Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **516,** 437–452 (2022)
Advance Access publication 2022 August 4

https://doi.org/10.1093/mnras/stac2196

Forecasting cosmological parameter constraints using multiple sparsity measurements as tracers of the mass profiles of dark matter haloes

P. S. Corasaniti,[1,2]★ A. M. C. Le Brun,[1,3]★ T. R. G. Richardson,[1] Y. Rasera,[1] S. Ettori,[4,5] M. Arnaud[3] and G. W. Pratt[3]

[1]*Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, F-92190 Meudon, France*
[2]*Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, F-75014 Paris, France*
[3]*AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Cité, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France*
[4]*INAF, Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Piero Gobetti 93/3, I-40129 Bologna, Italy*
[5]*INFN, Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy*

Accepted 2022 August 1. Received 2022 August 1; in original form 2022 April 14

ABSTRACT

The dark matter halo sparsity, i.e. the ratio between spherical halo masses enclosing two different overdensities, provides a non-parametric proxy of the halo mass distribution that has been shown to be a sensitive probe of the cosmological imprint encoded in the mass profile of haloes hosting galaxy clusters. Mass estimations at several overdensities would allow for multiple sparsity measurements, which can potentially retrieve the entirety of the cosmological information imprinted on the halo profile. Here, we investigate the impact of multiple sparsity measurements on the cosmological model parameter inference. For this purpose, we analyse *N*-body halo catalogues from the Raygal and M2Csims simulations and evaluate the correlations among six different sparsities from spherical overdensity halo masses at $\Delta = 200$, 500, 1000, and 2500 (in units of the critical density). Remarkably, sparsities associated to distinct halo mass shells are not highly correlated. This is not the case for sparsities obtained using halo masses estimated from the Navarro-Frenk-White (NFW) best-fitting profile, which artificially correlates different sparsities to order one. This implies that there is additional information in the mass profile beyond the NFW parametrization and that it can be exploited with multiple sparsities. In particular, from a likelihood analysis of synthetic average sparsity data, we show that cosmological parameter constraints significantly improve when increasing the number of sparsity combinations, though the constraints saturate beyond four sparsity estimates. We forecast constraints for the CHEX-MATE cluster sample and find that systematic mass bias errors mildly impact the parameter inference, though more studies are needed in this direction.

Key words: methods: numerical – galaxies: clusters: general – large-scale structure of Universe.

1 INTRODUCTION

At the time of writing, a heavy focus within the field of precision cosmology is set on constraining the parameters of the ΛCDM cosmological model, named according to its two main constituents: a cosmological constant Λ and Cold Dark Matter (CDM), while also exploring possible extensions to this model by investigating alternative scenarios of dark matter (see e.g. Boyarsky et al. 2019; Niemeyer 2020; Green & Kavanagh 2021) and dark energy (see e.g. Copeland, Sami & Tsujikawa 2006; Brax 2018).

To this effect, many probes have been devised and applied to a range of observations such as cosmic microwave background (CMB) experiments (e.g Fixsen et al. 1996; Komatsu et al. 2011; Planck Collaboration VI 2020), big bang Nucleo-synthesis estimates (e.g. Aver, Olive & Skillman 2015; Cooke, Pettini & Steidel 2018), large-scale structure observations through measurements of the clustering of matter measured with various probes, such as galaxies (see e.g. Percival et al. 2001; Tegmark et al. 2004; Cole et al. 2005; Beutler et al. 2017) or Lyman-α absorbing gas (Croft et al. 2016), gravitational lensing (e.g. Birrer et al. 2020; Wong et al. 2020; Gatti et al. 2021), measurements of the baryonic acoustic oscillation (BAO) scale (e.g. de Sainte Agathe et al. 2019; de Mattia et al. 2021; DES Collaboration 2022), and tests of the Hubble diagram from SN Ia standard-candles (e.g. Anand et al. 2022; Riess et al. 2022) to cite but a few examples.

In this regard, galaxy clusters have proven to be a useful asset. These structures, the most massive gravitationally bound in the Universe, exhibit multiple properties that can be used to test the cosmological paradigm. As an example, cosmological constraints have been inferred from measurements of their abundance (e.g. Planck Collaboration XX 2014a, Planck Collaboration XXIV 2016b; Pacaud et al. 2018; Bocquet et al. 2019; Lesci et al. 2022; To et al. 2021), their spatial clustering (Planck Collaboration XXI 2014b; Marulli et al. 2021) or the fraction of gas contained inside their potential wells (e.g. Ettori, Tozzi & Rosati 2003; Allen et al. 2008; Ettori et al. 2009; Mantz et al. 2014, 2022).

The possibility of extracting cosmological information from estimates of the mass profiles of galaxy clusters is another probes that, however, remains relatively unexplored (see e.g. Ettori et al. 2010).

★ E-mail: Pier-Stefano.Corasaniti@obspm.fr (PSC); amandine.le-brun@obspm.fr (AMCLB)

Historically, this approach has been built upon the remarkable result that density profiles of dark matter haloes from *N*-body simulation are described to a good approximation by a two-parameter universal function, the Navarro-Frenk-White profile (NFW; Navarro, Frenk & White 1997). However, the difficulty of obtaining accurate estimates of the concentration–mass relations from galaxy cluster observations has so far been the main limitation to the use of cluster mass profiles as cosmological proxy (see e.g. Mead et al. 2010; King & Mead 2011; Sereno et al. 2015).

In recent years, a novel approach in this direction has been developed around the concept of halo *sparsity*, which is the ratio of halo masses measured at radii enclosing different overdensities, as a non-parametric proxy for the internal halo mass distribution. In the seminal work of Balmès et al. (2014), it has been shown that halo sparsity depends on the characteristics of the underling cosmological model. Further investigation by Corasaniti, Giocoli & Baldi (2020) has found that the average sparsity is also sensitive to modified gravity scenarios and can therefore be used to constrain the latter. Recently, cosmological constraints using measurements of the average halo sparsity of galaxy cluster samples have yielded results competitive with other widely used probes (Corasaniti et al. 2018; Corasaniti, Sereno & Ettori 2021).

The average halo sparsity has been shown to possess a number of interesting features (see e.g. Corasaniti et al. 2018, 2021; Corasaniti & Rasera 2019). On one hand, it provides a simple link between measurements of the mass profile of an ensemble of galaxy clusters and cosmological model predictions derived from an integral relation, involving the halo mass function at the overdensities of interest. On the other hand, being a mass-ratio, the average sparsity is less impacted by the systematic errors known to affect the measurements of galaxy cluster masses (see e.g. Nagai, Vikhlinin & Kravtsov 2007; Meneghetti et al. 2010; Rasia et al. 2012; Velliscig et al. 2014; Sereno & Ettori 2015; Biffi et al. 2016). Furthermore, the properties that characterize the halo sparsity are independent of the specific form of the halo density profile. As such, the use of multiple sparsity measurements from non-parametric mass estimates opens the way to retrieving cosmological information encoded over the entire halo mass profile rather than from a single determination at two particular overdensities. However, because of the gravitational assembly processes shaping the mass distribution of haloes, we can expect these different sparsities to be correlated.

Here, we set to evaluate the minimum number of multiple average sparsity estimates that sample the halo mass profile at different overdensities while providing maximal constraints on a set of cosmological parameters. For this purpose, we perform a thorough analysis of average sparsities and their correlations using halo catalogues from large-volume high-resolution *N*-body simulations. Building upon this numerical study, we perform a Markov Chain Monte Carlo likelihood analysis on synthetic data sets to investigate the level of cosmological parameter constraints that can be inferred from different combinations of average sparsity measurements under different fiducial cosmologies and cluster mass measurement error assumptions.

The paper is organized as follows. In Section 2, we introduce the basic concepts, describe the *N*-body simulations, and present the results of the analysis of numerical halo catalogues. In Section 3, we describe the cosmological parameter inference from multiple average sparsity measurements for two distinct synthetic data sets, while in Section 4, we present a parameter forecast analysis for a realistic galaxy cluster sample. Finally, in Section 5, we discuss the conclusions.

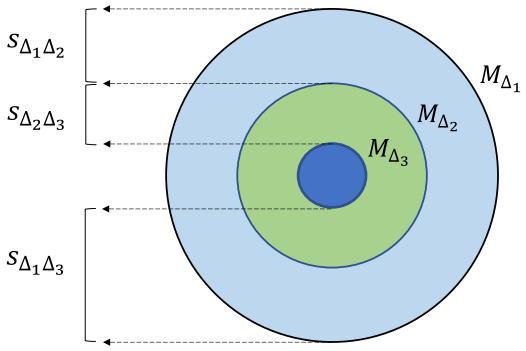

Figure 1. 2D illustration of the spherical mass profile of a halo as probed by multiple sparsity estimates at overdensities $\Delta_1 < \Delta_2 < \Delta_3$. Each sparsity s_{Δ_i,Δ_j} measures the fractional mass within the spherical shell comprised between r_{Δ_i} and r_{Δ_j} relative to the mass enclosed in the inner radius r_{Δ_j}.

2 COSMOLOGY WITH HALO SPARSITY

2.1 Definition and properties

Halo sparsity is defined as (Balmès et al. 2014):

$$s_{\Delta_1,\Delta_2} = \frac{M_{\Delta_1}}{M_{\Delta_2}}, \tag{1}$$

where M_{Δ_1} and M_{Δ_2} are halo masses at radii r_{Δ_1} and r_{Δ_2}, which enclose the overdensity Δ_1 and Δ_2, respectively, with $\Delta_1 < \Delta_2$ (with the overdensities in units of the critical ρ_c or background ρ_b density). This ratio can also be interpreted as the ratio of the mass ΔM_{12} within the radial shell $\Delta r = r_{\Delta_1} - r_{\Delta_2}$ and of the mass within the inner radius r_{Δ_2}, i.e. $s_{\Delta_1,\Delta_2} = \Delta M_{12}/M_{\Delta_2} + 1$. Hence, the values of sparsities at multiple overdensity pairs probe the fractional mass profile of the halo. As an example, in Fig. 1, we show a graphic illustration of the case of halo masses at overdensities Δ_1, Δ_2, and Δ_3, which allow to estimate three sparsity combinations s_{Δ_1,Δ_2}, s_{Δ_1,Δ_3}, and s_{Δ_2,Δ_3}.

Quite importantly, at any given redshift, the halo sparsity is largely independent of the outer halo mass M_{Δ_1} (Balmès et al. 2014; Corasaniti et al. 2018; Corasaniti & Rasera 2019); consequently, for a given pair of overdensities, the ensemble average value can be obtained by integrating the equality

$$\frac{\mathrm{d}n}{\mathrm{d}M_{\Delta_2}} = \frac{\mathrm{d}n}{\mathrm{d}M_{\Delta_1}} s_{\Delta_1,\Delta_2} \frac{\mathrm{d}\ln M_{\Delta_1}}{\mathrm{d}\ln M_{\Delta_2}}, \tag{2}$$

where $\mathrm{d}n/\mathrm{d}M_{\Delta_2}$ is the mass function at M_{Δ_2} of the ensemble of haloes with mass function $\mathrm{d}n/\mathrm{d}M_{\Delta_1}$ at M_{Δ_1} (i.e. mass functions of matched haloes), to obtain the average sparsity relation:

$$\int_{M_{\Delta_2}^{\rm min}}^{M_{\Delta_2}^{\rm max}} \frac{dn}{dM_{\Delta_2}} d\ln M_{\Delta_2} = \langle s_{\Delta_1,\Delta_2}\rangle \int_{\langle s_{\Delta_1,\Delta_2}\rangle M_{\Delta_2}^{\rm min}}^{\langle s_{\Delta_1,\Delta_2}\rangle M_{\Delta_2}^{\rm max}} \frac{dn}{dM_{\Delta_1}} d\ln M_{\Delta_1}. \tag{3}$$

Given the functional form of $\mathrm{d}n/\mathrm{d}M_{\Delta_1}$ and $\mathrm{d}n/\mathrm{d}M_{\Delta_2}$, the above equation can be solved numerically to obtain the value of the average sparsity. Corasaniti et al. (2018) has shown that equation (3) provides predictions of the average sparsity that are accurate to a few per cent level for $s_{200,\,500}$ and $s_{500,\,1000}$, thus providing the foundations to perform cosmological parameter inference using average sparsity measurements.

Notice that equation (3) is largely insensitive to the choice of the integration limits. Indeed, since at the high-mass end the mass function drops exponentially, the upper limit can be set to any

arbitrary large number; while at the low-mass end, given that the halo sparsity is nearly constant as a function of halo mass, the integral can be set without loss of generality to the minimum halo mass of the halo catalogues used for the calibration of the mass functions.

Hereafter, we will test the validity of equation (3) over a wider range of overdensities than those originally investigated in Corasaniti et al. (2018), which is a necessary step to infer cosmological parameter constraints from multiple average sparsity determinations.

