

This is a pre print version of the following article:

CHEX-MATE: Are we getting cluster thermodynamics right? / Seppi, R., Eckert, D., Rasia, E., Kay, S.T., Dolag, K., Biffi, V., Bahar, Y.E., Bourdin, H., De Luca, F., De Petris, M., Ettori, S., Gaspari, M., Gastaldello, F., Ghirardini, V., Lovisari, L., Mazzotta, P., Pratt, G.W., Pointecouteau, E., Rossetti, M., Sayers, J., et al.. - In: ASTRONOMY & ASTROPHYSICS. - ISSN 0004-6361. - (2026), pp. 1-19. [10.48550/arxiv.2603.03440]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

25/08/2026 02:14

CHEX-MATE: Are we getting cluster thermodynamics right?

R. Seppi^{1*}, D. Eckert¹, E. Rasia^{2,3}, S. T. Kay⁴, K. Dolag^{5,6,7}, V. Biffi^{2,8},
Y. E. Bahar⁹, H. Bourdin^{17,18}, F. De Luca^{17,18}, M. De Petris¹⁴, S. Ettori^{9,10}, M. Gaspari¹⁵, F. Gastaldello¹³, V.
Ghirardini⁹, L. Lovisari^{13,16}, P. Mazzotta^{17,18}, G. W. Pratt¹¹, E. Pointecouteau¹², M. Rossetti¹³, J. Sayers¹⁹, M.
Sereni^{9,10}, and G. Yepes^{20,21}

(Affiliations can be found after the references)

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Context. Galaxy clusters offer powerful insights into the large-scale structure of the Universe and the physics of baryons in hot state. Their scientific exploitation, however, hinges on our ability to accurately measure key thermodynamic properties.

Aims. In this work, we aim to assess the reliability of current analysis techniques in reconstructing these properties, with particular focus on samples similar to those observed in the Cluster HERitage project with XMM-Newton (CHEX-MATE).

Methods. We develop a suite of dedicated end-to-end simulations of CHEX-MATE-like clusters selected from large scale hydrodynamical simulations, and processed through a newly developed realistic XMM-Newton simulator. We apply a full X-ray data analysis pipeline to the mock datasets, including imaging, spectral fitting, and profile reconstruction.

Results. The gas density profiles can be robustly recovered across a wide radial range, when using azimuthal mean surface brightness profiles. Our reconstruction techniques are able to reproduce the intrinsic density profile with the correct scatter, with deviations of at most 10% between 0.1 and $1 \times R_{500c}$. The gas mass is reconstructed with better than 1% accuracy. Accurate measurement of temperature profiles is more challenging and possibly subject to biases, particularly in the presence of azimuthal variations and multi-temperature gas along the line of sight, which dominate over projection effects.

Conclusions. Our results highlight the need for caution in interpreting cluster temperature measurements and underscore the value of tailored mock observations for understanding observational systematics. These findings also suggest that biases in X-ray temperature measurements may alter the interpretation of the thermodynamical state of the intra-cluster medium, an outlook particularly relevant in light of recent low velocity measurements from the XRISM mission.

Key words. X-rays: galaxies: clusters - Galaxies: clusters: intracluster medium - Surveys - Cosmology: large-scale structure of Universe - Methods: data analysis

1. Introduction

Clusters of galaxies are the end point of the structure formation process throughout the history of the Universe and are located in the nodes of the cosmic web (Mo & White 2002; Springel 2005). They encode precious cosmological information about dark matter, driving the shaping of the large scale structure of the Universe, and dark energy, driving its accelerated expansion at late times (Allen et al. 2004; Kravtsov & Borgani 2012; Clerc & Finoguenov 2022; Ghirardini et al. 2024; Lesci et al. 2025). Massive galaxy clusters benefit from high signal to noise ratio observations at various wavelengths. In the optical band they are seen as a collection of their galaxy members (Rykoff et al. 2014; Abbott et al. 2020). However, only about 1% of their total mass resides in the galaxy population. Instead, about 90% is in form of dark matter, whose gravitational effect shows up as peaks in weak lensing convergence maps (Miyazaki et al. 2018). Finally, the majority of baryons is located in the hot gas that constitutes the intra-cluster medium (ICM), heated up by the process of gravitational collapse to high temperatures around 10^8 K. This allows detecting clusters in the millimeter band via the Sunyaev-Zeldovich (SZ) effect (Staniszewski et al. 2009; Planck Collaboration et al. 2014a), and in X-rays thanks to direct emission via thermal bremsstrahlung (Böhringer et al. 2004; Pratt et al. 2019). SZ surveys such as the Atacama Cosmology Telescope (ACT,

Hilton et al. 2021), the South Pole Telescope (SPT, Bleem et al. 2020), and *Planck* (Planck Collaboration et al. 2014b) are sensitive to most massive clusters up to high redshift. X-ray surveys from ROSAT (Böhringer et al. 2004), XMM-XXL (Pierre et al. 2016), and eROSITA (Predehl et al. 2021; Bulbul et al. 2024) are better suited to detect the low mass cluster population, but their sensitivity quickly drops at high redshift. The combination of multi-wavelength data is essential to obtain a clear view of galaxy clusters in the Universe (e.g., Beauchesne et al. 2024).

The Cluster HERitage project with XMM-Newton: Mass Assembly and Thermodynamics at the Endpoint of structure formation (CHEX-MATE¹, CHEX-MATE Collaboration et al. 2021) is a Heritage program to follow up a sample of 118 galaxy clusters selected from *Planck* with X-ray observations using XMM-Newton. The program covers three mega-seconds, with a median exposure time of 40 ks per object. The goal is to study the final products of structure formation, focusing on the most recent objects formed in time (Tier-1: $0.05 < z < 0.2$, $2 \times 10^{14} < M_{500c}/M_{\odot} < 9 \times 10^{14}$), and the most massive clusters in the Universe (Tier-2: $z < 0.6$, $M_{500c} > 7.25 \times 10^{14} M_{\odot}$). CHEX-MATE aims to combine the deep XMM observations with archival and follow up lensing and SZ data to tackle some open questions in cluster science, such as mass calibration, different selection processes, or the evolution of cluster properties throughout cosmic time. The CHEX-MATE collaboration has

* E-mail: riccardo.seppi@unige.ch

¹ xmm-heritage.oas.inaf.it

already produced several results relevant to this work. [Campitiello et al. \(2022\)](#) derived a dynamical-state indicator from X-ray morphological parameters, finding that CHEX-MATE clusters are generally more disturbed than X-ray-selected samples. [Bartalucci et al. \(2023\)](#) showed that surface-brightness profiles exhibit large diversity in the cores but converge with minimal scatter at $0.4\text{--}0.8 R_{500c}$. Temperature profiles for a representative subsample were presented by [Rossetti et al. \(2024\)](#), demonstrating the potential for statistical studies once the full sample is analysed. A pilot study of entropy profiles by [Riva et al. \(2024\)](#) revealed correlations between core entropy and dynamical state and deviations from self-similar scaling. CHEX-MATE data quality also enables advanced mass-modelling approaches: [Kim et al. \(2024\)](#) combining XMM-Newton and *Planck* data to obtain a 3D elliptical models ([Kim et al. 2024](#); [Chappuis et al. 2025](#)), or from galaxy dynamics ([Sereno et al. 2025](#)). Finally, [Muñoz-Echeverría et al. \(2025\)](#) introduced a joint fitting of universal pressure profiles and cluster masses to break their degeneracy, a method well suited to CHEX-MATE.

Rigorous validation of the X-ray analysis pipeline further strengthens the accuracy and precision of the recovered observables and the cluster properties inferred from them. The need for tools capable of producing realistic mock X-ray observations from hydrodynamical simulations emerged as a crucial step toward bridging the gap between theory and observation. Early works ([Evrard et al. 1996](#); [Mathiesen & Evrard 2001](#)) already emphasized that meaningful comparisons with X-ray data require simulated quantities that incorporate projection and instrumental effects, rather than purely theoretical profiles. This concept was operationally realized with the development of dedicated X-ray map simulators, such as X-MAS ([Gardini et al. 2004](#)), which for the first time reproduced the full observing process—including telescope response, background noise, and photon statistics—yielding realistic event files for instruments such as Chandra and XMM-Newton. Subsequent studies (e.g. [Rasia et al. 2006](#); [Nagai et al. 2007](#); [Rasia et al. 2008](#); [Biffi et al. 2012](#)) demonstrated that such simulated observations are essential to quantify observational biases, test data analysis pipelines, and ensure that theoretical predictions and observed cluster properties are compared on consistent grounds. These developments have firmly established X-ray map simulators as a cornerstone in the modern analysis of the ICM. The literature is rich with such examples, especially in the X-ray band (see also [Lau et al. 2009](#); [Battaglia et al. 2013](#); [Nelson et al. 2014](#); [Rasia et al. 2015](#); [Biffi et al. 2016](#)), but also in combination with other probes like lensing ([Meneghetti et al. 2010, 2011](#); [Rasia et al. 2012](#); [Euclid Collaboration et al. 2024](#); [Giocoli et al. 2025](#)) or the SZ effect ([Gupta et al. 2017](#); [Gianfagna et al. 2021](#); [Wicker et al. 2023](#)). Most of these studies focus on recovering intrinsic properties such as halo mass, providing estimates of the hydrostatic mass bias ranging between 10-30% ([Gianfagna et al. 2023](#); [Jennings & Davé 2023](#); [Muñoz-Echeverría et al. 2024](#)), with the bias increasing with cluster mass ([Braspenning et al. 2025](#)). Other works also suggest that part of the mass bias is encoded in X-ray temperature measurements ([Henson et al. 2017](#); [Pearce et al. 2020](#); [Barnes et al. 2021](#)). They assume various levels of complexity concerning the realism of the X-ray mock, starting from the simplest approach of analysing the hot gas particles in hydrodynamical simulations with dark matter and baryons.

Following the examples of [Rasia et al. \(2008\)](#); [Biffi et al. \(2013\)](#), our goal is to take one step further by selecting twin CHEX-MATE samples from three different hydrodynamical simulations: The Three Hundred (The300, hereafter) project ([Cui et al. 2018](#)), Magneticum ([Dolag 2015](#)), and

Table 1. Cosmological and numerical parameters describing the The300, Magneticum, and MACSIS simulations. For Magneticum the numbers in parentheses refer to Box2b, the others to Box2.

	The300	Magneticum	MACSIS
Box [Gpc h^{-1}]	1.0	0.352 (0.640)	2.17
Ω_M	0.307	0.272	0.307
Ω_B	0.048	0.0456	0.04825
Ω_Λ	0.693	0.728	0.693
σ_8	0.823	0.809	0.8288
H_0	67.8	70.4	67.77
n_s	0.96	0.963	0.9611
N particles	3840 ³	2×1584 ³ (2×2880 ³)	2520 ³
M_{DM} [$M_\odot h^{-1}$]	1.27×10 ⁹	6.9×10 ⁸	4.4×10 ⁹
M_{gas} [$M_\odot h^{-1}$]	2.36×10 ⁸	1.4×10 ⁸	8.0×10 ⁸

Notes. Box size: comoving length of the box size covered by the simulation, Ω_M : total matter density parameter, Ω_B : baryonic matter density parameter, Ω_Λ : dark energy density parameter, σ_8 : normalization of the linear matter power spectrum, H_0 : Hubble constant, n_s : initial slope of the linear matter power spectrum, N particles: total number of dark matter particles in the simulation, M_{DM} : mass of the dark matter particles, M_{gas} : initial mass of the gas matter particles.

MACSIS ([Barnes et al. 2017](#)). We generated realistic end to end XMM-like mock observations and analysed them with standard tools and pipelines widely used in the X-ray community. The high resolution of the simulations, together with realistic X-ray processing, enables a robust comparison between recovered and input thermodynamic profiles. This is essential for assessing whether our models can accurately reproduce X-ray measurements and recover intrinsic cluster properties without bias. We find that the gas density and gas mass reconstruction is robust within a few percent. The reconstruction of X-ray temperature is in agreement with expectations, but its direct link to mass is not straightforward: the results depend on the hydrodynamical simulation and temperature biases propagate to pressure and entropy at the 10 – 20% level.

Radial profiles are normalized to the true cluster radius, providing a consistent scale for input-output comparison, while centring and reconstruction are carried out independently through the X-ray analysis. A detailed investigation of the impact on mass and radius estimates is deferred to future work. The best-fit results are always median values accompanied by their 16th-84th percentile points. This article is organized as follows. In Sect. 2 we present the simulations used in this work. In Sect. 3 we explain the generation of the mock XMM-Newton data. In Sect. 4 we describe the X-ray analysis of the mock data. We present our results about X-ray observables in Sect. 5. We further discuss them and summarize our work in Sect. 6.

2. Hydrodynamical simulations

We briefly describe the hydrodynamical simulations used in this work, but refer to their presentation papers for a more in depth description. A summary is reported in Table 1.

2.1. The300

The300 simulations consists of 324 regions centred on large galaxy clusters simulated with full hydrodynamical processes ([Cui et al. 2018](#)). The clusters have been initially identified by the Rockstar halo finder ([Behroozi et al. 2013](#)) in the $1 h^{-1}$

Gpc dark-matter-only MDPL2 box (Klypin et al. 2016)². The dark matter haloes are selected to have virial mass larger than $8 \times 10^{14} h^{-1} M_{\odot}$.

In the resimulation process, the dark matter particles within each of the selected region of radius $\sim 15 h^{-1} \text{Mpc}$, are split into dark matter and gas, according to the cosmological baryon fraction and with an initial gas mass resolution of $2.36 \times 10^8 h^{-1} M_{\odot}$. The re-simulation was performed with GADGET-X (Beck et al. 2016) using smooth-particle hydrodynamics (SPH) and halos and sub-halos were identified by the AMIGA Halo Finder (AHF, Knollmann & Knebe 2009), which accounts for the baryonic components in the halo finding process. The baryonic physics includes models for gas cooling (Wiersma et al. 2009), stellar evolution (Tornatore et al. 2007), stellar feedback (Springel & Hernquist 2003), black hole growth and AGN feedback (Steinborn et al. 2015). Cosmological parameters are from Planck Collaboration et al. (2016). The300 project reproduces the baryon fraction and scaling relations of local galaxy clusters down to $10^{13} M_{\odot}$. Recent work from Rasia et al. (2025) highlighted the importance of properly modelling the hot gas fraction in galaxy clusters. In particular, The300 compares well with CHEX-MATE in terms of emission measure profiles (Bartalucci et al. 2023), temperature profiles and their inhomogeneities (Rossetti et al. 2024; Lovisari et al. 2024), and gas pressure and entropy for the most massive systems (Riva et al. 2024).

2.2. Magneticum

The Magneticum suite⁴ is a collection of full cosmological hydrodynamical simulations (Dolag et al. 2025). They are performed with the P-GADGET3 code (Springel 2005). They include baryonic process such as AGN feedback (Fabjan et al. 2010), star formation, supernovae explosions, galactic winds (Springel & Hernquist 2003), gas cooling (Wiersma et al. 2009), and enrichment (Tornatore et al. 2007). The cosmological parameters are taken from the WMAP results (Komatsu et al. 2011). In particular, we select galaxy clusters from the Box2 and Box2b, respectively with size of 352 and 640 Mpc h^{-1} , and dark-matter and gas particle mass of $6.9 \times 10^8 M_{\odot}$ and $1.4 \times 10^8 M_{\odot}$, respectively. Galaxies and clusters are identified using a Friends-of-Friends (FoF) algorithm combined with Subfind (Dolag et al. 2009).

