Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **489,** 401–419 (2019)
Advance Access publication 2019 August 5

doi:10.1093/mnras/stz2153

Cosmological constraints from *Planck* galaxy clusters with CMB lensing mass bias calibration

Íñigo Zubeldia[1,2]★ and Anthony Challinor[1,2,3]

[1] *Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK*
[2] *Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3] *DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK*

Accepted 2019 July 22. Received 2019 July 18; in original form 2019 April 16

ABSTRACT

We present a new cosmological analysis of the galaxy clusters in the *Planck* MMF3 cosmology sample with a cosmic microwave background (CMB) lensing calibration of the cluster masses. As demonstrated by *Planck*, galaxy clusters detected via the Sunyaev–Zel'dovich (SZ) effect offer a powerful way to constrain cosmological parameters such as $\Omega_{\rm m}$ and σ_8. Determining the absolute cluster mass scale is, however, difficult, and some recent calibrations have yielded cosmological constraints in apparent tension with constraints in the ΛCDM model derived from the power spectra of the primary CMB anisotropies. In order to calibrate the absolute mass scale of the full *Planck* cluster sample, we remeasure the masses of all 433 clusters through their weak lensing signature in the CMB temperature anisotropies as measured by *Planck*. We perform a joint Bayesian analysis of the cluster counts and masses taking as input the estimated cluster masses, SZ signal-to-noise ratios, and redshifts. Our analysis properly accounts for selection effects in the construction of the cluster sample. We find $\sigma_8(\Omega_{\rm m}/0.33)^{0.25} = 0.765 \pm 0.035$ and $1 - b_{\rm SZ} = 0.71 \pm 0.10$, where the mass bias factor $1 - b_{\rm SZ}$ relates cluster mass to the SZ mass that appears in the X-ray-calibrated cluster scaling relations. We find no evidence for tension with the *Planck* primary CMB constraints on ΛCDM model parameters.

Key words: galaxies: clusters: general – Cosmology – cosmic background radiation – cosmological parameters.

1 INTRODUCTION

Galaxy clusters, the largest gravitationally bound structures in the Universe, are powerful cosmological probes (Allen, Evrard & Mantz 2011). In particular, their abundance as a function of mass and redshift is very sensitive to the mean matter density of the Universe, which can be parametrized by $\Omega_{\rm m}$, and to the amplitude of the matter perturbations, which can be characterized by σ_8, the root mean square of the linear density fluctuations smoothed on a scale of $8\,h^{-1}$ Mpc. In recent years, large Sunyaev–Zel'dovich (SZ) cluster surveys, in which clusters are detected through their thermal-SZ (tSZ) signature (Sunyaev & Zeldovich 1972), have provided useful observations of this abundance from which cosmological constraints have been obtained (e.g. Staniszewski et al. 2009; Mantz et al. 2010; Hasselfield et al. 2013; Bleem et al. 2015; de Haan et al. 2016; Planck 2015 Results XXVII 2016; Bocquet et al. 2018; Salvati, Douspis & Aghanim 2018). SZ surveys are particularly interesting because the change in surface brightness of the cosmic microwave background (CMB) due to the SZ effect does not decrease with cluster redshift, allowing for the detection of high-redshift galaxy clusters given enough resolution. In addition, the observational selection of the cluster sample is typically straightforward to model, simplifying the extraction of cosmological information.

A crucial element in analysing the abundance of galaxy clusters ('cluster-count' analysis) is the observational determination of cluster masses. In SZ surveys, the mass of a galaxy cluster is typically estimated directly from CMB data through its mass-dependent SZ signal: the SZ signal is said to be a proxy of the galaxy cluster mass. However, the scaling relations between a cluster's mass and its SZ signal are not very well determined, and usually they need to be calibrated for each survey. This determination of cluster masses currently provides the largest source of uncertainty when obtaining cosmological information from galaxy clusters (Planck 2015 Results XXIV 2016; Pratt et al. 2019).

The *Planck* experiment detected about 1200 galaxy clusters via their SZ signature (Planck 2015 Results XXVII 2016). A subsample of 439 such clusters, known as the MMF3 (multi-frequency matched filter 3) cosmology sample, was used in a count analysis in order to constrain, amongst other parameters and models, $\Omega_{\rm m}$ and σ_8 within the context of a standard spatially flat Lambda cold dark

★ E-mail: iz221@cam.ac.uk

© 2019 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

matter (ΛCDM) cosmology (Planck 2015 Results XXIV 2016). This analysis proceeded as follows. Each cluster was characterized by two observables: SZ signal-to-noise ratio, q, and redshift, z. The signal-to-noise ratio q, a proxy of the cluster mass, was measured from the *Planck* data for each cluster. On the other hand, the redshifts of most clusters in the sample (433) were measured by follow-up observations. These two cluster observables were then binned on a grid in the q–z plane. The number of clusters within each cell was modelled as being independently Poisson distributed in a likelihood analysis, with the mean number of clusters being the theoretically predicted quantities dependent on cosmology.

