

Characterising galaxy cluster scaling relations as cosmic isotropy tracers using FLAMINGO simulations

Yujie He^{1,*}, Konstantinos Migkas^{1,2}, Joop Schaye¹, Joey Braspenning^{3,1}, and Matthieu Schaller^{4,1}

¹ Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

² SRON Netherlands Institute for Space Research, Niels Bohrweg 4, NL-2333 CA Leiden, The Netherlands

³ Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

⁴ Lorentz Institute for Theoretical Physics, Leiden University, P.O. box 9506, 2300 RA Leiden, The Netherlands

Received 1 April 2025 / Accepted 17 August 2025

ABSTRACT

The standard cosmological model, Λ CDM, assumes isotropy on large cosmic scales. However, recent studies using galaxy cluster scaling relations have reported an apparent H_0 anisotropy at 5.4σ that could be attributed to large bulk flows extending beyond 500 Mpc, which is in disagreement with Λ CDM. To quantify the statistical tension of the observational galaxy cluster data used in past studies with Λ CDM, we utilised the isotropic $(2.8 \text{ Gpc})^3$ run of the FLAMINGO (Λ CDM) simulations, the largest hydrodynamical cosmological simulation available to date. We created 1728 simulated lightcones and studied the apparent level of anisotropy traced by X-ray and thermal Sunyaev-Zeldovich scaling relations in the same cluster sample selection and methodology as in the past study. We find the probability of such apparent anisotropies randomly emerging in cluster scaling relations within a Λ CDM universe to be 0.12% (3.2σ). The discrepancy goes up to $\sim 3.6\sigma$ when modelled as a bulk flow at $z < 0.1$. We also find that statistical noise accounts for over 80% of the anisotropy amplitude in each lightcone, with large peculiar velocities contributing less than 20%. We also show that anisotropy amplitudes are highly sensitive to the intrinsic scatter in the scaling relations, with tighter relations providing stronger constraints. Nevertheless, the tension between the past results and Λ CDM persists, albeit at a lower significance than previously reported.

Key words. methods: numerical – methods: statistical – cosmology: miscellaneous – large-scale structure of Universe – X-rays: galaxies: clusters

1. Introduction

One of the most important underlying assumptions of our standard model of cosmology, Λ CDM, is the statistical homogeneity and isotropy of space and matter. This is known as the cosmological principle (CP). This model has been remarkably predictive in explaining various cosmic phenomena, including the power spectrum of the cosmic microwave background (CMB), the abundance of light elements, and the distribution of large-scale structures. However, recent observations have revealed several notable tensions with Λ CDM, including the long-standing H_0 tension (e.g. Riess et al. 2019, 2022; Planck Collaboration VI 2020; Valentino et al. 2021; Shah et al. 2021) and the S_8 tension (e.g. Joudaki et al. 2017; DES Collaboration 2022; Philcox & Ivanov 2022; McCarthy et al. 2023; Ghirardini et al. 2024). In addition, many studies have examined the validity of CP, and a number of anomalies have been reported, including the cosmic dipole (Rubart & Schwarz 2013; Tiwari et al. 2015, 2024; Secrest et al. 2021, 2022; Dam et al. 2023; Oayda & Lewis 2023; Mittal et al. 2024; da Silveira Ferreira & Marra 2024), SNIa anisotropy (Chang & Lin 2015; Mohayaee et al. 2021; Cowell et al. 2023; Hu et al. 2024a; Perivolaropoulos 2023; Mc Conville & Colgáin 2023), excessive bulk flows on large scales¹ (Kashlinsky et al.

2010; Salehi et al. 2021; Whitford et al. 2023; Watkins et al. 2023), and last but not least, anisotropic Hubble expansion (Migkas et al. 2021; Luongo et al. 2022; Pandya et al. 2024; Haridasu et al. 2024; Boubel et al. 2025). We refer to some reviews for more on cosmic tensions and anomalies (Bull et al. 2016; Abdalla et al. 2022; Perivolaropoulos & Skara 2022; Aluri 2023). While there are indications of discrepancies, no definitive answer or consensus has been reached on whether the anomalies are real tensions or statistical artefacts.

The CP provides the foundation of our cosmology theory and a cosmic reference frame for observations. The importance of robust tests of the CP cannot be overstated. As we enter an era of high-precision multi-messenger cosmology, with a wide range of experiments to probe local and high-redshift Universe, it is crucial to fully understand the possible systematic biases that creep into the past results and to quantify their true tension with Λ CDM.

The Λ CDM adoption of the CP posits that the matter rest frame should converge to the CMB rest frame (e.g. Ferreira 2021; Wagnveldt et al. 2024) at scales $\geq 150 \text{ Mpc}^2$ (Fixsen et al. 1996; Planck Collaboration XXVII 2014). On such scales, the Hubble flow is expected to be statistically uniform across the sky, with negligible bulk flow motions ($\lesssim 140 \text{ km s}^{-1}$, e.g. Watkins et al. 2023), as dictated by the statistical homogeneity of the cosmic matter distribution. However, if such bulk flows are ignored, they can result in an observed $\lesssim 3\%$ anisotropy in the redshift-distance relation and the Hubble constant (H_0) within the CMB rest frame (Migkas 2025). However, recent studies

² This scale is decided by the cosmological model and not the CP itself.

* Corresponding author: yujiehe@strw.leidenuniv.nl

¹ Some studies do not claim a strong tension with Λ CDM but find ‘atypical’ bulk flows (e.g. Hoffman et al. 2024, 3.4σ at $250 h^{-1} \text{ Mpc}$) or external bulk flows that do not converge on larger scales (Carrick et al. 2015, Fig. 9). Bulk flow studies tracing smaller volumes are more often consistent with Λ CDM (e.g. Nusser & Davis 2011, $< 100 h^{-1} \text{ Mpc}$).

suggest that the galaxy cluster rest frame does not converge to the CMB rest frame out to $z \sim 0.1\text{--}0.2$ (Migkas & Reiprich 2018; Migkas et al. 2020, 2021; Pandya et al. 2024). (For a detailed review, see Migkas 2025).

Galaxy clusters are the largest virialised structures in the Universe. These massive objects are typically observed at redshift $z \sim 0.05\text{--}0.5$ with a uniform distribution across the sky, rendering them a powerful cosmology probe into the late Universe. Migkas et al. (2021, hereafter M21) studied the cosmic (an)isotropy within $\lesssim 1$ Gpc using X-ray, microwave, and infrared data for an X-ray flux-limited cluster sample with 313 galaxy clusters introduced in Migkas et al. (2020, hereafter M20). They also utilised the ASCA cluster sample (Horner 2001) containing another ~ 160 independent objects. They constructed the multi-wavelength scaling relations $L_X\text{--}T$, $Y_{SZ}\text{--}T$, and $L_{BCG}\text{--}T$, where L_X is the X-ray luminosity, Y_{SZ} is the total integrated Compton parameter, L_{BCG} is the near-infrared cluster central galaxy luminosity, and T is the cluster X-ray temperature. While the determination of L_X , Y_{SZ} , and L_{BCG} depend strongly on the underlying cosmology, T is measured independently of any cosmological assumptions. Consequently, angular variations of cosmological parameters and cluster bulk flow motions are imprinted in the directional behaviour of these scaling relations. Exploiting this feature, M21 detected an apparent 9% dipole variation of H_0 pointing towards $(l, b) = (280^\circ \pm 35^\circ, -15^\circ \pm 20^\circ)$ at 5.4σ . Alternatively, they showed that their findings could be attributed to a bulk flow of ~ 900 km s $^{-1}$ towards the opposite direction that extends to $\gtrsim 500$ Mpc. Despite very different potential systematics relevant to each relation and sample, the apparent anisotropy was consistently detected in all three nearly independent scaling relations and the two cluster samples. An extensive set of tests was performed to ensure that the result did not originate from any known systematic biases. More recently, Pandya et al. (2024) used cluster velocity dispersion (σ_v) data and the $L_X\text{--}\sigma_v$ and $Y_{SZ}\text{--}\sigma_v$ scaling relations to probe cosmic isotropy. They detected a $(27.6 \pm 4.4\%) H_0$ dipole³ pointing towards $(295^\circ \pm 71^\circ, -30^\circ \pm 71^\circ)$ at a 3.6σ level, which is consistent with M21. The existence of such a dipole is recently confirmed independently by Boubel et al. (2025) using the galaxy Tully-Fisher relation, which finds a $3 \pm 0.7\%$ H_0 variation in $(l, b) = (322 \pm 30^\circ, -52 \pm 10^\circ)$, close to M21 and Pandya et al. (2024). We note that we report the lowest H_0 directions for these works.

Despite the assumed isotropy on large scales in Λ CDM, one may observe that anisotropies emerge simply due to cosmic and sample variance as well as from residual bulk flow motions. The amplitude and significance of such observed anisotropies could be further upscattered due to statistical noise and systematic biases in the data selection and applied statistics. Furthermore, one needs to accurately know the expected observed H_0 variation within a Λ CDM universe to properly assess the tension between the observational findings in M21 and the Λ CDM model. Arguably the most optimal way to achieve all of the above is to apply the same sample selection and statistical techniques of M21 to isotropic cosmological hydrodynamical simulations.

The Virgo Consortium’s FLAMINGO (Full-hydro Large-scale structure simulations with All-sky Mapping for the Interpretation of Next Generation Observations) simulations include

³ Pandya et al. (2024) have also reported that the level of their observed H_0 variation is overestimated due to the large scatter of the used scaling relations. They estimate that the ‘true’ H_0 dipole they detect on top of statistical noise is $\sim 8\%$.

the largest state-of-the-art hydrodynamical cosmological simulation to date. FLAMINGO is described in detail in Schaye et al. (2023), hereafter S23. These simulations are designed for studies of galaxy clusters and other large-scale structures. We used its flagship run, the $(2.8 \text{ Gpc})^3$ simulation box using a baryon particle mass of $1.1 \times 10^9 M_\odot$ mass resolution (named L2p8_m9), reaching $z = 0$ with 2.8×10^{11} particles. The $(1.0 \text{ Gpc})^3$ fiducial run (named L1_m9), which uses the same model and resolution but only a smaller box and lower particle count of 1.3×10^{10} , is sometimes used for comparison purposes. The FLAMINGO project incorporates rich baryonic physics and covers an extensive cosmic volume, rendering it an excellent laboratory for our purpose. Additionally, the cluster thermodynamic profiles (Braspenning et al. 2024) and scaling relations (S23) have been shown to adequately match observations.

In this work, we aim to quantify the probability that the anisotropies observed in M21 arise in a Λ CDM universe. To this end, we placed 1728 observers within the FLAMINGO simulation volume of $(2.8 \text{ Gpc})^3$ and constructed 1728 isotropic lightcone samples. We applied the same cluster selection criteria and statistical procedures as in M21, and we injected realistic scatter and measurement uncertainties into the simulated cluster catalogues. This enabled us to assess the true level of tension between the multi-wavelength galaxy cluster observations and the Λ CDM model, while accounting for potential systematics in data selection, anisotropy modelling, statistical methodology, and random noise. In addition to the M21 framework, we employed a Markov chain Monte Carlo (MCMC) based statistical approach for cross-validation. We further investigated systematics that may artificially inflate the amplitude or significance of the observed anisotropies in M21, such as peculiar velocities and scatter in scaling relations. This study is an essential step towards establishing cluster scaling relations as a robust probe of H_0 variation and bulk flows.

This paper is organised as follows: In Sect. 2, we describe the procedures used to obtain the lightcones and the cluster catalogues. In Sect. 3, we introduce the statistical methods employed to constrain anisotropy and bulk flows. In Sect. 4, we present the main results and assess the probability at which the M21 result reconciles with Λ CDM. In Sect. 5, we discuss the implications of our findings and the systematics that could affect our analysis. Finally, we conclude in Sect. 6.

We note that in calculating distances and the evolution factor, $E(z)$ (Sect. 3), we followed the Dark Energy Survey year three cosmology ($H_0 = 68.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.306$, $\Omega_\Lambda = 0.694$, DES Collaboration 2022), which the FLAMINGO fiducial runs (L1_m9, L2p8_m9) we used are based on. Under all circumstances, we modelled only the relative difference of H_0 , and the fiducial choice of H_0 therefore has no bearing on the result.

2. FLAMINGO lightcone generation and cluster sample construction

Typical numerical simulations, such as FLAMINGO, output snapshots of the simulation box at fixed cosmic times. However, observations are limited to an observer’s past lightcone, where objects are seen as they were when the light they emitted first reached the observer. To replicate this in simulations, mock lightcones are constructed by stitching together multiple snapshots and assigning each object a redshift and sky position based on the time it crossed the observer’s past lightcone. This process assumes a cosmological model and produces a lightcone catalogue that accounts for the evolution of structures over cosmic time, enabling direct comparison with observations.

The identification of galaxy clusters in FLAMINGO was performed using the VELOCIRaptor subhalo finder (Elahi et al. 2019), which implemented a Friends-of-Friends (FOF) algorithm. After this, the halo properties were computed using the Spherical Overdensity and Aperture Processor (SOAP)⁴ (McGibbon et al. 2025), a tool developed by the FLAMINGO team alongside the simulations to compute various (sub)halo properties (McGibbon, in prep.). We used the SOAP catalogue to obtain L_X , Y_{SZ} , T , the cluster gas mass M_{gas} , and the centre-of-mass velocity \mathbf{v} .

FLAMINGO produced eight lightcones at runtime for the $(2.8 \text{ Gpc})^3$ run. In runtime lightcones, the particle position between snapshots is available, allowing for a more accurate estimation of the particle location \mathbf{x} and redshift z at the moment of crossing the lightcone (S23). Despite the high quality of these data, eight lightcones are not sufficient to robustly assess the rarity of the M21 results. Consequently, through a post-processing procedure, we created 1728 evenly spaced lightcones, 12 on each side, in the $(2.8 \text{ Gpc})^3$ simulation box.

2.1. FLAMINGO lightcones construction

The observers were placed at fixed comoving coordinates $(iL/12, jL/12, kL/12)$, where i, j, k are integers from -6 to $+5$ and $L = 2.8 \text{ Gpc}$ is the box size. For consistency with M21, we considered only clusters at $z \leq 0.30$, corresponding to a comoving distance of $\sim 1.2 \text{ Gpc}$. Given SOAP's cadence $\Delta z = 0.10$, four snapshots are required to construct the lightcones to $z \leq 0.30$. This redshift range is well below the half-box size 1.4 Gpc , avoiding edge effects and box repetition. For our study, box repetition could introduce false statistics by duplicating the same scaling behaviour, biasing the scaling relation fit. Additionally, $z < 0.30$ is sufficient to include most of the flux-selected clusters, with a higher-redshift tail containing less than 1% of the sample, as seen in M20 and M21.

Lightcones were created as follows: for each halo in snapshot n , at cosmic redshift z_n , we calculated the comoving distance r from the cluster to the observer and compare it with $r(z_n) = \int_0^{z_n} \frac{cdz'}{H(z')}$. The cluster is included in the lightcone if its comoving distance falls within the shell $r(z_n - \Delta z/2) < r < r(z_n + \Delta z/2)$. We then solved for the precise cosmic redshift, z_{cos} , that satisfies the relation between comoving distance and light travel distance along radial null geodesics:

$$-\int_0^{z_{\text{cos}}} \frac{cdz'}{H(z')} + r(a) = 0. \quad (1)$$

After z_{cos} was found, all properties of the halo, including coordinates and velocities, were taken from snapshot n , for which z_n is closest to z_{cos} . To remove the galaxy groups and reduce the computational overhead, we applied a mass selection of $M_{500c} > 10^{13} M_{\odot}$ ⁵, where M_{500c} is the mass enclosed within a region of density 500 times the critical density of the Universe at the cluster redshift.

Once z_{cos} was determined, it was combined with the Doppler redshift (due to the peculiar velocity), z_{pec} , to compute the observed redshift, z_{obs} :

$$1 + z_{\text{obs}} = (1 + z_{\text{cos}})(1 + z_{\text{pec}}). \quad (2)$$

In our case, z_{obs} is equivalent to the CMB rest-frame redshift in observations, affected only by cosmic expansion and the peculiar

velocity of the source, and not by any peculiar motion of the observer within the cosmic rest frame. The relativistic Doppler effect was computed as

$$1 + z_{\text{pec}} = \gamma[1 + v_{\text{pec}} \cos \theta/c], \quad (3)$$

where θ is the angle between the cluster's peculiar motion and its line-of-sight to the observer, v_{pec} is the total peculiar velocity, $\gamma = [1 - (v_{\text{pec}}/c)^2]^{-1/2}$, and c is the speed of light. The observed redshift is ultimately used to model bulk flow and H_0 variation.

We used the SOAP properties computed at time steps $\Delta z = 0.10$ and did not perform any interpolation between snapshots. The general accuracy of phase-space interpolation of cluster velocities and positions in N -body simulations remains to be determined. The decision not to interpolate was previously adopted for galaxy lightcones (e.g. Kitzbichler & White 2007), where the dynamical time on $\sim 10 \text{ kpc}$ scales is much shorter than the time resolution of most simulations. At low redshift, the dynamical time scale for galaxy clusters is of the same order as the SOAP time step⁶. Whether interpolation would improve accuracy or introduce spurious structures for clusters is unclear. By not interpolating, our method preserves the local hydrodynamical structure at scales $\leq 400 \text{ Mpc}$ ⁷. Moreover, interpolation requires identifying the main branch of cluster evolution between snapshots, which in turn requires accessing the merger tree history. This may introduce additional uncertainties and pathological behaviours (see discussion in Smith et al. 2017). The physical properties are only offset by $\Delta z < 0.05$, which is well within acceptable limits for our purposes.

2.2. Sample selection and outlier removal

In each lightcone, we selected clusters based on their X-ray flux $f_{X,0.1-2.4 \text{ keV}} > 5 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the $0.1-2.4 \text{ keV}$ band (see Sect. 2.3.2 for details on the flux computation). A random Galactic plane was chosen for each lightcone, coordinates were transformed accordingly, and Galactic latitudes $|b| < 20^\circ$ were masked out to mimic the statistical behaviour of a Zone of Avoidance. To ensure no duplicated clusters, we then used the merger tree information to remove all duplicates⁸. After flux and latitude selection and duplicate removal, approximately 700–800 clusters per lightcone remained, whereas M21 contains ~ 350 . We note again that the reason for adopting the same selection process of M21 is to isolate its impact on the inferred dipole and does not reduce the predictive power of the simulation.

To understand why this discrepancy arises, we compared the distribution of projected X-ray concentration (c_X) between the selected simulated clusters and the sample used in M21. c_X is defined as $L_{\text{core}}/L_{\text{tot}}$ where L_{core} and L_{tot} are the X-ray luminosities within $< 0.15 R_{500c}$ and $< R_{500c}$ respectively. We find that FLAMINGO clusters are less concentrated than the observed ones, where a significant portion of clusters with $c_X > 0.5$ X-ray concentration is missing in FLAMINGO. The cause of this discrepancy is unclear, though it is diminished by accounting for projection effects. The issue is potentially related to the fact that X-ray surveys tend to pick up the most concentrated cool-core clusters due to their high X-ray emission coming from their

⁶ For an $R \sim 1 \text{ Mpc}$, $M \sim 10^{14} M_{\odot}$ cluster, the dynamical time scale $\delta t \sim \sqrt{R^3/GM} \sim 10^9 \text{ yr}$. The corresponding redshift change at low redshift is $\delta z = -H(z)/a(z) \delta t \sim -H_0 \delta t \sim 0.1$.

⁷ $\Delta z = 0.10$ corresponds to a comoving distance of 420 Mpc .

⁸ We do this by fetching all TopLeafIDs (the ID of the earliest progenitor in the main branch). If two clusters share the same TopLeafID, we retain only the one at lower redshift. Duplicates are very rare since no box repetition is used.

⁴ <https://github.com/SWIFTSIM/SOAP>

⁵ Clusters or galaxy groups with $M_{500c} < 10^{13} M_{\odot}$ rarely pass our flux selection (see Appendix B).

Table 1. Constants used for the scaling relation fitting.