Magneticum successfully reproduces the AGN luminosity function (Hirschmann et al. 2014), morphological properties of clusters (Teklu et al. 2015; Remus et al. 2017; Gupta et al. 2017), thermodynamical profiles and features of galaxy groups (Bahar et al. 2024; Popesso et al. 2024). It was exploited to study the hydrostatic mass bias and recovery of X-ray observables in the context of eROSITA (Scheck et al. 2023; ZuHone et al. 2023).

2.3. MACSIS

The MAssive ClusterS and Intercluster Structures (MACSIS, Barnes et al. 2017) dataset is a collection of 390 re-simulated

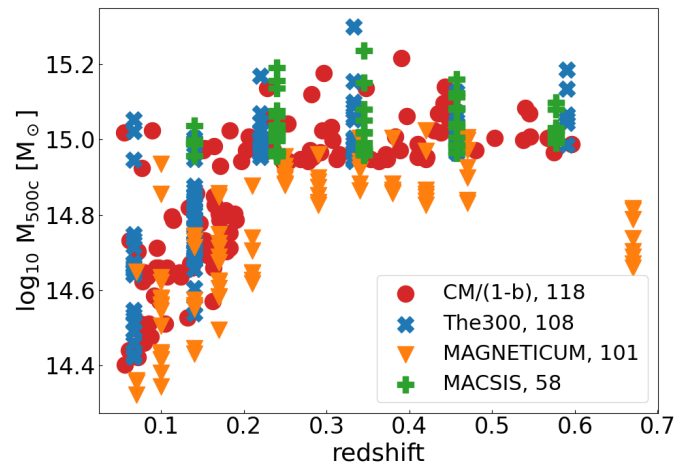


Fig. 1. Mass and redshift distribution of the CHEX-MATE clusters compared to the twin selected systems from The300, Magneticum, and MACSIS. The CHEX-MATE masses include the hydrostatic mass bias of 0.2.

regions carried out with full hydrodynamical models. The concept is similar to the one presented by The300 project in Sect. 2.1, with an initial selection on a large N -body simulation followed by a resimulation that includes the baryonic components. The parent simulation is a dark-matter-only cube of 3.2 Gpc, run with GADGET3 (Springel 2005). The cosmological parameters are taken from Planck Collaboration et al. (2014a). The dark matter particle mass is $5.43 \times 10^{10} h^{-1} M_{\odot}$. The MACSIS sample was selected from all haloes more massive than $10^{15} M_{\odot}$ at $z=0$, identified by a friend-of-friend algorithm. These haloes were grouped into mass bins of 0.2 dex. If a bin contained less than 100 haloes, all of them were selected, otherwise the bin was refined to 0.02 dex from which 10 objects were randomly selected.

These haloes were re-simulated with full hydrodynamical models following the prescriptions from BAHAMAS (McCarthy et al. 2017). Similarly to The300 and Magneticum, the baryonic model includes radiative cooling from different elements (Wiersma et al. 2009), star formation and feedback (Schaye & Dalla Vecchia 2008), as well as black hole seeding, growth, and feedback (Booth & Schaye 2009). The hot gas profiles from MACSIS show good agreement with observational data (Barnes et al. 2017; Riva et al. 2024).

2.4. Simulated cluster sample

We select clusters by picking a twin for each real CHEX-MATE cluster with the closest possible mass in the various redshift snapshots. For The300 and MACSIS, a 20% hydrostatic mass bias is applied when matching observed clusters to simulated analogues. Although this introduces heterogeneous assumptions on b_{HE} , it has the advantage that our analogue sample spans the full range from no bias to 20%. As a result, the reconstruction tests naturally assess the robustness of our methods across the plausible range of hydrostatic mass biases, rather than relying on a single fixed value. As explained in Sect. 1, CHEX-MATE is split into Tier 1 and Tier 2 clusters. Because of its construction containing exclusively massive systems, MACSIS is associated only to Tier 2 objects. Instead, for The300 and Magneticum we include matched samples for both Tier 1 and Tier 2. Fig. 1 shows the distribution of CHEX-MATE clusters (in red) in the mass-redshift space, together with the twin selected systems

² <https://skiesanduniverses.iaa.es/Simulations/MultiDark>, <https://www.cosmosim.org>

³ The virial mass is the total mass enclosed within the virial radius, i.e. a region encompassing an average overdensity that is equal to the critical density of the Universe at the cluster redshift multiplied by the virial overdensity (see Bryan & Norman 1998). At $z=0$ Δ_{vir} is about 99 times larger than the critical density and 330 times larger than the background matter density in vanilla ΛCDM .

⁴ <http://www.magneticum.org>

from The300 (in blue), Magneticum (in orange), and MACSIS (in green).

Because the snapshots of the simulations are saved at different redshifts, the matching with the redshifts of real CHEX-MATE objects may be more or less precise, especially at high redshift where the snapshot of the Magneticum Box2b are sparse, and the systems around redshift of 0.6 are matched to the snapshot at 0.67. This is not a limitation for our goals, as we do not aim at studying the properties of individual clusters, but we focus on a comparison at the population level to input properties using a global sample that is representative of CHEX-MATE.

3. Mock generation

We use the individual gas particles in the simulations to construct an emissivity model projected along the line of sight in a $30' \times 30'$ field of view (FoV). We fold the model through the response of XMM-Newton and generate mock events assuming a clean 25 ks exposure time. We then analyse the mock X-ray data extracting surface brightness and temperature profiles and applying a deprojection technique (details in Sect. 4).

3.1. Input data

To generate the X-ray emission from each gas particles we use the first unit of the code X-ray Map Simulator described in [Gardini et al. \(2004\)](#) with some modification as follows. The gas particles are selected within a temperature range of $[0.3 - 40]$ keV and with a gas density below the star formation threshold to guarantee that multi-phase particles are not included. We stress that this does not mean that emission below 0.3 keV is ignored, because gas elements at different temperatures also shine in the softest X-ray band. The field of view covers 30×30 arcmin² and it is sampled with a grid of 512×512 pixels. We create 495 narrow energy channels, linearly spaced between 0.1 and 10 keV and we sum and project along one random axis the emissivity of the individual particles. Specifically, the projected spectra assume thermal emission from a collisionally-ionized diffuse gas with a fixed metallicity of $0.3 Z_{\odot}$ ([Asplund et al. 2009](#)) and is corrected by an absorption term assuming a hydrogen column density of $5 \times 10^{20} \text{ cm}^{-2}$ (i.e. in Xspec: phabs(apec)). The final projected flux is stored in a data cube of size $512 \times 512 \times 495$. Effectively, this information is similar to that of an ideal integral field unit (IFU), where for each pixel we have access to the input global spectrum.

3.2. XMM simulation

We use the data cube obtained in the previous section as an input to the `xmm_simulator`⁵ software to generate realistic XMM-Newton simulations. We build a 2D model image for each energy channel. We compute the effective area on the pixel grid accounting for the detector quantum efficiency, filter transmission, CCD gaps, and the telescope vignetting by combining each component in the XMM current calibration files (CCF)⁶. We also include the non-X-ray background by loading filter wheel closed data from the CCF for each camera. Its intensity is constant on the detector surface, i.e. there are no soft protons. We model the sky background including contributions from the foreground, with the local hot bubble (an unabsorbed Apec model with a temperature of

0.11 keV), the galactic halo (an absorbed Phabs(Apec) model with a temperature of 0.22 keV), and the cosmic X-ray background with contribution from the faint, undetected AGN (an absorbed power law Phabs(Power) with a spectral index $\Gamma=1.46$). Finally we include the emission of AGN by randomly generating their position within the FoV. This means that no AGN clustering is present, but this is negligible given the size of the XMM FoV, and 1D cluster profile studies are not affected. The fluxes are drawn from the logN-logS distribution of [Lehmer et al. \(2012\)](#). The AGN spectral model is an absorbed power law with column density randomly drawn in the range $[10^{20} - 10^{23.5}] \text{ cm}^{-2}$ and a slope drawn from a Gaussian distribution with mean of 1.9 and variance 0.2 ([Ueda et al. 2014](#)).

For each energy channel, the total model image with the source, sky background, and AGN, is recast into XMM pixels, convolved with the PSF, and the Response Matrix File (RMF). Individual events are generated as a Poisson realization of the total model, including a separate particle background photon list that is merged with the event file. We generate mock EPIC events with an exposure time of 25 ks, which is representative for the CHEX-MATE observations.

4. X-ray analysis

We reduce the mock data with dedicated software to analyse XMM-Newton data XMM_SAS version 21.0⁷, routines from `pyproffit`⁸ ([Eckert et al. 2020](#)) and `hydromass`⁹ ([Eckert et al. 2022](#)). We note that this pipeline is not identical to that used for CHEX-MATE, although they share most of the underlying methodology. We ignored 4 (3, 3) clusters in The300 (Magneticum, MACSIS) with complex shape due to recent major mergers, where assumptions such as spherical symmetry fail.

4.1. Extraction of Images and Spectra

First we create images in the 0.7-1.2 keV band by combining the individual detectors MOS1, MOS2, and PN into a single EPIC XMM image. We also generate a single exposure map by summing together individual exposure maps, while multiplying the one relative to the PN detector by 3.42, that is the ratio of the PN to MOS effective area in this energy band. We use the *ewavelet* task to run a wavelet source detection algorithm on the mock images. This allows us to identify individual point sources to be masked during the X-ray analysis. We carefully visually inspect each region file obtained by the source detection process. Obviously the central detection relative to the simulated cluster needs to be removed. This is straightforward. However, sometimes the algorithm splits the clusters into multiple fake sources, or does not include the tails of the point source emission within the aperture corresponding to AGN detections. We manually modify each region mask to include the most amount of clean emission from the cluster, while minimizing the impact of subhaloes, nearby systems, and AGN leakage to the best of our abilities. This step does not involve any prior knowledge from the simulation and the mock is treated exactly as a real observation. From the simulation perspective, we can identify whether masked detections correspond to injected AGN or to non-AGN features, such as substructure or background fluctuations. Across the three simulations we find no significant variation in the substructure

⁵ github.com/domeckert/xmm_simulator

⁶ <https://www.cosmos.esa.int/web/xmm-newton/current-calibration-files>

⁷ cosmos.esa.int/web/xmm-newton/sas-news

⁸ pyproffit.readthedocs.io

⁹ hydromass.readthedocs.io

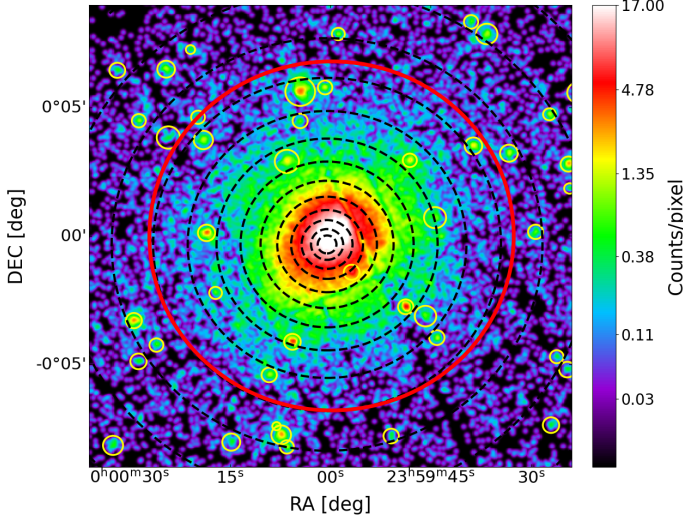


Fig. 2. Example of a simulated XMM-Newton EPIC image for one cluster. The black lines denote the regions used for spectral extraction to measure its temperature profile. The red circle denotes R_{500c} .

fraction: approximately 75% of masked sources are bona fide AGN with 5×10^{-14} erg/s/cm².

Secondly we prepare the regions to extract spectra to measure the radial temperature profile. By default, the radial profiles consist of one inner bin from 0 to $0.04 \times R_{500c}$, in addition to 12 more bins spanning from 0.04 to $1.1 \times R_{500c}$. We extract the background spectrum in a circular region located between 1.5 and $2.0 \times R_{500c}$. If the upper boundary overshoots the XMM FoV we manually set it to 15 arcmin. We carefully inspect each region automatically generated by our pipeline and modify it if needed. In some cases the background region is too large and falls out of the FOV. In other cases it ends up including a filament or part of a secondary nearby structure. For such cases, we extract the background spectrum in one of the FOV corners using a circular aperture of about 3 arcminutes according to the needs of each specific case. An example of the end result of the whole procedure is shown in Fig. 2. It displays the EPIC count map for one of the The300 clusters. The yellow circles denote the sources identified by the wavelet detection, the dashed black lines are the radial bins used for spectral extraction, and the red circle corresponds to R_{500c} .

Finally, we fit each spectrum using the X-COP pipeline (Ghirardini et al. 2019). The global model accounts for a high energy particle background (a broken power law with several Gaussian lines), the sky background including the CXB (an absorbed power law with photon index of 1.46), the Galactic halo (an absorbed apec model with temperature between 0.15 and 0.6 keV), and the local hot bubble (an unabsorbed apec model with temperature of 0.11 keV). We turned off the soft protons (an additional broken power law), since we do not simulate them. The source model is an absorbed apec model with temperature, normalization, and abundance free to vary. In the X-COP pipeline, background modelling is improved by simultaneously fitting the background extracted from the XMM field with ROSAT All-Sky Survey (RASS) background data, which are automatically retrieved in the region surrounding the source. To replicate this approach in our simulations, we generate synthetic X-ray background spectra composed of the same physical components used

in the pipeline, i.e. the cosmic X-ray background (CXB), Galactic halo emission, and the local hot bubble. Each component is simulated with the same temperatures, abundances, and normalizations adopted in the `xmm_simulator`, and convolved with the ROSAT PSPC response files. The simulated background spectrum is normalized to the extraction area, which by default corresponds to a circular annulus between 60 and 90 arcminutes from the source centre. To remain consistent with the simulation setup, we omit the cross-calibration correction factor between the ROSAT PSPC and the XMM-Newton EPIC camera that is otherwise applied in the X-COP pipeline. This procedure ensures a consistent background treatment and yields more robust temperature measurements, particularly in the outer radial bins where the signal is background-dominated. We use X-SPEC (v12.13.1 Arnaud 1996) with C-stat (Cash 1979). We refer the reader to Ghirardini et al. (2019) for additional details on the X-COP pipeline.

4.2. Measurement of X-ray properties

We extract surface brightness profiles from the EPIC mosaic and from its voronoi binned image. We refer to them as mean and median profiles respectively. The median profile is derived from the Voronoi-tessellated map, where each adaptive bin is assigned the median value of its constituent pixels before computing the radial profile. The mean profile is extracted directly from the original image by averaging the native pixels in each radial annulus, without any adaptive binning. Thus, the terms "median" and "mean" describe the construction of the underlying maps rather than different statistics applied to the same pixel set. We use the tessellation scheme from Diehl & Statler (2006), grouping the count maps into cells containing 25 events each to create the final voronoi image. We use the latter to generate azimuthal median profiles. This alleviates the issue of dense regions biasing the recovered gas density high due to its squared relation to emissivity. The surface brightness profile is modelled as a collection of Kings functions, which allows the computation of the 2D projected profile analytically. In particular, the cluster X-ray emissivity follows: $\epsilon_X = \sum_{n=1}^N \alpha_n \Phi_n(R) \propto n_e^2 \Lambda(T, Z)$, where $\Phi_n(R) = \left(1 + \frac{R^2}{R_{c,n}^2}\right)^{-3\beta}$ and $\Lambda(T, Z)$ is the cooling function, which depends on temperature and metallicity (Sutherland & Dopita 1993). The model is convolved with the PSF, and superimposed to a constant background profile that is estimated by computing a count rate value from the spectral model of the background obtained in Sect. 4.1. We refer the reader to Eckert et al. (2020) for a detailed description of the surface brightness measurement and modelling.

