temperatures measured by eFEDS and XCS. Finally, in §4.5 we generate luminosity-temperature relations, discuss implications of our findings and how they can be improved. Then in §4.6 we provide a final summary. The analysis code and samples are available in a GitHub repository\textsuperscript{1}.

Throughout this work we use a concordance $\Lambda$CDM cosmology where $\Omega_M=0.3$, $\Omega_\Lambda=0.7$, and $H_0=70$ km s$^{-1}$ Mpc$^{-1}$, consistent with the original eFEDS cluster analysis (and other XCS works).

\textsuperscript{1}Analysis code and samples
4.2 Comparison of the eFEDS Optically Confirmed and XXL-100-GC catalogues

The eFEDS cluster catalogue (Liu et al., 2021a) contains 542 candidates, 477 of which are considered to be optically confirmed (Klein et al., 2021) when assessed using the Multi-component Matched Filter Cluster Confirmation Tool (MCMF, Klein et al., 2018). All 542 X-ray candidates are accompanied by redshift ($z$) values, measured from galaxy photometric redshift distributions as part of the MCMF optical confirmation process. Soft-band (0.5-2.0 keV in the source frame) $L_X$ values have been measured for 91% of the X-ray cluster candidate sample. A smaller percentage, 21%, of $T_X$ values were obtained using spectra extracted from circular apertures centered on the eFEDS coordinates; 69 within 300 kpc, and 95 within 500 kpc (102 candidates have at least one temperature measurement).

The XXL survey (Pierre et al., 2016) covers $\sim$50 deg$^2$ of the sky (over two separate regions), making it the largest contiguous area survey in the XMM archive. It consists of 542 separate XMM observations with on-axis exposure times ranging from 10-20 ks. The contiguous nature of XXL makes it ideal to compare to eFEDS. As eFEDS and the XXL survey were taken in different parts of the sky there are no clusters in common. Although we note that the X-CLASS analysis of the XMM archive (up to August 2015) that made use of the XXL pipelines does contain some eFEDS candidates (Koulouridis et al., 2021). Comparisons are limited to ensemble distributions of the cluster samples. We made use of the XXL-100-GC sample (Pacaud et al., 2016), containing
Table 1: Summary of the samples defined in this work. $N_{cl}$ is the number of clusters, $N_{TeFEDS,500kpc}$ the number with eFEDS $T_{500kpc}$ values, and $N_{TXCS,500kpc}$ the number with XCS $T_{500kpc}$ values.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Description</th>
<th>$N_{cl}$</th>
<th>$N_{TeFEDS,500kpc}$</th>
<th>$N_{TXCS,500kpc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>eFEDS</td>
<td>The full eFEDS cluster candidates catalogue</td>
<td>542</td>
<td>95</td>
<td>-</td>
</tr>
<tr>
<td>eFEDS-XMM</td>
<td>eFEDS cluster candidates that fall on an XMM observation</td>
<td>62</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>eFEDS-XCS</td>
<td>eFEDS-XMM candidates available for analysis after inspection</td>
<td>37</td>
<td>8</td>
<td>28</td>
</tr>
</tbody>
</table>

the 100 brightest galaxy clusters observed in XXL, and the sample of 477 optically confirmed eFEDS candidates. The flux limits of the eFEDS and XXL-100-GC cluster samples are similar; $\sim 10^{-14}$ erg s$^{-1}$ cm$^{-2}$.

Figure 4.1a shows the redshift distributions of the clusters in the two samples (eFEDS and XXL-100-GC distributions are shown in red and cyan respectively) to be very similar overall, but that XXL-100-GC detects a higher proportion of clusters at low redshifts. Next, we compare the respective temperature distributions. Temperatures for the XXL-100-GC clusters were measured within a 300 kpc aperture (Giles et al., 2016), as were temperatures for eFEDS clusters, making a direct comparison of the distributions valid. Figure 4.1b plots the temperature distributions of the two samples, with the XXL-100-GC temperature distribution containing a significantly higher proportion of temperatures above $\sim 3.5$ keV. Liu et al. (2021a) note that eROSITA’s ability to measure temperatures for hot clusters at $\gtrsim 5$ keV is limited due to the reduced sensitivity of eROSITA at energies $> 3$ keV. This is a plausible reason for the increased number of higher temperature clusters in XXL-100-GC compared to eFEDS. The effective area of eROSITA is $\sim 150$ cm$^2$ at 5 keV, compared to $\sim 900$ cm$^2$ for EPIC-PN, see Figure 9 in Predehl et al. (2021) for a detailed comparison. Previous work by Lloyd-Davies et al. (2011) has also shown that more counts are required to constrain temperatures to the same level for hotter galaxy clusters. Furthermore, the temperature distribution could also be influenced by the selection functions of the two surveys, differing measurement methodology, or a systematic difference in temperatures measured by the eROSITA and XMM telescopes (we explore this in Section 4.4.4).

Finally we compare how well temperatures from the two samples are constrained, by comparing temperature uncertainties as a fraction of the absolute temperature value ($\Delta T_X / T_X$). Figure 4.1c shows that, on average, XXL achieves better temperature constraints than eFEDS, with the mean percentage uncertainties for XXL and eFEDS being 14% and 25% respectively. This is consistent with the findings of Lloyd-Davies et al. (2011), who showed that $\sim 1000$ background-subtracted soft-band (0.5-2.0 keV) counts are required to achieve a fractional temperature uncertainty of $\sim 0.1$
Figure 4.2: Footprint of eFEDS, given by the black solid line. Cluster candidates present in
the eFEDS X-ray catalogue are highlighted by red diamonds. The grey circles highlight XMM
observations, with a radius of 15′ (the approximate radius of the XMM FoV).

for a 3 keV cluster. In this regard the longer exposures of XXL compared to eFEDS would give
an advantage (especially in the deeper XMM-LSS fields, covering ∼11 deg² in the XXL-N field).

4.3 Understanding the eFEDS catalogue contamination frac-
tion

In this section, we make use of archival XMM data that overlaps with the eFEDS footprint to as-
semble samples for analysis, and make an estimate of the contamination fraction in the eFEDS
cluster candidate list. For this, we have used data products (images and source lists) generated
by the XMM Cluster Survey (XCS, Romer et al., 1999). The XCS source lists are constructed
by the XCS Automated Pipeline Algorithm (XAPA) source finder, and a full explanation of our
procedures can be found in Lloyd-Davies et al. (2011). We first determine which eFEDS X-ray
cluster candidates fall within the active area of an XMM observation (Section 4.3.1). For these, we
generate eROSITA cut-out images. We then compare, by eye, to the corresponding XMM cut-outs
(Section 4.3.2). We also compare to optical SDSS DR16 (Ahumada et al., 2020) images obtained
from the SDSS cutout server². As ∼70% of the eFEDS-XMM sample are at redshifts z<0.5 it
is generally appropriate to search for a red sequence using SDSS imagery, however for any can-

²SDSS Image Cutout Server
didate that we could not confirm with SDSS, we then examined images taken from the second data release of the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP PDR2, Aihara et al., 2019)\(^3\). The deeper data of HSC-SSP PDR2 (\(i\)-band limiting magnitude of 26.2 in the wide field, where the SDSS DR16 \(i\)-band limiting magnitude is 22.2) allow for detection of cluster galaxies at much higher redshifts. Our visual inspection allows us to categorise contaminating objects (Section 4.3.3) and to estimate the overall contamination fraction in the eFEDS sample (Section 4.3.5).

### 4.3.1 eFEDS Cluster Candidates in the XMM Footprint

Figure 4.2 shows the outline of the eFEDS footprint with eFEDS X-ray cluster candidates (Liu et al., 2021a) indicated by red diamonds. \(XMM\) observations taken within the eFEDS footprint are indicated by grey shaded circles, with a radius of 15\(^\prime\) (the approximate radius of the \(XMM\) field-of-view). There are a total of 143 \(XMM\) observations, covering \(\sim15\) deg\(^2\) (11\%) of the sky within the eFEDS footprint, accounting for overlapping \(XMM\) observations.

We used \(XMM\): Generate and Analyse (XGA\(^4\) Turner et al., 2022); a new, open-source, X-ray astronomy analysis module developed by XCS, to determine which of the 542 eFEDS cluster candidates listed in Liu et al. (2021a) have also been observed by \(XMM\). An initial search finds eFEDS candidates with central coordinates within 30\(^\prime\) of an \(XMM\) observation aimpoint (this is larger than the \(XMM\) field of view to account for any cases of low-\(z\) clusters with centroid offsets). We then refined the match so that at least 70\% of a 300 kpc aperture (centred on the eFEDS coordinate and assuming the eFEDS redshift) coincides with an \(XMM\) observation. Sixty-two eFEDS candidates met these criteria, and this sub-set is denoted the eFEDS-\(XMM\) sample (see Table 1). Fifty-three of the eFEDS-\(XMM\) candidates appear in the eFEDS optically confirmed sample (Klein et al., 2021).

The distribution of \(XMM\) exposure times for the eFEDS-\(XMM\) sample is shown in Figure 4.3. The light-blue distribution uses the best individual observation exposure time for each eFEDS-\(XMM\) candidate; the grey distribution is the total exposure time for each candidate. These are vignetting-corrected exposure times at the eFEDS coordinate (rather than at the respective observation aim-point). The typical eFEDS vignetting-corrected exposure (1.2 ks) is shown by the dashed red line for comparison. The majority of exposure times (individual or total) are longer for \(XMM\) than eFEDS.

\(^3\)HSC-SSP PDR2 Image Cutout Server

\(^4\)XMM: Generate and Analyse GitHub
Figure 4.3: Distribution of exposure times for eFEDS-\textit{XMM} cluster candidates, measured at the eFEDS coordinates. Exposures taken from 0.5-2.0 keV exposure maps, corrected for flaring and vignetting. Dashed line indicates the average vignetting corrected exposure of the eFEDS field reported by Liu et al. (2021a).
Figure 4.4: Comparison of eFEDS and XAPA central coordinates, for the subset of the eFEDS-XCS sample that have been detected by XAPA.
Figure 4.5: eFEDS-XMM cluster candidate (eFEDS ID 1644) identified as a pair interacting galaxies with ongoing AGN activity (see Section 4.3.3.1). The cross-hair indicates the eFEDS position. Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0822470101), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 750 kpc, HSC image has a half-side-length of 750 kpc (at the redshift provided by eFEDS).

Figure 4.6: eFEDS-XMM cluster candidate (eFEDS ID 3334) without an obvious corresponding source of emission (see Section 4.3.3.2). The cross-hair indicates the eFEDS position. Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0822470101), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 500 kpc, HSC image has a smaller half-side-length of 250 kpc (at the redshift provided by eFEDS).
Figure 4.7: Two eFEDS-XMM cluster candidates in the outskirts of a low redshift foreground AGN. A spurious eFEDS-XMM cluster candidate (eFEDS ID 8922) is indicated by the cross-hair (see Section 4.3.3.2). An eFEDS-XMM cluster candidate (eFEDS ID 16370) is indicated by the dashed circle. Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0655340133), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 1000 kpc, HSC image has a half-side-length of 1000 kpc (at the redshift for eFEDS ID 8922 provided by eFEDS).

Figure 4.8: eFEDS galaxy cluster split into two candidates by the source finder (see Section 4.3.3.3). Cross-hairs indicate one candidate (eFEDS ID 8602), and the white diamond indicates the other (eFEDS ID 1023). Left hand side is a combined PN+MOS1+MOS2 XMM image (ObsID 0761730501), centre is eROSITA, right hand side is HSC. Both XMM and eROSITA images are cutouts within a radius of 800 kpc, HSC image has a half-side-length of 800 kpc (at the redshift provided by eFEDS, which is the same for 8602 and 1023).
4.3.2 Constructing the eFEDS-XCS Sample

To judge the quality of the eFEDS candidates in the eFEDS-XMM sample, circular cut-out images, of radius 500 kpc, were generated from the eROSITA eFEDS data and XCS processed XMM data. We select the XMM observation with the highest signal-to-noise within a 300 kpc aperture centered on the eFEDS candidate coordinate, and use its XCS generated PN+MOS1+MOS2 image (Giles et al., 2022b). For the eROSITA cut-outs, the eSASS5 (Brunner et al., 2018) evtool software was used.

Both sets of images used a pixel size of 4.35″, but different energy ranges were used; 0.5-2.0 keV for XMM and 0.2-2.3 keV for eROSITA. These ranges reflect those used by the respective XCS and eFEDS source detection routines.

As shown in Predehl et al. (2021), the energy dependent effective area of XMM (assuming all three cameras are operating) is the same, or greater, than that of eROSITA. Therefore, we can expect most of the XMM cut-outs to have higher signal to noise than their eFEDS counterparts, even after accounting for the fact that the XMM background level is slightly higher than eROSITA’s. In 10 cases, however, we judged the XMM data to be inadequate for further analysis. This was either because the eFEDS candidate fell on the edge of the field of view and/or because the signal-to-noise was too low (see Table 1 for details). These 10 were excluded from further analysis, although it is noteworthy that in 5 cases, an obvious (by eye) extended source was visible in the

---

5Introduction to eSASS
corresponding eROSITA cut-out.

The remaining 52 eFEDS-XMM sources were then visually inspected side-by-side to judge the quality of the eFEDS candidates. Thirty-seven eFEDS-XMM candidates were confirmed as clusters suitable for X-ray analysis by this visual inspection, henceforth called the eFEDS-XCS sample. Generally speaking the XCS and eFEDS defined centroid positions were in good agreement (with an offset of less than 100 kpc), but several outliers are present (see Figure 4.4). The outliers were due to either low signal to noise eFEDS data, or to eFEDS point source contamination. Similar examples were noted in Klein et al. (2021).

The other eFEDS cluster candidates were classified as sample contaminants (11, Table 2), or as having their X-ray flux contaminated by other sources (4, Table 3). There are three broad categories of sample contamination, as described below (Section 4.3.3).