Relation	N	C_X	C_Y	γ	Configuration
L_X-T	313	4.2 keV	$10^{44} \text{ erg s}^{-1}$	-1	M21
	400	3.8 keV	$10^{44} \text{ erg s}^{-1}$	-1	MCMC
$Y_{\text{SZ}}-T$	260	3.7 keV	20 kpc^2	1	M21
	400	3.2 keV	20 kpc^2	1	MCMC
$M_{\text{gas}}-T$	300	4.2 keV	$3 \times 10^{13} M_{\odot}$	1	M21
	400	3.6 keV	$3 \times 10^{13} M_{\odot}$	1	MCMC

Notes. For each relation, we list the number of selected clusters N , the calibration constants C_X and C_Y , and the fixed redshift evolution index γ . Two configurations are shown: one matching the M21 observational setup using *Chandra*-like temperatures T_{Cl} , and one using the original SOAP T_{sl} values for the MCMC analysis.

dense cores. To create the most realistic catalogues and match statistics to M21, we choose to sort the clusters by X-ray concentration and consider only the first ~ 300 – 400 clusters for our analysis. The exact sample size depends on the relation and statistical methods used (Table 1). This c_X -based selection turns out to have no considerable effect on our final results as we show in Sect. 5.4. (For further discussion on the c_X discrepancy see Appendix D.1.) The c_X distribution and the differences therein are shown in Fig. D.1.

Strong scaling relation outliers can bias the obtained cosmological constraints, especially when small cluster samples are analysed in specific sky regions. Such outliers are absent in the M21 sample, as a sample cleaning was performed prior to the analysis (removing AGN-dominated clusters and highly disturbed mergers). To avoid this potential bias in our work, we consider the full-sky scaling relation fit and perform a $4\sigma_{\text{intr}}$ clipping from the best-fit scaling relation (see Sect. 3 for our power law fitting method) to remove outliers, where σ_{intr} is the fitted intrinsic scatter of each relation. For samples of $\lesssim 400$ clusters, it is unlikely that any object would scatter beyond the 4σ threshold by chance; thus, the removed clusters are likely to be peculiar cases not expected to follow universal scaling relations due to their physical characteristics. Outliers are removed for all three relations of interest, L_X-T , $Y_{\text{SZ}}-T$, and $M_{\text{gas}}-T$. Typically, the number of outliers per lightcone is <10 , i.e. $<3\%$ of the sample.

2.3. Simulation samples and cluster properties

After lightcone creation and flux-based sample selection, a total of 1728 cluster catalogues were obtained, each containing 700–800 clusters. Before evaluating their anisotropies, the robustness of the simulated samples must first be established. This section presents and compares our lightcone catalogues to the M20 sample and the FLAMINGO L1_m9 runtime lightcones. The runtime lightcones provide a more accurate account of these quantities by yielding more precise redshifts and extracting properties from snapshots closer to the lightcone crossing redshift. (For more details on the comparison of the samples, see Appendix D and the figures therein.)

2.3.1. Redshift distribution and sky coordinates

Generally, the properties of the simulated lightcone samples are in good agreement with the M21 sample. The average redshift distribution of the 1728 simulated samples is shown in Fig. 1

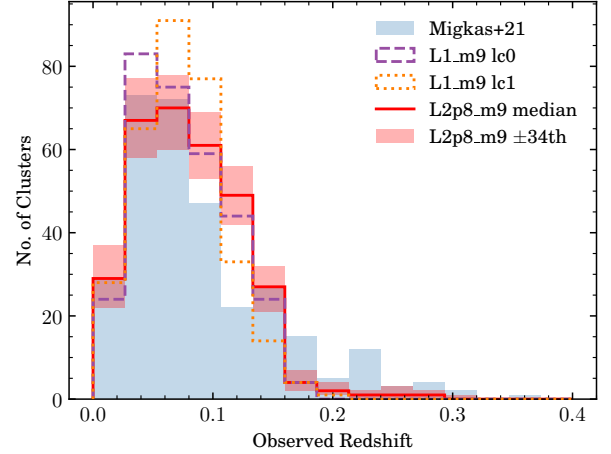


Fig. 1. Average cluster redshift distribution of the $1728 (2.8 \text{ Gpc})^3$ lightcones (red) compared with two runtime $(1.0 \text{ Gpc})^3$ lightcones (purple, orange) and the M21 sample (blue). The most X-ray concentrated clusters from the simulation were selected to match the M21 sample size ($N = 313$).

and compared with the M21 sample. The distribution extends up to $z \sim 0.30$ with a median of $z \sim 0.07$. Our simulated samples rarely extend above $z \sim 0.15$, while the M21 observed clusters exhibit a more pronounced tail at higher redshift. However, in both cases, clusters at $z \lesssim 0.15$ dominate the samples and thus primarily determine the probed scaling relation behaviour.

2.3.2. X-ray luminosity and flux

In SOAP, L_X is computed using the photoionisation code CLOUDY (Ferland et al. 2017), following the method of Ploeckinger & Schaye (2020). The implementation is described in detail in Braspenning et al. (2024). The final cluster L_X is calculated as the sum of the X-ray luminosities of all particles within R_{500c} . SOAP provides L_X values for three bands: 0.5–2 keV, 0.2–2.3 keV, and 2.3–8 keV.

To select samples consistent with M20 and M21, which used fluxes in the 0.1–2.4 keV band, we started from the SOAP-provided L_X in the 0.2–2.3 keV band and converted it to the 0.1–2.4 keV band using an unabsorbed apec model in XSPEC (Arnaud et al. 1999) and the cluster’s temperature (see Sect. 2.3.5). The observed flux was then computed from the converted L_X , using z_{obs} to account for the K-correction. This flux was used for sample selection. Although the flux-selected number of clusters remains higher than in M21, by this process we ensured that this discrepancy was not caused by incorrect flux estimation or the bandpass conversion.

However, for our analysis of the L_X-T relation, we used the original SOAP L_X values in the 0.2–2.3 keV band. While not strictly identical to the 0.1–2.4 keV band used in M21, the difference in L_X has a negligible impact on the angular variation of L_X-T , which is the focus of our study. Reprocessing the full pipeline using converted luminosities would introduce additional uncertainties and reduce reproducibility, without materially changing the anisotropy analysis. This choice avoids introducing additional uncertainties from model-based conversions and ensures that our results remain fully reproducible using the native SOAP outputs. The resulting f_X and L_X distributions of our simulated lightcone samples are in good agreement with both the L1_m9 runtime lightcones and the observational data

from M21 (see Appendix D.2 and figures therein). The typical⁹ L_X distribution spans $L_X = (6 \times 10^{41} - 2 \times 10^{45}) \text{ erg s}^{-1}$, with a median of $10^{44} \text{ erg s}^{-1}$.

2.3.3. Total integrated Compton Y_{SZ} parameter

The Y_{SZ} parameter was calculated by summing over the contribution of each gas particle inside the $5 \times R_{500c}$ aperture (Schaye et al. 2023):

$$Y_{SZ}[\text{kpc}^2] = \frac{\sigma_T}{m_e c^2} k_B \sum_i n_{e,i} T_{e,i} \frac{m_i}{\rho_i}, \quad (4)$$

where σ_T is the Thomson cross section, c is the speed of light, m_e is the electron mass, k_B is the Boltzmann constant, m_i and ρ_i are the mass and density of particle i , and $n_{e,i}$ and $T_{e,i}$ are its electron number density and temperature. This numerical implementation corresponds to the Y_{SZ} definition used in M21, where the integral is performed along the line of sight and over the area¹⁰.

In M21, 260 clusters with $Y_{SZ} \text{ S/N} > 2$ were selected for analysis. For comparison, we used the same number and selected the 260 highest X-ray concentration clusters. Similar to T , the simulated Y_{SZ} distribution lies also on the lower end compared to the M21 data. The typical distribution of simulated lightcones spans $Y_{SZ} = (0.2 - 484.8) \text{ kpc}^2$, with a median value of 20.8 kpc^2 , whereas the M21 sample has $Y_{SZ} = (1.3 - 891.6) \text{ kpc}^2$ and the median is 33.7 kpc^2 .

2.3.4. Gas mass

The M_{gas} of a cluster is defined in SOAP as the sum of the masses of all gas particles within R_{500c} . The typical distribution of M_{gas} in the lightcones has a range of $M_{\text{gas}} = (5.1 \times 10^{11} - 2.3 \times 10^{14}) M_\odot$ with a median of $2.5 \times 10^{13} M_\odot$. M_{gas} has not previously been used in any anisotropy analyses, but such work is forthcoming (Migkas et al., in prep.). In this study, the $M_{\text{gas}} - T$ relation is employed to probe cosmic anisotropy for the first time, providing a reference point for future investigations. We included all gas particles in the computation of M_{gas} to avoid any bias, even though cold gas is invisible in X-ray observations. However, we have verified that the difference introduced by this choice is $\lesssim 1\%$.

2.3.5. Temperature

The spectroscopic-like temperature was calculated following Mazzotta et al. (2004) as

$$T_{\text{sl}} = \frac{\sum_i \rho_i m_i T_i^{1/4}}{\sum_i \rho_i m_i T_i^{-3/4}}, \quad (5)$$

where ρ_i , m_i , and T_i are the density, mass, and temperature of particle i . We use T_{sl} calculated within a core-excised aperture of $0.15 R_{500c} < r < R_{500c}$. This was done to avoid the bias from cool-core clusters and to remain consistent with M21, which also excluded the cluster core using the $0.2 - 0.5 R_{500}$ annulus. Only

⁹ By ‘‘typical’’ we refer to the median of all distributions. The three reported numbers here and below are the medians of the minimum, maximum, and median values of the 1728 lightcones.

¹⁰ In M21, $Y_{SZ} = D_A^2 \int_{\text{l.o.s.}} d\Omega \int_{\text{l.o.s.}} dl \frac{\sigma_T}{m_e c^2} k_B n_e T_e$, where ‘‘l.o.s.’’ denotes the line of sight towards the cluster. The two definitions are equivalent. We neglect non-cluster line-of-sight contributions.

hot gas is visible in X-ray, and thus only particles with $T > 10^5 \text{ K}$ are summed over¹¹.

It is well established that a systematic difference exists between the *XMM-Newton* and *Chandra* X-ray telescopes in measuring T for the same clusters (Schellenberger et al. 2015; Migkas et al. 2024). We find that the slope values of scaling relations using SOAP T_{sl} closely resemble results from previous observational studies that used *XMM-Newton* temperatures. Thus, we consider T_{sl} to be equivalent to *XMM-Newton* T . However, M21 used *Chandra* T . To better match the scaling relation behaviour in M21, which is crucial for comparison, we converted T_{sl} to a *Chandra*-like temperature using $T_{\text{Cl}} = T_{\text{sl}}^{1/0.89}$, following Schellenberger et al. (2015). After this conversion, the M21 scaling relation behaviour is well replicated (see Sect. 4.1).

The above conversion was applied only when using the M21 statistical methodology (Sect. 3.1), in order to enable direct comparison with the M21 anisotropy results. When the alternative statistical procedure (MCMC) was used to explore anisotropies in FLAMINGO (Sect. 3.2), the default T_{sl} was retained.

The typical *XMM-Newton*-like T_{sl} distribution in our lightcones spans $(0.6 - 10.7) \text{ keV}$ with a median of 3.1 keV . M21’s 313 clusters have a higher median temperature of 4.5 keV . After converting T_{sl} to a *Chandra*-equivalent value as discussed above, the temperature difference was significantly reduced. The resulting distribution spans $T_{\text{Cl}} = (0.57 - 14.34) \text{ keV}$, with a median of 3.58 keV .

The remaining difference between the M21 and FLAMINGO temperatures could arise from several factors, including the different apertures used for temperature measurement, as well as systematic biases in temperature calibration. The fiducial selection of hot gas by $T > 10^5 \text{ K}$ might also contribute if the cut is set too high. Moreover, the under-representation of high-redshift clusters in our flux-limited, simulated samples, as previously discussed, also affects intrinsic cluster properties such as T , L_X , Y_{SZ} , M_{500c} , and M_{gas} , lowering their average values. We have verified that our high c_X selection (Sect. 2.2) does not cause the temperature offset, by examining the T distribution of the lowest- c_X clusters.

2.3.6. Velocities

The three-dimensional velocity \mathbf{v} of each cluster was taken to be the centre-of-mass velocity computed by SOAP, using all particles within R_{500c} . The typical peculiar velocity spans $|\mathbf{v}| = (41 - 1436) \text{ km s}^{-1}$, with a median of 452 km s^{-1} . The true bulk flow, in contrast to the values inferred from scaling relations, was estimated as the mean peculiar velocity of all clusters within a given redshift shell. The bulk flow profile as a function of redshift is presented in Sect. 4.3.

2.4. Matched scatter with M21

Our anisotropy analysis is highly sensitive to the scatter, as the inferred amplitude of a spurious cosmic anisotropy increases with larger scatter in the scaling relations (see Sect. 5.1). To minimise bias from a mismatch between the scatter in FLAMINGO and M21, we generated a complete set of separate lightcone catalogues with scaling relation scatter matched to that of M21.

Two sets of tests are presented in parallel throughout this work. The first is the default run, in which all values follow

¹¹ In contrast to M_{gas} , this selection is essential, as including all gas particles leads to T_{sl} values that are lower by 2–3 orders of magnitude, due to the strong weighting of cold, dense gas.

Table 2. Best-fit parameters of the scaling relations using the full lightcone sample.

Relation ($Y-X$)	A_{YX}	B_{YX}	σ_{intr}	$A_{YX}^{(s)}$	$B_{YX}^{(s)}$	$\sigma_{\text{intr}}^{(s)}$	$\sigma_{\text{tot}}^{(s)}$	Configuration
L_X-T	$1.770^{+0.041}_{-0.032}$	$2.220^{+0.070}_{-0.078}$	$0.164^{+0.012}_{-0.009}$	$1.762^{+0.049}_{-0.040}$	$2.215^{+0.081}_{-0.084}$	$0.240^{+0.018}_{-0.018}$	$0.251^{+0.018}_{-0.016}$	M21
	$1.754^{+0.016}_{-0.024}$	$2.500^{+0.063}_{-0.069}$	$0.154^{+0.012}_{-0.004}$	$1.738^{+0.032}_{-0.040}$	$2.506^{+0.069}_{-0.087}$	$0.228^{+0.012}_{-0.018}$	$0.243^{+0.012}_{-0.015}$	MCMC
$Y_{\text{SZ}}-T$	$1.094^{+0.015}_{-0.020}$	$2.441^{+0.030}_{-0.033}$	$0.110^{+0.006}_{-0.006}$	$1.081^{+0.030}_{-0.025}$	$2.448^{+0.051}_{-0.054}$	$0.156^{+0.012}_{-0.012}$	$0.189^{+0.010}_{-0.009}$	M21
	$1.102^{+0.015}_{-0.015}$	$2.748^{+0.032}_{-0.029}$	$0.108^{+0.004}_{-0.004}$	$1.096^{+0.020}_{-0.020}$	$2.763^{+0.048}_{-0.048}$	$0.156^{+0.012}_{-0.006}$	$0.195^{+0.009}_{-0.009}$	MCMC
$M_{\text{gas}}-T$	$1.079^{+0.010}_{-0.010}$	$1.827^{+0.039}_{-0.039}$	$0.080^{+0.006}_{-0.006}$	–	–	–	–	M21
	$1.086^{+0.010}_{-0.005}$	$2.050^{+0.039}_{-0.039}$	$0.078^{+0.004}_{-0.008}$	–	–	–	–	MCMC

Notes. All fits are performed using the reduced χ^2 regression method (Sect. 3.1.1). The reported uncertainties in A , B , and σ_{intr} correspond to the 16th and 84th percentiles across all lightcones. Superscript (s) denotes fits after injecting scatter to match the M21 sample (Sect. 2.4). The ‘Configuration’ column indicates the number of clusters and fitting constants used, as defined in Table 1.

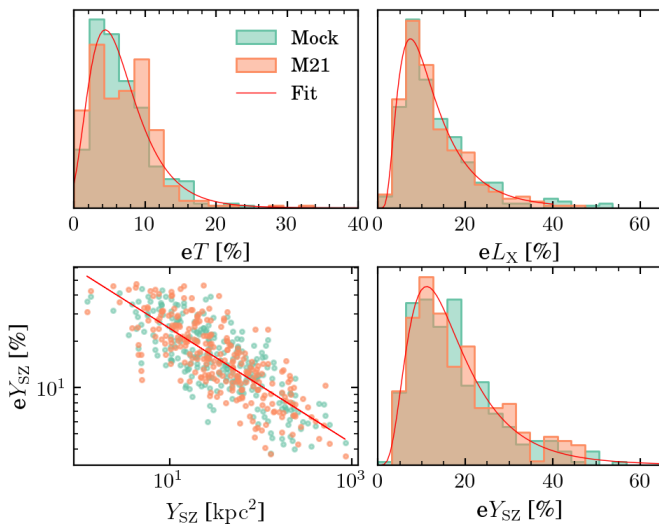


Fig. 2. Mock instrumental uncertainties for one lightcone in T (top left), L_X (top right), and Y_{SZ} (bottom right) compared with the M21 sample. The bottom-left panel shows the correlation between Y_{SZ} and its uncertainty eY_{SZ} . In each panel, the mock uncertainty (green) represents one realisation drawn from the fitted distribution (red).

the original simulation prescriptions. The second includes added scatter, matched to the uncertainties in the M21 catalogue. This matched scatter test was applied only to the L_X-T and $Y_{\text{SZ}}-T$ relations and not to $M_{\text{gas}}-T$; in other words, scatter was injected into L_X , Y_{SZ} , and T . The same uncertainty generation procedure is used for both $T_{\text{Chandra-like}}$ and T_{sl} .

Scatter was added using a two-step procedure. First, we amplified the deviation of the cluster property Y (L_X and Y_{SZ}) from the best-fit power law:

$$\log Y'_s = [\log Y' - \log Y'_{\text{th}}(X')] \frac{\sigma_{\text{intr},\text{M21}}}{\sigma_{\text{intr},\text{FL}}} + \log Y'_{\text{th}}(X'), \quad (6)$$

where $\log Y'_{\text{th}}(X') = \log A_{YX} + B_{YX} \log X'$ denotes the logarithmic, normalised Y from the fitted power law in terms of property X . The subscript s marks quantities with injected scatter. Here, $\sigma_{\text{intr},\text{FL}}$ denotes the median intrinsic scatter across FLAMINGO lightcones (Table 2), while $\sigma_{\text{intr},\text{M21}}$ refers to the best-fit intrinsic scatter reported in M21. This amplification step matches the intrinsic scatter to the observed level while preserving the physical correlations introduced by the FLAMINGO baryonic physics model.

Second, we assigned a percentage-based instrumental uncertainty drawn from the M21 distribution. For L_X and T , where no strong correlation is seen between the uncertainty and the property. We fitted a log-normal distribution to the M21 uncertainties and sampled from it. For Y_{SZ} , where a clear correlation is present, the dependence was explicitly modelled in the sampling. The measured value was then perturbed using the drawn uncertainty. The resulting uncertainty distributions and one example lightcone are shown in Fig. 2. The log-normal fit accurately captures the observed behaviour. We note that the uncertainties were generated as percentages of the underlying property values. Thus the lower absolute values of T_{sl} do not lead to systematically higher relative uncertainties compared to T_{Cl} .

3. Method

We fitted a cluster scaling relation between Y and X with the following power law form:

$$\frac{Y}{C_Y} E(z)^{\gamma_{YX}} = A_{YX} \left(\frac{X}{C_X} \right)^{B_{YX}}, \quad (7)$$

where C_X and C_Y are predefined pivot points for the quantities X and Y , A_{YX} is the normalisation of the relation, and B_{YX} is the slope. The constants C_X and C_Y were chosen to minimise the covariance between the fitted slope B_{YX} and normalisation A_{YX} in log space. Specifically, we set C_X and C_Y to the median values of X and Y . The term $E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$ is the evolution factor, that, together with the power index γ_{YX} , accounts for the redshift evolution of the scaling relation. The value of γ_{YX} is fixed to the self-similar value (e.g. Giodini et al. 2013). The precise value of γ_{YX} is unimportant for our direction-dependent comparisons, given the low- z nature of our sample and the nearly uniform redshift distribution across the sky. The constants used are summarised in Table 1. As in M21, we focus on relative angular variations and not absolute parameters, so selection effects can be safely ignored under the assumption of uniform sky coverage and complete detection above the flux threshold.

3.1. H_0 anisotropy: The M21 method

The first approach follows M21 and uses a χ^2 minimisation procedure to fit H_0 in different sky patches. The variation in the best-fit scaling relation normalisation is then translated into an apparent H_0 variation.

3.1.1. χ^2 minimisation

Consider a scaling relation $Y-X$ for a cluster sample, with each cluster labelled by an index i and described by properties Y_i and X_i . We fitted the parameters A_{YX} , B_{YX} , and the intrinsic scatter σ_{intr}^2 in logarithmic space by minimising the following expression:

$$\chi_Y^2 = \sum_{i=1}^N \frac{(\log Y'_i(z_{\text{obs},i}) - \log A_{YX} - B_{YX} \log X'_i)^2}{\sigma_{\text{intr}}^2 + \sigma_{\log Y_i}^2 + B^2 \times \sigma_{\log X_i}^2}, \quad (8)$$

where $z_{\text{obs},i}$ is the observed redshift in the cosmic rest frame (Eq. (2)), $Y'_i = E(z_{\text{obs},i})^{YX} Y_i / C_Y$, $X'_i = X_i / C_X$, and N is the number of clusters. The terms $\sigma_{\log Y_i}$ and $\sigma_{\log X_i}$ denote the mock instrumental uncertainties in logarithmic space¹². Following Maughan (2007), the intrinsic scatter $\sigma_{\text{intr},YX}^2$ is iteratively increased until the reduced chi-squared reaches unity, $\chi_{\text{res}}^2 \sim 1$ ¹³. We always take Y to be the quantity with stronger cosmology dependence, and only squared deviations along the Y -axis are minimised.

