

Bound or blown: The fate of hot gas in galaxy groups

R. Seppi^{1,*}, D. Eckert¹, J. Schaye², J. Braspenning³, M. Schaller^{2,4}, B. D. Oppenheimer⁵, E. O’Sullivan⁶,
F. Gastaldello⁷, L. Lovisari^{7,6}, M. A. Bourne^{8,9}, M. Sun¹⁰, A. Finoguenov¹¹, H. Khalil¹¹,
G. Gozaliasl^{11,12}, K. Kolokythas^{13,14,15}, Y. E. Bahar¹⁶, and R. Santra¹⁷

¹ Department of Astronomy, University of Geneva, Ch. d’Ecogia 16, CH-1290 Versoix, Switzerland

² Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

³ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

⁴ Lorentz Institute for Theoretical Physics, Leiden University, PO box 9506, 2300 RA Leiden, The Netherlands

⁵ University of Colorado, Center for Astrophysics and Space Astronomy, 389 UCB, Boulder, CO 80309, USA

⁶ Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

⁷ INAF, Istituto di Astrofisica Spaziale e Fisica Cosmica di Milano, via A. Corti 12, 20133 Milano, Italy

⁸ Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

⁹ Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

¹⁰ Department of Physics and Astronomy, University of Alabama in Huntsville, Huntsville AL35899, USA

¹¹ Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2A, Helsinki FI-00014, Finland

¹² Department of Computer Science, Aalto University, PO Box 15400, Espoo FI-00076, Finland

¹³ Centre for Radio Astronomy Techniques and Technologies, Department of Physics and Electronics, Rhodes University, PO Box 94, Makhanda 6140, South Africa

¹⁴ South African Radio Astronomy Observatory, Black River Park North, 2 Fir St, Cape Town 7925, South Africa

¹⁵ INAF – Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy

¹⁶ INAF, Osservatorio di Astrofisica e Scienza dello Spazio, via Piero Gobetti 93/3, I-40129 Bologna, Italy

¹⁷ International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560089, India

Received 23 March 2026 / Accepted 24 April 2026

ABSTRACT

The impact of feedback from active galactic nuclei (AGN) on the hot gas content of galaxy groups remains a key uncertainty in galaxy formation and its connection to the large-scale structure of the Universe. We aim to compare the *XMM-Newton* Group AGN Project (X-GAP) sample to the hydrodynamical FLAMINGO simulations, which span a wide range of AGN feedback prescriptions. We constructed X-GAP analogues by forward modelling the full selection function, including detection and observational systematics, and we generated end-to-end *XMM-Newton* mock observations analysed consistently with the data. We studied multiple observables, including the L - T and M_{gas} - T relations, the number of groups, mean temperature, and velocity dispersion, accounting for their covariance. The forward model accurately recovers input luminosities, gas masses, and core-excised temperatures for regular systems, enabling direct comparison in observable space. The normalisation of the scaling relations is the best discriminator between feedback models, while cosmic variance introduces $>20\%$ fluctuations in the number of detected systems, making counts alone a weak discriminator. Models with intermediate feedback strength provide the best agreement with X-GAP, with the $f_{\text{gas}} - 2\sigma$ model yielding the lowest tension of only 0.8σ , while the most extreme feedback scenario ($f_{\text{gas}} - 8\sigma$) is ruled out at $>4\sigma$. Our results indicate that the thermodynamic properties of galaxy groups favour feedback stronger than the fiducial FLAMINGO calibration, but they disfavour the most ejective models. This highlights the importance of combining forward modelling and multi-observable constraints to probe the fate of hot baryons in low-mass haloes.

Key words. methods: data analysis – surveys – galaxies: clusters: intracluster medium – galaxies: groups: general – large-scale structure of Universe – X-rays: galaxies: clusters

1. Introduction

Understanding galaxy formation requires a comprehensive view of how baryons are accreted, transformed, and redistributed within dark matter haloes across cosmic time (Wechsler & Tinker 2018). While only a small fraction of baryons condense into stars, the majority remain in diffuse gas phases, making the thermal and spatial distribution of baryons a key tracer of feedback and accretion processes in structure formation in the large-scale structure (LSS) of the Universe

(Fukugita et al. 1998; Fukugita & Peebles 2004; Ayromlou et al. 2023).

Galaxy groups, the most common hosts of galaxies in the Universe and occupying the intermediate-mass regime between isolated galaxies and rich clusters (Tinker et al. 2008), are particularly sensitive to baryonic processes (Mulchaey 2000; Giodini et al. 2009; Eckmiller et al. 2011). With halo masses of $M_{500c} \sim 10^{13} - 10^{14} M_{\odot}^1$, their relatively shallow potential makes

¹ M_{500c} is the mass enclosed within R_{500c} , i.e. the radius encompassing an average density that is 500 times larger than the critical density of the universe, $\rho_c = 3H^2(z)/8\pi G$, at a given redshift z .

* Corresponding author: riccardo.seppi@unige.ch

them especially susceptible to non-gravitational heating from supernovae and active galactic nuclei (AGNs; Padovani et al. 2017), allowing feedback to strongly modify their gas content and thermodynamic structure (McNamara & Nulsen 2007; McCarthy et al. 2010; Le Brun et al. 2014; Gaspari et al. 2020).

Understanding the thermodynamic state of the hot gas in galaxy groups is crucial for constraining models of galaxy formation and evolution. X-ray and Sunyaev-Zel'dovich (SZ) observations show that the gas in groups is less dense, more extended, and has higher entropy than expected from purely gravitational collapse, due to the role of AGN feedback in redistributing baryons (Sun et al. 2009; Pratt et al. 2009; Johnson et al. 2009; Stott et al. 2012; Etori et al. 2013; Lovisari et al. 2015; Gozaliasl et al. 2019). However, the precise impact of feedback on the intra-group medium (IGrM), and the extent to which simulations reproduce these effects, remain open questions (Ponman et al. 1999; Finoguenov et al. 2002; Oppenheimer et al. 2021; Bahar et al. 2024).

In recent years, cosmological hydrodynamical simulations have advanced substantially through increasingly sophisticated sub-grid models for star formation, metal enrichment, and AGN feedback. The latter is now recognised as a key ingredient to reproduce several observed properties, including the suppression of cooling flows in cluster cores (Brighenti & Mathews 2006; Fabian 2012; Hlavacek-Larrondo et al. 2022), the quenching of star formation (Silk & Rees 1998; Booth & Schaye 2009; Pillepich et al. 2018), and the scaling relations between black holes and their host galaxies (Magorrian et al. 1998; Kormendy & Ho 2013; Sahu et al. 2019; Habouzit et al. 2021). Modern simulations adopt different AGN feedback implementations. For example, cosmo-OWLS (Le Brun et al. 2014), EAGLE (Schaye et al. 2015), BAHAMAS (McCarthy et al. 2017), FLAMINGO (Schaye et al. 2023), and COLIBRE (Schaye et al. 2026) use isotropic thermal energy injection linked to gas accretion onto the supermassive black hole (SMBH; Booth & Schaye 2009). Other models include radio-mode bubbles (Illustris, FABLE; Vogelsberger et al. 2014; Henden et al. 2018), mechanical or radiative feedback channels (MAGNETICUM; Steinborn et al. 2015), or kinetic feedback schemes (HorizonAGN, IllustrisTNG, SIMBA; Dubois et al. 2016; Weinberger et al. 2017; Davé et al. 2019). Some simulations additionally implement directional jet feedback, including SIMBA, selected FLAMINGO runs, and hybrid AGN feedback models in COLIBRE (Huško et al. 2026). While these simulations are calibrated to reproduce the galaxy stellar mass function (GSMF) and, in some cases, properties of massive clusters, their predictions for lower-mass haloes remain sensitive to the details of feedback modelling. Galaxy groups therefore provide a powerful test-bed for assessing the realism of baryonic physics in simulations, which is closely tied to the efficiency with which feedback expels baryons from dark matter haloes. (Oppenheimer et al. 2021; Eckert et al. 2021; Gastaldello et al. 2021; Lovisari et al. 2021; Eckert et al. 2025).

Multiple studies suggest substantial ejection of hot gas in galaxy groups, although its strength remains uncertain (Gastaldello et al. 2007; Sun et al. 2009; Rasmussen & Ponman 2009; Gonzalez et al. 2013; Laganá et al. 2013; Sanderson et al. 2013; Akino et al. 2022). More recently, stacking analyses of eROSITA data around optically selected systems have inferred gas fractions of $\lesssim 5\%$ for haloes in the $10^{13.5}-10^{14.5} M_{\odot}$ range, below the predictions of most hydrodynamical simulations (Popesso et al. 2025). Independent constraints from entropy profiles and kinetic SZ (kSZ) measurements also indicate low baryon fractions in groups (Molendi et al. 2025; Siegel et al.

2025a; Roper et al. 2025). In particular, analyses combining ACT, DESI, and eROSITA data favour stronger AGN feedback models, including the most ejective FLAMINGO variants (Siegel et al. 2025a,b). However, recent XRISM observations suggest lower velocity dispersions and kinetic pressure than expected, potentially indicating that some feedback models may be overly ejective (XRISM Collaboration 2025). Similarly, analyses combining Planck and DES data find weaker power-spectrum suppression than predicted by simulations (Xu et al. 2025), highlighting ongoing tensions that sensitively depend on the adopted feedback prescription (e.g. including jets; Bourne & Yang 2023). By analysing the hot atmospheres of galaxy groups with a forward-modelling pipeline that reproduces the observational selection and observable measurements, we performed a direct comparison between simulations and data, thereby constraining the strength of baryon expulsion from massive haloes.

The *XMM-Newton* Group AGN Project (X-GAP; Eckert et al. 2024) is a sample of 49 galaxy groups observed with *XMM-Newton*, selected from the All-Sky Extended Sources project (AXES; Damsted et al. 2024; Khalil et al. 2024) using X-ray data from the ROSAT all-sky survey (RASS) and optical information from the Sloan Digital Sky Survey (SDSS), based on optically identified groups (Tempel et al. 2017). X-GAP includes peculiar systems such as SDSSTG-4436, which shows a high entropy floor and a kinked density profile, providing valuable insight into the cumulative impact of AGN activity across cosmic time (Eckert et al. 2025), as well as sources with extended radio emission (Santra et al. 2026). In a previous work, we forward modelled the selection process using end-to-end simulations that reproduce the detection scheme (Seppi et al. 2025). Eckert et al. (2026) present a first comparison between the thermodynamic properties of X-GAP groups and predictions from the hydrodynamical FLAMINGO simulations (Schaye et al. 2023; Kugel et al. 2023), which provide an important benchmark thanks to multiple runs with different feedback implementations.

For this study, we took a further step compared to Eckert et al. (2026) by generating and analysing detailed end-to-end *XMM-Newton* mock observations of X-GAP-like groups selected from FLAMINGO. We used the selection function from Seppi et al. (2025) to identify an X-GAP-like sample in FLAMINGO, which allowed for a direct one-to-one comparison with simulations accounting for selection effects in real observations. This framework enables a stringent test of whether state-of-the-art hydrodynamical models reproduce observations of galaxy groups in the nearby Universe. This paper is organised as follows. Section 2 describes the data and simulations, Sect. 3 the forward modelling and analysis of the *XMM-Newton* mocks, Sect. 4 presents the results, Sect. 5 discusses them in the context of recent complementary measurements, and Sect. 6 provides a summary and outlook. We adopted the FLAMINGO cosmology with a cosmological constant Λ and cold dark matter (CDM), i.e. a Λ CDM model with parameters from the Dark Energy Survey plus external constraints (DES Y3; Abbott et al. 2022): $H_0 = 68.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.306$, and $\Omega_{\Lambda} = 0.694$.

2. Data and simulations

In this section we describe the data and simulations used in this work: the X-GAP group sample, the FLAMINGO simulation suite, and the selection of simulated X-GAP analogues therein.

2.1. X-GAP and its selection

X-GAP² (Eckert et al. 2024) is an *XMM-Newton* large programme targeting galaxy groups (10^{13} – $10^{14} M_{\odot}$) to map IGrM properties out to R_{500c} . By comparing observed integrated quantities (T , L , M_{gas}), gas fractions, and thermodynamic profiles to hydrodynamical simulations, the project aims to constrain AGN feedback models. The sample originates from the AXES group catalogue (Damsted et al. 2024), which cross-matched wavelet-detected ROSAT sources (Käfer et al. 2019) with optically detected groups from SDSS galaxies ($r < 17.77$; Tempel et al. 2017, $z < 0.05$). In Seppi et al. (2025), we quantified AXES completeness and purity using end-to-end simulations of a full-sky light cone. We derived a robust selection function model based directly on observables, X-ray flux (F_X), galaxy velocity dispersion ($\sigma_{v,T}$), and redshift (z), rather than indirect halo properties. It reads

$$P_{\text{det}}(F_X, z, \sigma_{v,T}) = (1 + \exp[-\alpha_{F_X}(\log_{10} F_X - F_{X,0}) + \alpha_{\sigma v} \times \log_{10} \sigma_{v,T} + \alpha_z \times \log_{10} z])^{-1}, \quad (1)$$

where $F_{X,0}$, α_{F_X} , $\alpha_{\sigma v}$, and α_z are model parameters. We use the median point for each best-fit parameter in the posterior chains from Seppi et al. (2025).

X-GAP is designed to mitigate the problems in the optical identification of galaxy groups, such as projection effects and merging of unvirialised systems (Robotham et al. 2011; Old et al. 2014; Marini et al. 2025b) with purer X-ray selection based on the IGrM. The trade-off is that large X-ray surveys tend to be shallow, limiting the group identification to low redshift. Indeed, Seppi et al. (2025) showed that the cross-match with X-ray data serves as a cleaning step of the optical candidates, increasing purity but reducing completeness. The X-ray detection via wavelet filters is particularly beneficial for identifying the outskirts of galaxy groups, as it is not sensitive to the shape of the surface brightness profile, thereby reducing biases related to the presence of a cool core (Eckert et al. 2011; Seppi et al. 2021). This is especially critical at low masses, while high-mass, brighter systems are typically detected regardless of dynamical state, lower sensitivity at lower masses can limit detection to centrally peaked systems, introducing significant biases into the selection function modelling.

2.2. FLAMINGO

FLAMINGO³ is a large suite of cosmological hydrodynamical simulations (Schaye et al. 2023). It features various box sizes and resolutions. FLAMINGO is calibrated to reproduce the $z = 0$ GSMF and the gas mass fraction of low-redshift galaxy clusters (Kugel et al. 2023). In this work we use the high-resolution 1 Gpc box, i.e. with 3600^3 dark matter particles and gas particles mass (m_g) of $10^8 M_{\odot}$ (here after L1_m8), as a benchmark for the sample selection. In a second step, we use the intermediate resolution, with 2×1800^3 particles and m_g of $10^9 M_{\odot}$ (L1_m9), to test the various implementations of baryonic physics. The L1_m9 set occupies the lower-resolution/large volume parameter space compared to smaller simulations, such as EAGLE (Schaye et al. 2015), TNG (Pillepich et al. 2018), or SIMBA (Davé et al. 2019). Gas particle masses are not strictly fixed, as they incorporate mass and metals by neighbouring particles through outflows and feedback (Schaye et al. 2023). On top of the fiducial hydro model, FLAMINGO provides calibrations to

lower GSMF ($M_{*}-\sigma$), and to different gas fractions in clusters ($f_{\text{gas}} \pm N\sigma$). In particular, the $N\sigma$ label indicates the shift in the gas fraction used for calibration: more negative values (e.g. -2σ) correspond to stronger AGN feedback, which further depletes the gas content of haloes and reduces the gas fraction at a fixed mass. Stellar and AGN feedback parameters are jointly tuned to reproduce the chosen sets of observables (Kugel et al. 2023).

L1_m9 includes different AGN feedback implementations: thermal, isotropic energy injection (Booth & Schaye 2009), and kinetic feedback with AGN jets (Huško et al. 2022). FLAMINGO models feedback using a black hole (BH) energy reservoir that stores a fraction of the accreted rest-mass energy. In the thermal model, energy is injected once it exceeds a threshold corresponding to heating a neighbouring gas particle by ΔT_{AGN} (see Sect. 2.3.6 in Schaye et al. 2023). This parameter sets the temperature jump and thus regulates the burstiness of feedback events. In the jet model, the threshold corresponds to $m_g v_{\text{jet}}^2$, and once reached, two particles are kicked within a cone aligned with the BH spin at velocity v_{jet} , injecting kinetic energy in a collimated outflow. Both ΔT_{AGN} and v_{jet} are free calibration parameters. Haloes are identified with HBTHERONS (Forouhar Moreno et al. 2025), which locates centres at the most bound particle, and halo properties are computed with the Spherical Overdensity and Aperture Processor (SOAP; McGibbon et al. 2025).

2.3. Group sample selection

X-GAP groups are identified from SDSS galaxies with r -band magnitude $r < 17.77$. At very low redshifts $z < 0.02$, it includes galaxies with stellar masses below $10^9 M_{\odot}$, unresolved in L1_m9. We therefore use the higher-resolution L1_m8 run to properly resolve the X-GAP galaxy population.

We do not explicitly model spectroscopic incompleteness: the SDSS main galaxy sample is highly complete, with the dominant missing fraction of about 5% arising from fibre collisions (fibres cannot be placed within 55 arcsec of each other; Strauss et al. 2002; Eisenstein et al. 2011; Alam et al. 2015). For the nearby groups considered here, subtending large angular extents on the sky, this effect is expected to be modest. We provide full details for the m_r modelling in FLAMINGO in Appendix A.