2.2 *N*-body simulations

We use halo catalogues from two distinct sets of *N*-body simulations, characterized by different cosmological model parameters, but approximately similar mass resolution and generated with the same simulation code. This enables us to extend our investigation of the average sparsity correlations to the dependence upon the underlying cosmological model (around the ΛCDM model best fitting to the CMB data).

2.2.1 *RayGalGroupSims*

The RayGalGroupSims ΛCDM simulation, or simply Raygal, consists of a (2.6 Gpc h^{-1})^3 volume and sampled with 4096^3 particles (corresponding to a particle mass resolution $m_p = 1.88 \times 10^{10}\,\mathrm{M}_\odot\,h^{-1}$) realized with the adaptive mesh refinement (AMR) *N*-body code RAMSES (Teyssier 2002). The cosmological model parameters have been set consistently to the WMAP-7 year data analysis of a flat ΛCDM model (Komatsu et al. 2011): $\Omega_m = 0.2573$, $\Omega_b = 0.04356$, $h = 0.72$, $n_s = 0.963$, and $\sigma_8 = 0.801$. We refer interested readers to Breton et al. (2019) and Rasera et al. (2021) for a detailed description of the RayGalGroupSims suite. Full redshift snapshots have been stored at $z = 0.00$, 0.50, 0.66, 1.00, 1.14, 1.50, and 2.00.

2.2.2 *M2Csims ΛCDM simulation*

The M2Csims ΛCDM simulation suite consists of three (1 Gpc h^{-1})^3 volume boxes with 2048^3 particles (corresponding to a particle mass resolution $m_p = 1.02 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$) run with the AMR *N*-body code RAMSES (Teyssier 2002). The cosmological model parameters are set to the *Planck*-2015 ΛCDM cosmology (Planck Collaboration XIII 2016a) with $\Omega_m = 0.3156$, $\Omega_b = 0.0492$, $h = 0.6727$, $n_s = 0.9645$, and $\sigma_8 = 0.831$. We refer interested readers to Le Brun et al. (2018) for a more detailed description of the suite (a complete description will appear in Le Brun et al., in preparation). Note that the suite also contains high-resolution zooms for more than 450 massive galaxy clusters, which will not be used here as the most important requirement is the number of galaxy clusters over the resolution of their profiles. Full snapshots have been stored at $z = 0.00$, 0.125, 0.25, 0.30, 0.50, 0.60, 0.75, 0.80, 1.00, 1.25, and 1.50. In the analyses presented here, we use catalogues from only two of the M2Csims simulation suite corresponding to a total comoving volume of 2 (Gpc h^{-1})^3, slightly smaller than that of the Raygal simulation.

2.3 *N*-body halo catalogues

Halo catalogues for both simulations have been generated with the spherical overdensity (SO) algorithm (Lacey & Cole 1994) implemented in the parallel code `pSOD`.[1] The algorithm first evaluates the particle density in each cell, then starts from the cell with maximum density. In each candidate cell, the centre position is chosen to be that of the particle with the greatest number of neighbouring particles within a sphere of a given radius. Afterwards, the SO algorithm computes the particle density in spheres of increasing radii around that central particle until it reaches the overdensity threshold Δ. Hereafter, we will always refer to overdensities given in units of the critical density. We focus on haloes detected with an overdensity threshold $\Delta = 200$. For each halo in the catalogues, we estimate masses at overdensities $\Delta = 200$, 500, 750, 1000, 1500, 2000, and 2500, respectively. In order to be exempt of numerical resolution artefacts, we further select haloes with $M_{2500c} > 10^{13}\,\mathrm{M}_\odot\,h^{-1}$. This also guarantees us that we consider haloes with masses $M_{200c} > 10^{13}\,\mathrm{M}_\odot\,h^{-1}$, thus corresponding to haloes hosting galaxy groups and clusters.

Since we are interested in the application of multiple average sparsity measurements to galaxy cluster observations, we limit our analysis to halo catalogues in the redshift range $0 \leq z \leq 1.5$. This is because the detection of clusters at higher redshifts, as well as the estimation of the cluster masses at the level of accuracy required, seems currently unrealistic. Also, for consistency with the conventions of the galaxy cluster community, we focus on halo masses at overdensities $\Delta = 200$, 500, 1000, and 2500, respectively.

2.4 Halo mass function calibration

We compute the halo mass function for each of the mass overdensity definitions in the halo catalogues as:

$$\frac{\mathrm{d}n}{\mathrm{d}\ln M_\Delta} = \frac{N(M_\Delta)}{\Delta \ln M_\Delta}\frac{1}{L^3}, \tag{4}$$

where $N(M_\Delta)$ is the number of haloes in a logarithmic mass bin of size $\Delta \ln M_\Delta = 0.3$ centred at M_Δ and L is the size of the simulation box. We use the numerical estimates of the halo mass functions to calibrate, at each redshift snapshot, the coefficients of the Sheth–Tormen (ST; Sheth & Tormen 1999) formula $f_{\rm ST}$ as given by:

$$\frac{\mathrm{d}n}{\mathrm{d}M_\Delta} = \frac{\rho_m}{M_\Delta}\left(-\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}M_\Delta}\right) f_{\rm ST}(\sigma), \tag{5}$$

where ρ_m is the cosmic matter density, $\sigma(M_\Delta)$ is the root-mean-square fluctuation of the linear density field smoothed on a scale enclosing the mass M_Δ, and

$$f_{\rm ST}(\sigma) = A_\Delta \frac{\delta_c}{\sigma}\sqrt{\frac{2a_\Delta}{\pi}}\left[1 + \left(\frac{a_\Delta \delta_c^2}{\sigma^2}\right)^{-p_\Delta}\right] e^{-\frac{a_\Delta \delta_c^2}{2\sigma^2}}, \tag{6}$$

where A_Δ, a_Δ, and p_Δ are calibration parameters and δ_c is the linearly extrapolated spherical collapse threshold, which we compute using the formula by Kitayama & Suto (1996).

It is worth noticing that the functional form of the ST parametrization is the base of all the numerically calibrated formula that aim to predict the halo mass function for any given set of cosmological parameters (see e.g. Tinker et al. 2008; Bocquet et al. 2016; Despali et al. 2016; Castro et al. 2021). This is because such a form of multiplicity function manifests a high level of universality. Here, we have explicitly kept the dependence on linear spherical collapse threshold δ_c, which as shown in Courtin et al. (2011), it allows to better account for the cosmology dependence of the multiplicity

[1]The parallelization scheme has been adopted from the code `pFOF` (Roy, Bouillot & Rasera 2014).

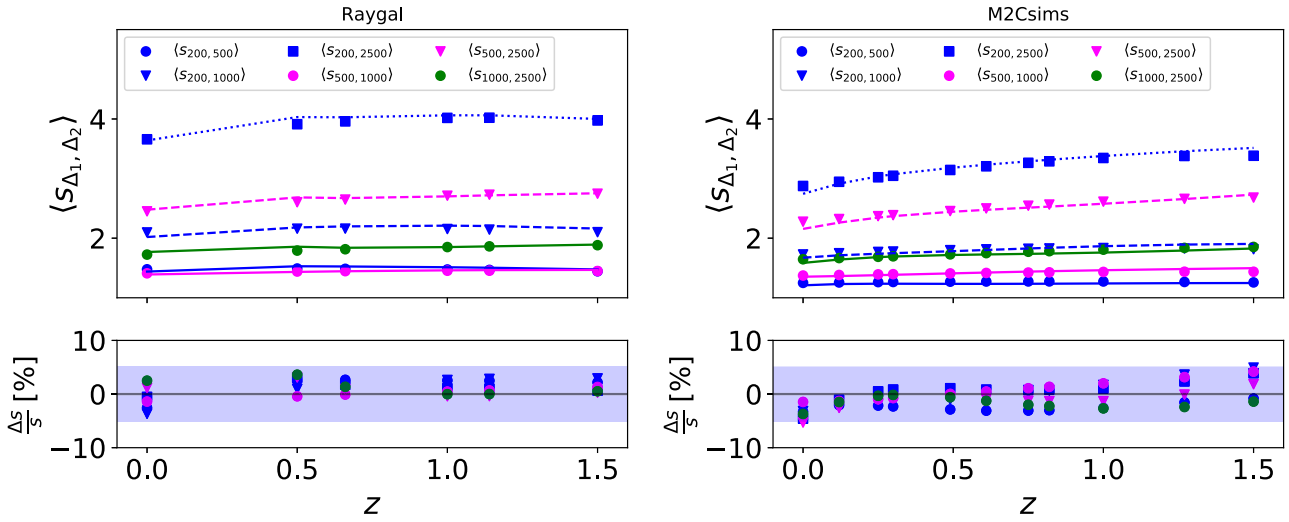

Figure 2. Average halo sparsity estimates for different overdensity configurations as a function of redshift for the Raygal (left-hand panel) and M2Csims (right-hand panel) halo catalogues. The data points in the plots correspond to the *N*-body estimates, while the various lines show the analytical predictions from the average sparsity relation of equation (3) that has been solved using the ST-parametrized mass functions calibrated on the simulations. The relative differences between the *N*-body and analytical results are shown in the bottom panels, where the shaded areas delimit the regions with less than 5 per cent relative differences. Differences between the average sparsity values from the Raygal and M2Csims halo catalogues for the same overdensity pairs pertain to the differences of the simulated cosmological models.

function. On the other hand, a key difference among the calibrated multiplicity functions discussed in the literature concerns the redshift parametrization of the ST parameters. Here, we follow the approach of Despali et al. (2016) and parametrize the redshift dependence of the ST coefficients in terms of an expansion in logarithmic powers of the overdensity Δ relative to the virial overdensity at the redshift of interest. The intent is to capture the redshift evolution of the halo mass function at different overdensities Δ for the sample of haloes detected at $\Delta = 200$. In particular, we assume a quadratic expansion, and we refer the reader to Appendix A for a detailed description of the fitting procedure of the halo mass functions. None the less, we would like to stress that our calibration substantially differs from that of Despali et al. (2016), who have provided fitting formula calibrated on halo samples detected with different overdensities. As such, their mass function cannot be used to predict the halo sparsity unless corrections are taken into account for the systematic due to effect of unmatched haloes as described in (Corasaniti et al. 2021).

Alternatively, the halo mass function for a given cosmological setup can be predicted from emulators. These are built using halo catalogues from suites of *N*-body simulations with different cosmological parameters (see e.g. McClintock et al. 2019; Nishimichi et al. 2019; Bocquet et al. 2020). In a similar manner, it should be possible to build emulators of the average halo sparsity, a possibility which we will investigate in a future study.

2.5 Average sparsities

Given a set of mass estimates M_{Δ_i} measured at n overdensities Δ_i (in units of the critical density), we can compute up to $N_s = \binom{n}{2} = \frac{n!}{2(n-2)!}$ distinct sparsities. As such, if we consider a number m of them with $m < N_s$, then the number of possible m sparsity combinations is given by $N_m = \binom{N_s}{m} = \frac{N_s!}{m!(N_s-m)!}$.

As already mentioned, here we restrict ourselves to $n = 4$ mass measurements at $\Delta = 200$, 500, 1000, and 2500 and for each halo in the catalogues, we focus on the following set of $N_s = 6$ halo sparsities: $s_{200,500}$, $s_{200,1000}$, $s_{200,2500}$, $s_{500,1000}$, $s_{500,2500}$, and $s_{1000,2500}$. Notice that in such a case, there is a total of $N_{\rm tot} = \sum_{m=1}^{N_s} N_m = 63$ possible permutations for any number m of sparsities used in the analysis. The factorial dependence of the number of combinations prohibits the full exploration of this parameter space; as such, in later sections, we will clearly quote which combinations are used. From this set, at each redshift snapshot, we evaluate the halo ensemble average sparsities by computing the arithmetic mean of the individual halo sparsities:

$$\langle s_{\Delta_1,\Delta_2}\rangle \equiv \left\langle \frac{M_{\Delta_1}}{M_{\Delta_2}} \right\rangle = \frac{1}{N_h}\sum_i^{N_h} s^i_{\Delta_1,\Delta_2}, \tag{7}$$

where $N_{\rm h}$ is the total number of haloes in a catalogs at a given redshift.

In Fig. 2, we plot the average halo sparsities for the Raygal (left-hand panel) and M2Csims (right-hand panel) halo catalogues respectively. In the same plots, we also show the values predicted by the solutions of equation (3) and the relative differences with respect to the *N*-body estimates (bottom panels). We can see that differences are well within 5 per cent level and in some cases even at sub-per cent level. In Appendix A, we also show the relative differences between the prediction from the Raygal calibrated mass functions for the M2Csims cosmology and the M2Csims average halo sparsity and vice versa, which we find to be $\lesssim 5$ per cent, consistent with those shown in Fig. 2.