**4.3.3 Categories of contaminating objects in the eFEDS X-ray cluster candidate catalogue**

Here we discuss categories for the different types of contaminating object that we discovered in the eFEDS X-ray cluster candidate catalogue. The figures that we use to illustrate these examples are not necessarily on the same scale, or centered on the same position, as those we used for visual inspection, and all figures use HSC imagery for clarity. The optical images we used for general inspection were SDSS, with HSC photometry used in cases were we needed to clarify our classification.

**4.3.3.1 Blended sources**

An example of this is eFEDS ID 1644, which is shown in Figure 4.5; the two sources at the centre of the XMM image (Figure 4.5 left; detected as separate point sources by XCS) in the XMM cut-out appear as a single object in the eROSITA image. The dominant X-ray source is discussed in Pfeifle et al. (2019). It is the result of AGN activity in a pair of interacting galaxies. The two galaxies can be seen in the corresponding HSC image in Figure 4.5 (right). The blending is likely a result of eROSITA’s 26” FOV average PSF half-energy width (HEW), which is larger than the XMM PN camera’s 16.5” PSF HEW (at 1.5 keV), in combination with the short eFEDS exposure time. This source was assigned a class of B2 during the MCMF classification process, which indicates point source contamination, but it is still retained in the optically confirmed catalogue.
4.3.3.2 Spurious sources

An example of this is shown in Figure 4.6. There does not appear to be a source at the eFEDS candidate centroid position (extended or otherwise) in either the XMM or eROSITA image (Figure 4.6, left and middle respectively). We note that the redshift provided by the eFEDS cluster catalogue for this candidate (eFEDS ID 3334) is very low (\(z=0.087\)) and so any X-ray cluster emission should be obvious, unless it is very low surface brightness and/or very extended. Nonetheless, we do not see evidence of a coincident population of galaxies in the HSC imagery (Figure 4.6, right).

Another example is eFEDS ID 8922, which is shown in Figure 4.7; this cutout is not centered on the spurious eFEDS cluster candidate, but on the bright source that causes it. The eFEDS candidate location (at the cross hairs) is in the outskirts of a bright source (eFEDS ID 3, not present in the cluster candidate catalogue), and not coincident with any distinct source in either the XMM or eROSITA image (Figure 4.7, left and middle respectively). There also does not appear to be an association of galaxies in HSC imagery (Figure 4.7, right). The dominant X-ray source is identified as an AGN in the Million Quasar catalogue (Flesch, 2021) located in a spiral galaxy visible in the corresponding optical image. It is also present in the eFEDS AGN catalogue (Liu et al., 2021b).

4.3.3.3 Fragmented sources

An example of this is shown in Figure 4.8. The white cross-hair indicating the position of one eFEDS candidate (eFEDS ID 8602), and the white diamond another (eFEDS ID 1023). The two candidates have almost identical redshifts (\(z=0.196\) and \(z=0.197\) respectively). Luminosity measurements for both, and a temperature estimate for eFEDS ID 1023, are given in Liu et al. (2021a). We discuss this system further in Appendix B.2. We note that it is not used during our luminosity (Section 4.4.2) and temperature comparisons (Section 4.4.3 and Section 4.4.4), or in our luminosity-temperature relation analysis (Section 4.5.1).

4.3.4 Clusters with contaminated X-ray emission

In these cases there is evidence, from the SDSS and/or HSC data, for a physical association of galaxies – which could in turn be responsible for an extended X-ray source due to emission from a hot ICM – however, we contend that any ICM emission present is significantly contaminated by other X-ray sources. One example (eFEDS ID 150) is shown in Figure 4.9, where the emission
detected by eROSITA appears to originate primarily from the central galaxy (alternative, but less likely explanations are that this is a fossil group or a system with a strong cool core). We note that similar examples were identified in eFEDS, candidates with IDs 3133 and 3008 (see Table 3). An example of a different type of contaminated emission is presented in Figure 4.7 (eFEDS ID 16370). The eFEDS candidate is highlighted by the white dashed circle. There is tentative evidence of X-ray emission in the XMM observation (especially when smoothing is applied), and the coordinates coincide with a collection of red galaxies in SDSS and HSC at an SDSS photo-$z$ of $z=0.44$ (matching the eFEDS catalogue’s $z$). However, due to the proximity of the eFEDS candidate to the bright AGN (as discussed in Section 4.3.3.2), the X-ray flux in this region will be contaminated by non ICM emission, so we exclude this cluster from the following analyses.

In this paper we focus on the comparison of the X-ray properties measured by eROSITA and XMM, so we do not include these four eFEDS candidates in the analyses presented Section 4.4. However, it would inappropriate to remove them from some other types of analyses – such as cluster number count cosmology based on optical/near-IR selection Klein et al. (2021) – because they are still associated with galaxy over-densities. All of the sources in Table 3 appear in the eFEDS optically confirmed sample.

**4.3.5 The eFEDS contamination fraction**

As discussed in Section 4.3.2 (and collated in Appendix B.1), 11 of the 62 candidates (18%, Table 2) in the eFEDS-XMM sample were not included in the eFEDS-XCS sample because they were classified as being in one of the three sample contaminant types described in Section 4.3.3. This should be viewed as a lower limit because, in 10 (of 62) cases (Table 1), it was not possible for us to confirm the validity of the eFEDS candidate using archival XMM data.

The eFEDS-XCS sample of eFEDS cluster candidates is an order of magnitude smaller than the full eFEDS X-ray cluster candidate catalogue. Moreover, several of the eFEDS-XCS clusters were the target of their respective XMM observations, and that has been shown to introduce selection bias (Giles et al., 2022b); this could influence XMM detections and thus the construction of our eFEDS-XCS sample. Even so, our result is consistent with simulations performed by the eRASS team that predicted a contamination level of $\sim20\%$ (Comparat et al., 2020). It is also consistent with the eFEDS optical counterparts study (Klein et al., 2021), which measured a contamination fraction of $17\pm3\%$.

We also investigated whether any of the 11 cluster candidates that we classed as sample contam-
inants (Table 2) are present in the sample of 477 candidates that Klein et al. (2021) consider to be optically confirmed. We find that 5 of the 11 are present therein. This compares to 53 in the overall eFEDS-XMM sample, indicating a minimum contamination fraction of \( \sim 9\% \) in the Klein et al. (2021) sample. This is slightly high compared to the value of \( 6 \pm 3\% \) reported by Klein et al. (2021). However, if we discount eFEDS ID 8602 and 3334 from consideration as sample contaminants (to be more consistent with approach taken in Klein et al., 2021), the contamination level drops to \( \sim 6\% \).

We note that one cluster of the eFEDS-XCS sample (eFEDS ID 5170) does not appear in the Klein et al. (2021) sample. This candidate was included in our sample because of its X-ray emission (in XMM and eROSITA images) and evidence of an over density of red galaxies in the SDSS and HSC photometry.

### 4.4 Comparisons of cluster properties measured by eFEDS and XCS

We use the XGA (Turner et al., 2022) SAS (Gabriel et al., 2004) and XSPEC (Arnaud, 1996) interfaces to generate spectra and fit models to them, for those clusters that have high enough quality XMM data. We then compare values to those presented in the eFEDS data release. Note that we do not re-analyse the eFEDS data, but compare to the measurements given by Liu et al. (2021a). We use XGA v0.2.1, SAS v17.0.0, and XSPEC v12.10.1.

#### 4.4.1 Fitting Procedure

Cluster spectra are extracted within a 500 kpc fixed aperture (as the eFEDS catalogue contains a greater number of 500 kpc temperatures than 300 kpc) and centered on the eFEDS position, and corresponding backgrounds are extracted within 1000-1500 kpc annuli. The use of fixed aperture radii risks the inclusion of differing levels of cluster emission in background regions, which alters the measured temperature and luminosity, but they are necessary to make our work comparable with the eFEDS catalogue. Non-cluster sources in both the 500 kpc apertures and background regions are identified using the XCS region files, and their corresponding events are removed during spectrum generation. The SAS evselect tool is used to generate spectra from all available XMM data for an eFEDS-XCS cluster (EPIC-PN, EPIC-MOS1, and EPIC-MOS2 cameras are all used; see Table 4 for the data used for each cluster). Spectra are re-binned using
Figure 4.10: Comparison of unabsorbed cluster luminosities within a 500 kpc aperture, in the 0.5-2.0 keV energy band, centered on eFEDS coordinates. Pale blue line indicates best fit power-law, with 68% confidence levels given by shaded region. Grey line indicates a power-law fit with slope set to 1 (with 68% confidence levels given by grey shaded region). Cyan diamond is for the split cluster discussed in Appendix B.2.
Figure 4.11: Temperature and fractional temperature error distributions of the eFEDS (red) and eFEDS-XCS (pale blue) samples, for measurements made within 500 kpc apertures, centered on eFEDS coordinates. The eFEDS sample plotted in red contains 95 eROSITA temperature measurements, and the eFEDS-XCS sample plotted in pale blue contains 28 XMM temperature measurements.

We fit absorbed (with \texttt{tbabs}, Wilms et al., 2000) plasma emission models (APEC, Smith et al., 2001) to the spectra; these models are standard for XCS analyses, but are also the same as those used in the eFEDS spectroscopic analysis. To maximise the similarity of our analysis to eFEDS we opt to use the abundance tables published by Asplund et al. (2009) when performing our spectral fits. The abundance parameter of the APEC model in all cases is frozen at 0.3 $Z_\odot$, the $nH$ parameter of the \texttt{tbabs} model is set from the full-sky HI survey by the HI4PI Collaboration et al. (2016) using the HEASoft \texttt{nh} tool and frozen. The redshift parameter is set to the eFEDS catalogue value and frozen. The temperature is initially set to 3 keV and the normalisation is initially set to 1 cm$^{-5}$, then both are allowed to vary.

Each individual spectrum (each instrument for each XMM observation) is first fit independently, and if the measured temperature is outside of the range $0.01 \text{ keV} \leq T_X < 20 \text{ keV}$, or either temperature uncertainty is $> 15 \text{ keV}$, then the spectrum will not be included in the final fit. A simultaneous fit is then performed using only the spectra that fulfil the requirements outlined.
above. A multiplicative constant is added to the model for the simultaneous fit, and is allowed to vary independently for each spectrum in the fit to account for different instrumental responses, whereas every other model parameter is tied together. Temperatures and unabsorbed luminosities are then determined from this joint fit. We used a given temperature measurement in further analyses (Sections 4.4.3, 4.4.4, and 4.5.1) if, a) the best fit value is less than 25 keV, b) the upper and lower uncertainties are both positive, and c) the larger uncertainty is less than three times the smaller. Likewise, fitted luminosities are used in our analyses (Sections 4.4.2 and 4.5.1), if both the upper and lower uncertainties are not greater than the best fit value, and if the upper and lower uncertainties are both positive.

For a more complete explanation of the spectral fitting process and comparisons of results with other XMM analyses that confirm the veracity of measurements produced by this procedure, see Turner et al. (prepa). All XMM measurements for the eFEDS-XCS sample can be found in Table 5, along with eFEDS ID, position, and redshift.

4.4.2 Luminosity Comparison

We compare luminosities measured with both XMM and eROSITA, since one of the main products of eRASS will be large catalogues of X-ray cluster luminosities. These will be used as the basis of various eROSITA science applications; for example, a mass-luminosity scaling relation (such as the one recently produced by Chiu et al., 2021) provides a way to estimate masses and overdensity radii of a given cluster, enabling X-ray cluster cosmology. Therefore, it is important to test the fidelity of eFEDS luminosities with XMM data.

The eFEDS analysis presents cluster luminosities measured via a forward-fitting analysis of 2D count-rate maps, including considerations of the morphology of the cluster, rather than by the fitting of emission models to spectroscopic data. In the context of eFEDS, this allows for the measurement of accurate luminosities for clusters that do not have high enough quality data to perform spectral fitting. We can directly compare XMM and eROSITA luminosities for 29 (~80%) of the eFEDS-XCS sample. We use a spectral fitting process to measure unabsorbed (corrected for hydrogen column absorption) luminosities in the soft (0.5-2.0keV) energy band for those clusters with a successful XMM temperature measurement.

We fit a power-law with the slope fixed at unity and another power-law with the slope left to vary to the luminosity comparison, finding the results of both to be entirely consistent with a one-to-one relation. The fits were performed in log space using the R package LInear Regression in
Astronomy(LIRA, Sereno, 2016b), fully described in Sereno (2016a). Figure 4.10 demonstrates an excellent soft-band luminosity agreement (including the two models) between eFEDS and XCS, especially considering the differing measurement methods. We also include the data point for the cluster eFEDS ID 1023 (discussed further in Appendix B.2), but do not include it in our comparison fit. Luminosities measured by eFEDS and XCS are similarly well constrained, though the XCS uncertainties tend to be slightly smaller.

4.4.3 Temperature Comparison

We have been able to measure XMM temperatures within a 500 kpc aperture for ~80% (28) of the eFEDS-XCS sample, though only ~30% (8) of those also have an eFEDS eROSITA temperature available (see Table 1 for a summary). We first compare the overall temperature, and fractional temperature uncertainty, distributions, as we did in Section 4.2 with the XXL-100-GC sample.

Figure 4.11a, which shows the overall distributions of the eFEDS eROSITA temperature and eFEDS-XCS XMM temperature samples, demonstrates that a larger proportion of XMM temperatures than eROSITA temperatures are above ~4 keV, similar to the behaviour in Figure 4.1b with the XXL-100-GC sample. It is likely that this is due to the difference in telescope sensitivity at high energies, as well as other selection effects resulting from targeted XMM exposures.

Figure 4.11b demonstrates that a larger proportion of the XMM temperatures from the eFEDS-XCS sample (compared to eROSITA measurements of the eFEDS sample) have a fractional temperature uncertainty of less than 20%, and as such the XMM temperature measurements are generally better constrained. However we also note that the temperature fractional error distribution of the eFEDS-XCS sample extends to larger values than eFEDS.

In summary, archival XMM observations can provide temperatures which are, on average, better constrained than eFEDS for those clusters that have been observed by XMM, and can also deliver more temperatures for hotter systems due to XMM’s greater sensitivity at high energies. As such, the XMM archive will be a very useful complement to the eRASS.