To select X-GAP analogues in FLAMINGO we use a single snapshot at $z = 0$, which extensively covers the proper distance within $z < 0.05$ corresponding to about 250 Mpc. Therefore, there is no need to concatenate different snapshots. Starting from the halo catalogue, we place an observer at the box centre and construct a light cone by converting Cartesian coordinates (X, Y, Z) to (RA, Dec, z) following the geometry of Seppi et al. (2025).

To predict X-ray fluxes for FLAMINGO haloes we use the true X-ray luminosities derived from the emissivity of gas particles within r_{500c} (Braspenning et al. 2024). A key methodological difference concerns the luminosity definition: in FLAMINGO it is computed as a spherical (3D) integral of the gas emissivity within r_{500c} (Braspenning et al. 2024), whereas in X-GAP (Eckert et al. 2024) and in our selection modelling (Seppi et al. 2025) L_{500c} is obtained from a cylindrical (projected) integral of the emission measure within a projected radius R_{500c} . The cylindrical definition includes additional emission along the line of sight, implying $L_{500c}^{\text{cyl}} > L_{500c}^{\text{sph}}$ for the same gas distribution.

Computing L_{500c}^{cyl} for every single halo in the FLAMINGO boxes is computationally expensive. Therefore, to place the

² <https://www.astro.unige.ch/xgap>

³ <https://flamingo.strw.leidenuniv.nl>

two on a common footing, we derive a multiplicative conversion using β -model gas density profiles. We denote by $n_e n_H \sim n_{e,0}^2 \left[1 + \left(\frac{r}{r_c} \right)^2 \right]^{-3\beta}$ the 3D emission measure density, while $EM(R)$ represents the line-of-sight emission measure (also known as emission integral). We compute both the spherical and cylindrical integrals,

$$\begin{aligned} L_X^{\text{sph}} &= \int_0^{r_{500c}} 4\pi r^2 n_e n_H(r) \Lambda(T, Z) dr \\ EM(R) &= 2 \int_R^\infty \frac{n_e n_H(r)}{\sqrt{r^2 - R^2}} dr \\ L_X^{\text{cyl}} &= \int_0^{R_{500c}} 2\pi R EM(R) \Lambda(T, Z) dR, \end{aligned} \quad (2)$$

where Λ is the cooling function that depends on temperature and metallicity (see Eckert et al. 2020). We interpolate the cumulative sum of the integrals to R_{500c} and evaluate the ratio $C_{\text{cyl}} = L_{500c}^{\text{cyl}}/L_{500c}^{\text{sph}}$ over a grid in slope β and core radius r_c . X-GAP groups are well described by β -models with β around 0.45 and r_c below $0.1 \times R_{500c}$. Exploring slopes between 0.4 and 0.5, and core radii below 0.05 on a fine grid we obtain a median correction of $C_{\text{cyl}} = 1.16$ with a scatter of 0.06.

Because feedback variants in FLAMINGO may modify the ICM profile shape (weaker feedback producing peakier cores and stronger feedback flatter distributions), we adopt a conservative conversion between spherical and cylindrical luminosities. We assume a log-normal distribution with median 0.1 dex ($C_{\text{cyl}} \approx 1.26$) and scatter 0.06 dex, truncated at $C_{\text{cyl}} = 1$ since the cylindrical luminosity must exceed the spherical one. For each halo we draw C_{cyl} and compute $L_{500c}^{\text{cyl}} = C_{\text{cyl}} L_{500c}^{\text{sph}}$. We verified that this correction does not significantly bias the reconstructed luminosity (see Appendix B.1). C_{cyl} maps FLAMINGO luminosities to the cylindrical aperture used by X-GAP while allowing for plausible profile variations. X-ray fluxes are then computed including a K-correction (see Seppi et al. 2025). Hereafter we refer to cylindrical luminosities as L_{500c} .

For each halo we identify member subhaloes, project their velocities along the line of sight to obtain peculiar velocities, and compute the velocity dispersion. The light cone construction, flux computation, and velocity-dispersion estimation follow Seppi et al. (2025).

We then apply the X-GAP selection using rejection sampling: for each halo we compute the detection probability (Eq. (1)) and retain the object if this exceeds a random number drawn in $[0, 1]$. Additional X-GAP cuts are imposed: (i) $0.02 < z < 0.05$, (ii) $R_{500c} < 15$ arcmin to ensure the system fits within the *XMM-Newton* field of view (FoV), and (iii) number of galaxies $N_{\text{gal,SDSS}} \geq 8$.

The second condition translates into an upper limit on the ROSAT luminosity at fixed redshift because R_{500c} was originally estimated from RASS luminosities (Eq. (1) in Eckert et al. 2026). We therefore calibrate the relation between L_{500c} and the ROSAT luminosity in the 0.5–2.0 keV band using X-GAP groups, in the form $\log_{10} L_{500c} = A + B \log_{10} L_{\text{ROSAT}} + \mathcal{N}(0, \sigma_{\text{int}})$, finding $B = 0.70$, $A = 0.51$, and $\sigma_{\text{int}} = 0.45$. This relation is applied to convert FLAMINGO luminosities before imposing the luminosity–redshift cut.

Finally, the requirement $N_{\text{gal,SDSS}} \geq 8$ cannot be applied directly because the number of members identified by the optical FoF algorithm in observations does not necessarily perfectly correspond to the intrinsic subhalo population of the parent dark matter halo. We therefore use the simulations developed

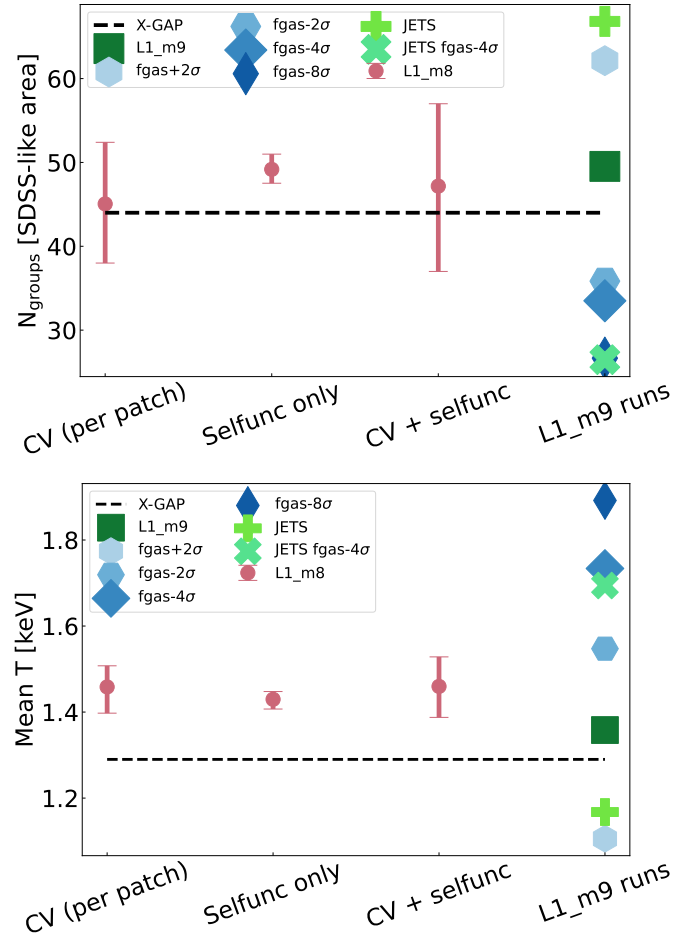


Fig. 1. Expected properties of an X-GAP-like sample selected from different FLAMINGO models. The L1_m8 includes tests for cosmic variance (CV) and uncertainties on the selection function. Top: Number of groups within an SDSS-like area of 7430 deg². Bottom: Median temperature of the selected sample. The latter is a promising discriminator between FLAMINGO models.

in Seppi et al. (2025) to calibrate the mapping between true and detected membership, enabling a forward-modelled implementation of this cut. Details are provided in Appendix A.

2.4. Selection uncertainties

We evaluate possible sources of systematics in our selected samples: cosmic variance, and uncertainties on the selection function. Their impact on the total number of selected groups and on the median temperature of the sample is shown in Fig. 1.

2.4.1. Cosmic variance

To estimate the impact of cosmic variance, we generate nine independent light cones by placing the observer at different positions within the simulation box of L1_m8. One observer is positioned at the centre of the box (500, 500, 500) Mpc, while the remaining eight are located at the centres of the eight sub-octants of the box, i.e. the midpoints of each corner cube of size 500 Mpc. This tiling strategy ensures that each light cone probes a unique volume. Given that the centres of the sub-octants are separated by 500 Mpc, the light cones originating from adjacent observers do not overlap within a radial distance of 250 Mpc,

corresponding to approximately $z = 0.05$ in the Abbott et al. (2022) cosmology, covering the redshift range relevant for the X-GAP sample without spatial redundancy. In addition, we divide each full-sky light cone into four quadrants, splitting the full RA, Dec plane in half. Then we count the number of haloes within an area sub-selected from each of the four quadrants that matches the SDSS survey area. To do this, we find a maximum value of Dec according to $\text{Dec}_{\text{max}} = \arcsin \frac{A_{\text{SDSS}}}{\Delta_{\text{RA}}}$, where $A_{\text{SDSS}} = 7430 \times (\pi/180)$ sr, and $\Delta_{\text{RA}} = \pi$ rad. We find $\text{Dec}_{\text{max}} = 46.09$ deg. We obtain a combination of four SDSS-like sky areas for each of the nine light cones, for a total of 36 independent areas to quantify cosmic variance (in combination with Poisson noise) in our selected sample.

The mean number of objects expected within an SDSS-like area is 45.1, with 16th and 84th percentiles of 38.0 and 52.4 (standard deviation 9.3), in excellent agreement with the 44 groups in X-GAP. This prediction depends on cosmology, in particular on $S_8 = \sigma_8 \sqrt{\Omega_M/0.3}$, which sets the halo abundance, and on H_0 , which affects both the survey volume and the X-ray fluxes through the luminosity distance. As shown in Appendix D.3, this changes the predicted counts by ~ 2 groups, subdominant compared to cosmic variance. The impact of cosmic variance on the temperature distribution is smaller, with mean 1.46 keV, 16th and 84th percentiles of 1.40 and 1.51 keV, and a standard deviation of 0.06 keV.

2.4.2. Uncertainty on the selection function

The best-fit parameters on the selection function model from Seppi et al. (2025) have error bars which naturally propagate as an uncertainty on the selected sample. To quantify such an effect, we randomly draw 20 models from the posterior chains, which allows us to properly account for the correlation between different parameters. We use a single area in one individual light cone to neglect cosmic variance and isolate the effect due to uncertainty in the selection function.

We obtain a mean number of 49 groups, with 16th and 84th percentiles at 47.5 and 50.9, for a standard deviation of 2.6. The temperature distribution is even less sensitive to uncertainties on the selection function: we find a mean of 1.43 keV, with 16th and 84th percentiles at 1.41 and 1.45, for a standard deviation of 0.02. We find that this effect is subdominant with respect to cosmic variance concerning both the number of groups and the temperature distribution of the selected sample.

We finally combine cosmic variance and selection function uncertainty by drawing 20 random selection function models in each sky area of the nine light cones, which gives us a total of 720 experiments to evaluate their combined effect. We obtain a total number (mean temperature) with respective 16th and 84th percentiles of 47, 37, 57 (1.46, 1.39, 1.53 keV), with a standard deviation of 10.2 (0.07 keV). In both cases the fluctuations are dominated by cosmic variance.

2.5. Hydro runs

After modelling the N_{gal} selection in L1_m8 (see Sect. 2.3), we identify corresponding haloes across the L1_m9 FLAMINGO runs, which share the same initial conditions with L1_m8, allowing consistent tracking of individual systems. We match each selected group to its counterpart within 1 Mpc, and replace the galaxy populations in lower-resolution runs with those from L1_m8, which fully resolve X-GAP members. This ensures accurate velocity dispersions and a consistent application of the

$N_{\text{gal}} \geq 8$ selection. This approach is justified as FLAMINGO simulations are calibrated to reproduce the observed GSMF, yielding consistent galaxy populations across feedback variants, while AGN feedback changes do not significantly affect galaxy properties at fixed resolution (see Fig. 9 in Schaye et al. 2023). While the L1_m8 exhibits a slight offset from L1_m9 at the high-mass end (see Fig. 8 in Schaye et al. 2023), this discrepancy is comparable to the systematic uncertainties arising from cosmic variance in the observational data. In addition, these deviations primarily affect the massive cluster regime, rather than the galaxy group scales analysed in this work. We exclude runs calibrated to different GSMFs. For each group, we compute the detection probability using Eq. (1) based on flux, redshift, and velocity dispersion (from L1_m8), and obtain X-GAP-like catalogues for all simulations.

We compare the samples selected for different model runs in FLAMINGO to the systematics tests from Sect. 2.4 in the last columns of Fig. 1. The number of selected groups depends on the feedback implementation. In an SDSS-like area, the model with strong jets provides the lowest number of selected groups with 27. The Jet model produces the most with 64, followed by the model with weak AGN $f_{\text{gas}} + 2\sigma$ at 60. The particle resolution does not affect this statistic: the number of groups in the low resolution run L1_m9 is 47, the same as in the higher resolution L1_m8. Given the uncertainty due to the combination of cosmic variance and selection function, the number of groups alone is not a great discriminator between different models, that are all encompassed within about 1.5σ compared to the fiducial L1_m8. Conversely, the temperature distribution shows smaller fluctuations, with the $f_{\text{gas}} + 2\sigma$ ($f_{\text{gas}} - 8\sigma$) at about 1.0 (1.8) keV being incompatible with the fiducial model at about 2.9 (5.8) σ . We conclude that temperature is an observable that potentially can discriminate FLAMINGO models compared to the real X-GAP sample (see also Sect. 4).

We evaluate the distribution of selected systems in the $f_{\text{gas}} - M_{500c}$ plane. The total gas mass is computed by summing the masses of all hot gas elements (see Sect. 3.1) within the true R_{500c} . The resulting relation is shown in Fig. 2, where solid lines and shaded regions indicate the selected samples, while dashed lines show the full halo populations. The top panel shows the mass distribution of the selected systems. As feedback strength increases, the selection shifts towards higher masses, since more massive haloes are required to retain enough baryons to produce X-ray luminosities that satisfy the X-GAP selection. The lower panel presents the ratio between the gas fractions of the selected and full samples. At the low-mass end the ratio exceeds unity due to incompleteness for faint systems, with deviations larger than 10% below $3 \times 10^{13} M_{\odot}$ (Seppi et al. 2025). In this regime the X-GAP selection preferentially includes systems that lie above the mean $f_{\text{gas}} - M_{500c}$ relation (i.e. up-scattered systems). Conversely, at the high-mass end the upper luminosity cut (imposed to ensure $R_{500c} < 15'$) results in a preferential selection of systems below the mean relation, with an impact below 5% at $2 \times 10^{14} M_{\odot}$. Overall, the X-GAP sample is broadly representative of the underlying group population, particularly near the typical mass scale of $(5-8) \times 10^{13} M_{\odot}$. We obtain similar results with a study of the impact of selection of the scaling relation between X-ray luminosity and temperature (see Appendix B and Fig. B.3).

Out of all the selected systems, we randomly sub-select 50 groups independently for each run, in order to have a statistical power that is similar to X-GAP. These are the systems we use for our full forward modelling. The various models show different gas fractions at fixed mass: the same selection function yields systems with different gas content as a function of mass based

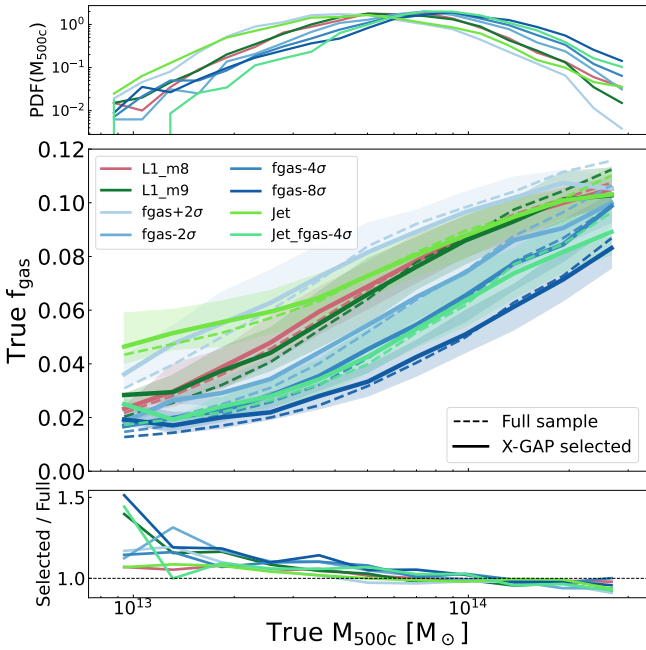


Fig. 2. Gas fraction as a function of mass for the selected systems (solid lines and shaded areas) and the full sample (dashed lines). Both are true input quantities. Top: Mass distribution of the selected systems: skewed to higher masses for strong feedback models. Bottom: Ratio between the gas fraction in the X-GAP-like selected sample and the whole population.

on the feedback model. Indeed this means that the gas content of the selected groups can discriminate between the FLAMINGO models (Kugel et al. 2023). For example, at $M_{500c} = 7 \times 10^{13} M_{\odot}$ the typical gas fraction changes by up to a factor of 2.3, with a value of 0.089 for $f_{\text{gas}} + 2\sigma$ and 0.038 for $f_{\text{gas}} - 8\sigma$. Milder values are shown by the fiducial L1_m8 with 0.076 and $f_{\text{gas}} - 2\sigma$ with 0.061. The fiducial L1_m9 is in excellent agreement with L1_m8 at 0.075. This is expected, because they are calibrated on the same target, and it confirms that particle resolution does not affect our result here.

