Quantifying the statistics of CMB-lensing-derived galaxy cluster mass measurements with simulations

Ínigo Zubeldia\textsuperscript{1,2,*} and Anthony Challinor\textsuperscript{1,2,3,†}

\textsuperscript{1}Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
\textsuperscript{2}Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK
\textsuperscript{3}DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
CMB lensing is a promising, novel way to measure galaxy cluster masses that can be used, e.g., for mass calibration in galaxy cluster counts analyses. Understanding the statistics of the galaxy cluster mass observable obtained with such measurements is essential if their use in subsequent analyses is not to lead to biased results. We study the statistics of a CMB lensing galaxy cluster mass observable for a Planck-like experiment with mock observations obtained from an $N$-body simulation. We quantify the bias, intrinsic scatter, and deviations from log-normality associated with this observable following two different approaches, one in which the signal due to the cluster and nearby correlated large-scale structure is isolated, and another one in which the variation due to uncorrelated large-scale structure is also taken into account. We briefly discuss how some of our results change for experiments with higher angular resolution and lower noise levels, such as the current generation of surveys obtained with ground-based, large-aperture telescopes.

Key words: cosmology – cosmic microwave background radiation – galaxies: clusters: general

1 INTRODUCTION
As the largest gravitationally-bound structures in the Universe, galaxy clusters are powerful cosmological probes (Allen et al. 2011; Pratt et al. 2019). In particular, their abundance as a function of mass and redshift, as given by the halo mass function, depends on the assumed cosmological model and on the values its parameters. Cosmological information can therefore be extracted from the observed galaxy cluster abundance in what is known as ‘cluster counts’ analyses. In a spatially-flat $\Lambda$CDM cosmology, this abundance is particularly sensitive to the mean matter density of the Universe, which can be parametrised by $\Omega_m$, and to the amplitude of the matter perturbations, which can be characterised by $\sigma_8$, the root mean square of the linear density fluctuations smoothed on a scale of $8 h^{-1}$ Mpc. Cluster masses are, however, not directly observable, so cluster counts studies must rely on one or several cluster observables known to scale with cluster mass and use them as ‘mass proxies’. These observables can come from observations across different regions of the electromagnetic spectrum and include optical richness, X-ray flux, Sunyaev–Zel’dovich (SZ) flux, and lensing mass estimates, all of which are known to trace cluster masses. In recent years, studies using such observables have yielded competitive cosmological constraints (e.g., Mantz et al. 2010; Hasselfield et al. 2013; Planck 2015 Results XXVII 2016; de Haan et al. 2016; Bocquet et al. 2019; Costanzi et al. 2019; Zubeldia & Challinor 2019; Kirby et al. 2019). In these analyses, scaling relations are needed in order to relate accurately the cluster observables to the cluster mass, since it is the dependence of the cluster abundance on the latter that can be theoretically predicted. These scaling relations need to be calibrated, and this is typically a difficult task which often brings in significant systematic uncertainty. Indeed, currently, the determination of cluster masses constitutes the largest source of uncertainty in cluster counts studies (see Pratt et al. 2019 for a recent review of cluster mass calibration).

Lensing observations can be very useful in this respect, since they provide almost bias-free estimates of cluster masses. In SZ counts analyses, in which an SZ-derived observable is used both to select the sample and as a precise (i.e., high signal-to-noise) cluster mass proxy for each cluster, the overall mass scale can be determined through exact (i.e., unbiased, or nearly unbiased) lensing mass estimates. These lensing-derived masses generally either have a significantly lower signal-to-noise ratio per cluster than the SZ
observable, are much more costly to obtain, or a combination of both. These lensing mass estimates can be obtained from galaxy lensing observations, a well-established practice (e.g., as in two of the calibrations used in the baseline analysis in Planck 2015 Results XXVII 2016, in de Haan et al. 2016, and in Bocquet et al. 2019), or from CMB observations, a more novel approach (as in the third calibration of the baseline analysis in Planck 2015 Results XXVII 2016, and in Zubeldía & Challinor 2019).

First considered in Seljak & Zaldarriaga (2000), CMB lensing by clusters has attracted some interest in recent years. Several mass estimators have been developed (see, e.g., Yoo & Zaldarriaga 2008; Melin & Bartlett 2015; Raghunathan et al. 2017; Horowitz et al. 2019), and the CMB lensing signal of galaxy clusters and, in general, halos, has been detected to moderate-to-high statistical significance in various recent works (Baxter et al. 2015; Madhavacheril et al. 2015; Planck 2015 Results XXVII 2016; Baxter et al. 2017; Raghunathan et al. 2019b; Zubeldía & Challinor 2019; Raghunathan et al. 2019a). CMB lensing cluster mass estimation has several intrinsic virtues, particularly in the context of SZ surveys. First, mass estimates can be obtained from the same data set as the SZ observable, and are relatively cheap to obtain, in contrast to most current galaxy-lensing-derived masses. This enables one to obtain CMB lensing mass estimates for the totality of the clusters in the SZ sample in a relatively straightforward way provided that redshift measurements of the clusters are available (Planck 2015 Results XXVII 2016; Zubeldía & Challinor 2019), as opposed to most current galaxy lensing mass determinations, in which typically the masses of only a small subset of clusters in the SZ sample are obtained. Moreover, the signal-to-noise of CMB lensing mass estimates does not decrease strongly with redshift (see, e.g., Melin & Bartlett 2015), whereas high-redshift galaxy lensing mass determinations can suffer from the lack of a sufficient number of background galaxies. In addition, the CMB lensing signal is not affected by the uncertainties in the photometric redshifts of the background galaxies, which are a limiting factor in galaxy lensing analyses. However, for an experiment like Planck, CMB lensing mass estimates are much noisier than state-of-the-art galaxy lensing mass estimates (e.g., von der Linden et al. 2014; Hoekstra et al. 2015), with typical signal-to-noise ratios of a fraction of unity for large clusters (Zubeldía & Challinor 2019). Furthermore, being a more novel technique, cluster CMB lensing mass estimation methods have not been tested to the extent that galaxy lensing methods have (see, e.g., Raghunathan et al. 2017 for a quantification of several possible systematics that can affect CMB lensing cluster mass measurements).

A fundamental element of a cluster cosmological analysis that needs to be determined accurately in order for the analysis to deliver unbiased cosmological information is the relation between the mass observable (or observables) and the true cluster mass. This involves determining both the scaling relation that relates some mean value of the observable(s) with the cluster true mass, accounting for any possible biases, and the statistical scatter that exists around the mean value of the observable(s), which typically has a complex origin and is difficult to predict from first principles (Allen et al. 2011). More rigorously, what needs to be specified is the conditional probability density function followed by the mass observable(s), obs, conditioned on the value of the true cluster mass, $M_{500}$, $P(\text{obs}|M_{500})$. We note that here and throughout we choose as a cluster’s ‘true mass’ the mass within a sphere within which the mean density is 500 times the critical density at the cluster’s redshift, and we denote it with $M_{500}$. We also note that this conditional probability density function can also be conditioned on other variables, such as redshift $z$; we omit these possible additional conditioning variables here for concision, but $P(\text{obs}|M_{500})$ should be thought of as potentially having more conditioning variables implicit. If a functional form is assumed for $P(\text{obs}|M_{500})$, its parameters can be self-calibrated with real data. However, doing this can have a negative impact on the statistical power of the cosmological analysis, especially if the calibration is carried out at the cosmological analysis level. An alternative approach is to use cosmological simulations in order to determine $P(\text{obs}|M_{500})$, or at least to inform the likely range of parameter values of some assumed functional form. This also makes it possible to assess whether the chosen functional form is a good description of $P(\text{obs}|M_{500})$. However, this approach is obviously limited by the accuracy of the simulations themselves, something that is difficult to quantify.

The distribution $P(\text{obs}|M_{500})$ has been widely studied in the literature from both simulations and observations for a number of cluster observables (e.g., Pratt et al. 2009; Becker & Kravtsov 2011; Angulo et al. 2012; Bahé et al. 2012; Rasia et al. 2012; Shirasaki et al. 2016; Geach & Peacock 2017). Nevertheless, it has never been studied for a CMB lensing mass observable. This is the aim of this paper: to quantify $P(\text{obs}|M_{500})$, where obs is the CMB lensing mass observable that is used in Zubeldía & Challinor (2019), the CMB lensing signal-to-noise $p_{\text{obs}}$. We follow the simulation approach, producing mock observations of $p_{\text{obs}}$ for all the clusters with $M_{500} > 2 \times 10^{14} M_\odot$ in two snapshots of a large N-body cosmological simulation, BigMDPL (Klypin et al. 2016). We think of $p_{\text{obs}}$ as consisting of three different contributions,

$$p_{\text{obs}} = p_c + \Delta p_a + \Delta p_n. \tag{1}$$

Here, $p_c$ is the contribution to the lensing mass observable from the cluster itself, which includes the variation due to cluster triaxiality, and from large scale structure (LSS) correlated with the cluster. The other two terms, $\Delta p_a$ and $\Delta p_n$, are the contributions originating from LSS uncorrelated with the cluster and from lensing reconstruction noise, respectively. We follow two different approaches in order to analyse our mock observations. In the first approach, which we call our deconvolution approach, we treat both $\Delta p_a$ and $\Delta p_n$ as noise and then use our mock observations in order to characterise $p_c$. In the second approach, which we call our extrapolation approach, we only treat $\Delta p_a$ as noise, and then characterise $p = p_c + \Delta p_a$ with our mock observations, extrapolating our results to the full line-of-sight from $z = 0$ back to CMB last scattering.

The main motivation for this work is to justify the choice of priors imposed on the parameters $1 - b_{\text{CMBlens}}$ and $\sigma_{\text{CMBlens}}$, which quantify, respectively, the bias and intrinsic scatter in the CMB lensing mass measurements, in the likelihood analysis of Zubeldía & Challinor (2019). In particular, this is done with our deconvolution approach, which is the one followed in the hierarchical model of Zubeldía & Challinor (2019). However, we also hope that our methods to
study $P(\text{obs}|M_{500})$ and that our main insights may be useful for future cluster counts analyses that may use CMB lensing masses as a mass proxy, which otherwise will have to determine their appropriate conditional probability distribution $P(\text{obs}|M_{500})$. Throughout this paper we consider an idealised experiment with specifications similar to those of Planck; however, we also briefly consider how some of the results change if a different experimental set-up is considered.

This paper is organised as follows. First, in Section 2 we give a brief introduction to CMB lensing. In Section 3 we describe the cosmological simulations we use in order to obtain our mock observations, and in Section 4 we explain how we make such mock observations. Next, in Section 5 we present our results, characterising $P(\text{obs}|M_{500})$ for a Planck-like experiment. In this section, we first explain our model of the mock observations in Section 5.1, and then in Sections 5.2 and 5.3 we present our results for our deconvolution and extrapolation approaches, respectively. In Section 6 we consider how the extrapolation approach results change if different experiment specifications are considered, and we finally conclude in Section 7.

2 BASICS OF CMB LENSING

Massive bodies deflect light due to the effect of their gravity, a phenomenon known as gravitational lensing. CMB photons coming from the last-scattering surface are therefore deflected, the observed net effect being a remapping of the CMB emissions coming from the last-scattering surface are therefore a weighted integral of the gravitational potential along the undeflected line of sight. This integrated matter distribution can be reconstructed from the CMB through the non-Gaussian features that are imprinted by lensing. Several methods to reconstruct the lensing convergence (or, equivalently, the lensing potential) exist, the most computationally simple being based on quadratic estimators (Hu 2001; Hu & Okamoto 2002).