3.1.2. Sky scanning, uncertainties, and H_0 variation

Cluster scaling relations were fitted within a cone of opening angle $\theta = 60^\circ$ (for $Y_{\text{SZ}}-T$) or 75° (for L_X-T and $M_{\text{gas}}-T$), and the resulting best-fit parameter was assigned to the cone centre $\hat{n}(l, b)$ to construct a sky map $p(\hat{n})$. The cone was scanned across the sky in steps of 4° in longitude and 2° in latitude. To enhance the contribution of clusters closer to the cone centre relative to those near its edge, each cluster was weighted by $\cos(\alpha_i)$, where α_i is its angular separation from the cone centre, following the method of M21. The cone size θ was chosen to balance the need for sufficient clusters in each patch and the desire for directional resolution. After the best-fit parameters are determined in all directions, uncertainties are estimated using bootstrap resampling with 500 realisations. The 1σ uncertainty σ_p was taken as the ± 34 th percentile around the best-fitting parameter p .

An anisotropy in H_0 affects all redshifts equally and does not alter the slope B of the scaling relation. For relations between a cosmology-dependent quantity (such as L_X , Y_{SZ} , or M_{gas}) and a cosmology-independent one (such as T), the variation in the normalisation $A(\hat{n})$ directly reflects directional variation in $H_0(\hat{n})$.

Taking L_X-T as an example, the luminosity is defined by $L_X = 4\pi D_L(z)^2 E(z)^\gamma f_X$, where the luminosity distance is $D_L(z) = (1+z) \int_0^z c dz' / [H_0 E(z')]$. On the other hand, the scaling relation implies that for a given T , the expected value is $L_X = A_{LT} T^{B_{LT}}$, which is cosmology-independent. Ignoring normalisation constants C_L and C_T for clarity, equating the two gives

$$L_X = 4\pi D_L(z)^2 f_X E(z)^\gamma = A_{LT} T^{B_{LT}}. \quad (9)$$

Thus,

$$A_{LT} \propto D_L(z)^2 \propto H_0^{-2}, \quad (10)$$

meaning any overestimation of L_X for a fixed T (reflected in a higher A_{LT}) implies an underestimated H_0 , meaning the true H_0 is higher.

¹² We present results both with and without mock instrumental uncertainties. When uncertainties are omitted, $\sigma_{\log Y_i} = \sigma_{\log X_i} = 0$ in Eqs. (8) and (16), and the intrinsic scatter equals the total scatter. The total scatter is defined as $\sigma_{\text{tot}} = \sqrt{\sigma_{\text{intr}}^2 + \sigma_{\log Y_i}^2 + B^2 \sigma_{\log X_i}^2}$.

¹³ The reduced chi-squared is defined as $\chi_{\text{res}}^2 \equiv \chi_Y^2 / (N - \text{d.o.f.})$, with $\text{d.o.f.} = 3$ corresponding to the free parameters A_{YX} , B_{YX} , and σ_{intr} .

The same argument applies to the $Y_{\text{SZ}}-T$ and $M_{\text{gas}}-T$ relations. Since $Y_{\text{SZ}} \propto D_A^2(z)$ (e.g. Giardini et al. 2013) and $M_{\text{gas}} \propto D_A^{5/2}(z)$ (e.g. Sasaki 1996), with $D_A(z) = (1+z)^{-2} D_L(z)$, we obtain:

$$A_{YT} \propto H_0^{-2}, \quad (11)$$

$$A_{MT} \propto H_0^{-5/2}. \quad (12)$$

The directional variation of H_0 can therefore be inferred from the normalisation map of any of the three scaling relations:

$$\frac{H'_0(\hat{n})}{H_0} = \left(\frac{A_{YX}(\hat{n})}{A_{YX,\text{all}}} \right)^\alpha, \quad (13)$$

where $\alpha = 1/2$, $1/2$, and $2/5$ for the L_X-T , $Y_{\text{SZ}}-T$, and $M_{\text{gas}}-T$ relations, respectively. Here, $H'_0(\hat{n})$ denotes the direction-dependent Hubble constant, H_0 the isotropic value, $A_{YX}(\hat{n})$ the local normalisation, and $A_{YX,\text{all}}$ the best-fit normalisation from all clusters in the same lightcone. We emphasise that we do not constrain the absolute value of H_0 from cluster scaling relations, but only its relative variation across the sky.

3.1.3. Dipole anisotropy statistical significance

To assess the statistical significance of a dipole anisotropy, we analysed the sky map of the scaling relation normalisation $A(\hat{n})$, along with its bootstrapped uncertainty $\sigma_A(\hat{n})$. The significance of a dipole pointing in a given direction \hat{n} was quantified as the difference in $A(\hat{n})$ and its value in the antipodal direction $-\hat{n}$, in units of the combined uncertainty:

$$n_\sigma(\hat{n}) = \frac{A(\hat{n}) - A(-\hat{n})}{\sqrt{\sigma_A^2(\hat{n}) + \sigma_A^2(-\hat{n})}}. \quad (14)$$

By definition, $n_\sigma(\hat{n}) = -n_\sigma(-\hat{n})$. A negative value indicates that A is lower in the direction \hat{n} than in its opposite region. The amplitude of the H_0 dipole for each lightcone is then defined as the maximum directional difference in H'_0/H_0 , i.e.

$$\Delta H_0[\%] = [H'_0(\hat{n}) - H'_0(-\hat{n})] / H_0, \quad (15)$$

where $\hat{n} = \arg \max n_\sigma(\hat{n})$ is the direction of maximal significance. Our approach differs slightly from that of M21, which compared each region to the rest of the sky rather than to its antipode. By focusing on direct dipole-level comparisons, our method is less prone to underestimating the amplitude of H_0 variation.

The statistical significance of the observed dipole was not taken from the maximum value of $n_\sigma(\hat{n})$ directly, as this would be biased by the look elsewhere effect (selecting the most extreme fluctuation across many directions). Instead, the significance is quantified by comparing against the FLAMINGO lightcones: we compute the fraction of them that yield a higher maximum $n_\sigma(\hat{n})$ (and corresponding ΔH_0) than the observed value. The statistical method for this comparison is described in Sect. 3.4. This defines a robust p -value that captures the probability of such an anisotropy arising in a statistically isotropic universe. The procedure mirrors that of M21, where mock catalogues were used to evaluate the significance of the detected dipole.

3.2. H_0 anisotropy: MCMC full likelihood analysis

The second approach uses a full-sky MCMC fit in which H_0 is allowed to vary within the parametrisation of each scaling relation. This serves as a cross-validation of the χ^2 sky-patch method

(the M21 method) described above. A dipole-like H_0 variation is modelled as $H'_0/H_0 = 1 + \delta \cos \alpha_i(\phi_H, \lambda_H)$, where δ is the relative amplitude of the anisotropy and α_i is the angular separation between the i -th cluster and the dipole direction (ϕ_H, λ_H) , given in simulated Galactic coordinates. A total of six parameters were constrained in the fit: δ , ϕ_H , λ_H , A_{YX} , B_{YX} , and σ_{intr} . We sampled the likelihood function

$$\ln \mathcal{L}_{YX} = -\frac{1}{2}\chi_Y^2 + \sum_i^N \ln(\sigma_{\text{intr}}^2 + \sigma_{\log Y_i}^2 + B^2 \times \sigma_{\log X_i}^2), \quad (16)$$

where χ_Y^2 is defined in Eq. (8), and N is the number of clusters in the fit.

For each parameter, we report the posterior median, with the 16th and 84th percentiles defining the uncertainty interval. The amplitude of the anisotropy is $\Delta H_0 = 2\delta$, to be consistent with the M21 method (Eq. (15)). The statistical significance n_σ of the anisotropy was estimated by dividing the median δ by its lower uncertainty. This reflects the number of standard deviations separating the result from isotropy ($\delta = 0$). As with the M21 method, the per-lightcone n_σ values were not used to infer significance directly, due to the look elsewhere effect (Sect. 3.1.3). Only the distribution of anisotropy amplitudes across all lightcones is compared with the M21 result to determine statistical relevance.

We used the EMCEE package developed by Foreman-Mackey et al. (2013) for MCMC sampling. The integrated autocorrelation time τ (Goodman & Weare 2010) was monitored to ensure convergence, and all chains were run to lengths exceeding 50 τ .

3.3. Modelling bulk flows

Peculiar velocities influence cluster scaling relations by altering the observed redshift z_{obs} away from the cosmological redshift z_{cos} . While random motions add to the intrinsic scatter without introducing directional effects, a coherent bulk flow can mimic an H_0 anisotropy in the scaling relations (Migkas et al. 2021; Migkas 2025). This degeneracy is strongest at low redshifts, where peculiar velocities are comparable in magnitude to the Hubble expansion. To isolate the effect of bulk flows, we assume an isotropic H_0 throughout the bulk flow analysis.

The bulk flow was parametrised by its amplitude, u_{bf} , and direction, $(l, b) = (\phi_{\text{bf}}, \lambda_{\text{bf}})$. For a given bulk flow, the observed redshift, z_{obs} , was corrected using (Dai et al. 2011)

$$z_{\text{cor}} = \frac{z_{\text{obs}} - z_{\text{bf}}}{1 + z_{\text{bf}}}, \quad (17)$$

where z_{cor} approximates the redshift in the absence of bulk motion. This equation follows from inverting Eq. (2) for z_{cos} . The bulk flow redshift shift, z_{bf} , was calculated from the non-relativistic Doppler formula, $z_{\text{bf}} = (u_{\text{bf}} \cos \alpha)/c$, where α is the angular separation between the bulk flow direction $(\phi_{\text{bf}}, \lambda_{\text{bf}})$ and the cluster sky position (ϕ, λ) . For bulk flow amplitudes $u_{\text{bf}} < 10^3 \text{ km s}^{-1}$, the accuracy¹⁴ of this non-relativistic treatment is sufficient.

We constrained the bulk flow parameters using MCMC to maximise the full likelihood function defined in Eq. (16). At each sampling step, z_{cor} was recalculated using Eq. (17), and the derived distances were used to modify the cosmology-dependent quantities: $L_X[D_L(z_{\text{cor}})]$, $Y_{\text{SZ}}[D_A(z_{\text{cor}})]$, and $M_{\text{gas}}[D_A(z_{\text{cor}})]$. The

¹⁴ Eq. (17) is not exact, as the cluster peculiar velocity includes both a coherent bulk flow and a residual component. Thus, $v_{\text{pec}} \neq u_{\text{bf}}$ and $z_{\text{pec}} \neq z_{\text{bf}}$, but the approximation remains accurate at the 10^{-5} level.

posterior median is quoted for each parameter, with the 16th and 84th percentiles giving the lower and upper uncertainty bounds.

3.4. Probability estimation

As mentioned in Sect. 3.1.3, the n_σ or ΔH_0 of a single lightcone (realisation) cannot be directly interpreted as the significance of a cosmic anisotropy due to the look-elsewhere effect. To assess whether the M21 result is in tension with Λ CDM, we instead consider the joint distribution of ΔH_0 and n_σ across all lightcones.

To estimate the probability of a certain outcome in the 2D parameter space $(\Delta H_0, n_\sigma)$, we used two complementary approaches: kernel density estimation (KDE) and an extreme value statistics (EVS)-based method (Davison & Smith 1990; Behrens et al. 2004; Hu et al. 2024b). KDE is more accurate when the target point is not extremely rare (e.g. $p \gtrsim 1/N$), whilst EVS is suitable to estimate much lower probability and can extrapolate to rare events. In EVS, we applied the Peak-Over-Threshold approach, relying on the theorem of Pickands (1975), along with Balkema & Haan (1974), which asserts that when the threshold is set high enough, the likelihood of surpassing it is governed by a generalized Pareto distribution (GPD). In our case of 1728 lightcones, a probability $\lesssim 1/1728 = 6 \times 10^{-4}$ will be better estimated by the EVS-POT method (hereafter EVS).

We applied KDE directly to the 2D distribution of ΔH_0 and n_σ of the detected anisotropies of all lightcones to estimate the rarity of the M21 findings within FLAMINGO¹⁵. The probability for a target point, P_0 , was calculated by integrating the probability mass above regions where the KDE value exceeds that at the target:

$$P_{\text{KDE},0} = 1 - \iint_{f>f_0} f(\Delta H_0, n_\sigma) d\Delta H_0 dn_\sigma. \quad (18)$$

Grid resolutions were tested to ensure numerical convergence. While effective near the main probability mass, KDE becomes unreliable in the sparsely populated tails, which is often our case, where the result depends heavily on grid resolutions and local distribution.

To address the limitations of KDE in the tails, we projected the 2D data onto 1D with some function $x = f(\Delta H_0, n_\sigma)$ and use the EVS approach to model the extreme values of x . We use a fiducial dimension reduction function, the lower uncertainty bound $x = \Delta H_0 - \Delta H_0/n_\sigma$. The GPD is fitted to exceedances $(x - u)$ above a high threshold u , and scaled to match the survival probability at the threshold $P(x = u)$. This ensures a seamless transition between the bulk and EVS in the tail. The probability of target point x_0 is then given by

$$P_{\text{EVS},0} = \begin{cases} P(x > x_0), & x_0 \leq u, \\ P(x > u)P_{\text{GPD}}(x > x_0|x > u), & x_0 > u, \end{cases} \quad (19)$$

where $P_{\text{GPD}}(x > x_0|x > u)$ is normalised as $P_{\text{GPD}}(x > x_0|x > u) = 1$, and $P(x > x_0)$ is the number of samples with $x > x_0$ divided by the total number of samples. The threshold is selected to give a good fit in the tail, usually the 95th percentile of x . The resulting estimation might depend on the choice of threshold, therefore we always vary the percentile to ensure convergence. Alternative dimension reduction functions were tested (see Appendix G) without a significant change

¹⁵ M21 used only n_σ to determine the p -value of the observed anisotropy, which can potentially underestimate the statistical significance of their findings.

in the conclusions. The projection and EVS combination was also applied for bulk flow analysis to assess the discrepancy between FLAMINGO and M21 bulk flows (see Sect. 4.3 for more details).

By combining KDE with the projected EVS method, we achieved a robust probability estimates across the entire range. This allows for a quantitative assessment of the rarity of the M21 anisotropy in FLAMINGO lightcones.

4. Results

4.1. General behaviour of scaling relations

Realistic scaling relations are a crucial aspect of our lightcone samples, they form the basis for evaluating anisotropy. The fitted parameters are summarised in Table 2. In this section, we compare our results with observations reported in the literature. We note that following the approach in M21, the default fitting parameters presented use the converted T_{Cl} instead of T_{sl} for all three relations (Sect. 2.3.5).

The L_X-T relation gives a slope of $B_{LT} = 2.220^{+0.070}_{-0.078}$, steeper than the simplest self-similar model ($B_{LT} = 1.5$ for the soft band L_X) and fully consistent with M20 (2.102 ± 0.064) and M21 ($2.086^{+0.073}_{-0.065}$). The slope is slightly shallower than studies using *XMM-Newton* temperatures (Chen et al. (e.g. 2007, 2.73 ± 0.13), Eckmiller et al. (2011, 2.52 ± 0.10), Lovisari et al. (2020, 3.110 ± 0.422)). Using the original SOAP T_{sl} yields $B_{LT} = 2.500^{+0.063}_{-0.069}$ that better matches the higher values obtained in the above studies. The matched scatter test returns a higher scatter of $\sigma_{\text{intr}}^{(s)} = 0.240^{+0.018}_{-0.018}$ ($\sigma_{\text{intr}}^{(s)} = 0.228^{+0.012}_{-0.018}$ for T_{sl}), reproducing the M21 scatter (0.233 ± 0.016) as well as independent observations (e.g. Pratt et al. (2009, 0.28), Eckmiller et al. (2011, 0.27), Lovisari et al. (2015, 0.26)) while keeping the normalisation and slope unchanged.

The $Y_{\text{SZ}}-T$ relation gives the slope $B_{YT} = 2.441^{+0.030}_{-0.033}$, consistent with M21 ($2.546^{+0.071}_{-0.067}$) and other *Chandra* temperature observations (e.g. Morandi et al. (2007, 2.64 ± 0.28)). The $Y_{\text{SZ}}-T$ slope of our lightcones is very close to the self-similar model prediction $B_{YT} = 2.5$. Using the original T_{sl} , the simulated $Y_{\text{SZ}}-T$ relation gives $B_{YT} = 2.748^{+0.032}_{-0.029}$ which is fully compatible with Planck Collaboration XI (2011, 2.82 ± 0.18). Similar to L_X-T , the simulated lightcones give 25% less scatter than observed in M21 (0.146 ± 0.013), with an intrinsic scatter of $\sigma_{\text{intr}} = 0.110^{+0.006}_{-0.006}$. With matched scatter, we obtain $\sigma_{\text{intr}}^{(s)} = 0.156^{+0.012}_{-0.012}$ for T_{Cl} and $0.156^{+0.012}_{-0.006}$ for T_{sl} which matches the intrinsic scatter of M21 and other observations (e.g. Morandi et al. (2007, 0.139), Planck Collaboration XI (2011, 0.14 ± 0.02)). The normalisation and slope remain the same.

For the $M_{\text{gas}}-T$ relation we obtain a slope of $B_{MT} = 1.827 \pm 0.039$, steeper than the self-similar model $B_{MT} = 1.5$. Using original SOAP T_{sl} , we find $B_{MT} = 2.050^{+0.039}_{-0.039}$, fully consistent with past *XMM-Newton* measurements (e.g. Arnaud et al. (2007, 2.10 ± 0.11), Croston et al. (2008, 1.986 ± 0.111), Zhang et al. (2008, 1.859 ± 0.187)). The intrinsic scatter for $M_{\text{gas}}-T$ in our lightcones is $\sigma_{\text{intr}} = 0.080^{+0.006}_{-0.006}$, making it the tightest relation of the three. As M21 does not measure M_{gas} , no matched scatter test is applied here.

As shown above, the slopes of our scaling relations are consistent with those of M21. This ensures that both the simulated and observed relations respond similarly to bulk flows or statistical noise. The matched scatter tests preserve the original slopes and normalisations, while successfully reproducing the intrinsic and total scatter levels reported in M21. This correction is essen-

tial for a direct and unbiased assessment of anisotropy amplitudes.

4.2. H_0 anisotropy

The anisotropy constraints from the FLAMINGO lightcones, in comparison with the M21 results, are shown in Figs. 3 and 4. Table 3 lists the p -values of the tension between the M21 detection and the FLAMINGO lightcones. The joint EVS distributions combining L_X-T and $Y_{\text{SZ}}-T$ are shown in Fig. 5, with per-relation analyses provided in Appendix G.

4.2.1. Constraints using L_X-T , $Y_{\text{SZ}}-T$, or $M_{\text{gas}}-T$

For the L_X-T relation, the M21 method yields a moderate ΔH_0 peak at 4.3% and a median dipole significance near 1.7σ (Fig. 3a). The observed M21 square lies above the main simulated region, with $p = 0.021$ (KDE) and $p = 0.025$ (2.25σ , EVS), suggesting a mild tension. However, once realistic scatter is added (Fig. 3c), the distribution shifts upward to $\sim 8\%$, overlapping the M21 point. This is reflected in the p -values, which rise to $p = 0.33$ (KDE) and $p = 0.23$ (1.20σ), indicating that the L_X-T dipole is consistent with Λ CDM under observational scatter in the L_X-T relation.

The MCMC analyses of L_X-T in our lightcones produce a more isotropic distribution (Fig. 3b), peaking at $\Delta H_0 \sim 3.0\%$ and $n_\sigma \sim 1.4$. The M21 circle lies moderately above this region, with $p = 0.055$ (KDE) and $p = 0.032$ (2.14σ , EVS). After adding scatter (Fig. 3d), the lightcone distribution shifts to $\sim 4\%$, but the p -values remain low: $p < 5.8 \times 10^{-4}$ (KDE) and $p = 0.085$ (1.72σ , EVS). In this case, scatter introduces more high ΔH_0 tails, but the M21 result stays just outside the bulk. The mild increase in tension compared to the M21 method may reflect the broader likelihood sensitivity of the MCMC approach (see Sect. 5.5 for a comparison between the two methods). The agreement with isotropy in L_X-T is expected, as M21 also found only a mild tension (2.4σ) in this relation.