The gas fraction has often been used in the context of AGN feedback, as stronger gas ejection would deplete haloes, therefore reducing the gas fraction at fixed halo mass (Eckert et al. 2021). Simulations, including FLAMINGO, have used the gas fraction as an anchor to calibrate their feedback model (Schaye et al. 2023). However, directly interpreting the observed $f_{\text{gas}}-M$ relation is challenging due to the inherent coupling between gas mass, total halo mass, and their integration volume. In observations, f_{gas} is integrated within a radius derived from the estimated halo mass, which is itself subject to uncertainties such as hydrostatic bias and projection effects. Consequently, errors in M_{500c} propagate into the determination of the integration radius of the gas mass, introducing non-trivial covariance in the $f_{\text{gas}}-M_{500c}$ relation. Observed trends thus reflect a combination of physical feedback effects and these measurement systematics, complicating comparisons between simulations and different observational datasets.

3. XMM-Newton forward modelling

In this section we describe the generation and analysis of XMM-Newton mock data starting from the selected galaxy groups. The complete workflow is depicted in Fig. 3.

3.1. The XMM mocks

We collect the gas particles in 3D within two times r_{200c} of each halo and isolate hot gas particles hotter than 3×10^5 K and null star formation rates. Cold gas and star forming particles are not expected to contribute to the X-ray emission of the IGrM. We also exclude gas particles that were heated by an AGN event in the last 100 Myr (comparable to a typical AGN duty cycle; Martini et al. 2004; Vantyghem et al. 2014), they are affected by numerical artefacts and unphysical temperatures and did not yet mix with the surrounding medium in a physical way. This makes the cores more realistic and typically removes less than 0.5% of the total luminosity. We assume that the hot gas is in collisional ionisation equilibrium and each hot gas particle emits following an APEC model (Smith et al. 2001). We do not introduce any assumption (such as the typical $0.3 Z_{\odot}$) on the average metallicity of the gas, we rather use the individual Z value for each particle. We use the SPH smoothed metallicity values readily available in the FLAMINGO gas particle data. We also account for the individual abundances of the following elements that are tracked in FLAMINGO: C, N, O, Ne, Mg, Si, and Fe. For non-tracked elements (S and Ca), we scale their solar contributions according to the total metallicity of particle Z/Z_{\odot} . This approach, which allows for non-solar abundance ratios, follows the methodology detailed in Braspenning et al. (2024). This follows the same recipe as the one described in Sect. 2.2–2.4 of Braspenning et al. (2024), who used cLOUDY to compute cooling rates. We verified in Appendix B that any intrinsic difference between cLOUDY and APEC does not affect our results. We estimate the X-ray spectrum of each particle using the pyXSIM software (ZuHone & Hallman 2016; Biffi et al. 2013). We generate idealised photon lists with a large effective area of $10\,000 \text{ cm}^2$ and exposure time of 2 Ms. To account for galactic absorption, we fix the neutral hydrogen column density to $N_{\text{H}} = 2 \times 10^{20} \text{ cm}^{-2}$. This value represents the typical median column density across the SDSS footprint, which is primarily located at high Galactic latitudes where N_{H} is relatively low.

To simulate realistic XMM-Newton images, we use the xmm_simulator⁴ package presented in Seppi et al. (2026). We developed a new interface to this software to generate realistic mock XMM-Newton event lists from idealised photon simulations, generated for example with pyXSIM. The interface accounts for the desired exposure time and the spatially varying effective area across the XMM field of view. As a first step, we convolve the idealised events with the instrument’s point spread function (PSF), modelled as a King profile. For each photon, we sample a radial offset from the PSF cumulative distribution function and a random azimuthal angle in $[0, 2\pi]$, which together define a blurred photon position. The three XMM detectors MOS1, MOS2, and PN are processed independently. Events falling outside each camera’s footprint are discarded. We then compute, for each pixel, the fraction of idealised photons that would actually be observed by the observational setup, given the ratios of effective areas and exposure times. This accounts for the varying instrument response across the FoV. A Poisson draw using a binomial distribution is used to select a subset of events accordingly. Finally, the redistribution matrix function (RMF) is applied: for each event, we sample a detected energy from its true energy using the appropriate RMF-derived probability distribution. The resulting event list closely mimics real XMM-Newton data products in format and instrumental characteristics. The xmm_simulator includes a realistic simulation of the X-ray foreground and background, with contributions from the local

⁴ github.com/domeckert/xmm_simulator

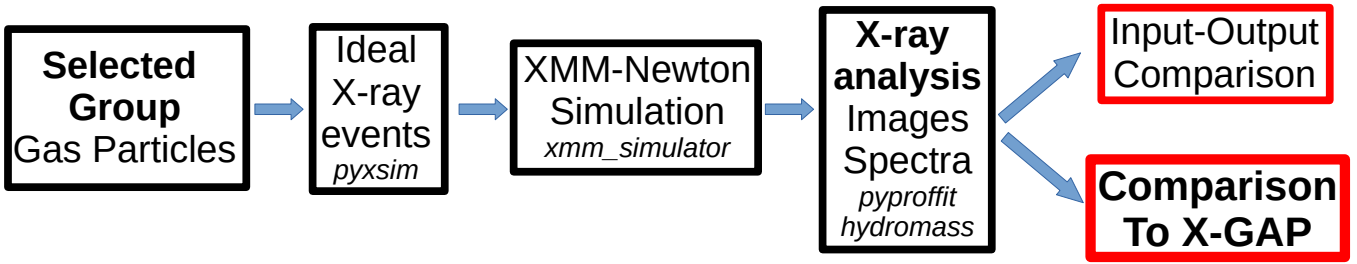


Fig. 3. Workflow for the end-to-end *XMM-Newton* simulation of FLAMINGO galaxy groups down to the direct comparison with X-GAP.

hot bubble, the galactic halo, and the cosmic X-ray background. We also simulate individual AGN in the FoV (see Seppi et al. 2026, for details). We do not specifically add X-ray binaries. They would only show up as a hard tail in the spectrum of the inner most bin in our analysis, and their contribution is expected to be around 10^{40} erg/s (Boroson et al. 2011), negligible in the soft X-rays compared to the hot gas luminosity in the regime of galaxy groups.

3.2. X-ray data analysis

We analyse the mock *XMM-Newton* images following the standard X-GAP procedures described in Eckert et al. (2025), starting from the count image, exposure map, and particle background map produced by the `xmm_simulator`. Point sources are detected with the `ewavelet` task using wavelet filtering. Compared to Seppi et al. (2026), we additionally perform a hard-band (2–7 keV) detection to identify heavily obscured AGN. PN-to-MOS count ratios of 3.42 and 2.74 are applied in the soft (0.7–1.2 keV) and hard (2–7 keV) bands, respectively. After visual inspection, the detections from both bands are combined and used to construct a mask for point sources.

Temperature profiles are derived from spectra extracted in 12 logarithmically spaced annuli centred on the X-ray peak, following Rossetti et al. (2024). The background spectrum is taken from the outermost annulus (12.5–15 arcmin) and jointly modelled with a simulated RASS background generated with the same parameters as in `xmm_simulator`. The background model includes three components: the local hot bubble ($kT = 0.11$ keV), the Galactic halo, and the cosmic X-ray background described by a power law with photon index 1.46. RASS spectra are rescaled by a cross-calibration factor of 0.12 (Rossetti et al. 2024). In each annulus the group emission is modelled as absorbed thermal plasma (`tbabs`×`APEC`), fixing the hydrogen column density to 2×10^{20} cm⁻² as assumed in the simulation, while fitting for temperature and metallicity.

3.3. X-ray modelling

We extracted surface brightness profiles from the 0.7–1.2 keV mock *XMM-Newton* EPIC images in bins of 3 arcsec using the `pyproffit` package (Eckert et al. 2020), accounting for the telescope vignetting and PSF. We use the measured 2D surface brightness and temperature profile to infer deprojected 3D thermodynamical profiles assuming that the systems are spherically symmetric with the `hydromass` package (we use the FORW model, see Eckert et al. 2022). It allows linking the pressure, temperature, and density profiles under the assumption of hydrostatic equilibrium to jointly model the dark matter halo mass pro-

file. We compute integrated quantities within the estimated R_{500c} and the core-excised ones by excluding the inner $0.15 \times R_{500c}$. The gas mass is obtained by integrating the gas number density profile weighted by the mean molecular weight and proton mass (Eckert et al. 2020; Seppi et al. 2026). Any bias in the estimation of R_{500c} would impact the integration radius to measure integrated quantities. This is expected to be negligible for X-ray luminosity and temperature, where most of the signal comes from the inner regions, but it is important for gas mass, that mostly lies in the outskirts (see Sect. 4.1.2). The integrated temperature follows the surface brightness weighting scheme from Lovisari et al. (2024), that has been shown to faithfully reproduce spectroscopic-like temperature (Mazzotta et al. 2004; Rasia et al. 2014). We also refer the reader to Eckert et al. (2025) for more details on the X-ray data analysis of X-GAP groups, which we implement in the same way on our mock groups from FLAMINGO.

4. Results

First, we verified that our X-ray analysis pipeline accurately recovers input quantities. For temperature and luminosity we exclude the central $0.15 R_{500c}$, where FLAMINGO suffers from low particle resolution and unrealistic metallicity profiles (Appendix B). The X-GAP data is treated identically. We find good agreement between input and recovered X-ray luminosities and gas masses. The `hydromass` pipeline estimates emissivities using radially dependent metallicity and temperature profiles derived from the spectral fits, enabling unbiased gas mass measurements (Appendices B and D). Temperature measurements are less robust for strong feedback models, where disturbed systems violate assumptions such as spherical symmetry or a single-temperature structure. After excluding such cases through visual inspection, the recovered temperatures remain consistent with the spectroscopic-like weighting scheme (Mazzotta et al. 2004; Seppi et al. 2026). Further details are provided in Appendix B. We now focus on the comparison between the samples selected from FLAMINGO and the real X-GAP.

4.1. Observable scaling relations

In this section we study the relations linking core-excised X-ray luminosity and gas mass within 400 kpc to core-excised temperature. The values are tabulated in Table E.1.

4.1.1. Luminosity – Temperature

For consistency with the core-excised temperature, we also consider core-excised X-ray luminosities. Similarly, we focus on the

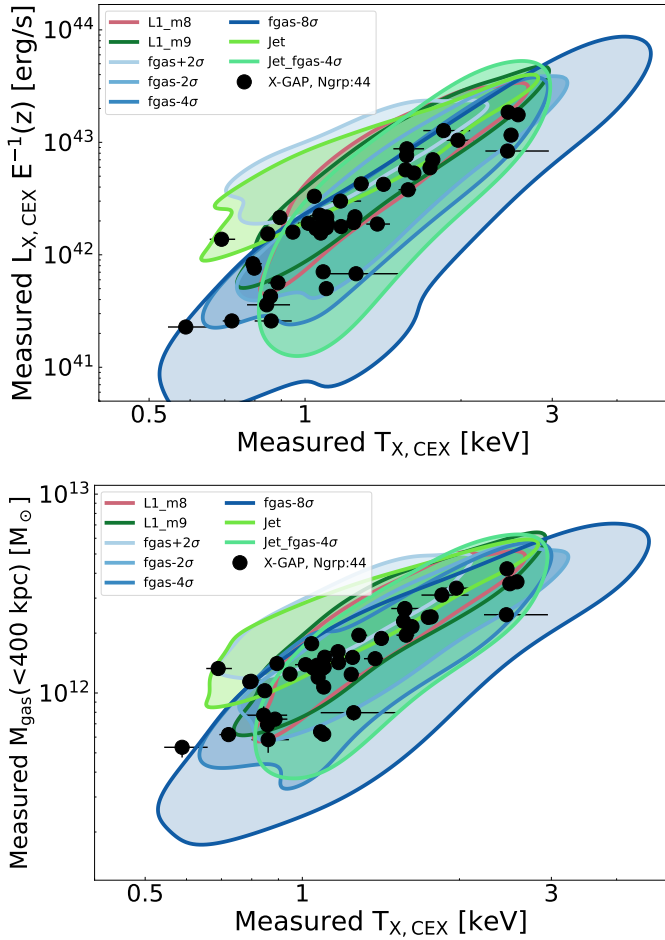


Fig. 4. Comparison between X-GAP and the FLAMINGO models using the scaling relations between observables: core-excised X-ray luminosity and core-excised temperature (top) and gas mass and core-excised temperature (bottom). The contours represent the 2σ distribution for each FLAMINGO model estimated with a Gaussian KDE.

region between 0.15 and $1.0 \times R_{500c}$. Although the inconsistency between input and recovered temperatures primarily affects the strongest feedback simulations, we retain all systems in the comparison with X-GAP. From an observational standpoint, such discrepancies would not be identifiable, as the true temperatures are not known for real data.

The relation between the core-excised X-ray luminosity and temperature obtained in the previous sections is shown in the top panel of Fig. 4. Both quantities are direct observables obtained from *XMM-Newton* mocks, so that a direct comparison between the FLAMINGO models selected with our selection function and the real X-GAP is possible. All samples occupy a similar region of this parameter space, roughly between 5×10^{41} and 3×10^{43} erg/s and 0.5 and 3.0 keV. However, increasing the strength of AGN feedback shifts groups towards the bottom-right: stronger feedback reduces the gas content of galaxy groups, lowering their X-ray luminosity. As a result, at fixed observed flux the selected systems correspond to intrinsically more massive and hotter haloes. It is important to account for the correlation between these two observables: temperature regulates the mass scale related to the groups selection whereas luminosity encodes the gas retained by these haloes.

We model the L - T relation as a power law of the form $\log L = B_{LT} + \alpha \log\left(\frac{T}{T_p}\right)$, where T_p is the pivot temperature of

1.3 keV, equivalent to the mean core-excised temperature in X-GAP. In particular, the normalisation B_{LT} reflects the gas content as traced by the X-ray luminosity at a characteristic temperature typical of the systems considered here. Hereafter we will refer to B_{LT} involving core-excised quantities. We model the relation using a Bayesian hierarchical framework that accounts for errors in both variables by treating the true temperature as a latent quantity. The fit is performed in base-10 logarithmic space. We adopt uniform priors on the slope (1 – 4) and normalisation (40 – 45), and include an intrinsic scatter with a broad half-Gaussian prior up to 0.8 dex. As all samples follow the same selection by construction, the power-law relation can be directly inferred. Posterior sampling is performed using the No-U-Turn sampler implemented in PyMC (Abril-Pla et al. 2023).

For X-GAP we obtain $B_{LT} = 42.43^{+0.03}_{-0.04}$. The closest FLAMINGO run is $f_{\text{gas}} - 2\sigma$ with $B_{LT} = 42.43^{+0.03}_{-0.02}$. The most extreme cases are $f_{\text{gas}} + 2\sigma$ with $B_{LT} = 42.91^{+0.03}_{-0.03}$ and $f_{\text{gas}} - 8\sigma$ with $B_{LT} = 42.06^{+0.05}_{-0.05}$. We obtain similar results for L1_m8 and L1_m9 with $B_{LT} = 42.63^{+0.02}_{-0.02}$ (see Table 1).

4.1.2. Gas mass – Temperature

We model the gas mass–temperature scaling relation similarly to the L - T relation described above. However, the true R_{500c} used in the input–output comparison (Sect. B) is not known observationally. We therefore measure the gas mass within a fixed aperture of 400 kpc, corresponding to the upper limit of the integral in Eq. (B.1). This radius lies close to or within R_{500c} for most selected systems, where the X-ray data provide good signal-to-noise ratio. The same quantity is measured for X-GAP using the best-fit hydromass models (e.g. Eckert et al. 2026, and Table E.1).

The result is shown in the bottom panel of Fig. 4. Similarly to the L - T relation, increasing the strength of AGN feedback shifts the points to the bottom right direction, ejecting gas and increasing the overall temperature. We fit the relation with the same method as for L - T , with priors on the slope (1 – 5) and normalisation (10 – 15). For X-GAP we obtain $B_{M_{\text{gas}}T} = 12.20^{+0.02}_{-0.02}$. The closest FLAMINGO run is $f_{\text{gas}} - 2\sigma$ with $B_{M_{\text{gas}}T} = 12.19^{+0.01}_{-0.01}$. The most extreme cases are $f_{\text{gas}} + 2\sigma$ with $B_{M_{\text{gas}}T} = 12.42^{+0.01}_{-0.01}$ and $f_{\text{gas}} - 8\sigma$ with $B_{M_{\text{gas}}T} = 12.00^{+0.02}_{-0.03}$. We find excellent agreement also for this relation between L1_m8 ($B_{M_{\text{gas}}T} = 12.27^{+0.01}_{-0.01}$) and L1_m9 ($B_{M_{\text{gas}}T} = 12.26^{+0.01}_{-0.01}$). Finally, we verified in Sect. D.1 that $B_{M_{\text{gas}}T}$ is recovered correctly independently of the gas particle metallicity, which is high in FLAMINGO (Braspenning et al. 2024).