Notice that there is a systematic difference between the values of the average sparsities obtained from the Raygal simulation and those from the M2Csims case. This pertains to the cosmological dependence of the sparsity originally pointed out in Balmès et al. (2014).

The average halo sparsity is mainly sensitive to a degenerate combination of Ω_m and σ_8 as given by $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$ (Corasaniti et al. 2018, 2021). In particular, the lower the level of clustering of a given cosmological model, i.e. the smaller the value of S_8, the higher the value of the average halo sparsity. This is because cosmic structures will form later in a cosmological model with lower S_8 than in a model with a larger value. Consequently, such structures will be

less concentrated or equivalently more sparse than in a model with a larger value of S_8.

Comparing the trends in Fig. 2, we can see that the Raygal average sparsities at any given redshift are systematically larger than the M2Csims values, which is consistent with the fact that the Raygal ΛCDM model has $S_8 = 0.742$, while in the case of the M2Csims we have $S_8 = 0.852$. It is also worth noticing that, among the different sparsity estimates, the one with the largest value and the largest variation with the underlying cosmology is associated to $\langle s_{200,\,2500}\rangle$. In particular, the maximum relative variation of $\langle s_{200,\,2500}\rangle$ with respect to the M2Csims case amounts to ~ 20 per cent. This is consistent with the results of Balmès et al. (2014), who have found that the cosmological dependence of the sparsity increases as the difference between Δ_1 and Δ_2 increases. On the other hand, we can also notice that sparsities for different overdensity pairs do not have the same sensitivity to the underlying cosmology. As an example, we find that $\langle s_{500,\,1000}\rangle$ and $\langle s_{1000,\,2500}\rangle$ vary by only a few per cent.

2.6 Average sparsities correlations

The cosmological information encoded in the estimated average sparsities is not independent. In fact, the gravitational processes that shape the mass assembly of the haloes correlate the properties of the mass distribution within different radial shells. For this reason, we use the data from the *N*-body halo catalogues to compute the correlation coefficients of the different sparsity estimates, which is given by:

$$r_{s_i,s_j} = \frac{\sum_{k=1}^{N_\mathrm{h}} \left(s_i^k - \langle s_i\rangle\right)\left(s_j^k - \langle s_j\rangle\right)}{\sqrt{\sum_{k=1}^{N_\mathrm{h}} \left(s_i^k - \langle s_i\rangle\right)^2 \sum_{k=1}^{N_\mathrm{h}} \left(s_j^k - \langle s_j\rangle\right)^2}}, \tag{8}$$

where the index $i, j = \{(200, 500), (200, 1000),..., (1000, 2500)\}$ with $i \neq j$. They are shown in Fig. 3 as a function of redshift for the Raygal and M2Csims halo catalogues, respectively. In order to facilitate the visualization of the low correlated pairs of sparsity configurations against the highly correlated one, we have adopted the *magma* colourmap for the colours of the various lines.

First of all, from Fig. 3, we may notice that all correlations increase from high-to-low redshifts both for the Raygal haloes and M2Csims ones. This is a direct consequence of the mass assembly process of haloes, which grow from inside out (Taylor 2011; Wang et al. 2011). As the haloes assemble their mass over cosmic time, the mass distributions within different mass shells become increasingly correlated. Secondly, we can see that the correlations are smaller for sparsities that sample the mass profile within mass shells that are at larger separations. As an example, $s_{200,\,500}$ and $s_{1000,\,2500}$ have a maximal ∼25 per cent correlation at $z = 0$, which is not the case for sparsities probing the mass distribution in close mass shells (or even overlapping ones) with correlation greater than 50 per cent. We find the redshift evolution of the correlation coefficients is to be well-approximated by a linear regression, which we provide in Appendix B for practical applications.

Notice that the correlation coefficients from the Raygal catalogues slightly differ from those of the M2Csims ones. Again, this is a direct consequence of the differences between the simulated cosmological models. In particular, at any given redshift, the correlations from M2Csims are slightly larger than that from Raygal, which is consistent with the fact that the former has a larger S_8 value than the latter. Nevertheless, these differences are too small to have an impact on the cosmological parameter inference, as we will discuss in Section 3.

In contrast, we would like to highlight the fact that if the density profile of haloes was exactly described by the NFW formula (Navarro et al. 1997), then all the information on the halo mass profile would be fully encoded in the values of the concentration parameter and the overall halo mass, given that the halo mass at any other overdensities can be derived from these two quantities. Furthermore, because of the one-to-one relation between halo sparsity and concentration for NFW haloes (see Balmès et al. 2014), this would imply that a single sparsity estimate would carry all the information on the mass profile. Hence, if the density profile of haloes exactly were to follow the NFW model, one should find that different sparsities, even those probing distant mass shells, are highly correlated. One may argue that, given the fact that for each halo the best-fitting value of the concentration is a stochastic variable characterized by a scatter and a mean that varies with M_{200c}, the correlation among the NFW inferred sparsities, e.g. s_{200,Δ_2} and s_{Δ_3,Δ_4} (with $\Delta_2 \neq \Delta_3 \neq \Delta_4 > 200$), may not be exactly one. However, because the functional form of the mass profile has to follow NFW, these should still be close to unity. This is indeed what we find when we compute the correlations among sparsities that have been computed using NFW inferred masses for each halo in the catalogues, as shown in Fig. 4. More specifically, for each halo with a given mass M_{200c} in the M2Csims catalogues, we have fit its density profile with the NFW function and deduced the corresponding best-fitting NFW concentration parameter c_{200c}. Then, given the values of M_{200c} and c_{200c}, we have calculated the NFW halo mass at $\Delta =$ 500, 1000, and 2500 (in units of the critical density) and computed the associated sparsities. Finally, we have estimated the correlation coefficients among the various sparsities using equation (8). As we can see in Fig. 4, the correlation coefficients among these NFW estimated sparsities are all close to unity. This is in sharp contrast with what we found from the analysis of the *N*-body halo masses shown in Fig. 3.

Indeed, the fact that differently from the NFW case, the correlations among different sparsities are not all clustered around $r = 1$, but spread over a larger interval of values, as shown in Fig. 3, is indicative of the fact that on average *N*-body haloes are not exactly described by the NFW formula. Moreover, it clearly shows that there is additional information about the halo mass profile, which is not captured by the NFW profile but can be extracted using multiple sparsity measurements, thus potentially providing additional constraints on the cosmological parameters.

3 SYNTHETIC DATA ANALYSIS

We seek to investigate the constraints on cosmological parameters that can be inferred from multiple sparsity measurements. To this end, we use the average sparsity estimates from the *N*-body halo catalogues as a synthetic data set and perform a Markov Chain Monte Carlo likelihood analysis under different average sparsity error model assumptions. Our goal is twofold. On one hand, we want to test to which extent the analysis recovers the fiducial cosmological parameters of the simulated cosmologies. On the other hand, we aim to study how the inferred parameter uncertainties vary for different sparsity configurations, uncertainties, and fiducial cosmologies.

3.1 Sparsity configurations and uncertainties

We consider the following set of average sparsity combinations:

S1) $\langle s_{200,\,2500}\rangle$;
S2) $\langle s_{200,\,500}\rangle$, $\langle s_{200,\,2500}\rangle$;
S3) $\langle s_{200,\,500}\rangle$, $\langle s_{200,\,2500}\rangle$, $\langle s_{500,\,2500}\rangle$;
S4) $\langle s_{200,\,500}\rangle$, $\langle s_{200,\,1000}\rangle$, $\langle s_{200,\,2500}\rangle$, $\langle s_{500,\,2500}\rangle$;
S5) $\langle s_{200,\,500}\rangle$, $\langle s_{200,\,1000}\rangle$, $\langle s_{200,\,2500}\rangle$, $\langle s_{500,\,1000}\rangle$, $\langle s_{500,\,2500}\rangle$;

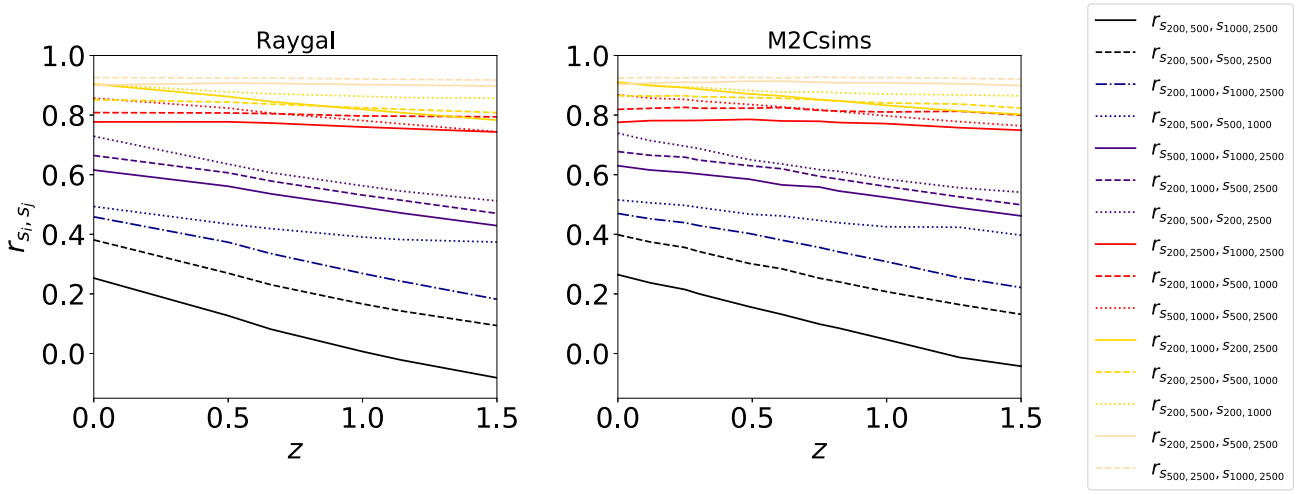

Figure 3. Sparsity correlation coefficients from the Raygal (left-hand panel) and M2Csims (right-hand panel) halo catalogues, respectively. As we can see over the redshift range $0 < z < 1.5$, sparsity combinations probing the halo mass distribution in close (overlapping) mass shells are highly correlated ($r \gtrsim 0.5$). This is not the case of sparsities associated to mass shells that are at larger separations ($r \lesssim 0.5$).

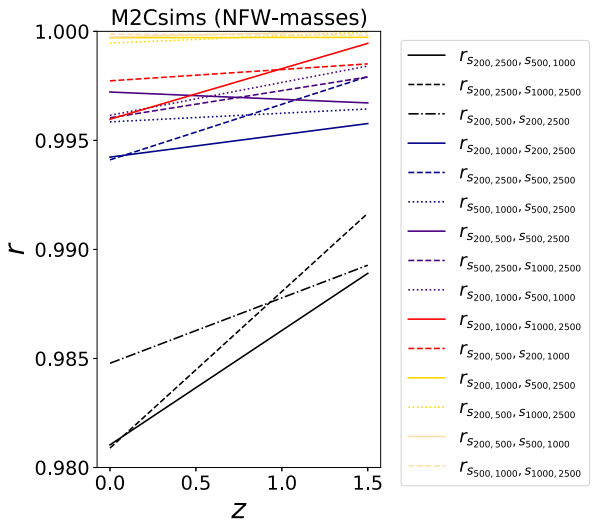

Figure 4. Sparsity correlation coefficients obtained using the mass estimates from the best-fitting NFW profiles of the M2Csims halo catalogues. As we can see, contrary to the correlations among sparsities estimated from the SO halo masses, assuming that the density profile of haloes is described by NFW profile artificially correlates sparsities to order $r \approx 1$, even those probing mass distribution in mass shells at large separations.

S6) $\langle s_{200,500}\rangle$, $\langle s_{200,1000}\rangle$, $\langle s_{200,2500}\rangle$, $\langle s_{500,1000}\rangle$, $\langle s_{500,2500}\rangle$, $\langle s_{1000,2500}\rangle$;

where starting from the single sparsity $\langle s_{200,2500}\rangle$, we explore multiple sparsity configurations up to S6, which corresponds to the maximal number of sparsities N_s that can be obtained from the estimation of halo masses at four different overdensities.