The $Y_{\text{SZ}}-T$ relation, by contrast, exhibits a stronger anisotropic signal in all cases. Without added scatter, the M21 square lies far outside the simulated region, with the KDE estimation bounded only from above, $p < 5.8 \times 10^{-4}$ (KDE)¹⁶ and $p = 1.8 \times 10^{-10}$ (6.38σ , EVS), both are extremely low. The lightcone distribution itself peaks at $\Delta H_0 \sim 4.0\%$, significantly below the M21 value of 14%, as shown in Fig. 3a. With added scatter, the lightcones shift up to $\sim 8\%$ (Fig. 3c), and the tension is notably reduced: $p = 0.0013$ (3.22σ) in EVS and $p < 5.8 \times 10^{-4}$ with KDE. The KDE result remains disproportionately low due to local sparsity in that region, despite the smoothed contours in Fig. 3c visually resembling a 3σ level. This highlights a limitation of KDE. Its sensitivity to the density of neighbouring points makes it less reliable when evaluating tail probabilities. In this case, the projected EVS method provides the more robust interpretation.

The MCMC results for $Y_{\text{SZ}}-T$ are similar in trend. Without added scatter, the lightcones peak at $\Delta H_0 \sim 2.2\%$ (Fig. 3b), and the M21 circle is well beyond the contours: $p = 2.8 \times 10^{-4}$ (3.64σ). After scatter matching (Fig. 3d), the peak shifts to $\sim 3\%$, and the M21 circle sits closer to the lightcone bulk. The p -values reduce to $p = 0.0049$ (2.82σ , EVS) and $p = 0.0014$ (KDE).

¹⁶ The M21 point lies too far from the main probability mass, making KDE unreliable. We can place only an upper bound of $p < 5.8 \times 10^{-4}$ corresponding to the fraction $1/1728$, as none of the FLAMINGO lightcones exhibits a higher anisotropy.

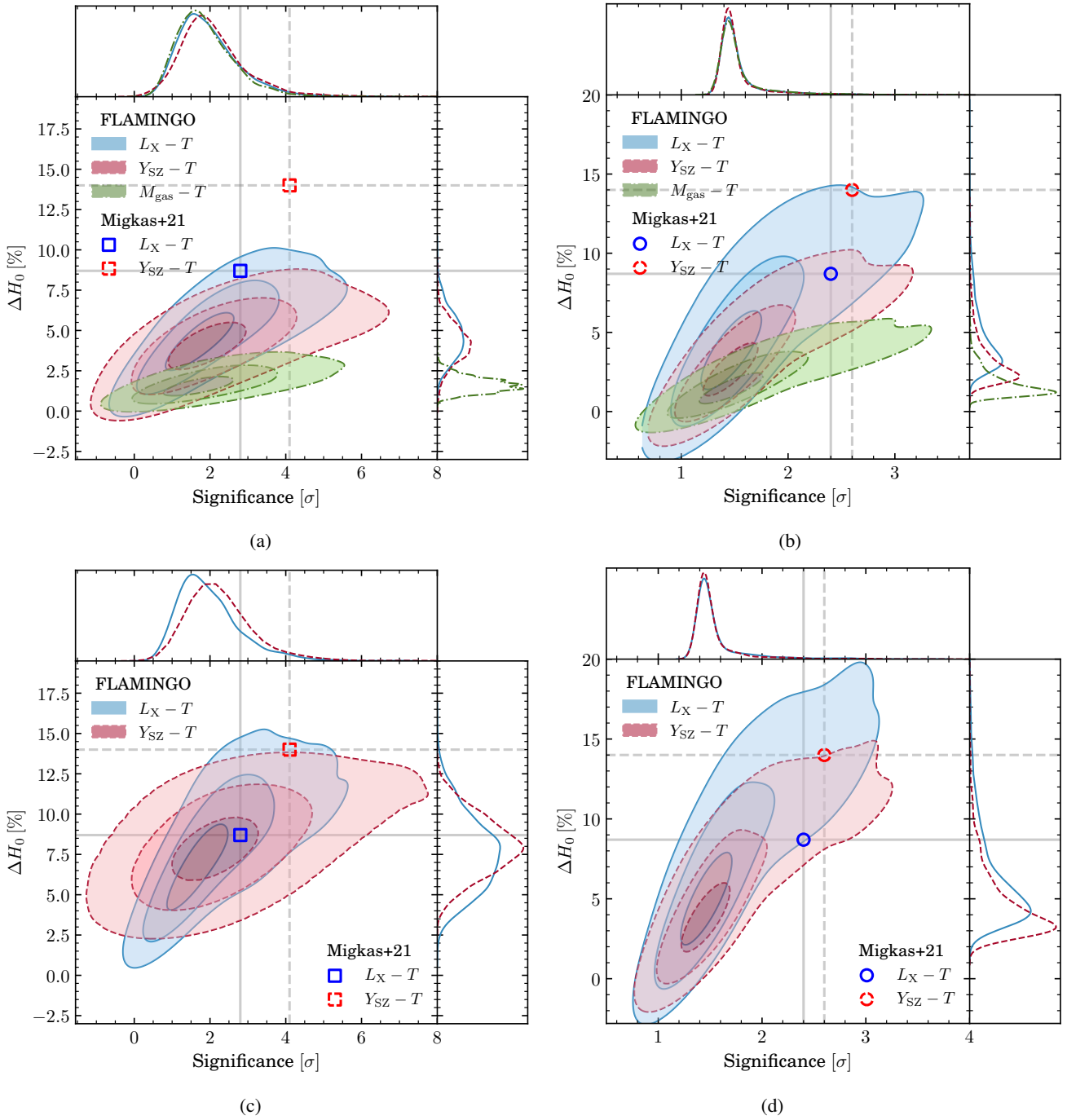


Fig. 3. Distribution of H_0 variation and its statistical significance from 1728 lightcones using the M21 scanning method (left) and the MCMC method (right) with (bottom) and without (top) mock scatter. The blue solid, red dashed, and green dash-dotted contours correspond to constraints from the L_X-T , $Y_{SZ}-T$, and $M_{\text{gas}}-T$ relations, respectively. Contour levels indicate 39%, 86%, and 98.9% probability regions (equivalent to 1, 2, and 3σ in two dimensions). Blue (red) squares mark the M21 results from L_X-T ($Y_{SZ}-T$), including their measured statistical significance. Blue (red) circles show the M21 results obtained from isotropic Monte Carlo realisations for L_X-T ($Y_{SZ}-T$). While the L_X-T results fall within the 3σ contours in every configuration, the $Y_{SZ}-T$ results fall well outside the contours for the no-scatter case, and roughly at the 3σ limit for the matched scatter case.

The rarity slightly exceeds the isotropic mock sample analyses in M21 which reported 2.6σ .

The $M_{\text{gas}}-T$ relation, included for the first time in such anisotropy analysis, consistently shows the lowest ΔH_0 values and the narrowest distributions on the isotropic lightcones. In the M21 method (Fig. 3a), ΔH_0 peaks at 1.5%, and the dipole significance remains below 2σ . Similarly, in the MCMC method

(Fig. 3b), ΔH_0 peaks at 1.2% and $n_\sigma \sim 1.4$, with a visibly tighter contour. These results identify $M_{\text{gas}}-T$ as the most stable tracer of isotropy among the three relations. Since the relation was not presented in M21, we do not add scatter or compute p -values, but its low scatter and minimal anisotropy make it a promising probe for future studies, especially given the anisotropy amplitude being sensitive to scatter.

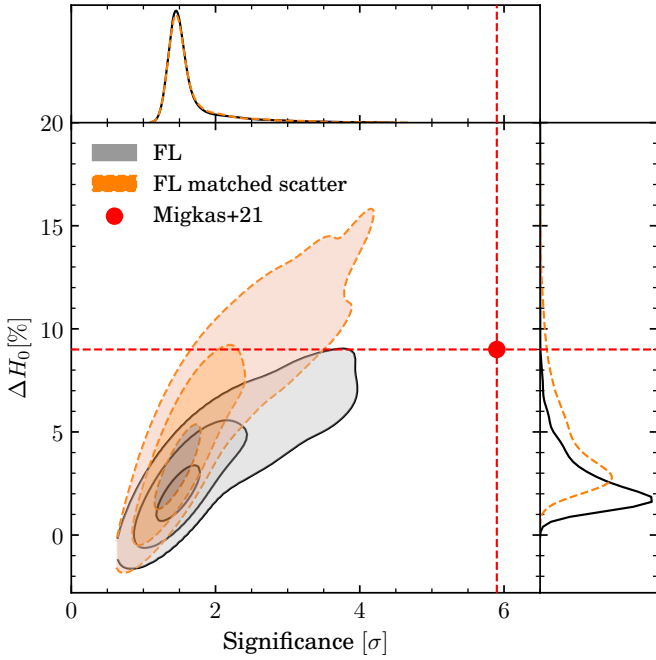


Fig. 4. Joint MCMC constraint from the L_X-T and $Y_{SZ}-T$ relations compared to the combined M21 result (red dot). The orange contour includes injected scatter, while the grey does not. Contours show the 39%, 86%, and 98.9% confidence regions.

Altogether, these results show that scatter in cluster scaling relations significantly contributes to the fitted apparent anisotropy. For L_X-T , scatter could largely account for the discrepancy. For $Y_{SZ}-T$, while scatter mitigates much of the tension, the remaining signal still places the M21 measurement in the upper tails of the lightcone distribution. $M_{\text{gas}}-T$ offers a promising future prospect as a more sensitive anisotropy probe. The overall trends are consistent across methods (M21 or MCMC) and probability estimators (KDE or EVS), supporting the conclusion that scatter matters and that tension with Λ CDM remains in the $Y_{SZ}-T$ detection.

4.2.2. Joint analysis of both L_X-T and $Y_{SZ}-T$

As different scaling relations may be affected by distinct systematics but all reflect the same underlying cosmology, a joint analysis can provide a more robust detection of a true cosmological anisotropy. However, past studies have shown a positive correlation between the scatter in L_X and Y_{SZ} at fixed mass (Nagarajan et al. 2019) or fixed temperature (M21), likely due to the interplay of physical processes inside of clusters. As a result, if a cluster is upscattered in L_X-T , it is also likely to be upscattered in $Y_{SZ}-T$. This correlation can cause random scatter to align and thus resemble a cosmological signal. M21 accounted for this by injecting the expected scatter correlation into their simulated samples used for significance estimation. When this correlation was included, the statistical significance of the dipole was found to be 5.9σ . Ignoring the correlation led to an underestimated significance of 5.4σ .

In this work, the correlation is built-in with the cluster physics in FLAMINGO. Therefore, we straightforwardly perform a joint likelihood MCMC analysis by sampling the combined log-likelihood function,

$$\log \mathcal{L} = \log \mathcal{L}_{\text{LT}} + \log \mathcal{L}_{\text{YT}}, \quad (20)$$

where $\log \mathcal{L}_{\text{LT}}$ and $\log \mathcal{L}_{\text{YT}}$ are given by Eq. (16). This result can be directly compared to the final constraints reported in M21 that are obtained by similarly multiplying the likelihoods of L_X and Y_{SZ} . In our samples and M21, the anisotropy signals are boosted by correlated scatter in the same manner. Therefore, the effect is cancelled to allow for a probability assessment unbiased to this effect.

The MCMC joint constraint is shown in Fig. 4, alongside the M21 result of $\Delta H_0 = 9\%$ at 5.9σ . The EVS analyses are presented in Figs. 5a and 5c. The M21 point lies well outside the 3σ region spanned by the lightcones (grey contours). Using projected EVS (Sect. 3.4), we estimate the probability of recovering the M21 result within FLAMINGO lightcones to be $p = 1.5 \times 10^{-4}$, corresponding to a 3.78σ tension (Fig. 5a). When the same test is repeated including injected scatter, the tension reduces to $p = 0.020$ (2.32σ with EVS, Fig. 5c). Although the added scatter brings the result visibly closer to M21, the KDE still yields a low p -value. As seen in Fig. 4, high ΔH_0 and n_σ are attainable in the MCMC method, but the combination of high n_σ and low ΔH_0 remains rare. However, since the M21 MC result was obtained using a slightly different statistical method (isotropic MC samples rather than full-sky MCMC), it is not guaranteed to follow the same 2D distribution. Thus, KDE-based probabilities with MCMC, especially in the tails, should be treated with caution in this comparison.

A joint probability from the M21 method results can also be obtained. Strictly following M21 and creating 10^5 isotropic samples for each lightcone is computationally unfeasible. Instead, we expand the projection to an average of L_X-T and $Y_{SZ}-T$, which naturally accounts for the potential correlation between the two relations. To incorporate directional alignment, we define a dimension-reduction function $x = |\mathbf{x}_{\text{LT}} + \mathbf{x}_{\text{YT}}|/2$, where $|\mathbf{x}_{\text{YX}}| = \Delta H_0 - \Delta H_0/n_\sigma$ and the vector direction corresponds to the anisotropy dipole. This reduces the five-dimensional space¹⁷ to a single variable, allowing us to apply the EVS method for probability estimation. The corresponding KDE estimation would require integrating the full five-dimensional space, which is beyond the scope of the method. The two entries in Table 4 for the Joint M21 analyses using KDE are thus empty.

Without added scatter, the M21 result¹⁸ lies beyond the numerical reach of even the EVS method, as shown in Fig. 5b. The corresponding entry is therefore left empty in Table 4. When matched scatter is included, the estimated probability becomes $p = 0.0012$ (3.24σ , EVS, Fig. 5d), indicating a moderate tension with Λ CDM. This constitutes our primary constraint on the H_0 anisotropy scenario: under identical statistical treatment and matched scatter, the M21 anisotropy result combining L_X-T and $Y_{SZ}-T$ differs from the FLAMINGO lightcones at 3.24σ .

In interpreting these estimations, one should note that the joint result in M21 incorporates 110–170 additional clusters from the ASCA Cluster Catalogue (ACC; Horner 2001) and an extra scaling relation, $L_{\text{BCG}}-T$, where L_{BCG} is the infrared luminosity of the brightest central galaxy per cluster. However, the contribution of $L_{\text{BCG}}-T$ is minor (with individual constraint at 1.9σ dipole significance and 0.8σ from mock samples), and the statistical impact of the ACC sample is compensated in our MCMC analyses by using $N = 400$, which better matches the combined data size of 481 for L_X-T and 373 for $Y_{SZ}-T$ in M21.

¹⁷ ΔH_0 and n_σ for L_X-T and $Y_{SZ}-T$, plus their angular separation.

¹⁸ As a rough estimation, the M21 result is 6.7 standard deviations away from the median of the lightcones; however, this cannot be taken at face value and translated to a p -value because the distribution (Fig. 5b) is not Gaussian.

Table 3. Summary of the probability analyses quantifying the tension between the M21 results and the FLAMINGO lightcones using different scaling relations (L_X-T , $Y_{SZ}-T$, or their combination), statistical methods (M21 sky scanning or MCMC), and probability estimators (KDE or projected EVS).

Relation	M21 (KDE)	M21 (projected EVS)	MCMC (KDE)	MCMC (projected EVS)
L_X-T	$p = 0.021$	$p = 0.025$ (2.25 σ)	$p = 0.055$	$p = 0.032$ (2.14 σ)
$Y_{SZ}-T$	$p < 5.8 \times 10^{-4}$	$p = 1.8 \times 10^{-10}$ (6.38 σ)	$p < 5.8 \times 10^{-4}$	$p = 2.8 \times 10^{-4}$ (3.64 σ)
Joint analysis	–	–	$p < 5.8 \times 10^{-4}$	$p = 1.5 \times 10^{-4}$ (3.78 σ)
L_X-T (scatter)	$p = 0.33$	$p = 0.23$ (1.20 σ)	$p < 5.8 \times 10^{-4}$	$p = 0.085$ (1.72 σ)
$Y_{SZ}-T$ (scatter)	$p < 5.8 \times 10^{-4}$	$p = 0.0013$ (3.22 σ)	0.0014	$p = 0.0049$ (2.82 σ)
Joint analysis (scatter)	–	$p = 0.0012$ (3.24 σ)	$p < 5.8 \times 10^{-4}$	$p = 0.020$ (2.32 σ)

Notes. The p -values denote the probability that a randomly placed observer in a Λ CDM universe would measure an anisotropy equal to or more extreme than that seen in M21, with equivalent (1D) Gaussian significance given in parentheses where applicable. ‘Scatter’ refers to the mock samples with matched M21-level instrumental uncertainties (see Sect. 2.4). Dashes indicate configurations beyond the scope of the probability estimators.

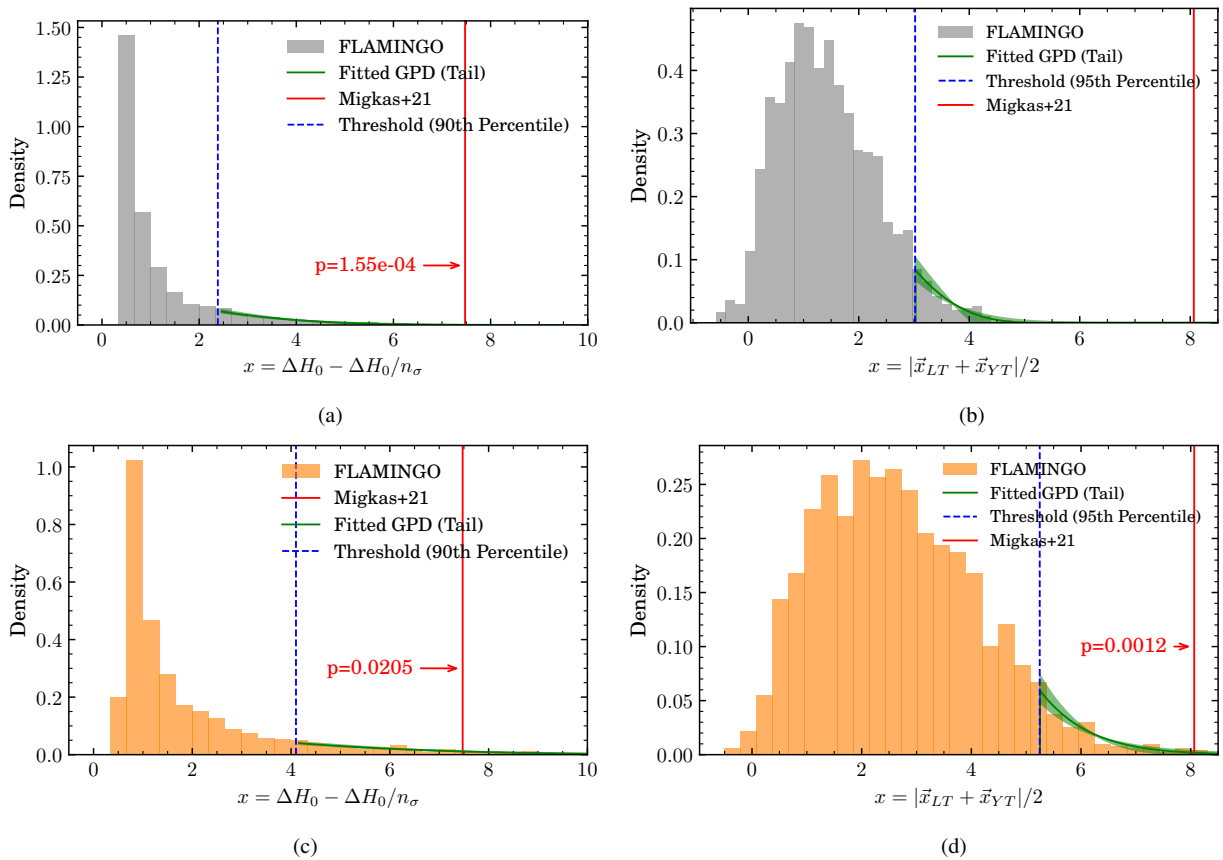


Fig. 5. Extreme value statistics analyses of joint H_0 anisotropy in the FLAMINGO lightcones using the MCMC (left) and M21 (right) methods without (top) and with (bottom) matched scatter. Grey and orange histograms show the lower-bound projection of all 1728 lightcones without and with scatter, respectively. The red vertical line marks the M21 result under the same projection. The blue dashed line indicates the GPD threshold (90th or 95th percentile), above which the tail is modelled by a GDP, shown in green. The small shaded green region around the best fit line represents the 16th to 84th percentile region of bootstrapping error of the fit. Annotated in red is the EVS probability of obtaining the M21 result in FLAMINGO.