4.2. Combining observables

As we have shown in Sect. 2.4, the number of groups and the median temperature of the samples are potential discriminators between the FLAMINGO models. Here we combine the normalisation of the L - T and $M_{\text{gas}}-T$ relations obtained in the previous section with the number of selected groups, the mean of the measured core-excised temperature, and an estimate of the mean velocity dispersion of the groups. For each observable, we construct its distribution across the simulation ensemble. We reference this process as the MonteCarlo procedure: for the L - T and $M_{\text{gas}}-T$ normalisations, B_{LT} and $B_{M_{\text{gas}}T}$, the distribution is obtained directly from the posterior samples of the PyMC fit described in the previous section. The temperature distribution is estimated via 1000 bootstrap resamples of the

Table 1. Summary statistic about the comparison between X-GAP and various FLAMINGO flavours.

Model	B_{LT}	$B_{M_{gasT}}$	Mean T [keV]	N_{grps}	Mean σ_v [km/s]	Combined significance
X-GAP	42.43 ^{+0.04} _{-0.04}	12.20 ^{+0.02} _{-0.02}	1.30 ^{+0.06} _{-0.07}	44	433 ⁺¹⁸ ₋₁₈	–
L1_m8	42.63 ^{+0.03} _{-0.03}	12.26 ^{+0.01} _{-0.01}	1.45 ^{+0.05} _{-0.06}	47 ⁺¹⁰ ₋₁₀	431 ⁺¹⁵ ₋₁₄	3.1 σ
L1_m9	42.63 ^{+0.03} _{-0.03}	12.27 ^{+0.01} _{-0.01}	1.44 ^{+0.05} _{-0.06}	47 ⁺¹⁰ ₋₁₀	432 ⁺¹⁴ ₋₁₄	3.1 σ
$f_{gas} + 2\sigma$	42.92 ^{+0.02} _{-0.02}	12.43 ^{+0.01} _{-0.01}	1.26 ^{+0.06} _{-0.05}	60 ⁺¹¹ ₋₁₁	421 ⁺¹⁶ ₋₁₄	3.6 σ
$f_{gas} - 2\sigma$	42.42 ^{+0.03} _{-0.03}	12.19 ^{+0.01} _{-0.01}	1.49 ^{+0.05} _{-0.06}	39 ⁺⁹ ₋₉	444 ⁺¹² ₋₁₃	0.8 σ
$f_{gas} - 4\sigma$	42.28 ^{+0.03} _{-0.03}	12.12 ^{+0.02} _{-0.02}	1.45 ^{+0.06} _{-0.06}	33 ⁺⁸ ₋₇	473 ⁺¹⁸ ₋₁₉	2.4 σ
$f_{gas} - 8\sigma$	42.05 ^{+0.05} _{-0.05}	12.00 ^{+0.02} _{-0.03}	1.57 ^{+0.09} _{-0.09}	26 ⁺⁷ ₋₆	475 ⁺²⁰ ₋₁₈	4.4 σ
Jet	42.83 ^{+0.02} _{-0.02}	12.38 ^{+0.01} _{-0.01}	1.24 ^{+0.06} _{-0.06}	64 ⁺¹³ ₋₁₃	410 ⁺¹⁵ ₋₁₅	3.7 σ
Jet_ $f_{gas} - 4\sigma$	42.23 ^{+0.05} _{-0.05}	12.11 ^{+0.03} _{-0.03}	1.52 ^{+0.06} _{-0.06}	27 ⁺⁷ ₋₇	478 ⁺¹³ ₋₁₄	3.3 σ

Notes. The columns report: the group sample, the normalisation of the core-excised luminosity-temperature relation (B_{LT}) and the gas mass-core-excised temperature ($B_{M_{gasT}}$), the mean temperature, the number of groups, the mean velocity dispersion, and the combined tension to X-GAP.

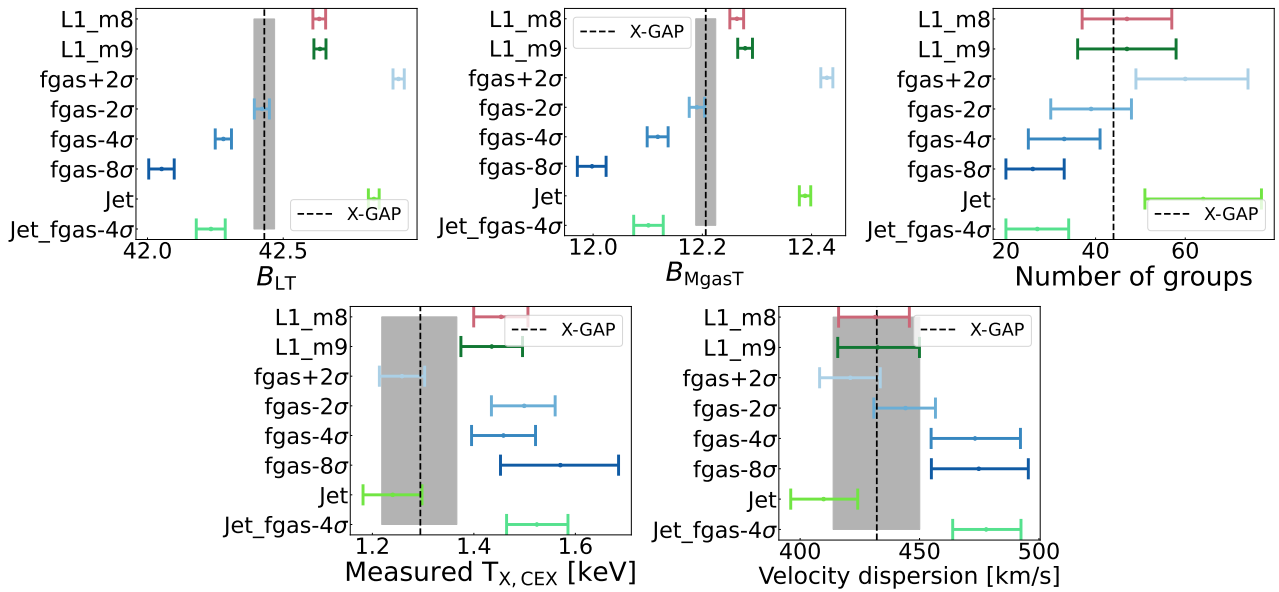


Fig. 5. Observables used for the comparison between X-GAP and various FLAMINGO models: the normalisation of the scaling relation between X-ray luminosity and temperature (both core-excised), between a gas mass within 400 kpc and core-excised temperature, the total number of groups, the mean temperature, and the galaxy member velocity dispersion. X-GAP shows the best agreement with the $f_{gas} - 2\sigma$ model.

values measured for the end-to-end simulations for each individual model. Regarding the velocity dispersion, we adopt the calibrated relation between true and measured velocity dispersion derived in Seppi et al. (2025), who performed an end-to-end simulation also in the optical band (see their Sect. 6 and Eqs. 17–18). For each object, a measured velocity dispersion is drawn from the corresponding log-normal distribution defined by that model, using the true velocity dispersion computed in Sect. 2.3. Similarly to temperature, the distribution is obtained via bootstrap sampling. These distributions include measurement uncertainties on a single mock X-GAP realisation. For the number of groups we apply to all FLAMINGO runs the procedure described in Sect. 2.4 for L1_m8. We name this process the N-light cone generation. This provides, for each model, a distribution of the selected number of groups that self-consistently incorporates cosmic variance, Poisson noise, and uncertainties in the selection function.

The results are collected in Fig. 5. B_{LT} (top-left panel) shows a preference for the $f_{gas} - 2\sigma$ model, the only one that is compatible within 1σ with X-GAP at 0.3σ . The deviations are largest for the $f_{gas} + 2\sigma$, Jet, and $f_{gas} - 8\sigma$, with tensions at the 11.6,

9.6, and 6.3σ levels. $B_{M_{gasT}}$ (top-central panel) shows a preference for the $f_{gas} - 2\sigma$ model, again the only one that is compatible within 1σ with X-GAP at 0.7σ . The largest deviations are in the $f_{gas} + 2\sigma$, Jet and $f_{gas} - 8\sigma$, with tensions at the 10.6, 8.8, and 6.7σ levels. The distributions of the number of groups (top-left panel) are broad due to cosmic variance, therefore most models agree within 2σ with X-GAP. The vertical dashed line marks the realised X-GAP catalogue size of 44 systems. Unlike the model distributions, which include Poisson fluctuations, cosmic variance, and selection-function uncertainties, this value is not assigned an additional error bar, since it corresponds to the fixed observed sample within the survey footprint. $f_{gas} - 8\sigma$ and Jet_ $f_{gas} - 4\sigma$ fall above this threshold by 2.7 and 2.4σ respectively. The mean temperature of the sample (bottom-left panel) broadly shows a good agreement between models and data, with $f_{gas} + 2\sigma$ and Jet being the only ones under-predicting the mean X-GAP temperature, but agreeing with the data within 0.4 and 0.6σ . Instead $f_{gas} - 2\sigma$, $f_{gas} - 8\sigma$, Jet_ $f_{gas} - 4\sigma$ over-predicting it by 2.1 , 2.2 , and 2.4σ . We note that the measured temperature in Jet_ $f_{gas} - 4\sigma$ and $f_{gas} - 8\sigma$ tends to be underestimated as reported in Fig. B.5, which reduces the constraining power

of temperature on its own. In addition, the underlying temperature distribution also depends on the slope of the temperature mass scaling relation in the group regime, which is below the mass range where the FLAMINGO models were calibrated (see [Braspenning et al. 2024](#), and the discussion in Sect. 6). $\text{Jet-}f_{\text{gas}} - 4\sigma$ and $f_{\text{gas}} - 8\sigma$ show the largest tension for the mean velocity dispersion (bottom-right panel), in tension with the X-GAP measurement by 2.4 and 2.7σ . Finally, all the observables presented in Fig. 5 are in excellent agreement between the high resolution L1_m8 and L1_m9. This is also a validation for the flagship FLAMINGO simulation L2p8_m9 that uses the same resolution, but covers a larger box of 2.8 Gpc.

These observables are not statistically independent: for example, hotter systems tend to have higher velocity dispersions, potentially affecting the scaling relation as well as the total number of groups. Therefore, combining the individual p-values would overestimate the overall tension. Instead, we explicitly account for their covariance. In Appendix C we provide full details on our method. The final σ_{eq} quantifies the statistical significance of the difference between X-GAP and each FLAMINGO model, accounting for the covariance between the observables.

The final result is collected in Table 1 and shown in Fig. C.3. We conclude that $f_{\text{gas}} - 2\sigma$ provides the best comparison to X-GAP, with a global tension of 0.8σ . Given our framework, this means that there is a 52% probability of observing X-GAP if $f_{\text{gas}} - 2\sigma$ is the right model. We find a worse statistical agreement for all other models. $f_{\text{gas}} - 4\sigma$ is already too ejective, with lower scaling relation normalisations (Fig. 5) and at 2.4σ global tension. All other models are ruled out with more than 3σ confidence, with $f_{\text{gas}} - 8\sigma$ showing the largest discrepancy at 4.4σ . In particular, our result is in tension with recent kSZ measurements that prefer $f_{\text{gas}} - 8\sigma$ ([Bigwood et al. 2025](#); [Siegel et al. 2025a](#); [McCarthy et al. 2025](#); [Kovač et al. 2025](#)). However, kSZ studies do not necessarily match the scales probed by our galaxy groups (see also discussion in Sect. 5.2).

5. Discussion

The impact of baryons on dark matter and the LSS is an open question in modern astronomy. In this work we investigated this problem through the impact of AGN feedback on galaxy groups. Here we further discuss our results.

5.1. Scaling relations and selection

In Sect. 4.2 and Fig. 5 we found that FLAMINGO generally tends to overestimate the X-GAP temperatures. The weak-feedback models $f_{\text{gas}} + 2\sigma$ and Jet are the only ones consistent with the data within 1σ , while models such as L1_m8 and $f_{\text{gas}} - 2\sigma$ show discrepancies at the $\sim 2\sigma$ level. This behaviour is likely driven by the interplay between the underlying L_X-M and $T-M$ scaling relations and the X-ray flux-limited selection function ([Seppi et al. 2025](#)). Because the selection effectively operates in the L_X-M plane, even small changes in the slope of this relation modify the mass distribution of the selected systems, which in turn shifts the mean temperature of the sample.

For instance, [Braspenning et al. \(2024\)](#) show (their Fig. 2) that the $L-M$ relation is slightly flatter in the Jet model compared to the fiducial case. A flatter slope increases the relative luminosity of lower-mass groups (see also the larger gas fractions for the Jet model in Fig. 10 of [Schaye et al. 2023](#)), effectively lowering the mass scale probed by a flux-limited survey. Similarly, their Fig. 3 shows a modest steepening of the $T-M$ relation towards

the low-mass regime, with most models lying below the fiducial model. The combined effect of these slope variations is therefore non-trivial: shifts in the mass distribution of the selected haloes propagate into shifts in the temperature distribution, even when the intrinsic scaling relations differ only mildly.

Importantly, FLAMINGO was not calibrated in the low-mass group regime, where observational constraints remain sparse. Small differences in the slopes of scaling relations can therefore be amplified by the flux-limited selection, producing systematic shifts in the mean temperature of an X-GAP-like sample.

5.2. Comparison to other works

Our results are broadly consistent with [Eckert et al. \(2026\)](#), who compared FLAMINGO and X-GAP with population models and excluded the strongest feedback models. In particular $f_{\text{gas}} - 8\sigma$ showed a 5.7σ tension, and the best agreement was provided by the fiducial feedback model. In this work we refined the analysis with a full end-to-end forward modelling of the observations and accounting for the covariance between observables. We find evidence for stronger feedback than the fiducial model, with $f_{\text{gas}} - 2\sigma$ showing the lowest tension to X-GAP, but still exclude the strongest model by more than 4σ .

Both the results presented in this paper and the ones from [Eckert et al. \(2026\)](#) are in tension with kSZ measurements combined with X-ray cluster gas masses presented in [McCarthy et al. \(2025\)](#), [Kovač et al. \(2025\)](#), [Bigwood et al. \(2025\)](#), [Siegel et al. \(2025a\)](#), which show a preference for the strongest FLAMINGO feedback schemes. A direct comparison with stacked kSZ measurements is, however, non-trivial. X-GAP is located in the local Universe at redshifts less than 0.05, while measurements of the stacked kSZ are sensitive to a much wider mass and redshift range. In addition, the X-ray cluster gas mass fractions may be subject to systematics relative to mass calibration and therefore aperture-related uncertainties on the integrated quantities. These differences suggest that the apparent tension may reflect a combination of redshift evolution, selection effects, and measurement systematics rather than a single discrepancy in feedback strength.

Altogether, these results show that we currently lack a complete understanding of AGN triggering throughout cosmic history. For example, [Costello et al. \(2025\)](#) used the FLAMINGO simulations and found an anti-correlation between BH mass and gas fraction in the galaxy regime, however this reverses and becomes positive for galaxy groups: more massive haloes form early, host more massive BHs, eject quickly and reaccrete later, increasing the gas fractions. Conversely, [Marini et al. \(2025a\)](#) do not find evidence for such a trend in the Magneticum simulation. This hints to the possibility that the exact feedback evolution history remains unclear.

A variation of the AGN feedback efficiency across cosmic time, for example due to environmental effects, may change group properties as a function of redshift. In addition, the exact interplay between AGN feedback and the surrounding medium, both in thermal, kinetic, and jet modes, has an impact on the duty cycle ([Sotira et al. 2025, 2026](#)). The shape of the $f_{\text{gas}}-M$ relation at the group regime likely explains the differences between the fiducial models and the Jet run. Although both models are calibrated to match the same gas fraction at cluster scales, the extrapolation towards the group regime probed in this work predicts higher gas fractions for the Jet model than for the fiducial case, while they converge at higher masses (Fig. 10 in [Schaye et al. 2023](#)). This highlights that different prescriptions for AGN feedback as a function of halo mass can have distinct

effects depending on the halo population under consideration (groups, clusters, or galaxies) which may also contribute to tensions with studies based on the kSZ effect.

5.3. Impact on cosmology

On top of all observables used in this work, together with kSZ and gas fractions, another useful feedback probe is the cross correlation between X-ray photons and cosmic shear, that can potentially discriminate between FLAMINGO models (McDonald et al. 2026). Disentangling all these effects in comparison to a local sample such as X-GAP requires deep, high-resolution observations of galaxy groups, a particularly relevant outlook in sight of NewAthena (Cruise et al. 2025).

Finally, the relevance of group-scale feedback extends beyond halo astrophysics. The distance up to which one would find all the baryons ejected by feedback, effectively recovering the universal baryon fraction, depends on feedback intensity and efficiency. The closure radius, i.e. the radius at which the enclosed baryon fraction reaches the cosmic mean, was shown to vary from one to several R_{200c} for group-scale systems Ayromlou et al. (2023), which implies a substantial decoupling between baryons and dark matter. This can be probed by large scale X-ray observations of galaxy groups at low redshifts. The redistribution of baryons modifies the matter distribution and suppresses small-scale power in the matter power spectrum (van Daalen et al. 2011; Schneider et al. 2019; van Daalen et al. 2020; Gebhardt et al. 2026). The suppression is mainly driven by the shape of the gas density profiles of galaxy groups, such that deep observations of local groups are highly complementary to LSS cosmological probes (see Debackere et al. 2020; Mead et al. 2021; Grandis et al. 2024; Schneider et al. 2025; Schaller et al. 2025). This has relevant consequences on cosmological experiments on the LSS (Semboloni et al. 2011; Aricò et al. 2023), meaning that a deeper understanding of feedback paves the way towards a more accurate and robust knowledge of cosmology (see e.g. DES Collaboration 2026).