We then model the gas and mass profiles under the assumptions of hydrostatic equilibrium and spherical symmetry, which allows linking the gas pressure to the total mass. Gas density, temperature, and thermal pressure are related by the ideal gas equation of state. The formalism reads:

$$\begin{aligned} \frac{dP_{\text{gas}}}{dr} &= -\rho_{\text{gas}} \frac{GM_{\text{TOT}}(<r)}{r^2}, \\ P_{\text{gas}} &= \frac{k_B}{\mu m_p} \rho_{\text{gas}} T, \\ M_{\text{TOT}}(<r) &= -\frac{rk_B T(r)}{G\mu m_p} \left(\frac{d \log T}{d \log r} + \frac{d \log n_{\text{gas}}}{d \log r} \right). \end{aligned} \quad (1)$$

One can then model the data using Eq. 1 from different points of view.

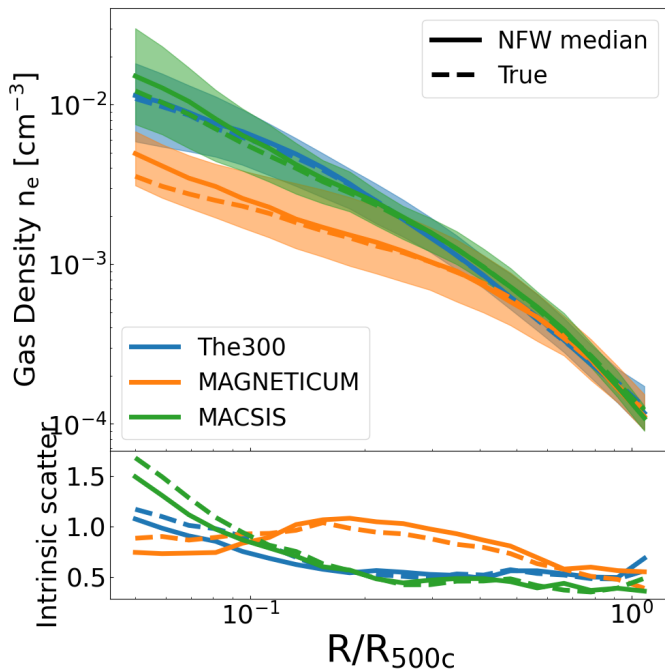


Fig. 3. Gas density profiles in our simulations. Top panel: comparison between the measured (solid lines and shaded areas) and true (dashed lines) gas density profiles in different simulations (in various colours), using the NFW reconstruction. Bottom panel: intrinsic scatter as a function of radius.

(i) Navarro-Frenk-White model (NFW, Navarro et al. 1996) or model of the mass profile: given the definition of a mass model, it is possible to derive pressure by integrating the hydrostatic equilibrium equation. Temperature is inferred using the ideal gas equation of state, while gas density is related to the surface brightness via the cooling function.

(ii) Non-parameteric model (NP) or model of the temperature profile: a linear combination of log-normal functions is used to describe the 3D temperature profile. It is then projected along the line of sight and fitted to the measurement.

(iii) Forward model (FM) of the pressure profile: a generalized NFW is used to describe the pressure profile, which allows computing analytically its gradient and the mass profile.

In any of these three cases, the fit is performed on gas density and temperature. The model is fitted to total (gas and dark matter) mass. We refer the reader to Eckert et al. (2022) for full details on the modelling. Finally, one can derive the entropy from the 3D temperature and density profiles as $K = k_B T n_{\text{gas}}^{-2/3}$.

5. Results

In this section we compare the profiles obtained from the analysis in Sect. 4 to the input quantities measured directly using the properties of the gas particles in the hydrodynamical simulations.

5.1. Gas density profiles

Gas density is the most fundamental thermodynamic quantity accessible in X-rays, as the surface brightness directly probes n_e^2 and is therefore less model dependent than temperature. It is key to infer cluster properties such as gas mass, pressure, and entropy, and is essential for comparisons with SZ and lensing. We

study the density reconstruction on a population level. We compare input-output profiles on a per-cluster basis, i.e. we always show the distribution of individual-cluster ratios, rather than the ratio of the median reconstructed and true profiles. This holds for all results across the article.

Figure 3 shows the recovered and true density profiles for the three simulation suites. Overall, the agreement is good: discrepancies are limited to the innermost $0.1 R_{500c}$ in *Magneticum*, and remain within the scatter of the measurements. *The300* and *MACSIS* exhibit similar core shapes and normalisations, while *Magneticum* shows flatter and lower-density cores, due to the lower typical masses and the distinct baryonic-physics implementation (see Rasia et al. 2025). A simple beta model¹⁰ fit is close to the expectation $\beta=2/3$, but the flatter profile in *Magneticum* causes a larger core radius in the best fit. We perform the fitting with the `curve_fit` package in `scipy` (Virtanen et al. 2020). We report the values in Table A.1. The best fit parameters are always compatible within uncertainties between the input and recovered profiles. A single β -model cannot capture the full complexity of all clusters, so the fitted parameters should be regarded as descriptive summaries rather than ground truths, used to illustrate the flexibility of the reconstruction. Small differences in the core are likely due to miscentering (see Appendix E): the observed profiles are computed from using the peak of the X-ray emission in the XMM mocks, while the input profiles are computed from the position of the most bound particle in each cluster.

The bottom panel of Fig. 3 shows the intrinsic scatter of the profiles presented in the top panel. We define the intrinsic scatter as the difference between the 84th and 16th percentiles of the profile distribution, divided by the median. Among the simulations, *MACSIS* exhibits the largest scatter in the core region (within $0.1 \times R_{500c}$), reaching values up to 1.5, whereas *The300* remains below 1 across the same range. *Magneticum* displays the most diverse behaviour at intermediate radii, with a scatter peaking at 1.2 around $0.2 \times R_{500c}$, while *The300* and *MACSIS* are closer to 0.6 at this radius. In the outskirts beyond $0.8 \times R_{500c}$, all simulations converge to a similar relative scatter of about 0.5. Importantly, the dashed lines (true profiles) closely track the solid ones (measured profiles): the reconstructed gas density profiles not only match the true median profiles, but also reproduce the true population scatter as a function of radius.

We find excellent agreement between different reconstruction methods, with the median profiles from the voronoi images showing the greatest precision (see Fig. A.2 and discussion in A). In Appendix A we also show that the integration of gas density profiles to measure gas masses also recovers the input gas mass accurately with the voronoi approach to better than 1% precision, while the standard processing tends to overestimate gas mass by about 5%.

5.2. Temperature profiles

Temperature is a fundamental X-ray observable of galaxy clusters. As a quantity obtained directly from spectral fitting, it traces the depth of the gravitational potential and underpins hydrostatic mass estimates. Its radial profile constrains the thermodynamic structure of the ICM and, together with pressure, calibrates the scaling relations used in cluster cosmology. Biases in the recovered temperature arising from multi-temperature structure, pro-

¹⁰ $n_e(r) = n_0 \left(1 + \frac{r^2}{r_c^2}\right)^{-3/2\beta}$

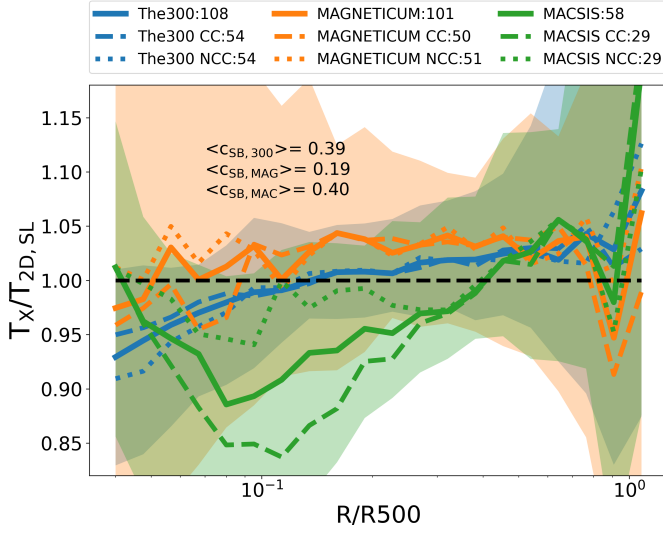


Fig. 4. Ratio between the result of the X-ray spectral fitting and the input spectroscopic-like temperature profile. The profiles are also split between cool core (CC) and non cool core (NCC) according to the median surface brightness concentration reported as text in the panel.

jection, or instrumental effects therefore propagate directly into mass estimates and the interpretation of ICM physics.

5.2.1. Temperature weighting schemes

Unlike gas density, for which the true profile is easily defined using the particles, the notion of a true temperature profile is inherently ambiguous: although each gas element has a well-defined internal energy and thus a temperature, any radial profile or integrated quantity requires adopting a weighting scheme. The inferred temperature is sensitive to this choice (Mazzotta et al. 2004; Rasia et al. 2014), especially in the presence of multiphase gas. Observationally, spectral fitting preferentially weights cooler, denser phases of the ICM with stronger X-ray emission. To reflect these effects in our comparison, we consider two standard definitions. The mass-weighted (MW) temperature follows the assumption that more massive gas elements contribute proportionally more to the thermal energy budget. It is a physically motivated thermodynamic average linked to the gravitational potential and should be the benchmark for comparisons to observations based on combinations to the SZ effect. However, it does not fully capture the complex dependence of X-ray emissivity on both gas density and temperature, particularly in the presence of temperature inhomogeneities or multiphase structures (Gaspari et al. 2020), also because the typical response of X-ray instruments is higher at soft energies around 2 keV and lower for hotter gas above 3 keV. Instead, the spectroscopic-like (SL) scheme, introduced by Mazzotta et al. (2004) is constructed to emphasize denser, cooler gas phases that dominate the observed X-ray emission in XMM-Newton and Chandra data. It yields robust estimates for high-temperature plasma hotter than about 3 keV, making it particularly suitable for galaxy clusters in the mass range considered here. They are defined as follows:

$$T_{\text{MW}} = \frac{\sum_i m_i T_i}{\sum_i m_i},$$

$$T_{\text{SL}} = \frac{\sum_i n_i^2 / T_i^{3/4} T_i}{\sum_i n_i^2 / T_i^{3/4}}, \quad (2)$$

where m , n are respectively the mass and the density of each individual gas particle, and the index i identifies each particle within its volume, i.e. a spherical shell in case of a 3D profile, and a cylindrical shell for 2D profiles. In this work, we test these weighting schemes with a full forward modelling of the XMM observations for the first time.

First we compare the radial profile of the measured temperature obtained by spectral fitting in annuli (T_X) to the expected T_{SL} . The result is shown in Fig. 4 with solid lines. This is close to one for The300 and Magneticum, but it is underestimated in the cores of MACSIS, within about $0.2 \times R_{500c}$ by at most about 10% (although the one-to-one line is still within the scatter of MACSIS profiles). We exclude that this is due to PSF effects, because the median PSF scale of 30 arcsec corresponds to about $0.11 \times R_{500c}$ in MACSIS, compared to about 0.08 (0.105) in The300 (Magneticum). Instead, we find this effect to correlate with cool cores (see dashed and dotted lines in Fig. 4). Following Campitiello et al. (2022), we quantify surface-brightness concentration¹¹ (their Eq. 1 and Sect. 4.1) and classify clusters as CC or NCC using the sample median. We obtain median c_{SB} values of 0.39, 0.19, and 0.54 for The300, Magneticum, and MACSIS, respectively. The300 agrees with Campitiello et al. (2022), while the lower value in Magneticum reflects its flatter cores. We then compare the T_X/T_{SL} ratio for CC and NCC subsamples. In Magneticum, concentrations are generally low (quartiles Q1-Q3=0.15–0.22), and the two populations show no significant difference. In The300, the distribution is broader (Q1-Q3=0.23–0.51), but temperature profiles are flat (Fig. 5), resulting in only a $\sim 5\%$ core difference. In MACSIS, both the concentration distribution (Q1-Q3=0.27–0.59) and the temperature gradients are larger, producing a clear T_X/T_{SL} offset: NCC resemble The300, but CC are underestimated by up to $\sim 15\%$. Indeed, Mazzotta et al. (2004) defined the weight of 0.75 in T_{SL} to accommodate a combination of various plasma phases, although some variations are expected: their Fig. 9 shows that the larger the difference between the temperature of two plasma, the larger the deviations in T_{SL} compared to T_X . This is particularly relevant for MACSIS: the temperature profiles are steep, so the 2D projection of the core contains plasma phases around 4–5 keV as well as 8–9 keV, where deviations around the 10% level are expected. Finally, Rossetti et al. (2024) showed that the CHEX-MATE temperature profiles are in agreement with The300 and do not exhibit the steep profiles that are typical in MACSIS. This suggests that T_{SL} is in an excellent proxy for T_X in the real CHEX-MATE sample. For more complex systems and strong cool cores, multi-temperature fits are required. In any case, the agreement is excellent also for MACSIS outside of $0.2 \times R_{500c}$, so that total mass measurements more sensitive to local temperature and temperature gradient around R_{500c} are not affected by the discrepancy in the core. Nonetheless, we verified that the temperature modelling from the observer’s perspective is indeed able to reconstruct the measured temperature profile (see Fig. B.1).

In the main panels of Fig. 5 we report the reconstructed two-dimensional temperature profiles from the NFW, NP, and FM models, compared to the true T_{SL} and T_{MW} profiles for each simulation. The profiles were rescaled to T_{500c} adopting the X-COP cluster sample calibration, following Ghirardini et al. (2019) (see their Eq. 10). The smaller panels show output-input ratios: the top rows compare 2D reconstructed temperature to T_{SL} , and the bottom rows show the corresponding ratios for the deprojected 3D profiles. In each panel we also report the median T_{500c} for

¹¹ $c_{\text{SB}} = \text{SB}(< 0.15 R_{500c}) / \text{SB}(< R_{500c})$

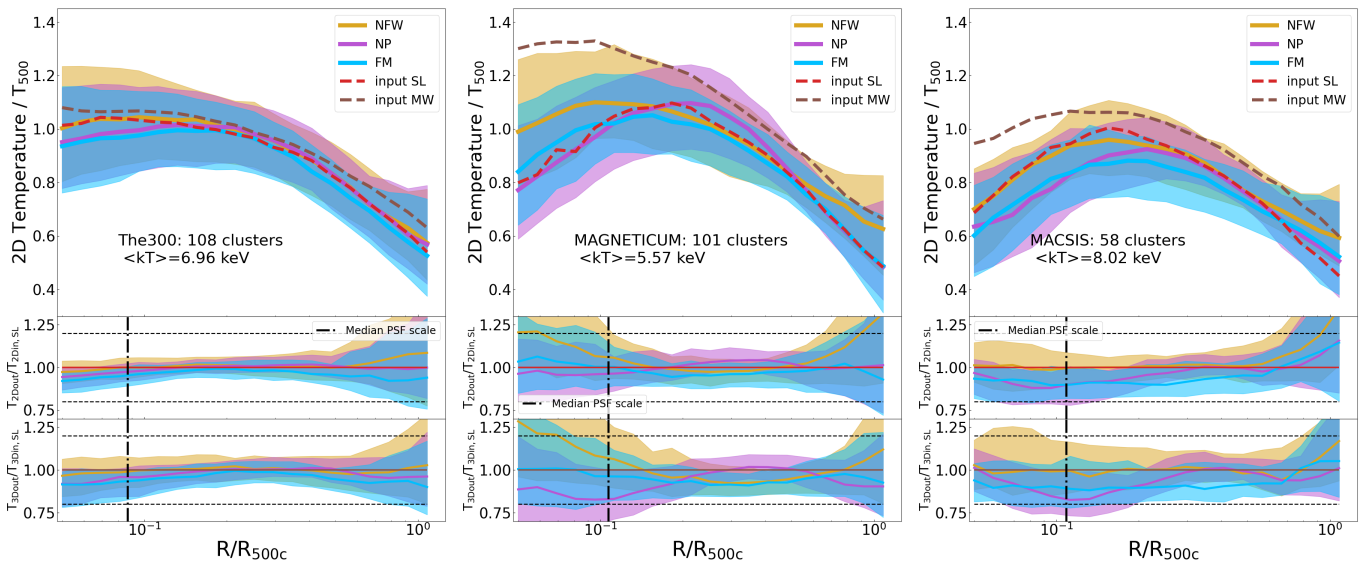


Fig. 5. Gas temperature profiles. Each panel corresponds to one simulation: *The300* on the left, *Magneticum* in the centre, and *MACSIS* on the right. The colours denote the three reconstruction methods. The smaller panels show the ratio between the reconstructed 2D and deprojected 3D profiles.

each sample. We find this to be low for *Magneticum* at 5.57 keV compared to 6.96 keV for *The300* and 8.02 keV for *MACSIS*. This is expected because the sample from *Magneticum* has intrinsically less massive systems (Sect. 2.4). The black lines denote a median PSF scale of 30 arcsec given the radius and redshift of each system.