In principle, for the configurations S1–S5, we have a total of 62 possible sparsity configurations to study. Rather than a brute force investigation, for any $m < 6$ number of sparsities, we have adopted a physically motivated strategy to explore among the various possibilities. This relies upon the observation that the cosmological differences among single average sparsity measurements $\langle s_{\Delta_1,\Delta_2}\rangle$ at a given redshift are maximized when the differences between Δ_1 and Δ_2 are the largest (Balmès et al. 2014). Hence, given the range of overdensities Δ we have considered, we set S1 to be the sparsity associated to the largest overdensity separation $\Delta_1^{\rm min} = 200$ and $\Delta_2^{\rm max} = 2500$. For the S2 configuration, we proceed by adding the average sparsity that probes the average mass distribution in a mass shell at an intermediate overdensity between $\Delta_1^{\rm min}$ and $\Delta_2^{\rm max}$. In our case, we have chosen $\Delta = 500$ and considered $\langle s_{200,500}\rangle$. This is done with the intent of investigating how the cosmological parameter constraints vary with the addition of information encoded within an intermediate mass shell with respect to the one already accounted by the previous sparsity configuration. For the configuration S3, we consider the average sparsity associated to the overdensities with the second largest separations among those considered at S2, which is $\langle s_{500,2500}\rangle$. Then, we proceed in a similar manner for S4 and S5.

For each of the sparsity combinations in the list, we consider two distinct synthetic data sets that consist of the average sparsity estimates at different redshifts from the Raygal and M2Csims catalogues, respectively. Given the larger number of redshift snapshots of the M2Csims simulations, this enable us to asses the impact of additional average sparsity estimates for a larger number of redshifts in the same redshift interval $0 \leq z \leq 1.5$.

Here, we account for statistical uncertainties on average sparsity measurements and propagate the effect of systematic errors due to the mass function model uncertainties in predicting the redshift and cosmological model dependence of the average sparsity. In Section 4, we extend the analysis of systematics and present the result of a forecast parameter inference analysis for realistic cluster survey configurations.

Statistical errors on average sparsities are the consequence of the propagation of the uncertainties of cluster mass measurements.[2] Following Corasaniti et al. (2018), we model the error on the average sparsity $\langle s_{\Delta_1,\Delta_2}\rangle$ at redshift z as:

$$\sigma_{\langle s(z)\rangle} = \frac{\langle s_{\Delta_1,\Delta_2}(z)\rangle}{\sqrt{N_{\rm cl}(z)}}\sqrt{e^2_{M_{\Delta_1}} + e^2_{M_{\Delta_2}}}, \tag{9}$$

[2]We neglect possible correlations among cluster mass determination, a choice which makes the assumed errors on the average sparsity only more conservative. In fact, given that the sparsity is a mass ratio, neglecting the correlations r_{Δ_1,Δ_2} between the determination of the masses M_{Δ_1} and M_{Δ_2} is equivalent to overestimating the errors on the sparsity by a factor $\sim 1/\sqrt{1-r}$.

where $e_{M_{\Delta_1}}$ and $e_{M_{\Delta_2}}$ are the fractional error on the mass measurements at overdensities Δ_1 and Δ_2, respectively, and $N_{\rm cl}(z)$ is the number of clusters in the bin centred at redshift z.

We focus on a simplified configuration and consider two distinct cases for the statistical errors: $\sigma_{\langle s(z)\rangle} = 0.3$ and 0.1. The former is a rather conservative choice, corresponding to having ∼50 clusters per redshift bin with 30 per cent fractional mass measurement errors, while the latter is a more optimistic assumption corresponding to having ∼100 clusters per redshift bin with 10 per cent precision on the estimated masses.

We treat the discrepancies between the average sparsity predictions and the *N*-body found in Section 2.5 and Appendix A as an intrinsic systematic error $\sigma^{\rm sys}_{s_{\Delta_1,\Delta_2}}(z)$ on the average sparsity $\langle s_{\Delta_1,\Delta_2}\rangle$ obtained by solving equation (3).

3.2 Priors and likelihood

We specifically focus our analysis on Ω_m and σ_8, the cosmological parameters to which the sparsity is most sensitive, while setting h, Ω_b, and $n_{\rm s}$ to their fiducial values. We assume uniform priors on $\Omega_m \sim U(0.1, 0.5)$ and $\sigma_8 \sim U(0.2, 1.2)$.

We perform a Markov Chain Monte Carlo (MCMC) sampling of the log-likelihood function:

$$-2\ln\mathcal{L} = \sum_{i,j=1}^{m}\sum_{k=1}^{N_z}\Delta s_i(z_k)\cdot C^{-1}_{s_i,s_j}(z_k)\cdot\Delta s_j(z_k), \tag{10}$$

with $m \leq N_{\rm s}$ the number of sparsity configurations considered, N_z the number of redshift bins and

$$\Delta s_i(z_k) = \langle s_i^{\rm mf}(z_k)\rangle(1+\tilde{Y}^{s_i}_{z_k}) - \langle s_i(z_k)\rangle,$$

where $\langle s_i(z_k)\rangle$ is the synthetic data point at the k-th redshift bin z_k for the i-th configuration of overdensities, while $\langle s_i^{\rm mf}(z_k)\rangle$ is the average sparsity predicted by the mass function model equation (3) with $\tilde{Y}^{s_i}_{z_k} \sim N(0, \sigma^{\rm sys}_{s_i}(z_k))$ being a Gaussian random variable which we marginalize over, characterized by zero mean and standard deviation $\sigma^{\rm sys}_{s_i}(z)$. The latter being the sum of the intrinsic scatter with respect to the *N*-body average sparsities discussed in Section 2.5 and Appendix A. The covariance matrix reads as:

$$C_{s_i,s_j}(z_k) = \sigma^2_{\langle s(z_k)\rangle} r_{s_i,s_j}(z_k), \tag{11}$$

where $r_{s_i,s_j}(z_k)$ is the correlation matrix at redshift z_k, which we have previously computed using the Raygal and M2Csims catalogues in Section 2.6 and $\sigma_{\langle s(z_k)\rangle}$ is the statistical uncertainty on the average sparsity estimates.

3.3 Results

We use the MCMC chains to infer marginal constraints on Ω_m, σ_8, and S_8. For this purpose, we have implemented a Metropolis–Hastings algorithm and tested the convergence of the chains with the Gelman–Rubin diagnostics (see Roy 2020, for a review). We have analysed the chains using the publicly available package GETDIST[3] (Lewis 2019). The results are summarized in Table 1 where we quote the marginalized 1σ errors on the parameters as obtained from the various cases we have considered. We plot the corresponding 1 and 2σ credibility contours in the $\Omega_m - \sigma_8$ from the Raygal synthetic data analysis in Fig. 5 and for the M2Csims data set in Fig. 6.

[3]https://getdist.readthedocs.io/

3.3.1 *Raygal synthetic data set*

First, we find that for all sparsity configurations S1–S6, the best-fitting parameters of the MCMC likelihood analysis coincide with the values of the Raygal fiducial cosmology. Unsurprisingly, the constraints on Ω_m and σ_8 (and consequently S_8) quoted in Table 1 show that, for a given sparsity configuration, the inferred parameter errors obtained assuming statistical errors of $\sigma_{\langle s(z)\rangle} = 0.1$ are systematically smaller than those obtained for $\sigma_{\langle s(z)\rangle} = 0.3$. This can also be seen by the different size of the 1 and 2σ credibility contours shown in Fig. 5. Quite interestingly, we notice that, in both cases, the contours shrink from S1 to S4, thus indicating that using additional sparsity measurements does improve the cosmological constraints. On the other hand, we can see that, in the case of a greater number of sparsity configurations S5 and S6, the contours do not shrink further. Quite the opposite, from the marginalized error values quoted in Table 1, we find that the constraints on the cosmological parameters slightly degrade. This is most likely due to the fact that the additional sparsities considered in S5 and S6, namely $\langle s_{500,\,1000}\rangle$ and $\langle s_{1000,\,2500}\rangle$, do not vary significantly with the cosmological parameters. As we have seen in Section 2.5, these have variation of the order of per cent level around the fiducial cosmology, which is of the same order of the accuracy of the cosmological model predictions given by equation (3), using the numerical ST calibrated mass functions. This trend can be better seen in the inset plots, where we show the marginalized 1σ error on S_8 as a function of the sparsity configurations considered. As we can see, σ_{S_8} diminishes as a function of the number of sparsity configuration considered, reaching a minimum value for S4, while increasing for S4 and S5 a possible consequence of the $\Omega_m - \sigma_8$ degeneracy. Indeed, taking as figure of merit, the values of the area within the 1σ credibility contours in the $\Omega_m - \sigma_8$ plane highlights better the saturation of the constraints on the cosmological parameters beyond S4. As we can infer from the values quoted in Table 2, the area diminishes from S1 to S4 and then remains constant.

Quantitatively, we find that in the case of $\sigma_{\langle s(z)\rangle} = 0.3$, the marginalized 1σ error on Ω_m improves by approximately a factor of 20 from S1 to S4; similarly, the uncertainties on σ_8 improves by a factor ∼15. Even, considering three sparsity measurements, such as in the S3 configuration, leads to an improvement of a factor ∼5 on σ_{Ω_m} and a factor ∼4 on σ_{σ_8}. In the case of $\sigma_{\langle s(z)\rangle} = 0.1$, we find that the σ_{Ω_m} reduces by a factor of ∼9 from S1 to S4 and σ_{σ_8} by a factor of ∼8. On the other hand, it is worth noticing that, for S4, the constraints do not significantly improve when reducing the statistical errors on the average sparsities by a factor of 3 (i.e. from $\sigma_{\langle s(z)\rangle} =$ 0.3 to 0.1). This suggests that the use of fours sparsity measurements can mitigate the need for improved mass measurements.

3.3.2 *M2Csims synthetic data set*

The likelihood analysis of the MC2sims data set shows trends that are similar to those we have found using the Raygal data set. Again, we have that, in all the cases we considered, the best-fitting parameters coincide with the values of the MC2sims fiducial cosmology. Moreover, as it can be noticed from the values quoted in Table 1 for $\sigma_{\langle s(z)\rangle} = 0.3$ and 0.1, also in these cases we find that the uncertainties on Ω_m, σ_8 (and S_8) decrease to a minimum value as the number of sparsities increases from S1 to S4, while they slightly increase for S5 and S6 configurations. This can also be seen in Fig. 6, where the credibility contours shrink from S1 to S4 as we have already found in the Raygal case. The inset plot shows σ_{S_8} as a function of the number of sparsity configurations, which reaches a

Table 1. Marginalized 1σ errors on Ω_m, σ_8, and S_8 from the MCMC likelihood analysis of the Raygal and M2Csims synthetic data inferred assuming average sparsity errors of $\sigma_{\langle s(z)\rangle} = 0.3$ and 0.1, respectively, for the various sparsity configurations S1–S6. As we can see, the constraints on the cosmological parameters improve for increasing number of sparsity configurations, reaching a minimum for S4.

Configuration	Raygal ($\sigma_{\langle s(z)\rangle} = 0.3$)			Raygal ($\sigma_{\langle s(z)\rangle} = 0.1$)			M2Csims ($\sigma_{\langle s(z)\rangle} = 0.3$)			M2Csims ($\sigma_{\langle s(z)\rangle} = 0.1$)		
	σ_{Ω_m}	σ_{σ_8}	σ_{S_8}	σ_{Ω_m}	σ_{σ_8}	σ_{S_8}	σ_{Ω_m}	σ_{σ_8}	σ_{S_8}	σ_{Ω_m}	σ_{σ_8}	σ_{S_8}
S1	0.039	0.124	0.086	0.019	0.061	0.038	0.068	0.100	0.042	0.043	0.060	0.018
S2	0.037	0.109	0.072	0.016	0.049	0.030	0.061	0.086	0.037	0.034	0.047	0.015
S3	0.007	0.027	0.019	0.003	0.012	0.008	0.016	0.025	0.008	0.007	0.010	0.004
S4	0.002	0.008	0.005	0.002	0.010	0.006	0.007	0.009	0.002	0.006	0.008	0.001
S5	0.005	0.020	0.015	0.002	0.011	0.007	0.011	0.017	0.007	0.006	0.007	0.003
S6	0.005	0.020	0.015	0.002	0.011	0.007	0.011	0.017	0.007	0.006	0.007	0.003