Lensing by galaxy clusters produces variations typically of order $10 \mu K$ in the measured CMB temperature (e.g., Lewis & Challinor 2006), which are large enough to be probed by experiments like Planck in a statistical way. If a cluster density profile is assumed and the cluster redshift is known, the cluster mass can be estimated by, e.g., fitting an expected cluster convergence profile to a non-parametrically reconstructed convergence, or by fitting the cluster model directly to the CMB maps.

3 SIMULATION AND CONVERGENCE MAPS

In our study we use two snapshots of the BigMDPL simulation, a large, state-of-the-art simulation part of the publicly-available MultiDark simulation suite (Klypin et al. 2016). It consists of $3.840^3$ particles of mass $2.4 \times 10^{10} h^{-1} M_{\odot}$ in a simulation box of $2.5h^{-1}$ Gpc with periodic boundary conditions and evolved in a flat $\Lambda$CDM cosmology with $H_0 = 67.8 \text{km s}^{-1}\text{Mpc}^{-1}$, $\Omega_m = 0.307$, baryon density parameter $\Omega_b = 0.048$, $\sigma_8 = 0.829$, and scalar spectral index $n_s = 0.96$. We consider the snapshots at $z = 0.23$ and $z = 0.52$.

We use the halo positions as determined with the BDM ("Bound Density Maximum") technique (Klypin & Holtzman 1997; Ribe et al. 2013) that are publicly available in the CosmoSim database. We then measure the spherical overdensity mass of each halo, $M_{500}$, using as density contrast $\Delta = 500$ defined with respect to the critical density. We note that we do not remove unbound particles in this process. We refer to $M_{500}$ as the halo (or cluster) ‘true mass’. We only consider clusters with $M_{500} > 2 \times 10^{14} M_{\odot}$ and that are not subhalos of another halo, which yields a cluster catalogue with a total of 60,391 clusters in the $z = 0.23$ snapshot and of 29,561 clusters in the $z = 0.52$ snapshot. This cluster range corresponds to the most massive clusters of the Universe and spans most of the clusters in the Planck MF3 cosmology sample, the sample used in the cosmological analysis in Planck 2015 Results XXVII (2016) and in Zubeldia & Challinor (2019); see Planck 2015 Results XXVII (2016) for how this sample is constructed.

In order to obtain our CMB lensing measurements, the key quantity that needs to be produced is the lensing convergence around the location of each cluster in our catalogue and along a given direction of observation. We generate convergence maps using the Born approximation; under this approximation, for a spatially-flat cosmology the lensing convergence can be obtained with Eq. (3).
We produce square convergence maps centred at the location of each cluster in our catalogue for both snapshots. Each map has an angular size of 128 arcmin along both directions perpendicular to the line of sight, and we always take the $\gamma$-direction, as defined within the simulation, as our line-of-sight direction. Obviously, with our simulation snapshots we cannot compute the total lensing convergence from $\chi = 0$ all the way back to last scattering, $\chi = \chi_{\gamma}$, as in Eq. (3). We instead restrict our integration limits to a relatively small interval of comoving distance. Specifically, we take each cluster to be at the comoving distance given by the redshift of the snapshot to which it belongs, $\chi_{\gamma}$, and we then consider seven sets of integration limits in Eq. (3), all of them centred at $\chi_{\gamma}$ but with an increasing comoving length $\chi_L = [\chi_{\gamma} - \chi_{\gamma}/2, \chi_{\gamma} + \chi_{\gamma}/2]$. We take $\chi_{\gamma}$, which hereafter we will refer to as ‘integration length’, to be logarithmic spaced between 5 Mpc and 400 Mpc, thus taking values of 5, 10.4, 21.5, 44.7, 92.8, 192.7, and 400 Mpc (comoving). We do this in order to investigate how the statistics of our CMB lensing mass observable depend on the relative amount of large scale structure (LSS) correlated and uncorrelated with the cluster that is present in the convergence maps. We note that the way in which we compute our convergence maps from the simulation snapshots neglects the temporal evolution of $\delta$ along the line of sight, since in a snapshot all the particle positions are given at the same time. The error we introduce by doing this should be negligible for the relatively small integration lengths (on cosmic scales) that we are considering. We do, however, take into account the change of the geometrical lensing kernel in Eq. (3) along the line of sight.

To summarise, for each cluster in our catalogue at any of the two considered redshifts we compute seven convergence maps, each of which corresponds to a different integration length. We refer to these maps as ‘cluster convergence maps’ at redshift $\chi$ with integration length $\chi_L$.

Furthermore, for each of the two snapshots we compute $10^5$ square convergence maps centred at random locations in the simulation box, with the same angular size as the cluster convergence maps (128 arcmin each direction), and with the two largest integration lengths used in the computation of the cluster convergence maps. We refer to these maps as ‘random convergence maps’ at redshift $\chi$ with integration length $\chi_L$, and describe their interpretation and use in Section 5.

4 MOCK OBSERVATIONS

In this section we describe our CMB lensing cluster mass observable and how we obtain mock measurements from our cluster convergence maps in order to produce a mock data set, which we then analyse in Section 5.

We think of each cluster convergence map with a given integration length as the ‘true’ convergence map around the location of that cluster due to the mass distribution contained within the considered integration length, and we denote it with $\kappa_\gamma$. As mentioned in Section 2, a fixed lensing convergence remaps the CMB anisotropies in a way that makes it possible for it to be estimated (or ‘reconstructed’) from CMB observations alone. Throughout this work, except in Section 6, we consider an idealised Planck-like CMB experiment with Gaussian beam with full-width at half maximum (FWHM) of 5 arcmin and with Gaussian instrumental noise with temperature noise levels of 45 $\mu$K arcmin. We do not study the impact of uncleaned or residual foregrounds on our results. As already noted above, in Section 6 we investigate how some of our results change with different experimental specifications.

We consider the simplest of the CMB lensing reconstruction techniques, the quadratic estimators (Hu 2001; Hu & Okamoto 2002), thus called because they are quadratic in the CMB fields. Specifically, we only use the $TT$ quadratic estimator, which produces a reconstructed convergence map out of two copies of a CMB temperature map. We make this choice because this estimator is close to the optimal lensing estimator for a Planck-like experiment. Using the flat-sky approximation, which is accurate given the small angular size of our convergence maps, the reconstructed convergence map that corresponds to one of our cluster convergence maps, $\hat{\kappa}$, can be written in Fourier space as

$$\hat{\kappa}(L) = \kappa_\gamma(L) + n(L).$$

Here, $\kappa_\gamma(L)$ is the Fourier transform of the ‘true’ convergence (the cluster convergence map), and $n(L)$ is the Fourier transform of the reconstruction noise, which in our simulated experiment has contributions from the primary CMB fluctuations and from instrumental noise. This noise is approximately Gaussian (Lewis & Challinor 2006). Following Hu et al. (2007), in our quadratic estimator implementation we impose a Fourier-space top-hat low-pass filter on the gradient leg of the estimator, zeroing the gradient map for modes with $L > 2000$. This ensures that the quadratic estimator remains unbiased around regions of large convergence, e.g., in the central regions of galaxy clusters. In addition, as is customary, we use the lensed CMB power spectrum in the weighting of the gradient leg of the estimator and in the normalisation, rather than the unlensed CMB power spectrum, as this gives (approximately) the correct response of the estimator $\hat{\kappa}(L)$ to lenses at wavevector $L$ averaged over all other lensing modes (Lewis et al. 2011; Hanson et al. 2011). This is the same form of the quadratic estimator that we used in our analysis of real Planck data in Zubeldia & Challinor (2019). It provides, to a good approximation, unbiased reconstructed convergence maps, i.e., $\langle \hat{\kappa}(L) \rangle = \kappa_\gamma(L)$, where angular brackets denote ensemble averaging over reconstruction noise, that is, over CMB and instrumental noise fluctuations.

In our study we do not produce reconstructed maps from our cluster convergence maps, $\kappa_\gamma$, but instead we directly use the cluster convergence maps themselves. The reasons for doing this are the following. First, as just noted, to a good approximation our lensing reconstruction method produces unbiased convergence estimates for the experimental specifications considered. In addition, the scatter of the final lensing mass observable arising from reconstruction noise is already well understood. Indeed, since to a good approximation the reconstruction noise is Gaussian, it is fully described by its variance, which can be written as

$$\langle \hat{\kappa}(L) \hat{\kappa}^*(L') \rangle = \delta^{(2)}(L - L') N_{\kappa}(L),$$

where $N_{\kappa}(L)$ is well approximated in our case by $N^{\text{opt}}(L)$, the normalisation of the quadratic estimator (see e.g., Hu 2001 for an analytic expression). As described below, our
mass observable is linear in the convergence; thus, its scatter due to reconstruction noise is also Gaussian to a good approximation.

The next step of our measurement pipeline is to match-filter our cluster convergence maps with a cluster convergence model in order to obtain a mass observable measurement for each cluster. We use exactly the same matched-filter implementation that was used in Zubeldia & Challinor (2019), and which follows Melin & Bartlett (2015). As a cluster model we adopt a truncated NFW profile (Navarro et al. 1997)

$$\rho(r) = \begin{cases} \frac{\rho_0}{r/r_s}\theta_\text{r}^2, & \text{if } r \leq R_\text{trunc}, \\ 0, & \text{if } r > R_\text{trunc}, \end{cases}$$

(6)

where \(r_s\) is a characteristic scale radius, \(R_\text{trunc}\) is the truncation radius, and \(\rho_0\) is a characteristic density, given by

$$\rho_0 = \rho_c(z) \frac{c_{500}^3}{c_{500}^3} \left( \frac{3}{\ln(1 + c_{500}) - c_{500}/(1 + c_{500})} \right),$$

(7)

where \(\rho_c(z)\) is the critical density at the cluster’s redshift, and \(c_{500}\) is the concentration parameter, which is defined as \(c_{500} = R_{500}/r_s\). As in Zubeldia & Challinor (2019), we choose \(R_{\text{trunc}} = 500 c_{500}/r_s\) and we fix \(c_{500} = 5\). Thus, in our model, a cluster is completely specified by two parameters, e.g., \(M_{500}\) and redshift, \(z\). We use this parameterisation throughout this paper, and refer to \(M_{500}\) and \(z\) as the ‘true cluster parameters’.