Similarly to gas density, the three simulations show different behaviours also in the temperature profiles. *The300* exhibits flat cores with temperatures very close to T_{500c} within $0.3 \times R_{500c}$, while *MACSIS* and *Magneticum* produce clusters with lower temperatures in the core. In particular, *MACSIS* has strong cool cores with temperature being about 30% lower than T_{500c} at $0.05 \times R_{500c}$. We find evidence that the stronger cooling in *MACSIS* suppresses the global ICM temperature, in fact, using the T_{500c} normalization from Ghirardini et al. (2019), the average $T(r)/T_{500c}$ profile in *MACSIS* remains below unity over the full radial range examined. In comparison, *The300* and *Magneticum* reach (and locally exceed) unity, though with different profile shapes. These trends imply that, at fixed halo mass, *MACSIS* yields cooler clusters overall—consistent with enhanced cooling efficiency lowering the thermal energy of the ICM.

The top panels of Fig. 5 also show that there is better agreement between the reconstructed profiles with the SL input models compared to the MW ones. This holds both for the 2D and deprojected 3D profiles. The largest differences are in the core within $0.15 \times R_{500c}$, where the MW scheme overestimates the reconstructed profiles by more than 20% (40%) in *MACSIS* (*Magneticum*). The disagreement is not as pronounced in *The300*, which points to a smoother ICM temperature distribution, as we further explore in Sect. 5.2.2. Instead, the SL scheme shows a good agreement with the one-to-one ratio to the reconstructed profiles. The best result is provided by the NP model, which is the one that reconstructs temperature directly. Its flexibility provides almost a perfect one-to-one recovery of T_{SL} in *The300* and *Magneticum*. In *MACSIS* instead the agreement is excellent outside of $0.2 \times R_{500c}$, but not in the core, where we find that the temperature is underestimated by 5-10%. This is due to the discrepancy in the temperature measurement investigated in Fig. 4. *The300* shows the best consistency between differ-

ent reconstruction methods, whereas *Magneticum* exhibits the largest scatter between NFW, NP, or FM reconstructions. Such differences may be caused by general departures from spherical symmetry or specific assumptions within each model, such as the NFW parametrisation, or likely different levels of multi-temperature structures in these simulations (see Sect. 5.2.2).

Finally, the trends observed in the ratios between the 2D and deprojected 3D temperature profiles are consistent across the different simulations and reconstruction methods, for both the T_{SL} and T_{MW} . This consistency suggests that the deprojection procedure does not introduce any significant bias in the reconstructed 3D profiles.

We provide full details and figures about pressure and entropy in Appendix B. Pressure is systematically underestimated in all simulations, although at different levels. At $0.3 \times R_{500}$ the bias is about 5% in *The300* and increases to about 20% in *Magneticum* and *MACSIS*. Comparable trends are seen near R_{500} , albeit with some method-dependent scatter. Entropy is recovered to within 5% across most radii in *The300*, whereas *Magneticum* and *MACSIS* show a larger deficit (15-20%) in the inner regions. This reduces to 5-10% near R_{500c} . We also study the 2D cluster temperature distribution by generating and analysing voronoi binned temperature maps. Similarly to the density case, we find that the voronoi technique is less affected by clumpiness on a population level. In particular, after removing outliers, temperatures are raised by about 15-20%, underscoring once again the effect of multi-temperature structure on radial temperature profiles. The same analysis on a 2 Ms simulation shows that the individual profiles are compatible with the radial spectral fitting within uncertainties (see Appendix D).

5.2.2. Multi-temperature gas

The discrepancies observed between the reconstructed temperature profiles with T_{SL} and T_{MW} , as well as the intrinsic differences between these schemes themselves, can arise from several factors: the presence of multi-phase gas due to mixing or substructures, which induces azimuthal temperature variations within a given radial bin, and projection effects affecting the

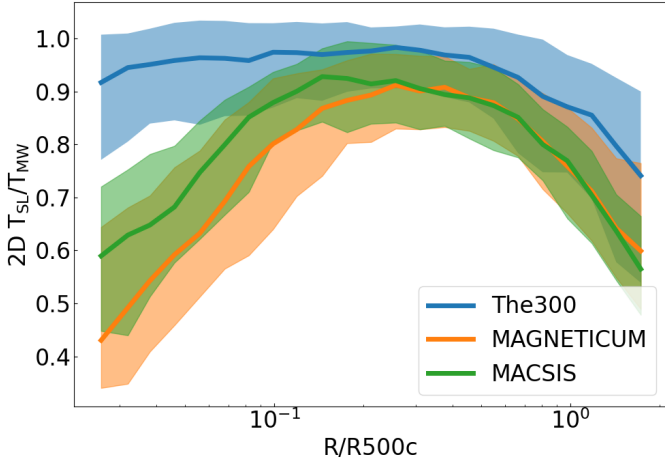


Fig. 6. Ratio between the spectroscopic-like and mass-weighted temperature profiles in different simulations.

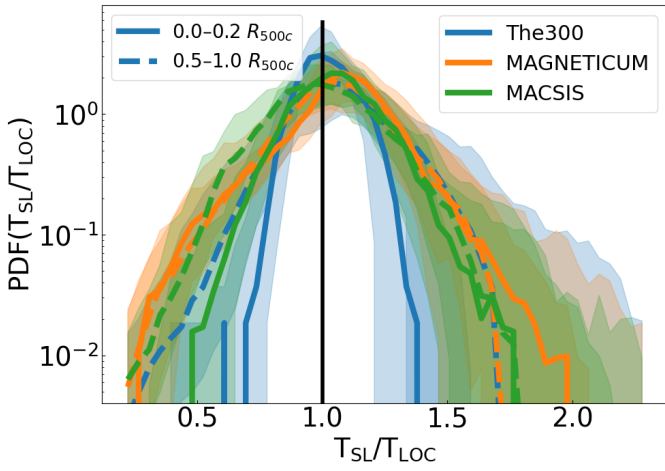


Fig. 7. Comparison between the distribution of the local spectroscopic-like temperature fluctuations in 2D maps within different apertures for different simulations. T_{LOC} is the 1D radial profile at the position of each pixel in the map.

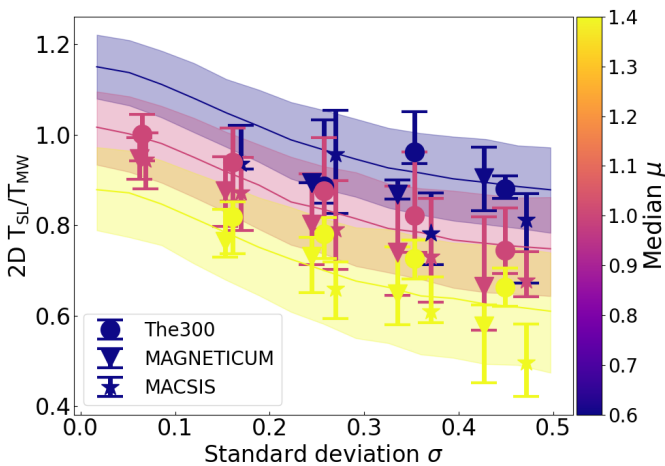


Fig. 8. Ratio between the spectroscopic-like and mass weighted temperature profiles as a function of the standard deviation of the spectroscopic-like temperature fluctuations. The points are colour coded by the median of the fluctuation distributions.

3D reconstruction of the temperature profile. In Appendix C we used a toy model with one single radial profile distribution and find that projection effects should yield a ratio between SL and MW temperatures below one in the core. However, in Fig. 6, we observe a different trend: T_{MW} remains consistently higher than T_{SL} across all radii. While the discrepancy in the core is expected, the growing divergence in the outskirts suggests that systems simulated in a cosmological environment exhibit significant temperature inhomogeneities, departing from a simple individual temperature profile. This indicates that multi-temperature gas, driven by substructures, mixing, or accretion-related shocks, has a non-negligible impact on the projected temperature profiles of galaxy clusters. This is also in agreement with the picture described by Lovisari et al. (2024), who found a small effect on the temperature profile of CHEX-MATE clusters due to highly deviating regions in the core, but larger effects up to 20% moving towards the outskirts. This also relevant for estimates of the total cluster mass, that depends on the temperature normalisation and gradient at a given radius (Eq. 1) (see also discussions in Kawahara et al. 2007; Rasia et al. 2006, 2012, 2014; Pearce et al. 2020; Barnes et al. 2021).

To further study azimuthal temperature variations in the ICM, we use 2D maps of T_{SL} to analyse its distribution normalized by the local 1D radial profile in a given radial annulus interpolated at the position of each pixel of the 2D map (T_{LOC}). This approach isolates relative temperature fluctuations within clusters, allowing for a direct comparison of the spatial variability of the thermal structure, independent of the overall temperature normalization. The use of this normalized quantity is motivated by several key trends observed in the simulations. In particular, *Magneticum* exhibits a more pronounced difference between T_{SL} and T_{MW} compared to *The300* (see Fig. 6). Similarly, *Magneticum* shows a stronger deviation between temperature profiles reconstructed from median voronoi temperature maps and those obtained via spectral fitting in radial bins (see Fig. D.1). These discrepancies suggest that *Magneticum* may exhibit stronger thermal inhomogeneities and multiphase structure in the ICM, which can enhance azimuthal variations and bias projected temperatures. This is consistent with differences in the underlying hydrodynamic schemes: SPH formulations with limited mixing are known to preserve cool substructures and generate larger temperature contrasts (e.g. Rasia et al. 2014). More recent SPH implementations, such as the improved scheme in *GADGET-X* used for *The300*, or the OWLS-based subgrid physics adopted in *MACSIS* (Schaye et al. 2010) promote more efficient mixing and therefore tend to produce smoother temperature fields. By computing the PDF of the normalised T_{SL} across clusters as a function of radius, we obtain a population-level diagnostic that is sensitive to such internal fluctuations. A broader distribution in this statistic reflects stronger local deviations from the median, and thus captures the degree of thermal asymmetry, clumpiness, or substructure. This method removes the need for assuming a fixed hydrostatic mass bias or one-to-one matching between clusters, it provides a flexible and robust way to compare the structural complexity of the ICM across different simulation sets and physical models. The result is shown in Fig. 7, focusing on the core and outskirts in the three simulations. We find that *The300* has the tightest distribution in the core, where fluctuations are at the 40% level at most. The distribution is broader in *MACSIS* and especially in *Magneticum*, reaching also a factor of two. In the outskirts there is better agreement between the simulations, especially at the positive fluctuations end. We also notice that the PDF in *Magneticum* tend to peak at values slightly larger than

Table 2. Best fit parameters of the relation between the ratio of SL and MW temperatures and the width of the temperature distribution (see Eq. 3).

Simulation	A	b	log P	σ_{intr}
The300	0.20 ± 0.01	-0.35 ± 0.01	1.09 ± 0.06	0.055 ± 0.001
Magneticum	0.26 ± 0.02	-0.38 ± 0.01	1.08 ± 0.08	0.082 ± 0.001
MACSIS	0.31 ± 0.03	-0.37 ± 0.01	0.85 ± 0.09	0.083 ± 0.001
COMBINED	0.31 ± 0.01	-0.33 ± 0.01	0.76 ± 0.05	0.083 ± 0.001

one, possibly pointing to non-gaussian temperature fluctuation distribution.

We take a further step by directly measuring the first, second, and third orders of the temperature fluctuations, that are the median, the standard deviation, and the skewness, in 30 bins equally distributed between 0 and $1.05 \times R_{500c}$. In each radial bin we also keep track of the ratio between T_{SL} and T_{MW} . We find an excellent correlation between the standard deviation of these distributions with the ratio between the 2D T_{SL} and T_{MW} in different bins, as shown in Fig. 8. Combining the three simulations we get a Pearson correlation coefficient of -0.76. This is due to the presence of gas chunks with different temperatures compared to the mean, increasing the difference between T_{SL} and T_{MW} . In addition, we find a secondary trend in the temperature ratio as a function of the median μ at fixed standard deviation σ . The Pearson correlation coefficient is -0.49. The ratio is closer to one for small μ and decreases as μ increases. Although we do not find a clear correlation between the skewness of the distribution and the ratio between T_{SL} and T_{MW} (Pearson correlation coefficient of -0.04), this secondary μ -trend is partially encoded in the skewness as well: distributions with a long tail at low temperatures tend to peak at higher values, so that a PDF with mean larger than one describes a gas distribution skewed to colder temperature, as we can see from the core of Magneticum clusters in Fig. 7. The SL scheme increases the weight of the cold gas, reducing the ratio to the T_{MW} . The opposite happens for distributions that are more symmetric or even skewed towards hotter gas, with the ratio being closer to one. In general, for tight and symmetric distributions, with μ close to one and low σ , T_{SL} is an unbiased tracer of T_{MW} . As the distribution broadens, gas chunks that deviate from the median are weighted differently by the SL and MW schemes, which progressively lowers the temperature ratio. We build a model that includes the scatter of the temperature ratio around the expectation value computed from the width and the median. It reads:

$$\frac{T_{\text{SL}}}{T_{\text{MW}}}(A, b, p) = \mathcal{N}\left(1.35 + b\mu - A \frac{\sigma^p}{\sigma^p + 0.22^p}, \sigma_{\text{intr}}\right), \quad (3)$$

where $\mathcal{N}$ is a Gaussian distribution, and σ_{intr} the relative scatter around the expectation value. The best fit parameters are reported in Table 2 and their posterior distributions are shown in Fig. B.2. Given a measurement of the median and the standard deviation of the temperature distribution at a given radial distance, this model provides a direct mapping from the spectroscopic-like temperature to the mass weighted one, which is the temperature more directly related to the gravitational potential itself. The correction provided by this model allows X-ray analyses to account for multi-temperature structure, reducing the expected hydrostatic mass bias that is not due to departures from hydrostatic equilibrium but to spectroscopic temperature bias in multi-phase gas, as already pointed out by Pearce et al. (2020); Barnes et al. (2021). We leave the quantitative evaluation of this effect to future (see also Ansarifard et al. 2020).