Theory predicts the number of clusters as a function of their redshift and true mass [we use M_{500}, the mass within a radius R_{500}, inside which the mean matter density is 500 times that of the critical density at the cluster's redshift, $\rho_c(z)$]. To connect these predictions to the expected counts in the q–z plane, the analysis of Planck 2015 Results XXIV (2016) used the SZ–mass scaling relations from Arnaud et al. (2010), with parameters calibrated with X-ray mass estimates of a subsample of clusters of the MMF3 cosmology sample (see Planck 2013 Results XXIX 2014 and Planck 2015 Results XXVII 2016 for the calibration details). However, it is known that the X-ray mass estimates are typically biased low. To account for this, a mass bias factor was introduced, such that in the X-ray-calibrated scaling relations used in the likelihood, the X-ray-derived masses, were replaced by the scaled true masses, $(1-b)M_{500}$, where $1-b$ parametrizes the mass bias. There are several possible sources of this bias: it is known that X-ray cluster mass estimates are typically biased low at a significant level due to their being obtained under the assumption of hydrostatic equilibrium within the cluster, an assumption that can be violated in several scenarios (Nagai, Kravtsov & Vikhlinin 2007; Piffaretti & Valdarnini 2008; Meneghetti et al. 2010); there can also be observational systematic effects in the X-ray mass estimates and selection effects biasing X-ray-selected samples relative to SZ-selected samples (Planck 2015 Results XXIV 2016).

To obtain cosmological constraints, it is necessary to calibrate the mass bias parameter $1-b$ since this determines the overall cluster mass scale. For example, an increase in σ_8, which increases the number of clusters above a given true mass, can be offset by a reduction in $1-b$, which makes the true masses of the observed clusters larger, preserving the expected number of observed clusters. In the analysis of Planck 2015 Results XXIV (2016), the mass bias was estimated from three independent calibrations. In each of these calibrations, lensing mass estimates of the clusters of a subsample of the MMF3 cosmology sample were directly compared to the corresponding masses that are obtained from the X-ray-calibrated SZ–mass scaling relations if no mass bias is assumed. Lensing probes the total cluster mass distribution without relying on any of the assumptions that underlie X-ray mass estimates, and so is arguably the most robust way to measure cluster masses to date. Two of the calibrations used in Planck 2015 Results XXIV (2016) were from galaxy weak lensing mass estimates of a small number of clusters in the MMF3 cosmology sample: $1-b = 0.688 \pm 0.072$ from the Weighing the Giants (WtGs; von der Linden et al. 2014) programme; and $1-b = 0.78 \pm 0.07$ from the Canadian Cluster Comparison Project (CCCP; Hoekstra et al. 2015). Both galaxy weak lensing calibrations support a significant mass bias. The third calibration used CMB lensing mass estimates of most of the clusters in the MMF3 cosmology sample (433) and reported $1/(1-b) = 0.99 \pm 0.19$ (Planck 2015 Results XXIV 2016), which is consistent with their being no mass bias. These three different calibrations were then used as priors on the mass bias parameter in the cluster-count likelihood analysis. They affect the results significantly, particularly the constraints on σ_8. As shown in fig. 7 of Planck 2015 Results XXIV (2016), the absence of a mass bias, a scenario suggested by the CMB lensing calibration, leads to parameter constraints in the Ω_m–σ_8 plane that are in tension with the corresponding constraints derived from the *Planck* CMB power spectra in the ΛCDM model. On the other hand, the two galaxy weak lensing mass bias calibrations alleviate the tension significantly. Model-dependent constraints on the mass bias can be obtained by combining cluster counts with the CMB power spectra. Within ΛCDM, the combination of the *Planck* temperature and polarization power spectra (the *Planck* 2015 TT,TE,EE+lowP likelihood) and the *Planck* cluster-count likelihood gives a posterior distribution $1-b = 0.58 \pm 0.04$ (Planck 2015 Results XXIV 2016). This confirms that a significant mass bias is required if the ΛCDM cosmology favoured by the primary CMB anisotropies is to be consistent with the observed counts of galaxy clusters.

The mass calibration from joint analysis of the cluster counts and primary CMB power spectra was recently updated in Planck 2018 Results VI (2018) to use the *Planck* 2018 TT,TE,EE+lowE+lensing likelihood (the cluster-count likelihood was unchanged from Planck 2015 Results XXIV 2016). The main relevant change is the improved constraints on large-scale polarization due to the use of the data from the Planck high-frequency instrument rather than the less precise data from the low-frequency instrument used in the 2015 likelihoods. This change favours a lower central value for the optical depth to reionization, τ, with the new constraint also being around twice as precise. Lowering τ reduces the fluctuation amplitude to preserve the amplitude of the CMB spectra on smaller scales and leads to a lower predicted value of σ_8. Consistency with the observed cluster counts then requires a larger value of $1-b$ (i.e. less mass bias), with Planck 2018 Results VI (2018) reporting $1-b = 0.62 \pm 0.03$.

In this work, we revisit the *Planck* cluster-count analysis for the particular case of the CMB lensing mass bias calibration. After measuring again the masses of all the clusters in the *Planck* CMB lensing calibration sample through their CMB lensing signature, we argue that a mass bias calibration like the one presented in Planck 2015 Results XXIV (2016) is biased by several effects.