As detailed in Zubeldia & Challinor (2019), we can then write the convergence due to our cluster model at an angular separation from the cluster centre \(\theta\) as

$$\kappa_\text{cl}(\theta) = k_0 \kappa_1 (\theta; \theta_s).$$

(8)

Here, \(k_0\) is a normalisation such that \(k_0 \theta_s^2\) is proportional to \(M_{500}\) and \(\kappa_1 (\theta; \theta_s)\) is a circularly-symmetric template function that depends only on \(\theta/\theta_s\), where \(\theta_s\) is the angular size of the scale radius \(r_s\), \(\theta_s = r_s/\theta_c(z)\), with \(\theta_c(z)\) the angular diameter distance to the cluster. The template is normalised to unity at \(\theta = \theta_s\). Given an estimate of the convergence around a cluster, \(\hat{\kappa}(L)\), and a fiducial value of the cluster angular size, \(\theta^{\text{fid}}\), an estimator for \(k_0\) can be written as

$$\hat{k}_0 = \frac{1}{\sigma_{\kappa_1}^2} \left( \frac{\int \hat{\kappa}(L) \kappa_1^2(L)}{2 \pi N_e(L)} \right)^{-1} \int \hat{\kappa}(L) \kappa_1^2(L) \frac{d\theta^2}{2 \pi N_e(L)}.$$  

(9)

where \(N_e(L)\) is the variance of the reconstruction noise of \(\hat{\kappa}(L)\), and where the dependence of \(\kappa_1\) on \(\theta^{\text{fid}}\) is left implicit. This inverse-variance weighting ensures that the estimator is minimum-variance, which can be written as

$$\sigma_{\kappa_1}^2 \equiv \left( \frac{\kappa_0 - \kappa_0^2}{\sigma^2} \right)^2 = \frac{1}{\sqrt{2 \pi}} \left( \frac{\int \hat{\kappa}(L) \kappa_1^2(L)}{2 \pi N_e(L)} \right)^{-1} \int \hat{\kappa}(L) \kappa_1^2(L) \frac{d\theta^2}{2 \pi N_e(L)}. $$

(10)

We use each cluster’s true mass, \(M_{500}\), as the model fiducial mass, from which we derive its fiducial angular scale \(\theta^{\text{fid}}\):

$$M^{\text{fid}} = 500 \frac{4\pi}{3} \left( \frac{\left[\kappa_1(\theta_s)\right]^2}{c_{500}^2}\right)^2 \rho_c(z).$$

(11)

As discussed in Zubeldia & Challinor (2019), from this estimator \(\hat{k}_0\) a mass estimator \(M_{500}\) can be obtained in a straightforward way, but the signal-to-noise on \(\hat{k}_0\) (or, equivalently, on \(M_{500}\)) turns out to have better properties as an observable. This is defined as

$$p_{\text{obs}} \equiv \hat{k}_0/\sigma_{\kappa_1}.$$  

(12)

As can be seen from Eq. (9), for a given cluster \(p_{\text{obs}}\) is an unbiased estimator of the cluster CMB lensing signal-to-noise only if the convergence model, \(\kappa_1\), is equal to the true cluster convergence, \(\kappa_c\). As shown in Zubeldia & Challinor (2019), \(p_{\text{obs}}\) is more immune to mismatch between \(\kappa_1\) and \(\kappa_c\). In particular, it is much less dependent on the choice of \(\theta^{\text{fid}}\). We therefore use \(p_{\text{obs}}\) as our cluster observable; \(p_{\text{obs}}\) is also the observable used in Zubeldia & Challinor (2019).

As mentioned above, the statistics of the lensing reconstruction noise are well understood, and due to this we do not produce reconstructed convergence maps of each cluster convergence map, but instead we use each cluster’s true convergence, \(\kappa_c\) in place of \(\hat{k}_c\) in the matched-filtering process. That is, what we compute is the mean CMB lensing signal-to-noise, \(p \equiv \langle p_{\text{obs}} \rangle\), where angular brackets denote averaging over reconstruction noise, \(n(L)\), for each cluster in our catalogue and for all our seven integration lengths:

$$p = \sqrt{2 \pi} \left[ \frac{\int d^2L |\kappa(L)|^2}{2 \pi N_e(L)} \right]^{-1/2} \int d^2L \kappa_c(L)\kappa_c^*(L).$$

(13)

In part of our analysis – specifically, in Section 5.2 – we also treat as noise the contribution to the cluster convergence maps \(\kappa_c\) coming from large scale structure (LSS) uncorrelated with the cluster. In this case, the variance of the reconstructed convergence, after averaging over reconstruction noise and this uncorrelated LSS, is therefore \(\sigma_{\text{LSS}}^2\), where \(C_{\text{LSS}}^2\) is the convergence power spectrum. We also make measurements of \(p\) for all our cluster convergence fields using this variance in the matched filter.

In summary, we obtain two sets of mock measurements of our cluster CMB lensing mass observable, \(p\), for all our clusters, which are at two different redshifts, and for the seven integration lengths \(\chi_L\) considered. Each set has a different understanding of what is thought of as noise in the matched-filtering process. In the first set, noise is understood as being solely due to reconstruction noise; this data set, which we refer to as cluster data set \(B\), is analysed in Section 5.3. In the second case, the variation due to uncorrelated LSS is also included in the noise budget; this data set, which we refer to as cluster data set \(A\), is studied in Section 5.2.

We also apply our measurement pipeline to our random convergence maps. The idea is to obtain the response of our measurement method to an observation that is not centred on a cluster, but where only the convergence due to random LSS is present. The use of these observations will become clear in Section 5. For each random convergence map, we obtain a set of measurements following our measurement pipeline, using as matched-filter fiducial mass \(M^{\text{fid}}\) the centres of 90 equally-spaced subdivisions of the interval \((2-20) \times 10^{14} M_\odot\), and as redshift the corresponding snapshot redshift. Our \(10^5\) random convergence fields at each snapshot redshift yield \(10^5\) mock measurements for each mass bin, redshift, and integration length. They constitute two sets of measurements of \(p\) as a function of filter fiducial mass and integration length, which we refer to as random data set \(A\) (in which the LSS power spectrum is included in the matched filter inverse-variance weighting) and \(B\) (in which only the reconstruction noise power spectrum is included in the matched filter inverse-variance weighting).
5 STATISTICS OF THE CMB LENSING CLUSTER MASS OBSERVABLE

In this section we analyse our mock observations, which we obtain as detailed in Section 4, in order to characterise the statistics of our CMB lensing cluster mass observable. First, in Section 5.1, we describe how we understand our observations in terms of random variables. We then study our mock observations following the deconvolution and extrapolation approaches in Sections 5.2 and 5.3, respectively.

5.1 Model of observations

Let us consider our noisy CMB lensing cluster mass observable, $p_{\text{obs}}$, with either of our two choices of matched-filter inverse-variance weighting, and for a given integration length $\chi_L$. Let us also consider a point in true cluster parameter space, $M_{500}^0 - z$. We can think of $p_{\text{obs}}$ as a random variable, with variability arising from different sources. First, there is reconstruction noise, which arises from the CMB and instrumental noise fluctuations. In addition, there is a contribution coming from variation in the true lensing convergence. Indeed, clusters with a given true mass $M_{500}^0$ and at a given redshift $z$ yield, in general, different lensing convergences: they have different shapes, being in general triaxial, and different large scale structure correlated and uncorrelated with them, which, in projection along the line of sight, also contributes to the convergence. Since $p_{\text{obs}}$ is linear in the lensing convergence, this variability in the lensing convergence translates directly into variability in $p_{\text{obs}}$.

We can write $p_{\text{obs}}$ at a given point in true cluster parameter space, $M_{500}^0 - z$, as a sum of three random variables (as in Eq. 1, repeated here for convenience),

$$p_{\text{obs}} = p_c + \Delta p_u + \Delta p_n.
$$

Here, $\Delta p_n$ denotes the contribution to $p_{\text{obs}}$ coming from the lensing reconstruction noise. It is the random variable associated with the response of our matched filter at the given true cluster parameters to the reconstruction noise $n$ (see Eq. 4). Next, $\Delta p_u$ denotes the contribution coming from LSS uncorrelated with the cluster (hereafter and for concision, ‘uncorrelated LSS’). It is the random variable associated with the response of our matched filter at the given true cluster parameters to our random convergence maps with the considered integration length $\chi_L$. Finally, $p_c$ denotes the contribution to $p_{\text{obs}}$ arising from the cluster itself (that is, from cluster morphology and orientation) and from LSS along the line of sight correlated with the cluster (hereafter, ‘correlated LSS’). It is defined as the variable that arises from subtracting $\Delta p_u$ from $p_c$ and taking the second moment of Eq. (13). The variance of $p_{\text{obs}}$ over $\Delta p_n$ is expected to be close to Gaussian. We can calculate the variance of $\Delta p_n$ by taking the second moment of Eq. (13) to find

$$\sigma_{\Delta p_n}^2 = \left[ \int \frac{d^2L}{2\pi} |\kappa_0(L)|^2 \right]^{-1} \int \frac{d^2L}{2\pi} C_{\kappa\kappa}^L \frac{C_\kappa^L}{2\pi N_c(L)},
$$

where $C_{\kappa\kappa}^L$ is the lensing convergence power spectrum, and where, recall, $N_c(L) = N_c^{(0)}(L) + C_{\kappa\kappa}^L$. For $z = 0.23$ we find $\sigma_{\Delta p_u}/\sigma_{\Delta p_n} = 0.19$ for $M_{500} = 2 \times 10^{14} M_\odot$, and $\sigma_{\Delta p_u}/\sigma_{\Delta p_n} = 2$.

5.2 Deconvolution approach

5.2.1 Method

In this approach, our matched filter has both the reconstruction noise and the lensing convergence power spectra in the inverse-variance weighting (see Eq. 9). Let us first think of a hypothetical set of simulated observations in which the variation due to lensing by uncorrelated LSS is present from $z = 0$ back to last scattering, and not just within a box of a given length along the line of sight, and where reconstruction noise is also present. In order to study $P(p_{\text{obs}} | M_{500}^0 - z ; \chi_s)$, we divide it into two layers:

The first layer is $P(p_{\text{obs}} | p_c ; M_{500}^0 - z ; \chi_s)$, which is approximately a Gaussian distribution centred on $p_c$ with unit standard deviation. Indeed, first, both $\Delta p_n$ and $\Delta p_u$ have zero expected values. We empirically check that $\Delta p_n$ has an expected value consistent with zero as a function of $M_{500}$ with our random data sets, which can be thought of as realisations of $\Delta p_n$ at a given integration length (see below). In addition, reconstruction noise is approximately Gaussian, and therefore so is $\Delta p_n$, which is linear in it. As we show below, for our experimental specifications, the standard deviation of $\Delta p_n$, $\sigma_{\Delta p_n}$, is significantly larger than the standard deviation of $\Delta p_u$, $\sigma_{\Delta p_u}$, across the mass range considered, and so the variance of $\Delta p_u + \Delta p_n$ is dominated by the reconstruction noise. For modest levels of non-Gaussianity of $\Delta p_u$ – expected given the long integration length $\chi_s$ – the relatively high redshift of the lenses involved, and the effective low-pass filtering due to the inverse-variance weighting in the matched filter – we therefore expect the distribution of $\Delta p_u + \Delta p_n$ to be close to Gaussian. We can calculate the variance of $\Delta p_u$ by taking the second moment of Eq. (13) to find

$$\sigma_{\Delta p_u}^2 = \left[ \int \frac{d^2L}{2\pi} |\kappa_0(L)|^2 \right]^{-1} \int \frac{d^2L}{2\pi} C_{\kappa\kappa}^L \frac{C_\kappa^L}{2\pi N_c(L)},
$$

where $C_{\kappa\kappa}^L$ is the lensing convergence power spectrum, and where, recall, $N_c(L) = N_c^{(0)}(L) + C_{\kappa\kappa}^L$. For $z = 0.23$ we find $\sigma_{\Delta p_u}/\sigma_{\Delta p_n} = 0.19$ for $M_{500} = 2 \times 10^{14} M_\odot$, and $\sigma_{\Delta p_u}/\sigma_{\Delta p_n} = 2$.

Here, $\chi_s$ denotes that the integral along the line of sight is performed from $z = 0$ to $\chi_s$ (i.e., to last scattering); a semicolon is used to stress the fact that, unlike the other two conditioning variables, $\chi_s$ is not a cluster-related variable.
0.22 for $M_{500} = 10^{15} M_\odot$; for $z = 0.52$ we find $\sigma_{\Delta p_u}/\sigma_{\Delta p_n} = 0.15$ for $M_{500} = 2 \times 10^{14} M_\odot$, and $\sigma_{\Delta p_u}/\sigma_{\Delta p_n} = 0.18$ for $M_{500} = 10^{15} M_\odot$. Finally, our matched-filter inverse-variance weighting guarantees that $\Delta p_u + \Delta p_n$ has unit standard deviation and hence so does the distribution $P(p_{\text{obs}}|p_c, M_{500}, z; \chi^\star)$. In the context of this deconvolution approach, we refer to $P(p_{\text{obs}}|p_c, M_{500}, z; \chi^\star)$ as observational scatter.