In addition, we assess the correlation of the L_X and Y_{SZ} scatter with respect to T in FLAMINGO. We find that the two quantities exhibit a correlation with a Pearson’s correlation coefficient $r = 0.25$ (0.31 with matched scatter; see Appendix E). This correlation is weaker than what is found in M21 ($r = 0.67$). However, this should not bias our result significantly because in M21 L_X-T has ~ 100 more clusters than $Y_{SZ}-T$, meaning ~ 100 samples are not common, and therefore the effect of a stronger correlation is mitigated.

4.3. Bulk flow constraints

Assuming isotropic Hubble expansion, we estimate the bulk flow using the MCMC implementation described in Sect. 3.3. The bulk flow is constrained within three z spheres: $z < 0.07$, $z < 0.10$, and $z < 0.13$. These are chosen to balance the need for a sufficient number of clusters with maintaining a relatively uniform redshift distribution within each sphere. Additionally, they allow for direct comparisons with M21.

Table 4. Summary of bulk flow results from the FLAMINGO lightcones for different scaling relations and redshift ranges in direct comparison with the observed values from M21.

Relation	Redshift Range	No scatter u_{bf} (km s ⁻¹)	Matched scatter u_{bf} (km s ⁻¹)	FLAMINGO $ \langle \mathbf{v} \rangle $ (km s ⁻¹)	M21 observed u_{bf} (km s ⁻¹)	p -value of M21 (σ)	p -value of M21 (σ) matched scatter
$L_X - T$	$z < 0.07$	315^{+262}_{-131}	414^{+263}_{-169}	86^{+40}_{-35}	1100^{+410}_{-410}	^a 0.0067 (2.71 σ)	^a 0.0081 (2.65 σ)
	$z < 0.10$	315^{+260}_{-131}	415^{+260}_{-172}	75^{+36}_{-30}	1080^{+380}_{-380}	0.0082 (2.64 σ)	0.0067 (2.72 σ)
	$z < 0.13$	313^{+263}_{-130}	417^{+261}_{-172}	68^{+34}_{-27}	–	–	–
$Y_{\text{SZ}} - T$	$z < 0.07$	164^{+119}_{-56}	264^{+170}_{-83}	86^{+40}_{-35}	1060^{+390}_{-390}	3.2×10^{-6} (4.66 σ)	5.8×10^{-4} (3.44 σ)
	$z < 0.10$	164^{+119}_{-56}	264^{+166}_{-81}	75^{+36}_{-30}	1020^{+380}_{-380}	3.5×10^{-6} (4.64 σ)	0.0011 (3.26 σ)
	$z < 0.13$	164^{+117}_{-56}	264^{+172}_{-82}	68^{+34}_{-27}	–	–	–
$M_{\text{gas}} - T$	$z < 0.07$	158^{+129}_{-73}	–	86^{+40}_{-35}	–	–	–
	$z < 0.10$	158^{+130}_{-73}	–	75^{+36}_{-30}	–	–	–
	$z < 0.13$	158^{+130}_{-72}	–	68^{+34}_{-27}	–	–	–
$L_X - T + Y_{\text{SZ}} - T$	$z < 0.07$	–	–	–	–	^a 5.0×10^{-5} (4.05 σ)	^a 3.6×10^{-4} (3.57 σ)
Combined	$z < 0.10$	–	–	–	–	4.5×10^{-5} (4.08 σ)	3.8×10^{-4} (3.56 σ)

Notes. The columns show the fitted bulk flow amplitudes, u_{bf} (with and without matched scatter), the actual bulk peculiar velocity in FLAMINGO ($|\langle \mathbf{v} \rangle|$), the observed M21 values, and the resulting p -values (with Gaussian-equivalent significance in parentheses). Results are reported for three redshift shells and for the combined $L_X - T + Y_{\text{SZ}} - T$ analysis. Dashes indicate configurations not evaluated. ^(a) Precisely, the $L_X - T$ bulk flow in M21 is measured in $z < 0.06$; however, this difference is negligible.

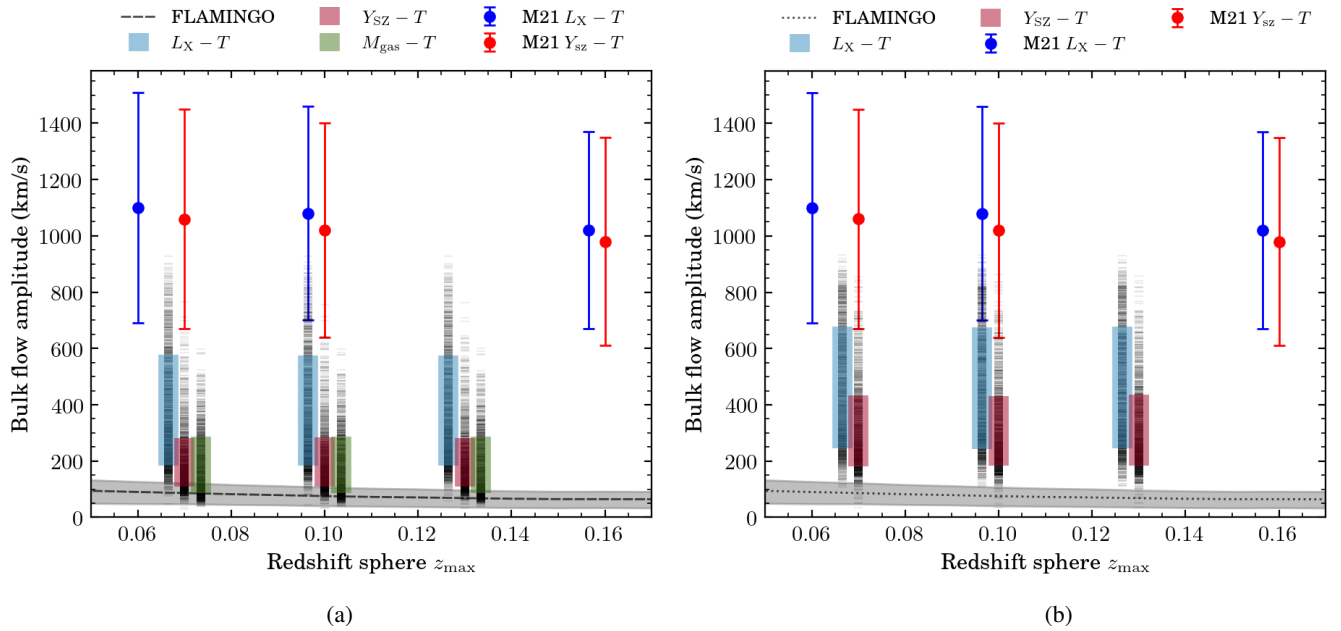


Fig. 6. The MCMC bulk flow amplitude constraints for $z < 0.07$, $z < 0.10$, and $z < 0.13$ shown for the $L_X - T$, $Y_{\text{SZ}} - T$, and $M_{\text{gas}} - T$ relations. The right panel includes injected scatter and the left panel does not. Black dashes represent the results from individual lightcones, while the thick coloured bars indicate the 16th to 84th percentile range across the lightcones. For comparison, the M21 results for $L_X - T$ and $Y_{\text{SZ}} - T$ are shown with their original error bars, corresponding to redshift spheres of $z < 0.06$, $z < 0.07$, $z < 0.10$, and $z < 0.16$. To avoid visual overlap, the points for $L_X - T$ and $M_{\text{gas}} - T$ are slightly offset along the x -axis. The black dashed line represents the median true average bulk motion of the clusters, with the grey shaded region indicating the 16th to 84th percentile range of the lightcones. (a) No scatter. (b) Matched scatter.

The bulk flow results are presented in Fig. 6 and summarised in Table 4, alongside the M21 measurements and the true FLAMINGO bulk flow. The FLAMINGO bulk flow is estimated by the average of cluster velocities of our sample, $|\langle \mathbf{v} \rangle| = |\sum_i \mathbf{v}_i|/N$. As expected, we find that the true bulk flows are $\lesssim 100$ km s⁻¹ across our redshift range, consistent with Λ CDM predictions (Qin et al. 2021; Watkins et al. 2023). Using the $L_X - T$, $Y_{\text{SZ}} - T$, and $M_{\text{gas}} - T$ relations, one finds a $\approx 315^{+262}_{-131}$ km s⁻¹, 164^{+119}_{-56} km s⁻¹, and 158^{+130}_{-73} km s⁻¹ bulk flow respectively, in all three redshift spheres. It is evident that $L_X - T$ returns the most overestimated bulk flow compared to the true

FLAMINGO value (black dashed line in Fig. 6) due to its larger scatter. On the other hand, the low-scatter $Y_{\text{SZ}} - T$ and $M_{\text{gas}} - T$ relations return values much closer to the true underlying bulk flow ($< 1.5\sigma$ difference). When scatter is matched to M21, the recovered bulk flow increases to about 415^{+260}_{-172} km s⁻¹ for $L_X - T$ and 264^{+166}_{-81} km s⁻¹ for $Y_{\text{SZ}} - T$ (at $z < 0.10$, similar at other redshift shells), amplifying the overestimation compared to the true value. The systematic overestimation of the bulk flow by all three relations and its increase with mock scatter indicate that the bulk flow constraints also depend on the scatter in the cluster scaling relations.

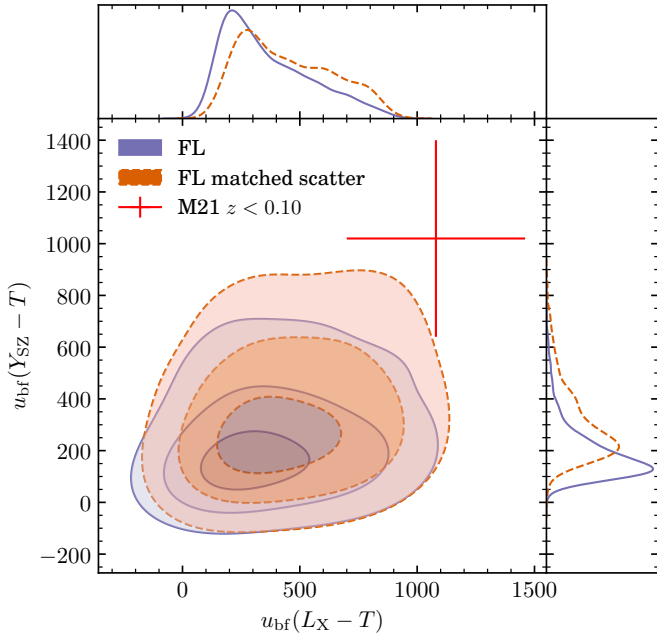


Fig. 7. Distribution of the L_X-T and $Y_{SZ}-T$ bulk flow constraints at $z < 0.10$ along with the M21 result at the same redshift with its error bars (red). The orange contour includes injected scatter, while the purple does not. Contours of different shades give the 39%, 86%, and 98.9% regions.

Compared to M21, we find no prominent bulk flow exceeding the level of M21 ($\sim 900 \text{ km s}^{-1}$). However, the dispersion of our bulk flow constraints and the observational uncertainties of M21 are both large. We estimate the statistical significance of this discrepancy while accounting for the large uncertainties by adopting the same EVS-based projection method used in the H_0 analysis. We define the effective 1D variable $x = u_{\text{bf}} - \sigma_{\text{low}}$, where σ_{low} is the lower 1σ uncertainty of each lightcone fit, and assess the probability of recovering the M21 value under this projection using the EVS method (Sect. 3.4).

The resulting p -values at $z < 0.07$ and $z < 0.10$ are given in Table 4 (corresponding EVS plots are included in Appendix G). At $z < 0.07$, using L_X-T , the probability of obtaining the M21-level bulk flow from an isotropic lightcone is $p = 0.0067$ (2.71σ) without scatter and $p = 0.0081$ (2.65σ) with matched scatter. For $Y_{SZ}-T$, the tension is stronger: $p = 3.2 \times 10^{-6}$ (4.66σ) (no scatter) and $p = 5.8 \times 10^{-4}$ (3.44σ) (with scatter). The $z < 0.10$ results are similar, yielding 2.72σ (L_X-T) and 3.26σ ($Y_{SZ}-T$) with scatter.

A joint $L_X-T + Y_{SZ}-T$ bulk flow distribution is shown in Fig. 7. Compared to the H_0 case, increasing the scatter has a smaller effect. This implies that bulk flow amplitudes are less sensitive to scaling relation scatter in our implementation. We estimate the joint significance by taking the average bulk flow vector amplitude as $x = |x_{LT} + x_{YT}|/2$, following the same prescription as in the H_0 analysis (Sect. 4.2.2). This results in a final tension of $p = 3.6 \times 10^{-4}$ (3.57σ) at $z < 0.07$ and $p = 3.8 \times 10^{-4}$ (3.56σ) at $z < 0.10$, under matched scatter. Without scatter injection, the tension is only slightly higher at 4.05σ ($z < 0.07$) and 4.08σ ($z < 0.10$). While no single lightcone reaches $u_{\text{bf}} \sim 900 \text{ km s}^{-1}$, such values remain rare even after accounting for scatter. The resulting tension with ΛCDM , at the $\sim 3.6\sigma$ level, exceeds that of the H_0 anisotropy ($2.3-3.2\sigma$). This may point towards a more localised, low-redshift departure from ΛCDM predictions.

In addition, we notice that the bulk flow tension seems to be less affected by scatter compared to the H_0 anisotropy. Here the tension only drops by 0.4σ with matched scatter, whereas the joint H_0 tension dropped $>1.4\sigma$ with MCMC and more with the M21 method.

We note that for the bulk flow, the M21 observational analysis employed a χ^2 minimisation over the full sky, while we adopt a full-likelihood maximisation via MCMC. In addition, M21 estimated uncertainties using bootstrap resampling, whereas we derive them from the posterior distributions of the MCMC samples. These methodological differences may introduce variations in the estimated uncertainties. However, as χ^2 minimisation is mathematically equivalent to maximising the likelihood under Gaussian assumptions, the impact on the central values is expected to be minimal. If anything, the use of MCMC could lead to slightly narrower uncertainty estimates. Since we evaluate the statistical tension using the lower 1σ bound of each lightcone's distribution, any underestimation of uncertainties would in fact strengthen the inferred tension with ΛCDM .

5. Discussion

5.1. Cosmological implications

By generating 1728 lightcones in the FLAMINGO (2.8 Gpc^3) box, we find that the M21 results are unlikely to occur within a ΛCDM universe. In joint analyses of the L_X-T and $Y_{SZ}-T$ relations, the amplitude of H_0 variation and the significance reported by M21 lie beyond the range of simulated lightcones. We estimate the overall tension between the M21 findings and ΛCDM to be $\sim 2.3-3.2\sigma$, depending on the exact statistical method and probability estimator. The alternative explanation based on a coherent bulk flow also shows significant tension: at $z < 0.1$, the M21-level bulk flow constitutes a 3.6σ deviation. Notably, none of our 1728 lightcones produce such a large bulk flow, even when observational scatter is included. The fact that the statistical tension increases for bulk flow model z compared to H_0 suggests that the observed anisotropy is at a higher discrepancy locally at $z < 0.1$.

Overall, these results suggest that both scenarios proposed in M21 to explain their findings (H_0 anisotropy or a significant bulk flow) are in tension with ΛCDM and, although substantially reduced, cannot be fully attributed to statistical noise or cosmic variance. Assuming no major undiscovered systematic biases as yet in the M21 results or the FLAMINGO simulations, this work implies that either the ΛCDM model needs modifications at $z < 0.2$ or that we inhabit a statistically rare region of the Universe. Any of the two scenarios can potentially have significant implications for other low- z cosmological tensions, such as the Hubble tension. Local H_0 measurements typically assume isotropy in the expansion rate and ΛCDM -predicted bulk flows. If either assumption fails, as our findings suggest may be the case, it could systematically bias the inferred H_0 values (Migkas 2025). This opens the possibility that part of the Hubble tension arises from local anisotropies rather than global physics.

Another key finding with direct cosmological relevance is the role of scatter in cluster scaling relations. We show that the observed anisotropy amplitude is boosted by scatter in L_X-T and $Y_{SZ}-T$, even when the input cosmology is isotropic. Tighter relations, such as $M_{\text{gas}}-T$, exhibit lower bias and are therefore better suited for future anisotropy studies. For example, with matched scatter, we find that FLAMINGO lightcones overestimate ΔH_0 by $\sim(7-8)\%$ when L_X-T or $Y_{SZ}-T$ are used alone (see Fig. 3). This implies that the $\Delta H_0 = 9\%$ reported in M21

may correspond to a true underlying anisotropy of only $\sim 3\%$ ¹⁹ or so. Properly addressing this bias would require injecting an artificial anisotropy signal into simulations and comparing the measured with the input values (see Sect. 5.6.3). If a strong signal exists, the bias from noise may be smaller. Future work will provide a more conclusive assessment.

Finally, to aid the interpretation of our results, we summarise the important assumptions that we adopt. First and most importantly, we assume that the shape of the anisotropies is dipole-like and study only dipole anisotropies, either dipole H_0 variations or a single bulk flow component. We do this because we focus on understanding the results of M21, which found a dipole-shaped anisotropy. Higher order anisotropies are beyond the scope of this work. Second, our interpretation that translates the probability of an anisotropy in the simulations to the probability of that in a Λ CDM universe assumes that the FLAMINGO simulation (with its physical modelling and numerical implementations) is an accurate realisation of the Λ CDM cosmology. And third, the scatter injection method as described in Sect. 2.4 assumes that the scatter is mass independent and with a log-normal distribution.

5.2. Origin of apparent anisotropies: Statistical noise or peculiar velocities

We now discuss the possible origins of the apparent anisotropies observed in the FLAMINGO lightcones. One key contributor is bulk flow. Coherent peculiar motions can introduce real anisotropy signals that manifest consistently across different scaling relations. As shown in Fig. 6, FLAMINGO reproduces the expected low-amplitude large-scale bulk flows of Λ CDM. Such motions are predicted to induce apparent H_0 anisotropies at the $\sim 3\%$ level (Migkas 2025), which is close to the amplitudes found in our simulated lightcones.

A second major source is statistical noise, particularly random upscattering of ΔH_0 in individual lightcones. In this scenario, different scaling relations would yield weakly correlated anisotropies in the same lightcone. Especially ones that are less correlated in scatter (Appendix E). The two sources can be disentangled by examining the degree of alignment between anisotropies inferred from different relations. Distinguishing between bulk flow and statistical noise is crucial for guiding future studies of cosmic anisotropy and for interpreting the significance of observed signals.

5.2.1. Directional correlation between different relations

To investigate whether the apparent anisotropies arise from a genuine bulk flow, we examine whether different scaling relations exhibit their maximum dipole in similar directions. We calculate the angular separation between the directions of the strongest dipole detected in each relation. Our results indicate that most relation pairs do not show significant directional alignment. The only pair with a mild angular correlation is L_X-T and $M_{\text{gas}}-T$, whose dipoles peak around a separation angle of $\sim 20^\circ$. By contrast, $Y_{\text{SZ}}-T$ and $M_{\text{gas}}-T$ show only weak alignment, and L_X-T and $Y_{\text{SZ}}-T$ exhibit no consistent directional correlation. We also test whether a larger ΔH_0 is associated with better directional agreement, i.e. smaller angular separation, but find no such correlation.

¹⁹ The M21 detection was $\Delta H_0 = 14\%$ at 4.1σ using $Y_{\text{SZ}}-T$ and 8.7% at 2.8σ using L_X-T . The $Y_{\text{SZ}}-T$ detection likely retains an amplitude above 1%.

These findings suggest that the apparent anisotropies in the simulated lightcones are most likely predominantly driven by statistical noise rather than by peculiar velocity fields consistent with Λ CDM. This interpretation is further supported by the scatter correlation patterns (Appendix E), where relations that align in scatter also tend to align in dipole direction. (For a visualisation of the angular separation distributions and their cross-correlations, see Appendix C.)

5.2.2. Effect of peculiar velocities in apparent anisotropies

Peculiar velocities contribute to anisotropy signals in two ways: random motions increase the scatter in scaling relations, while coherent bulk flows can mimic an H_0 dipole. To assess their impact, we repeat the anisotropy analyses using cosmological redshifts z_{cos} instead of the observed redshifts z_{obs} (see Sect. 2.1). In this setup, clusters are effectively placed at their Hubble-flow distances, completely removing the influence of peculiar velocities.

We find that this replacement leads to a mild reduction in anisotropy amplitudes. For L_X-T , the median ΔH_0 decreases from 3.94% to 3.35%; for $Y_{\text{SZ}}-T$, from 2.71% to 2.24%; and for $M_{\text{gas}}-T$, from 1.50% to 1.26%. Hence, peculiar velocities–bulk flows included–account for only $\sim 15\text{--}17\%$ of the apparent anisotropy, with the remaining contribution attributed to statistical noise.