6. Conclusion

In this section we provide a summary of this work and the future outlook.

6.1. Summary

In this work we investigated the impact of AGN feedback on the hot atmospheres of galaxy groups by comparing simulations and observations. We used the FLAMINGO suite, which includes several AGN feedback prescriptions, and compared them with the X-GAP group sample. Selection effects were taken into account using the detection probability derived in Seppi et al. (2025), which forward modelled the RASS detection process. We found cosmic variance to be a large source of systematic uncertainty for what concerns the number of selected groups, with relative variations larger than 20% (see Fig. 1 and also Appendix B). Therefore, the number of groups is not a driver in the comparison between data and models in this case, due to the relatively small volume (about 0.07 Gpc^3) probed by X-GAP.

We further applied the X-GAP selection criteria of at least eight galaxy members and $R_{500c} < 15'$ to construct simulated X-GAP analogues. We then generated end-to-end *XMM-Newton* mock observations, including instrumental effects, and analysed them with the same imaging and spectral procedures used for the real data (Fig. 3).

We then analysed the X-ray simulation with the same method as for the real X-GAP sample (Eckert et al. 2025). This framework enabled a direct comparison between simulations and observations in observable space, where temperature traces the halo mass scale and X-ray luminosity probes the gas content (Fig. 4).

We used the $L-T$, $M_{\text{gas}}-T$ normalisations, mean temperature, mean velocity dispersion, and the number of detected groups as discriminating observables, accounting for their covariance with a Gaussian copula approach (Fig. C.3). The $f_{\text{gas}} - 2\sigma$ model provides the best agreement with X-GAP, deviating by only 0.8σ , while the strongest feedback model ($f_{\text{gas}} - 8\sigma$) is excluded at 4.4σ . The $f_{\text{gas}} - 2\sigma$ model provides an X-GAP-like population with a typical gas fraction of 6.1%, while the $f_{\text{gas}} - 8\sigma$ yields emptier groups with a typical gas fraction of 3.8% (see Sect. 2.5). Overall, our end-to-end framework provides a direct constraint on the fate of hot gas in galaxy groups, favouring a feedback scheme that is stronger than the fiducial FLAMINGO calibration but weaker than the most ejective scenarios.

6.2. Outlook

In future work, we plan to further expand the comparison to more simulations, to enrich the comparison including jets in this article. In this context, the coupling between the IGrM and AGN feedback is a function of the distance to the central SMBH: different recipes in hydrodynamical simulations may cause variations of the efficiency to which AGN feedback impacts the hot medium. This information is encoded in the shape of the entropy profiles, which provides a promising avenue for future tests.

Simulations at higher resolution such as COLIBRE (Schaye et al. 2026), in addition to well-resolved AGN jets (Bourne & Yang 2023) will help shed light on this topic. For example, Correa Magnus et al. (2026) show that surface brightness concentration can be another discriminator between FLAMINGO models. This is directly related to the slope of the profile β . However, at the current resolution level, the inner core of X-GAP-like groups is resolved with only a few tens of gas elements, and core measurements are not fully reliable from the observer perspective. This could be achieved with more massive clusters in FLAMINGO where the inner $0.1 \times R_{500c}$ is resolved with hundreds of gas elements, but the AGN impact on the hot ICM is milder. Finally, our framework will allow us to correct group mass measurements in the real X-GAP and therefore also their relation to the gas fraction, accounting for their covariance, measurement uncertainties and systematics, and selection effects.

Acknowledgements. RS and DE are supported by Swiss National Science Foundation project grant #200021_212576. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. M.A.B. is supported by a UKRI Stephen Hawking Fellowship (EP/X04257X/1). LL acknowledges support from INAF grant 1.05.24.05.15. The authors thank the anonymous referee for their helpful comments on this article

References

- Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, *Phys. Rev. D*, **105**, 023520
- Abril-Pla, O., Andreani, V., Carroll, C., et al. 2023, *PeerJ Comput. Sci.*, **9**
- Akino, D., Eckert, D., Okabe, N., et al. 2022, *PASJ*, **74**, 175
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, **219**, 12
- Aricò, G., Angulo, R. E., Zennaro, M., et al. 2023, *A&A*, **678**, A109

- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
- Ayromlou, M., Nelson, D., & Pillepich, A. 2023, *MNRAS*, 524, 5391
- Bahar, Y. E., Bulbul, E., Ghirardini, V., et al. 2024, *A&A*, 691, A188
- Biffi, V., Dolag, K., & Böhringer, H. 2013, *MNRAS*, 428, 1395
- Bigwood, L., Yamamoto, M., Siegel, J., et al. 2025, arXiv e-prints [arXiv:2510.15822]
- Blanton, M. R., & Roweis, S. 2007, *AJ*, 133, 734
- Bocquet, S., Grandis, S., Krause, E., et al. 2025, *Phys. Rev. D*, 111, 063533
- Booth, C. M., & Schaye, J. 2009, *MNRAS*, 398, 53
- Boroson, B., Kim, D.-W., & Fabbiano, G. 2011, *ApJ*, 729, 12
- Bourne, M. A., & Yang, H.-Y. K. 2023, *Galaxies*, 11, 73
- Braspenning, J., Schaye, J., Schaller, M., et al. 2024, *MNRAS*, 533, 2656
- Brighenti, F., & Mathews, W. G. 2006, *ApJ*, 643, 120
- Correa Magnus, L., Kay, S. T., Schaye, J., & Schaller, M. 2026, *MNRAS*, 545, staf2073
- Costello, E. E., McCarthy, I. G., Salcido, J., et al. 2025, arXiv e-prints [arXiv:2510.17980]
- Cruise, M., Guainazzi, M., Aird, J., et al. 2025, *Nat. Astron.*, 9, 36
- Damsted, S., Finoguenov, A., Lietzen, H., et al. 2024, *A&A*, 690, A52
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, *MNRAS*, 486, 2827
- Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2020, *MNRAS*, 492, 2285
- DES Collaboration (Abbott, T. M. C., et al.) 2026, arXiv e-prints [arXiv:2601.14559]
- Dubois, Y., Peirani, S., Pichon, C., et al. 2016, *MNRAS*, 463, 3948
- Eckert, D., Molendi, S., & Paltani, S. 2011, *A&A*, 526, A79
- Eckert, D., Finoguenov, A., Ghirardini, V., et al. 2020, *Open J. Astrophys.*, 3, 12
- Eckert, D., Gaspari, M., Gastaldello, F., Le Brun, A. M. C., & O'Sullivan, E. 2021, *Universe*, 7, 142
- Eckert, D., Etori, S., Pointecouteau, E., van der Burg, R. F. J., & Loubser, S. I. 2022, *A&A*, 662, A123
- Eckert, D., Gastaldello, F., O'Sullivan, E., et al. 2024, arXiv e-prints [arXiv:2403.17145]
- Eckert, D., Gastaldello, F., Lovisari, L., et al. 2025, *A&A*, 701, A127
- Eckert, D., Seppi, R., Braspenning, J., et al. 2026, *A&A*, 709, L4
- Eckmiller, H. J., Hudson, D. S., & Reiprich, T. H. 2011, *A&A*, 535, A105
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, 142, 72
- Etori, S., Pratt, G. W., de Plaa, J., et al. 2013, arXiv e-prints [arXiv:1306.2322]
- Fabian, A. C. 2012, *ARA&A*, 50, 455
- Finoguenov, A., Jones, C., Böhringer, H., & Ponman, T. J. 2002, *ApJ*, 578, 74
- Forouhar Moreno, V. J., Helly, J., McGibbon, R., et al. 2025, *MNRAS*, 543, 1339
- Fukugita, M., & Peebles, P. J. E. 2004, *ApJ*, 616, 643
- Fukugita, M., Hogan, C. J., & Peebles, P. J. E. 1998, *ApJ*, 503, 518
- Gaspari, M., Tombesi, F., & Cappi, M. 2020, *Nat. Astron.*, 4, 10
- Gastaldello, F., Buote, D. A., Humphrey, P. J., et al. 2007, *ApJ*, 669, 158
- Gastaldello, F., Simionescu, A., Mernier, F., et al. 2021, *Universe*, 7, 208
- Gebhardt, M., Anglés-Alcázar, D., Genel, S., et al. 2026, arXiv e-prints [arXiv:2601.06258]
- Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, *A&A*, 689, A298
- Giodini, S., Pierini, D., Finoguenov, A., et al. 2009, *ApJ*, 703, 982
- Gonzalez, A. H., Sivanandam, S., Zabludoff, A. I., & Zaritsky, D. 2013, *ApJ*, 778, 14
- Gozaliasl, G., Finoguenov, A., Tanaka, M., et al. 2019, *MNRAS*, 483, 3545
- Grandis, S., Aricò, G., Schneider, A., & Linke, L. 2024, *MNRAS*, 528, 4379
- Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, *MNRAS*, 503, 1940
- Henden, N. A., Puchwein, E., Shen, S., & Sijacki, D. 2018, *MNRAS*, 479, 5385
- Hlavacek-Larrondo, J., Li, Y., & Churazov, E. 2022, in *Handbook of X-ray and Gamma-ray Astrophysics*, eds. C. Bambi, & A. Sanganello, 5
- Huško, F., Lacey, C. G., Schaye, J., Schaller, M., & Nobels, F. S. J. 2022, *MNRAS*, 516, 3750
- Huško, F., Lacey, C. G., Schaye, J., et al. 2026, *MNRAS*, 547, stag324
- Johnson, R., Ponman, T. J., & Finoguenov, A. 2009, *MNRAS*, 395, 1287
- Käfer, F., Finoguenov, A., Eckert, D., et al. 2019, *A&A*, 628, A43
- Khalil, H., Finoguenov, A., Tempel, E., & Mamon, G. A. 2024, *A&A*, 690, A212
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, 51, 511
- Kovač, M., Nicola, A., Bucko, J., et al. 2025, *JCAP*, 2025, 046
- Kügel, R., Schaye, J., Schaller, M., et al. 2023, *MNRAS*, 526, 6103
- Laganá, T. F., Martinet, N., Durret, F., et al. 2013, *A&A*, 555, A66
- Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, *MNRAS*, 441, 1270
- Lovisari, L., Reiprich, T. H., & Schellenberger, G. 2015, *A&A*, 573, A118
- Lovisari, L., Etori, S., Gaspari, M., & Giles, P. A. 2021, *Universe*, 7, 139
- Lovisari, L., Etori, S., Rasia, E., et al. 2024, *A&A*, 682, A45
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, 115, 2285
- Marini, I., Popesso, P., Dolag, K., et al. 2025a, *A&A*, 698, A191
- Marini, I., Popesso, P., Dolag, K., et al. 2025b, *A&A*, 694, A207
- Martini, P. 2004, in *Coevolution of Black Holes and Galaxies*, ed. L. C. Ho, 169
- Mazzotta, P., Rasia, E., Moscardini, L., & Tormen, G. 2004, *MNRAS*, 354, 10
- McCarthy, I. G., Schaye, J., Ponman, T. J., et al. 2010, *MNRAS*, 406, 822
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, *MNRAS*, 465, 2936
- McCarthy, I. G., Amon, A., Schaye, J., et al. 2025, *MNRAS*, 540, 143
- McDonald, W., Schaye, J., Kuijken, K., et al. 2026, arXiv e-prints [arXiv:2602.02484]
- McGibbon, R., Helly, J., Schaye, J., Schaller, M., & Vandenbroucke, B. 2025, *J. Open Source Software*, 10, 8252
- McNamara, B. R., & Nulsen, P. E. J. 2007, *ARA&A*, 45, 117
- Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, *MNRAS*, 502, 1401
- Molendi, S., Balboni, M., Bartalucci, I., et al. 2025, *A&A*, 699, A112
- Mulchaey, J. S. 2000, *ARA&A*, 38, 289
- Old, L., Skibba, R. A., Pearce, F. R., et al. 2014, *MNRAS*, 441, 1513
- Oppenheimer, B. D., Babul, A., Bahé, Y., Butsky, I. S., & McCarthy, I. G. 2021, *Universe*, 7, 209
- Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, *A&ARv*, 25, 2
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077
- Planck Collaboration VI. 2020, *A&A*, 641, A6
- Ploekinger, S., & Schaye, J. 2020, *MNRAS*, 497, 4857
- Ponman, T. J., Cannon, D. B., & Navarro, J. F. 1999, *Nature*, 397, 135
- Popesso, P., Marini, I., Dolag, K., et al. 2025, *A&A*, 704, A277
- Pratt, G. W., Croston, J. H., Arnaud, M., & Böhringer, H. 2009, *A&A*, 498, 361
- Rasia, E., Lau, E. T., Borgani, S., et al. 2014, *ApJ*, 791, 96
- Rasmussen, J., & Ponman, T. J. 2009, *MNRAS*, 399, 239
- Regamey, M., Eckert, D., Seppi, R., et al. 2026, *A&A*, 708, A260
- Robotham, A. S. G., Norberg, P., Driver, S. P., et al. 2011, *MNRAS*, 416, 2640
- Roper, F. A., Cai, Y. C., & Peacock, J. A. 2025, arXiv e-prints [arXiv:2510.12553]
- Rossetti, M., Eckert, D., Gastaldello, F., et al. 2024, *A&A*, 686, A68
- Sahu, N., Graham, A. W., & Davis, B. L. 2019, *ApJ*, 876, 155
- Sanders, J. S. 2023, arXiv e-prints [arXiv:2301.12791]
- Sanderson, A. J. R., O'Sullivan, E., Ponman, T. J., et al. 2013, *MNRAS*, 429, 3288
- Santra, R., Kale, R., Kolokythas, K., et al. 2026, *A&A*, 710, A1
- Schaller, M., Schaye, J., Kügel, R., Broxterman, J. C., & van Daalen, M. P. 2025, *MNRAS*, 539, 1337
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521
- Schaye, J., Kügel, R., Schaller, M., et al. 2023, *MNRAS*, 526, 4978
- Schaye, J., Chaikin, E., Schaller, M., et al. 2026, *MNRAS*, 548, stag375
- Schneider, A., Teyssier, R., Stadel, J., et al. 2019, *JCAP*, 2019, 020
- Schneider, A., Kovač, M., Bucko, J., et al. 2025, *JCAP*, 2025, 043
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, *MNRAS*, 417, 2020
- Seppi, R., Comparat, J., Nandra, K., et al. 2021, *A&A*, 652, A155
- Seppi, R., Eckert, D., Finoguenov, A., et al. 2025, *A&A*, 699, A206
- Seppi, R., Eckert, D., Rasia, E., et al. 2026, *A&A*, submitted, [arXiv:2603.03440]
- Siegel, J., Amon, A., McCarthy, I. G., et al. 2025a, arXiv e-prints [arXiv:2509.10455]
- Siegel, J., Bigwood, L., Amon, A., et al. 2025b, arXiv e-prints [arXiv:2512.02954]
- Silk, R., & Rees, M. J. 1998, *A&A*, 331, L1
- Smith, R. K., Brickhouse, N. S., Liedahl, D. A., & Raymond, J. C. 2001, *ApJ*, 556, L91
- Sotira, S., Vazza, F., & Brighenti, F. 2025, *A&A*, 697, A138
- Sotira, S., Bourne, M. A., Sijacki, D., Vazza, F., & Brighenti, F. 2026, *A&A*, submitted [arXiv:2601.14391]
- Steinborn, L. K., Dolag, K., Hirschmann, M., Prieto, M. A., & Remus, R.-S. 2015, *MNRAS*, 448, 1504
- Stott, J. P., Hickox, R. C., Edge, A. C., et al. 2012, *MNRAS*, 422, 2213
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, 124, 1810
- Sun, M., Voit, G. M., Donahue, M., et al. 2009, *ApJ*, 693, 1142
- Tempel, E., Tuvikene, T., Kipper, R., & Libeskind, N. I. 2017, *A&A*, 602, A100
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, 688, 709
- van Daalen, M. P., Schaye, J., Booth, C. M., & Dalla Vecchia, C. 2011, *MNRAS*, 415, 3649
- van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, *MNRAS*, 491, 2424
- Vantyghem, A. N., McNamara, B. R., Russell, H. R., et al. 2014, *MNRAS*, 442, 3192
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Wechsler, R. H., & Tinker, J. L. 2018, *ARA&A*, 56, 435
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, *MNRAS*, 465, 3291
- XRISM Collaboration (Audard, M., et al.) 2025, arXiv e-prints [arXiv:2510.06322]
- Xu, J., Eifler, T., Krause, E., et al. 2025, arXiv e-prints [arXiv:2510.25596]
- ZuHone, J. A., & Hallman, E. J. 2016, pyXSIM: Synthetic X-ray observations generator, Astrophysics Source Code Library [record ascl:1608.002]

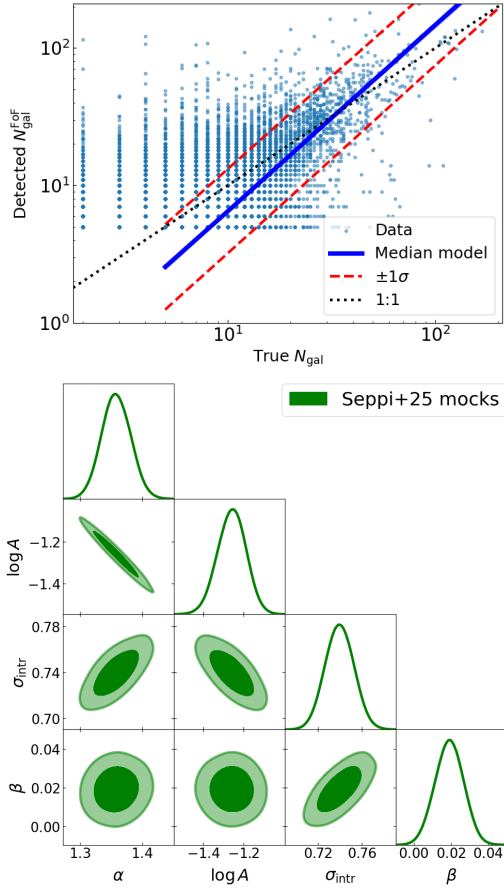


Fig. A.1. Relation between the number of true galaxy members and the ones identified by the optical finding process. Top: Data from the simulations developed by Seppi et al. (2025) and the best-fit model described in Eq. A.2. Bottom: Marginalised 1D posterior distribution of the best-fit parameters.