The other layer in which we decompose the scatter of our observable is $P(p_c|M_{500}, z; \chi^\star)$. Within the context of this approach, we refer to this variability as intrinsic scatter. We study this conditional distribution with our mock observations. First, we assume that $P(p_c|M_{500}, z; \chi^\star = 400 \, \text{Mpc})$ is, to a good approximation, equal to $P(p_c|M_{500}, z; \chi)$; we remind that $\chi^\star = 400 \, \text{Mpc}$ is the largest integration length that we have considered. Indeed, beyond 200 Mpc from the cluster centre, the contribution from correlated LSS to the convergence should be negligible (see Section 5.3 for a quantitative discussion). Thus, $P(p_c|M_{500}, z; \chi^\star = 400 \, \text{Mpc})$ is what needs to be determined. In the following, we will denote this pdf simply with $P(p_c|M_{500}, z)$.

Unfortunately, our mock observations do not provide us with samples from $P(p_c|M_{500}, z)$. However, our cluster data set $A$ can be thought of as consisting of samples from $P(p, M_{500}|z; \chi)$, where $p = p_c + \Delta p_u$, and where $z$ can be either of our two redshifts ($z = 0.23$ and $z = 52$), and $\chi$ either of our seven integration lengths (see Section 3). These observations for $\chi^\star = 400 \, \text{Mpc}$ and $z = 0.23$ are shown, for illustrative purposes, in the upper panel of Fig. 1. In addition, our random data set $A$ can be thought of as consisting of samples from $P(\Delta p_u|\chi^\star = 400 \, \text{Mpc})$ at the considered redshifts $z$, integration lengths $\chi$, and true masses $M_{500}$ (see Section 4). Our route to obtain $P(p_c|\chi^\star = 400 \, \text{Mpc})$ is first to determine $P(p_c|M_{500}, z; \chi)$ as a function of $M_{500}$ as slices in $M_{500}$ of $P(p, M_{500}|z; \chi)$, and then, since $p = p_c + \Delta p_u$, to (formally) deconvolve it with $P(\Delta p_u|M_{500}, z; \chi)$. We describe this procedure, which takes its name from the deconvolution step, in detail in the rest of this section.

We first group our mock measurements of $p$ for $\chi^\star = 400 \, \text{Mpc}$ at $z = 0.23$ into 90 subsets by binning their corresponding values of $M_{500}$ into 90 equally-spaced bins between $2 \times 10^{14} M_\odot$ and $10^{15} M_\odot$. For $z = 0.52$ we apply a similar binning but then combine the 33 bins of the high-mass end into groups of three, yielding a total of 68 bins. We do this in order to compensate for the very small number of clusters per original bin at the high-mass end for $z = 0.52$, which in some cases are below 10. We then assume that the distribution followed by $p$ in each of these bins is $P(p|M_{500}, z; \chi^\star = 400 \, \text{Mpc})$, with $M_{500}$ equal to the central value of the bin. This is certainly true in the limit in which the bins are infinitesimal, and should be a good approximation if they are sufficiently small. For illustrative purposes, an estimate of $P(p|M_{500}, z; \chi^\star = 400 \, \text{Mpc})$ as a function of $M_{500}$ can be seen in the lower panel of Fig. 1 for $z = 0.23$ (the same case as the upper panel). We used a kernel density estimation method, fastKDE (O’Brien et al. 2014, 2016); we note that this illustrative estimate is done with no previous binning.

We treat each of the $M_{500}$ bins independently. Let us consider one of our snapshot redshifts $z$ and one of our $M_{500}$ bins. We can write $P(p|M_{500}, z; \chi^\star = 400 \, \text{Mpc})$ as

$$ P(p|M_{500}, z; \chi^\star = 400 \, \text{Mpc}) = \int_0^\infty P(p|p_c, M_{500}, z; \chi^\star = 400 \, \text{Mpc})P(p_c|M_{500}, z)dp_c, \quad (16) $$

where $P(p|p_c, M_{500}, z; \chi^\star = 400 \, \text{Mpc})$ is equal to $P(\Delta p_u|M_{500}, z; \chi^\star = 400 \, \text{Mpc})$ evaluated at $\Delta p_u = p - p_c$, and where $P(p_c|M_{500}, z)$ is the distribution of intrinsic scatter that we want to characterise (note that we have dropped the dependence on $\chi$ following the convention introduced above, as this pdf for $\chi^\star = 400 \, \text{Mpc}$ ought to be very close to that for $\chi^\star = \chi^\star$). That is, as anticipated, the pdf of $p$ can be seen as a convolution of the pdf of $p_c$ with the pdf of $\Delta p_u$. We then assume that $P(p_c|M_{500}, z)$ is approximately log-normal. Log-normality is a common assumption for the intrinsic scatter of cluster observables, even if evidence for, e.g., some skewness has been found in some studies (see, e.g., Becker & Kravtsov 2011). In particular, we model $P(\ln p_c|M_{500}, z)$3 with an Edgeworth series, truncating it.
after the first three expansion terms. That is, we assume that
\[ P(\ln p_c | M_{500}, z) = \frac{1}{\sigma_c} \left( \phi - \frac{\lambda_3 c}{6} \phi(3) + \frac{\lambda_4 c}{24} \phi(4) + \frac{\lambda_5 c}{72} \phi(6) \right) . \]

(17)

Here, \( \phi \) is the standard normal distribution evaluated for
\( \ln p_c - \mu_c / \sigma_c \), and \( \phi(n) \) is the \( n \)-th derivative of the standard normal distribution, also evaluated for \( \ln p_c - \mu_c / \sigma_c \). The quantities \( \mu_c, \sigma_c, A_{3c}, \) and \( A_{4c} \) are, respectively, the mean, standard deviation, skewness, and kurtosis of \( \ln p_c \).

The skewness and kurtosis, in particular, are defined in terms of the cumulants \( k_n \) of \( \ln p_c \) as \( A_{3c} = k_3 / \sigma_c^3 \) and \( A_{4c} = k_4 / \sigma_c^4 \), respectively. As can be seen from Eq. (17), a Gaussian distribution has \( A_{3c} = A_{4c} = 0 \) (and all higher-order cumulants vanish also). We note that \( A_{4c} \) is sometimes referred to as excess kurtosis. We refer to this model of the intrinsic scatter as the log-Edgeworth model of the intrinsic scatter.

Assuming this model, \( P(\ln p_c | M_{500}, z) \) is characterised by four parameters, which we choose to be \( \beta_c, \sigma_c, A_{3c}, \) and \( A_{4c} \). Here, \( \beta_c \) is a lensing mass bias parameter that is used in substitution of \( \mu_c \) and is defined as follows. As in Zubeldia & Challinor (2019), we introduce the model mean signal-to-noise at true cluster parameters \( M_{500} \) and \( z \), \( \bar{\rho}(\beta_c, M_{500}, z) \), as

\[ \bar{\rho}(\beta_c, M_{500}, z) = \frac{\sigma_0(\beta_c, M_{500}, z)}{\sigma_0}, \]

where \( \sigma_0(\beta_c, M_{500}, z) \) is the convergence of the truncated NFW model at the scale radius for a cluster of mass \( \beta_c M_{500} \), \( \sigma_0 \) is the noise of the matched filter in this case, and \( \sigma_{M_{500}}(\beta_c, M_{500}, z) \) is the noise of the \( M_{500} \) matched filter estimator, given by Eq. (26) of Zubeldia & Challinor (2019). Following Zubeldia & Challinor (2019), the lensing mass bias parameter \( \beta_c \) is determined by demanding that \( \ln \bar{\rho}(\beta_c, M_{500}, z) \) is the mean of \( \ln p_c \) for true clusters of mass \( M_{500} \) at redshift \( z \), i.e., \( \ln \bar{\rho}(\beta_c, M_{500}, z) = \mu_c \). As we discuss further below, deviations of \( \beta_c \) from unity arise both from mismatch between the true mean cluster (plus correlated LSS) convergence and the truncated NFW model at the same mass, and from intrinsic scatter. Although the model mean cluster signal-to-noise is not the expected value of any of our random variables, for small \( \sigma_c \) it is approximately equal to the expected value of \( \sigma_c \), in which case the bias is determined by profile mismatch. We note that in Zubeldia & Challinor (2019) \( \beta_c \) is denoted with \( 1 - b_{\text{CMBlens}} \) and is assumed to be independent of mass and redshift.

The mean signal-to-noise \( \bar{\rho} \), as defined in Eq. (18), is shown in Fig. 2 for our reference Planck-like experiment as a function of \( M_{500} \) for the two redshifts and the two matched filter inverse-variance weightings considered, one including the contribution from uncorrelated LSS (dashed lines), and the other one consisting only of reconstruction noise (solid lines). We remind that the former weighting is the one considered in this deconvolution approach; the other choice is also shown for comparison. It can be seen that including LSS in the inverse-variance weighting reduces the mean signal-to-noise at any given mass, as the observations are understood as being noisier. This decrement, however, is small, since reconstruction noise dominates over the variation due to uncorrelated LSS in our reference experiment. We note that \( \bar{\rho} \) in Fig. 2 is computed assuming \( \beta_c = 1 \) in Eq. (18); \( \bar{\rho} \) can be evaluated at any biased mass \( \beta_c M_{500} \) through simple interpolation.

Finally, we determine the remaining factor of the integrand of Eq. (16), \( P(\Delta p_{\text{nl}} | M_{500}, z; \chi_l = 400 \text{Mpc}) \), with our random data set A. Recall, this data set consists of \( 10^5 \) mock measurements of \( \Delta p_{\text{nl}} \) for each mass bin, redshift, and integration length considered. Specifically, for each of our \( M_{500} \) bins we estimate \( P(\Delta p_{\text{nl}} | M_{500}, z; \chi_l = 400 \text{Mpc}) \) with the corresponding \( 10^5 \) mock measurements for \( \chi_l = 400 \text{Mpc} \) in a non-parametric way using the kernel density estimation library fastKDE (O’Brien et al. 2014, 2016).

Assuming our log-Edgeworth model for the intrinsic scatter and with the scatter due to uncorrelated LSS determined with kernel density estimation, Eq. (16) can be thought of as defining the likelihood of our mock measurements of \( p \) for \( \chi_l = 400 \text{Mpc} \) for the four parameters, \( \beta_c, \sigma_c, A_{3c}, \) and \( A_{4c} \). Adopting wide flat priors for all four parameters, we explore the corresponding posterior distributions across our \( M_{500} \) bins and for our two snapshot redshifts with the emcee package, which performs affine-invariant MCMC sampling, generating a total of \( 10^5 \) samples for each bin. In addition, we consider a second case in which we take the intrinsic scatter \( P(\ln p_c | M_{500}, z) \) to be log-normal. This model is the particular case of our log-Edgeworth model in which \( A_{3c} = A_{4c} = 0 \). In this case, the likelihood has only two parameters, \( \beta_c \) and \( \sigma_c \). Similarly adopting wide flat priors on the parameters, we explore the corresponding posteriors with the emcee package, also generating \( 10^5 \) samples for each bin.
Figure 3. Measured (median) values of $\beta_c$ (lensing mass bias), $\sigma_c$ (scatter), $\lambda_3,c$ (skewness), and $\lambda_4,c$ (kurtosis) as a function of $M_{500}$ for $z = 0.23$ (left) and $z = 0.52$ (right) obtained from the cluster and random data sets $A$ following our deconvolution approach (see Section 5.2). The parameter values obtained assuming our log-Edgeworth model are shown in green, whereas those obtained assuming log-normality are shown in red.
Figure 4. Histograms of the values of $\ln p$ from the cluster data set $A$ for $\chi_l = 400\,\text{Mpc}$ for several selected $M_{500}$ bins and at the two considered snapshot redshifts, $z = 0.23$ (left panels), and $z = 0.52$ (right panels). In addition, corresponding predictions from the log-Edgeworth model of the intrinsic scatter for the median parameter values as obtained from the MCMC samples are shown in green, and analogous predictions from the log-normal model of the intrinsic scatter (see Section 5.2.1) are shown in red.
5.2.2 Results and discussion

Figure 3 shows the measured values of $\beta_c$, $\sigma_c$, $\lambda_{c,4}$, and $\lambda_{c,4}$ as a function of $M_{500}$ for our two snapshot redshifts that we obtain with our deconvolution approach. Specifically, for each $M_{500}$ bin the median value of each parameter as obtained from the MCMC samples is shown. The green curves are obtained assuming the log-Edgeworth model of the intrinsic scatter described in Section 5.2.1, whereas the red curves are obtained taking the intrinsic scatter to be log-normal. By construction, $\lambda_{c,4} = \lambda_{c,4} = 0$ for the log-normal case; these constant zero lines are not shown for clarity. The curves are, as expected, noisy, due to the fact that there is only a finite number of clusters in each $M_{500}$ bin. In general, the noise visibly increases with mass, as the number of clusters per bin decreases. For $z = 0.52$, however, the noise decreases at the high-mass end due to the use of wider bins.