In order to account for these effects, we present an alternative approach in which the CMB lensing mass estimates of each cluster, along with q and z, are directly incorporated into a likelihood that is able to constrain jointly Ω_m, σ_8, and $1-b$ at once. We present the constraints obtained from our analysis within the ΛCDM model, finding good agreement with the constraints from the *Planck* CMB power spectra.

This paper is organized as follows. First, in Section 2 we detail the cluster sample used in our analysis. In Section 3 we briefly describe the basics of CMB lensing. Next, in Section 4 we give a detailed account of the pipeline that we follow in order to estimate the masses of the clusters in our sample via CMB lensing. In Section 5, we explain why a mass bias calibration such as the one presented in Planck 2015 Results XXIV (2016) is, in general, biased, and in Section 6 we present a likelihood that, by combining the CMB lensing mass estimates of each cluster in our sample with the corresponding q and z, has the power to constrain Ω_m, σ_8, and $1-b$ jointly in an unbiased way. We validate this approach with simulated data in Section 7. Our parameter results are presented and discussed in Section 8. We conclude in Section 9. Finally, in Appendix A, a method for dealing with uncertainty in cluster

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

centring in future analyses is proposed, and in Appendix B the relation of our likelihood to a simple Poisson count likelihood in the space of cluster observables is discussed.

2 DATA

For our cosmological analysis, we use the clusters of the *Planck* MMF3 cosmology sample. This sample contains a total of 439 clusters and is the sample that was used in the baseline analysis in Planck 2015 Results XXIV (2016). Details of the construction of this sample are given in Planck 2015 Results XXVII (2016) and Planck 2013 Results XXIX (2014). Briefly, clusters are detected through their tSZ signature across the six highest frequency *Planck* channels, and selected by imposing that their tSZ signal-to-noise ratio $q > 6$, and that they are in the area of the sky left unmasked in the analysis (65 per cent of the sky). The cluster catalogue is publicly available in the *Planck* Legacy Archive (hereafter, PLA).[1]

As a measure of the SZ signal for each cluster in the sample, we follow Planck 2015 Results XXIV (2016) and use the signal-to-noise ratio, q. This was measured by the Planck Collaboration for all the clusters of the MMF3 cosmology sample through their MMF3 pipeline (Planck 2015 Results XXVII 2016). We similarly use the cluster redshifts obtained by the Planck Collaboration from ancillary data and follow-up observations, as described in Planck 2015 Results XXVII (2016).

In order to calibrate the mass bias, we also incorporate in our analysis CMB lensing mass estimates of all the clusters in the MMF3 cosmology sample with measured redshift. This subsample consists of a total of 433 clusters, and it is the same sample as the one used in the CMB lensing calibration presented in Planck 2015 Results XXIV (2016). In the following, we will refer to it as the *Planck* CMB lensing calibration sample. We directly estimate the masses of all the clusters in this sample from *Planck* 2015 full-mission temperature maps. A detailed account of this process is described in Section 4. For those clusters with redshift information, the Planck Collaboration also provide estimates of each cluster's mass, $M_{\rm SZ}$. These are obtained by combining measurements of the cluster SZ observables θ_{500}, the cluster angular scale, and Y_{500}, the SZ flux within a θ_{500} aperture, which are generally strongly correlated given the resolution of *Planck*, with X-ray-calibrated fiducial scaling relations to break the degeneracy; see Planck 2015 Results XXVII (2016) for further details. From this conditional estimate of Y_{500}, the mass proxy $M_{\rm SZ}$ can be derived from the same scaling relations. As emphasized in Planck 2015 Results XXVII (2016), $M_{\rm SZ}$ should be regarded as the expected *hydrostatic mass* of a cluster, given the assumed scaling relations, cluster redshift, and distribution of Y_{500} and θ_{500} derived from the data. In this work, we only use $M_{\rm SZ}$ to provide an initial angular scale for the matched filter used to estimate the cluster mass from CMB lensing data (Section 4.3), and in Section 5 where we compare with the lensing masses to provide a simple (but biased) estimate of the hydrostatic mass bias following Planck 2015 Results XXIV (2016).

3 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 fluctuations on the sky by some deflection field $\boldsymbol{\alpha}(\hat{\boldsymbol{n}})$ (see Lewis & Challinor 2006, for a general review of CMB lensing).

Let X be an 'unlensed' CMB field, i.e. a CMB field as it would have been observed if there was no lensing, where X can be T (the CMB temperature), Q, or U (the two linear polarization Stokes' parameters). Lensing remaps the CMB fields so that the lensed field $\tilde{X}(\hat{\boldsymbol{n}})$ along the line-of-sight direction $\hat{\boldsymbol{n}}$ is the unlensed field at $\hat{\boldsymbol{n}} + \boldsymbol{\alpha}(\hat{\boldsymbol{n}})$, i.e. $\tilde{X}(\hat{\boldsymbol{n}}) = X(\hat{\boldsymbol{n}} + \boldsymbol{\alpha})$.