4.4.4 Temperature Calibration

Motivated by the known temperature offset between XMM and Chandra (Schellenberger et al., 2015), we test for a difference in temperatures measured by XMM and eROSITA. We use 8 XMM
Figure 4.12: Comparison of eFEDS and XCS cluster temperatures within 500 kpc, centered on eFEDS coordinates. Pale blue line indicates best fit power-law (slope free to vary), with 68% confidence levels given by shaded region. Grey line indicates a power-law fit with fixed slope of unity (with 68% confidence levels given by the grey shaded region). Cyan diamond is for the split cluster discussed in Appendix B.2.

Table 2: The normalisation, slope, and intrinsic scatter values of the fitted temperature calibration models for 500 kpc apertures. $A_{TT}$ is normalisation, $B_{TT}$ is slope, and $\sigma_{T_{\text{eROSITA}}/T_{\text{XMM}}}$ the intrinsic scatter.

<table>
<thead>
<tr>
<th>Calibration Name</th>
<th>$A_{TT}$</th>
<th>$B_{TT}$</th>
<th>$\sigma_{T_{\text{eROSITA}}/T_{\text{XMM}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Law</td>
<td>0.88$^{+0.37}_{-0.29}$</td>
<td>0.89$^{+0.25}_{-0.24}$</td>
<td>0.04$^{+0.06}_{-0.03}$</td>
</tr>
<tr>
<td>Power Law Fixed Slope</td>
<td>0.75$^{+0.10}_{-0.08}$</td>
<td>1</td>
<td>0.04$^{+0.05}_{-0.02}$</td>
</tr>
</tbody>
</table>
temperatures that we have measured for a subset of the eFEDS-XCS cluster sample that have eROSITA temperatures presented in the eFEDS cluster catalogue to perform a comparison.

The comparison of 8 clusters with measured XMM and eROSITA (eFEDS) temperatures is given in Figure 4.12, and shows a systematic offset between the two telescopes. We also plot the data point for the cluster with eFEDS ID 1023 (discussed further in Appendix B.2), but do not include it in our comparison fit. All but one of the clusters are below the one-to-one line, indicating that the eROSITA temperatures are systematically lower than their XMM counterparts. To model this, we fit a power law of the form

\[
\log (T_{eROSITA}^{500kpc}) = \log (A_{TT}) + B_{TT} \log (T_{XMM}^{500kpc}) + \sigma_{T_{eROSITA}/T_{XMM}},
\]

(4.1)

where \(A_{TT}\) is the normalisation, \(B_{TT}\) the slope and \(\sigma_{T_{eROSITA}/T_{XMM}}\) the intrinsic scatter. The fits were performed in the same way as those in Section 4.4.2. The best fit parameters are given in Table 2.

First, we fit the power law with the slope left free to vary, probing whether the observed offset evolves with temperature (as found in Schellenberger et al., 2015, comparing between Chandra and XMM). We measure a slope value of \(0.89^{+0.25}_{-0.24}\), indicating that the calibration evolves with system temperature. We note however that due to the large errors, the value of \(B_{TT}\) is consistent with 1 (within 1\(\sigma\)). The measured intrinsic scatter of both fits is very low (essentially consistent with zero), which is as expected. Due to the large errors on the measured slope, we re-fit the power-law with the slope fixed at unity. This allows us to measure a single overall normalisation that describes the average difference in temperatures measured by the two telescopes. We measure a normalisation of \(0.75^{+0.10}_{-0.08}\), meaning that (on average) eROSITA measures a temperature \(\sim 25\%\) cooler than those measured by XMM for the same cluster.

As we have not re-analysed eROSITA data and measured our own eROSITA temperatures for the eFEDS-XCS cluster sample, the observed temperature offset could be the result of some mismatch in our respective methodologies. This may be supported by the analyses of Sanders et al. (2021b) and Whelan et al. (2021), who generated temperature profiles from eROSITA, XMM, and Chandra, and found that XMM temperatures in their profiles were higher than eROSITA; the opposite of our findings. It is possible that a bias has been introduced by fitting a single-temperature plasma model to the spectra, as was explored by Schellenberger et al. (2015), or by a more basic analysis decision such as the choice of background region. We also measure temperatures using a simultaneous fit of all available XMM data, a mixture of PN, MOS1, and MOS2 spectra, which have have been shown
Table 3: The normalisation, slope, and residual scatter values from the LIRA fits of the different datasets, for the $L_{X,0.5-2.0}^{500\text{kpc}} - T_{X}^{500\text{kpc}}$ scaling relation.

<table>
<thead>
<tr>
<th>Relation Name</th>
<th>$A_{LT}$</th>
<th>$B_{LT}$</th>
<th>$\sigma_{L/T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>eFEDS</td>
<td>$1.03^{+0.09}_{-0.08}$</td>
<td>$2.31^{+0.13}_{-0.13}$</td>
<td>$0.19^{+0.03}_{-0.03}$</td>
</tr>
<tr>
<td>eFEDS-XCS</td>
<td>$0.78^{+0.12}_{-0.11}$</td>
<td>$1.58^{+0.22}_{-0.23}$</td>
<td>$0.23^{+0.05}_{-0.04}$</td>
</tr>
<tr>
<td>eFEDS-XCS Calibrated</td>
<td>$1.04^{+0.15}_{-0.14}$</td>
<td>$1.76^{+0.25}_{-0.26}$</td>
<td>$0.22^{+0.06}_{-0.04}$</td>
</tr>
</tbody>
</table>

to measure different temperatures (Schellenberger et al., 2015). The accuracy of the measured offset is also limited by small number statistics, as very few eFEDS selected clusters have both an eROSITA and an XMM temperature. However, we have provided evidence of an offset that requires further investigation to understand the mechanism behind it.

4.5 Discussion

In this work we have presented a measurement of the temperature offset between eROSITA and XMM for a sample of galaxy clusters. Here we discuss potential impacts of this offset on the derived scaling relations and how the temperature calibration can be improved.

4.5.1 Comparison of eROSITA and XMM X-ray scaling relations

We have explored the impact on X-ray scaling relations in light of the temperature offset measured in Section 4.4.4. We focus on the luminosity-temperature relation derived from eFEDS and XCS data. We use 28 eFEDS-XCS clusters with a successful 500 kpc XMM temperature ($T_{X}^{500\text{kpc}}$) and soft-band luminosity ($L_{X,0.5-2.0}^{500\text{kpc}}$) measurement (instead of limiting the analysis to the 8 clusters used in Section 4.4.4 for temperature calibration), and all available eFEDS candidates with eROSITA measurements and compare the relations.

The $L_{X,0.5-2.0}^{500\text{kpc}} - L_{X}^{500\text{kpc}}$ relations for eFEDS (eROSITA) (grey points) and eFEDS-XCS (XMM) (green points) are shown in Figure 4.13; the eFEDS relation uses data from 94 clusters. We fit both sets of data using a power law of the form,

$$\log\left(\frac{L_{X,0.5-2.0}^{500\text{kpc}}}{E(z)L_0}\right) = \log(A_{LT}) + B_{LT}\log\left(\frac{T_{X}^{500\text{kpc}}}{T_0}\right) \pm \sigma_{L/T},$$

where $A_{LT}$ denotes the normalisation, $B_{LT}$ the slope, and $\sigma_{L/T}$ the intrinsic scatter of the relation. We calculate $E(z)$ using the redshift supplied in the eFEDS catalogue and our chosen concordance
Figure 4.13: Soft-band (0.5-2.0 keV) luminosity-temperature relations for the eFEDS and eFEDS-XCS samples. Properties measured within a 500 kpc fixed aperture centered on the eFEDS positions. eFEDS data points are green crosses and the model fit is green, eFEDS-XCS data points are black diamonds with a grey model fit.
Figure 4.14: Corner plot of the 1σ and 2σ confidence contours of the $L_{500\text{kpc}}^{x,0.5-2.0} \cdot T_{500\text{kpc}}^{500\text{kpc}}$ relation parameters, for the eFEDS (green contours), eFEDS-XCS (grey contours) and eFEDS-XCS calibrated (blue contours) samples. The diagonal shows the posterior densities of each parameter.
Figure 4.15: Soft-band luminosity-temperature relations for eFEDS, eFEDS-XCS, and calibrated eFEDS-XCS. Properties measured within a 500 kpc fixed aperture centered on the eFEDS positions.
cosmology. The fits are performed using the LIRA package. We set normalisation values for luminosity and temperature to approximate median eFEDS values; $L_0 = 3.0 \times 10^{43}$ erg s$^{-1}$ and $T_0 = 2.3$ keV.

Figure 4.13 shows the best-fit relations using the eFEDS (green line) and eFEDS-XCS (grey line) samples respectively. The best-fit values are given in Table 3, with their distributions illustrated in Figure 4.14. While the distributions highlight that the parameters of the relation are consistent within their 2$\sigma$ contours, the difference in the central value of the slope warrants further discussion.

We explore whether this difference can be reduced by accounting for the observed temperature offset found in Section 4.4.4. We measure a third version of the scaling relation, designed to test the effect of the temperature calibration quantified in Section 4.4.4. We determine a “calibrated” luminosity-temperature relation by using the power law model (with slope free to vary), with parameter values provided in Table 2, to convert the eFEDS-XCS XMM temperature values to predicted eROSITA values.

Figure 4.15 shows the model fits for all three relations (with data points omitted for clarity), and shows that the eFEDS-XCS calibrated scaling relation has a steepened slope and increased normalisation when compared to the original eFEDS-XCS relation. Figure 4.14 shows a shift of the contours and distributions (the eFEDS-XCS calibrated parameters are given by the blue contours) towards eFEDS (blue contours). The normalisation of the calibrated eFEDS-XCS relation is fully
consistent with the eFEDS relation, with the tension in the measured slopes reduced.

The slopes of all the luminosity-temperature scaling relations presented in Table 3 are lower than those which are typically measured (Maughan et al., 2012; Giles et al., 2016, 2022b). Only one of the relations (calibrated eFEDS-XCS) is consistent with the self-similar prediction of a slope of 2 within $1\sigma$, though most recent work derives a steeper luminosity-temperature relation than that prediction.

Bahar et al. (2022) have constructed a variety of scaling relations using eFEDS eROSITA properties measured within $R_{500}$, including luminosity-temperature relations. They found their relation to be in agreement with past work using XMM, though with a slightly steeper slope than the XXL (Giles et al., 2016) result. Our relations indicate a larger difference in slope, but we do not account for selection effects, we use properties measured within 500 kpc rather than $R_{500}$, and our analyses are not completely consistent with eFEDS as has been discussed elsewhere. The use of fixed apertures removes some information about the scale of each galaxy cluster, and is not common in the construction of scaling relations; as such there are no previous works to compare these results to (though as these relations were meant only as a test of the eFEDS sample, that is not important). It has been shown that biases in the construction of cluster samples, particularly luminosity limits, can significantly flatten the slope of luminosity-temperature scaling relations (see Figure A1 of Mantz et al., 2010).

4.5.2 Future work to improve the calibration of the XMM to eROSITA temperature offset

The measured temperature discrepancy described in Section 4.4.4 and shown in Figure 4.12 is based on only 8 sets of measurements, of which only 3 are more than one sigma from the one-to-one relation. Further investigation is required to quantify the level of a temperature offset between the XMM and eROSITA. For this, we plan three complementary approaches:

- Re-analyse the eFEDS cluster candidate observations using an identical spectroscopic methodology to the XMM analysis. This way, we will have control of all aspects of the analysis, including which regions are used to mask the observations for spectrum generation (see a related discussion in Appendix B.2). This will be done for all 94 clusters in the eFEDS luminosity-temperature analysis (Figure 4.13), which includes the eight clusters featured in Figure 4.12.
• Propose *XMM* follow-up observations of a representative sample of eFEDS clusters with robustly (i.e. percentage error less than 25%) measured *eROSITA* temperatures. There are 43 such examples in the eFEDS data release that are not already included in the analysis shown in Figure 4.12. We will preferentially select clusters that fill gaps at the high and low temperature ends, to better constrain a temperature dependent slope in the calibration relation (if one exists).

• Repeat the analysis herein after the next *eROSITA*-DE data release (due in Q4 2022\(^7\)). This will cover the whole Southern sky (red area in Figure 4.16) and thus overlap with many more archival *XMM* observations than did eFEDS (grey points and regions in Figure 4.16). The next data release will have an exposure time eight times less shorter than eFEDS (one pass), but, with ~150 times the area, we can still expect it to yield roughly 1000 robust temperature measurements.

### 4.6 Summary

In this work we have performed the first comparison between cluster properties measured by the eFEDS survey and those measured by *XMM* surveys, both directly and for ensembles of clusters. A comparison of XXL-100-GC and the eFEDS optically confirmed sample indicated that the two samples have very similar redshift distributions, that XXL-100-GC contained proportionally more temperature measurements above 3.5 keV, and that XXL-100-GC temperatures are, on average, better constrained than eFEDS temperatures (14% vs 25% average percentage uncertainties respectively).

We have located and analysed eFEDS cluster candidates that have a counterpart in an *XMM* observation; as part of this process we visually inspected eFEDS cluster candidates that are within an *XMM* observation and rejected any that had no ICM emission at the eFEDS coordinates, had obviously contaminated ICM emission, or too low quality *XMM* data. The eFEDS-*XMM* sample contains 62 eFEDS cluster candidates that have been observed by *XMM*, and the eFEDS-XCS sub-sample contains the 37 clusters that pass our visual inspection (Table 1). During visual inspection we found that 10 candidates (Table 1) did not have sufficient *XMM* data to confirm or deny a cluster, 4 had X-ray flux contamination (Section 4.3.4, Table 3), and 11 were sample contaminants (Section 4.3.3, Table 2). We found that the majority of eFEDS-*XMM* candidates had a longer exposure time (at their eFEDS position) in *XMM* than in eFEDS. We also found that the majority of

---

\(^7\) *eROSITA*-DE Data Release Schedule
eFEDS centroid positions for the eFEDS-XCS sample are within 100 kpc (at the eFEDS redshift) of the XCS centroid positions (for clusters with an XCS match), though some outliers exist due to low signal-to-noise eFEDS data.