A similar trend is seen in the bulk flow fits. For L_X-T , the amplitude decreases slightly from 313 km s^{-1} to 303 km s^{-1} across all redshift bins. The $M_{\text{gas}}-T$ relation shows a similar drop from 161 km s^{-1} to 151 km s^{-1} , and $Y_{\text{SZ}}-T$ exhibits a stronger decline from 165 km s^{-1} to 135 km s^{-1} . Even when peculiar velocities are fully removed, the reduction in signal remains below $\sim 18\%$, reinforcing the interpretation that statistical noise dominates the recovered anisotropies. This suggests that our bulk flow analysis is primarily dominated by noise when the underlying bulk flow amplitude is $u_{\text{bf}} \lesssim 100 \text{ km s}^{-1}$ at $z \sim 0.10$ (Table 4).

5.3. Effects of scaling relation scatter in apparent anisotropies

We find that the scatter in scaling relations directly affects the apparent anisotropies. Fig. 4 shows that injecting scatter boosts ΔH_0 while leaving n_σ nearly unchanged. This behaviour holds across individual relations and with the M21 method (Fig. 3). The reason is that scatter amplifies both the anisotropy amplitude and its uncertainty, keeping the statistical significance approximately stable. As a result, ΔH_0 is more sensitive to observational systematics than n_σ , and our projection-based methods may not fully capture this effect, potentially underestimating the true tension.

Importantly, it is not the total scatter of the relation itself but the scatter in ΔH_0 that matters. For instance, with injected scatter, the $Y_{\text{SZ}}-T$ relation yields a comparable ΔH_0 to L_X-T under the M21 method (Fig. 3c), despite having $\sim 25\%$ lower σ_{tot} (Table 2). In this sense, the MCMC method is more robust to noise, as its contour shapes remain less disturbed and low-scatter relations give lower ΔH_0 consistently (Figs. 3d and 4). The $M_{\text{gas}}-T$ relation, which shows the lowest H_0 scatter, yields the most stable results. Future studies should prioritise low-scatter relations to reduce bias, and apply caution when interpreting anisotropy amplitudes from high-scatter relations.

Finally, we note that although the simulated scatter is generally matched to the observational values, there remains

uncertainty in the scatter, the scatter may still be slightly underestimated if the uncertainties are not well matched. For the L_X-T ($Y_{SZ}-T$) relation, the typical fitting uncertainty in the intrinsic scatter is ~ 0.015 (0.012) dex²⁰ in our lightcones, which closely matches the M21 uncertainties of 0.016 and 0.013 dex, respectively. However, we find a slight underestimation in the uncertainty of the total scatter, 0.015 dex for L_X-T and 0.010 dex for $Y_{SZ}-T$, compared to 0.018 and 0.015 dex in M21. This could result in a marginal underestimation of the scatter, even after our scatter-matching. Nonetheless, the effect should be small. The bootstrapped scatter distributions in both M21 and FLAMINGO are approximately Gaussian and do not exhibit strong tails. Anomalies caused by a few strongly scattered clusters would therefore be reproduced in the lightcones. Moreover, it was carefully examined in both M20 and M21 that there is no strong outlier causing the observed anisotropy, further reducing this effect.

5.4. Effects of the highest X-ray concentration selection

As we noted in Sect. 2.2, for the same selection criteria, FLAMINGO yields more clusters than the observed M21, but with a lower c_X distribution. The origin of this discrepancy requires further investigation as we discuss in Appendix D.1. For our analysis, we adopted a temporary solution by selecting the highest c_X clusters from the simulated samples. Since the selection is performed independently of the sky position of clusters and should have no direct impact on our results. Nevertheless, there might be secondary systematic effects that can potentially bias our results (e.g. via affecting the intrinsic scatter of scaling relation).

To check whether our final results are robust against the applied c_X selection, we repeat the MCMC tests with random cluster selection independently of the c_X values. Specifically, after the flux and zone-of-avoidance selections (described in Sect. 2.2), for each lightcone, we randomly drew 313 (400) clusters for the M21 (MCMC) statistical approach, out of the 700–800 remaining clusters. Using the random cluster selection in each lightcone, we quantify the tension between the M21 findings and Λ CDM to be 2.27σ for L_X-T , 3.38σ for $Y_{SZ}-T$, and 3.42σ when the two relations combined. The L_X-T tension increased by 0.1σ while the $Y_{SZ}-T$ tension decreased by 0.2σ , and the combined tension slightly reduces by $\sim 0.4\sigma$. With matched scatter, the L_X-T tension is approximately unchanged at 1.74σ , while the $Y_{SZ}-T$ tension is at 2.60σ , slightly (0.2σ) lower than when the c_X -bases selection was used. When combined in the joint probability, the overall tension with matched scatter is 2.32σ , which is essentially the same with the default analysis (2.31σ , Table 3). When modelled as a bulk flow and including the injected scatter, we find a joint (combining L_X-T and $Y_{SZ}-T$, as described in Sect. 4.3) 3.63σ (3.56σ) for $z < 0.07$ ($z < 0.10$), similar to the default results shown in Table 4. Without matched scatter, the joint tension exceeds 4σ , and the corresponding p -value is so low that even the EVS method cannot reliably estimate it.

Based on the above tests, we conclude that although the c_X discrepancy between observational data and the FLAMINGO simulation warrants further investigation, it does not noticeably impact the main conclusions of this work. The $\geq 3\sigma$ tension between our simulation lightcones and M21 is neither caused

by nor sensitive to our top c_X selection during sample selection (Sect. 2.2).

5.5. Comparing the M21 method to MCMC

Constraints on cosmic anisotropy from galaxy clusters depend sensitively on the statistical method used, underscoring the importance of robust approaches with minimal noise. We have presented results from both the M21 and MCMC methods separately, and now compare them directly by correlating outcomes for the same lightcones. In the absence of scatter, ΔH_0 values from M21 and MCMC show only weak correlation when using the same scaling relation, with Pearson coefficients of ~ 0.5 for L_X-T , $Y_{SZ}-T$, and $M_{\text{gas}}-T$. A similarly weak correlation is found in their respective n_σ values ($r \lesssim 0.5$; see Appendix F), consistent with the interpretation that apparent anisotropies in the isotropic mock lightcones are dominated by statistical noise rather than bulk flows.

Without a true cosmic anisotropy or a bulk flow mimicking one, the MCMC method yields less noisy constraints, with narrower distributions in both ΔH_0 and n_σ (Fig. 3) across all three scaling relations and more stable to injected noise (Sect. 5.3). It shows better angular consistency between dipoles inferred from different relations (see Sect. 5.2.1). Additionally, the MCMC method is computationally more efficient, more versatile, and can be extended to non-dipole anisotropies with relative ease.

These advantages, however, are not sufficient to conclude that MCMC outperforms the M21 method in all contexts. First, the MCMC method produces significantly more outliers, which strongly boosts any probability estimation (e.g. EVS) based on the M21 distribution. In future observations, this may increase the risk of reporting statistically significant anisotropies from data that are otherwise consistent with isotropy. While comparison to simulations or other mock samples can mitigate this effect, the heavy-tailed nature of the MCMC result remains challenging to model robustly. Second, our implementation of the MCMC method corresponds to applying the M21 approach with a 90° cone. Using larger cones tends to smooth out both real signals and statistical fluctuations, lowering both ΔH_0 and n_σ (see M20, their Table 2). Third, part of the reduced noise in MCMC may be attributed to the larger cluster sample size: 400 clusters for MCMC versus 313 for M21. Nonetheless, the MCMC-based method offers a promising avenue for future observational applications.

It is worth emphasising that the two methods differ in more than just sampling techniques. The MCMC method uses the full likelihood rather than χ^2 , MCMC sampling instead of bootstrap resampling, full-sky fitting rather than cone scanning, and forward modelling rather than a direct A to H_0 translation. Any difference in performance must therefore be understood in the context of these broader methodological distinctions.

5.6. Future prospects

5.6.1. Other cluster catalogues

While the statistical significance of the cosmic dipole reported in M21 is lowered after improved treatment of cosmic variance, a notable tension with Λ CDM remains at $z \lesssim 0.1$ (3.6σ). A genuine cosmological signal should appear consistently across datasets, independent of scaling relation, selection function, or instrument. To mitigate potential systematics and selection biases, it is essential to repeat such analyses on independent cluster samples with alternative selection methods

²⁰ The values differ from Table 2 because here we consider the uncertainty in each individual lightcone, rather than the distribution across all lightcones, which is what is reported in Table 2.

(e.g. SZ selection), aided by dedicated mock catalogue studies. We plan to apply our approach to the eROSITA first All-Sky Survey (eRASS1), which comprises over 12 000 clusters across $z = 0.01\text{--}1.2$ (Bulbul et al. 2024). Other catalogues, including X-CLASS (Koulouridis et al. 2021) and XCS (Mehrtens et al. 2012; Giles et al. 2022), also provide promising avenues for future investigation. A consistent detection of a dipole across such independent datasets would offer a robust test of any large-scale anisotropy of cosmological origin.

5.6.2. Improvement on lightcone independence

Our 1728 lightcones do not represent entirely 1728 realisations of the Λ CDM cosmology. While each lightcone places the observer at a different location and the FLAMINGO simulation spans an unprecedented scale, the overlap between lightcones is unavoidable. Neighbouring observers are ~ 230 cMpc apart. As a result, similar structures will appear across neighbouring lightcones at $z \gtrsim 0.05$. The effect on our analysis is kept to a minimum by random rotations and zone-of-avoidance masking (Sect. 2.2). For each lightcone, a different random galactic plane is masked to reduce repetitive structures. Future analyses can look to address this limitation by developing better techniques for building the samples. For example, using the dense sphere packing technique instead of a cubic grid that we use, the number of independent lightcones can increase by 40% under the same volume. To further increase the statistical power, larger simulation volumes are needed.

5.6.3. Response to injected anisotropy

All the tests in this work were performed on datasets with no or small (e.g. from bulk flows) signals. We have concluded that the M21 findings are rare if only noise is present. It is then important to study, if exists, whether the amplitude of such an anisotropy is overestimated in M21. In this work, we showed how statistical noise can cause a spurious anisotropy signal. It is likely the same mechanism could boost up an existing signal. Future work should include controlled signal injection to assess the ability of this methodology to recover known input anisotropies. Such tests would also allow for direct comparison with M21 and newer observational results and provide a more systematic validation of the use of cluster scaling relations in probing cosmic anisotropy.

6. Conclusion

In this work, we have quantified the tension between previously observed cosmic anisotropies using galaxy clusters and the Λ CDM model using the fully hydrodynamic cosmological FLAMINGO simulations. We created 1728 isotropic lightcones and analysed cluster scaling relations $L_X\text{--}T$, $Y_{SZ}\text{--}T$, and $M_{\text{gas}}\text{--}T$ using two independent statistical approaches: one following M21 and another using MCMC inference with full likelihood. We also increased the intrinsic scatter of the cluster scaling relations and assigned instrumental uncertainties that match M21 data. We have provided a robust assessment of cosmic variance, statistical noise, and systematic biases inherent in the observational findings.

Our analysis reveals that the observed 9% H_0 dipole detected by M21 is rare within a Λ CDM framework. For the first test using the original FLAMINGO data (without matched scatter), M21 is at a strong tension with the Λ CDM simulation. Using the same methodology with $L_X\text{--}T$ and $Y_{SZ}\text{--}T$, the results differ

by 6.7σ between M21 and the simulation lightcones. Using the MCMC method yields 3.8σ (4.0σ) when the anisotropy is modelled as an H_0 variation (bulk flow). However, scatter in the scaling relations plays a significant role in the results, which the simulations underestimate by $\sim 25\%$. Therefore, we also performed tests with scatter injected.

When the intrinsic scatter is matched to observations and mock instrumental uncertainties are assigned, the joint analysis of the $L_X\text{--}T$ and $Y_{SZ}\text{--}T$ scaling relations shows tension between data and the standard cosmological model at $2.3\text{--}3.6\sigma$. Repeating the exact M21 method, the M21 finding is 3.2σ more anisotropic than the FLAMINGO lightcones. Additionally, we examined the bulk flow hypothesis as an alternative explanation for the observed anisotropies. While the simulated lightcones exhibit Λ CDM-like bulk flows, no bulk flow motion comparable to the 900 km s^{-1} flow observed by M21 was detected in our 1728 lightcones. A 3.6σ tension between the M21 bulk flow and FLAMINGO lightcones was found at $z < 0.07$ and $z < 0.10$. This indicates that the Λ CDM predictions are less compatible with the M21 cluster data at low redshift. Although this finding does not alleviate the dipole found in M21, the statistical tension with Λ CDM is significantly reduced from 5.4σ (M21) to 3.2σ (this work).

We find that the amplitude of the detected anisotropies in FLAMINGO is dominated by statistical noise in the cluster scaling relations at the $>80\%$ level, while the remaining $<20\%$ is attributed to peculiar velocities (bulk flows included). We have shown that in the case of dominant statistical noise, different scaling relations yield anisotropy in different directions with no strong correlation between them. This has important implications for future studies, as by utilising independent relations and cluster samples, one can distinguish statistical noise from true cosmological effects by looking at the correlation of the detected dipole directions. The fact that tight direction correlation is very rare also stresses that the M21 finding with three relations and two independent catalogues all pointing to a $<10^\circ$ cone is unlikely to be a coincidence. Last but not least, we find that the inferred ΔH_0 in the isotropic lightcones increase significantly with a larger scatter in the scaling relation (injected or not). The $M_{\text{gas}}\text{--}T$ relation has the smallest scatter and thus shows a better constraining power than $L_X\text{--}T$ and $Y_{SZ}\text{--}T$. Lower σ_{intr} relations will play an important role in future studies.

By carefully accounting for cosmic variance and statistical noise, the tension of the detected M21 anisotropy and Λ CDM is substantially reduced but remains significant. This work highlights the necessity of simulations as an integral part of interpreting statistical significance and isolating genuine cosmological anomalies from sample variance. As next-generation datasets from surveys such as eRASS become available, extending this analysis with real data alongside simulated cluster catalogues will be crucial for corroborating these results and refining our understanding of the observed cosmological anisotropies.

Acknowledgements. We thank the anonymous referee for their valuable comments that helped us improve our paper. K.M. acknowledges support in the form of the X-ray Oort Fellowship at Leiden Observatory. 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.

References

Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. 2022, *J. High Energy Astrophys.*, 34, 49

- Aluri, P., Cea, P., Chingangbam, P., et al. 2023, *Class. Quant. Grav.*, **40**, 094001
- Andrade-Santos, F., Jones, C., Forman, W. R., et al. 2017, *ApJ*, **843**, 76
- Arnaud, K., Dorman, B., & Gordon, C. 1999, *Astrophysics Source Code Library* [record ascl:9910.005]
- Arnaud, M., Pointecouteau, E., & Pratt, G. W. 2007, *A&A*, **474**, L37
- Balkema, A. A., Haan, L., & d., 1974, *Ann. Probab.*, **2**, 792
- Behrens, C. N., Lopes, H. F., & Gamerman, D. 2004, *Stat. Modell.*, **4**, 227
- Boubel, P., Colless, M., Said, K., & Staveley-Smith, L. 2025, *J. Cosmol. Astropart. Phys.*, **2025**, 066
- Braspenning, J., Schaye, J., Schaller, M., et al. 2024, *MNRAS*, **533**, 2656
- Bulbul, E., Liu, A., Kluge, M., et al. 2024, *A&A*, **685**, A106
- Bull, P., Akrami, Y., Adamek, J., et al. 2016, *Phys. Dark Universe*, **12**, 56
- Carrick, J., Turnbull, S. J., Lavaux, G., & Hudson, M. J. 2015, *MNRAS*, **450**, 317
- Chang, Z., & Lin, H.-N. 2015, *MNRAS*, **446**, 2952
- Chen, Y., Reiprich, T. H., Böhringer, H., Ikebe, Y., & Zhang, Y. Y. 2007, *A&A*, **466**, 809
- Cowell, J. A., Dhawan, S., & Macpherson, H. J. 2023, *MNRAS*, **526**, 1482
- Croston, J. H., Pratt, G. W., Böhringer, H., et al. 2008, *A&A*, **487**, 431
- da Silveira Ferreira, P., & Marra, V. 2024, *J. Cosmol. Astropart. Phys.*, **2024**, 077
- Dai, D.-C., Kinney, W. H., & Stojkovic, D. 2011, *J. Cosmol. Astropart. Phys.*, **2011**, 015
- Dam, L., Lewis, G. F., & Brewer, B. J. 2023, *MNRAS*, **525**, 231
- Davison, A. C., & Smith, R. L. 1990, *J. R. Stat. Soc., B: Stat. Methodol.*, **52**, 393
- DES Collaboration (Abbott, T. M. C., et al.) 2022, *Phys. Rev. D*, **105**, 023520
- Eckmiller, H. J., Hudson, D. S., & Reiprich, T. H. 2011, *A&A*, **535**, A105
- Elahi, P. J., Cañas, R., Poulton, R. J. J., et al. 2019, *PASA*, **36**, e021
- Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, *Rev. Mex. Astron. Astrofis.*, **53**, 385
- Ferreira, P., & d. S. & Quartin, M., 2021, *Phys. Rev. Lett.*, **127**, 101301
- Fixsen, D. J., Cheng, E. S., Gales, J. M., et al. 1996, *ApJ*, **473**, 576
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, *A&A*, **689**, A298
- Giles, P. A., Romer, A. K., Wilkinson, R., et al. 2022, *MNRAS*, **516**, 3878
- Giodini, S., Lovisari, L., Pointecouteau, E., et al. 2013, *Space Sci. Rev.*, **177**, 247
- Goodman, J., & Weare, J. 2010, *Commun. Appl. Math. Comput. Sci.*, **5**, 65
- Haridasu, B. S., Salucci, P., & Sharma, G. 2024, *MNRAS*, **532**, 2234
- Hoffman, Y., Valade, A., Libeskind, N. I., et al. 2024, *MNRAS*, **527**, 3788
- Horner, D. J. 2001, Ph.D. Thesis, University of Maryland College Park
- Hu, J. P., Wang, Y. Y., Hu, J., & Wang, F. Y. 2024a, *A&A*, **681**, A88
- Hu, C., Swallow, B., & Castro-Camilo, D. 2024b, arXiv e-prints [arXiv:2401.15703]
- Joudaki, S., Blake, C., Heymans, C., et al. 2017, *MNRAS*, **465**, 2033
- Kashlinsky, A., Atrio-Barandela, F., Ebeling, H., Edge, A., & Kocevski, D. 2010, *ApJ*, **712**, L81
- Kitzbichler, M. G., & White, S. D. M. 2007, *MNRAS*, **376**, 2
- Koulouridis, E., Clerc, N., Sadibekova, T., et al. 2021, *A&A*, **652**, A12
- Kugel, R., Schaye, J., Schaller, M., et al. 2024, *MNRAS*, **534**, 2378
- Lovisari, L., Reiprich, T., & Schellenberger, G. 2015, *A&A*, **573**, A118
- Lovisari, L., Forman, W. R., Jones, C., et al. 2017, *ApJ*, **846**, 51
- Lovisari, L., Schellenberger, G., Sereno, M., et al. 2020, *ApJ*, **892**, 102
- Luongo, O., Muccino, M., Colgáin, E. Ó., Sheikh-Jabbari, M. M., & Yin, L. 2022, *Phys. Rev. D*, **105**, 103510
- Maughan, B. J. 2007, *ApJ*, **668**, 772
- Mazzotta, P., Rasia, E., Moscardini, L., & Tormen, G. 2004, *MNRAS*, **354**, 10
- Mc Conville, R., Colgáin, Ó., & E., 2023, *Phys. Rev. D*, **108**, 123533
- McCarthy, I. G., Salcido, J., Schaye, J., et al. 2023, *MNRAS*, **526**, 5494
- McGibbon, R., Helly, J. C., Schaye, J., Schaller, M., & Vandenbroucke, B. 2025, *JOSS*, **10**, 8252
- Mehrtens, N., Romer, A. K., Hilton, M., et al. 2012, *MNRAS*, **423**, 1024
- Migkas, K. 2025, *Philos. Trans. R. Soc. A*, **383**, 20240030
- Migkas, K., & Reiprich, T. H. 2018, *A&A*, **611**, A50
- Migkas, K., Schellenberger, G., Reiprich, T. H., et al. 2020, *A&A*, **636**, A15
- Migkas, K., Pacaud, F., Schellenberger, G., et al. 2021, *A&A*, **649**, A151
- Migkas, K., Kox, D., Schellenberger, G., et al. 2024, *A&A*, **688**, A107
- Mittal, V., Oayda, O. T., & Lewis, G. F. 2024, *MNRAS*, **527**, 8497
- Mohayaee, R., Rameez, M., & Sarkar, S. 2021, *Eur. Phys. J. Spec. Top.*, **230**, 2067
- Morandi, A., Ettori, S., & Moscardini, L. 2007, *MNRAS*, **379**, 518
- Nagarajan, A., Pacaud, F., Sommer, M., et al. 2019, *MNRAS*, **488**, 1728
- Nusser, A., & Davis, M. 2011, *ApJ*, **736**, 93
- Oayda, O. T., & Lewis, G. F. 2023, *MNRAS*, **523**, 667
- Pandya, A., Migkas, K., Reiprich, T. H., et al. 2024, *A&A*, **691**, A355
- Perivolaropoulos, L. 2023, *Phys. Rev. D*, **108**, 063509
- Perivolaropoulos, L., & Skara, F. 2022, *New Astron. Rev.*, **95**, 101659
- Philcox, O. H. E., & Ivanov, M. M. 2022, *Phys. Rev. D*, **105**, 043517
- Pickands, J. 1975, *Ann. Stat.*, **3**, 119
- Planck Collaboration XI. 2011, *A&A*, **536**, A11
- Planck Collaboration XXVII. 2014, *A&A*, **571**, A27
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Ploeckinger, S., & Schaye, J. 2020, *MNRAS*, **497**, 4857
- Pratt, G. W., Croston, J. H., Arnaud, M., & Böhringer, H. 2009, *A&A*, **498**, 361
- Qin, F., Parkinson, D., Howlett, C., & Said, K. 2021, *ApJ*, **922**, 59
- Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, *ApJ*, **876**, 85
- Riess, A. G., Yuan, W., Macri, L. M., et al. 2022, *ApJ*, **934**, L7
- Rubart, M., & Schwarz, D. J. 2013, *A&A*, **555**, A117
- Salehi, A., Yarahmadi, M., Fathi, S., & Bamba, K. 2021, *MNRAS*, **504**, 1304
- Sasaki, S. 1996, *PASJ*, **48**, L119
- Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, **526**, 4978
- Schellenberger, G., Reiprich, T. H., Lovisari, L., Nevalainen, J., & David, L. 2015, *A&A*, **575**, A30
- Secrest, N. J., von Hausegger, S., Rameez, M., et al. 2021, *ApJ*, **908**, L51
- Secrest, N. J., von Hausegger, S., Rameez, M., Mohayaee, R., & Sarkar, S. 2022, *ApJ*, **937**, L31
- Shah, P., Lemos, P., & Lahav, O. 2021, *A&ARv*, **29**, 9
- Smith, A., Cole, S., Baugh, C., et al. 2017, *MNRAS*, **470**, 4646
- Tiwari, P., Kothari, R., Naskar, A., Nadkarni-Ghosh, S., & Jain, P. 2015, *Astropart. Phys.*, **61**, 1
- Tiwari, P., Schwarz, D. J., Zhao, G.-B., et al. 2024, *ApJ*, **975**, 279
- Valentino, E. D., Mena, O., Pan, S., et al. 2021, *Class. Quant. Grav.*, **38**, 153001
- Wagenveld, J. D., Klöckner, H.-R., Gupta, N., et al. 2024, *A&A*, **690**, A163
- Watkins, R., Allen, T., Bradford, C. J., et al. 2023, *MNRAS*, **524**, 1885
- Whitford, A. M., Howlett, C., & Davis, T. M. 2023, *MNRAS*, **526**, 3051
- Zhang, Y. Y., Finoguenov, A., Böhringer, H., et al. 2008, *A&A*, **482**, 451