The deflection field can be written as $\boldsymbol{\alpha} = \nabla_{\hat{\boldsymbol{n}}}\psi$, where $\nabla_{\hat{\boldsymbol{n}}}$ denotes the angular derivative (covariant derivative on the unit sphere, or, in the flat-sky approximation, partial derivative with respect to the two local angular variables) and where ψ is known as the lensing potential. For a flat universe, the lensing potential can be written as (Lewis & Challinor 2006)

$$\psi(\hat{\boldsymbol{n}}) = -2\int_0^{\chi_\star} \mathrm{d}\chi\, \frac{\chi_\star - \chi}{\chi_\star \chi}\Psi(\chi\hat{\boldsymbol{n}}, \eta_0 - \chi), \tag{1}$$

where $\chi_\star$ is the comoving distance to last scattering ($\chi_\star \approx 14$ Gpc), η_0 is the current conformal time, and Ψ is the Newtonian gravitational potential (or, in a general relativistic framework, the Weyl potential). The lensing potential is therefore a weighted integral of the gravitational potential along the undeflected line of sight.

It is often useful to work with the convergence, κ, which is given by the two-dimensional Laplacian of the lensing potential,

$$\kappa(\hat{\boldsymbol{n}}) = -\frac{1}{2}\nabla^2_{\hat{\boldsymbol{n}}}\psi(\hat{\boldsymbol{n}}). \tag{2}$$

Thus, the convergence is a weighted integral of the matter overdensity along the line of sight. Therefore, the integrated matter distribution along the line of sight determines how the CMB photons are lensed, which allows estimation of this integrated matter distribution from CMB observations alone. 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 the order of 10 μK in the measured CMB temperature (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.

4 ESTIMATION OF THE *PLANCK* CLUSTER MASSES THROUGH CMB LENSING

In this section, we describe the procedure we follow to estimate the masses of the galaxy clusters in the *Planck* CMB lensing calibration sample using the *Planck* data. This is similar to that followed in the CMB lensing cluster mass calibration presented in Planck 2015 Results XXIV (2016), but there are likely minor differences in implementation. (Planck 2015 Results XXIV 2016 do not provide full details of their implementation.)

4.1 Production of clean temperature maps

4.1.1 Cluster fields, masks, and apodization

We use the six highest frequency (100–857 GHz) Full-Mission 2015 *Planck* temperature maps, from which we cut square patches of side 256 arcmin centred on the positions of the 433 clusters of the *Planck*

[1] http://pla.esac.esa.int/pla/

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

CMB lensing calibration sample. We will refer to these cut-outs as the 'cluster frequency fields'. Since the *Planck* maps are in the HEALPIX pixelization (Górski et al. 2005), we extract the cluster frequency fields as Cartesian projections centred at the cluster positions, setting the number of pixels of each field to be 512×512, with a pixel size of 0.5 arcmin. For the cluster central positions, we use the Galactic coordinates estimated by the Planck collaboration through their cluster detection pipeline, which are publicly available in the PLA. We only use temperature maps, given that, for an experiment with the resolution and noise levels of *Planck*, the *TT* lensing quadratic estimator offers the best performance in terms of signal-to-noise (Hu & Okamoto 2002).

We cosine apodize the edges of each cluster frequency field over a width of 12.8 arcmin to suppress spectral leakage when taking Fourier transforms. We find, using simulated CMB observations, that with such apodization applied to a flat-sky, non-periodic CMB realization with resolution and noise levels similar to *Planck*, we can recover the CMB power spectrum after simple scaling to account for the apodized mask. We also mask point sources using the union of the *Planck* 100 GHz and 143 GHz 2015 temperature point-source masks provided in the PLA, and inpaint the masked regions with the method presented in Gruetjen et al. (2017) to reduce spectral leakage in later stages of our pipeline.

Next, we deconvolve the cluster frequency fields with the corresponding *Planck* isotropic instrumental beam at each field frequency (all of them available in the PLA) and with the HEALPIX pixel transfer function. We perform these operations as simple multiplications in harmonic space, which for the small fields we consider reduces to Fourier space.

4.1.2 *Foreground suppression: constrained internal linear combination*

Since we want to investigate how the primary CMB is deflected by each cluster, we need to suppress all the significant contributions to the observed signal that are on top of the primary, lensed CMB. We also need to reduce the instrument noise as much as possible. The most important foregrounds around a cluster are the thermal and kinetic SZ effects. They add noise and, more troublingly, bias to the mass measurements (e.g. Melin & Bartlett 2015). The tSZ effect has the larger amplitude, in many cases larger than the cluster lensing signal by about an order of magnitude (Lewis & Challinor 2006; Yoo, Zaldarriaga & Hernquist 2010). However, it has a very characteristic frequency dependence, different from the thermal CMB spectrum, which is independent of the cluster temperature in the non-relativistic limit (Rephaeli 1995). In this limit, the SZ spectrum is the same for all galaxy clusters. Thus, multifrequency CMB observations can be combined in order to estimate and subtract the tSZ signal if the cluster gas is, to a good approximation, non-relativistic. On the other hand, the kinetic SZ signal, a Doppler-shift-induced contribution proportional to the cluster velocity along the line of sight, has the same spectral dependence as the primary CMB, so it cannot be removed with multifrequency observations. Its amplitude can be of the same order of magnitude as the cluster lensing signal (Seljak & Zaldarriaga 2000) and, for asymmetric clusters, may be expected to bias the mass measurement at some level. However, it has been argued through simulations that the bias is not significant for an experiment with the resolution and sensitivity of *Planck* and for clusters with realistic peculiar velocities (Melin & Bartlett 2015). The kSZ does, however, add additional noise to the mass measurements.