In conclusion, identifying and quantifying multi-temperature structures is crucial for accurately interpreting observed temperature profiles and understanding ICM thermodynamics. One promising avenue is the use of high-resolution X-ray spectroscopy from missions like the ongoing XRISM or Athena (Cruise et al. 2025) in the next decades. In addition to fitting a global temperature from the thermal bremsstrahlung continuum, the spectral resolution of the microcalorimeters onboard these missions allows for detailed diagnostics using line ratios, particularly of the FeXXV and FeXXVI lines around 6.4–6.7 keV. These lines are sensitive to different ionization states and thus to the plasma temperature: in a hot, single-phase plasma the FeXXVI line is prominent and comparable to FeXXV, while in cooler plasma the FeXXV line dominates (Xrism Collaboration et al. 2025a). Importantly, the effective area of XRISM is sharply peaked at soft energies, where the folded continuum spectra of hot and cool clusters can appear deceptively similar due to instrumental response. In contrast, the iron line ratio method probes a narrow energy window where the ARF is approximately constant, reducing the impact of effective area variation and making it a more direct diagnostic of the thermal structure of the plasma. A discrepancy between temperatures inferred from the continuum and from line ratios would potentially indicate the presence of multi-T gas components. Such works would still need a careful treatment of systematics, as new spectral data with high signal to noise ratio and resolution may uncover limitation of the current models (see Chatzigiannakis et al. 2025). An alternative and complementary approach involves combining X-ray and SZ observations (Pointecouteau et al. 2002; Kitayama et al. 2004; Ruppel et al. 2018; Eckert et al. 2019; De Luca et al. 2021; Chappuis et al. 2025; Gavidia et al. 2025, De Luca et al. subm). The SZ signal is proportional to the integrated electron pressure, whereas the X-ray emission is proportional to the emission measure, i.e. the gas density squared. As a result, X-ray temperatures are weighted more heavily toward denser, cooler regions, while SZ-derived temperatures reflect a more volume-averaged (or mass-weighted) thermal state. Comparing the SZ-inferred temperature, obtained by dividing the SZ pressure by the X-ray electron density or directly from the relativistic SZ signal, to the X-ray spectroscopic temperature can thus reveal discrepancies arising from multi-T structure, clumping, or non-thermal pressure support (Kay et al. 2024). In both approaches, a consistent difference between the temperature diagnostics offers a valuable observational pathway to identifying thermal complexity in the ICM.

6. Summary and conclusions

Within the context of the CHEX-MATE project (CHEX-MATE Collaboration et al. 2021), our primary goal is to assess whether standard X-ray analysis techniques can accurately recover the true thermodynamic properties of massive galaxy clusters. To this end, we selected CHEX-MATE-like cluster samples from three state-of-the-art hydrodynamical simulations: The300 (Cui et al. 2018), Magneticum (Dolag et al. 2017), and MACSIS (Barnes et al. 2017). Using the gas particles in each simulated cluster, we generated idealised X-ray emission maps and folded them through the instrumental response of XMM-Newton using our newly developed tool, `xmm_simulator`, to create high-fidelity, end-to-end synthetic EPIC observations (see Sect. 3).

We then applied standard X-ray analysis tools, including routines from `pyproffit` (Eckert et al. 2020) and `hydromass` (Eckert et al. 2022), to extract surface brightness and temperature profiles, perform deprojection, and derive intrinsic 3D ther-

modynamic profiles (see Sect. 4). This workflow allows for a direct comparison between the output of our X-ray analysis and the corresponding "true" profiles from the hydrodynamical simulations.

We find that the gas density profiles are robustly reconstructed across all methods and simulations, with the correct intrinsic scatter. In particular, using azimuthal median surface brightness profiles derived from voronoi-tessellated images improves accuracy: other than some discrepancies in the core because of mis-centering effects between the peak of the X-ray emitting gas and the dark matter, the gas density profiles are reconstructed within at most 2% in all simulations. The gas masses are reconstructed to better than 1% uncertainty. The voronoi technique suppresses the influence of localised surface brightness enhancements due to cool gas clumps, which otherwise bias the inferred density high by a few percent.

While the gas density is well-defined in simulations, defining a representative true temperature is more complex. We compare two standard weighting schemes: the spectroscopic-like (SL) and mass-weighted (MW) temperatures. The SL weighting provides an accurate approximation of temperatures derived from X-ray spectral fitting in annular bins. However, we observe more scatter in the temperature reconstruction compared to the density case, and a systematic difference between T_{SL} and T_{MW} profiles, a difference that varies across simulations. This suggests that these weighting schemes carry information about the underlying physical state of the ICM. If these effects were only due to projection effects, we would expect the T_{MW} to be higher in the core and lower in the outskirts compared to T_{SL} . However, in our simulations, we consistently find T_{MW} to be higher than T_{SL} at all radii. This deviation follows expectations and points to the presence of unresolved multi-temperature gas structures, which have a significant impact on the reconstructed profiles. Indeed, even though the input and output density and temperature profiles appear consistent, the derived thermal pressure and entropy profiles differ. This discrepancy may result from an unaccounted multi-temperature distribution in the X-ray spectral modelling, combined with possible non-thermal pressure contributions.

These findings raise important implications for the interpretation of hydrostatic mass estimates in galaxy clusters (see discussions in Rasia et al. 2006, 2012; Biffi et al. 2016; Pearce et al. 2020). A long-standing issue in X-ray cluster cosmology is the hydrostatic mass bias, where masses derived under the assumption of hydrostatic equilibrium are systematically low. This discrepancy has often been attributed to non-thermal pressure support from bulk motions or turbulence in the ICM. However, recent high-resolution spectroscopic measurements from XRISM have revealed remarkably low gas velocity dispersions in relaxed systems, suggesting that non-thermal pressure may not be sufficient to account for the full bias (XRISM Collaboration et al. 2025; Xrism Collaboration et al. 2025a,b; XRISM Collaboration et al. 2025; Fujita et al. 2025), although constrained to central regions so far. Our results suggest an alternative and complementary explanation in line with findings from Henson et al. (2017); Pearce et al. (2020); Barnes et al. (2021): the presence of unresolved multi-temperature structures in the ICM can bias spectroscopic temperature measurements, leading to a suppression of the inferred thermal pressure. If the temperature is underestimated due to projection effects or local cooling structures, the derived hydrostatic mass will also be biased low, even in systems that are otherwise in equilibrium. These results highlight the importance of temperature reconstruction methods that mitigate these biases, and stress the importance of multi-wavelength, spatially resolved, and dynamical analyses to disentangle the true

thermodynamic state and equilibrium conditions of the ICM. Looking ahead, our findings emphasise the importance of isolating the distinct sources of bias affecting X-ray-derived cluster mass estimates. In future work, we will extend this analysis to directly compare hydrostatic mass estimates with true masses, enabling a detailed assessment of how much bias arises from assumptions of hydrostatic equilibrium, from the use of single-temperature models in a multi-phase medium, and from projection effects such as triaxiality and line-of-sight substructure (Kim et al. 2024; Saxena et al. 2025; Chappuis et al. 2025). This will be key to understanding whether the observed hydrostatic mass bias truly reflects a breakdown of equilibrium in the ICM, or whether it stems, at least in part, from systematic limitations in our current X-ray analysis methodologies.

Acknowledgements. RS and DE are supported by Swiss National Science Foundation project grant #200021_212576. ER is partially supported by the NASA -EGIP Grant 80NSSC25K8009. EP acknowledges support of CNES, the French space agency and from the French Agence Nationale de la Recherche (ANR), under grant ANR-22-CE31-0010. VB acknowledges partial support from the INAF Grant 2023 "Origins of the ICM metallicity in galaxy clusters". This research was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #565 (*Multi-Wavelength Studies of the Culmination of Structure Formation in the Universe*). We acknowledge the financial contribution from the contracts Prin-MUR 2022 supported by Next Generation EU (M4.C2.1.1, n.20227RNLY3 *The concordance cosmological model: stress-tests with galaxy clusters*), and from the European Union's Horizon 2020 Programme under the AHEAD2020 project (grant agreement n. 871158). MDP acknowledges financial support from PRIN-MUR grant 20228B938N *Mass and selection biases of galaxy clusters: a multi-probe approach* funded by the European Union Next generation EU, Mission 4 Component 2 CUP B53D23004790006. MS acknowledges financial contributions from contract ASI-INAF n.2017-14-H.0, contract INAF mainstream project 1.05.01.86.10, INAF Theory Grant 2023: Gravitational lensing detection of matter distribution at galaxy cluster boundaries and beyond (1.05.23.06.17). MR acknowledges the financial contribution from INAF grant 1.05.24.02.10 LL acknowledges the financial contribution from the INAF grant 1.05.12.04.01. MG acknowledges support from the ERC Consolidator Grant *BlackHoleWeather* (101086804). JS was supported by NASA Astrophysics Data Analysis Program (ADAP) Grant 80NSSC21K1571. HB, FDL, PM acknowledge the support by INFN through the InDark initiative.

References

- Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2020, *Phys. Rev. D*, 102, 023509
- Allen, S. W., Schmidt, R. W., Ebeling, H., Fabian, A. C., & van Speybroeck, L. 2004, *MNRAS*, 353, 457
- Ansarfard, S., Rasia, E., Biffi, V., et al. 2020, *A&A*, 634, A113
- Arnaud, K. A. 1996, in *Astronomical Society of the Pacific Conference Series*, Vol. 101, *Astronomical Data Analysis Software and Systems V*, ed. G. H. Jacoby & J. Barnes, 17
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
- Bahar, Y. E., Bulbul, E., Ghirardini, V., et al. 2024, *A&A*, 691, A188
- Barnes, D. J., Kay, S. T., Henson, M. A., et al. 2017, *MNRAS*, 465, 213
- Barnes, D. J., Vogelsberger, M., Pearce, F. A., et al. 2021, *MNRAS*, 506, 2533
- Bartalucci, I., Molendi, S., Rasia, E., et al. 2023, *A&A*, 674, A179
- Battaglia, N., Bond, J. R., Pfrommer, C., & Sievers, J. L. 2013, *ApJ*, 777, 123
- Beauchesne, B., Clément, B., Hibon, P., et al. 2024, *MNRAS*, 527, 3246
- Beck, A. M., Murante, G., Arth, A., et al. 2016, *MNRAS*, 455, 2110
- Behroozi, P., Wechsler, R., & Wu, H.-Y. 2013, *ApJ*, 762, 109
- Biffi, V., Borgani, S., Murante, G., et al. 2016, *ApJ*, 827, 112
- Biffi, V., Dolag, K., & Böhringer, H. 2013, *MNRAS*, 428, 1395
- Biffi, V., Dolag, K., Böhringer, H., & Lemson, G. 2012, *MNRAS*, 420, 3545
- Bleem, L. E., Bocquet, S., Stalder, B., et al. 2020, *ApJS*, 247, 25
- Böhringer, H., Schuecker, P., Guzzo, L., et al. 2004, *A&A*, 425, 367
- Booth, C. M. & Schaye, J. 2009, *MNRAS*, 398, 53
- Braspenning, J., Schaye, J., Schaller, M., Kugel, R., & Kay, S. T. 2025, *MNRAS*, 536, 3784
- Bryan, G. L. & Norman, M. L. 1998, *ApJ*, 495, 80
- Bulbul, E., Liu, A., Kluge, M., et al. 2024, *arXiv e-prints*, arXiv:2402.08452
- Campitiello, M. G., Ettori, S., Lovisari, L., et al. 2022, *A&A*, 665, A117
- Cash, W. 1979, *ApJ*, 228, 939
- Chappuis, L., Eckert, D., Sereno, M., et al. 2025, *A&A*, 699, A141