Appendix A: Two 1 Gpc runtime lightcones

Applying the same analysis to the two runtime lightcones from the L1_m9 run, we find a maximum dipole of $3.15 \pm 0.78\%$ at 4.02σ in lightcone 1, and $2.50 \pm 0.71\%$ at 3.55σ in lightcone 2, using the M21 method (see Fig. A.1). These runtime lightcones are consistent with the 1728 post-processed lightcones analysed in the main text. No detection exceeds 3σ from the main probability distribution of the 1728 samples. This supports the robustness of our post-processed lightcones.

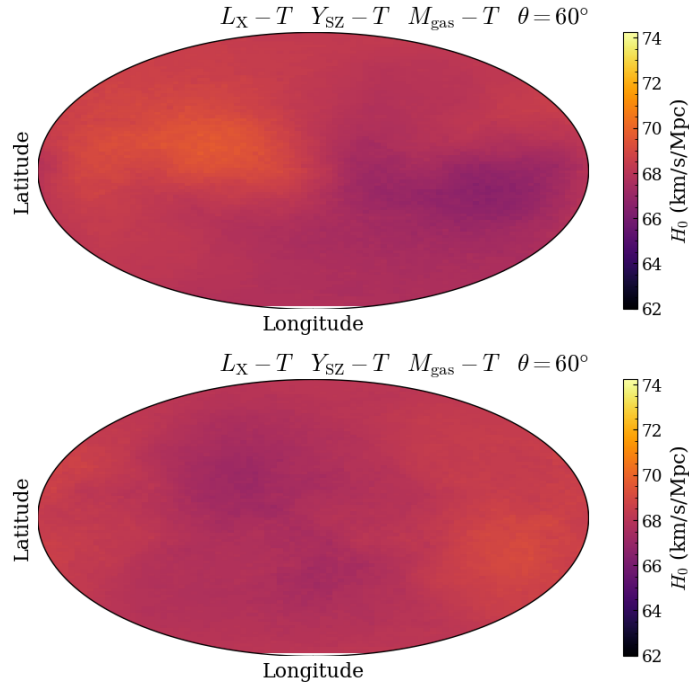


Fig. A.1. H_0 maps from the M21 method for the two runtime lightcones in the 1.0 Gpc fiducial run (L1_m9), using the combined L_X-T , $Y_{SZ}-T$, and $M_{\text{gas}}-T$ relations. Lightcone 1 (top) and lightcone 2 (bottom).

Appendix B: Details on the fiducial mass cut

To reduce computational cost, we have applied a fiducial mass cut of $M_{500c} > 10^{13} M_{\odot}$ (Sect. 2.1). We verify here that this selection does not exclude clusters that would otherwise satisfy the flux threshold. Using the publicly available eRASS1 catalogue (Bulbul et al. 2024), we apply the same flux selection of $f_{X,0.2-2.3\text{keV}} > 5 \times 10^{-12} \text{erg s}^{-1} \text{cm}^{-2}$. The resulting mass distribution is shown in Fig. B.1. Out of 12247 clusters detected by eROSITA, 152 meet the flux threshold, none of which have $M_{500c} < 10^{13} M_{\odot}$.

On the simulation side, applying the same flux threshold to our lightcones results in, on average, fewer than five clusters within the range $10^{13} M_{\odot} < M_{500c} < 10^{13.5} M_{\odot}$ (see Fig. D.2). These systems lie at the low-mass tail. Thus, even if the mass cut excludes a few rare clusters that could surpass the flux threshold, they form an extremely small subset and do not contribute meaningfully. We conclude that the mass cut does not introduce a selection bias.

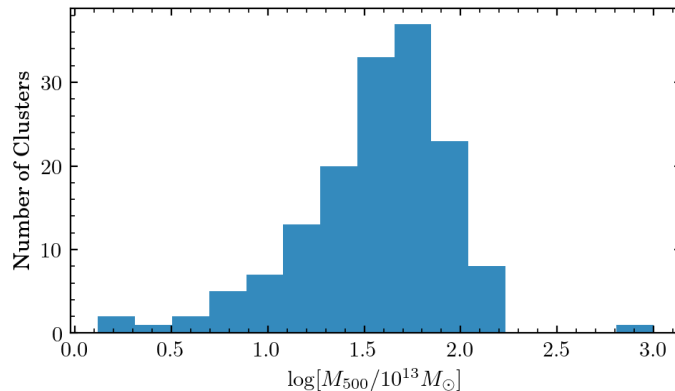


Fig. B.1. Mass distribution of eRASS1 clusters satisfying $f_{X,0.2-2.3\text{keV}} > 5 \times 10^{-12} \text{erg s}^{-1} \text{cm}^{-2}$. None fall below the fiducial mass cut of $10^{13} M_{\odot}$.

Appendix C: Correlation of results using different relations: angular separation

See Fig. C.1. We compute the angular separation on the sphere between best-fit dipole directions derived from different scaling relations. Comparing with Fig. E.1, one observes that greater scatter correlation results in closer dipole alignment. Relations with higher Pearson correlation coefficients yield smaller angular separations in our lightcones. Even for the best-aligned pair, L_X-T and $M_{\text{gas}}-T$ (via MCMC), only 77% of lightcones have separation angle $\theta < 60^\circ$. This drops to 56% for $Y_{\text{SZ}}-T$ and $M_{\text{gas}}-T$ and 39% for L_X-T and $Y_{\text{SZ}}-T$. The figure also confirms consistency between the two statistical methods: the choice of relation affects the dipole direction more strongly than the method used. Comparing the bottom panel and the blue curves, no significant change is seen when scatter is injected. This is expected—no intrinsic signal is present, and the injected scatter has no cross-relation correlation between L_X-T and $Y_{\text{SZ}}-T$ (see Appendix E).

For uncorrelated directions, the expected angular separation distribution is $P(\theta) = \frac{1}{2} \sin \theta$. This follows from uniform sampling on the unit sphere: the probability of a second point lying between angles θ and $\theta + d\theta$ from a fixed point is proportional to the area of the corresponding spherical ring, which scales as $2\pi \sin \theta d\theta$. The normalisation factor $1/2$ ensures unit probability.

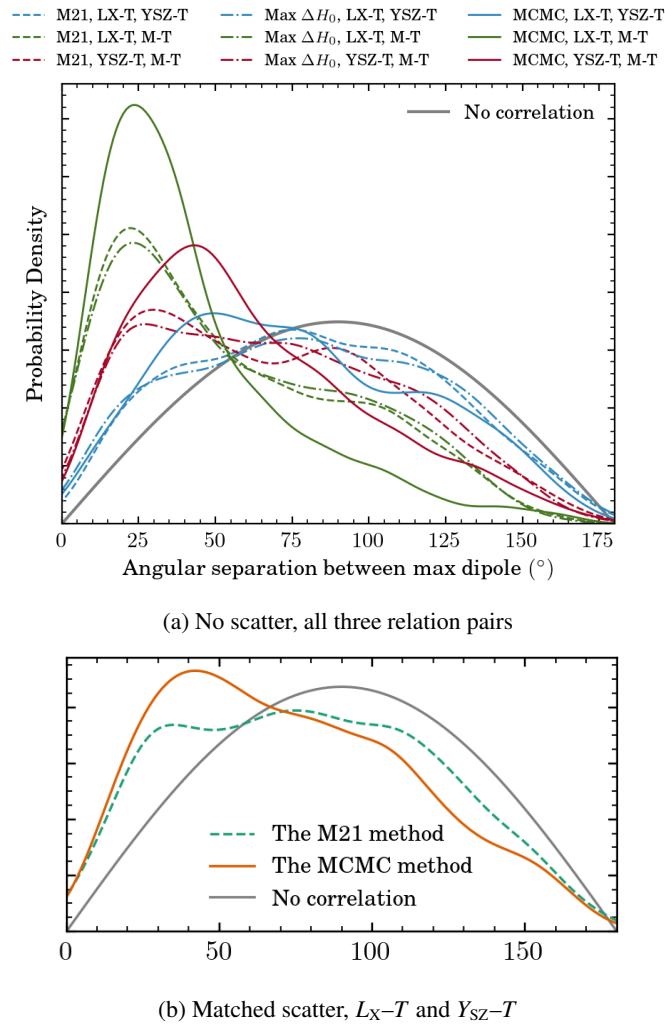


Fig. C.1. Angular separation between the best-fit dipoles derived from different relations, without (top) and with (bottom) injected scatter. Linestyles indicate statistical methods: solid for MCMC; dashed and dash-dotted for the M21 method, differing slightly in implementation. Dashed lines define the dipole by highest significance (our default for the M21 method); dash-dotted lines by highest amplitude. The two give nearly identical results. The grey line indicates the expected distribution for uncorrelated dipoles.

Appendix D: Sample comparison between FLAMINGO and M21

D.1. X-ray concentration

The distribution of c_X , defined as the fraction of emission coming from the central 15% of the cluster's radius ($L_{r < 0.15R_{500c}} / L_{r < R_{500c}}$), is an easily estimated metric that helps us identify potentially relaxed clusters with high central emission. Large c_X values often correlate with cool-core presence in clusters (Andrade-Santos et al. 2017; Kugel et al. 2024). c_X exhibits the most prominent difference between FLAMINGO and the M21 sample (see Fig. D.1). Kugel et al. (2024) investigated the distribution of the same quantity and found the same behaviour when the projection effect was not taken into account. However, Fig. D.1 shows that the difference is

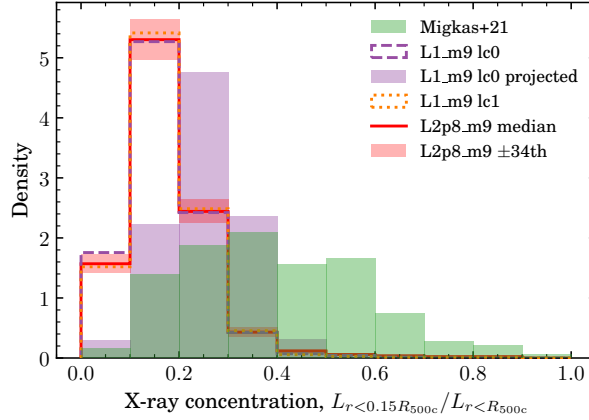


Fig. D.1. X-ray concentration distribution of 1728 (2.8 Gpc)³ lightcones (red), two 1.0 Gpc runtime lightcones (purple dashed, orange dotted), and the M21 sample (green). Purple bars show the 2D projected concentration of a 1.0 Gpc lightcone.

much reduced, but not eliminated, by the projection effect. The projected L_X are calculated by summing particles inside cylindrical regions instead of spherical ones. The FLAMINGO samples' projected c_X is centred at ~ 0.25 whereas the M21 sample's distribution is centred at $c_X \sim 0.35$. There are few FLAMINGO clusters with $c_X > 0.35$. This might be caused by an under-representation of high X-ray concentration clusters in FLAMINGO in addition to non-detections of diffuse, low c_X clusters in X-ray flux-limited samples, especially close to the flux limit. Moreover, highly disturbed clusters (expected to return lower c_X values) were removed from the M20/M21 sample. Nevertheless, cluster samples selected through their SZ signal (e.g. Lovisari et al. 2017) also show a relatively higher c_X distribution than FLAMINGO. In any case, further investigation of this discrepancy is needed.

D.2. Cluster properties comparison

Fig. D.2 summarises the distributions of f_X , L_X , T_{Cl} , M_{500c} , Y_{SZ} , and M_{gas} for FLAMINGO, the runtime lightcones, and M21. Simulated samples are selected as the top 313 X-ray concentrated clusters to match the number of available M21 entries; results are robust to changing this number to 250–400. Simulated L_X values are in the original SOAP band (0.2–2.3 keV; see Sect. 2.3.2), with M21 values converted accordingly. Agreement between FLAMINGO and M21 is generally good across properties, though T_{Cl} is systematically lower in FLAMINGO (see Sect. 2.3.5). In addition, L_X , f_X , and Y_{SZ} in FLAMINGO also tend to lie slightly on the lower end, which may be partly due to an under-representation of high-redshift clusters. The runtime lightcones are consistent with the simulated samples.

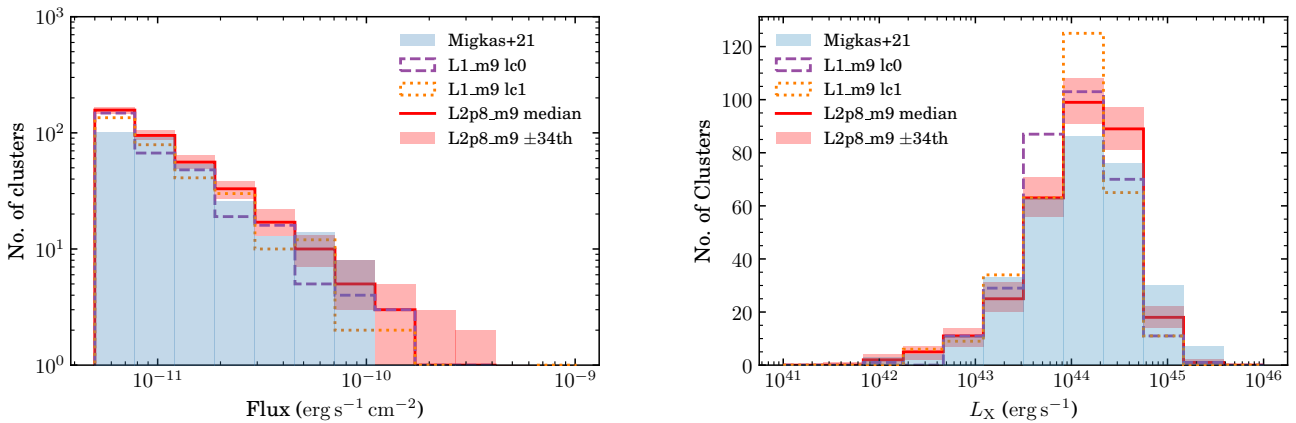


Fig. D.2. Distribution of various cluster properties of the simulated samples, runtime lightcones, and the M21 sample, where available (continued next page).

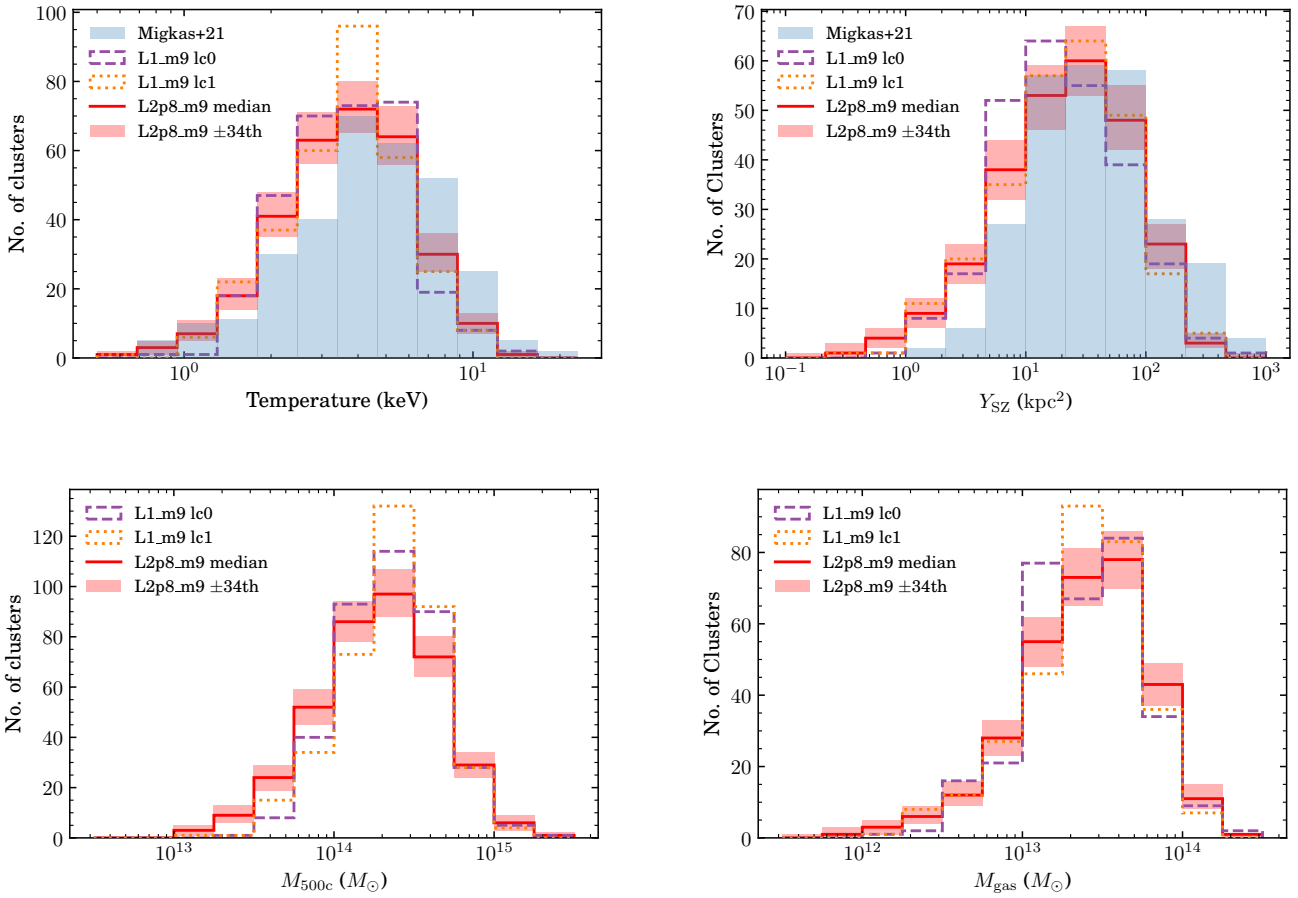


Fig. D.2. Continued.