In addition, Fig. 4 shows the histograms of $\ln p$ for our simulated clusters (with $\chi^2 = 400\, \text{Mpc}$) in several selected $M_{500}$ bins at both snapshot redshifts. The predictions of our two models, Eq. (16) with log-Edgeworth or log-normal intrinsic scatter, evaluated at the respective median parameter values obtained from our MCMC samples, are also shown as the green and red curves, respectively.

Several remarks can be made about Figs 3 and 4. First, in Fig. 3 it can be seen that the log-Edgeworth model detects significant non-zero (positive) skewness, $\lambda_{c,4}$, and kurtosis, $\lambda_{c,4}$, across most of our mass bins and for both redshifts. At fixed redshift, $\lambda_{c,4}$ is observed to decrease with mass, with no significant detection at the high mass end. On the other hand, $\lambda_{c,4}$ is observed to depend less strongly on mass and to be non-zero at all masses. Neither of these parameters exhibit a strong redshift dependence. The bias, $\beta_c$, and scatter, $\sigma_c$, obtained assuming the log-Edgeworth model take values close to those obtained assuming log-normality. A small systematic difference is observed at lower masses, especially for $z = 0.23$, which we attribute to the detection of non-zero skewness and kurtosis, since these two parameters have some degeneracy with $\beta_c$ and $\sigma_c$.

For both models, $\beta_c$ is observed to depend on both mass and redshift. At fixed redshift, it decreases with mass. At fixed mass, it is observed consistently to take larger values at $z = 0.52$ than at $z = 0.23$, the difference being larger at smaller masses. The scatter, $\sigma_c$, on the other hand, does not have a strong dependence on mass for either model. However, there does seem to be a small dependence with redshift; at a given mass, $\sigma_c$ generally takes larger values at $z = 0.52$.

We recall that $\sigma_c$ is the standard deviation of $\ln p_c$, the logarithm of the contribution to the signal-to-noise associated with the cluster itself and to correlated LSS. It is therefore a measure of how large the fractional scatter on $p_c$, $\sigma_p_c/p_c$, is, where $\sigma_p_c$ is the standard deviation of $p_c$. Thus, we find that the standard deviation of $p_c$ due to the cluster itself and to correlated LSS increases with mass at roughly the same rate as $p_c$ does.

The bias, $\beta_c$, on the other hand, has a less straightforward interpretation. It depends both on the mismatch between the convergence model used in the matched filter (in our case, a truncated NFW profile; see Section 4) and the true mean convergence at the given $M_{500}$ and $z$, and on the scatter (and higher moments) at that point in true cluster parameter space. This is a consequence of $\beta_c$ being defined through the expected value of $\ln p_c$, $\mu_c$ (and not of, e.g., $p_c$). We remind that $\mu_c = \ln \hat{p}(\beta_c M_{500}, z)$, where $\hat{p}(\beta_c M_{500}, z)$ is the model mean signal-to-noise defined in Eq. (18). In order to understand this better, let us assume that the intrinsic scatter is log-normal. In this case,

$$
\ln \hat{p}(\beta_c) = \ln \langle p_c \rangle - \frac{\sigma_c^2}{2},
$$

where $\langle p_c \rangle$ is the expected value of $p_c$, which only depends on the true mean convergence at the point in cluster parameter space considered, and which is equal to the true expected value of $p$, since uncorrelated large scale structure has zero expected value. Note that to simplify the notation, we have only made explicit the dependence of $\hat{p}$ on $\beta_c$ in Eq. (19). Thus, in general, at fixed $M_{500}$ and $z$, $\beta_c$ depends on both the true mean cluster convergence (through $\langle p_c \rangle$) and the scatter, $\sigma_c$. However, in our case $\sigma_c \approx 0.15$, so $\sigma_c^2/2 \approx 0.01$, and $\langle p_c \rangle$ is typically 0.2, so $\ln \langle p_c \rangle \sim -1.6$. Therefore, $\ln \hat{p}(\beta_c) \approx \ln \langle p_c \rangle$, i.e., the bias, $\beta_c$, essentially corrects for the mismatch between the assumed convergence model used in the matched filter and the true mean convergence. This is also true if we approximate the intrinsic scatter with our Edgeworth series; indeed, in this case we find

$$
\ln \hat{p}(\beta_c) = \ln \langle p_c \rangle - \frac{\sigma_c^2}{2} - \ln \left(1 + \frac{\lambda_{c,4}}{6} \sigma_c^3 + \frac{\lambda_{c,4}}{24} \sigma_c^4 + \frac{\lambda_{c,4}}{72} \sigma_c^6\right).
$$

For $\sigma = 0.15$, $\lambda_{c,4} = 1$, and $\lambda_{c,4} \approx 2$, the new (logarithm) term is $O(10^{-4})$, and therefore also negligible. This also means that, if the scatter is small, the skewness and kurtosis also have little impact on the bias. We will see in Section 5.2.4 that using a model that better fits the true mean convergence yields $\beta_c \approx 1$. In addition, in Section 5.3, where we consider the extrapolation approach, we will meet an example of the scatter not being negligible and the bias being sensitive to it as a result.

Finally, in Fig. 4 it can be seen that both models fit our mock data rather well across the mass bins and for both redshifts. Visually, the log-Edgeworth model (green curves) fits the mock data better than the log-normal model (red curves), as expected since the latter is a special case of the former and since statistically-significant skewness and kurtosis are detected throughout the mass bins. This is most apparent in the histograms corresponding to lower masses, which are less noisy as they contain a larger number of clusters (see, e.g., the upper-left panel). We note that rigorous analysis of the goodness of fit of each model and model comparison between the two models, which could be done by, e.g., comparing their Bayesian evidences as a function of mass and redshift, are beyond the scope of this paper.

5.2.3 Sufficiency of log-normal approximation

In the previous section we presented evidence for the intrinsic scatter not being log-normal (see Fig. 3). A question that arises as a consequence is whether assuming log-normal intrinsic scatter is a good enough approximation. Here we show that for Planck clusters log-normal intrinsic scatter is

\footnote{http://dfno.io/encee/current/}
Indeed a good approximation that yields unbiased estimates of $M_{500}$. We show that assuming log-normal intrinsic scatter is adequate as follows. First, we fit a quadratic polynomial to our median values of $\beta_c$ and $\sigma_c$ obtained assuming log-normal intrinsic scatter (red curves in Fig. 3). We do this by minimising the mean square error, taking the noise associated to each data point (i.e., to each $M_{500}$ bin) to be inversely proportional to the square root of the number of halos within that bin. This produces smooth estimates of $\beta_c$ and $\sigma_c$ as a function of $M_{500}$ which, for each $M_{500}$ bin, are much less sensitive to the specific random fluctuation in that bin than the original data point is. Then, for each $M_{500}$ bin we use the likelihood that we used to determine $\beta_c$ and $\sigma_c$, defined in Eq. (16), assuming log-normal intrinsic scatter. This function, fed with our smooth estimates of $\beta_c$ and $\sigma_c$ (which we evaluate at the central mass of the bin) and with the corresponding mock measurements of $p$, can be seen as a posterior for $M_{500}$ (after taking the prior on $M_{500}$ to be flat and wide). Since it is a one-dimensional posterior, we explore it by evaluating it for a range of values of $M_{500}$ and we compute the expected value, which we denote with $\langle M_{500} \rangle$. Figure 5 shows this mean estimated mass over the true mass $M_{500}$, $\langle M_{500} \rangle/M_{500}$, as a function $M_{500}$ for our two snapshot redshifts, along with the corresponding standard deviation, $\sigma_{M_{500}}/M_{500}$ (depicted with the dotted lines). No evidence for a bias is seen.

5.2.4 Impact of the choice of cluster convergence model

As argued in Section 5.2.2, at a given point in true cluster parameter $(M_{500}, z)$ space, the mismatch between the lensing convergence model used in the matched filter and the true mean lensing convergence profile is the main contribution to the bias, $\beta_c$. In particular, if the model matches the true mean lensing profile, then $\ln p = \ln (p_c)$ and therefore $\beta_c = 1$. In principle, any reasonable model can be used, and any mismatch between it and the true mean convergence profile will make $\beta_c \neq 1$; its precise value can be determined as we have described in Sections 5.2.1 and 5.2.2. However, in a matched filter the signal-to-noise is optimal when the model matches exactly the true signal to be filtered; therefore, using a different template may yield suboptimal results. In this section, we validate numerically that $\beta_c \approx 1$ when the model fits the true mean convergence well enough by studying a new set of mock match-filtered observations produced with a more realistic convergence model and we investigate this potential issue of suboptimality.

The model that we use in the matched filter in order to produce the mock measurements studied in this paper is, as described in Section 4, an NFW profile with fixed concentration $c_{500} = 3$ (critical) and truncated at $R_{500} = 5 h^{-1} Mpc$. This was also the model used in Zubeldia & Challinor (2019). Here we consider a more realistic cluster convergence model, which we write as

$$\kappa_{\text{m}}(M_{500}, z; b, \chi_{\text{trunc}}) = \kappa_{1h}(M_{500}, z; b, \chi_{\text{trunc}}) + \kappa_{2h}(M_{500}, z). \quad (21)$$

Here, $\kappa_{1h}(M_{500}, z; b, \chi_{\text{trunc}})$ is the convergence of a truncated NFW profile with a concentration $c_{500} = 2.6$, which is closer to the values reported by Diemer & Kravtsov (2015) for the mass and redshift range considered here (see the left panel of their Fig. 2), evaluated at a biased mass $b M_{500}$ and truncated at $R_{\text{trunc}} = \chi_{\text{trunc}} b^{1/3} R_{500}$. The second term in the model convergence, $\kappa_{2h}(M_{500}, z)$, is the two-halo term, which accounts for LSS correlated with the halo. We compute it following Oguri & Hamana (2011) and using the Tinker et al. (2010) model of the halo bias. We refer to the model $\kappa_{\text{m}}$ as our fitted truncated NFW+$2h$ model; for comparison purposes, in this section, we refer to our original model as our truncated NFW model. For each $M_{500}$ bin, we determine $b$ and $\chi_{\text{trunc}}$ by fitting $\kappa_{\text{m}}(M_{500}, z; b, \chi_{\text{trunc}})$ to the corresponding mean convergence as obtained from our cluster convergence fields. More specifically, for each $M_{500}$ bin we obtain a mean convergence map by averaging over the convergence maps of all the clusters within it, and then we azimuthally average this mean convergence map over 40 annuli linearly spaced between the halo centre and an angular radius of 90 arcmin (recall that the simulated convergence fields have side lengths of 180 arcmin). We then fit for $b$ and $\chi_{\text{trunc}}$ by minimising the mean square error between this binned convergence and the prediction from our model, $\kappa_{\text{m}}(M_{500}, z; b, \chi_{\text{trunc}})$, taking the variance of each annulus to be inversely proportional to its area. The azimuthally-averaged lensing convergence profile as a function of angular separation from the halo centre measured from our cluster convergence fields can be seen in Fig. 6 for three selected $M_{500}$ bins (solid curves) at our two snapshot redshifts, along with the corresponding predictions from our truncated NFW model (dotted curves) and our fitted truncated NFW+$2h$ model (dash-dot curves). Visually, the latter model clearly fits the measured average convergence better.