We treat the tSZ effect as a signal to be explicitly removed from our fields and the instrumental noise and all the other foregrounds as noise on top of the primary, lensed CMB signal. In order to suppress the tSZ signal and to reduce the noise of our maps, we perform a constrained internal linear combination (CILC) with the six beam- and pixel-deconvolved cluster frequency fields, as described in Remazeilles, Delabrouille & Cardoso (2011). A CILC is a linear combination of a set of different frequency maps in which the weights are chosen in order to extract a signal with a known spectrum, minimizing the variance of the output map with the constraint that another signal with a known spectral signature is completely removed. In this case, the signal to be extracted is the primary, lensed CMB, the signal to be removed is the tSZ signal, and the output is a single, tSZ-free temperature map around each cluster. We shall refer to these maps as 'cluster temperature fields'.

In more detail, in a CILC the observations (in our case, the six cluster frequency fields around a given cluster) are modelled as Remazeilles et al. (2011)

$$\boldsymbol{x}(p) = \boldsymbol{a}s(p) + \boldsymbol{b}y(p) + \boldsymbol{n}(p). \tag{3}$$

Here, $\boldsymbol{x}(p)$ is a six-dimensional vector containing the values of the six frequency fields at the pixel p; $s(p)$ is the CMB map to be extracted, which also includes the kSZ signal as it has the same spectral dependence as the CMB; $y(p)$ is the tSZ contribution, to be removed; and $\boldsymbol{n}(p)$ denotes all the other unmodelled signals plus the instrumental noise. The fixed vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ describe the frequency dependence of the CMB (and kSZ) and tSZ signals, respectively, as integrated across the frequency bands of the experiment.

In units of the CMB temperature, the CMB vector $\boldsymbol{a} = (1, 1, 1, 1, 1, 1)^T$. In the non-relativistic limit, the tSZ vector $\boldsymbol{b}$ is the same for all clusters and has components for the *Planck* frequency bands that are given in Planck 2015 Results VIII (2016).

Once $\boldsymbol{a}$ and $\boldsymbol{b}$ are known, a weight vector $\boldsymbol{w}$ can be constructed such that $\hat{s}(p) = \boldsymbol{w}^T\boldsymbol{x}(p)$ has unit response to the signal $s(p)$ and zero response to the contaminant $y(p)$. These impose the constraints $\boldsymbol{w}^T\boldsymbol{a} = 1$ and $\boldsymbol{w}^T\boldsymbol{b} = 0$. Further demanding that the effective noise variance be minimized leads to weights (Remazeilles et al. 2011)

$$\boldsymbol{w}^T = \frac{(\boldsymbol{b}^T\mathbf{R}^{-1}\boldsymbol{b})\boldsymbol{a}^T\mathbf{R}^{-1} - (\boldsymbol{a}^T\mathbf{R}^{-1}\boldsymbol{b})\boldsymbol{b}^T\mathbf{R}^{-1}}{(\boldsymbol{a}^T\mathbf{R}^{-1}\boldsymbol{a})(\boldsymbol{b}^T\mathbf{R}^{-1}\boldsymbol{b}) - (\boldsymbol{a}^T\mathbf{R}^{-1}\boldsymbol{b})^2}, \tag{4}$$

where $\mathbf{R} = \langle \boldsymbol{n}\boldsymbol{n}^T \rangle$ is the inter-frequency covariance matrix of the six frequency maps. Here, we assume this is constant across each cluster frequency field and estimate it empirically from the data by summing over pixels.[2]

4.2 Lensing convergence estimation: quadratic estimators

The next step is to use our set of cluster temperature fields to estimate the lensing convergence around the clusters in our sample. We perform this estimation with a modified version of the well-known CMB lensing *TT* quadratic estimator, first described in Hu (2001), since for an experiment of the resolution and sensitivity of *Planck*, the *TT* quadratic estimator is very close to the optimal lensing estimator (Raghunathan et al. 2017).

CMB lensing by a fixed convergence $\kappa(\boldsymbol{x})$ destroys the statistical isotropy of the CMB, coupling Fourier modes with different wavevectors $\boldsymbol{l}$ in a κ-dependent way. The lensing quadratic estimators take advantage of this fact in order to estimate κ. The input

[2]Note that adding (symmetric) outer products to $\mathbf{R}$ involving only $\boldsymbol{a}$ and/or $\boldsymbol{b}$ does not change the weights.

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

of a quadratic estimator is a pair of observed CMB fields defined in the same region of the sky, which can consist of any combination of T, Q, or U; the input pair of the TT estimator is simply two copies of the same temperature map. Its output is an estimate $\hat{\kappa}$ of the convergence, which is unbiased at linear order in κ.