- Chatzigiannakis, D., Pillepich, A., Simionescu, A., Truong, N., & Nelson, D. 2025, arXiv e-prints, arXiv:2503.01983
- CHEX-MATE Collaboration, Arnaud, M., Ettori, S., et al. 2021, A&A, 650, A104
- Clerc, N. & Finoguenov, A. 2022, arXiv e-prints, arXiv:2203.11906
- Croston, J. H., Pratt, G. W., Böhringer, H., et al. 2008, A&A, 487, 431
- Cruise, M., Guainazzi, M., Aird, J., et al. 2025, Nature Astronomy, 9, 36
- Cui, W., Knebe, A., Yepes, G., et al. 2018, MNRAS, 480, 2898
- De Luca, F., De Petris, M., Yepes, G., et al. 2021, MNRAS, 504, 5383
- Diehl, S. & Statler, T. S. 2006, MNRAS, 368, 497
- Dolag, K. 2015, in IAU General Assembly, Vol. 29, 2250156
- Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, MNRAS, 399, 497
- Dolag, K., Mevius, E., & Remus, R.-S. 2017, Galaxies, 5, 35
- Dolag, K., Remus, R.-S., Valenzuela, L. M., et al. 2025, arXiv e-prints, arXiv:2504.01061
- Eckert, D., Ettori, S., Pointecouteau, E., van der Burg, R. F. J., & Loubser, S. I. 2022, A&A, 662, A123
- Eckert, D., Finoguenov, A., Ghirardini, V., et al. 2020, The Open Journal of Astrophysics, 3, 12
- Eckert, D., Ghirardini, V., Ettori, S., et al. 2019, A&A, 621, A40
- Euclid Collaboration, Giocoli, C., Meneghetti, M., et al. 2024, A&A, 681, A67
- Evrard, A. E., Metzler, C. A., & Navarro, J. F. 1996, ApJ, 469, 494
- Fabjan, D., Borgani, S., Tornatore, L., et al. 2010, MNRAS, 401, 1670
- Fujita, Y., Fukushima, K., Sato, K., Fukazawa, Y., & Kondo, M. 2025, arXiv e-prints, arXiv:2507.00126
- Gardini, A., Rasia, E., Mazzotta, P., et al. 2004, MNRAS, 351, 505
- Gaspari, M., Tombesi, F., & Cappi, M. 2020, Nature Astronomy, 4, 10
- Gavdia, A., Kim, J., Sayers, J., et al. 2025, arXiv e-prints, arXiv:2507.10857
- Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, A&A, 689, A298
- Ghirardini, V., Eckert, D., Ettori, S., et al. 2019, A&A, 621, A41
- Gianfagna, G., De Petris, M., Yepes, G., et al. 2021, MNRAS, 502, 5115
- Gianfagna, G., Rasia, E., Cui, W., et al. 2023, MNRAS, 518, 4238
- Giocoli, C., Despali, G., Meneghetti, M., et al. 2025 [arXiv:2501.14019]
- Gupta, N., Saro, A., Mohr, J. J., Dolag, K., & Liu, J. 2017, MNRAS, 469, 3069
- Henson, M. A., Barnes, D. J., Kay, S. T., McCarthy, I. G., & Schaye, J. 2017, MNRAS, 465, 3361
- Hilton, M., Sifón, C., Naess, S., et al. 2021, ApJS, 253, 3
- Hirschmann, M., Dolag, K., Saro, A., et al. 2014, MNRAS, 442, 2304
- Jennings, F. & Davé, R. 2023, MNRAS, 526, 1367
- Kawahara, H., Suto, Y., Kitayama, T., et al. 2007, ApJ, 659, 257
- Kay, S. T., Braspennig, J., Chluba, J., et al. 2024, MNRAS, 534, 251
- Kim, J., Sayers, J., Sereno, M., et al. 2024, A&A, 686, A97
- Kitayama, T., Komatsu, E., Ota, N., et al. 2004, PASJ, 56, 17
- Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
- Knollmann, S. R. & Knebe, A. 2009, ApJS, 182, 608
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
- Kravtsov, A. V. & Borgani, S. 2012, ARA&A, 50, 353
- Lau, E. T., Kravtsov, A. V., & Nagai, D. 2009, ApJ, 705, 1129
- Lehmer, B. D., Xue, Y. Q., Brandt, W. N., et al. 2012, ApJ, 752, 46
- Lesci, G. F., Marulli, F., Moscardini, L., et al. 2025, arXiv e-prints, arXiv:2507.14285
- Lovisari, L., Ettori, S., Rasia, E., et al. 2024, A&A, 682, A45
- Mathiesen, B. F. & Evrard, A. E. 2001, ApJ, 546, 100
- Mazzotta, P., Rasia, E., Moscardini, L., & Tormen, G. 2004, MNRAS, 354, 10
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936
- Meneghetti, M., Fedeli, C., Zitrin, A., et al. 2011, A&A, 530, A17
- Meneghetti, M., Rasia, E., Merten, J., et al. 2010, A&A, 514, A93
- Miyazaki, S., Oguri, M., Hamana, T., et al. 2018, PASJ, 70, S27
- Mo, H. J. & White, S. D. M. 2002, MNRAS, 336, 112
- Muñoz-Echeverría, M., Macías-Pérez, J. F., Pratt, G. W., et al. 2024, A&A, 682, A147
- Muñoz-Echeverría, M., Pointecouteau, E., Pratt, G. W., et al. 2025, arXiv e-prints, arXiv:2510.18578
- Nagai, D., Vikhlinin, A., & Kravtsov, A. V. 2007, ApJ, 655, 98
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
- Nelson, K., Lau, E. T., Nagai, D., Rudd, D. H., & Yu, L. 2014, ApJ, 782, 107
- Pearce, F. A., Kay, S. T., Barnes, D. J., Bower, R. G., & Schaller, M. 2020, MNRAS, 491, 1622
- Pierre, M., Pacaud, F., Adami, C., et al. 2016, A&A, 592, A1
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014a, A&A, 571, A16
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014b, A&A, 571, A20
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
- Pointecouteau, E., Hattori, M., Neumann, D., et al. 2002, A&A, 387, 56
- Popesso, P., Marini, I., Dolag, K., et al. 2024, arXiv e-prints, arXiv:2411.17120
- Pratt, G. W., Arnaud, M., Biviano, A., et al. 2019, Space Sci. Rev., 215, 25
- Pratt, G. W., Arnaud, M., Maughan, B. J., & Melin, J.-B. 2022, A&A, 665, A24
- Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
- Rasia, E., Borgani, S., Murante, G., et al. 2015, ApJ, 813, L17
- Rasia, E., Ettori, S., Moscardini, L., et al. 2006, MNRAS, 369, 2013
- Rasia, E., Lau, E. T., Borgani, S., et al. 2014, ApJ, 791, 96
- Rasia, E., Mazzotta, P., Bourdin, H., et al. 2008, ApJ, 674, 728
- Rasia, E., Meneghetti, M., Martino, R., et al. 2012, New Journal of Physics, 14, 055018
- Rasia, E., Tripodi, R., Borgani, S., et al. 2025, arXiv e-prints, arXiv:2505.21624
- Remus, R.-S., Dolag, K., Naab, T., et al. 2017, MNRAS, 464, 3742
- Riva, G., Pratt, G. W., Rossetti, M., et al. 2024, A&A, 691, A340
- Rossetti, M., Eckert, D., Gastaldello, F., et al. 2024, A&A, 686, A68
- Ruppin, F., Mayet, F., Pratt, G. W., et al. 2018, A&A, 615, A112
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104
- Saxena, H., Sayers, J., Gavdia, A., et al. 2025, arXiv e-prints, arXiv:2505.23005
- Schaye, J. & Dalla Vecchia, C. 2008, MNRAS, 383, 1210
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, MNRAS, 402, 1536
- Scheck, D., Sanders, J. S., Biffi, V., et al. 2023, A&A, 670, A33
- Sereno, M., Maurogordato, S., Cappi, A., et al. 2025, A&A, 693, A2
- Springel, V. 2005, MNRAS, 364, 1105
- Springel, V. & Hernquist, L. 2003, MNRAS, 339, 289
- Staniszewski, Z., Ade, P. A. R., Aird, K. A., et al. 2009, ApJ, 701, 32
- Steinborn, L. K., Dolag, K., Hirschmann, M., Prieto, M. A., & Remus, R.-S. 2015, MNRAS, 448, 1504
- Sutherland, R. S. & Dopita, M. A. 1993, ApJS, 88, 253
- Teklu, A. F., Remus, R.-S., Dolag, K., et al. 2015, ApJ, 812, 29
- Tornatore, L., Borgani, S., Dolag, K., & Matteucci, F. 2007, MNRAS, 382, 1050
- Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
- Wicker, R., Douspis, M., Salvati, L., & Aghanim, N. 2023, A&A, 674, A48
- Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009, MNRAS, 399, 574
- XRISM Collaboration, Audard, M., Awaki, H., et al. 2025, arXiv e-prints, arXiv:2505.06533
- Xrism Collaboration, Audard, M., Awaki, H., et al. 2025a, ApJ, 985, L20
- Xrism Collaboration, Audard, M., Awaki, H., et al. 2025b, ApJ, 982, L5
- XRISM Collaboration, Audard, M., Awaki, H., et al. 2025, Nature, 638, 365
- Zuhone, J., Bahar, Y. E., Biffi, V., et al. 2023, A&A, 675, A150

Appendix A: Gas density and gas mass

We further investigate the reconstruction of gas density profiles and gas mass in this Appendix. Fig. A.1 shows the gas density reconstruction for one of The300 clusters. The true profile is shown in black, the reconstructed ones, including the voronoi binning technique, are shown in light blue and red. The bottom panel shows the ratio to the true profile, highlighting the better precision of the profile reconstructed from the voronoi binned image, with local deviations of at most 5%.

A detailed comparison is presented in Fig. A.2, where each of the three main panels corresponds to one simulation set. In each case, the top panel displays the gas density profiles reconstructed using the NFW, NP, and FM methods. The middle panel shows the ratio between the azimuthal median reconstructed and true profiles for each method, as derived from the voronoi-binned images. The bottom panel presents the ratio between the profiles measured directly from the mock EPIC images and the true profiles. Overall, we find excellent agreement among the density profiles recovered with the three methods. The median surface brightness profiles extracted from the voronoi images closely follow the one-to-one relation with the true profiles. In contrast, the direct analysis of the mock EPIC images tends to overestimate the gas density by approximately 5%. This result confirms that the voronoi-based approach enables an unbiased recovery of the gas density profile, owing to its reduced sensitivity to surface brightness fluctuations caused by substructures or colder gas clumps whose emission is not prominent enough to be masked during preprocessing. The profiles in *Magneticum* tend to have lower normalisation and a flatter slope. This is expected given the slightly different mass selection (see Sect. 2.4), because more massive clusters typically show steeper profiles (see e.g. Croston et al. 2008; Pratt et al. 2022). We conclude that despite the intrinsic differences between individual simulations, our methods are able to properly reconstruct the gas density profiles independently of their specific shape.

We estimate the total gas mass content in our mock clusters and compare it to the true value in the hydrodynamical simulations. The gas mass is computed by integrating the electron number density, following:

$$M_{\text{gas}} = 4\pi\mu_e m_p \int_0^{R_{500c}} n_e(r) r^2 dr, \quad (\text{A.1})$$

where μ_e is the mean electron molecular weight and m_p is the proton mass. In particular, we convert the gas particle number density to total gas mass in units of solar masses by computing the mean molecular weight per electron in a flexible way based on the chosen abundance table with the latest version of *hydromass*. For consistency with the generated boxes (see Sect. 3) we use the abundances from Asplund et al. (2009), which gives $\mu_e = 1.146$. Fig. A.3 shows the comparison between the measured and true total gas mass within the true R_{500c} in all three simulations. The panel refers to the gas mass estimated from the integral of the voronoi azimuthal median (mean surface brightness) density profile. When combining the three simulation, the median ratio between measured and true gas mass is 0.993, with 16th and 84th percentile points at 0.916, and 1.067, meaning that the gas mass reconstruction is precise within 1% and a cluster to cluster scatter variation of about 7.5%. We run a linear regression algorithm using *scipy* on the logarithm of the gas masses, accounting for a zero point of 14.0 on both axis. Overall, we find excellent agreement with the one to one relation. The best fit slope is closer to one especially for The300 and *Magneticum* using the median voronoi profiles. The fit is less accurate for

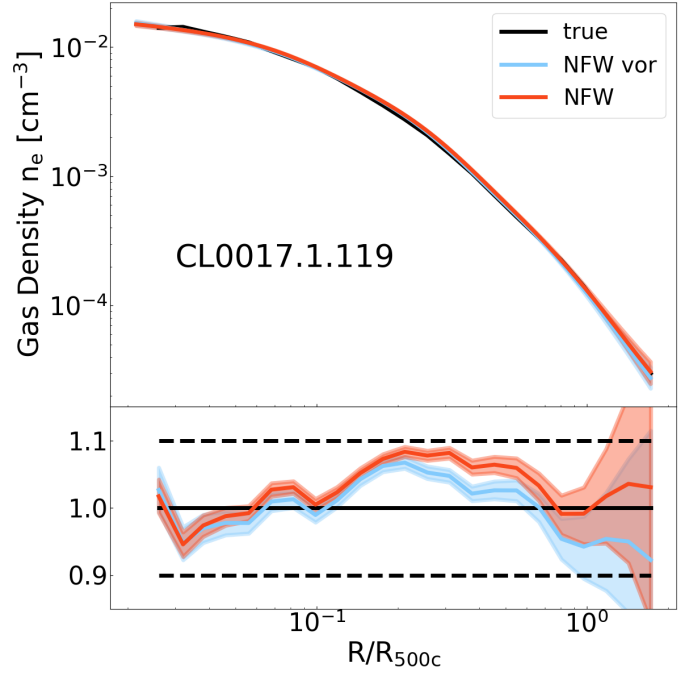


Fig. A.1. Density profile reconstruction of CL0017.1.119 from The300 simulation.

Table A.1. Best fit beta model parameters of the gas density profiles in our simulations.

Simulation	β	$r_c [R_{500c}]$
The300	0.63 ± 0.03	0.10 ± 0.02
The300 true	0.67 ± 0.03	0.13 ± 0.02
Magneticum	0.66 ± 0.03	0.29 ± 0.02
Magneticum true	0.69 ± 0.03	0.33 ± 0.02
MACSIS	0.70 ± 0.03	0.20 ± 0.02
MACSIS true	0.72 ± 0.03	0.21 ± 0.02

Table A.2. Best linear fit parameters of the gas mass recovery in our simulations.

Simulation	Slope	Intercept
The300 median	1.01 ± 0.01	0.01 ± 0.01
The300 mean	1.02 ± 0.02	0.04 ± 0.01
Magneticum median	1.05 ± 0.02	0.01 ± 0.01
Magneticum mean	1.07 ± 0.02	0.03 ± 0.02
MACSIS median	0.94 ± 0.05	0.05 ± 0.01
MACSIS mean	0.96 ± 0.07	0.06 ± 0.01

Notes. The relation is fitted on the base 10 logarithm of the gas masses, and the mass on the x-axis was normalised by 10^{14} to have a better handle on the correlation between slope and intercept.

MACSIS, which lacking the Tier1 mass objects has a limited mass range. We find the residuals to scatter around the mean relation by 0.027 (0.021, 0.022) dex in The300 (*Magneticum*, MACSIS). The gas mass is overestimated by about 5% using the standard count rate images, this is expected as we also showed that the gas density is overestimated at similar levels in the previous section. We report all parameters in Table A.2. We conclude that our methods are able to properly reconstruct the total gas mass in CHEX-MATE-like cluster samples.

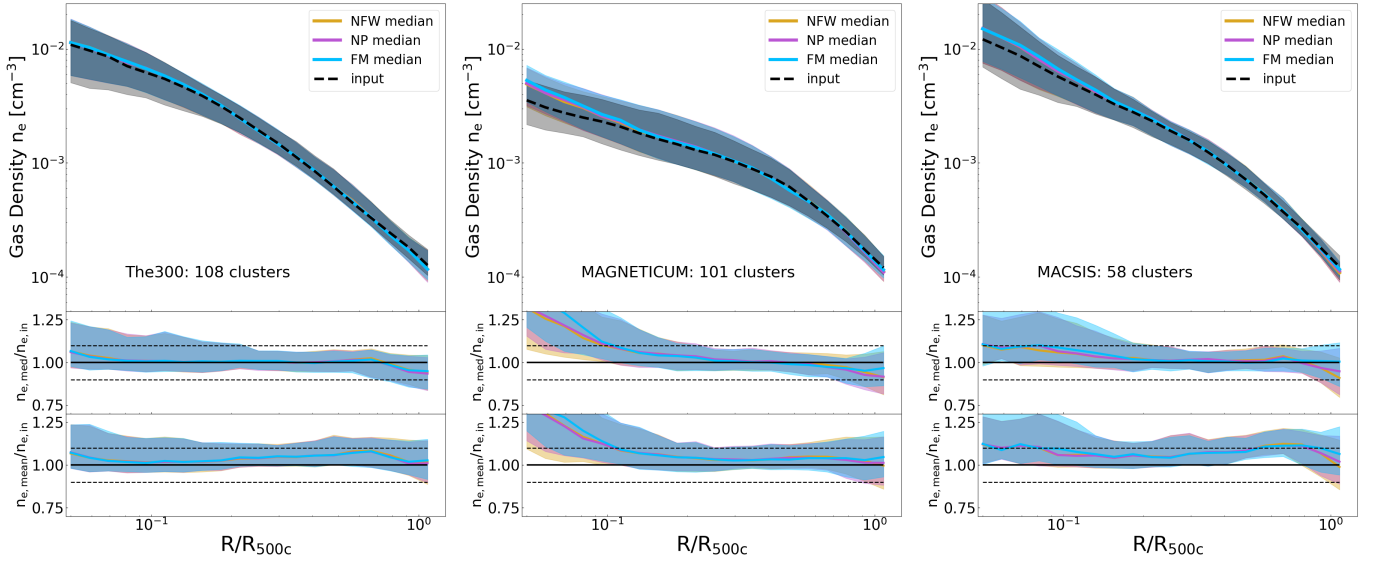


Fig. A.2. Deprojected 3D gas density profiles. Each panel corresponds to one simulation: *The300* on the left, *Magneticum* in the centre, and *MACSIS* on the right. The colours denote the three reconstruction methods. The smaller central (bottom) panels show the ratio between the median voronoi (the standard mean) profile and the true input one.

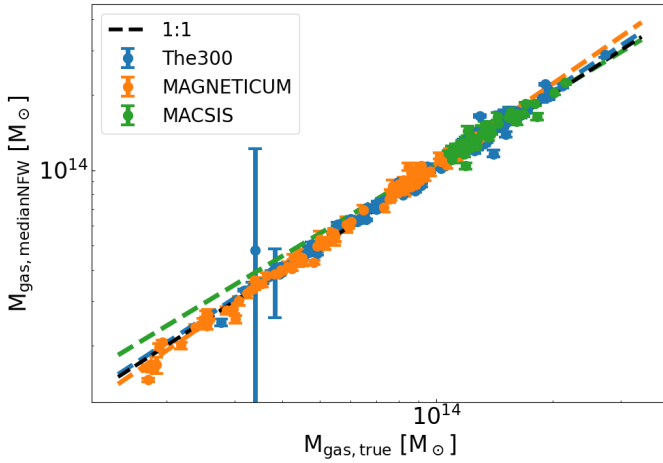


Fig. A.3. Comparison between the measured and true gas mass within R_{500c} . The panel refers to the masses inferred from the voronoi tessellated images. The dashed coloured lines denote the best linear fit to each simulation.