Our visual inspections of the eFEDS cluster candidates that fall on an XMM field (using eROSITA, XMM, SDSS and HCS images) have shown that there are some aspects of eROSITA’s source finding and confirmation steps that can introduce spurious sources into their catalogues, which in turn could impact their cosmological analyses. Our inspection process finds that the eFEDS-XMM sample is (at minimum) \( \sim 18\% \) contaminated, and that the optically confirmed sample is \( \sim 9\% \) (\( \sim 6\% \) if made more consistent with the eFEDS sample selection method) contaminated. This is consistent with predictions from simulations (Comparat et al., 2020) and eFEDS measured values (Klein et al., 2021).

We have presented comparisons between cluster luminosities and temperatures measured with eROSITA and XMM. Our analysis finds excellent agreement between soft-band luminosities measured by eFEDS and XCS for the eFEDS-XCS sample (Section 4.4.2), which is very encouraging for future eRASS cosmology analyses. Such analyses will rely almost exclusively on mass-luminosity relations (such as the recent eFEDS-HSC work, Chiu et al., 2021), as even full depth eRASS will not be able to measure temperatures for enough galaxy clusters to use mass-temperature relations for cosmology. An ensemble cluster comparison was performed (Section 4.4.3) between the whole eFEDS cluster candidate sample (that had successful temperature/luminosity measurements), and the eFEDS-XCS sample (that had successful temperature/luminosity measurements). It showed similar results to the XXL-100-GC comparison, with a greater fraction of XMM temperature measurements being above 4 keV than eFEDS temperatures, and better average temperature constraints for XMM measurements.

A discrepancy between cluster temperatures measured by eFEDS and XCS has been found and quantified (Section 4.4.4), with eROSITA temperatures being (on average) \( 25\pm 9\% \) cooler than those measured by XMM for the same cluster. This could hint at the need for a calibration function between the eROSITA and XMM telescopes (as has been necessary between XMM and Chandra). Several variables need to be better controlled before we can definitively state that the observed discrepancy is entirely due to a required temperature calibration. Alternative explanations including biases introduced by fitting a single temperature plasma model to a multi-phase emission, or a combination of analysis choices, are made more likely by initial studies of temperature profiles finding eROSITA temperatures to be higher than XMM temperatures.
We also fit and compare luminosity-temperature scaling relations using data from the eFEDS catalogue and XMM data for the eFEDS-XCS sample, using them to compare scaling relations measured with eROSITA and XMM data. This has particular relevance for anyone wishing to use an XMM generated scaling relation in an eROSITA analysis, or vice versa, as we find a distinct tension between scaling relations from the eFEDS and eFEDS-XCS samples. We also generate a second, calibrated, eFEDS-XCS scaling relation which is in better agreement with the eFEDS relation; normalisation becomes entirely consistent, and the slope measurement tension is reduced. This again may suggest that a temperature scaling between XMM and eROSITA is necessary. It is likely, however, that a large part of the observed tension between the scaling relations is due to selection effects.

Our comparisons have shown that we can expect a great deal of useful data from the full eRASS catalogues, and that XMM-Newton still has a significant part to play as followup instrument. Its archive of 20 years worth of observations is still extremely valuable to the eROSITA team, and the X-ray astronomy community as a whole, and there will be many excellent opportunities for synergies between the two telescopes.
Chapter 5: Conclusions and future work

This final chapter is used to give high-level overviews of projects undertaken during the PhD that do not have entire chapters dedicated to them. Then a summary of plans for future work related to this thesis, and final discussion and conclusions are presented.

5.1 Other Projects

5.1.1 Artificial XMM observations of simulated galaxy clusters

This section gives an overview of the creation and uses of the XCSim software for creating artificial XMM observations of simulated galaxy clusters.

5.1.1.1 Aims and motivations

This work involves the creation of software tools to create realistic artificial XMM observations of simulated galaxy clusters from cosmological hydrodynamical simulations. Such artificial observations have a myriad of uses, from injecting simulated objects with known properties into existing observations to assess selection effects of the XAPA (Section 1.4.3) source finder, to helping to investigate the systematics involved in measurement of galaxy cluster masses with the hydrostatic mass method. These artificial observations can also provide value to the simulation teams, as making measurements from realistic simulated data products allows for comparisons of bulk X-ray properties of clusters taken from the simulations to real clusters.
5.1.1.2 Which simulations are used?

There are a limited number of simulation suites that are applicable to the kind of artificial observations that XCS is interested in. All dark-matter only simulations cannot be used for the creation of X-ray observations using XCSim, as the temperature, density, metallicity, and velocity fields of gas particles are all used to model the X-ray emission of the cluster. Another caveat is that the hydrodynamical simulation has to be run down to a redshift of zero (or very nearly zero), as galaxy cluster formation occurs at relatively low redshifts. The simulations also have to either: a) be run in large enough boxes that large-scale structure and thus clusters are able to form, or b) be ‘zoom simulations’ of galaxy clusters, where a portion of a dark matter only run containing a galaxy cluster is re-simulated with full gas physics.

As such this project has only made use of two sets of simulations so far; firstly, the highest resolution, 300 Mpc per side, IllustrisTNG (Springel et al., 2018; Pillepich et al., 2018; Marinacci et al., 2018; Nelson et al., 2018; Naiman et al., 2018). Specifically the same subset of 370 simulated galaxy clusters from the redshift zero snapshot selected by (Barnes et al., 2018), using a simple mass cut of $M_{500} > 10^{13.75} M_\odot$. Secondly a set of zoom-simulated clusters called the 300 Project (Cui et al., 2018), so-called because the re-simulated regions contain the three-hundred most massive clusters from the large dark matter only simulation, Multi Dark Planck 2 (MDPL2).

5.1.1.3 How is the X-ray emission simulated?

X-ray photons are generated from the hydrodynamical simulations using a Python module called pyXSIM (ZuHone et al., 2014). This process does not generate X-ray photons from the entire space of the simulations, but rather for a spherical region centered on the coordinates of the galaxy cluster being ‘observed’, with the radius of the region controlled by the user. This ensures that assumptions about the emission mechanism are not imposed on gas for which they are not valid, as well as avoiding unnecessary computation of photons that do not contribute to the end goal of ‘observing’ the cluster.

The premise of the pyXSIM module is simple, with the generation (note that the photons are not ‘observed’ in this step, see Section 5.1.1.4 for more information on that) of photons having two main steps:

1. **Generating Photons** - The temperature, density, and metallicity fields of the physical simulations are used to generate a large, 3D collection of photons, each with an energy, position,
and velocity vector. This is achieved by assuming a model of galaxy cluster intra-cluster medium emission for each particle of the simulated cluster. In this case the APEC model is used, which XCS uses to fit to galaxy cluster spectra. This model incorporates the ICM continuum emission from bremsstrahlung (which uses the temperature and density information of the particle), as well as transition emission lines from the metal content of the ICM. Either specific tracked metal fields can be used for this purpose, or a global metallicity field, depending on what is available for a particular simulation. Photons generated for each particle are combined into the final 3D photon sample, thus properly representing the composite emission from different parts of the galaxy cluster. This step requires the user to choose a length of time for which to generate photons, which is normally set to a value several times longer than a typical XMM observation exposure time (500 ks for instance), to provide a large, representative, sample of photons to sub-sample.

2. **Projecting Photons** - Processing the photons generated in the first step, and projecting them from a 3D distribution to a 2D distribution by projecting onto an imaginary plane of the sky. The line of sight direction is chosen by the user, allowing for the same simulated galaxy cluster to be observed from many directions. This can be beneficial for some use cases, such as machine learning techniques where you want to account for every possible direction clusters could be viewed from to maximise your training samples. Also, X-ray photons emitted from real galaxy clusters are partially absorbed by the neutral hydrogen column of the Milky Way in the direction of absorption, and as such an absorption model (tbabs) is applied assuming the observed column density of the RA-Dec coordinates where the simulated cluster will be placed. Once a projected set of photons is available, it can be ‘observed’ by any number of artificial X-ray telescopes, in any number of configurations.

### 5.1.1.4 Constructing an artificial version of XMM

Once a projected photon list generated from intra-cluster medium emission models is available, XCSim has to ‘observe’ it with an artificial telescope that copies the properties of the XMM-Newton cameras. This is performed using a forked, modified, version of the SOXS\(^1\) (Simulated Observations of X-ray Sources) Python module, which is written by the same person as pyXSIM. This module can imitate the observations of most X-ray telescopes, but at the time of the project did not apply realistic enough instrumental effects to meet the needs of the project (updates since then have added some of the required functionality).

---
\(^1\)[https://hea-www.cfa.harvard.edu/soxs/](https://hea-www.cfa.harvard.edu/soxs/)
Figure 5.1: Sets of polygons that map boundaries of XMM detectors for observation 0201903501. The maps are in XMM detector coordinates, one of the internal XMM coordinate systems. Left hand side shows the detector boundary map for MOS2, the right hand side shows it for PN.

XCSim uses an existing XMM observation as a ‘donor’, and proceeds to emulate the state of the telescope at the time of that observation. This includes any lessening of the sensitivity of the detectors due to their age, any damage that the cameras have sustained, and (optionally) the exposure time used. The most complex parts of quantifying the detector are:

1. Mapping the positions, boundaries, and health of all the CCDs that make up XMM’s three cameras. This part of the pipeline uses an XMM exposure map and an edge finding algorithm to construct the sets of chips and dead pixels that make up the XMM detectors. Figure 5.1 shows the sets of polygons measured to describe the MOS2 and PN detectors as they were at the time of observation 0201903501.

2. Applying the XMM point spread functions (PSF). Most previous artificial X-ray observation efforts make the assumption that the PSF is Gaussian, and while this approximates the truth for observations that sit near the centre of the pointing, it rapidly becomes a poor assumption as you move to the outskirts of the detector; a Gaussian model also doesn’t take the spoke features caused by the support structure of the optics into account (see Figure 2.8 for examples). As such XCSim generates instances of the Read et al. (2011) XMM PSF model (ELLBETA) at various positions on each detector; they are treated as a discrete multivariate PDF, and then inverse-transform sampling is used to generate new detector positions for the projected artificial photons.
3. Adding realistic changes in detector sensitivity across the field of view of the artificial detector. In a similar vein to the PSF, most artificial X-ray observations use one effective area curve (usually one generated for the central region of the detector) to describe the sensitivity of entire field of view. As past work on this front has placed clusters at the centre of the pointing, this assumption was valid, but XCSim is able to place clusters anywhere in the field of view, as it will be used to evaluate the XCS selection function. As such XCSim generates a grid of Auxiliary Response Files for different regions of the donor observation’s detectors. The effective area curve is used to accept or reject candidate photons, but this now happens for each of the regions in the grid independently, adding a granular variable sensitivity.

4. XCSim can be used to make perfect images of simulated clusters, with no background signal to add noise to our measurements, but unfortunately that does not reflect reality. It is necessary to introduce a realistic X-ray background so that the artificial observations are analogous to the real world. Once again XCSim makes use of the donor observation, and randomly sub-samples the existing X-ray data to create an X-ray background suitable for the chosen exposure time of the artificial observation (the only limit on this is that the artificial exposure cannot be longer than the donor exposure if we want a background).
Figure 5.3: A galaxy cluster luminosity-temperature relation for a sample of real galaxy clusters analysed by XCS (in black, with a purple fit line), and properties measured from artificial XMM observations of a sample of simulated clusters taken from IllustrisTNG. The simulated cluster measurements are shown in blue, with a red fit line.

The artificial data products produced by XCSim are indistinguishable from real XMM data products, and as such can be inserted into XCS’s existing source detection and analysis pipelines with no changes (including XGA). An example of this is given in Figure 5.2, where XAPA has successfully detected an IllustrisTNG cluster added to an existing COSMOS field.

### 5.1.1.5 Preliminary results

Very little scientific output has resulted from this project so far, and no XCSim simulations have been used for scientific analysis in any published papers. A very limited comparison between a sample of XCS-measured galaxy cluster properties (Giles et al., 2022b) and measurements of $T_X$ and $L_X$ for samples of simulated clusters has been performed. This comparison shows that IllustrisTNG galaxy clusters are more luminous than observed clusters with the same temperature. Figure 5.3 demonstrates this issue, showing a markedly different luminosity-temperature scaling relation slope than the real clusters.
Figure 5.4: A galaxy cluster luminosity-temperature relation for a sample of real galaxy clusters analysed by XCS (in black, with a purple fit line), and properties measured from artificial XMM observations of a sample of simulated clusters taken from the 300 project. The simulated cluster measurements are shown in blue, with a red fit line.
An equivalent analysis was performed for artificial XMM observations of simulated clusters taken from the 300 project (comparison shown in Figure 5.4), and though all the clusters are much more massive (and thus hotter and more luminous) than IllustrisTNG clusters, the comparison shows far better agreement with observations. The XCS fit in this instance was performed only on real clusters whose temperature was between the coolest and hottest 300 project clusters.

5.1.2 Building samples of Pea galaxies with machine learning

This section gives an overview of the project, starting with the reasons why larger samples of Pea galaxies might be useful, then giving some information on the techniques that have been used so far, as well as the difficulties of this project.

5.1.2.1 What are Pea galaxies?

Peas galaxies are a rare type of compact, highly star forming galaxy. They were first discovered as ‘Green Peas’ by volunteers taking part in the original Galaxy Zoo project classifying SDSS sources and were first reported by Cardamone et al. (2009). Green peas are so called because, when RGB images were created using SDSS’s gri filters, they appeared as compact, round, and vibrantly green objects, see Figure 5.5 for images of some of the first Green Peas discovered, and compared with a random elliptical galaxy.