We use the following form of the TT quadratic estimator in order to estimate κ around a given cluster:

$$\hat{\kappa}(\boldsymbol{L}) = -N(\boldsymbol{L})\frac{2}{L^2}\int \frac{d^2\boldsymbol{x}}{2\pi} \mathrm{e}^{-\mathrm{i}\boldsymbol{L}\cdot\boldsymbol{x}} \nabla \cdot [F_1(\boldsymbol{x})\nabla F_2(\boldsymbol{x})]\,, \tag{5}$$

where $N(\boldsymbol{L})$ is a normalization that can be written as

$$N(\boldsymbol{L}) = \frac{L^4}{4}\left(\int \frac{d^2\boldsymbol{l}}{(2\pi)^2} \frac{\left[(\boldsymbol{L}-\boldsymbol{l})\cdot\boldsymbol{L}C^{\tilde{T}\tilde{T}}_{|\boldsymbol{l}-\boldsymbol{L}|} + \boldsymbol{l}\cdot\boldsymbol{L}C^{\tilde{T}\tilde{T}}_{l}\right]^2}{2C^{\hat{T}\hat{T}}_{l}C^{\hat{T}\hat{T}}_{|\boldsymbol{l}-\boldsymbol{L}|}}\right)^{-1}. \tag{6}$$

Here, $C^{\tilde{T}\tilde{T}}_l$ is the power spectrum of the lensed CMB temperature, and $C^{\hat{T}\hat{T}}_l$ is the (total) power spectrum of the corresponding cluster temperature field. We have also defined two filtered temperature fields,

$$F_i(\boldsymbol{l}) \equiv f_i(l)\hat{T}(\boldsymbol{l}), \tag{7}$$

for $i = 1, 2$, where $\hat{T}(\boldsymbol{l})$ is the corresponding cluster temperature field and the filters are defined as

$$f_1(l) \equiv \frac{1}{C^{\hat{T}\hat{T}}_l}, \tag{8}$$

and

$$f_2(l) \equiv \begin{cases} \frac{C^{\tilde{T}\tilde{T}}_l}{C^{\hat{T}\hat{T}}_l} & \text{if } l \le l_f, \\ 0 & \text{if } l > l_f. \end{cases} \tag{9}$$

Note that $F_1(\boldsymbol{l})$ is the inverse-variance-filtered temperature field, $F_2(\boldsymbol{l})$ is the Wiener-filtered temperature field (for $l \le l_f$), and the quadratic estimator (5) involves the product of the former with the gradient of the latter.

Our TT quadratic estimator differs from the version originally described in Hu (2001) in two respects. First, we choose $f_2(l)$ to be zero for $l > l_f$. We take $l_f = 2000$. This ensures that the estimator remains unbiased in regions around galaxy clusters, where second and higher orders of κ become important and the usual filtering yields biased convergence estimates (Hu, DeDeo & Vale 2007). In addition, we use the lensed CMB power spectrum, $C^{\tilde{T}\tilde{T}}_l$, in $f_2(l)$ and in the normalization, rather than the unlensed CMB power spectrum, as this gives (approximately) the correct response of the estimator $\hat{\kappa}(\boldsymbol{L})$ to lenses at wavevector $\boldsymbol{L}$ averaged over all other lensing modes (Hanson et al. 2011; Lewis, Challinor & Hanson 2011).

We estimate $C^{\hat{T}\hat{T}}_l$ empirically for each of our cluster temperature fields to account for the variation of residual foregrounds and instrumental noise across the sky. We calculate $C^{\tilde{T}\tilde{T}}_l$ with CAMB[3] assuming our fiducial cosmology. We then extract the central 256×256 pixels from our cluster temperature fields, and it is on these 128 arcmin $\times$ 128 arcmin fields ('reduced cluster temperature fields') that we reconstruct the lensing convergence. The reason for doing this is as follows. We first produce larger cluster temperature fields in order to obtain less noisy local power spectrum estimates. We then choose smaller fields on which to perform the lensing reconstruction to reduce the effect of neighbouring clusters and large-scale structure that is correlated with the cluster, which could contaminate the mass measurement of the central cluster.

Before performing the lensing reconstruction, we cosine apodize each reduced cluster temperature field over a width of 6.4 arcmin and then apply the same point-source mask used to mask the corresponding set of frequency maps. We inpaint the source-masked regions with constrained Gaussian realizations, as detailed in Benoit-Lévy et al. (2013), so that the response of the quadratic estimator is the same as if there was no point-source masking. We account for the apodization of the edges of the reduced cluster temperature fields by multiplying the reconstructed lensing convergence by a factor of $1/\sqrt{f_4}$, where f_4 is given by (Benoit-Lévy et al. 2013)

$$f_4 = \frac{1}{N_{\rm pix}}\sum_i w_i^4. \tag{10}$$

Here, $N_{\rm pix}$ is the number of pixels in our reduced cluster temperature fields, w_i is the value of the apodization mask at pixel i, and the sum is over all the pixels. Despite this normalization being appropriate for measuring the power spectrum of statistically isotropic lensing fields, we verified with simulations that we accurately recover the cluster convergence with this choice.