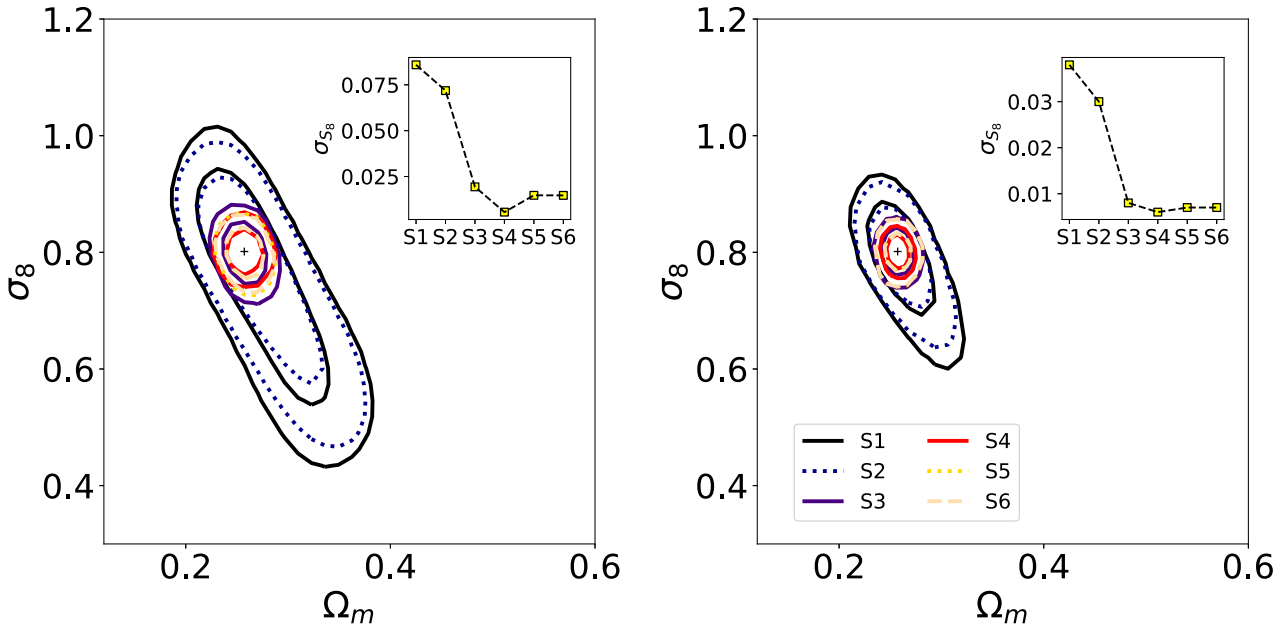

Figure 5. 1 and 2σ credibility contours in the $\Omega_m - \sigma_8$ plane for the Raygal synthetic data sets. The different lines correspond to the sparsity configurations S1–S6 described in Section 3.1, assuming statistical errors on the average sparsity estimates of $\sigma_{\langle s(z)\rangle} = 0.3$ (left-hand panel) and $\sigma_{\langle s(z)\rangle} = 0.1$ (right-hand panel). The cross corresponds to the cosmological parameter values of the fiducial Raygal cosmology. As the best-fitting parameters for the different sparsity configurations coincide with those of the fiducial model, we do not show them in the plot to avoid cluttering. The inset plot shows the 1σ error on S_8 as a function of the number of average sparsity configurations S1–S6. As we can see the uncertainties saturate beyond S4.

minimum value for S4 and slightly increase for S5 and S6. Again, we can better appreciate the saturation of the cosmological parameter constraints at S4 from the values of the area enclosed within the 1σ credibility contours quoted in Table 3.

Quantitatively, from the values quoted in Table 1, we find an improvement of a factor of $\sim$10 on σ_{Ω_m} and σ_{σ_8} for $\sigma_{\langle s(z)\rangle} = 0.3$, and a factor of $\sim$7 for $\sigma_{\langle s(z)\rangle} = 0.1$.

Notice that in addition to the contours from the analysis of the configurations S1–S6, in Fig. 6, we also plot the results of two additional cases we have investigated for the S1 configuration. In particular, we have performed an analysis of the M2Csims synthetic average sparsity data using the covariance matrix from the Raygal simulation such as to evaluate the impact of the cosmological dependence of the covariance on the cosmological parameter constraints. For this purpose, we have evaluated the covariance at the M2Csims redshift bins using the parametrization of the Raygal sparsity correlation coefficients given in Appendix B. We have also performed an analysis of the M2Csims synthetics data set limited to $N_z = 6$ redshift bins at $z = 0.00$, 0.49, 0.61, 1.00, 1.27, and 1.50 (as in the Raygal case) to evaluate the impact of additional redshift bins on the cosmological parameter inference. In the former case, we find that there is no effect of using the Raygal covariance for the analysis of the M2Csims data, which suggests that the cosmological dependence of the covariance discussed in Section 2.6 is too small to have an impact on the cosmological parameter inference for the level of average sparsity uncertainty we have assumed. In the latter case, we do find that increasing the number of redshift bins improves the constraints on the parameters, though not significantly when compared to the effect of using multiple sparsity measurements. As an example, for the case $\sigma_{\langle s(z)\rangle} = 0.3$ with $N_z = 6$, we find $\sigma_{\Omega_m} = 0.074$ and $\sigma_{\sigma_8} = 0.118$, while in the case with $N_z = 11$, we have $\sigma_{\Omega_m} = 0.068$ and $\sigma_{\sigma_8} = 0.100$. Similarly, for the case with $\sigma_{\langle s(z)\rangle} = 0.1$ and $N_z = 6$, we have $\sigma_{\Omega_m} = 0.047$ and $\sigma_{\sigma_8} = 0.072$, while in the case with $N_z = 11$, we have $\sigma_{\Omega_m} = 0.043$ and $\sigma_{\sigma_8} = 0.060$.

Overall, comparing the results from the Raygal analysis and those obtained from the M2Csims, we find that the inferred parameter constraints do depend on the underlying fiducial cosmology. As summarized by the 1σ errors on S_8, we have that for a given sparsity

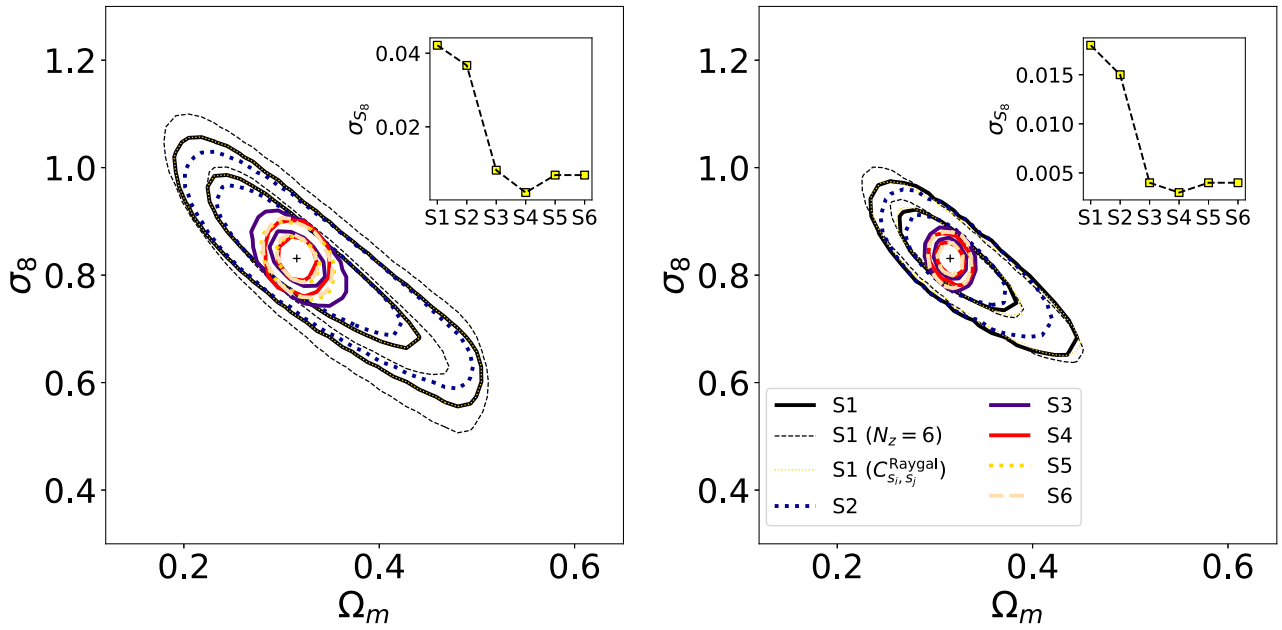

Figure 6. As in Fig. 5 but for the M2Csims case. In addition to the sparsity combinations S1–S6, we also show the 1 and 2σ contours for the S1 case with $N_z = 6$ redshift bins in the redshift interval $0 < z < 1.5$ (thin black-dashed lines) rather than the nominal $N_z = 11$ (black solid line), and with covariance from the Raygal simulation (yellow-dotted line).

Table 2. Area within the 1σ credibility contours shown in Fig. 5 for the various sparsity configuration in the case of the Raygal data analysis with ($\sigma_{\langle s(z)\rangle} = 0.3$) and 0.1 statistical uncertainties. We may notice that the area diminishes for increasing number of sparsity configurations and saturates at S4.

Configuration	Raygal analysis $A_{1\sigma}(\sigma_{\langle s(z)\rangle} = 0.3)$	$A_{1\sigma}(\sigma_{\langle s(z)\rangle} = 0.1)$
S1	0.059	0.021
S2	0.048	0.017
S3	0.009	0.005
S4	0.006	0.004
S5	0.006	0.004
S6	0.006	0.004

Table 3. As in Table 2 for the M2Csims data analysis.

Configuration	M2Csims analysis $A_{1\sigma}(\sigma_{\langle s(z)\rangle} = 0.3)$	$A_{1\sigma}(\sigma_{\langle s(z)\rangle} = 0.1)$
S1	0.058	0.030
S2	0.050	0.024
S3	0.011	0.004
S4	0.007	0.003
S5	0.007	0.003
S6	0.007	0.003

configuration and given level of statistical uncertainty on the average sparsity measurements, the value of σ_{S_8} is systematically smaller in the M2Csims case than in the Raygal case by approximately a factor of 2. Such dependence on the fiducial cosmology of forecast parameter error analysis is not new (e.g. we refer the readers to the appendix B of Mukherjee et al. 2006, for a detailed discussion). It simply reflects the amplitude of the variation of the observable (the average sparsity in our case) across the cosmological parameter space relative to the amplitude of the observational uncertainties at the observed data points. This justifies the need for parameter forecast studies performed under different model assumptions.

4 CHEX-MATE CLUSTERS FORECAST ANALYSIS

We forecast cosmological parameter constraints from multiple average sparsity measurements for a realistic galaxy cluster data sample. We specifically focus on cluster mass measurements as expected from the CHEX-MATE project (CHEX-MATE Collaboration 2021), which consists of a sample of 118 clusters from the Planck-SZ catalogue in the redshift range $0 < z < 0.6$. These are the targets of a dedicated X-ray observing program on the *XMM*-Newton satellite, which is expected to provide accurate measurements of the cluster mass distributions and gas properties. For each cluster in the sample, mass estimates at different overdensities will be obtained under the hydrostatic equilibrium (HE) hypothesis. Similarly to the study presented in Section 3, we perform a likelihood MCMC analysis of a synthetic data set with characteristics and mass measurement errors expected from the CHEX-MATE sample to infer constraints on Ω_m and σ_8. In the following, we set the fiducial cosmological model to the flat ΛCDM best fitting to the Planck 2015 data (Planck Collaboration XIII 2016a).

In order to build a synthetic data set of average sparsity measurements that are consistent with the characteristics of the CHEX-MATE sample, we first bin the CHEX-MATE clusters in equally spaced redshift bins of size $\Delta z = 0.1$, the corresponding number counts $N(z)$ are shown in Fig. 7. Then, we generate a sample of synthetic average sparsity data $\langle s_{200,\,500}\rangle$, $\langle s_{200,\,1000}\rangle$, $\langle s_{200,\,2500}\rangle$, and $\langle s_{500,\,2500}\rangle$ by solving equation (3) at the central redshift of the different bins using the M2Csims mass function parametrizations discussed in Appendix A. We also estimate the average sparsity errors using equation (9), where we assume cluster mass uncertainties expected from the analysis of the CHEX-MATE observations. In particular, thanks to the observational strategy adopted in CHEX-MATE,

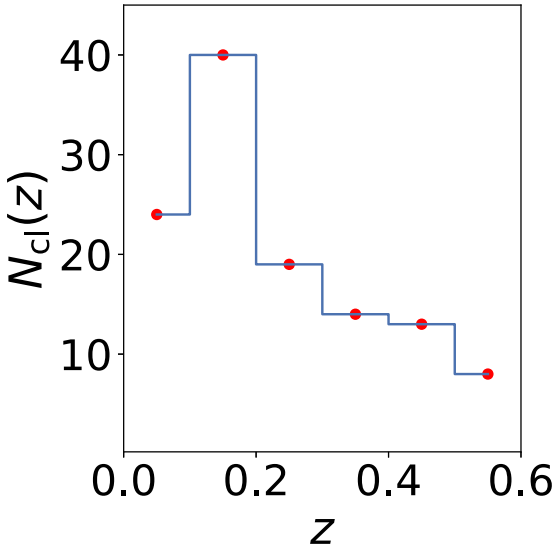

Figure 7. CHEX-MATE binned cluster counts in equally spaced redshift bins of size $\Delta z = 0.1$ in the range $0 < z < 0.6$.