We then produce a new set of mock observations from our cluster and random convergence maps for $\chi_{\text{m}} = 400$ Mpc in exactly the same way that we produced our cluster and...
Statistics of CMB-lensing-derived galaxy cluster masses

Figure 6. Azimuthally-averaged convergence as a function of angular separation from the cluster centre, $\theta$, for three representative $M_{500}$ bins, each identified by a different colour, for $z = 0.23$ (upper panel) and $z = 0.52$ (lower panel). The solid curves show the empirical convergence as obtained from our convergence maps extracted from the simulation snapshots; the dotted curves show the prediction from our truncated NFW model; and the dash-dotted curves show the prediction from our fitted truncated NFW+2h model. The latter model visibly matches the mean simulation convergence better than the former one.

Figure 7. Ratio between the expected value of $p_c$, $\langle p_c \rangle$, if the truncated NFW convergence model is used in the matched filter and the analogous quantity if the fitted truncated NFW+2h convergence model is used instead, as a function of $M_{500}$ for our two snapshot redshifts. As expected, the fitted truncated NFW+2h model yields a higher mean signal-to-noise, since it fits the signal to be matched by the matched filter better; the difference is larger for $z = 0.52$. Next, we apply our deconvolution approach, as described in Section 5.2.2, to these new mock observations in order to obtain the values of $\beta_c$ and $\sigma_c$ for this convergence model; we only consider log-normal scatter. These values are shown in Fig. 8 as a function of $M_{500}$ for our two snapshot redshifts. As anticipated in Section 5.2.2, $\beta_c = 1$ throughout our mass bins. At high masses, $\sigma_c$ takes similar values to those that we obtain with our truncated NFW model (see Fig. 3), but at low masses it takes visibly higher values, especially for $z = 0.52$. In fact, for this redshift, at low masses $\beta_c$ is visibly smaller than 1, which we attribute to the corresponding larger values of $\sigma_c$ (as detailed in Section 5.2.2, at fixed mean convergence, a larger value of $\sigma_c$ yields a lower value of $\beta_c$). We attribute this to the fact that it is more spatially extended, and therefore more sensitive to variations around the mean convergence due to correlated LSS.

The fact that the mean signal-to-noise is larger for the fitted truncated NFW+2h model than it is for the truncated NFW model, but that the intrinsic scatter is also larger for the fitted truncated NFW+2h model at lower masses, poses...
the question of which model is more optimal, i.e., which one yields, after calibration, tighter constraints on $M_{500}$? We investigate this question by using new mock observations drawn as samples from our hierarchical model. Let us consider one of our two convergence models. First, for each redshift we fit a polynomial to the corresponding measurements of $\beta_c$ and $\sigma_c$ as a function of $M_{500}$ obtained assuming log-normality (the posterior median values; i.e., the values shown in Figs 3 and 8). This yields a smooth estimate of $\beta_c$ and $\sigma_c$ as a function of $M_{500}$ for each redshift, which we take to be the true values of these parameters. This smoothing of $\beta_c$ and $\sigma_c$ is necessary for the subsequent posteriors of $M_{500}$ to be smooth (see below). Then, using these smooth estimates and Eq. (18), we compute $\bar{p}$ at the central value of each of our $M_{500}$ bins. Next, for each bin we draw $10^4$ samples from a Gaussian centred at $\ln \bar{p}$ with standard deviation equal to the smooth estimate of $\sigma_c$ at the bin centre. We exponentiate these samples, which yields a set of samples of $p_c$ for each bin, and we finally add, to each of them, a sample from a Gaussian with zero mean and standard deviation unity. This procedure generates a set of mock observations of $p_{\text{obs}}$ drawn according to our hierarchical model (with log-normal intrinsic scatter). We apply this procedure to both convergence models, using the same random seed for each model so that the random fluctuations are the same.

Next, we use these mock observations of $p_{\text{obs}}$ for each convergence model in order to constrain, for each mass bin, its corresponding value of $M_{500}$. We do this by considering the likelihood associated with the hierarchical model used to generate these mock observations: fixing the values of $\beta_c$ and $\sigma_c$ to their (smooth) input values, this likelihood can be seen as a function of $M_{500}$. Assuming a wide flat prior, it yields a posterior for $M_{500}$ for each mass bin. This is very similar to the way in which we obtained a posterior for $M_{500}$ in Section 5.2.3, where we tested the impact of assuming log-normal intrinsic scatter; the only difference is that the pdf with which we convolve the intrinsic scatter pdf is the unit standard-deviation Gaussian that describes the full observational scatter, and not the pdf associated with the uncorrelated LSS within a given integration length, as is the case in Section 5.2.3. Each posterior is one-dimensional, so we explore them by evaluating them at a range of values of $M_{500}$ and then we compute their standard deviations, which we denote with $\sigma_{\text{NFW}}^{M_{500}}$ for the truncated NFW convergence model and with $\sigma_{\text{NFW+2h}}^{M_{500}}$ for the fitted truncated NFW+2h convergence model and which we use as a metric of constraining power.

We find that at a given point in true cluster parameter ($M_{500}-z$) space, $\sigma_{\text{NFW}}^{M_{500}}$ and $\sigma_{\text{NFW+2h}}^{M_{500}}$ are within a few percent of each other. The exact value of their ratio as a function of $M_{500}$ is quite sensitive to the details of the polynomial fit to $\beta_c$ and $\sigma_c$ (specifically, the polynomial degree and the weighting of each bin), especially at the high-mass end. However, we consistently find some trends. For $z = 0.23$, $\sigma_{\text{NFW+2h}}^{M_{500}}$ is about 1–2 % larger than $\sigma_{\text{NFW}}^{M_{500}}$ at low masses, and about 0.5–1 % smaller at intermediate and high masses. That is, the original truncated NFW model slightly outperforms the fitted truncated NFW+2h model at low masses, but becomes less optimal at higher masses. We argue that this is due to the intrinsic scatter $\sigma_c$ being larger at low masses for the fitted truncated NFW+2h model (see Figs 3 and 8). At higher masses, however, the scatter takes similar values, and therefore the small difference in mean signal-to-noise (see Fig. 7), which favours the fitted truncated NFW+2h model, becomes the main source of difference in constraining power. For $z = 0.52$, the results are more sensitive to the fitting details, since there are fewer clusters, especially at the high-mass end, where we cannot determine any trend conclusively. At low masses, however, we find $\sigma_{\text{NFW+2h}}^{M_{500}}$ to be about 5–10 % larger than $\sigma_{\text{NFW}}^{M_{500}}$, and about 1–3 % at intermediate masses. As at $z = 0.23$, we attribute this to the (now much larger) intrinsic scatter difference between the two models (see Figs 3 and 8).

In summary, we find that, at low masses, the fitted truncated NFW+2h convergence model, despite yielding higher signal-to-noise measurements, performs slightly worse than the truncated NFW convergence model due to it having a larger associated intrinsic scatter. For $z = 0.23$, the situation is reversed at intermediate and high masses, where the intrinsic scatter is similar. These differences, however, are small, at the few percent level.
5.2.5 Priors in Zubeldia & Challinor (2019)

As already noted in Section 5.2.1, the approach to understanding the observations considered in this section underlies the hierarchical model of Zubeldia & Challinor (2019), where $1 - b_{\text{CMBlens}}$, $\sigma_{\text{CMBlens}}$, and the variable $p_t$ correspond to $\beta_c$, $\tau_c$, and $p_k$ here, respectively. We note, however, that $1 - p_{\text{CMBlens}}$ in Zubeldia & Challinor (2019) is also intended to account for the few-percent negative bias due to cluster miscentering. In that work, the intrinsic scatter is assumed to be log-normal, which, as shown in Section 5.2.3, is a sufficiently good approximation, and priors are imposed on $1 - p_{\text{CMBlens}}$ and $\sigma_c$. These are Gaussian priors centred at $1 - p_{\text{CMBlens}} = 0.93$ and $\sigma_{\text{CMBlens}} = 0.2$, both with standard deviation of 0.05.

The cluster sample used in Zubeldia & Challinor (2019) is the Planck MMF3 cosmology sample, which consists of 439 SZ-detected clusters (see Planck 2015 Results XXVII 2016). In it, most clusters are at low redshift (see Fig. 1 of Planck 2015 Results XXVII 2016) it can be seen that most cluster masses take values lying around the centre of our considered mass range at lower redshifts, and values towards our high-mass end at higher redshifts. The latter is particularly true for clusters with redshifts close to our high-redshift snapshot. Therefore, most of the clusters in the sample lie within the region of mass-redshift parameter space where the priors imposed on $1 - b_{\text{CMBlens}}$ and $\sigma_{\text{CMBlens}}$ in Zubeldia & Challinor (2019) are consistent with our measurements of $\beta_c$ (plus a few percent decrement due to miscentering; see Zubeldia & Challinor 2019) and $\tau_c$, respectively.

5.3 Extrapolation approach

5.3.1 Statistics as a function of integration length

In the extrapolation approach, only the reconstruction noise power spectrum is present in the matched-filter inverse-variance weighting (see Eq. 9). As in Section 3.3, let us first think of a hypothetical set of simulated observations with reconstruction noise and with LSS present from $z = 0$ back to last-scattering. Since reconstruction noise is approximately Gaussian, as a consequence of our choice of matched-filter weighting $P(p_{\text{obs}}, p, M_{500}; \chi_{500})$ is approximately a univariate Gaussian centred on $p$. We remind that $p = p_t + \Delta p$, and that $\chi_{500}$ denotes that integration is performed from $z = 0$ back to last-scattering. In order to fully determine the statistics of $p_{\text{obs}}$ at given true cluster parameters, $M_{500}$ and $\chi_{500}$, what remains to be characterised is $P(p|p_{\text{obs}}, M_{500}; \chi_{500})$. In the context of this approach we refer to the variability due to reconstruction noise, $P(p_{\text{obs}}|p, M_{500}; \chi_{500})$, as observational scatter, and to the remaining variability, due to the cluster itself, correlated and uncorrelated LSS, $P(p|M_{500}; \chi_{500})$, as intrinsic scatter. We note that this way of understanding the scatter is different from that made in Section 5.2, in which the variation due to uncorrelated LSS is understood as being part of the observational scatter.

The aim of the rest of this section is to determine $P(p|M_{500}; \chi_{500})$, which we do with our mock observations, namely with our cluster and random data sets $B$. We first study the statistics of $p$ as a function of $M_{500}$, $\chi_{500}$, and integration length, $\chi_{500}$. We can do this with the mock observations of cluster data set $B$, which can be thought of as samples from $P(p, M_{500}|\chi_{500})$, where $\chi_{500}$ can be any of the seven integration lengths that we consider; $P(p|\chi_{500})$ can be obtained as slices in $M_{500}$ of $P(p, M_{500}|\chi_{500})$. In particular, for each of our two snapshot redshifts and each of our seven values of $\chi_{500}$, we divide our data points into 90 subsets, binning $M_{500}$ in 90 equally-spaced bins between $2 \times 10^{14} M_\odot$ and $10^{15} M_\odot$. As in Section 5.2, we think of the value of $p$ of each data point falling within a given bin as a sample of $P(p|M_{500}; \chi_{500})$, $M_{500}$ being the mass at the bin centre.