A mean field has to be subtracted from the lensing reconstruction to account for non-lensing sources of statistical anisotropy, which the quadratic estimator mistakes for lensing, thereby biasing the reconstruction (e.g. Benoit-Lévy et al. 2013; Namikawa, Hanson & Takahashi 2013).

The sources of the mean field include inhomogeneous noise across the sky, anisotropic instrumental beams, and masks. All of these sources are present, to a certain degree, in our fields: *Planck* noise levels vary across the sky as does the residual foreground power in the CILC; the *Planck* (effective) instrumental beams are not perfectly isotropic; and our apodization mask breaks isotropy. (The apodization actually suppresses a larger mean field that would be present if there was no apodization at all, due to the non-periodicity of the CMB cluster fields.) In order to subtract the mean field, we define the following modified quadratic estimator:

$$\hat{\kappa}_{\rm mod} = \frac{1}{\sqrt{f_4}}\left(\hat{\kappa} - \hat{\kappa}_{\rm mf}\right), \tag{11}$$

where $\hat{\kappa}$ is the quadratic estimator presented above and $\hat{\kappa}_{\rm mf}$ is an estimate of the mean field.

The mean fields will differ across the cluster fields due to variation in the noise, effective beams, and residual foreground levels. (We note that the largest contributions to the mean fields are from the masking, but this is approximately the same for all cluster fields, differing only in the point-source masking/inpainting.) We deal with this by estimating local mean fields from the data using reduced cluster temperature fields at random locations on subdivisions of the sky. In detail, we divide the sky into 20 regions in bands of ecliptic latitude, since the instrument noise and beam contributions to the mean field are expected to vary significantly with latitude due to the *Planck* scan strategy. We identify each of the 10 bands falling in the northern-ecliptic hemisphere with the corresponding band located at its mirror location in the southern-ecliptic hemisphere, yielding a total of 10 distinct bands. We then estimate a mean field within each band and subtract this from the raw convergence for clusters located with the band. To estimate the mean fields, we randomly select 16 384 fields with the same size and pixelization as the cluster frequency fields, but centred on randomly located points within each band and within the 2015 *Planck* 60 per cent Galactic mask, which was the mask used in the construction of

[3] http://camb.info/

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

the MMF3 cosmology catalogue. We apply the pipeline described above to the six *Planck* frequency maps for each randomly centred frequency field and average the resulting estimated convergences (estimated using $\hat{\kappa}$) within each band. We tested with simulated CMB realizations that the number of random fields used to estimate the mean fields is large enough to allow us to ignore statistical errors in the mean fields in the mass estimates.

The estimated κ map around each cluster ('cluster reconstructed potential field') is therefore obtained by applying the quadratic estimator $\hat{\kappa}$ to the corresponding apodized and inpainted reduced cluster temperature field, subtracting an estimate of the mean field, and finally renormalizing by $1/\sqrt{f_4}$. At fixed cluster convergence, the variance of this modified estimator is approximately given by

$$\left\langle \Delta\hat{\kappa}_{\rm mod}\left(\boldsymbol{L}\right)\Delta\hat{\kappa}^{*}_{\rm mod}\left(\boldsymbol{L}'\right)\right\rangle = \delta^{(2)}\left(\boldsymbol{L}-\boldsymbol{L}'\right)N_{\kappa}(\boldsymbol{L}), \tag{12}$$

where

$$N_{\kappa}(\boldsymbol{L}) = C_{L}^{\kappa\kappa} + N^{(0)}(L) + N^{(1)}(L). \tag{13}$$

Here, $C_L^{\kappa\kappa}$ is the lensing potential power spectrum (which we obtain from CAMB) and describes the variance from large-scale structure that is uncorrelated with the cluster; $N^{(0)}(L)$, known as the $N^{(0)}$ bias, describes Gaussian fluctuations of the CMB and noise that mimic the effects of lensing; and $N^{(1)}(L)$, known as the $N^{(1)}$ bias, depends linearly on $C_L^{\kappa\kappa}$ and arises from alternative couplings of the lensed CMB four-point function with the two copies of the lensing quadratic estimator (Kesden, Cooray & Kamionkowski 2003). The $N^{(0)}$ bias is approximately equal to the estimator normalization $N(\boldsymbol{L})$ given in equation (6), and we calculate the $N^{(1)}$ bias in the flat-sky limit following Kesden et al. (2003).

We perform the quadratic lensing reconstruction with the help of `quicklens`, a freely available Python CMB lensing package.[4]

4.3 Matched filtering

We estimate a cluster mass from each cluster reconstructed convergence field using a matched filter, as presented in Melin & Bartlett (2015) and used in the *Planck* analysis in Planck 2015 Results XXIV (2016). The method consists of matched filtering the estimated convergence, $\hat{\kappa}_{\rm mod}$, with a model for the cluster convergence (the lensing potential is matched filtered in Melin & Bartlett 2015, but both procedures are equivalent).