Figure 5.5: Example g.r.i composite colour 50′′ x 50′′ SDSS images classified by Pea hunters, with the r-band representing green light. The distinctly green colour and compact morphology makes the Peas (left 3 images) easily distinguishable from the classical elliptical (right image). The elliptical galaxy is clearly red and has a smooth profile, while the Peas are r-band dominated and unresolved in these images, appearing like stellar point sources. All objects shown here are at $z \sim 0.2$, and this figure is taken from Cardamone et al. (2009).

Since the original work, other colours of Pea galaxy have been discovered, this is because dis-
Distinctive colour of the Green Peas comes from a very powerful [OIII] (5007 Å) emission line that dominates SDSS’s r band and increases the luminosity with respect to the g and i bands. Due to an emission line being responsible for the appearance, Peas will only be green within a redshift range of approximately $0.112 \leq z \leq 0.360$. At lower redshifts they change colour to purple and blue, and at high redshifts they tend to present as redder objects. Figure 5.6 shows an image and SDSS spectrum for a Green Pea, and for a Purple Pea.

Figure 5.6: The top left figure shows an image of a Green Pea galaxy, taken by the SDSS photometric survey, and the top right figure is the corresponding spectrum for that Green Pea galaxy, taken with the BOSS spectrograph. The bottom left figure shows an image of a Purple Pea galaxy (at lower redshift than the Green Pea), taken by the SDSS photometric survey, and the bottom right figure is the corresponding spectrum for that Purple Pea galaxy, taken with the SDSS spectrograph. Both spectra have filter curves overlaid to indicate which filter then emission lines fall within.
5.1.2.2 Aims and motivation

The main aim of this project is to locate large samples of Pea galaxies in optical and near-infrared surveys, and then provide them to the astronomy community for further analysis. This will also enable the construction of multi-wavelength Pea galaxy samples, as a greater base sample of Pea galaxies provides greater scope for serendipitous detection in other wavelengths (in X-ray with XMM for instance). The original SDSS sample was made up of 251 Green Peas, and the number of identified Peas is still of the order $10^2$, so any increase in sample size could be impactful.

Peas galaxies are studied because it's thought that they may give insight into the “Epoch of Reionisation” (EOR), when the transition of the Universe’s baryons from neutral hydrogen to ionised plasma occurred. This transition allowed baryonic matter to play a more significant role in structure formation (whereas before dark matter dominated) and is thought to have happened soon after the formation of the first galaxy and star populations.

The cause of reionisation is one of the most important astronomical phenomena being investigated today, and will be the subject of a great deal of study by the James Webb Space Telescope (JWST). Studying Pea galaxies provides a complementary route to understanding the EoR, as these compact, highly star-forming, metal poor galaxies are considered to be possible local analogues of the first types of galaxies that formed in the Universe. The escape of ionising radiation in the form of Lyman continuum photons (LyC) is a possible candidate for the source of reionisation, and if Pea galaxies can be shown to emit a certain level of ionising photons, it would lend credibility to the claim that early galaxies did as well. There is also limited evidence to suggest that Pea galaxies have a greater than expected X-ray emission (Svoboda et al., 2019), which could allow a larger fraction of LyC to escape the galaxies. The paper by (Svoboda et al., 2019) was the inspiration behind this project, as XCS’ experience with serendipitous X-ray analyses may allow the investigation of the X-ray properties of a larger sample of Pea galaxies.

5.1.2.3 Main difficulties

The original SDSS sample was made up of 251 objects that had been observed both photometrically and spectroscopically, allowing the authors to confirm that the objects had galaxy-type spectra. The spectra available for this sample also led to the discovery of extreme emission lines as the cause of the visual appearance of Pea galaxies. These emission lines introduce difficulties in the construction of Pea galaxy samples, as the redshift of the source has to be known to ensure
Figure 5.7: Comparisons of photometric redshifts and spectroscopic redshifts measured by SDSS. The left hand side shows the comparison for a sample of Green Pea galaxies selected from SDSS, and the right hand side shows the comparison for a sample of randomly selected galaxies in the same redshift range as the Green Peas.

that a Pea is not being confused with a different type of source at a different redshift. Pea galaxies are highly separable from other objects in the same redshift bin, with occupying a distinctive area in colour space, but only when compared to objects in the same redshift bin. Unfortunately only a small fraction of photometrically observed SDSS galaxies are followed up by spectra, and in other large scale surveys (such as DES) the lack of a dedicated companion spectrograph means there is effectively no large scale spectroscopy available; making it hard to expand to other surveys with current methods. This reliance on spectra cannot be broken with current photometric techniques, as they perform poorly on the atypical Pea galaxies. Figure 5.7 demonstrates this by comparing photometric and spectroscopic redshifts of a sample of Green Peas, and a sample of galaxies selected randomly from SDSS. The scatter on photometric redshifts compared to spectroscopic is much larger for Pea galaxies.

During the course of this project it has also become clear that, presumably due to the very compact and bright nature of these galaxies, the SDSS photometric classification pipeline misidentifies many (∼80% of Green Peas) of the candidates as stars. This can mean that photometric redshift measurement codes aren’t even run on them, and they may be less likely to be selected for spectroscopic follow-up.
5.1.2.4 Developing machine learning methods for locating Pea galaxies

The original paper that described Green Peas (Cardamone et al., 2009) constructed cuts on colour space that separated Green Peas from other SDSS objects in the same redshift range; Figure 5.8 shows the cuts chosen by the paper, and demonstrates that Green Peas are quite clearly very separable from other objects in colour space. A recreation of the original selection method, taking advantage of the significant increase in galaxy spectra collected by SDSS since the original analysis, produced a sample of \( \sim 1000 \) objects after visual inspection. After the original work by (Cardamone et al., 2009), a new method designed to find other colours of Pea galaxy was proposed and used by (Izotov et al., 2011). This method used the relative luminosities of various emission lines to identify Peas across various redshift bins. Making use of the bulk fitting of SDSS and BOSS spectra by (Thomas et al., 2013) this method was repeated, and small samples of all colours of Peas were constructed. It was noted that this method did not recover all of the Green Peas present in the updated sample using the original method, but the emission line method provided the samples required to experiment with a machine learning method called a Support Vector Machine (SVM).

SVMs are an algorithm that use simple machine learning to essentially replicate the process taken
Figure 5.9: A corner plot showing the feature space used to train a Green Pea classifying SVM. The blue points and histograms represent the training set of Green Peas, whereas the orange points and histograms represent random galaxies drawn from the same redshift range.

by the authors of the first Green Pea paper, and uses training sets to define an N-dimensional hyperplane (or hyperplanes) that achieve the maximum separation of different classes. This method supports as many dimensions as there are features of the different classes, but it is a ‘supervised’ machine learning method, meaning that the training set has to have been labelled beforehand. Using the emission line method, as well as follow-up visual inspection, a labelled dataset had been created.

The feature-set comprised of all available SDSS band magnitudes, as well as Petrosian radii measured in those bands; Figure 5.9 shows a feature hyperplane that was used to train and verify an SVM for Green Pea classification, with clear separation between the property distributions of the two classes. As another indication of how separable Pea galaxies are from other objects, Figure 5.10 demonstrates the results of projecting the hyperplane into a two-dimensional space, with two distinct clusters forming (representing Peas and not Peas). A random sampling of other SDSS objects with photometric and spectroscopic data was added to the Pea sample to balance the data-
Figure 5.10: This figure shows the hyperplane in Figure 5.9 projected into two dimensions using the UMAP (McInnes et al., 2018) dimensionality reduction technique. Once again the blue points represent Green Peas, and the orange points represent randomly selected galaxies from the same redshift range.
set, so that the method would learn how to identify what was and wasn’t a Pea galaxy. As Pea galaxies appear distinct from other objects in their redshift band, the first experiments with this method only trained classifiers for objects that had been pre-selected to be within the approximate redshift range of the type of Pea they concerned. So separate SVMs were created to find Green Peas, Purple Peas, and Red Peas. Spectroscopic redshifts were still required, as Figure 5.7 demonstrates that photometric redshifts for Pea galaxies are not necessarily reliable.

Once trained, the SVM classifiers for single Pea types are used to classify the test set, a portion of the whole set of the relevant type of Pea galaxy and random galaxies that is set aside and not used in training or validation, and should provide an unbiased measure of the performance of the classifier. Figure 5.11 shows the confusion matrix produced from the test set for the Green Pea SVM, and shows exceptional accuracy; a prediction of Pea is correct 99.85% of the time, and a prediction of not Pea is correct 97.4% of the time. This demonstrates the power of SVMs when applied to this problem, even though when applied to hundreds of thousands of candidates this will result in thousands of misclassifications, it is extremely impressive accuracy.

An attempt was also made to train a classifier that did not require a redshift as an input, and instead would be able to separate Pea galaxies from different redshift bands, and random galaxies
from those same redshift bands, purely based on the feature set presented in Figure 5.9. The benefits of successfully training such a classifier would be significant, as it could be applied to large scale photometric surveys without spectroscopic follow-up (like the Dark Energy Survey) to locate even more Pea galaxies. Figure 5.12 shows the confusion matrix for this classifier which, while producing unexpectedly good performance, does not achieve the same level of accuracy as the individual classifiers.

### 5.1.2.5 Future plans for the project

The main next step for this project is to converge on a final Pea galaxy catalogue that makes use of a combination of SDSS photometry/spectroscopy and SVM classifiers trained for individual classes of Pea galaxy. Once a sample like this has been assembled and visually inspected, it can both be made available to the community as a whole, and be used as the first building block of the next steps.

These next steps will be multi-faceted; with the first main goal being to increase the size of Pea catalogues by developing methods that can identify Pea galaxies without spectroscopic redshifts (thus increasing the possible sample pool in SDSS and allowing for expansion to other optical/NIR surveys). This could take the form of attempting to fine-tune a photometric redshift method to provide better results for Pea galaxies, but efforts will mostly be focused on building so-called ‘ensemble’ machine learning methods to identify Peas. These will use the existing trained SVMs to perform an initial classification and narrow down which candidates might be Peas, and then a deep neural network that has been retrained to classify Pea galaxies to make the final determination. Such a method could be applied without much modification to images from the Dark Energy Survey, drastically increasing the reach of the Pea galaxy search.

The second main goal is to assemble complementary catalogues of Pea galaxy properties from other wavelengths. Considering the inspiration for this project was work analysing the X-ray properties of some Green Pea galaxies, and that one of the main goals of this work has always been to expand that X-ray sample, the first complementary analyses will be performed in the X-ray band. This will likely take the form of determining which, if any, of the new Pea galaxies have been detected serendipitously in X-ray observations (initially *XMM*, but expanding to other X-ray telescopes), then measuring basic properties such as flux, and trying to fit emission models to spectra generated for those candidates. UV observations of Pea galaxies are also of particular interest, as given the typical redshifts of Pea galaxies, extreme UV observations should contain Ly-
Figure 5.12: The confusion matrix for a multi-Pea SVM classifier trained to separate Pea galaxies from other galaxies drawn randomly from the same redshift range as each Pea type. The classifier is ignorant of the redshifts of the galaxies passed into it, though redshift is used in the construction of the training samples. This confusion matrix was generated from the testing set, and as such the data were not used during training.
man continuum emission, giving information about the escape fraction for these ionising photons. Observations in other wavelengths will be able to tell us about other properties of the Pea galaxies, but have not yet been planned; it is likely that collaboration with experts in galaxy observations in other wavelengths will be sought.

5.1.3 *XMM* properties of LoTSS DR2 radio sources

![LoTSS DR2 Source Number Density](image)

Figure 5.13: A figure showing the number density of LoTSS DR2 radio sources per square degree. The LoTSS DR2 catalogue was constructed from two separate observation fields, which are evident in this Mollweide projection sky plot. LoTSS DR2 source positions were binned using HEALPix to calculate source density on the sky, with the number of sides set to 64. Dark blue areas indicate that no LoTSS DR2 sources are located in that part of the sky.

This section gives some information on a small project to find *XMM* properties of radio sources from the recently released LoTSS DR2 catalogue.

5.1.3.1 Aims and motivations

The LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2017) is a wide and deep radio survey in the 120-168 MHz band. The full survey will cover the entire northern sky, and will be complemented by other wide area surveys performed using LOFAR; the low frequency 42–66
MHz LOFAR Low Band Antenna Sky Survey (LoLSS; de Gasperin et al., 2021) and the even lower frequency 14–30 MHz LOFAR Decametre Sky Survey (LoDSS). The first LoTSS data release (Shimwell et al., 2019) covered 424 deg$^2$, approximately 2% of its final area coverage. The second LoTSS data release (referred to as LoTSS DR2), has substantially increased the coverage area. LoTSS DR2 (Shimwell et al., 2022a) covers a total of 5635 deg$^2$ (27% of final area), in two fields totalling 4178 deg$^2$ and 1457 deg$^2$ respectively. The LotSS DR2 radio catalogue contains 4396228 sources, with Figure 5.13 showing the number density of LoTSS DR2 sources per square degree. Source detection is performed on Stokes I maps generated at an 6$''$ resolution.

The science goals of the LOFAR telescope and its wide area surveys (LoTSS, LoLSS, LoDSS) are broad, ranging from the detection and investigation of radio galaxies at high ($z > 6$), to the analysis of radio halos associated with the evolution of galaxy clusters and large scale structure. Radio halos associated with clusters can help to characterise sources of non-thermal pressure (Cassano et al., 2010a) in the intra-cluster medium (ICM), which is crucial to understanding the evolution of clusters and accounting for non-thermal pressure support in X-ray mass measurements (Ettori et al., 2013; Eckert et al., 2019b). The lower-frequency LOFAR surveys will also be able to provide insights into cosmic magnetic fields (Beck et al., 2013), including those within galaxy clusters. The LoTSS will also substantially advance the study of Active Galactic Nuclei (AGN), increasing sample sizes of all types of AGN, as well as finding higher redshift systems.