Appendix A: Modelling the N_{gal} selection

X-GAP is selected applying a lower cut equal to eight group members identified in SDSS. In this appendix we elaborate on the modelling of this cut, an ingredient necessary to complete the sample selection (Sect. 2.3). To reproduce the $N_{\text{gal,SDSS}}$ selection we retrieve the r -band luminosity and convert it to apparent magnitude according to Eq. (A.1):

$$\begin{aligned} M_r &= -2.5 \times \log_{10} \frac{L_r}{L_{r0}} \\ K_r &= -2.5 \times \log_{10} \left(\frac{1}{1+z} \right) \\ m_r &= M_r + 5 \log_{10} [D_L / \text{pc}] - 5 + K_r, \end{aligned} \quad (\text{A.1})$$

where L_r is the optical r -band luminosity, and L_{r0} is the reference luminosity corresponding to 4.64 Mag, and D_L is the luminosity distance in units of parsec. Given the low redshift of our sample, the K-correction is small and we assume a flat galaxy SED in Eq. (A.1) (see Blanton & Roweis 2007, Sect. 4 and Eq. 15). For each halo we count members with $m_r < 17.77$. Although L1_m8 misses galaxies below $M_{\star} \sim 10^8 M_{\odot}$, these are mostly isolated, leading to a mild deficit in the total galaxy number density at $z < 0.03$ compared to the UchuuSDSS mocks (Seppi et al. 2025). However, the number density of group members is consistent and requires no correction. The simulated luminosities correspond to integrated stellar emission and therefore approximate,

Table A.1. Priors and posteriors of the relation between the number of true galaxy members and the ones recovered by the FoF.

Parameter	Prior	Posterior
α	$\mathcal{U}(0, 2)$	1.36 ± 0.03
$\log A$	$\mathcal{U}(-5, 5)$	-1.26 ± 0.07
σ_{intr}	$\mathcal{U}(0.01, 5.0)$	0.74 ± 0.01
β	$\mathcal{U}(-5, 5)$	0.02 ± 0.01

Notes. The symbol $\mathcal{U}(M, N)$ denotes a uniform prior between the values M and N . Posterior values are reported from the third column onwards, for each case labelled in the top row.

but do not necessarily exactly reproduce, SDSS Petrosian magnitudes. At the low redshifts considered here ($z < 0.05$), such differences are expected to be subdominant relative to the empirical calibration between intrinsic and FoF-detected group membership described below. We then account for the mismatch between true members and those identified by the FoF finder using the end-to-end mocks of Seppi et al. (2025). These are used to calibrate the relation between the true number of members N_{gal} and the detected number $N_{\text{g,FoF}}$. Since only systems with $N_{\text{g,FoF}} \geq 5$ are retained, we model this relation with a truncated likelihood, i.e. the probability of observing $N_{\text{g,FoF}}$ given N_{gal} and the selection cut. The model depends on four parameters: A , α , σ_{intr} , and β :

$$\log \mu = \log A + \alpha \log N_{\text{gal}},$$

$$\sigma_{\text{ToT}} = \sigma_0 \cdot N_{\text{gal}}^{-\beta},$$

$$P(\log N_{\text{g,FoF}} | \log \mu) = \mathcal{N}(\log N_{\text{g,FoF}} | \log \mu, \sigma_{\text{ToT}}),$$

$$\begin{aligned} P(N_{\text{g,FoF}} \geq 5 | \log \mu) &= \int_{\log 5}^{\infty} \mathcal{N}(\log y' | \log \mu, \sigma_{\text{ToT}}) d \log y' = \\ &= 1 - \Phi \left(\frac{\log 5 - \log \mu}{\sigma_{\text{tot}}} \right), \end{aligned}$$

$$\mathcal{L}_i = P(\log N_{\text{g,FoF}}^{(i)} | \log \mu, N_{\text{g,FoF}}^{(i)} \geq 5) = \frac{P(\log N_{\text{g,FoF}}^{(i)} | \log \mu^{(i)})}{P(N_{\text{g,FoF}}^{(i)} \geq 5 | \log \mu^{(i)})},$$

$$\begin{aligned} \log \mathcal{L}_i &= \log [P(\log N_{\text{g,FoF}}^{(i)} | \log \mu^{(i)})] - \log [P(N_{\text{g,FoF}}^{(i)} \geq 5 | \log \mu^{(i)})], \\ \log \mathcal{L} &= \sum_{i=1}^{N_{\text{groups}}} \log \mathcal{L}_i, \end{aligned} \quad (\text{A.2})$$

where μ is the mean expected $N_{\text{g,FoF}}$ at fixed N_{gal} , \mathcal{N} is a normal distribution, Φ is the cumulative normal distribution function and we assumed that $P(\log N_{\text{g,FoF}}^{(i)}, N_{\text{g,FoF}}^{(i)} \geq 5 | \log \mu^{(i)}) = P(\log N_{\text{g,FoF}}^{(i)} | \log \mu^{(i)})$, because since the group is in the catalogue we know that it has more than 5 detected members. The relation between the true number of galaxy members, N_{gal} , and the FoF-detected number, $N_{\text{g,FoF}}$, is modelled probabilistically, accounting for intrinsic scatter and selection effects. We assume a log-normal distribution for $N_{\text{g,FoF}}$ conditioned on N_{gal} , with a scatter in log-space that varies with richness, reflecting larger relative fluctuations in poorer systems. We account for the detection threshold $N_{\text{g,FoF}} \geq 5$ using a truncated log-normal likelihood, including the survival probability $P(N_{\text{g,FoF}}^{(i)} \geq 5 | N_{\text{gal}}^{(i)})$. The total log-likelihood is summed over all groups. This framework naturally incorporates both the selection cut and measurement scatter.

We apply the correction between number of true members and optical members using the best fit model in Eq. A.2 using

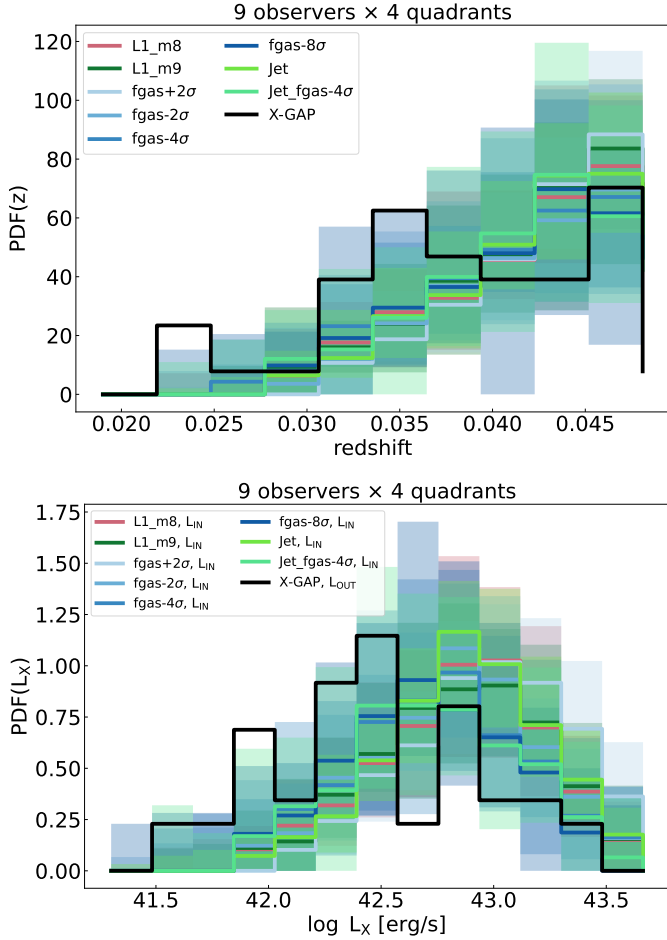


Fig. B.1. Redshift and luminosity distribution in the selected FLAMINGO samples and X-GAP.

the parameters in Table A.1. This gives us a predicted number of optically identified galaxy members starting from the number of true galaxy members accounting for the full observational selection process (Damsted et al. 2024; Eckert et al. 2024; Seppi et al. 2025).

Figure A.1 shows the final result, with the relation between true and detected number of members, along with the best-fit model described above. The bottom panel shows the full marginalised posterior distributions for the parameters in Eq. A.2.

Appendix B: X-GAP-FLAMINGO validation

In this appendix we quantify the impact of cosmic variance on the selected X-GAP-like samples across the various FLAMINGO models and validate the recovery of intrinsic X-ray observables through our analysis pipeline.

After applying the selection described in Sect. 2.3, we compare the redshift and luminosity PDFs of the selected groups to X-GAP (Fig. B.1), where shaded regions indicate the 16–84th percentiles across observers and light cones. X-GAP appears over(under)-dense at $z \sim 0.035$ ($z \sim 0.043$), leading to an excess (deficit) of systems around $L_X \sim 2 \times 10^{42}$ erg s⁻¹ ($\sim 10^{43}$ erg s⁻¹), consistent with the flux-limited selection. As shown in Fig. 1, such variations are dominated by cosmic variance, and it is unlikely that FLAMINGO reproduces the local Universe exactly. While LSS fluctuations may affect observable dis-

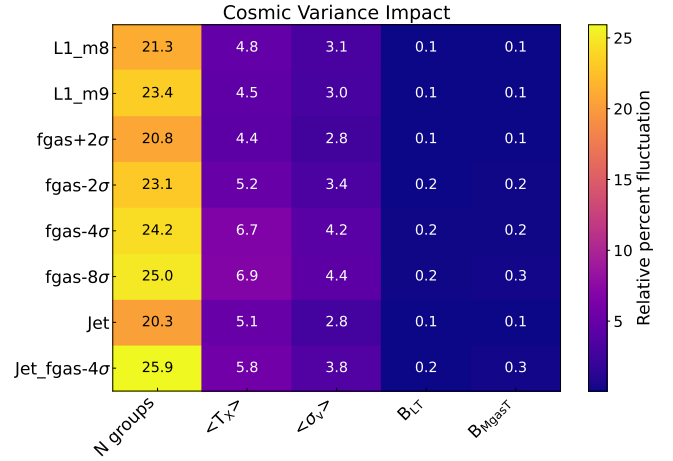


Fig. B.2. Relative fractional variation of each observable estimated from the cosmic variance tests.

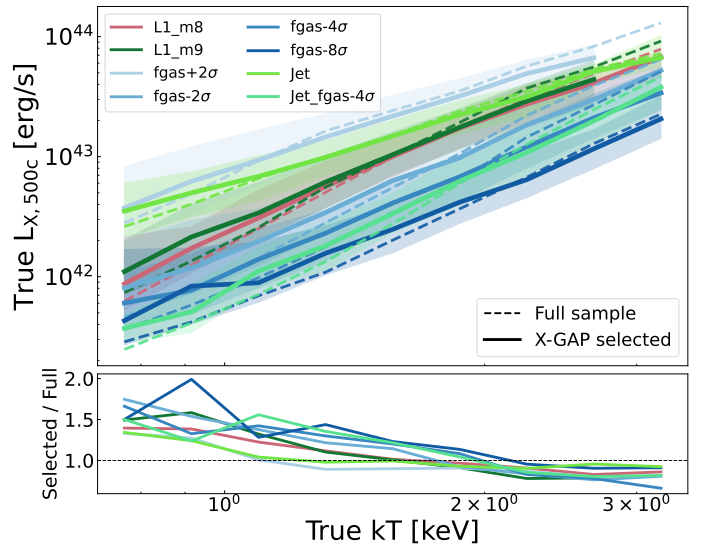


Fig. B.3. Luminosity-temperature relation for the selected systems (solid lines and shaded areas) and the full sample (dashed lines). Both are true input quantities.

tributions, this is mitigated by the copula approach (Sect. 4.2 and Appendix C). Moreover, the scaling-relation normalisations ($M_{\text{gas}}, L_{X,\text{CEX}}$), evaluated at fixed temperature, are largely unaffected and dominate the constraints.

We verify this using the cosmic variance and selection-function tests of Sect. 2.4. Relative variations, defined as $(p_{84} - p_{16}) / (2 p_{50})$, are shown in Fig. B.2. The number of groups varies by $> 20\%$, while mean temperature and velocity dispersion vary at the 5% and 4% level, respectively. The scaling-relation normalisations at 1.3 keV are highly stable, with variations of 0.2%.

We also examine the impact of selection on the $L-T$ relation, analogous to the $f_{\text{gas}}-M$ relation shown in Fig. 2. The result is presented in Fig. B.3. The selection bias follows a similar pattern: incompleteness at low masses and a down-scattered population at high masses caused by the upper radius cut. However, the effect is more pronounced for the $L-T$ relation because gas fraction is less directly tied to the selection observable (X-ray flux). We measure deviations of $\sim 30\%$ around 1 keV and $\sim 10\%$ near 3 keV.

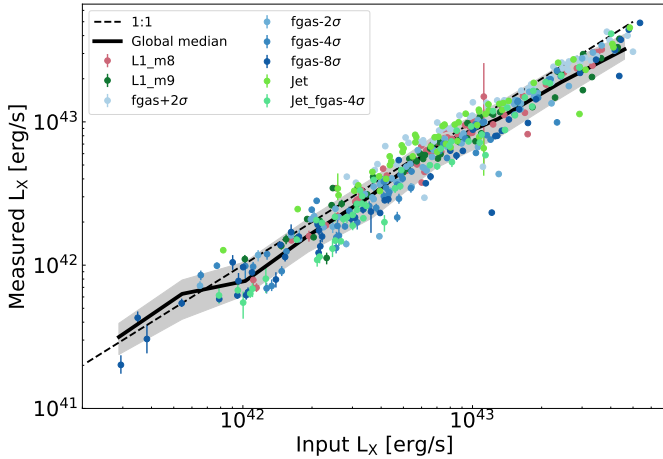


Fig. B.4. Comparison between the input X-ray luminosity and the output luminosity measured and modelled with hydromass.

We now focus the recovery of input X-ray luminosities, temperatures, and gas masses in the FLAMINGO groups.

B.1. X-ray luminosity

In our framework, X-ray luminosity is not computed as a simple photon count within an aperture. Instead, hydromass jointly models the gas density and temperature profiles to reproduce the observed surface brightness, which depends on the line-of-sight integral of n_e^2 and on temperature and metallicity through the cooling function (Ploekinger & Schaye 2020). Once the surface brightness and projected temperatures are fitted, the resulting 3D profiles yield the plasma emissivity, which is integrated along the line of sight to obtain the luminosity. The integration radius is critical, particularly for groups with flatter profiles, as it directly impacts the total luminosity. We therefore adapt hydromass to integrate out to the true R_{500c} rather than the value inferred from the data.

We compare our measured luminosities to the ones tabulated by Braspenning et al. (2024, see their Sects. 2.3 and 2.4), and rescaled by the cylindrical correction C_{cyl} obtained in Sect. 2.3, i.e. the ones used for the selection function and as inputs to the `xmm_simulator`. The result is shown in Fig. B.4. The black line shows the global measured luminosity at a given true value. The shaded area includes the 16th-84th percentiles. Although the solid line appears to be just below the 1:1, with a global median that is 6% lower than the input value, the scatter on the median is well within the one-to-one line. We conclude that we do not find evidence for significant bias in the luminosity measurement. It means that any difference in the computation of the cooling rates using `cldy` (in Braspenning et al. 2024) or `apex` (this work) does not impact the measurement of X-ray luminosity in our framework. In addition, a slight offset and a portion of the observed scatter may originate from the cylindrical correction (see Eq. 2). The correction is applied to the reference true luminosities, whereas the reconstructed values are derived independently from the mock photon data. Therefore, any inaccuracy in the correction would manifest as a departure from the one-to-one relation. The close agreement observed here suggests that the cylindrical correction is indeed robust.

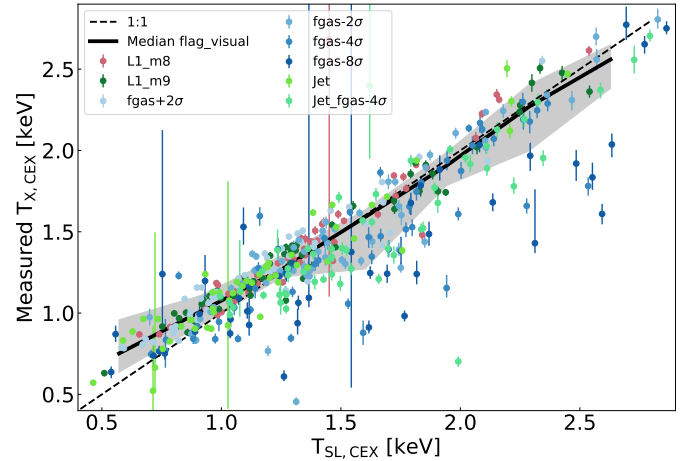


Fig. B.5. Comparison between the input spectroscopic-like temperature and the output temperature measured and modelled with hydromass. Both quantities are core-excised, using the estimated R_{500c} as a reference.