Appendix E: Scatter correlation between the relations

Scatter is computed as $\log Y - \log Y_{\text{fit}}$, where $\log Y_{\text{fit}} = B \log X/C_X + \log A + \log C_Y - \gamma_{YX} \log E(z)$. We evaluate scatter for L_X-T , $Y_{\text{SZ}}-T$, and $M_{\text{gas}}-T$, and fit a linear relation between each pair across all 1728 lightcones. The resulting scatter points and fits are shown in Fig. E.1, with Pearson's correlation coefficient r labelled in each panel. The uncertainty is estimated from the standard deviation of the fit results across all lightcones.

Using T_{Cl} as the common variable for comparison with M21, we find $r = 0.25 \pm 0.04$ between L_X-T and $Y_{\text{SZ}}-T$ in FLAMINGO. This is significantly lower than the $r = 0.668 \pm 0.089$ found in M21. To assess the effect of injected scatter, we repeated the analysis with artificial scatter applied. Since the injection procedure does not induce cross-correlation, the result remains similar: $r = 0.31 \pm 0.03$. Nagarajan et al. (2019) reported $r = 0.47^{+0.24}_{-0.35}$ for the L_X-M_{500} and $Y_{\text{SZ}}-M_{500}$ relations at fixed mass. Such correlation could arise from baryonic physics linking X-ray and SZ emission, which may not be fully captured in our modelling.

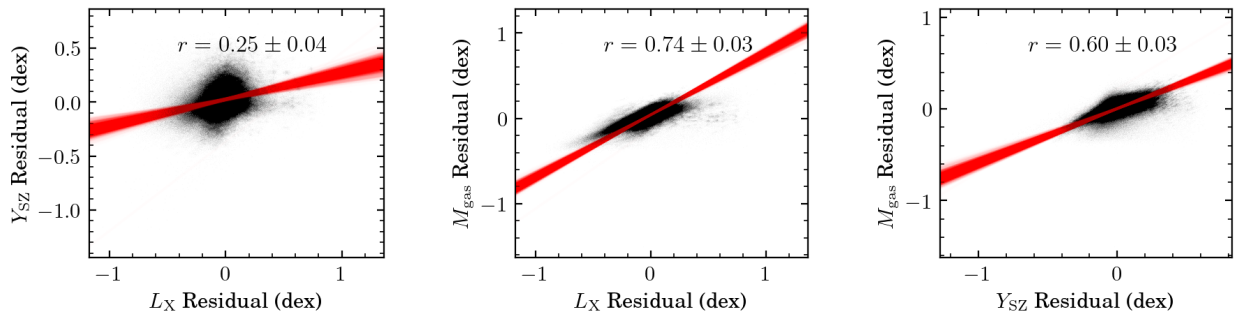


Fig. E.1. Correlation of scatter between L_X-T and $Y_{\text{SZ}}-T$ (left), L_X-T and $M_{\text{gas}}-T$ (middle) and $Y_{\text{SZ}}-T$ and $M_{\text{gas}}-T$ (right). Black dots are individual clusters. Red lines represent the linear fit of the lightcones. All of the 1728 lightcones are overlaid. r is Pearson's correlation coefficient.

In contrast, we find stronger correlations between L_X and M_{gas} , and between Y_{SZ} and M_{gas} at fixed T , with $r = 0.74 \pm 0.03$ and $r = 0.60 \pm 0.03$, respectively. Future studies using M_{gas} -based relations should be cautious when combining observables or constructing mock samples.

Appendix F: Correlation of anisotropy using the M21 or MCMC method

The correlation of amplitude ΔH_0 and statistical significance n_σ between the M21 method and the MCMC approach is shown in Fig. F.1. A uniform mild correlation of $r \sim 0.4$ – 0.5 is observed across all six comparisons, covering the three scaling relations and both anisotropy metrics. This test is carried out without injected scatter. The scatter around the one-to-one line suggests that the two methods do not fully agree on the strength of anisotropy for individual realisations, but they tend to rank lightcones similarly in terms of dipole strength. This level of agreement supports the overall consistency between methods, while highlighting that method-dependent systematics could still influence the detailed interpretation of individual detections.

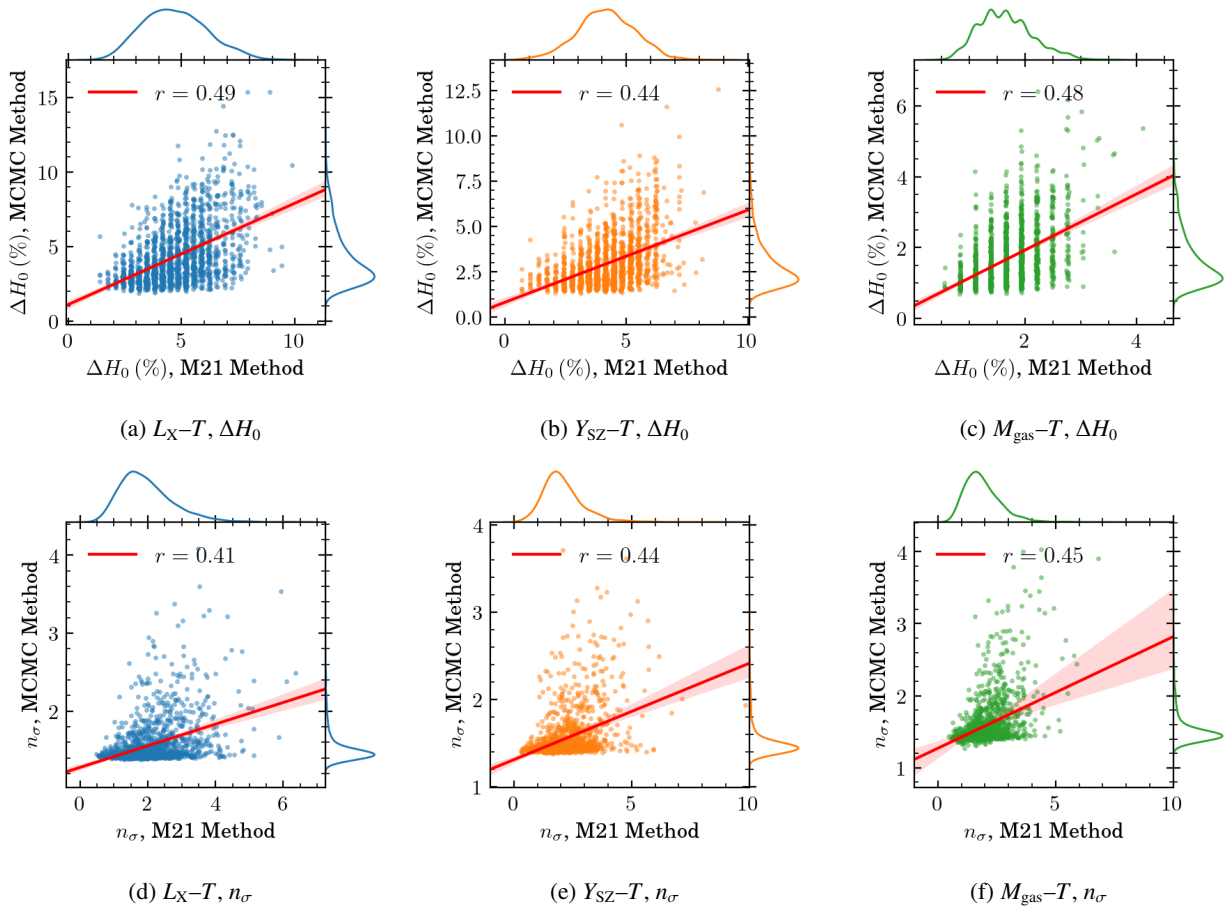


Fig. F.1. Comparing ΔH_0 (top) and n_σ (bottom) predicted by the MCMC method versus the M21 method, using the L_X-T (left), $Y_{\text{SZ}}-T$ (middle), or $M_{\text{gas}}-T$ (right) relation. Each data point represents the maximum dipole ΔH_0 or n_σ on a lightcone. The red line shows linear fit with error. r represents the Pearson correlation coefficient.

Appendix G: Extreme value statistics

Here we show the EVS estimations for other than the joint H_0 (Fig. 5). Every entry of EVS p -values in Table 3 and 4 is accompanied by a figure. For Table 3 the per relation EVS distributions are presented in Fig. G.1. For the L_X-T , M21 or MCMC method entries the reference point from M21 is $(\Delta H_0, n_\sigma) = (8.7\%, 2.8\sigma)$ or $(8.7\%, 2.4\sigma)$. For $Y_{\text{SZ}}-T$ we compare to $(14\%, 4.1\sigma)$ with the M21 method or $(14\%, 2.6\sigma)$ with MCMC. Same as in Fig. 5 the tension is greatly relieved by increased scatter, where the projected x value (Sect. 3.4) is shifted closer to the M21 points and with increased scatter (see Fig. G.1f in comparison to Fig. G.1b, for example).

The joint EVS analyses for bulk flow (Table 4) are summarised in Fig. G.2, and results for individual relations in Fig. G.3. As with H_0 , the projected distributions shift upward when scatter is added, though the effect is visibly weaker. In particular, the tension in L_X-T remains largely unchanged.

Other than $x = \Delta H_0 - \Delta H_0/n_\sigma$, several other projection functions were tested for the H_0 matched scatter analysis (see Table G.1). We see that the choice of projection has a noticeable effect on the results. The choice of projection has a clear impact on the

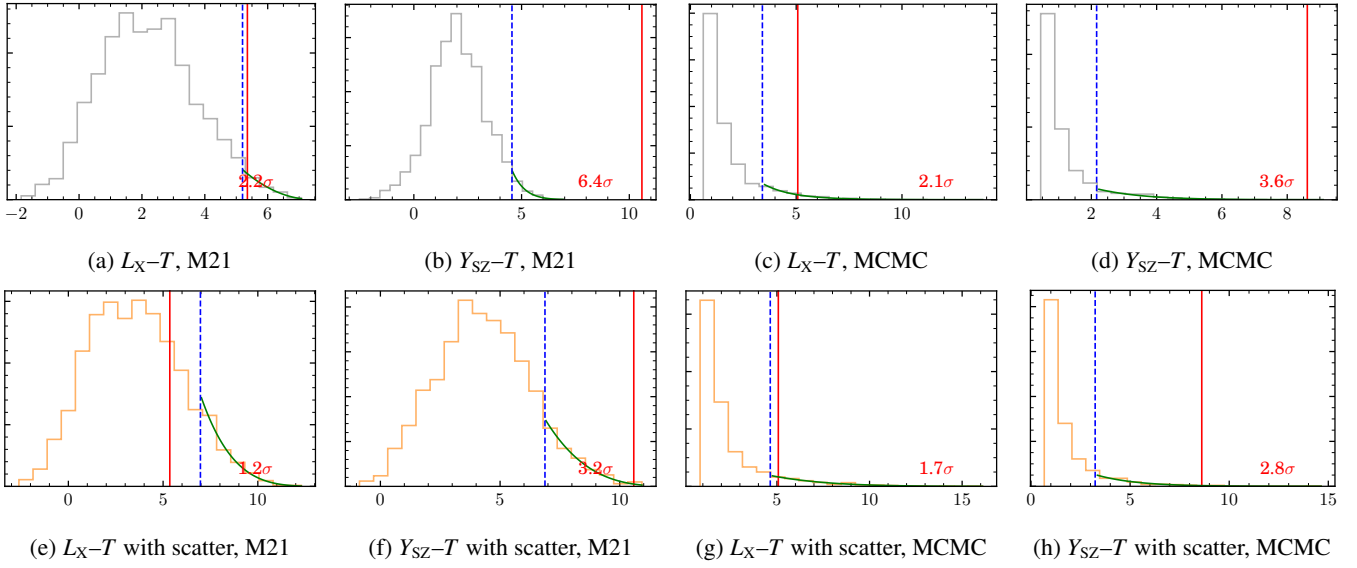


Fig. G.1. H_0 anisotropy EVS probability distributions for individual scaling relations. Legend as in Fig. 5. Corresponding p -values are listed in Table 3.

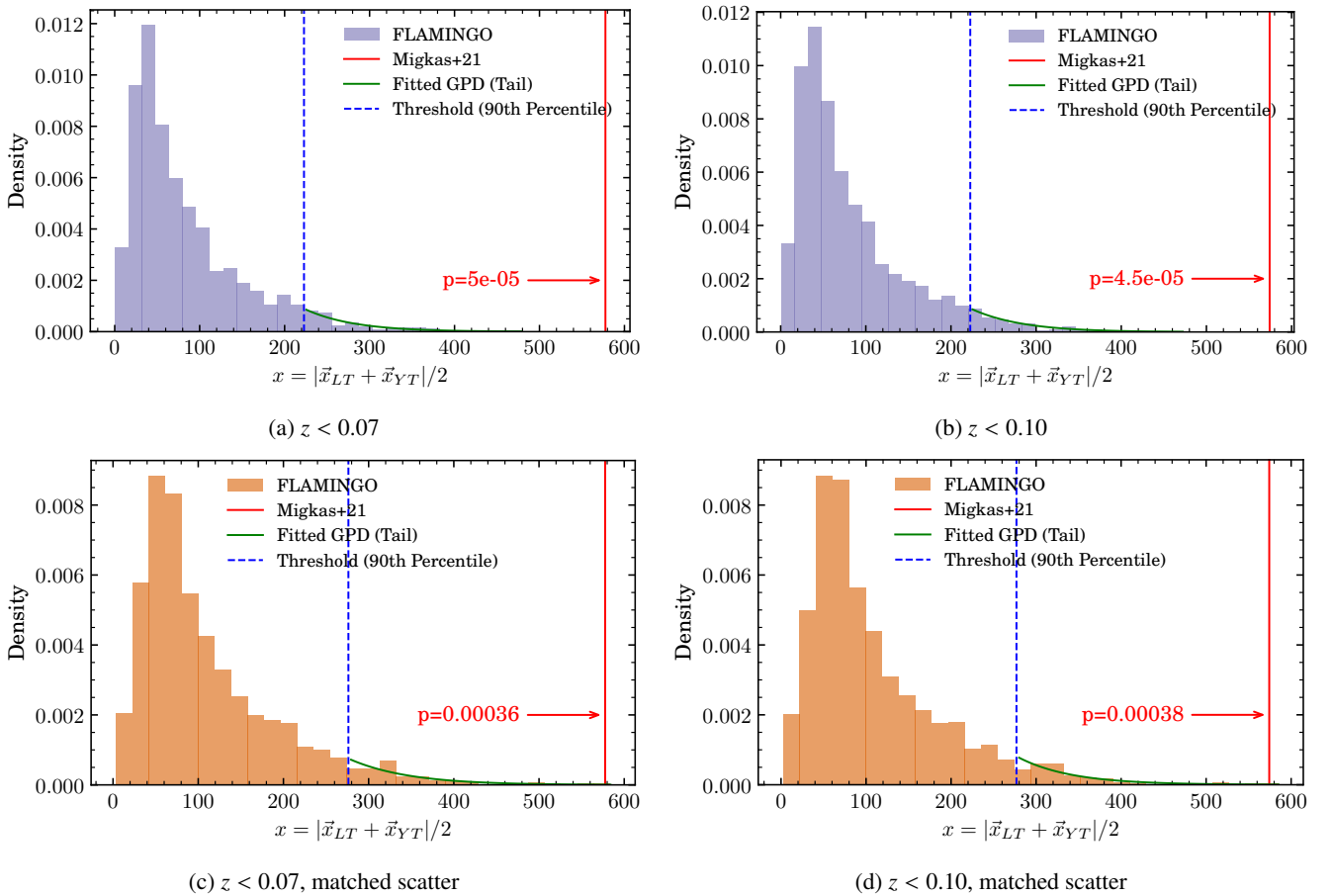


Fig. G.2. Joint EVS analyses for bulk flows, before (top) and after (bottom) matching scatter. Legend as in Fig. 5. Corresponding p -values are listed in Table 3.

resulting p -values, underscoring the need for a physically motivated and consistent definition. Our adopted metric, the lower one-sigma bound, was chosen to incorporate both anisotropy amplitude and uncertainty in a single conservative estimate. It ensures that a modest but well-constrained dipole (e.g. $5 \pm 1\%$) is treated comparably to a larger but poorly constrained one (e.g. $15 \pm 11\%$), avoiding overinterpretation of noisy outliers.

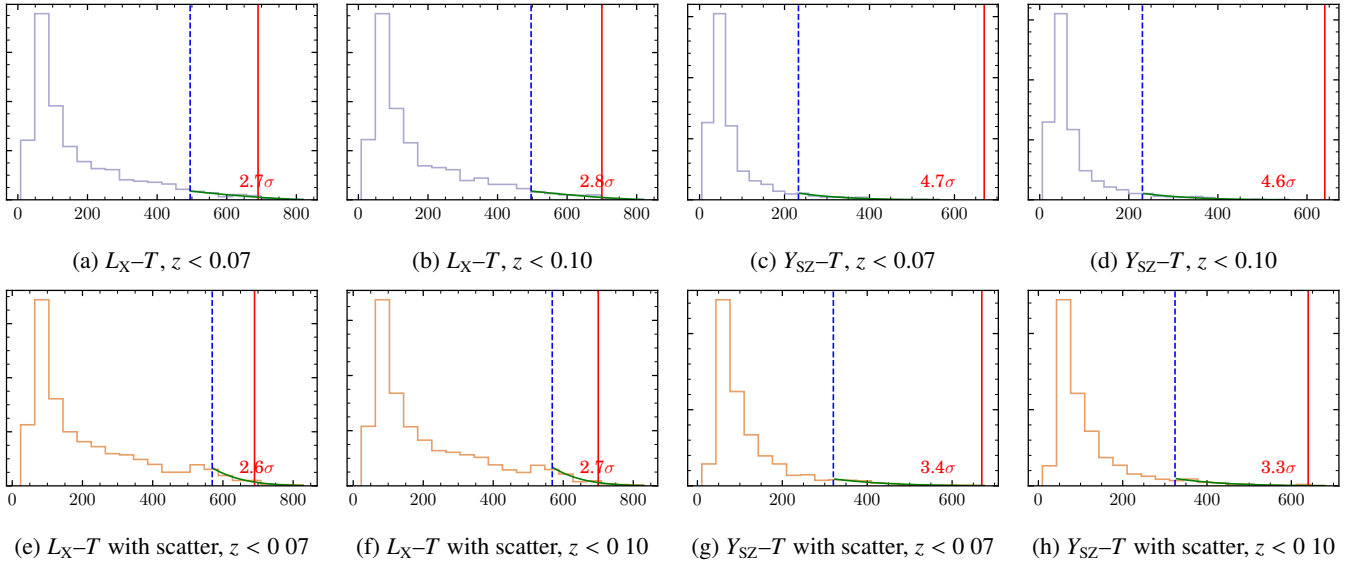


Fig. G.3. Bulk flow EVS probability distributions for individual scaling relations. Legend as in Fig. 5. Corresponding results are summarised in Table 4.

Table G.1. Effect of different projection functions on the EVS p -values.

Projection	M21 L_X-T	M21 $Y_{SZ}-T$	M21 joint	MCMC L_X-T	MCMC $Y_{SZ}-T$	MCMC joint
$\Delta H_0 - \Delta H_0/n_\sigma$	0.231 (1.2 σ)	0.00129 (3.22 σ)	0.00118 (3.24 σ)	0.0848 (1.72 σ)	0.00487 (2.82 σ)	0.0205 (2.32 σ)
$\Delta H_0/4 + n_\sigma$	0.232 (1.2 σ)	0.0205 (2.32 σ)	0.00619 (2.74 σ)	0.0877 (1.71 σ)	0.00576 (2.76 σ)	0.00079 (3.36 σ)
$\Delta H_0/2 + n_\sigma$	0.250 (1.15 σ)	0.0139 (2.46 σ)	0.00486 (2.82 σ)	0.114 (1.58 σ)	0.00447 (2.84 σ)	0.00491 (2.81 σ)
$\Delta H_0 + \Delta H_0/n_\sigma$	0.394 (0.85 σ)	0.000297 (3.62 σ)	0.00235 (3.04 σ)	0.237 (1.18 σ)	0.00251 (3.02 σ)	0.142 (1.47 σ)