Many of the sources that LOFAR and the LOFAR Two-metre Sky Survey were designed to investigate are likely to have counterpart X-ray emission. X-ray emission from AGN is ubiquitous and has been the subject of significant study; Bianchi et al. (2009) constructed a catalogue of AGN from the XMM-Newton archive, and Liao et al. (2020) explored the connection between radio and X-ray luminosity of AGN, finding a strong correlation. X-ray emission from galaxy clusters is also well documented (Piffaretti et al., 2011; Martino et al., 2014; Pacaud et al., 2016; Liu et al., 2021a; Giles et al., 2022b), giving insights into the properties and morphology of the ICM. Many cool-core galaxy clusters are also known to host radio mini-halos (Giacintucci et al., 2019), with multi-wavelength X-ray-radio investigations (Ignesti et al., 2020) providing insights into non-thermal components of the ICM.

Clearly there are significant opportunities for synergy between X-ray and radio telescopes. Different X-ray telescopes offer different advantages and downsides; XMM-Newton offers sensitivity and a two-decade science archive, Chandra also has a two-decade archive of observations as well as excellent spatial resolution (though is less sensitive than XMM), and eROSITA very similar sensitivity to XMM combined with a large field of view (FoV). The goals of this work are to use XGA
alongside XCS processed XMM data and region files to; a) determine which radio sources from the second LoTSS data release are on an XMM observation (serendipitous or otherwise); b) use LoTSS shape information to extract X-ray fluxes for the sources; and c) perform preliminary matching to XCS source regions.

5.1.3.2 Using XGA for simple investigations of massive samples

There are over four million sources in LoTSS DR2 catalogue, something that currently makes it unfeasible to setup an XGA sample object (Section 2.3.1), largely due to computational constraints, but also due to the fact that there will be many different types of astrophysical source in the LoTSS catalogue, and XGA samples are designed to analyse objects of a single type.

However, there are certain functions of XGA that can be used without setting up sources or samples, and one of them is very useful in this case. The XMM observation matching functions (Section 2.3.4) that are used to locate the relevant data for XGA sources are implemented separately from XGA source class, and as such can be called by passing RA-Dec coordinates, without the overhead of setting up a sample.

When we wish to determine whether sources in a very large sample fall on an XMM observation, the on_xmm_match function can be used; this triggers a two step process that ends with the output of an array containing either a null entry, if no XMM data are found, or a list of XMM ObsIDs on which the coordinate of a particular source fall. The two steps are as follows:

1. An initial search to find XMM observations with an aimpoint within 15' of each source is performed. The observations must have at least one instrument that was not in a calibration mode, and any observations located are output with a unique identifier to match them to the relevant source. Any sources that don’t have any relevant data are flagged for the next step. This process is fully parallelised, something that can become necessary when searching for XMM data for a large input catalogue.

2. All input sources that have been flagged as not being near any XMM data are disregarded at this point. Next, all ObsIDs identified as being near to at least one of the input sources are collated into a list of unique ObsIDs. An XGA NullSource (used for the bulk generation of photometric products for input ObsIDs, see Section 2.3.2) is declared and used to generate XMM exposure maps for all the ObsIDs. Finally each input source coordinate is checked against its initially associated ObsID(s) from the last step, fetching the value of the exposure.
map at the coordinate; if it is zero then the coordinate is off of the observation. These processes are also fully parallelised.

5.1.3.3 Preliminary results

This project is only partially complete, so results are limited. However a sub-sample of LoTSS sources that have some XMM data has been assembled, and unsophisticated matches to the XCS master source list (Section 1.4.4) performed. Of the 4396228 radio sources presented in the LoTSS DR2 catalogue, 172374 fall on at least one XMM observation. Figure 5.14 shows the positions of those sources.

Figure 5.14: A figure showing the locations of LoTSS DR2 radio sources that have been found to fall on at least one XMM observation. They do not necessarily have corresponding detections in the XCS master source list.

A simple matching technique was used to identify possible XCS detections of LoTSS DR2 selected sources. The LoTSS DR2 sub-sample of sources that have some XMM data were matched to the latest XCS master source list, with a match occurring if any XCS source was within 10'' of a LoTSS DR2 source position. This results in 9105 (5.25% of the LoTSS DR2 with XMM data sub-sample) sources matching to an XCS source.

In the future this matching criteria will be refined further, taking into account delineations between
point and extended samples. Side-by-side \textit{XMM} and LoTSS DR2 images of the matching sources will also be generated, in case a visual inspection of any of the sources is required. Additionally flux measurements will be made for every LoTSS DR2 source with \textit{XMM} data, regardless of whether an XCS match was found or not; these fluxes will be measured within the region defined by the LoTSS DR2 catalogue semi-major and semi-minor axis entries. Finally, the results of this work will be made publicly accessible, hopefully enabling radio astronomers who wish to explore X-ray properties of these sources to start their research projects more easily.

5.2 Conclusions and discussion

This thesis has presented work that aims to maximise the scientific potential of the \textit{XMM}-Newton science archive, and to fully exploit it to measure X-ray properties and masses of galaxy clusters, as well as making X-ray analysis easy and accessible to everyone. The \textit{XMM} Cluster Survey’s processed version of the \textit{XMM} archive has been pivotal in this work, along with samples of galaxy clusters selected from the SDSS, DES, and ACT.

In Chapter 2 a new open-source Python module for the large-scale exploitation of X-ray astrophysics archival data, \textit{X-ray: Generate and Analyse} (\textit{XGA}) was introduced. This is one of the most significant outputs of this PhD, and provides extremely flexible software tools for the analysis of samples of hundreds of objects observed by \textit{XMM}, either completely interactively or in an automated, hands off fashion. The design, structure, and capabilities of the various parts of \textit{XGA} are explained in detail, as well as the underlying methodologies and new techniques developed for \textit{XGA}. While this module was developed to fulfil the scientific goals of this PhD, namely to be able to measure the properties (and most importantly masses) of samples of galaxy clusters, it is completely generalised, and is capable of analysing any type of X-ray emitting source (see Appendix A.1 for an example). It is only the second code capable of measuring galaxy cluster hydrostatic masses to be made publicly available; the only other being CLMASS (Nulsen et al., 2010), which is no longer actively maintained, and only provides mass fitting in \textit{XSPEC} without the supporting infrastructure and ability to generate annular spectra that \textit{XGA} provides. \textit{XGA} is a unique and powerful tool which will allow the rapid X-ray analysis follow-up of the many objects located and studied by the large-scale surveys (such as LSST) that will begin in the next few years. It has also powered the majority of research presented in this thesis, as well as supported some work by other research groups.

Chapter 3 \textit{XGA} was applied to three samples of galaxy clusters, the SDSSRM-XCS sample (Sec-
tion 1.6.1), the XXL-100-GC sample (Section 3.2.3), and finally the LoCuSS High-$L_X$ sample (Section 3.2.4). Initially they are used to validate XGA galaxy cluster property measurements, proving the veracity of XGA measurements and that it is suitable to apply to the study of galaxy clusters. First of all, comparisons made with the SDSSRM-XCS sample’s published $T_X$ and $L_X$ values (Giles et al., 2022b) demonstrate that XGA is consistent with XCS work, as we would expect considering essentially the same event lists and region files were used. The LoCuSS High-$L_X$ and XXL-100-GC samples are also used to compare temperatures and luminosities, demonstrating XGA’s flexibility by matching the original analysis choices of those samples, and demonstrating good agreement between temperatures, and general agreement between luminosities, though with offsets that are likely the result of different background regions. Next the LoCuSS High-$L_X$ and XXL-100-GC samples are used to test XGA measurements of gas mass, finding general agreement for XXL-100-GC (though with a similar trend in offset to that shown by the XXL luminosity comparison), and while XGA LoCuSS High-$L_X$ gas masses measured within literature $R_{500}$ agree with the original analysis, XGA measures slightly larger $R_{2500}$ gas masses. LoCuSS hydrostatic masses are then compared and found to be in broad agreement with XGA, though again XGA tends to measures slightly high masses than LoCuSS. Finally, new measurements of mass for 102 clusters from the SDSSRM-XCS sample, as well as new scaling relations constructed with those masses, are presented. Then measurements of mass (and other properties) of clusters from the DESY3RM-XCS cluster sample, and the newly constructed ACTDR5-XMM sample are presented, with 87 masses for DES and 245 for ACTDR5. The total number of unique masses measured for the three samples is 334, accounting for overlap in the contents of the samples; this combined sample is by far the largest collection of hydrostatic masses ever measured, with just the ACTDR5 masses being a larger sample than previous work by Lovisari et al. (2020), who measured 120 masses.

Finally, Chapter 4 describes work performed to demonstrate the veracity of galaxy cluster analyses performed on data from the eROSITA commissioning survey. This was motivated by a desire to a) assess the sample contamination and ability of eROSITA to detect clusters, and b) to investigate how measurements of cluster properties differ between XMM and eROSITA. We were the first team to provide external validation of eFEDS analyses, largely due to the speed at which a sub-sample of eFEDS selected clusters could be created and then analysed with XGA. Sixty-two eFEDS cluster candidates fell on XMM observation, and a detailed examination of their XMM, eROSITA, and HSC-SSP images helped to locate some rare failure modes of the eFEDS cluster finder, as well as confirming the eFEDS measurement of the contamination fraction. Once a sample of confirmed clusters was constructed through visual inspection, XGA was used to measure cluster $T_X$ and $L_X$ values. Excellent agreement between eFEDS and XGA luminosities was found, despite the
differences in source regions and technique. An offset between \textit{XMM} and \textit{eROSITA} temperatures was also discovered (\textit{XMM} temperatures were 25\% hotter), though the small sample size, the differences in analysis techniques, and previous evidence of \textit{eROSITA} temperature being hotter than \textit{XMM} mean that the offset may not be due to calibration errors in instruments.

\section*{5.3 Future work}

Each chapter contains discussion of what future steps could be taken to improve analyses, or achieve new results related to the contents of the chapter. However, short summaries of thoughts and plans for future work will be presented here.

While \textit{XGA} is a fully-featured and very powerful piece of software, there are several plans for immediate improvement; whether that’s increasing the sophistication of existing analyses, or adding entire new sets of features. The plan with the most potential for widespread impact and usefulness is to support the analysis of data from X-ray telescopes other than \textit{XMM}. This will extend the same convenience of quickly searching for serendipitous detections that \textit{XGA} provides for \textit{XMM}, as well as setting up the infrastructure for exciting new joint analyses. Plans are also in place for the generation and fitting of 2D spectral property maps, and the generation and analysis of time-domain data products such as light curves (extending \textit{XGA}'s usefulness).

The work on measuring cluster masses demonstrated the power of \textit{XGA} applied to galaxy cluster analysis, and will enable a significant amount of follow-up work. Some of that will be focused on examining the sources of error and systematic effects associated with the measurement of hydrostatic masses, and will fold in the artificial observations of simulated cluster work discussed in Section 5.1.1. The hydrostatic mass analyses implemented in \textit{XGA} will also be enhanced, with support for corrections taking into account non-thermal pressure support. \textit{XGA} will also be used to provide hydrostatic mass measurements for clusters selected by LSST.

Finally, there is a great deal of potential for more work related to the eFEDS sample; not only is the pathfinder sample for a mission that will produce tens of thousands of X-ray galaxy cluster measurements, but the limited size of our sample left several questions. The primary subject that will be the focus of future work is the temperature calibration of \textit{XMM} and \textit{eROSITA}. A much more detailed comparison could be made, even with the current data, particularly with a spectral reanalysis of \textit{eROSITA} data; this will be made very easy after \textit{eROSITA} support has been implemented in \textit{XGA}. 


A census of cool-core galaxy clusters in IllustrisTNG. *MNRAS*, 481(2):1809–1831. 192


Klein, M., Oguri, M., Mohr, J. J., Grandis, S., Ghirardini, V., Liu, T., Liu, A., Bulbul, E., Wolf,


Lloyd-Davies, E. J., Romer, A. K., Mehrtens, N., Hosmer, M., Davidson, M., Sabirli, K., Mann,


model of the point spread functions of the XMM-Newton EPIC telescopes: spurious source suppression and improved positional accuracy. A&A, 534:A34. 72, 119, 194


Exploring scaling relations and completeness of the Dark Energy Survey Year 3 RedMaPPer cluster catalogue. 28


van Weeren, R. J., de Gasperin, F., Akamatsu, H., Brüggen, M., Feretti, L., Kang, H., Stroe,


Appendix A: Contributions to other work

This appendix gives some information on some work that I contributed to during the course of the PhD, but did not lead or play a very significant role in. Each of these projects has resulted in a paper that has either been accepted and published, or has been submitted to a journal.

A.1 X-ray confirmation of low-mass AGN candidates

The Python module X-ray: Generate and Analyse (XGA) that was developed as part of this PhD is a completely generalised X-ray analysis package, as detailed in Chapter 2. As such it is not restricted only to the analysis of galaxy clusters, but can be used for other types of X-ray source as well. As such we used XGA to provide X-ray confirmation of 11 dwarf AGN candidates selected from the Dark Energy Survey (Burke et al., 2021). This was a simple analysis that consisted of checking whether the candidates had any XMM data, and whether they were detected by XAPA. We found that all 11 of the candidates fell upon an XMM observation, and six had an XCS point source detection from XAPA. One AGN candidate with no XCS detection appeared to coincide with point-like emission in stacked XMM observations generated by XGA, and was detected by another XMM serendipitous survey Webb et al. (2020) which uses a different energy band (XCS performs source detection on 0.5-2.0 keV, optimised for cluster finding). X-ray luminosities were also measured for several of the candidates by fitting an absorbed power-law to spectra, but these results were not included in the paper.

A.2 Dynamical state of a cluster using MeerKAT and XMM

We also used XGA to provide X-ray analysis for a galaxy cluster being studied using the Square Kilometre Array (SKA) pre-cursor telescope MeerKAT. The cluster in question was selected from the ACT-DR5 cluster catalogue, and follow-up observations using MeerKAT were performed. The
The goal of the work was to assess the dynamical state of the galaxy cluster. XGA was initially used to assess which objects in a sample of MeerKAT galaxy clusters had XMM observations, which then helped to inform the decision of which galaxy cluster to focus on in the work; ACT-CL J0019.6+0336 (Pillay et al., 2021).