In order to characterise the distribution $P(p|M_{500}; \chi_{500})$, we first compute its empirical mean, which we denote with $\langle p \rangle$, with the values of $p$ within each mass bin. We also compute the empirical mean, standard deviation, skewness, and kurtosis of $P(\ln p|M_{500}; \chi_{500})$, which we denote with $\mu$, $\sigma$, $A_3$, and $A_4$, respectively, as a function of $M_{500}$ and $\chi_{500}$ for our two snapshots. Analogously to what is done in Section 5.2, $A_3$ and $A_4$ are defined in terms of the cumulants of $P(\ln p|M_{500}; \chi_{500})$, $k_n$, as $A_3 = k_3/\sigma^3$ and $A_4 = k_4/\sigma^4$, respectively. Thus, as in Section 5.2, instead of presenting our results in terms of $\mu$, we introduce a bias parameter, $\beta$, which is defined in an analogous way to $\beta_c$ in Section 5.2, i.e., by imposing that $\ln \beta p(M_{500}; \chi_{500}) = \mu$, where $\beta p(M_{500}; \chi_{500})$ is given by Eq. (18), noting that $\sigma_{\mu|M_{500}}$ is now the noise of the matched filter with only reconstruction noise in the inverse-variance weighting.

Figure 9 shows $(p)$ and $\beta$ as a function of integration length for three selected mass bins for the $z = 0.23$ snapshot. We obtain the error bars with bootstrap resampling from our cluster data points $(M_{500}, p)$, with a $(M_{500}, p)$ pair taken as a single data point so that the correlation structure is not lost. As expected, the errors are larger for higher masses, as there are fewer clusters per bin. In addition, our measurements of $\beta$, $\sigma$, $A_3$, and $A_4$ are shown in Fig. 10 as a function of $M_{500}$ and integration length for our two redshifts. For the sake of clarity, only four integration lengths have been included.

Since $p$ is linear in the lensing convergence (see Eq. 13), $\langle p \rangle$ is linear in the mean lensing convergence within a given integration length at given true cluster parameters and is insensitive to other moments of the convergence. This is illustrated in the upper panel of Fig. 9: $(p)$ responds to the change in the mean convergence as the integration length is increased starting from its lowest value, increasing as more correlated LSS is added, but this growth plateaus at large integration lengths, where effectively only zero-mean uncorrelated LSS is being added.

The bias parameter $\beta$, however, defined in terms of $\ln p$, is not linear in the convergence. In general it feels both the effect of the mean convergence at a given integration length and of the variance, and higher moments, around the mean convergence. In Section 5.2.2 we saw that the analogous parameter defined in the context of the deconvolution approach, $\beta_c$, is effectively only sensitive to the mean convergence, due to the fact that the intrinsic scatter is small; that is, this analogous parameter essentially corrects the mismatch between the true mean convergence and the model used in the matched filter. Following an argument similar to that in Section 5.2.2, and assuming log-normality (as shown
in Eq. 20, the effect of the skewness and the kurtosis is higher in the scatter and therefore negligible.

\[
\ln(\beta M_{500, z}) = \ln \langle p \rangle - \frac{\sigma^2}{2}.
\]  

(22)

At low integration lengths, \(\sigma\) is small, as can be seen in Fig. 10, and, as in the analogous case in Section 5.2.2, negligible. The bias, \(\beta\), is therefore just sensitive to the mean convergence profile through \(\ln \langle p \rangle\), which at low integration lengths changes with integration length, as there is still correlated LSS that is being added. In particular, we can see in Fig. 9 that \(\beta\) initially increases with integration length. As the integration length increases, however, the mean convergence changes progressively less, which can be seen in the plateauing of \(\langle p \rangle\) in Fig. 9, and more uncorrelated LSS is added, with \(\sigma\) increasing as a result (see Fig. 10). Equation (22) shows that at fixed \(\langle p \rangle\), if \(\sigma\) increases \(\beta\) decreases and, as a consequence, so does \(\beta\). This fall with integration length is observed at high integration lengths in Figs 9 and 10, being more pronounced at lower masses.

In Fig. 9 it can also be seen that \(\beta\) and \(\sigma\) decrease with \(M_{500}\) for both redshifts, analogously to \(\beta_c\) and \(\sigma_c\) in Section 5.2. In addition, the skewness, \(\lambda_3\), and kurtosis, \(\lambda_4\), exhibit a similar qualitative behaviour as a function of integration length across mass and for both redshifts: they both take lower values for \(\chi_l = 5\) Mpc, the shortest integration length considered, grow with \(\chi_l\) to reach some maximum value, and then decrease beyond this as effectively only uncorrelated LSS is added. That is, above some value of \(\chi_l\), as integration length increases the intrinsic scatter, due to the cluster itself, correlated and uncorrelated LSS, becomes increasingly log-normal. As in Section 5.2, the skewness and the kurtosis are generally lower for \(z = 0.52\) than for \(z = 0.23\).

5.3.2 Extrapolation

The fact that for large values of \(\chi_l\), at which effectively only uncorrelated LSS is added on increasing \(\chi_l\) further, \(P(p|M_{500}, z; \chi_l)\) becomes increasingly log-normal motivates the following approach to obtain \(P(p|M_{500}, z; \chi_\star)\), the ultimate goal of this section.

For a given integration length \(\chi_l\), \(p = p_c + \Delta p_u\). Since \(p_c\) and \(\Delta p_u\) are independent random variables, the cumulants of \(p\) are equal to the cumulants of \(p_c\) plus the cumulants of \(\Delta p_u\). In particular, the mean of \(p\), \(\langle p \rangle\), is equal to the mean of \(p_c\), since \(\Delta p_u\) has zero mean, and \(\sigma_p^2 = \sigma_{p_c}^2 + \sigma_{\Delta p_u}^2\), where \(\sigma_{p_c}^2\) and \(\sigma_{\Delta p_u}^2\) are the variances of \(p_c\) and \(\Delta p_u\), respectively.

For \(\chi_l = 400\) Mpc we measure \(\langle p \rangle\) and \(\sigma_{p_c}^2\) from our cluster data set \(B\) and \(\sigma_{\Delta p_u}^2\) from our random data set \(B\) across the 90 \(M_{500}\) bins. We then obtain \(\sigma_{\Delta p_u}^2\) by subtraction. Since this integration length effectively contains all LSS correlated with the cluster, our measured values of \(\langle p \rangle\) and \(\sigma_{p_c}^2\) are the same as the values of such cumulants for \(\chi_l = \chi_\star\), which we denote with \(\langle p \rangle\) and \(\sigma_{p_c}^2\), respectively. Thus, we just need to determine \(\sigma_{\Delta p_u}^2\), the variance associated with LSS from \(z = 0\) to last-scattering.

In general, \(\sigma_{\Delta p_u}^2(\chi_l)\) can be written as

\[
\sigma_{\Delta p_u}^2(\chi_l) = \int \frac{d^2L}{2\pi} \frac{|\chi(L)|^2}{N_c(L)} \int \frac{d^2L}{2\pi} \frac{|\chi(L)|^2}{N_c(L)} C_L^{\phi}(\chi_l),
\]  

(23)

where \(C_L^{\phi}(\chi_l)\) is the power spectrum of the lensing convergence within the comoving distance interval specified by \(\chi_l\), and the other variables are the same as in Eq. (13). We compute \(C_L^{\phi}(\chi_l)\) with CAMB\(^5\) using the nonlinear matter power spectrum and the Limber approximation. We find that for \(\chi_l = 400\) Mpc and for \(\chi_l = 192.7\) Mpc, \(\sigma_{\Delta p_u}^2(\chi_l)\) computed this way is within a few percent of the value we measure across our \(M_{500}\) bins and our two redshifts. We also find that, for a given redshift and mass bin, the small fractional disagreement for \(\chi_l = 400\) Mpc is very similar to that for \(\chi_l = 192.7\) Mpc. This motivates us to compute \(\sigma_{\Delta p_u}^2(\chi_\star)\) using Eq. (23) with a simple rescaling in each mass bin and at each redshift by the appropriate factor obtained empirically for \(\chi_l = 400\) Mpc. We add this rescaled value of \(\sigma_{\Delta p_u}^2(\chi_\star)\) to \(\sigma_{p_c}^2(\chi_\star)\) to obtain finally \(\sigma_{\Delta p_u}^2(\chi_\star)\)

\(^5\) www.camb.info
Figure 10. Measured values of $\beta$ (bias), $\sigma$ (scatter), $\lambda_3$ (skewness), and $\lambda_4$ (kurtosis) as a function of $M_{500}$ and integration length, $\chi_I$, for $z = 0.23$ (left) and $z = 0.52$ (right), as obtained from cluster data set $B$ (see Section 5.3).
We then assume that $P(p|M_{500}; \chi*)$ is log-normal. Denoting the mean and standard deviation of $P(\ln p|M_{500}; \chi*)$ with $\mu_e$ and $\sigma_e$, respectively, we compute them by imposing that the mean and variance of $P(p|M_{500}; \chi*)$ have the values $(p)(\chi*)$ and $\sigma_p^2(\chi*)$ that we determine as described above. Following Sections 5.2 and 5.3.1, instead of presenting our results in terms of $\mu_e$ we introduce a bias parameter, $\beta_e$, defined such that $\ln p(\beta_e M_{500}; z) = \mu_e$, with $\beta$ given by Eq. (18) (where $\sigma_{M_{500}}$ is now computed with only the reconstruction noise power spectrum in the inverse-variance matched-filter weighting).

The bias, $\beta_e$, and scatter, $\sigma_e$, obtained this way are shown in Fig. 11. As expected, the addition of a large amount of uncorrelated LSS significantly decreases $\beta_e$ compared to our previous measurements of $\beta$ for smaller integration lengths (see Fig. 10 and Eq. 22). This is true across our $M_{500}$ range and for both redshifts, the effect being more significant at lower masses. Similarly, the scatter has increased substantially with respect to the values observed for smaller integration lengths (see also Fig. 10), being dominated by uncorrelated LSS.6

We remark that this extrapolation approach relies on the assumption of log-normality of $P(p|M_{500}; \chi*)$, an assumption that, although motivated by our mock observations (see, e.g., Fig. 10), cannot be checked empirically. A more rigorous approach to study the full line-of-sight intrinsic scatter would involve considering full lightcone simulations of the lensing convergence from $z = 0$ back to last-scattering (see, e.g., Giocoli et al. 2016; Takahashi et al. 2017), something that is beyond the scope of this paper.

In summary, this extrapolation approach treats the total scatter due to the cluster and to correlated and uncorrelated LSS as log-normal, and reconstruction noise as Gaussian. In contrast, the deconvolution approach (Section 5.2) treats the scatter due to the cluster and to correlated LSS as log-normal (or close to log-normal), and the scatter due to uncorrelated LSS and reconstruction noise as Gaussian. For an experiment like Planck, both approaches are essentially equivalent. Indeed, as shown in Section 5.2, the scatter due to the cluster and to correlated LSS is approximately log-normal, at least for mid to high-mass clusters, and the scatter due to uncorrelated LSS and reconstruction noise is roughly Gaussian, since reconstruction noise is almost Gaussian and dominates over uncorrelated LSS (which itself is approximately Gaussian on the scales relevant for cluster mass estimation with Planck). In addition, as argued in this section, the total scatter due to the cluster and to correlated and uncorrelated LSS is approximately log-normal for a Planck-like experiment.