B.2. X-ray temperature

A single definition of the true global hot gas temperature in simulations is not trivial and requires an assumption of weights for each gas element. The introduction of the spectroscopic-like weighting scheme bridged a gap between temperature definition in simulations and temperatures inferred from X-ray spectra (Mazzotta et al. 2004), although some scatter is still expected (Seppi et al. 2026). Dense and cool regions dominate the emissivity in X-ray spectra, which biases the temperature to such regions. An additional effect is given by metallicity: contribution from line emission is extremely relevant in groups where the temperature is not hot enough to fully ionise all elements (e.g. Sanders 2023). The FLAMINGO simulations produced metal rich cores, with abundances that can reach super solar values in the inner region on average as reported by Braspenning et al. (2024). The X-ray spectral analysis in such a regime that is extreme in comparison to observations is challenging. In addition, the particle resolution allows to resolve the core within the inner $0.1 \times R_{500c}$ with tens of particles at best in the L1_m9 simulations. For these reasons, we compute global temperatures by excising the core and focus on the region between 0.15 and $1.0 \times R_{500c}$ (we do the same for observations). From the simulation perspective, we compute spectroscopic-like temperatures within this ring using the apertures defined with the estimated R_{500c} to be consistent with the temperature estimated from the mock observations.

The comparison is shown in Fig. B.5. We find the majority of the points to align close to the one-to-one line. However, we also obtain a cloud of points lying below the one-to-one line, made up mostly of systems in the strongest feedback simulations: $f_{gas} - 8\sigma$ and $Jet_{fgas} - 4\sigma$. From the visual inspection the majority of these systems showcase a complex IGrM, clearly disrupted by the strong AGN activity. We define a boolean flag, `FLAG_VISUAL`, which is set to True for systems that do not display obvious features likely to compromise the X-ray analysis, such as clear merger signatures, highly disturbed morphologies, very clumpy IGrM, strong asymmetries, or evident deviations from spherical symmetry. The median temperature for these systems is shown by the black solid line in Fig. B.5. The black shaded area denotes the 16th-84th percentiles. It encompasses the one-to-one line for the majority of the temperature range

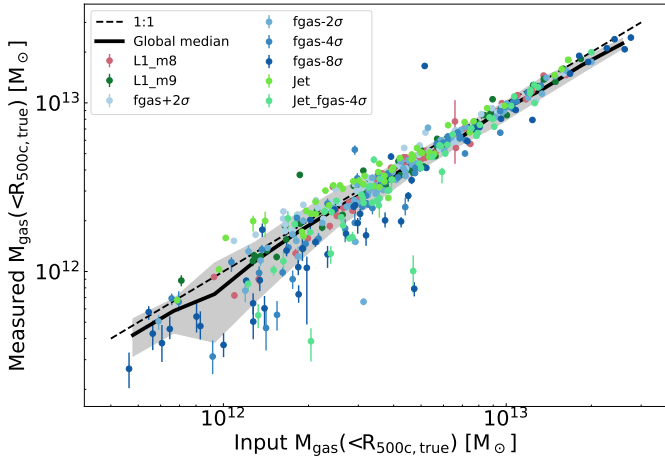


Fig. B.6. Comparison between the input and output gas mass measured and modelled with hydromass. Both quantities are integrated within the same radius (true R_{500c}).

considered here. A small disagreement at low temperature is likely due to the hydrogen column density, that absorbs the softer part of the spectrum, and the effective area, that is higher in the 1-2 keV range. Their combination may slightly bias our temperatures high in this regime, although the difference is not statistically significant ($< 2\sigma$). Overall, Fig. B.5 shows that for this subsample we find good agreement with the input spectroscopic-like temperature. For the global comparison to X-GAP, we use the full samples, otherwise we would bias the selection relative to the data.

B.3. Gas mass

The gas mass is measured by integrating the deprojected density profile following

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

where μ_e is the mean electron molecular weight and m_p is the proton mass. In particular, the mean molecular weight per electron is computed in a flexible way based on the chosen abundance table (Asplund et al. 2009, the one used for the X-GAP analysis) within hydromass. The pipeline computes a conversion factor that relates electron number density and emissivity, which therefore includes flexible dependencies on temperature and metallicity.

The result is shown in Fig. B.6. The black line shows the global measured gas mass at a given true value. The shaded area includes the 16th-84th percentiles. We find excellent agreement with the one-to-one relation: our framework allows accurate gas mass measurements. Importantly, this agreement persists even though the FLAMINGO simulations produce inner metallicity profiles that are systematically higher than those observed (Schaye et al. 2023; Braspenning et al. 2024). This indicates that our analysis pipeline is sufficiently flexible to recover unbiased gas masses, owing to the adaptive conversion between gas density and X-ray emissivity implemented in our modelling.

Appendix C: The statistical comparison to X-GAP

In this appendix we detail the framework related to the statistical comparison between X-GAP and FLAMINGO in Sect. 4.

C.1. Observables correlation

From the N-light cone generation procedure (explained in Sect. 4.2) we also store the input temperature and input velocity dispersion. Although these are not identical to the measured quantities, they provide a physically motivated baseline to quantify correlations between observables driven by halo population and selection effects. We also fit the normalisation of the scaling relation between input luminosity and temperature as explained in the previous section. This provides a data vector with B_{LT} , B_{MgasT} , N_{groups} , mean temperature, and mean velocity dispersion for each experiment. From the ensemble we compute the correlation matrix $C_{i,j}$ between the observables, shown in Fig. C.1. Results are consistent across models. We find a strong covariance between mean temperature and velocity dispersion (correlation ~ 0.8), as both trace halo mass. The scaling-relation normalisations are also correlated (0.5–0.7), reflecting the link between gas mass and X-ray luminosity (e.g. Eckert et al. 2020). In contrast, the number of groups shows no significant correlation with other observables, being dominated by cosmic variance. Similarly, the L - T normalisation is also largely independent: it describes luminosity variations at fixed temperature, so changes in temperature or velocity dispersion do not directly affect its normalisation. The same holds for the $M_{\text{gas}}-T$ normalisation.

C.2. The copula approach

For each observable individually, we construct its empirical cumulative distribution function (CDF) \mathcal{F}_i from the Monte Carlo procedure, without assuming any analytic form. For temperature, velocity dispersion, and B_{LT} , the Monte Carlo distributions primarily capture measurement and modelling uncertainties (bootstrap scatter and posterior sampling). However, the light-cone ensemble shows that field-to-field (cosmic variance and selection-function) fluctuations can also contribute to the scatter of sample-averaged quantities. To ensure a consistent comparison with X-GAP, we add this additional variance in quadrature to the measurement-driven scatter when it is not already dominant. The final empirical distributions \mathcal{F}_i therefore represent the total expected uncertainty for each observable, combining measurement effects and sample-to-sample fluctuations.

We use a Gaussian copula to model the observables covariance. A Gaussian copula provides a flexible framework to model the joint distribution of multiple variables by separating the individual properties of each observable (often named marginal distributions) from the way they co-vary (often named dependence structure). This approach is motivated by the need to combine observables with diverse, empirical distributions while preserving their intrinsic covariance. By mapping the marginals onto a latent multivariate Gaussian space, we can generate synthetic realisations that honour the measured correlation matrix regardless of the underlying shape of the individual 1D distributions. Operationally, this is implemented as follows: we draw random Gaussian vectors z_j from the correlation matrix $C_{i,j}$. Each component is transformed into a uniform variable using the standard normal CDF: $u_j = \Phi(z_j)$. The u_j values span the range between 0 and 1 but retain the expected correlation from $C_{i,j}$. For each observable, u_j is mapped onto a physical value $x_j = \mathcal{F}_j^{-1}(u_j)$ using the inverse empirical CDF of that observable. This procedure produces synthetic realisations that follow the empirical 1D distributions of each observable, and reproduce the measured correlation structure between them. Finally, since the empirical

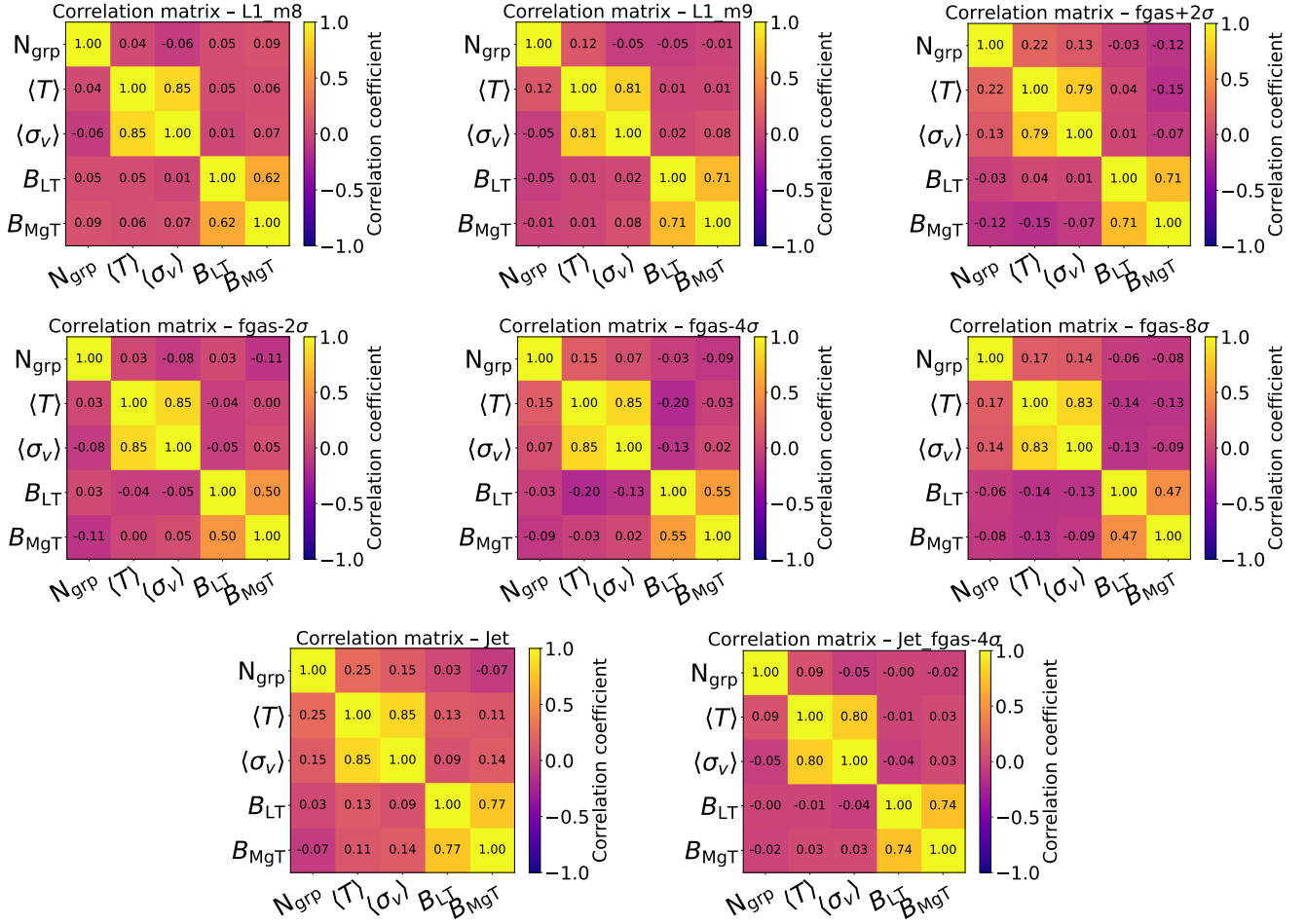


Fig. C.1. Correlation matrix between observables estimated from various light cones and selection function models for the various FLAMINGO models: the number of groups, the mean core-excised temperature, the mean velocity dispersion, and the normalisation of the scaling relations.

CDF only contain scatter due to the measurement process, we add the scatter estimated from the N -light cone generation to also account for systematics due to cosmic variance and selection function uncertainties (see also Fig. B.2 about the relative impact on each observable). These samples represent the distribution of outcomes expected from that model under the same survey selection and cosmic variance as X-GAP.

Now, for each observable we compute a two-sided tail probability p_j and define the summary statistic S following the Fisher method such that

$$p_j = 2 \times \min(\mathcal{F}_j(x_j), 1 - \mathcal{F}_j(x_j)),$$

$$S = -2 \sum_{i=1}^5 \log p_j. \quad (\text{C.1})$$

Because the p_j are correlated, we do not expect S to necessarily follow the standard χ^2 distribution. We show the distribution of the S statistic for each model in Fig. C.2 (a method similar to Regamey et al. 2026). We also calculate the S statistic for the X-GAP sample itself and indicate its position with a vertical line of the corresponding colour. The discriminating power of the test lies in the percentile at which the X-GAP value intersects the model's CDF. The CDFs are very similar across the FLAMINGO runs: the discriminating power arises primarily from the location of the X-GAP statistic within each distribution. The model $f_{\text{gas}} - 2\sigma$ places the X-GAP value close to the

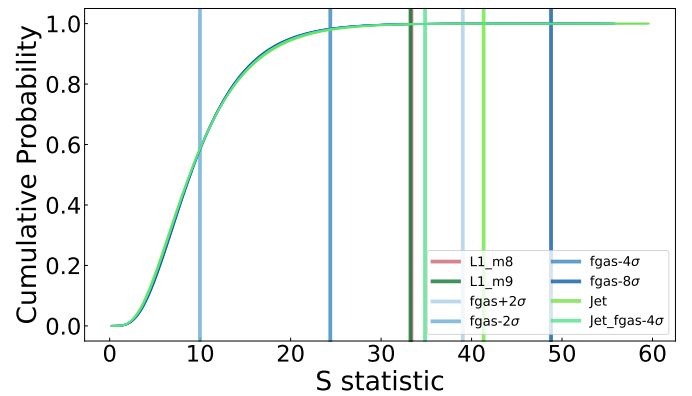


Fig. C.2. Cumulative distribution functions of the summary statistic S for the different FLAMINGO models and for the X-GAP sample. The curved lines represent the distribution of S generated via a Gaussian copula to account for the covariance between observables in the FLAMINGO realisations. The vertical lines indicate the position of the observed X-GAP data within each model distribution. A model is in good agreement if the vertical line falls near the median of the CDF (like $f_{\text{gas}} - 2\sigma$), while it is statistically disfavoured if the vertical line falls in the extreme tail (like $f_{\text{gas}} - 8\sigma$).

median of the distribution and provides the best agreement with the data, whereas the fiducial runs (L1_m8 and L1_m9) lie at

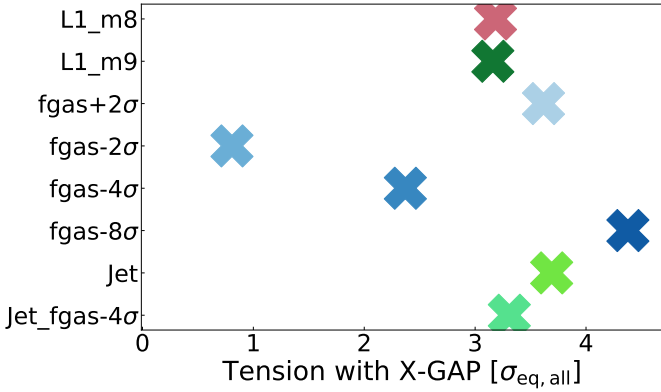


Fig. C.3. Final comparison to X-GAP. It includes the normalisation of the L - T (both core-excised) and $M_{\text{gas}}(<400 \text{ kpc})$ - T (core-excised) scaling relations, the expected number of groups, the mean core-excised temperature, and velocity dispersion. The tension expressed in Gaussian equivalent significance accounts for the intrinsic correlation between observables as explained in Sect. 4.2. The values are also reported in Table 1.

higher percentiles and are mildly disfavoured. Models with very strong feedback (e.g. $f_{\text{gas}} - 8\sigma$) place the X-GAP value in the extreme tail of the distribution and are strongly inconsistent with the observations. The joint probability p_{joint} is given empirically by the fraction of S larger than the one computed for X-GAP, i.e. the fraction of simulated realisations that are at least as extreme as X-GAP in the combined 4D space. Finally, we convert the joint probability into an equivalent two-sided Gaussian significance

$$\sigma_{\text{eq}} = \Phi^{-1}\left(1 - \frac{p_{\text{joint}}}{2}\right), \quad (\text{C.2})$$

where Φ^{-1} is the inverse standard normal cumulative distribution function. The σ_{eq} quantifies how unusual the X-GAP measurements are relative to each model, while fully accounting for covariance between observables (Fig. C.3 and Table 1).

Appendix D: Potential systematics

Here we address three potential sources of systematics: the gas particle metallicity the sample size, and the cosmological model.

D.1. Gas particle metallicity

FLAMINGO predicts metal abundances higher than observed, reaching super-solar values in group cores (Braspenning et al. 2024). In this mass regime, where the gas is not fully ionised, line emission contributes significantly to the total luminosity, making metallicity a key driver of L_X . Higher abundances can therefore boost the luminosity at fixed gas mass. This may affect the results in Fig. 5. The L - T normalisation is largely unaffected, as luminosity drives the selection. In contrast, the $M_{\text{gas}}-T$ relation is more sensitive: an enhanced L_X would require lower M_{gas} to match the observations.