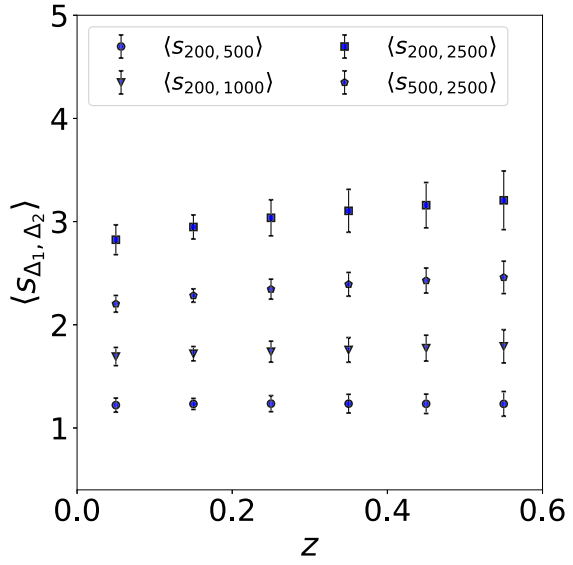

Figure 8. Synthetic average sparsity data $\langle s_{200,500}\rangle$ (circles), $\langle s_{200,1000}\rangle$ (triangles), $\langle s_{200,2500}\rangle$ (squares), and $\langle s_{500,2500}\rangle$ (pentagons). The error bars indicate the amplitude of statistical errors due to the propagation of mass measurement uncertainties.

homogenous exposures with *XMM-Newton* for the entire sample will guarantee to reach a relative error of about 15 per cent on the hydrostatic masses measured at $\Delta = 500$. Hence, by interpolating and scaling the relative errors on hydrostatic masses obtained at different overdensities in the X-COP project (Ettori et al. 2019), we can reasonably assume fractional mass errors of $e_{M_\Delta} = 0.23, 0.15, 0.11$ and 0.10 at $\Delta = 200, 500, 1000$, and 2500, respectively. The synthetic data sets are shown in Fig. 8.

We consider two distinct cases: single average sparsity measurements $\langle s_{200,2500}\rangle$ (S1); four average sparsity measurements $\langle s_{200,500}\rangle$, $\langle s_{200,1000}\rangle$, $\langle s_{200,2500}\rangle$, and $\langle s_{500,2500}\rangle$ (S4). In the latter case, we evaluate the covariance matrix using equation (11), where we estimate the correlation coefficients for the different average sparsities at the different redshifts using the linear regression obtained from the analysis of the M2Csims halo catalogues and presented in Appendix B.

Table 4. Percentage bias shift of the average sparsities due to the HE mass bias (first row) and the impact of baryons (second row).

	$\Delta b_{200,500}$	$\Delta b_{200,1000}$	$\Delta b_{200,2500}$	$\Delta b_{500,2500}$
HE mass bias	0.03	0.02	0.03	0.04
Baryon mass bias	0.04	0.10	0.15	0.10

We assume the log-likelihood function as given by equation (10). Similarly to the analysis presented in Section 3, we propagate the effect of systematic uncertainties by marginalizing over the Gaussian random variable with zero mean and standard deviation corresponding to the sum of all systematic errors we account for. Here, in addition to the intrinsic scatters due to our data model, we also propagate the impact of mass biases on the average sparsity caused by the presence of baryons.

We infer constraints for three different error configurations:

(a) statistical errors due to the propagation of mass-measurement uncertainties as estimated by equation (9) in combination with the intrinsic data model errors;

(b) statistical errors in combination with the intrinsic systematic errors of our data model and the systematic uncertainties due to the effects of hydrostatic mass bias on sparsity estimates;

(c) statistical errors in combination with intrinsic systematic data model errors and systematic uncertainties due to the effects of baryons on sparsity estimates based on dark matter only masses.

In cases (b) and (c), we estimate the impact of mass biases on the average sparsity by evaluating the percentage bias shift: $\Delta b_{\Delta_1,\Delta_2} = \Delta s_{\Delta_1,\Delta_2}/\langle s_{\Delta_1,\Delta_2}\rangle$.

We assume the percentage bias shifts due to HE mass bias estimated in Richardson & Corasaniti (2022), which have been obtained from the analysis of *N*-body/hydro simulations of galaxy clusters from Biffi et al. (2016). We quote these systematic bias shifts in Table 4. Instead, we evaluate the impact of baryons on the average sparsity estimates from dark matter-only masses using the results of the mass biases found in Velliscig et al. (2014) from the analysis of a combination of the OverWhelmingly Large Simulations (OWLS; Schaye et al. 2010) and cosmo-OWLS (Le Brun et al. 2014) for the feedback model AGN 8.0 that reproduce the observed X-ray profiles of clusters (Le Brun et al. 2014). The corresponding percentage bias shifts on different sparsity estimates have been estimated in Corasaniti et al. (2018) as a function of cluster mass $M_{200c} > 10^{13}$ $M_\odot\, h^{-1}$ (see their Fig. 7). Here, we conservatively assume the largest absolute values from Corasaniti et al. (2018), which we quote in Table 4.

We assume priors and evaluate the likelihood as specified in Section 3.2. From the MCMC chains, we derive the marginal constraints on Ω_m, σ_8, and S_8 for the different sparsity configurations and error assumptions. These are quoted in Table 5, while in Fig. 9, we plot the corresponding 1 and 2σ credibility regions in the $\Omega_m - \sigma_8$ plane.

First of all, we find that in all the cases the best-fitting model parameters coincide with those of the fiducial cosmological model, shown as a cross in Fig. 9. In the S1 case, we may notice that the inclusion of the systematic errors due to the HE or baryon biases only allow to infer an upper bound on σ_8 and a lower bound on Ω_m. This is because, for the assumed errors, a single sparsity measurements over the range of redshift considered only constrains the degenerate parameter combination given by S_8. In particular, we find $\sigma_{S_8} = 0.06$

Table 5. Marginalized 1σ errors on Ω_m, σ_8, and S_8 for the different sparsity configurations and error assumptions. In the last column, we quote the values of the area under the 1σ credibility countour. Notice that, in the S1 case, the propagation of systematic uncertainties due to the HE bias or the effect of baryons only allow to infer an upper bound on σ_8 and a lower bound on Ω_m.

	σ_{Ω_m}	σ_{σ_8}	σ_{S8}	$A_{1\sigma}$
S1 (Stats + Intrinsic Sys.)	0.07	0.08	0.04	0.014
S1 (Stats + Intrinsic Sys. + HE Bias)	–	–	0.06	–
S1 (Stats + Intrinsic Sys. + Baryon Bias)	–	–	0.10	–
S4 (Stats + Intrinsic Sys.)	0.04	0.05	0.02	0.021
S4 (Stats + Intrinsic Sys. + HE Bias)	0.04	0.05	0.02	0.022
S4 (Stats + Intrinsic Sys. + Baryon Bias)	0.04	0.05	0.02	0.024

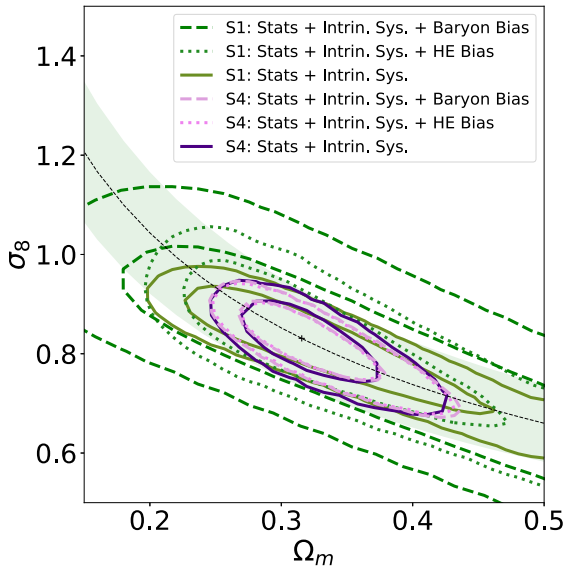

Figure 9. 1 and 2σ credibility regions in the $\Omega_m - \sigma_8$ plane from the analysis of the synthetic data set with different error assumptions for the S1 and S4 cases, respectively. The best-fitting values of Ω_m and σ_8 from the different parameter inferences coincide with the values of the fiducial cosmology marked by the cross-point. The dashed line and the green-shaded area corresponds to curves of constant $S_8 = 0.852 \pm 0.104$ values, which is the mean and standard deviation of S_8 from the MCMC chains of S1 case with intrinsic systematic errors and baryon bias. We can see that using multiple sparsity estimates breaks the S_8 degeneracy.

for the HE bias and $\sigma_{S_8} = 0.10$ for the baryon bias. The latter case is shown in Fig. 9 as the green-shaded region around the curve of constant best-fitting value of $\hat{S}_8 = 0.85$. Such a result is consistent with the constraint obtained in Corasaniti et al. (2018) from the analysis of $s_{500,\,1000}$ of a sample of $\sim$100 X-ray clusters.

In the S4 case, the constraints significantly improve when compared to the single sparsity measurements. As we can see in Fig. 9, the use of additional sparsities indeed breaks the S_8 degeneracy. From the values quoted in Table 5, we notice that accounting for the HE and baryon bias slightly alter the area under the 1σ credibility contours, with the baryon bias case corresponding to the larger value and that with the intrinsic systematics only corresponding to the smallest value. This is consistent with the difference in amplitude of the systematic shifts between the HE and baryon case, respectively. Nevertheless, we can see that such difference have no impact on the marginalized 1σ errors on the cosmological parameters.

Nevertheless, it is important to notice that such bias effects were estimated using results of *N*-body/hydro simulations that were not specifically devoted to the study of the halo sparsity. Hence, we advocate for a more in-depth study of the influence of baryonic processes on the mass profile of haloes as traced by sparsity measurements, which we leave for a future study.

Finally, we would like to stress that for this type of cosmological parameter inference to be possible, independent cluster mass measurements at multiple overdensity need to be carried out. This implies adopting new methodologies that abandon the two-parameters NFW fitting profile in favour of more general, non-parametric approaches (see e.g. Ettori et al. 2013, for a review). Recent examples of these procedures to infer the galaxy cluster mass profile with a non-parametric method have been presented in Bartalucci et al. (2018) to derive NFW-independent sparsity estimates of a sample of high-redshift clusters, and in Bartalucci et al. (2019) to test against the standard forward/backward NFW methods. More recently, Eckert et al. (2022) have developed a forward non-parametric method to derive mass profiles independent from any functional form of the potential. Weak-lensing observations can also provide non-parametric estimates of the mass profile (and consequently of the cluster sparsity) through mass aperture measurements (see e.g. Debackere et al. 2022, for a recent study).

5 SUMMARY AND DISCUSSION

The gravitational mass assembly process that leads to the formation of dark matter haloes, which host galaxy groups and clusters, imprints cosmological information on the halo mass profiles. This can be retrieved through measurements of the halo sparsity, i.e. the ratio between halo masses enclosing two different overdensities, which has been shown to provide a non-parametric proxy for the halo internal mass distribution (Balmès et al. 2014). In the past few years, cosmological constraints have been inferred from measurements of the average sparsity of galaxy cluster samples using HE masses at $\Delta = 500\rho_c$ and $1000\rho_c$ from X-ray observations (Corasaniti et al. 2018) and weak-lensing masses at $\Delta = 200\rho_c$ and $500\rho_c$ (Corasaniti et al. 2021). However, cosmological information is encoded over the entire halo mass profile, rather than at only two overdensities.

Here, we have investigated the use of multiple sparsity measurements from halo mass estimates at several overdensity as a probe of the cosmological imprint on the halo mass profile. For this purpose, we have analysed *N*-body halo catalogues from the Raygal and M2Csims simulations and estimated the correlation among different sparsities as a function of redshift. In particular, we have focused on halo masses evaluated at four different overdensities, thus allowing to estimate a total of six sparsities. Interestingly, we find that, among these sparsities, those associated with the mass distribution in distinct spherical halo shells are not highly correlated. Thus, indicating that there is additional cosmological information encoded in the average halo mass profile, which can be exploited through multiple sparsity measurements. In contrast, sparsities obtained using mass estimates derived from the NFW best-fitting density profile to the *N*-body haloes result in correlations that are close to unity and significantly different from those inferred from the analysis of the SO *N*-body halo masses. This suggests that imposing a NFW profile to haloes performs a strong compression that misses cosmological information imprinted on different regions of the halo mass profile.

To assess the constraining power of multiple sparsity measurements, we have performed an MCMC likelihood analysis of synthetic generated data sets from the Raygal and M2Csims simulations consisting of different number of sparsities, from a single sparsity

case up to a total of six, and inferred cosmological parameter constraints on Ω_m and σ_8. We find the constraints to improve as the number of sparsities used increases, with a maximal effect for the case with four sparsities. Instead, the constraints saturates beyond the four sparsity case, which suggest that the additional sparsity estimates only provide redundant information. We have also performed a forecast analysis for a synthetic data set of four average sparsity measurements generated assuming the characteristic of a realistic cluster sample such as that from the CHEX-MATE project. We have inferred cosmological parameter constraints for different errors assumptions, including the impact of systematic effects on sparsities due to baryons or deviations from the HE, from mass bias estimates obtained from past studies using N-body/hydrodynamical simulations. The results show that these effects only mildly impact the cosmological parameter inference, although dedicated numerical studies are still needed to derive more accurate predictions for baryon systematics on sparsity measurements.