This equivalence may no longer be valid for future experiments, which will have lower reconstruction noise, i.e., higher signal-to-noise per cluster, and probe the convergence on smaller angular scales. If reconstruction noise becomes comparable with the scatter due to uncorrelated LSS, approximating the summed scatter as Gaussian may no longer be accurate. The extrapolation approach, therefore, may be preferable in such cases. We note that the extrapolation approach does not include the LSS convergence power spectrum in the inverse-variance weighting of the matched filter, and, as a consequence, does not optimise for it, as opposed to the deconvolution approach. There is therefore some signal-to-noise loss in the extrapolation approach with respect to the deconvolution approach. For Planck, which is reconstruction-noise dominated, this loss is small, but it may be more significant in lower reconstruction noise scenarios.

6 DIFFERENT OBSERVATIONAL SPECIFICATIONS

We briefly consider how the results obtained with our extrapolation approach vary if we assume a different experimental setup. In particular, we consider an idealised experiment analogous to the Planck-like experiment considered throughout this paper but with a FWHM = 1.4 arcmin and with temperature noise levels of 7 μK arcmin; this is plate. Indeed, to investigate this, we have obtained $\sigma_e$ as a function of $M_{500}$ for other choices of convergence model (e.g., our fitted truncated NFW+2h model), and found that, in general, there is a dependency on redshift.

---

6 We note that $\sigma_e$ takes similar values for the two snapshot redshifts across the mass range considered. This seems to be a coincidence specific to our choice of matched filter convergence tem-
intended to resemble AdvACT. We consider two different lensing quadratic estimators, the $TT$ quadratic estimator, which is the one used throughout this paper for our Planck-like experiment, and the $EB$ quadratic estimator. For both quadratic estimators we use a maximum multipole of $l = 2000$ in the gradient leg and of $l = 3000$ in the field leg.

The $EB$ quadratic estimator differs from the $TT$ quadratic estimator in that it takes as input one map of the (lensed) $E$-mode and one map of the (lensed) $B$-mode CMB polarisation in order to estimate the lensing convergence, instead of two copies of the (lensed) temperature anisotropies. (See Hu & Okamoto 2002 for the detailed construction of the $EB$ estimator.) The $E$-modes and $B$-modes are a useful basis-independent description of the (linear) polarisation of the CMB (Seljak & Zaldarriaga 1997; Kamionkowski et al. 1997). Lensing reconstruction with polarisation is significantly noisier than reconstruction with temperature for an experiment like Planck (Planck 2018 results VIII 2018). Indeed, for this experiment, for a cluster with $M_{500} = 5 \times 10^{14} M_\odot$ at redshift $z = 0.23$ and with the convergence profile of our matched-filter template (Eq. 8), we find, using Eq. (18) with $\beta_e = 1$, a mean signal-to-noise of $(p) = 0.21$ if the $TT$ quadratic estimator is used. If the $EB$ quadratic estimator is used instead, we find $(p) = 0.03$. The situation is, however, different for our AdvACT-like experiment, for which, for the same cluster, we find $(p) = 0.42$ if the $TT$ estimator is used, and $(p) = 0.41$ if the $EB$ estimator is used. The $EB$ estimator therefore provides a signal-to-noise similar to that of the $TT$ estimator for this experimental set-up. It also has the additional interest that it is less sensitive to extragalactic foregrounds than the $TT$ estimator. In the cluster context, the $TT$ estimator in particular suffers from contamination due to the thermal and kinetic $SZ$ effects from the cluster itself, the latter being difficult to remove given that it has the same frequency dependence as the CMB anisotropies, while the polarisation signal from such effects is a negligible contaminant (Raghu et al. 2017). We note that the other three quadratic estimators, the $TE$, the $TB$, and the $EE$ estimators, yield, in this scenario, smaller mean signal-to-noise ratios, $(p) = 0.22$, $(p) = 0.14$, and $(p) = 0.24$, respectively. We therefore do not consider them here.

The extrapolated bias, $\beta_e$, and scatter, $\sigma_e$, that we obtain using the $TT$ and $EB$ quadratic estimators for our AdvACT-like set-up are shown in Fig. 12. It can be observed that the values of $\beta_e$ and $\sigma_e$ as a function of $M_{500}$ and $z$ depend on the choice of quadratic estimator and, also, on the experiment specifications (compare with Fig. 11, where analogous results are shown for the Planck-like experiment). This implies that, although the calibration approaches presented in this paper may be applied to other lensing estimation techniques and experimental set-ups, their numerical results require case-by-case consideration.

7 CONCLUSION

We have studied the statistics of a CMB lensing galaxy cluster mass observable, $p$, for a Planck-like experiment with mock observations obtained from an $N$-body simulation, characterising the biased mean signal and the scatter, and deviations from log-normality, due to the variation associated with the cluster and with correlated and uncorrelated LSS. This characterisation is essential for a cosmological analysis that may make use of this mass observable (e.g., Zubeldia & Challinor 2019) to deliver unbiased results.

We have followed two alternative routes in order to quantify the statistics of this mass observable. First, in our deconvolution approach (Section 5.2), we have treated the variation due to uncorrelated LSS as noise in the matched-filtering process, where it adds to the reconstruction noise, and then characterised the mean signal and the variation due to the cluster itself and to correlated LSS (what we call intrinsic scatter) with our mock observations. We find this intrinsic scatter to be roughly log-normal, although significant skewness and kurtosis are detected, as in similar studies of galaxy weak lensing cluster mass measurements (e.g., Becker & Kravtsov 2011, where significant positive skewness is found). This approximate log-normality and our measured values of the bias, $\beta_e$, and the scatter, $\sigma_e$, serve as justification for the priors on the analogous parameters used in Zubeldia & Challinor (2019). Second, in our extrapolation approach (Section 5.3) we have considered the variation due to the cluster itself and to both correlated and uncorrelated LSS as intrinsic scatter. We find with our mock observations that this scatter becomes increasingly log-normal as we integrate along longer paths, which motivates us to ex-
extrapolate our results in order to incorporate the scatter due to uncorrelated LSS from \( z = 0 \) back to CMB last-scattering assuming log-normality.

We have also considered, for illustration, how our extrapolation results change if a different experimental set-up is assumed (an AdvACT-like experiment; see Section 6). While the qualitative trends with mass and redshift are similar, numerical results for \( \beta_k \) and \( \sigma_e \) differ significantly from those obtained for the Planck-like experiment. This implies that ongoing and future experiments that may want to use our CMB lensing mass observable for, e.g., mass calibration in a cluster counts analysis, will have to quantify its statistics for the particular case of their experiment specifications. Our numerical results, for both a Planck-like and an AdvACT-like experiment, are not transferable to other experiments. Different cluster mass observables (e.g., the observable proposed in Raghu

\textit{Acknowledgements}

The authors would like to thank Debora Sijacki, William Handley, and, especially, Ewald Puchwein for useful discussions. IZ is supported by the Isaac Newton Studentship from the University of Cambridge. AC acknowledges support from the UK Science and Technology Facilities Council (grant numbers ST/N000927/1 and ST/S000623/1). Part of the computational work of this paper was done using the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (http://www.csd3.cam.ac.uk/), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).

\textbf{REFERENCES}

Abazajian K. N., et al., 2016, preprint, (arXiv:1610.02743)

Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046

Bahé Y. M., McCarthy I. G., King L. J., 2012, MNRAS, 421, 1073

Barnes D. J., Kay S. T., Henson M. A., McCarthy I. G., Schaye J., Jenkins A., 2016, Monthly Notices of the Royal Astronomical Society, 465, 213

Baxter E. J., et al., 2015, ApJ, 806, 247

Baxter E. J., et al., 2017, J. Cosmology Astropart. Phys.,

Becker M. R., Kravtsov A. V., 2011, ApJ, 740, 247

Boquet S., et al., 2019, ApJ, 878, 55

Costanzi M., et al., 2019, Monthly Notices of the Royal Astronomical Society, 486, 637

Diemer B., Kravtsov A. V., 2015, ApJ, 799, 108

Geach J. E., Peacock J. A., 2017, Nature Astronomy, 1, 795

Giocoli C., et al., 2016, MNRAS, 461, 209

Hanson D., Challinor A., Efstathiou G., Bielewicz P., 2011, Phys. Rev. D, 83, 043005

Hasselfield M., et al., 2013, J. Cosmology Astropart. Phys., 7, 008

Henden N. A., Puchwein E., Sijacki D., 2019, Monthly Notices of the Royal Astronomical Society, p. 2230

Hoekstra H., Herbonnet R., Muzzin A., Babul A., Madau A., Viola M., Cacciato M., 2015, MNRAS, 449, 685

Horowitz B., Ferraro S., Sherwin B. D., 2019, MNRAS, 485, 3919

Hu W., 2001, ApJ, 557, L79

Hu W., Okamoto T., 2002, ApJ, 574, 566

Hu W., DeDeo S., Vale C., 2007, New Journal of Physics, 9, 441

Kamionkowski M., Kosowsky A., Stebbins A., 1997, Phys. Rev. Lett., 78, 2058

Kirby M., Rozo E., Morris R. G., Allen S. W., Costanzi M., Mantz A. B., Rykoff E. S., von der Linden A., 2019

Klypin A., Holtzman J., 1997, arXiv Astrophysics e-prints,

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Lewis A., Challinor A., 2006, Phys. Rep., 429, 1

Lewis A., Challinor A., Hanson D., 2011, J. Cosmology Astropart. Phys., 3, 018

Louis T., Alonso D., 2017, Phys. Rev. D, 95, 043517

Maddhavacheril M., et al., 2015, Phys. Rev. Lett., 114, 151302

Mantz A., Allen S. W., Rapetti D., Ebeling H., 2010, MNRAS, 406, 1759

Melin J.-B., Bartlett J. G., 2015, A&A, 578, A21

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

O’Brien T. A., Collins W. D., Rauscher S. A., Ringler T. D., 2014, Computational Statistics & Data Analysis, 79, 222

O’Brien T. A., Kashlani K., Cavanaugh N. R., Collins W. D., O’Brien J. P., 2016, Computational Statistics & Data Analysis, 101, 148

Oguri M., Hamana T., 2011, MNRAS, 414, 1851

Planck 2015 Results XXVII 2016, A&A, 594, A27

MNRAS 000, 1–21 (2019)
Statistics of CMB-lensing-derived galaxy cluster masses

Planck 2018 results VIII 2018, arXiv e-prints, p. arXiv:1807.06210
Planck S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 467, 3827
Pratt G. W., Croston J. H., Arnaud M., Boehringer H., 2009, A&A, 498, 361
Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., 2019, Space Sci. Rev., 215, 25
Raghunathan S., et al., 2017, J. Cosmology Astropart. Phys., 8, 030
Raghunathan S., et al., 2019a, arXiv e-prints,
Raghunathan S., et al., 2019b, ApJ, 872, 170
Rasia E., et al., 2012, New Journal of Physics, 14, 055018
Riebe K., et al., 2013, Astronomische Nachrichten, 334, 691
Seljak U., Zaldarriaga M., 1997, Phys. Rev. Lett., 78, 2054
Seljak U., Zaldarriaga M., 2000, ApJ, 538, 57
Shaw L. D., Holder G. P., Dudley J., 2010, ApJ, 716, 281
Shirasaki M., Nagai D., Lau E. T., 2016, MNRAS, 460, 3913
Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shiroyama K., 2017, ApJ, 850, 24
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Truong N., et al., 2017, Monthly Notices of the Royal Astronomical Society, 474, 4089
Wallace D. L., 1958, Ann. Math. Statist., 29, 635
Yoo J., Zaldarriaga M., 2008, Phys. Rev. D, 78, 083002
Zubeldia I., Challinor A., 2019, MNRAS, 489, 401
de Haan T., et al., 2016, ApJ, 832, 95
von der Linden A., et al., 2014, MNRAS, 439, 2

This paper has been typeset from a TeX/\LaTeX file prepared by the author.