Stacked X-ray images created by XGA were then combined with XGA masks to remove contaminating sources, and the resulting masked images were smoothed. The smoothed images were then used to measure several X-ray cluster morphology metrics; this includes the concentration parameter (a measure of how much emission is concentrated in the centre of the cluster), centroid shift (difference between the X-ray peak and X-ray centroids), and X-ray power ratios (multipole decomposition of the X-ray surface brightness image). We concluded that ACT-CL J0019.6+0336 is likely to be in a post-merging phase of evolution, with sub-structure indicated in the X-ray and Dark Energy Survey photometry.

### A.3 Upper limit X-ray luminosities of galaxy clusters

Part of this PhD involved developing a new pipeline (which will be integrated into XGA) to measure galaxy cluster luminosities from X-ray images, rather than XCS’ typical approach of using fitted spectral models. Fitting models to spectra is preferable in some ways, because it also constrains the temperature of the galaxy cluster. Luminosity measurements from X-ray photometry can be performed when a cluster is not detected however (giving an upper-limit on the luminosity), thus the required data quality is lower. Given a galaxy cluster detection in another wavelength, with an accompanying redshift estimate, an emission model can be assumed to convert the count rate within an analysis region to a luminosity. Knowledge of the redshift is vital, as that (along with a chosen cosmology), allow for the distance to the cluster to be calculated. The assumed emission model (APEC) is multiplied by an absorbing model (tbabs), with neutral hydrogen column density set to the value at the cluster coordinates as measured by HI4PI Collaboration et al. (2016).

This pipeline was used to provide upper limit luminosity measurements for galaxy clusters detected by redMaPPer run on the DESY3, that also fall on XMM observation but are not detected by XAPA (Wetzell et al., 2022). A similar set of undetected galaxy clusters will also be presented in a DESY3 paper on the X-ray properties of DES clusters (Jeltema et al., prep). Finally, this pipeline was used to provide upper limit X-ray luminosities of galaxy clusters in the XXL fields detected by GAMA (Giles et al., 2022a).
Appendix B: eFEDS-XMM cluster properties

B.1 Excluded Cluster Candidates

This section details the eFEDS-XMM X-ray cluster candidates that were not included in eFEDS-XCS sample, as discussed in Section 4.3. Basic information about the samples used in this work is available in Table 1. Table 1 contains candidates that were not included due to the low quality of the XMM data available, Table 3 contains galaxy clusters whose X-ray emission has been significantly contaminated by another X-ray source and as such were not included in the eFEDS-XCS sample. Table 2 contains sample contaminants that were not included in the eFEDS-XCS sample (see Section 4.3.2).

B.2 eFEDS Candidate 1023

This galaxy cluster has been been split into two sources by the eFEDS source finder; a visual inspection confirmed a single extended X-ray source (in both eROSITA and XMM images) and a single projected distribution of red galaxies (Section 4.3.3.3). XCS also detected this as a single extended source.

One of the two eFEDS catalogue entries that make up this cluster has measured eROSITA $T_X$ and $L_X$ values. These values will be impacted by the masking of emission from the other component. Therefore it would not be appropriate to include those values in the comparisons presented in Figures 4.10 or 4.12.

However, we have attempted to mimic the eROSITA values using XMM data. This involves manually adding a region to be excluded when the XMM spectra are generated. This region is centered on the eFEDS X-ray candidate catalogue coordinates for eFEDS-8602, and uses the ‘extent’ value for that candidate published in the optical counterpart catalogue as the radius of the new exclusion.
Table 1: eFEDS-XMM galaxy cluster candidates excluded from further analysis due to one or more XMM-Newton data quality issues. The EL and DL columns correspond to the extent likelihood (EXT_LIKE) and detection likelihood (DET_LIKE) columns in the eFEDS catalogue. † indicates that the candidate was present in the optically confirmed sample from Klein et al. (2021).

<table>
<thead>
<tr>
<th>eFEDS ID</th>
<th>RA</th>
<th>Dec</th>
<th>z</th>
<th>EL</th>
<th>DL</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>8094†</td>
<td>133.644</td>
<td>-1.677</td>
<td>0.595</td>
<td>13.56</td>
<td>38.71</td>
<td>On the edge of the XMM field of view.</td>
</tr>
<tr>
<td>7700</td>
<td>133.669</td>
<td>-2.159</td>
<td>0.472</td>
<td>6.95</td>
<td>32.39</td>
<td>On the edge of the XMM field of view.</td>
</tr>
<tr>
<td>1797†</td>
<td>133.876</td>
<td>-1.11</td>
<td>0.754</td>
<td>40.55</td>
<td>74.33</td>
<td>On the edge of the XMM field of view (eROSITA image confirms presence of extended source).</td>
</tr>
<tr>
<td>11836†</td>
<td>135.272</td>
<td>-1.424</td>
<td>0.405</td>
<td>17.78</td>
<td>29.38</td>
<td>On the edge of the XMM field of view, and the XMM observation is shallow (1023 s exposure at the eFEDS coordinates). The eROSITA image confirms presence of an extended source.</td>
</tr>
<tr>
<td>9877†</td>
<td>136.04</td>
<td>0.642</td>
<td>0.311</td>
<td>12.22</td>
<td>18.62</td>
<td>Low signal-to-noise XMM image (7489 s exposure at the eFEDS coordinates) which has been affected by flaring.</td>
</tr>
<tr>
<td>2757†</td>
<td>134.756</td>
<td>1.114</td>
<td>0.162</td>
<td>33.65</td>
<td>97.42</td>
<td>Low signal-to-noise XMM image (6870 s exposure at the eFEDS coordinates).</td>
</tr>
<tr>
<td>5858</td>
<td>136.687</td>
<td>1.19</td>
<td>0.441</td>
<td>9.78</td>
<td>60.28</td>
<td>Low signal-to-noise XMM image (15020 s exposure at the eFEDS coordinates). eROSITA image confirms presence of extended source.</td>
</tr>
<tr>
<td>2074†</td>
<td>136.971</td>
<td>1.569</td>
<td>0.163</td>
<td>9.43</td>
<td>17.90</td>
<td>Low signal-to-noise XMM image (5331 s exposure at the eFEDS coordinates).</td>
</tr>
<tr>
<td>11837†</td>
<td>138.201</td>
<td>0.413</td>
<td>0.308</td>
<td>7.06</td>
<td>17.43</td>
<td>Low signal-to-noise XMM image (29479 s exposure at the eFEDS coordinates) which has been affected by flaring. The eROSITA image confirms presence of an extended source.</td>
</tr>
<tr>
<td>1376†</td>
<td>133.23</td>
<td>-1.627</td>
<td>0.343</td>
<td>7.35</td>
<td>14.90</td>
<td>XMM data too shallow for confirmation. The eROSITA and SDSS data indicate a likely cluster, but eFEDS coordinate is offset from the extended emission and SDSS galaxies.</td>
</tr>
</tbody>
</table>
Table 2: eFEDS-XMM galaxy cluster candidates classed as contaminants during our visual inspection of XMM, eROSITA, and SDSS images. The EL and DL columns correspond to the extent likelihood (\textit{EXT\_LIKE}) and detection likelihood (\textit{DET\_LIKE}) columns in the eFEDS catalogue. † indicates that the candidate was present in the optically confirmed sample from Klein et al. (2021).

<table>
<thead>
<tr>
<th>eFEDS ID</th>
<th>RA</th>
<th>Dec</th>
<th>z</th>
<th>EL</th>
<th>DL</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1644†</td>
<td>130.396</td>
<td>1.031</td>
<td>0.507</td>
<td>16.55</td>
<td>139.61</td>
<td>Blend: In XMM image two point sources are detected, due to XMM’s smaller PSF effect. The source is the target of the XMM observation and is associated with an interacting pair of active galaxies. (see Figure 4.5).</td>
</tr>
<tr>
<td>3334†</td>
<td>130.508</td>
<td>0.995</td>
<td>0.087</td>
<td>7.52</td>
<td>9.95</td>
<td>Spurious: There is not an X-ray source at this location in either the XMM or eROSITA images, nor do there appear to be any associated galaxies in the SDSS/HSC images.</td>
</tr>
<tr>
<td>8602†</td>
<td>132.593</td>
<td>0.269</td>
<td>0.196</td>
<td>13.40</td>
<td>18.24</td>
<td>Fragmented: The ICM emission from a single cluster that has been classified as coming from two eFEDS candidates, ID 8602 and 1023.</td>
</tr>
<tr>
<td>5909†</td>
<td>133.83</td>
<td>-1.721</td>
<td>0.365</td>
<td>12.90</td>
<td>42.53</td>
<td>Spurious: There is an X-ray source at the eFEDS candidate location, but it is a defined as point source by XCS in the higher signal to noise XMM data. There are no associated galaxies in the SDSS or HSC images.</td>
</tr>
<tr>
<td>8922†</td>
<td>134.067</td>
<td>-1.663</td>
<td>0.514</td>
<td>10.56</td>
<td>22.47</td>
<td>Spurious: In eFEDS, this is a spurious detection of the outskirts of the emission from an X-ray bright spiral galaxy. In the higher resolution XMM image, there is no source at this location.</td>
</tr>
<tr>
<td>9463</td>
<td>136.753</td>
<td>1.176</td>
<td>0.799</td>
<td>10.29</td>
<td>23.07</td>
<td>Blend: In higher signal to noise (18501 s exposure) XMM image, two point sources detected.</td>
</tr>
<tr>
<td>13484</td>
<td>136.766</td>
<td>1.132</td>
<td>0.307</td>
<td>7.90</td>
<td>12.57</td>
<td>Spurious: There is no obvious extended X-ray emission in either the eROSITA or XMM data. There are no associated galaxies in the SDSS or HSC images.</td>
</tr>
<tr>
<td>13299</td>
<td>138.691</td>
<td>0.439</td>
<td>0.348</td>
<td>14.78</td>
<td>17.75</td>
<td>Spurious: In eFEDS, this is a spurious detection of the outskirts of the emission from an X-ray bright star.</td>
</tr>
<tr>
<td>11754</td>
<td>140.018</td>
<td>1.007</td>
<td>0.033</td>
<td>19.95</td>
<td>35.88</td>
<td>Spurious: Spurious detection in outskirts of nearby eFEDS candidate ID 150 (z=0.017).</td>
</tr>
<tr>
<td>5702</td>
<td>130.295</td>
<td>0.867</td>
<td>0.415</td>
<td>7.13</td>
<td>39.44</td>
<td>Spurious: In the higher signal to noise (130445 s exposure) XMM image, a point source is detected which appears to be associated with a blue (i.e. likely AGN) object in the SDSS and HSC images.</td>
</tr>
<tr>
<td>6840</td>
<td>135.597</td>
<td>1.868</td>
<td>0.561</td>
<td>6.75</td>
<td>43.23</td>
<td>Spurious: In the XMM image, XCS detects a point source which appears to be associated with a blue (i.e. likely AGN) object in the SDSS and HSC images.</td>
</tr>
</tbody>
</table>
Table 3: eFEDS-\textit{XMM} galaxy cluster candidates which appear to be galaxy clusters whose X-ray emission is significantly contaminated by another source. The EL and DL columns correspond to extent (\texttt{EXT\_LIKE}) and detection likelihood (\texttt{DET\_LIKE}) columns in the eFEDS catalogue. † indicates that the candidate was present in the optically confirmed sample from Klein et al. (2021).

<table>
<thead>
<tr>
<th>eFEDS ID</th>
<th>RA</th>
<th>Dec</th>
<th>z</th>
<th>EL</th>
<th>DL</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>16370†</td>
<td>134.098</td>
<td>-1.604</td>
<td>0.425</td>
<td>11.86</td>
<td>17.74</td>
<td>eFEDS candidate coincident with a collection of galaxies in SDSS/HSC; however X-ray emission is contaminated by low redshift spiral galaxy.</td>
</tr>
<tr>
<td>150†</td>
<td>140.009</td>
<td>1.039</td>
<td>0.017</td>
<td>179.59</td>
<td>1049.64</td>
<td>SDSS/HSC indicates the presence of a group of galaxies, however, the X-ray emission originates primarily from the central galaxy.</td>
</tr>
<tr>
<td>3133†</td>
<td>140.649</td>
<td>-0.412</td>
<td>0.055</td>
<td>30.21</td>
<td>49.42</td>
<td>SDSS/HSC indicates the presence of a group of galaxies, however, the X-ray emission originates primarily from the central galaxy.</td>
</tr>
<tr>
<td>3008†</td>
<td>130.451</td>
<td>0.82</td>
<td>0.078</td>
<td>16.78</td>
<td>80.87</td>
<td>SDSS/HSC indicates the presence of a group of galaxies, however, the X-ray emission originates primarily from the central galaxy.</td>
</tr>
</tbody>
</table>
region. Doing this, we find the \textit{XMM} determined $L_X$ and $T_X$ values are consistent with those presented in Liu et al. (2021a); see Figures 4.10 and 4.12 (cyan diamond).

\section*{B.3 eFEDS-XCS Data and Measurements}

In Table 4 we present information on the \textit{XMM} data that were used for each eFEDS-XCS cluster, including the unique \textit{XMM} observation identifier and which instruments had usable data. We also include information on which instruments of which observations were contributed to the final luminosity and temperature measurements of each eFEDS-XCS cluster. In Table 5 we present temperature and luminosity values measured for the clusters in the eFEDS-XCS sample. We use XGA to generate spectra and run XSPEC fits for these clusters. The fitting procedure is discussed in more detail in Section 4.4.1. All measurements are centered on the eFEDS coordinates from the X-ray cluster candidate catalogue, with redshift information also taken from that catalogue.
Table 4: The *XMM* data used in the analysis of the eFEDS-XCS sample, individual clusters denoted by their unique eFEDS ID. ObsID contains the unique identifier(s) of the *XMM* observation(s) used. T denotes true, F denotes false, - denotes that either no successful spectral fit was performed, or the data for that cameras was not available. Columns with a subscript A (e.g. PN$_A$) indicate whether that instrument is available for an ObsID. Columns with a subscript radius (e.g. PN$_{500\text{kpc}}$) indicate whether that instrument’s data contributed to the final XSPEC fit from which we extract temperature and luminosity information.