It has been long considered that cosmological information encoded in the halo density profile can be retrieved through measurements of the concentration–mass relation (see e.g. Ettori et al. 2010). The observational challenges posed by the necessity of having accurate measurements of the concentration parameter of galaxy clusters has been the primary limitation for the use of such an approach (Mead et al. 2010; King & Mead 2011; Sereno et al. 2015).

Our study not only shows that the use of halo sparsity provides a more direct and simpler way to access such information as already discussed in past analyses (Corasaniti et al. 2018, 2021), but also that multiple sparsity measurements can fully exploit the cosmological signal imprinted in the mass profile, which would be otherwise missed if the halo density profile was assumed to be NFW. Because of this, we encourage the development of methodologies capable of providing independent mass estimates at different overdensities free of the assumption of the NFW profile.

ACKNOWLEDGEMENTS

The authors are thankful to Romain Teyssier for his role in the development of the M2Csims simulations suite. AMCLB is grateful to Christian Arnold and Baojiu Li for granting access to their Dirac allocation on COSMA for running pSOD on the M2Csims simulations. AMCLB was supported by the French Agence Nationale de la Recherche under grant ANR-11-BS56-015 and by the European Research Council under the European Union Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement number 340519 while conducting the M2Csims simulation programme. AMCLB is currently supported by a fellowship of PSL University at the Paris Observatory. PSC, TR and YR acknowledge support from the DIM ACAV of the Region Ile-de-France. SE acknowledges financial contribution from the contracts ASI-INAF Athena 2019-27-HH.0, 'Attività di Studio per la comunità scientifica di Astrofisica delle Alte Energie e Fisica Astroparticellare' (Accordo Attuativo ASI-INAF n. 2017-14-H.0), INAF mainstream project 1.05.01.86.10, and from the European Union Horizon 2020 Programme under the AHEAD2020 project (grant agreement no. 871158). This research was supported by the Munich Institute for Astro- and Particle Physics (MIAPP), which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311. The work presented here was granted access to HPC resources of TGCC/CINES through allocations made by GENCI (Grand Equipement National de Calcul Intensif) under the allocations 2016-042287, 2017-A0010402287, 2018-A0030402287, 2019-A0050402287, and 2020-A0070402287 for RayGal and of CINES under allocations 2015-047350, 2016-047350, 2017-A002047350, 2018-A004047350, 2019-A006047350, and 2020-A008047350 for M2Csims, respectively. 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.

DATA AVAILABILITY

The data which were used in the study presented here will be made available upon reasonable request to the corresponding authors. Data from the Raygal simulation suite, including light-cone halo catalogues accounting for relativistic effects, are publicly available at https://cosmo.obspm.fr/public-datasets/.

REFERENCES

Allen S. W., Rapetti D. A., Schmidt R. W., Ebeling H., Morris R. G., Fabian A. C., 2008, MNRAS, 383, 879
Anand G. S., Tully R. B., Rizzi L., Riess A. G., Yuan W., 2022, ApJ, 932, 15
Aver E., Olive K. A., Skillman E. D., 2015, J. Cosmol. Astropart. Phys., 2015, 011
Balmès I., Rasera Y., Corasaniti P. S., Alimi J. M., 2014, MNRAS, 437, 2328
Bartalucci I., Arnaud M., Pratt G. W., Le Brun A. M. C., 2018, A&A, 617, A64
Bartalucci I., Arnaud M., Pratt G. W., Démoclès J., Lovisari L., 2019, A&A, 628, A86
Beutler F. et al., 2017, MNRAS, 466, 2242
Biffi V. et al., 2016, ApJ, 827, 112
Birrer S. et al., 2020, A&A, 643, A165
Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
Bocquet S. et al., 2019, ApJ, 878, 55
Bocquet S., Heitmann K., Habib S., Lawrence E., Uram T., Frontiere N., Pope A., Finkel H., 2020, ApJ, 901, 5
Boyarsky A., Drewes M., Lasserre T., Mertens S., Ruchayskiy O., 2019, Prog. Part. Nucl. Phys., 104, 1
Brax P., 2018, Rep. Prog. Phys., 81, 016902
Breton M.-A., Rasera Y., Taruya A., Lacombe O., Saga S., 2019, MNRAS, 483, 2671
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
CHEX-MATE Collaboration, 2021, A&A, 650, A104
Castro T., Borgani S., Dolag K., Marra V., Quartin M., Saro A., Sefusatti E., 2021, MNRAS, 500, 2316
Cole S. et al., 2005, MNRAS, 362, 505
Cooke R. J., Pettini M., Steidel C. C., 2018, ApJ, 855, 102
Copeland E. J., Sami M., Tsujikawa S., 2006, Int. J. Mod. Phys. D, 15, 1753
Corasaniti P. S., Rasera Y., 2019, MNRAS, 487, 4382
Corasaniti P. S., Ettori S., Rasera Y., Sereno M., Amodeo S., Breton M. A., Ghirardini V., Eckert D., 2018, ApJ, 862, 40
Corasaniti P. S., Giocoli C., Baldi M., 2020, Phys. Rev. D, 102, 043501
Corasaniti P.-S., Sereno M., Ettori S., 2021, ApJ, 911, 82
Courtin J., Rasera Y., Alimi J. M., Corasaniti P. S., Boucher V., Füzfa A., 2011, MNRAS, 410, 1911
Croft R. A. C. et al., 2016, MNRAS, 457, 3541
de Mattia A. et al., 2021, MNRAS, 501, 5616
de Sainte Agathe V. et al., 2019, A&A, 629, A85
DES Collaboration, 2022, Phys. Rev. D, 105, 043512
Debackere S. N. B., Hoekstra H., Schaye J., Heitmann K., Habib S., 2022, MNRAS, 515, 3383
Despali G., Giocoli C., Angulo R. E., Tormen G., Sheth R. K., Baso G., Moscardini L., 2016, MNRAS, 456, 2486

Eckert D., Ettori S., Pointecouteau E., van der Burg R. F. J., Loubser S. I., 2022, A&A, 662, A123
Ettori S., Tozzi P., Rosati P., 2003, A&A, 398, 879
Ettori S., Morandi A., Tozzi P., Balestra I., Borgani S., Rosati P., Lovisari L., Terenziani F., 2009, A&A, 501, 61
Ettori S., Gastaldello F., Leccardi A., Molendi S., Rossetti M., Buote D., Meneghetti M., 2010, A&A, 524, A68
Ettori S., Donnarumma A., Pointecouteau E., Reiprich T. H., Giodini S., Lovisari L., Schmidt R. W., 2013, Space Sci. Rev., 177, 119
Ettori S. et al., 2019, A&A, 621, A39
Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, ApJ, 473, 576
Gatti M. et al., 2021, MNRAS, preprint (arXiv:2110.10141)
Green A. M., Kavanagh B. J., 2021, J. Phys. G Nucl. Phys., 48, 043001
King L. J., Mead J. M. G., 2011, MNRAS, 416, 2539
Kitayama T., Suto Y., 1996, MNRAS, 280, 638
Komatsu E. et al., 2011, ApJS, 192, 18
Lacey C., Cole S., 1994, MNRAS, 271, 676
Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, MNRAS, 441, 1270
Le Brun A. M. C., Arnaud M., Pratt G. W., Teyssier R., 2018, MNRAS, 473, L69
Lesci G. F. et al., 2022, A&A, 659, A88
Lewis A., 2019, preprint (arXiv:1910.13970)
Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., von der Linden A., Schmidt R. W., 2014, MNRAS, 440, 2077
Mantz A. B. et al., 2022, MNRAS, 510, 131
Marulli F., Veropalumbo A., García-Farieta J. E., Moresco M., Moscardini L., Cimatti A., 2021, ApJ, 920, 13
McClintock T. et al., 2019, ApJ, 872, 53
Mead J. M. G., King L. J., Sijacki D., Leonard A., Puchwein E., McCarthy I. G., 2010, MNRAS, 406, 434
Meneghetti M., Rasia E., Merten J., Bellagamba F., Ettori S., Mazzotta P., Dolag K., Marri S., 2010, A&A, 514, A93
Mukherjee P., Parkinson D., Corasaniti P. S., Liddle A. R., Kunz M., 2006, MNRAS, 369, 1725
Nagai D., Vikhlinin A., Kravtsov A. V., 2007, ApJ, 655, 98
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Niemeyer J. C., 2020, Prog. Part. Nucl. Phys., 113, 103787
Nishimichi T. et al., 2019, ApJ, 884, 29
Pacaud F. et al., 2018, A&A, 620, A10
Percival W. J. et al., 2001, MNRAS, 327, 1297
Planck Collaboration XX, 2014a, A&A, 571, A20
Planck Collaboration XXI, 2014b, A&A, 571, A21
Planck Collaboration XIII, 2016a, A&A, 594, A13
Planck Collaboration XXIV, 2016b, A&A, 594, A24
Planck Collaboration VI, 2020, A&A, 641, A6
Rasera Y. et al., 2022, A&A, 661, A90
Rasia E. et al., 2012, New J. Phys., 14, 055018
Richardson T. R. G., Corasaniti P. S., 2022, MNRAS, 513, 4951
Riess A. G. et al., 2022, ApJ, 934, L7
Roy V., 2020, Annu. Rev. Stat. Appl., 7, 387
Roy F., Bouillot V. R., Rasera Y., 2014, A&A, 564, A13
Schaye J. et al., 2010, MNRAS, 402, 1536
Sereno M., Ettori S., 2015, MNRAS, 450, 3633
Sereno M., Giocoli C., Ettori S., Moscardini L., 2015, MNRAS, 449, 2024
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Taylor J. E., 2011, Adv. Astron., 2011, 604898
Tegmark M. et al., 2004, ApJ, 606, 702
Teyssier R., 2002, A&A, 385, 337
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
To C. et al., 2021, Phys. Rev. Lett., 126, 141301
Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, MNRAS, 442, 2641
Wang J. et al., 2011, MNRAS, 413, 1373
Wong K. C. et al., 2020, MNRAS, 498, 1420

APPENDIX A: HALO MASS FUNCTION BEST-FITTING COEFFICIENTS

We use the halo mass functions from the Raygal and M2Csims halo catalogues to fit the coefficients of the mass function parametrization given by equation (6), which we determine using a Levenberg–Marquardt minimization scheme. The multiplicity functions at $\Delta = 200$, 500, 1000, and 2500 estimated from the Raygal and M2Csims halo catalogues are shown in Fig. A1. Given the proximity of the simulated cosmological models, we can see that the estimated multiplicity functions are in good agreement with one another. Notice that since the Raygal simulation probes a slightly larger cosmic volume than the M2Csims simulations, the corresponding multiplicity functions extend over larger $\ln \sigma^{-1}$ values. Conversely, the M2Csims simulations have slightly better mass resolution, thus probing smaller $\ln \sigma^{-1}$ values than Raygal multiplicity functions.

In order to predict the average sparsity at redshifts different from those probed by the simulation snapshots using equation (3), we introduce the redshift-dependent variable $x = \log_{10}(\Delta/\Delta_{\rm vir}(z))$, where $\Delta_{\rm vir}(z)$ is the virial overdensity as given by the formula derived in Bryan & Norman (1998), then following Despali et al. (2016), we parametrize the redshift evolution of the best-fitting ST coefficients as a quadratic function of x:

$$\theta_\Delta = c_0 + c_1 \cdot x + c_2 \cdot x^2, \tag{A1}$$

where $\theta_\Delta = \{A_\Delta, a_\Delta, p_\Delta\}$. In Table A1 and A2, we quote the values of the quadratic parametrizations for the ST-coefficients.

In Fig. A2, we plot the predictions from equation (3) for the Raygal cosmology using the M2Csims calibrated multiplicity functions against the average halo sparsity estimates from the Raygal halo catalogues (left-hand panel) and the predictions for the M2Csims cosmology using the Raygal calibrated multiplicity functions against the average halo sparsity estimates from the M2Csims halo catalogues (right-hand panel). As we can see differences are $\lesssim 5$ per cent, consistent with those shown in Fig. 2.