We recompute the input X-ray luminosities with pyXSIM by fixing the metallicity of gas particles within R_{500c} to $0.3 Z_{\odot}$ for haloes with $M_{500c} > 5 \times 10^{12} M_{\odot}$. This requires collecting the gas particles around each halo and recalculating their emissivity at fixed abundance. Because the procedure is computationally expensive, we restrict the analysis to a single L1_m8 light cone at $z \leq 0.05$. We then apply the selection described in Sect. 2.3

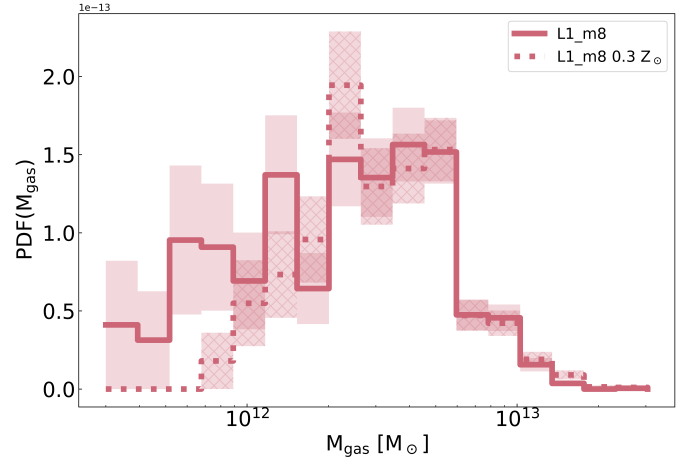


Fig. D.1. Comparison of the gas mass distribution between the sample selected from L1_m8 as described in Sect. 2.3 and the variation assuming $Z=0.3$ for all gas elements.

using these recomputed luminosities instead of the catalogue values. Since Appendix B shows that our pipeline accurately recovers gas masses, we do not generate new mock *XMM-Newton* observations but directly compare the gas-mass distributions of the original L1_m8-selected sample and the variant computed with $0.3 Z_{\odot}$.

The result is shown in Fig. D.1. The solid line with the shaded area denotes the standard X-GAP-like sample from L1_m8, while the dotted line with hatched shaded area corresponds to the $Z=0.3 Z_{\odot}$ variation. Overall we find good agreement between the two, especially for average to higher gas masses where the distributions are almost indistinguishable. The agreement is within 1σ down to about $M_{\text{gas}} = 10^{12} M_{\odot}$. Below such a threshold the standard L1_m8 contains more objects, which skews the median gas mass to lower values compared to the $Z=0.3$ variation. This is in agreement with fainter groups being boosted by the higher metallicity. The median gas mass in the two samples is $4.5 \pm 0.2 \times 10^{12}$ and $4.7 \pm 0.2 \times 10^{12} M_{\odot}$. The uncertainties are estimated from 10000 bootstrap samples. The two values differ by about 5%, although they are compatible at the 0.9σ level. Therefore, we do not find a statistical disagreement between the two.

In addition, the normalisation of the $M_{\text{gas}}-T$ is even less affected overall, because it is computed at fixed temperature. We find $B_{M_{\text{gas}}T} = 12.45 \pm 0.01$ for L1_m8 and $B_{M_{\text{gas}}T} = 12.46$ for the $Z=0.3$ variation. We conclude that even though a particle selection with lower metallicity would not allow selecting a few low gas mass objects, the normalisation of the $M_{\text{gas}}-T$ relation, which brings constraining power to our framework, is not significantly affected by this systematic.

D.2. Sample size

We study whether the relatively limited sample size of about 50 groups in X-GAP and the one used in this work for comparison to the models has an impact on our results. We investigate the 2D distribution of gas mass and temperature, and X-ray luminosity and temperature. We compare the sample from L1_m8 used for comparison to X-GAP to the full combination of all light cones and observers used for the cosmic variance tests. We use the input values, because we do not have the full end-to-end

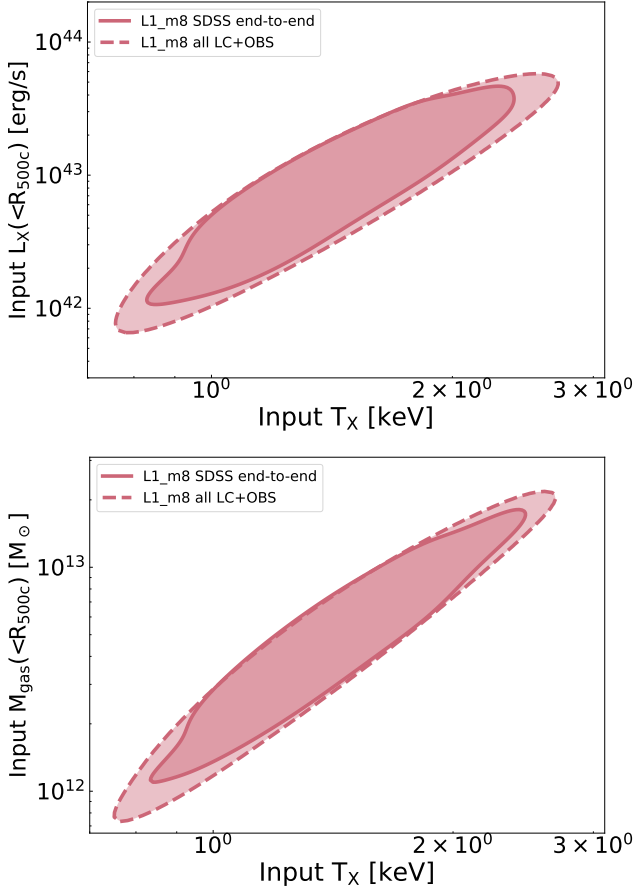


Fig. D.2. Comparison of the input $M_{\text{gas}}-T$ and $L-T$ relations between the L1_m8 sample (solid lines) used for comparison to X-GAP and the full combination of all light cones and observers used for the cosmic variance tests (dashed lines).

pipeline for all light cones and observers. In total we collect 2368 haloes.

The results are shown in Fig. D.2. The two populations occupy a very similar region of the parameter space. The larger sample combining all light cones and observers is more likely to contain rarer objects that deviate from the mean relations, therefore the dashed contours extend to slightly larger regions of the parameter space.

The median values of each quantity are very close in the two samples (T of 1.40 and 1.41 keV, M_{gas} of 4.65 and $4.48 \times 10^{13} M_{\odot}$, L_X of 7.48 and 7.39×10^{42} erg/s). The 1D 16th percentiles move from 1.79 to 1.81 keV (7.59 to $8.41 \times 10^{13} M_{\odot}$, 3.1 to 2.7×10^{42} erg/s). The 1D 84th percentiles move from 1.15 to 1.07 keV (2.5 to $2.1 \times 10^{13} M_{\odot}$, 1.71 to 1.73×10^{43} erg/s).

Finally, the normalisation of the two scaling relations move from 12.56 to 12.55 for $M_{\text{gas}}-T$ and from 42.76 to 42.75 for $L-T$. We conclude that the sample size does not significantly affect our results.

D.3. Cosmology

The number of selected groups is expected to also depend on cosmology. To study this effect, we follow a methodology similar to Regamey et al. (2026) and integrate the halo mass function from Tinker et al. (2008) up to $z=0.05$ and in the 10^{13} – $10^{14} M_{\odot}$ range. We account for an SDSS-like sky coverage. We

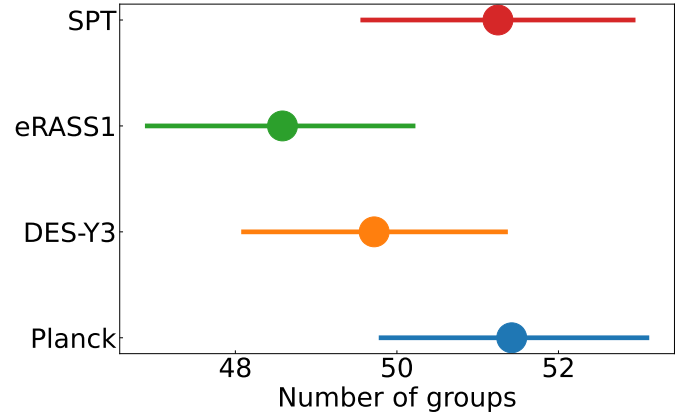


Fig. D.3. Comparison of the predicted number of X-GAP-like selected groups for different cosmological models.

rescale the total number of generated haloes by the ratio between the X-GAP selected systems and the number of all haloes in the L1_m8 light cone, that is equal to about 6%. The result is a rough estimate of the total number of selected groups. We then vary the cosmological model and study the impact on the final number of groups. We use the best-fit parameters from Planck (Planck Collaboration VI 2020), DES-Y3 (Abbott et al. 2022), eRASS1 clusters (Ghirardini et al. 2024), and SPT clusters (Bocquet et al. 2025).

The result is shown in Fig. D.3. The errorbars are estimated from 1000 samples generated from a different random seed for each cosmology. We obtain a median number of 51.4 ± 1.8 for Planck, 49.7 ± 1.7 for DES-Y3, 48.5 ± 1.7 for eRASS1, and 51.3 ± 1.7 for SPT. All results are compatible within 1σ . In addition, the statistical uncertainty and the variation between different cosmological models are about one order of magnitude lower than the uncertainty due to cosmic variance estimated in Sect. 2.4. We conclude that our results are not significantly affected by the choice of a specific cosmological model.

Appendix E: X-GAP measurements

Table E.1 collects the individual values of gas mass within 400 kpc, core-excised X-ray luminosity and temperature measured for each individual X-GAP group. These three observables have been measured for this work and have not been published before. We refer the reader to Eckert et al. (2025) for full details about the X-ray analysis pipeline, and Eckert et al. (2024) for the X-GAP main master table.

Table E.1. Measurements of gas mass within 400 kpc, core-excised temperature, and core-excised X-ray luminosity for X-GAP groups. See Eckert et al. (2024) for full properties.

Group ID	redshift	$\log_{10} M_{\text{gas}}(<400 \text{ kpc}) [M_{\odot}]$	$\log_{10} L_{\text{X,CEX}} [\text{erg/s}]$	$T_{\text{X,CEX}} [\text{keV}]$
8050	0.047	$12.38^{+0.01}_{-0.01}$	$42.85^{+0.01}_{-0.01}$	$1.76^{+0.04}_{-0.06}$
10842	0.040	$12.33^{+0.01}_{-0.01}$	$42.74^{+0.01}_{-0.01}$	$1.62^{+0.03}_{-0.03}$
885	0.047	$12.55^{+0.01}_{-0.01}$	$43.07^{+0.01}_{-0.01}$	$2.50^{+0.08}_{-0.08}$
1011	0.046	$12.36^{+0.01}_{-0.01}$	$42.76^{+0.01}_{-0.01}$	$1.56^{+0.03}_{-0.03}$
9695	0.038	$12.38^{+0.01}_{-0.01}$	$42.78^{+0.00}_{-0.00}$	$1.74^{+0.02}_{-0.02}$
4654	0.022	$11.84^{+0.02}_{-0.02}$	$41.64^{+0.01}_{-0.01}$	$0.86^{+0.02}_{-0.02}$
1695	0.039	$12.53^{+0.02}_{-0.02}$	$43.03^{+0.02}_{-0.02}$	$1.97^{+0.15}_{-0.15}$
15641	0.027	$12.10^{+0.00}_{-0.00}$	$42.21^{+0.01}_{-0.01}$	$0.95^{+0.03}_{-0.03}$
3128	0.032	$12.03^{+0.01}_{-0.01}$	$42.25^{+0.01}_{-0.01}$	$1.10^{+0.02}_{-0.02}$
2620	0.039	$12.42^{+0.01}_{-0.01}$	$42.89^{+0.01}_{-0.01}$	$1.57^{+0.02}_{-0.02}$
12349	0.036	$12.25^{+0.01}_{-0.01}$	$42.53^{+0.00}_{-0.00}$	$1.04^{+0.01}_{-0.01}$
16150	0.032	$11.79^{+0.05}_{-0.04}$	$41.71^{+0.06}_{-0.08}$	$1.10^{+0.02}_{-0.02}$
3460	0.043	$12.18^{+0.02}_{-0.02}$	$42.35^{+0.02}_{-0.02}$	$1.25^{+0.02}_{-0.03}$
9771	0.044	$12.17^{+0.02}_{-0.02}$	$42.28^{+0.02}_{-0.03}$	$1.38^{+0.08}_{-0.07}$
1601	0.034	$12.08^{+0.01}_{-0.01}$	$42.20^{+0.01}_{-0.02}$	$1.07^{+0.01}_{-0.01}$
10094	0.031	$11.81^{+0.02}_{-0.02}$	$41.86^{+0.02}_{-0.02}$	$1.08^{+0.03}_{-0.03}$
6159	0.024	$11.79^{+0.05}_{-0.05}$	$41.42^{+0.04}_{-0.04}$	$0.72^{+0.04}_{-0.03}$
8102	0.033	$12.13^{+0.01}_{-0.01}$	$42.34^{+0.01}_{-0.01}$	$1.10^{+0.02}_{-0.02}$
9178	0.040	$11.89^{+0.11}_{-0.17}$	$41.56^{+0.10}_{-0.18}$	$0.84^{+0.09}_{-0.07}$
39344	0.028	$12.01^{+0.01}_{-0.01}$	$42.19^{+0.01}_{-0.01}$	$0.85^{+0.01}_{-0.01}$
5742	0.034	$12.06^{+0.02}_{-0.03}$	$41.93^{+0.02}_{-0.03}$	$0.79^{+0.03}_{-0.02}$
6058	0.045	$11.87^{+0.04}_{-0.07}$	$41.76^{+0.04}_{-0.07}$	$0.89^{+0.05}_{-0.04}$
1162	0.044	$12.27^{+0.01}_{-0.01}$	$42.64^{+0.01}_{-0.02}$	$1.42^{+0.02}_{-0.02}$
2424	0.040	$12.15^{+0.02}_{-0.02}$	$42.26^{+0.02}_{-0.02}$	$1.17^{+0.03}_{-0.02}$
46701	0.042	$12.13^{+0.02}_{-0.02}$	$42.24^{+0.02}_{-0.03}$	$1.05^{+0.05}_{-0.05}$
35976	0.036	$12.29^{+0.02}_{-0.02}$	$42.64^{+0.02}_{-0.02}$	$1.28^{+0.05}_{-0.04}$
9399	0.035	$12.18^{+0.02}_{-0.02}$	$42.29^{+0.02}_{-0.02}$	$1.10^{+0.03}_{-0.03}$
4436	0.046	$12.29^{+0.01}_{-0.01}$	$42.59^{+0.02}_{-0.01}$	$1.58^{+0.06}_{-0.07}$
9647	0.023	$12.06^{+0.02}_{-0.02}$	$41.89^{+0.02}_{-0.02}$	$0.80^{+0.02}_{-0.02}$
4936	0.042	$11.77^{+0.12}_{-0.14}$	$41.42^{+0.11}_{-0.10}$	$0.86^{+0.08}_{-0.06}$
828	0.046	$12.62^{+0.01}_{-0.01}$	$43.28^{+0.01}_{-0.01}$	$2.47^{+0.06}_{-0.06}$
15776	0.036	$12.14^{+0.04}_{-0.04}$	$42.29^{+0.03}_{-0.04}$	$1.01^{+0.07}_{-0.05}$
9370	0.038	$12.12^{+0.03}_{-0.04}$	$42.27^{+0.03}_{-0.03}$	$1.09^{+0.05}_{-0.04}$
1398	0.046	$12.56^{+0.01}_{-0.01}$	$43.25^{+0.01}_{-0.01}$	$2.58^{+0.07}_{-0.07}$
40241	0.049	$12.15^{+0.02}_{-0.02}$	$42.34^{+0.01}_{-0.02}$	$0.89^{+0.01}_{-0.01}$
11844	0.038	$12.12^{+0.05}_{-0.06}$	$42.15^{+0.04}_{-0.04}$	$0.69^{+0.04}_{-0.04}$
10159	0.031	$11.73^{+0.07}_{-0.11}$	$41.36^{+0.06}_{-0.09}$	$0.59^{+0.07}_{-0.04}$
11320	0.045	$12.39^{+0.01}_{-0.01}$	$42.93^{+0.02}_{-0.02}$	$2.46^{+0.50}_{-0.23}$
11631	0.046	$12.42^{+0.02}_{-0.02}$	$42.95^{+0.01}_{-0.01}$	$1.57^{+0.12}_{-0.09}$
16393	0.046	$12.09^{+0.02}_{-0.02}$	$42.30^{+0.02}_{-0.03}$	$1.24^{+0.07}_{-0.04}$
22635	0.034	$12.21^{+0.02}_{-0.03}$	$42.48^{+0.02}_{-0.03}$	$1.17^{+0.11}_{-0.06}$
28674	0.037	$12.14^{+0.01}_{-0.01}$	$42.36^{+0.01}_{-0.02}$	$1.07^{+0.06}_{-0.05}$
3513	0.036	$11.90^{+0.07}_{-0.09}$	$41.84^{+0.07}_{-0.07}$	$1.25^{+0.28}_{-0.17}$
3669	0.048	$12.49^{+0.01}_{-0.01}$	$43.11^{+0.02}_{-0.02}$	$1.85^{+0.23}_{-0.16}$