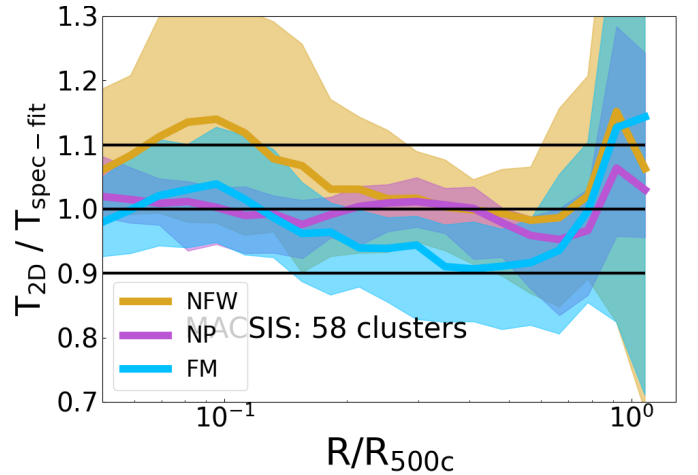


Fig. B.1. Ratio between the reconstructed 2D temperature profile with NFW, NP, and FM models to the result of the X-ray spectral fitting in *MACSIS*.

Appendix B: Thermodynamical profiles

In this Appendix we collect additional results about X-ray observables, starting with temperature. Fig. B.1 is about the temperature profile reconstruction in *MACSIS* (Sect. 5.2): it is the ratio between the model of the temperature profile and the measured one, for different reconstruction methods. It shows how the modelling is able to reproduce the measurement, and confirms that the temperature bias in the core of *MACSIS* (Fig. 5) is due to the measurement (see also Fig. 4) and not due to the modelling itself. Fig. B.2 is about corner plot showing the best fit parameters of the model describing the ratio between mass weighted and spectroscopic-like temperatures in Sect. 5.2.2 (see Eq. 3).

We then focus on derived observables, i.e. the ones that are not directly measured on the X-ray data, such as surface brightness and temperature. We focus on gas pressure and entropy. They are respectively reported in Fig. B.3 and B.4.

We find that both the thermal pressure and entropy profiles recovered from the X-ray analysis are systematically underestimated relative to the true values from the simulations. Among the three simulation suites, *The300* exhibits the smallest discrepancies: the thermal pressure is underestimated by approximately 10% across the entire radial range, while the entropy profile closely matches the true values beyond $0.2 \times R_{500c}$ and is underestimated by about 10% in the core. Both quantities remain consistent with the one-to-one relation within 1σ uncertainties. In *Magneticum*, the discrepancies are more pronounced. Thermal pressure is systematically underestimated by about 20% across the full profile when using the NFW and FM reconstruction methods, while the NP approach shows a smaller discrepancy (about 10%) around $0.4R_{500c}$ but exhibits even larger deviations in both the core and outskirts. A similar pattern is observed for entropy, although the reconstructed values converge more closely to the true values near R_{500c} . In *MACSIS*, the thermal pressure profile is moderately biased low by about 15%,

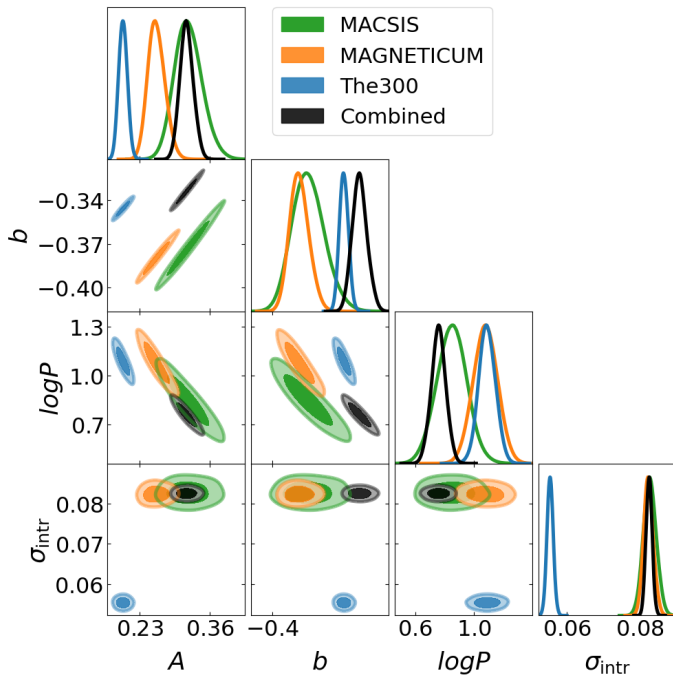


Fig. B.2. Corner plot with the marginalised 1d posterior distribution of the parameters describing the ratio between spectroscopic-like and mass weighted temperature as a function of median and width of the temperature fluctuation distribution.

while the entropy profile shows a level of disagreement similar to that seen in *Magneticum*. However, the three reconstruction methods yield more consistent results. In summary, the degree of agreement between the inferred and true thermodynamic profiles varies across simulations and partially depends on the reconstruction method employed. These results highlight the sensitivity of X-ray-derived pressure and entropy profiles to both physical modelling in simulations and methodological assumptions in the analysis pipeline. Understanding the origin of these discrepancies is essential, as both quantities play a central role in characterising the thermodynamic state of the ICM and in deriving cluster masses under the assumption of hydrostatic equilibrium.

The thermal pressure profile is proportional to the electron number density and the gas temperature, through the Boltzmann constant under the assumption of the ideal gas law. Since our density profiles are accurately recovered, as shown in Section 5, the pressure bias is more likely to be driven by either a bias in the reconstructed temperature, or by simplistic assumptions in the whole deprojection strategy and modelling. The hydrostatic mass modelling used in the temperature reconstruction assumes that the ICM is in equilibrium with the gravitational potential, neglecting any non-thermal pressure support from turbulence, bulk motions, or cosmic rays. If non-thermal pressure contributes significantly to the total pressure budget, the inferred temperature profile from a purely hydrostatic model will be suppressed to compensate, leading to an underestimation of thermal pressure. However, in this case the input pressure is only thermal, i.e. the one computed from the ideal gas law starting from density and temperature. Therefore, any inconsistency in the recovery of thermal pressure is not due to the non-thermal component in this work. In addition, the assumption of spherical symmetry in the hydrostatic analysis may not hold in detail: triaxiality of the halo and projection along preferential axes can distort both

the density and temperature gradient estimates and thus affect the pressure reconstruction.

Entropy is directly (inversely) proportional to temperature (density to the two-thirds power). It is a derived quantity that inherits biases from both temperature and density. Entropy is particularly sensitive to the thermal history of the ICM and is therefore more susceptible to localised structures, cooling clumps, or feedback-induced perturbations that may not be properly accounted for in the deprojection process under the assumption of single-temperature ICM within the framework of hydrostatic mass modelling.

Future work will focus on dissecting these contributions in more detail by comparing hydrostatic mass reconstructions to the true simulation masses, investigating the impact of different temperature reconstruction schemes, and exploring projection effects due to triaxiality and substructure. Such an analysis is essential to calibrate cluster thermodynamics for cosmological applications and to refine our interpretation of the hydrostatic mass bias.

Appendix C: Multi-T gas: a toy model

To explore projection effects on the cluster temperature profile in a controlled setting, we construct a toy model under the assumption of a spherically symmetric, single-temperature structure. This does not imply that the temperature profile is flat, but rather that the gas follows a single, smooth, and radially dependent profile without azimuthal or multi-phase complexity. We then project this 3D temperature distribution along the line of sight using both the SL and MW weighting schemes (see Eq. 2). The gas density profile is modelled with a standard β -model assuming $\beta = 2/3$ and a core radius of $r_c = 0.1 \times R_{500c}$. Our toy model includes a cool core and an outer decline and is parametrised by:

$$T(r) = T_0 \times (1 - A \exp[-(r/r_{c1})^{1.5}]) \times (1 + (r/r_{c2})^2)^{-0.3}, \quad (\text{C.1})$$

with $T_0 = 7$ keV, $A = 0.5$, $r_{c1} = 0.1$, and $r_{c2} = 0.5$. The projection is performed by looping over the radial grid, computing the line-of-sight depth corresponding to each projected radius, and integrating the temperature profile in Eq. C.1 with the appropriate SL or MW weights within each shell, so that the difference between the weighting schemes only applies to the projection.

The result is shown in Fig. C.1. We find that in the central region within $0.1 \times R_{500c}$, the MW projection yields higher temperatures than the SL projection. This behaviour arises because the SL weighting scheme gives more emphasis to lower-temperature gas, which is prevalent in the core, thereby biasing the projected temperature downward. Moving toward the outskirts, the density profile steepens and the line-of-sight integration length shortens, reducing the total amount of gas contributing to the projection. In this regime, the SL weighting becomes increasingly sensitive to the denser, hotter gas located in the inner portion of the shell due to its scaling with gas density squared. In contrast, the MW scheme continues to weight gas linearly by mass, giving more equal importance across the shell. As a result, the SL temperature can exceed the MW value beyond the core. At large radii beyond $3 \times R_{500c}$, the line-of-sight depth becomes small and the integrated signal is dominated by local gas, making both weighting schemes converge and the MW/SL ratio approach unity. These trends arise purely from projection effects applied to a spherically symmetric, single-temperature model.

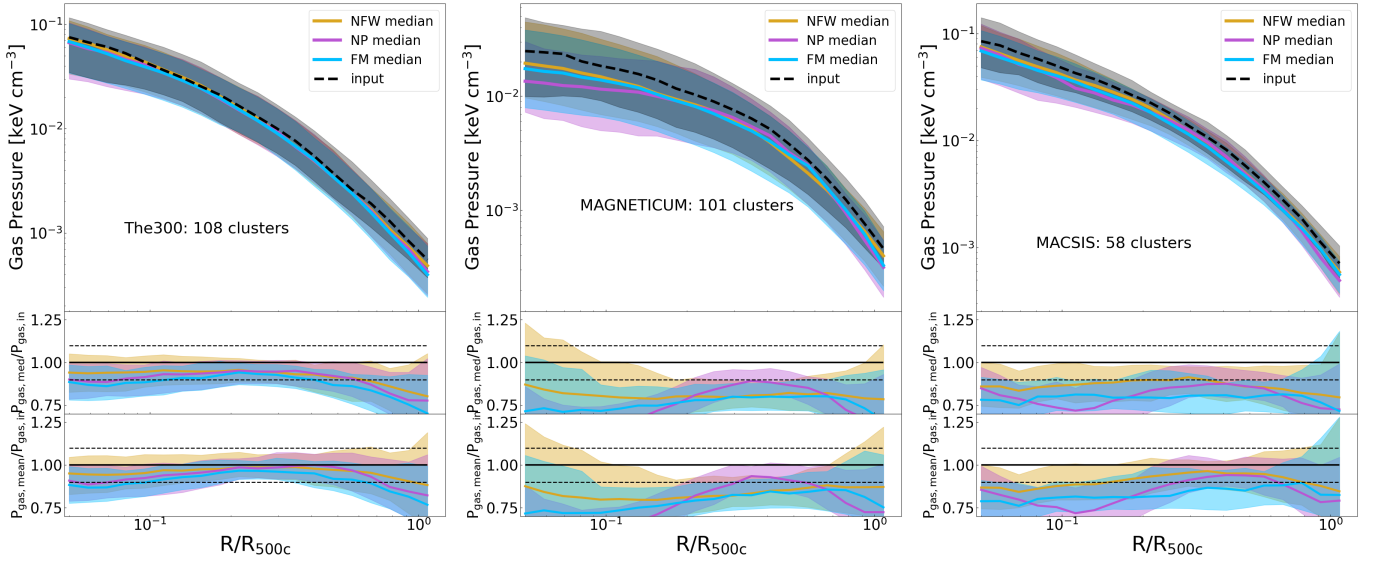


Fig. B.3. Deprojected 3D electron pressure profiles. Each panel corresponds to one simulation: *The300* on the left, *Magneticum* in the centre, and *MACSIS* on the right. The colours denote the three reconstruction methods. The smaller central (bottom) panels show the ratio between the median voronoi (the standard mean) profile and the true input one.

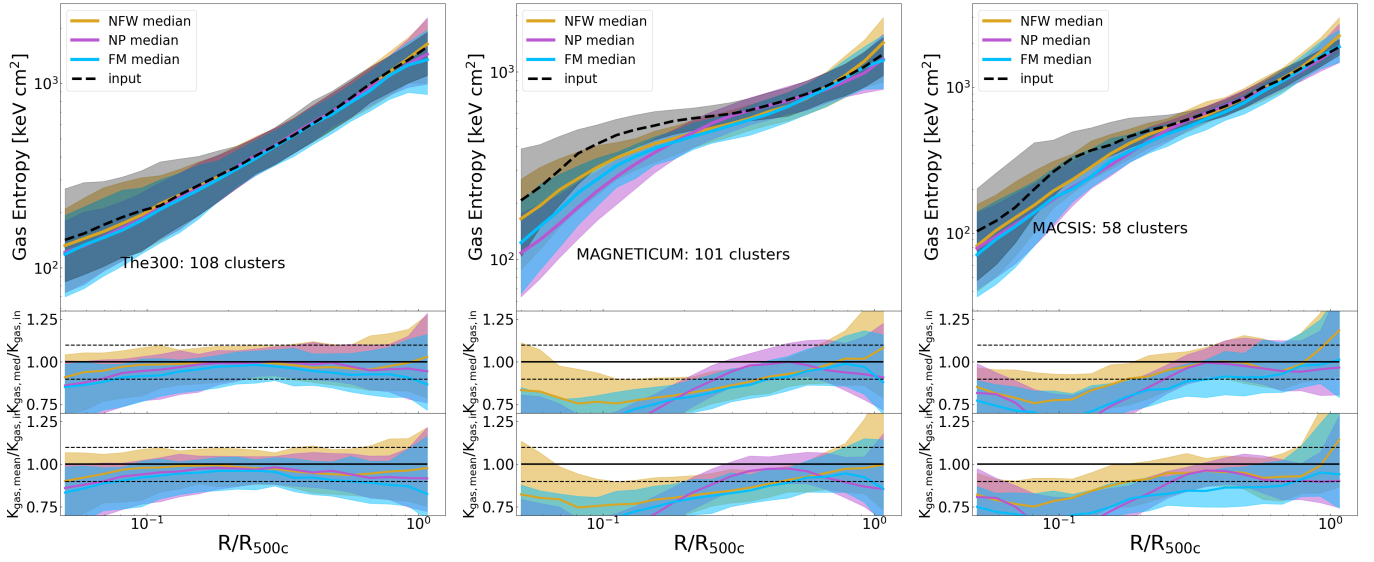


Fig. B.4. Deprojected 3D gas entropy profiles. Each panel corresponds to one simulation: *The300* on the left, *Magneticum* in the centre, and *MACSIS* on the right. The colours denote the three reconstruction methods. The smaller central (bottom) panels show the ratio between the median voronoi (the standard mean) profile and the true input one.