<table>
<thead>
<tr>
<th>eFEDS ID</th>
<th>ObsID</th>
<th>PN$_A$</th>
<th>MOS1$_A$</th>
<th>MOS2$_A$</th>
<th>PN$_{500\text{kpc}}$</th>
<th>MOS1$_{500\text{kpc}}$</th>
<th>MOS2$_{500\text{kpc}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6605</td>
<td>0202940101</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0202940201</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>144</td>
<td>0650381601</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>7831</td>
<td>0784350101</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>1023</td>
<td>0761730501</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>6125</td>
<td>0761730501</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>339</td>
<td>0655340137</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>4810</td>
<td>0655340137</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>1458</td>
<td>0655340135</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>2079</td>
<td>0651170301</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>569</td>
<td>0783881001</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>1385</td>
<td>0783881001</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>8857</td>
<td>0725290142</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3171</td>
<td>0725300134</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725290144</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725290145</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>8881</td>
<td>0725290139</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725290146</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>1104</td>
<td>0655340160</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0804410201</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>4232</td>
<td>0655340160</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0804410201</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>5655</td>
<td>0725300158</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725300159</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725300136</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>1712</td>
<td>0725300140</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725300132</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725300131</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>eFEDS ID</td>
<td>ObsID</td>
<td>PN</td>
<td>MOS1</td>
<td>MOS2</td>
<td>PN500kpc</td>
<td>MOS1500kpc</td>
<td>MOS2500kpc</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>5774</td>
<td>0725300157</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0725310131</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3590</td>
<td>0725310152</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725300160</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12660</td>
<td>0725290131</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725290154</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>3585</td>
<td>0725310149</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725310150</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725310131</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5170</td>
<td>0725300152</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725300153</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>9359</td>
<td>0725300145</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725300146</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>3259</td>
<td>0725300144</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725300151</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>7086</td>
<td>0725310133</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0725310147</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0725310148</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0725310157</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0402780801</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5219</td>
<td>0725310158</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>0725310147</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>0725310159</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>7084</td>
<td>0725310157</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>885</td>
<td>0725310142</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0725310141</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>2004</td>
<td>0800400501</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>3523</td>
<td>0800400501</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>372</td>
<td>0602830401</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4253</td>
<td>0602830401</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>-</td>
<td>T</td>
</tr>
<tr>
<td>534</td>
<td>0804410101</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>0650381801</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>100</td>
<td>0804410501</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>857</td>
<td>0823710301</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>12565</td>
<td>0802220601</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 5: eFEDS-XCS galaxy cluster XGA measured values, RA, Dec, and redshift are taken from the eFEDS X-ray cluster candidate catalogue. $T_{x,500kpc}^{XGA}$ are temperatures within 500 kpc apertures, given in keV. $L_{x,500kpc}^{XGA,52}$ and $L_{x,500kpc}^{XGA,bol}$ are 0.5-2.0 keV bolometric luminosities within a 500 kpc apertures, in units of $10^{43}$ erg s$^{-1}$. All uncertainties calculated from 68% confidence limits, equivalent to 1σ.

<table>
<thead>
<tr>
<th>eFEDS ID</th>
<th>RA</th>
<th>Dec</th>
<th>z</th>
<th>$T_{x,500kpc}^{XGA}$</th>
<th>$L_{x,500kpc}^{XGA,52}$</th>
<th>$L_{x,500kpc}^{XGA,bol}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6605</td>
<td>130.353</td>
<td>0.777</td>
<td>0.41</td>
<td>1.61$^{+0.17}_{-0.13}$</td>
<td>2.32$^{+0.16}_{-0.17}$</td>
<td>4.81$^{+0.46}_{-0.41}$</td>
</tr>
<tr>
<td>144</td>
<td>131.37</td>
<td>3.461</td>
<td>0.33</td>
<td>5.8$^{+0.59}_{-0.48}$</td>
<td>19.18$^{+0.84}_{-0.67}$</td>
<td>67.1$^{+3.88}_{-3.88}$</td>
</tr>
<tr>
<td>7831</td>
<td>132.272</td>
<td>2.243</td>
<td>0.4</td>
<td>1.61$^{+0.11}_{-0.11}$</td>
<td>2.82$^{+0.21}_{-0.15}$</td>
<td>5.87$^{+0.44}_{-0.43}$</td>
</tr>
<tr>
<td>1023†</td>
<td>132.616</td>
<td>0.251</td>
<td>0.2</td>
<td>3.08$^{+0.14}_{-0.13}$</td>
<td>3.55$^{+0.04}_{-0.05}$</td>
<td>9.29$^{+0.25}_{-0.22}$</td>
</tr>
<tr>
<td>6125</td>
<td>132.627</td>
<td>0.558</td>
<td>0.19</td>
<td>2.39$^{+0.33}_{-0.31}$</td>
<td>0.98$^{+0.05}_{-0.05}$</td>
<td>2.35$^{+0.2}_{-0.18}$</td>
</tr>
<tr>
<td>339</td>
<td>133.071</td>
<td>-1.025</td>
<td>0.46</td>
<td>5.57$^{+0.81}_{-0.63}$</td>
<td>18.3$^{+0.71}_{-0.76}$</td>
<td>62.82$^{+4.09}_{-6.48}$</td>
</tr>
<tr>
<td>4810</td>
<td>133.13</td>
<td>-1.208</td>
<td>0.55</td>
<td>2.06$^{+2.12}_{-0.72}$</td>
<td>4.56$^{+1.06}_{-1.85}$</td>
<td>10.14$^{+3.84}_{-4.1}$</td>
</tr>
<tr>
<td>1458</td>
<td>133.554</td>
<td>-2.357</td>
<td>0.38</td>
<td>4.47$^{+0.78}_{-0.72}$</td>
<td>11.59$^{+0.53}_{-0.75}$</td>
<td>35.63$^{+3.34}_{-4.39}$</td>
</tr>
<tr>
<td>2079</td>
<td>133.696</td>
<td>-1.359</td>
<td>0.35</td>
<td>3.27$^{+0.56}_{-0.38}$</td>
<td>4.59$^{+0.23}_{-0.27}$</td>
<td>12.23$^{+0.86}_{-0.98}$</td>
</tr>
<tr>
<td>569</td>
<td>134.086</td>
<td>1.78</td>
<td>0.72</td>
<td>4.94$^{+0.16}_{-0.16}$</td>
<td>29.96$^{+0.55}_{-0.44}$</td>
<td>96.29$^{+2.02}_{-1.94}$</td>
</tr>
<tr>
<td>1385</td>
<td>134.113</td>
<td>1.705</td>
<td>0.73</td>
<td>6.47$^{+1.49}_{-1.11}$</td>
<td>7.06$^{+0.4}_{-0.37}$</td>
<td>26.9$^{+2.66}_{-3.5}$</td>
</tr>
<tr>
<td>8857</td>
<td>134.658</td>
<td>1.449</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3171</td>
<td>135.269</td>
<td>1.279</td>
<td>0.25</td>
<td>1.7$^{+0.24}_{-0.2}$</td>
<td>2.64$^{+0.26}_{-0.63}$</td>
<td>5.69$^{+0.76}_{-1.3}$</td>
</tr>
<tr>
<td>8881</td>
<td>135.314</td>
<td>0.844</td>
<td>0.31</td>
<td>3.27$^{+4.56}_{-1.79}$</td>
<td>1.13$^{+0.34}_{-0.36}$</td>
<td>3.02$^{+1.2}_{-1.7}$</td>
</tr>
<tr>
<td>1104</td>
<td>135.372</td>
<td>-1.648</td>
<td>0.31</td>
<td>5.96$^{+0.32}_{-0.32}$</td>
<td>10.94$^{+1.11}_{-0.81}$</td>
<td>38.75$^{+4.67}_{-4.06}$</td>
</tr>
<tr>
<td>4232</td>
<td>135.443</td>
<td>-1.632</td>
<td>0.29</td>
<td>1.95$^{+0.53}_{-0.3}$</td>
<td>0.99$^{+0.1}_{-0.14}$</td>
<td>2.19$^{+0.36}_{-0.33}$</td>
</tr>
<tr>
<td>5655</td>
<td>135.735</td>
<td>1.774</td>
<td>0.12</td>
<td>0.97$^{+0.11}_{-0.12}$</td>
<td>0.02$^{+0.11}_{-0.02}$</td>
<td>0.04$^{+0.19}_{-0.04}$</td>
</tr>
<tr>
<td>1712</td>
<td>135.74</td>
<td>0.805</td>
<td>0.52</td>
<td>2.08$^{+1.74}_{-0.63}$</td>
<td>6.12$^{+2.32}_{-1.85}$</td>
<td>13.68$^{+4.61}_{-4.53}$</td>
</tr>
<tr>
<td>5774</td>
<td>136.036</td>
<td>1.432</td>
<td>0.84</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3590</td>
<td>136.078</td>
<td>2.112</td>
<td>0.81</td>
<td>4.99$^{+8.64}_{-2.3}$</td>
<td>10.17$^{+1.35}_{-3.57}$</td>
<td>32.84$^{+7.6}_{-13.35}$</td>
</tr>
<tr>
<td>12660</td>
<td>136.08</td>
<td>-1.077</td>
<td>0.31</td>
<td>2.68$^{+1.51}_{-0.81}$</td>
<td>0.81$^{+0.21}_{-0.14}$</td>
<td>2.01$^{+0.55}_{-0.47}$</td>
</tr>
<tr>
<td>3585</td>
<td>136.42</td>
<td>1.539</td>
<td>0.64</td>
<td>4.0$^{+1.46}_{-0.94}$</td>
<td>8.81$^{+2.0}_{-1.89}$</td>
<td>25.52$^{+6.27}_{-6.63}$</td>
</tr>
<tr>
<td>5170</td>
<td>136.473</td>
<td>0.379</td>
<td>0.37</td>
<td>1.22$^{+0.14}_{-0.18}$</td>
<td>2.33$^{+1.38}_{-0.46}$</td>
<td>4.42$^{+0.74}_{-0.94}$</td>
</tr>
<tr>
<td>9359</td>
<td>136.502</td>
<td>-0.423</td>
<td>0.3</td>
<td>1.99$^{+1.46}_{-0.51}$</td>
<td>1.0$^{+0.66}_{-1.4}$</td>
<td>2.3$^{+1.89}_{-1.4}$</td>
</tr>
<tr>
<td>3259</td>
<td>136.504</td>
<td>0.015</td>
<td>0.2</td>
<td>2.53$^{+1.06}_{-0.59}$</td>
<td>1.22$^{+0.33}_{-0.19}$</td>
<td>2.96$^{+0.87}_{-0.68}$</td>
</tr>
<tr>
<td>7086</td>
<td>136.654</td>
<td>1.148</td>
<td>0.79</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5219</td>
<td>136.977</td>
<td>0.961</td>
<td>0.74</td>
<td>4.74$^{+2.17}_{-1.44}$</td>
<td>9.92$^{+1.76}_{-1.6}$</td>
<td>31.21$^{+5.25}_{-7.5}$</td>
</tr>
<tr>
<td>7084</td>
<td>137.025</td>
<td>1.331</td>
<td>0.66</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>885</td>
<td>137.308</td>
<td>-0.204</td>
<td>0.31</td>
<td>2.99$^{+0.32}_{-0.37}$</td>
<td>5.65$^{+0.49}_{-0.44}$</td>
<td>14.52$^{+1.38}_{-1.55}$</td>
</tr>
<tr>
<td>eFEDS ID</td>
<td>RA</td>
<td>Dec</td>
<td>z</td>
<td>$T_{\text{XGA,500kpc}}$</td>
<td>$L_{\text{XGA,52,500kpc}}$</td>
<td>$L_{\text{XGA, bol,500kpc}}$</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>-----</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>2004</td>
<td>137.314</td>
<td>-1.018</td>
<td>0.82</td>
<td>$5.57^{+4.09}_{-1.96}$</td>
<td>$6.17^{+2.72}_{-3.96}$</td>
<td>$21.12^{+10.09}_{-10.65}$</td>
</tr>
<tr>
<td>3523</td>
<td>137.386</td>
<td>-0.839</td>
<td>1.13</td>
<td>$5.61^{+2.19}_{-1.11}$</td>
<td>$26.33^{+2.91}_{-3.09}$</td>
<td>$93.83^{+11.49}_{-17.28}$</td>
</tr>
<tr>
<td>372</td>
<td>138.723</td>
<td>4.27</td>
<td>0.14</td>
<td>$2.46^{+0.84}_{-0.67}$</td>
<td>$1.8^{+0.11}_{-0.09}$</td>
<td>$4.35^{+0.5}_{-0.4}$</td>
</tr>
<tr>
<td>4253</td>
<td>138.801</td>
<td>4.585</td>
<td>0.36</td>
<td>$5.18^{+5.18}_{-3.76}$</td>
<td>$1.18^{+0.3}_{-1.18}$</td>
<td>$3.96^{+5.96}_{-1.84}$</td>
</tr>
<tr>
<td>534</td>
<td>139.042</td>
<td>-0.397</td>
<td>0.32</td>
<td>$6.04^{+0.35}_{-0.35}$</td>
<td>$13.86^{+0.32}_{-0.34}$</td>
<td>$49.42^{+1.96}_{-1.85}$</td>
</tr>
<tr>
<td>100</td>
<td>140.338</td>
<td>3.291</td>
<td>0.33</td>
<td>$7.83^{+0.31}_{-0.31}$</td>
<td>$19.66^{+0.25}_{-0.23}$</td>
<td>$79.71^{+0.85}_{-1.93}$</td>
</tr>
<tr>
<td>857</td>
<td>140.55</td>
<td>-0.459</td>
<td>0.32</td>
<td>$5.86^{+0.24}_{-0.23}$</td>
<td>$9.51^{+0.13}_{-0.11}$</td>
<td>$33.41^{+0.66}_{-0.84}$</td>
</tr>
<tr>
<td>12565</td>
<td>142.473</td>
<td>0.467</td>
<td>0.15</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>