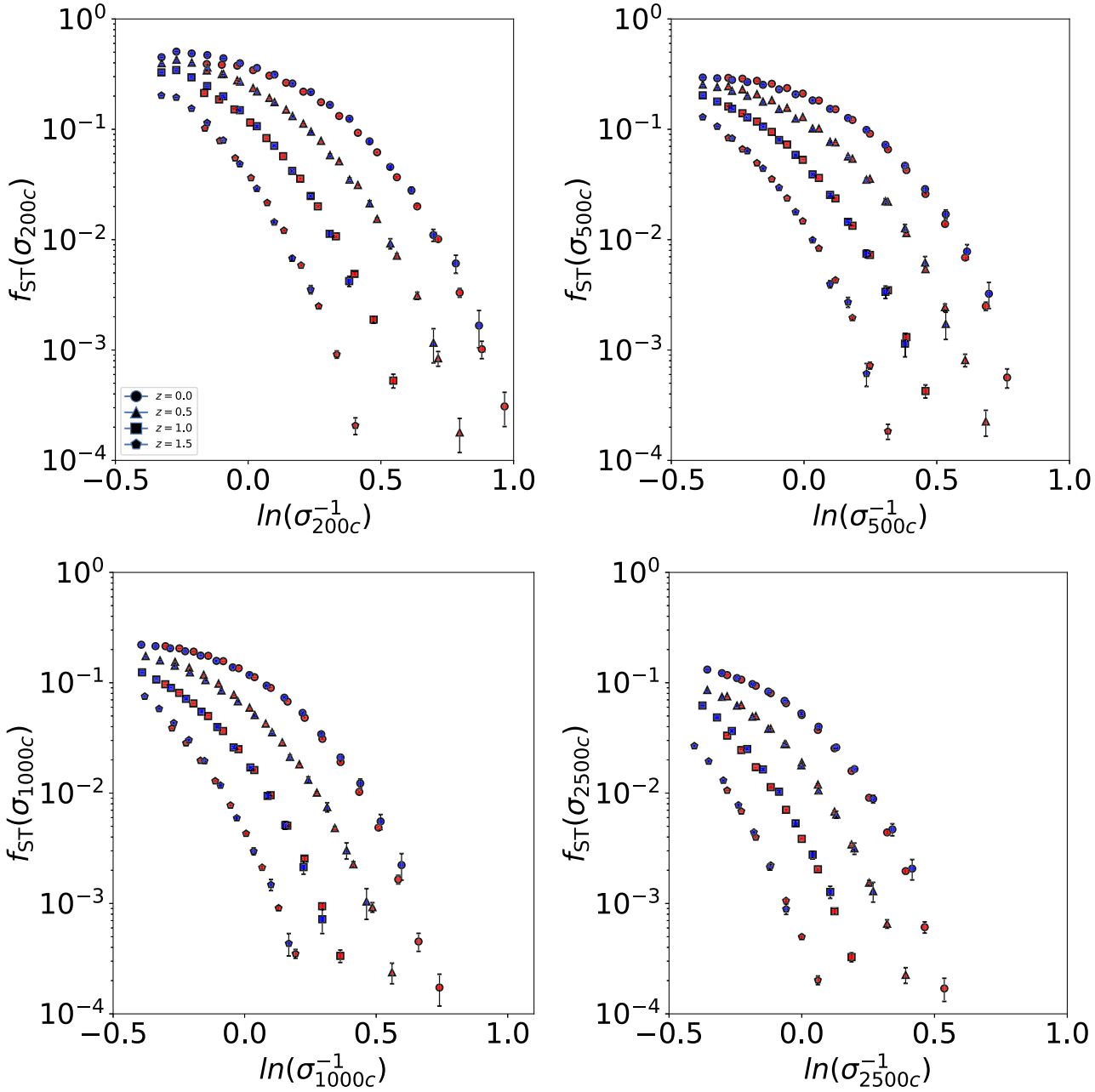

Figure A1. Multiplicity function at $\Delta = 200$ (top-left panel), 500 (top-right panel), 1000 (bottom-left panel), and 2500 (bottom-right panel) from the Raygal (red points) and M2Csims (blue points) simulations, respectively at $z = 0.0$ (circles), 0.5 (triangles), 1.0 (squares), and 1.5 (pentagons). Given the proximity of the simulated cosmologies, the multiplicity functions estimated from the Raygal and M2Csims halo catalogues are consistent with one another within Poisson errors over the common range of masses probed by the simulations.

Table A1. Coefficients of the quadratic function of x parametrizing the redshift evolution of the ST parameters for the Raygal halo mass functions.

	c_0	c_1	c_2
A_{200c}	−0.134134392	5.018016486	−10.7419621
a_{200c}	0.989977371	−0.5291517036	4.430318949
p_{200c}	−1.688969793	10.0338655815	22.480766636
A_{500c}	−0.52885598	2.4986942	−1.91309251
a_{500c}	−0.408103836	3.5426469558	−1.249831194
p_{500c}	0.380460633	−2.894905482	1.9075409115
A_{1000c}	−1.20684329	3.07924433	−1.65792698
a_{1000c}	−1.9777704111	4.8809416194	−1.1738202399
p_{1000c}	1.856153925	−4.089186069	1.5184593426
A_{2500c}	7.1510462212	−10.4752748702	3.96694467
a_{2500c}	−9.5732094279	13.1786821425	−3.257783121
p_{2500c}	52.2138579936	−77.6915082432	28.3514283168

Table A2. As in Table A1 but for the M2Csims halo mass functions.

	c_0	c_1	c_2
A_{200c}	−0.1392442736	5.2091790188	−11, 1511797962
a_{200c}	0.9996830315	−0.534339383	4.4737534485
p_{200c}	−1.8134201988	10.7732030454	−24.1372441776
A_{500c}	−1.88354029	7.56649037	−6.128838
a_{500c}	1.32586081	−2.18316668	2.90408494
p_{500c}	−6.74606335	21.7532933	−17.97632164
A_{1000c}	−5.13368556	11.95696884	−6.5251012
a_{1000c}	3.98834671	−7.94206751	5.54853335
p_{1000c}	−20.43015395	45.10712761	−25.21860361
A_{2500c}	−8.03447658	12.44996314	−4.704367
a_{2500c}	3.06066127	−4.70146171	3.01797036
p_{2500c}	−42.49723101	64.44575098	−24.79541975

APPENDIX B: SPARSITY CORRELATION COEFFICIENT FITTING FUNCTIONS

The redshift evolution of the average sparsity correlation coefficients shown in Fig. 3 is well approximated by a linear relation:

$$r_{s_1,s_2}(z) = q + m \cdot z, \tag{B1}$$

where the coefficients of the linear regression are given in Tables B1 and B2 for the Raygal and M2Csims halo catalogues, respectively.

Table B1. Linear regression parameters of sparsity correlation coefficients from the Raygal halo catalogues.

Raygal	m	q
$r_{s_{200,500},s_{200,1000}}$	-0.032 ± 0.005	0.898 ± 0.005
$r_{s_{200,500},s_{200,2500}}$	-0.14 ± 0.01	0.71 ± 0.01
$r_{s_{200,500},s_{500,1000}}$	-0.08 ± 0.01	$0.48 + 0.01$
$r_{s_{200,500},s_{500,2500}}$	-0.19 ± 0.01	0.37 ± 0.01
$r_{s_{200,500},s_{1000,2500}}$	-0.22 ± 0.01	0.24 ± 0.01
$r_{s_{200,1000},s_{200,2500}}$	-0.081 ± 0.002	0.902 ± 0.002
$r_{s_{200,1000},s_{500,1000}}$	-0.011 ± 0.002	0.810 ± 0.002
$r_{s_{200,1000},s_{500,2500}}$	-0.132 ± 0.003	0.666 ± 0.003
$r_{s_{200,1000},s_{1000,2500}}$	-0.188 ± 0.004	0.461 ± 0.004
$r_{s_{200,2500},s_{500,1000}}$	-0.030 ± 0.002	0.855 ± 0.002
$r_{s_{200,2500},s_{500,2500}}$	-0.003 ± 0.003	0.905 ± 0.003
$r_{s_{200,2500},s_{1000,2500}}$	-0.024 ± 0.005	0.784 ± 0.004
$r_{s_{500,1000},s_{500,2500}}$	-0.076 ± 0.002	0.858 ± 0.001
$r_{s_{500,1000},s_{1000,2500}}$	-0.127 ± 0.003	0.619 ± 0.003
$r_{s_{500,2500},s_{1000,2500}}$	-0.006 ± 0.001	0.927 ± 0.001

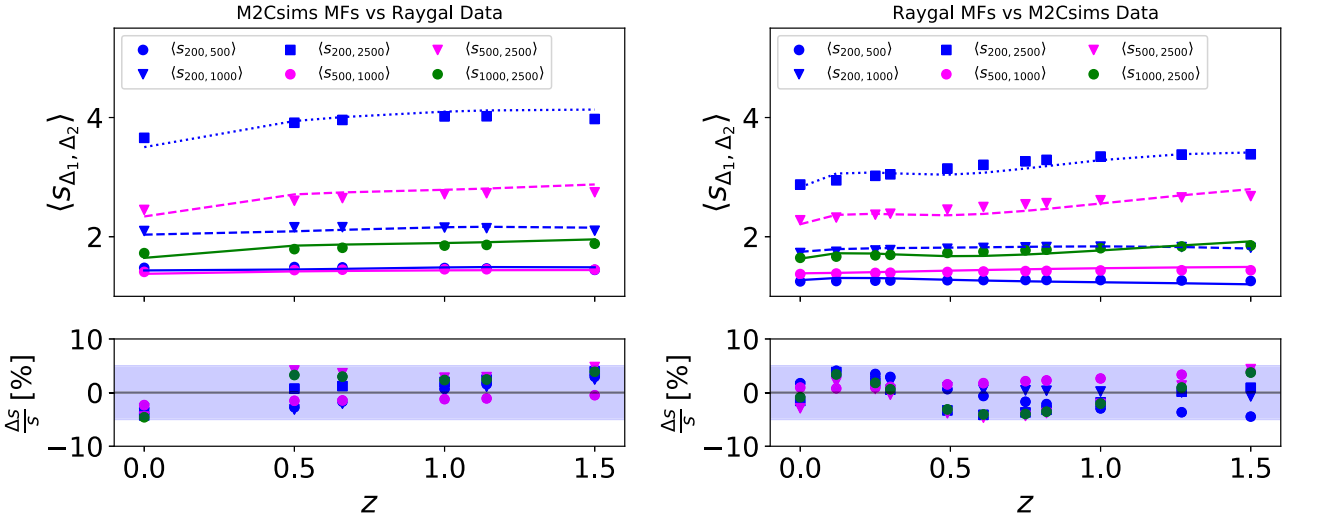

Figure A2. Left-hand panel: average halo sparsity for different overdensity configurations as a function of redshift predicted by solving equation (3) using the M2Csims calibrated multiplicity functions for the Raygal cosmology plotted against the average sparsities from the Raygal halo catalogues. Right-hand panel: average halo sparsities predicted by solving equation (3) using the Raygal calibrated multiplicity functions for the M2Csims cosmology plotted against the average sparsities from the M2Csims halo catalogues. The lower panels show the relative difference between the predictions and the *N*-body halo catalogue estimates.

Table B2. Linear regression parameters of sparsity correlation coefficients from the M2Csims halo catalogues.

M2Csims	m	q
$r_{s_{200,500},s_{200,1000}}$	-0.028 ± 0.003	0.901 ± 0.002
$r_{s_{200,500},s_{200,2500}}$	-0.134 ± 0.006	0.726 ± 0.005
$r_{s_{200,500},s_{500,1000}}$	-0.079 ± 0.004	$0.48 + 0.01$
$r_{s_{200,500},s_{500,2500}}$	-0.182 ± 0.003	0.395 ± 0.003
$r_{s_{200,500},s_{1000,2500}}$	-0.212 ± 0.004	0.262 ± 0.003
$r_{s_{200,1000},s_{200,2500}}$	-0.074 ± 0.001	0.909 ± 0.001
$r_{s_{200,1000},s_{500,1000}}$	-0.014 ± 0.003	0.826 ± 0.003
$r_{s_{200,1000},s_{500,2500}}$	-0.123 ± 0.003	0.685 ± 0.003
$r_{s_{200,1000},s_{1000,2500}}$	-0.170 ± 0.004	0.478 ± 0.003
$r_{s_{200,2500},s_{500,1000}}$	-0.028 ± 0.002	0.870 ± 0.002
$r_{s_{200,2500},s_{500,2500}}$	-0.003 ± 0.003	0.910 ± 0.002
$r_{s_{200,2500},s_{1000,2500}}$	-0.019 ± 0.004	0.787 ± 0.003
$r_{s_{500,1000},s_{500,2500}}$	-0.070 ± 0.001	0.869 ± 0.001
$r_{s_{500,1000},s_{1000,2500}}$	-0.112 ± 0.003	0.635 ± 0.002
$r_{s_{500,2500},s_{1000,2500}}$	-0.002 ± 0.001	0.926 ± 0.001

This paper has been typeset from a TEX/LATEX file prepared by the author.