Appendix D: Voronoi temperature maps

To further investigate on the 2D temperature distribution in our simulations, we assess the intrinsic azimuthal variations in the gas temperature profiles within each radial shell (Rasia et al. 2014; Lovisari et al. 2024). To this end, we generate voronoi-tessellated images for each cluster, adopting a signal-to-noise ratio threshold of 30 per bin. We then extract and fit the X-ray spectrum in each voronoi region within R_{500c} , following the procedure outlined in Sect. 4. This yields temperature maps based on voronoi tessellation, which we use to characterize the azimuthal median temperature distribution at each radius. By comparing this to the results from direct X-ray spectral fitting, we can identify potential biases in the temperature measurement analogous to the analysis performed for surface brightness and gas density. The comparison is shown in Fig. D.1, where we plot the

median ratio between the temperature profiles obtained from the azimuthal median of the voronoi temperature maps in concentric annuli and those derived from direct X-ray spectral fitting. Shaded regions represent the 16th to 84th percentile spread of the distribution. In the innermost region within $0.1 \times R_{500c}$, the voronoi-derived temperature profile is very close or systematically higher than the one obtained from spectral fitting in annuli. This is expected, as the annular spectral fit is particularly sensitive to bright, dense, and cool gas components that dominate the X-ray emission in the core, biasing the temperature low. In contrast, the azimuthal median computed from the voronoi tessellation gives a more equal statistical weight to each spatial bin and is thus less affected by localized cool substructures. At such small radii the voronoi maps do not contain enough resolution elements to provide a meaningful comparison. Since they are on average less massive, reaching a given signal to noise

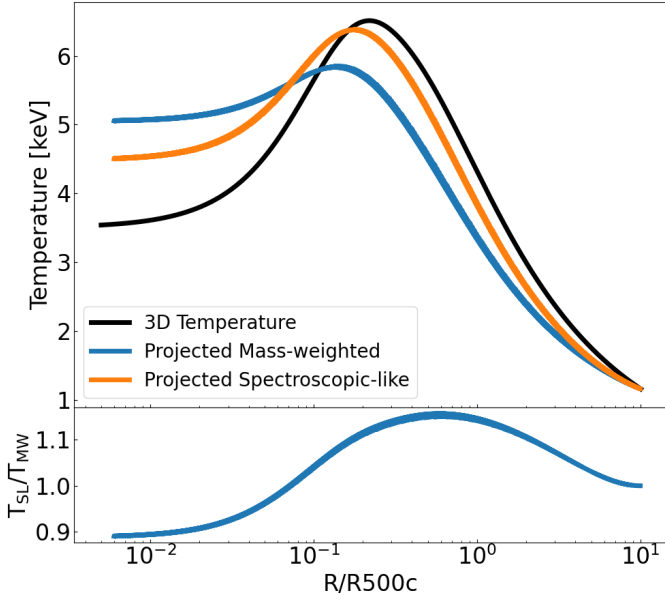


Fig. C.1. Spherical isothermal toy model of the gas temperature profile, including its projections along the line of sight assuming the spectroscopic-like and mass-weighted temperature schemes.

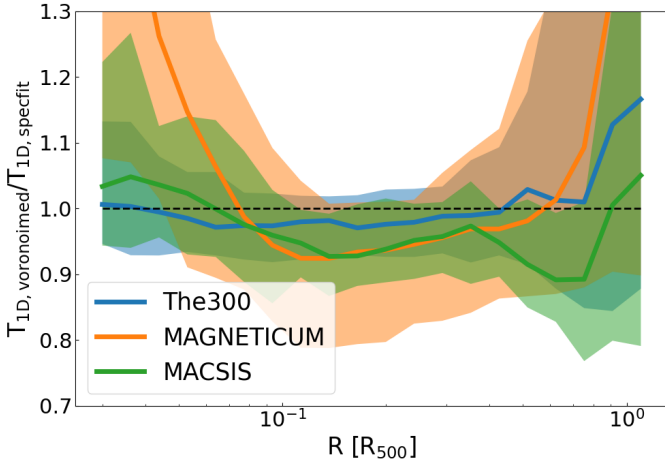


Fig. D.1. Comparison between the temperature profiles from X-ray spectral fitting and azimuthal median profiles obtained from voronoi tessellated temperature maps.

cut requires integrating larger regions. In the voronoi temperature maps, we resolve the innermost $0.05 \times R_{500c}$ with a median of 4 resolution elements, compared to 12 in MACSIS and 25 in The300. Between 0.1 and about $0.6 \times R_{500c}$, the trend reverses, with the voronoi temperatures falling below the annular spectral fits. Similarly to the projection effects discussed for the SL and MW schemes, this may reflect a stronger sensitivity of the annular spectral fitting to high-emissivity, denser, and hotter gas that dominates the X-ray signal outside the core, where both the temperature and density profiles decrease. The voronoi-based method, which downplays such emissivity-driven biases by construction, results in lower average temperatures. Beyond about $0.7\text{--}1 \times R_{500c}$, the scatter in the temperature ratio increases significantly. This is largely driven by the larger voronoi cells required to maintain the desired signal-to-noise threshold in the outskirts, where the surface brightness is low. The reduced spatial resolu-

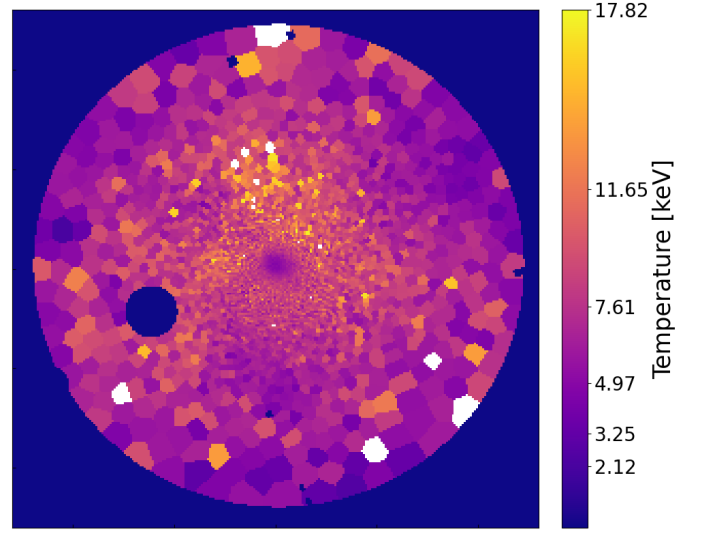


Fig. D.2. Voronoi-tessellated temperature map from the 2Ms XMM-Newton mock observation of the CL0030.1.115 cluster in The300 simulation.

tion in these regions weakens the reliability of the comparison and introduces larger uncertainties in the azimuthal statistics.

To assess the robustness of our technique and rule out potential systematics, such as the limited spatial resolution of the voronoi bins, low signal-to-noise ratio, or contamination from AGN, we generate a dedicated high-exposure simulation (as described in Sect. 3.2) for a single cluster. This mock observation features a much deeper exposure time of 2 Ms and excludes AGN emission. We apply the same X-ray analysis pipeline and find that all thermodynamic profiles are in excellent agreement with those recovered from the standard 25 ks simulation well within the uncertainties. We also produce the voronoi-tessellated temperature maps for this deep mock observation. Given the computational expense of this process, which involves extracting and fitting spectra in 6002 voronoi bins, we perform it for only one system. This serves as a baseline for comparing the median temperature profile derived from the 2 Ms voronoi map to that obtained from the standard 25 ks exposure. The full map extracted within R_{500c} is shown in Fig. D.2.

The median profiles extracted from the maps and the ones from X-ray spectral fitting in annular bins are reported in the upper panel of Fig. D.3, together with the input SL and MW ones. The bottom panel shows the relative intrinsic scatter in the voronoi median profiles extracted from the 25 ks and 2 Ms simulations. It is computed as the difference between the 84th and 16th percentile points of the distribution within each bin, divided by its median. This cluster has an $R_{500c} = 1491.5$ kpc and is located at $z = 0.33$, so that R_{500c} covers an angular scale of about 5.05 arcmin. We find good agreement between the result of the X-ray spectral fitting in annular bins between the 25 ks and 2 Ms simulations, with the latter being smoother and with lower uncertainties than the former by about one order of magnitude, the typical error bar is about 0.05 keV compared to 0.5 keV. This specific system shows a small inconsistency with the SL temperature in the inner core within 0.8 arcmin, corresponding to about $0.15 \times R_{500c}$. However, the same trend is visible in both the 25 ks and 2 Ms simulations. Similarly to the spectral fit in radial bins, also the median temperature profiles extracted from the voronoi maps are in agreement, with a discrepancy of about 6% at most, but still very well within the 1σ interval. Finally, in

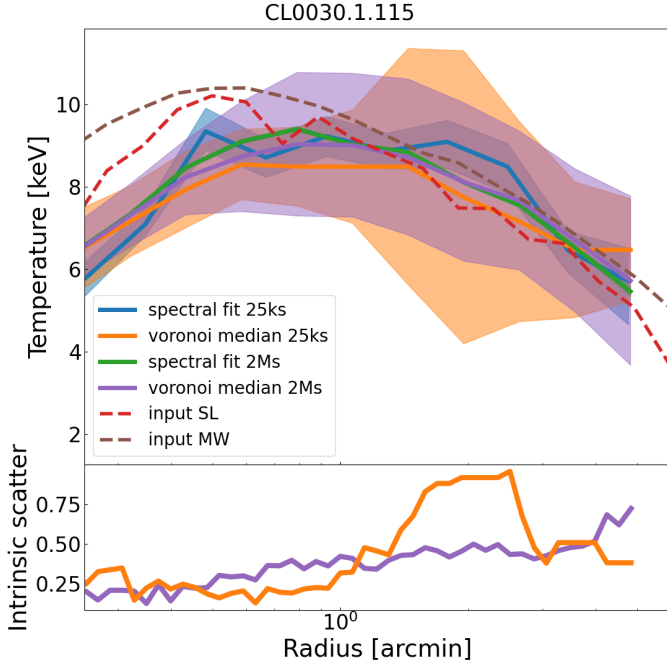


Fig. D.3. Temperature profiles extracted from the direct spectral fit and from the median of the voronoi maps for the 25ks and 2Ms simulations of CL0030.1.115.

both simulation we see an increasing trend of the intrinsic scatter as a function of radius, which is about 0.25 in the cluster core and get closer to 0.5 towards R_{500} . This is expected, as small regions in the core are more likely to include gas that is well described by a single temperature, whereas larger bins towards the outskirts are more likely to contain multi-phase gas, especially accounting for azimuthal variations in a single bin. The trend is smoother for the 2 Ms simulation, while the 25 ks one shows a big jump in the intrinsic scatter at about 2 arcmin, likely due to the low spatial resolution as a consequence of the large voronoi bins to achieve the desired signal to noise ratio. Nonetheless, the median properties of the profiles are overall in agreement between the 25 ks and the 2 Ms simulations.

Appendix D.1: Removing temperature fluctuations

We test the impact of removing the most deviating regions from the 2D temperature distribution on the projected radial temperature profile. This is an attempt of replicating a similar strategy applied to CHEX-MATE by [Lovisari et al. \(2024\)](#) from the perspective of the simulation. The authors identified regions deviating more than 1σ from the ratio between the local 2D temperature map and the projected 1D profile. Such regions are likely associated to hot and cold gas clumps. We replicate the approach using 2D T_{SL} maps and 1D T_{SL} profiles, removing any systematic related to the temperature measurement by construction. In practice, for each pixel in the T_{SL} maps we compute the quantity $S_i = \frac{T_{1D,SL} - T_{1D,SLmap}}{\sigma_{T2D}}$, where σ_{T2D} is the standard deviation of the temperature map at a given radius. We then select and remove pixels with $|S_i| > 1$ and recompute the 1D median profile without outliers. The result is shown in Fig. D.4. We find that after removing outliers the temperature estimate increases by about 15-20%, which means that we are preferentially removing clumps of cold gas that are easier to detect. This matches the result ob-

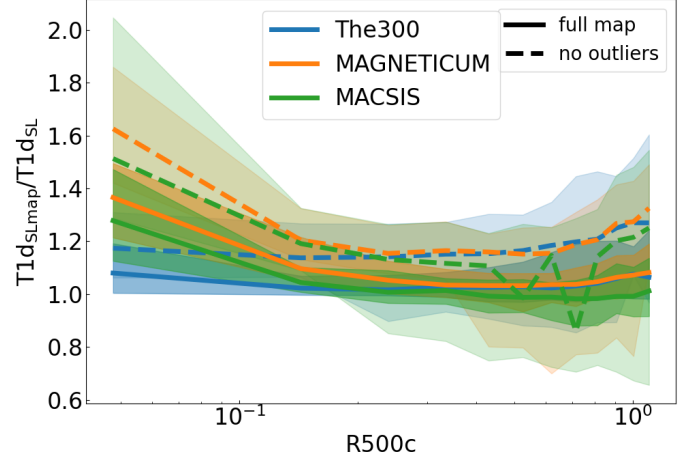


Fig. D.4. Comparison between the median of the 2D spectroscopic-like temperature maps and the 1D radial profile, before and after removing the most deviating regions.

tained by [Lovisari et al. \(2024\)](#), who observed a similar trend with effects up to 10-20%.

Appendix E: Mis-centering

In this appendix we test whether any inconsistencies in the reconstruction of gas density profiles in Sect. 5 is due to mis-centering. In fact, from the observer's perspective, the profile is computed from the peak of the X-ray emission, while the input profiles use the position of the most bound particle, i.e. the deepest point of the potential well, as centre.

From the full population in each simulation we subselect well-centred systems where the offset between the peak of the X-ray emission and the dark matter is $\Delta_X < 0.007 \times R_{500c}$ and study the ratio between inferred and true density profiles. We obtain respectively 25, 15, and 15 well-centred systems in The300, Magneticum, and MACSIS. The result is shown in Fig. E.1, with the solid lines denoting the full samples and the dashed ones the well centred ones. The shaded areas denote the 16th-84th percentile points of the selected population. The ratio is basically unchanged in The300, which does not show significant inconsistencies even with the full sample. The issue is fully solved in MACSIS, where the ratio to true density is within 5% down to $0.06 \times R_{500c}$, whereas it crosses this threshold already at about $0.15 \times R_{500c}$ for the full sample. For Magneticum the discrepancy decreases drastically, reaching a radius of about $0.08 \times R_{500c}$ where the gas density is within 5% of the true value, compared to about $0.13 \times R_{500c}$ for the full sample. Additional differences may be due to deviations from sphericity. In any case, the gas mass enclosed in such regions is a small fraction of the total one and does not impact the total gas mass reconstruction, as shown in Fig. A.3.

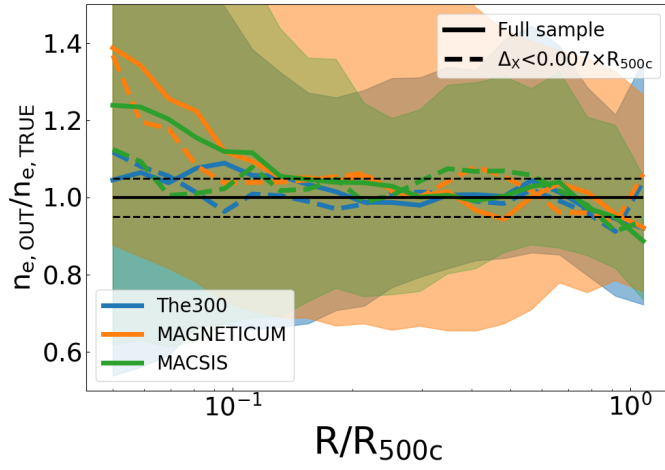


Fig. E.1. Ratio between reconstructed and true density profiles for all three simulations. The solid lines denote the full samples, the dashed ones refer only to the well centred objects, where Δ_X is the offset between the identified X-ray centre and the dark matter one. The black dashed lines denote the $\pm 5\%$ ratio.