4.3.1 Cluster model

Following Planck 2015 Results XXIV (2016), we choose for the cluster mass model a truncated NFW profile (Navarro, Frenk & White 1997):

$$\rho(r) = \begin{cases} \frac{\rho_0}{(r/r_{\rm s})(1+r/r_{\rm s})^2} & \text{if } r \leq R_{\rm trunc}, \\ 0 & \text{if } r > R_{\rm trunc}, \end{cases} \tag{14}$$

where ρ_0 is a characteristic density, $r_{\rm s}$ is a characteristic scale radius, and $R_{\rm trunc}$ is the truncation radius.

In order to be consistent with Planck 2015 Results XXIV (2016), we define the cluster mass in terms of M_{500}, the mass contained with a radius R_{500} such that the mean enclosed mass is 500 times that of the critical density at the cluster's redshift, $\rho_{\rm c}(z)$. Imposing the definition of M_{500},

$$\rho_0 = \rho_c(z)\frac{500}{3}\frac{c_{500}^3}{\ln(1+c_{500}) - c_{500}/(1+c_{500})}, \tag{15}$$

where the concentration parameter is defined by $c_{500} = R_{500}/r_{\rm s}$. As in Melin & Bartlett (2015), we assume that $c_{500} = 3$, although it is known that it actually varies weakly with cluster mass and redshift (Muñoz-Cuartas et al. 2011). We further follow Melin & Bartlett (2015) in their choice of truncation radius: $R_{\rm trunc} = 5R_{500}$. Fixing c_{500}, for our purposes a cluster is completely specified by two parameters, e.g. M_{500} and redshift z.

The lensing convergence of the cluster at sky position $\boldsymbol{\theta}$ from its centre, $\kappa_{\rm cl}(\boldsymbol{\theta})$, is related to the (physical) surface density at projected position $\boldsymbol{r}$ from the centre, $\Sigma_{\rm cl}(\boldsymbol{r})$, through

$$\kappa_{\rm cl}(\boldsymbol{\theta}) = \Sigma_{\rm cl}(\boldsymbol{r})/\Sigma_{\rm crit}(z), \tag{16}$$

where the critical surface density at redshift z is

$$\Sigma_{\rm crit}(z) = \frac{c^2}{4\pi G}\frac{d_{A,\rm CMB}}{d_{A,\rm cl}d_{A,\rm cl\text{-}CMB}}, \tag{17}$$

which involves the physical angular diameter distances to the CMB last-scattering surface, to the cluster, and between the cluster and the CMB. The projected position at the cluster, $\boldsymbol{r}$, is related to the angular position, $\boldsymbol{\theta}$, through $\boldsymbol{r} = d_A(z)\boldsymbol{\theta} = d_{A,\rm cl}\boldsymbol{\theta}$. The projected surface density can be written as

$$\Sigma_{\rm cl}(\boldsymbol{r}) = 2\rho_0 r_{\rm s}\int_{r/r_{\rm s}}^{5c_{500}}\frac{{\rm d}x}{(1+x)^2\sqrt{x^2-(r/r_{\rm s})^2}}, \tag{18}$$

for $|\boldsymbol{r}| \leq R_{\rm trunc}$. For a fixed concentration, the integral only depends on $\theta/\theta_{\rm s} = |\boldsymbol{r}|/r_{\rm s}$, where the angular scale radius is $\theta_{\rm s} = r_{\rm s}/d_A(z)$. It follows that the cluster convergence can be written in the form

$$\kappa_{\rm cl}(\boldsymbol{\theta}) = \kappa_0\kappa_{\rm t}(\boldsymbol{\theta};\theta_{\rm s}), \tag{19}$$

where the circularly symmetric template function $\kappa_{\rm t}(\boldsymbol{\theta};\theta_{\rm s})$ depends only on $\theta/\theta_{\rm s}$. We choose the normalization such that $\kappa_{\rm t} = 1$ at the scale radius, so that κ_0 is the cluster convergence there. It follows that

$$\kappa_0 = \frac{2\rho_0 r_{\rm s}}{\Sigma_{\rm crit}(z)}\int_1^{5c_{500}}\frac{{\rm d}x}{(1+x)^2\sqrt{x^2-1}}, \tag{20}$$

and so, at fixed concentration,

$$\kappa_0\theta_{\rm s}^2 \propto \frac{M_{500}}{\Sigma_{\rm crit}(z)d_A^2(z)}, \tag{21}$$

i.e. the integrated convergence within an aperture of radius $\theta_{\rm s}$ is proportional to the cluster mass.

4.3.2 Matched filter

Given an estimate of a cluster angular size $\theta_{\rm s}$, we filter its reconstructed convergence field with the template $\kappa_{\rm t}$ to form an estimator for κ_0:

$$\hat{\kappa}_0 = \left[\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{|\kappa_{\rm t}(\boldsymbol{L})|^2}{N_{\kappa}(\boldsymbol{L})}\right]^{-1}\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{\hat{\kappa}(\boldsymbol{L})\kappa_{\rm t}^{*}(\boldsymbol{L})}{N_{\kappa}(\boldsymbol{L})}, \tag{22}$$

where $N_{\kappa}(\boldsymbol{L})$ is the variance of $\hat{\kappa}(\boldsymbol{L})$ given in equation (13). The inverse-variance weighting of $\hat{\kappa}$ ensures this estimator has minimum variance, which can be calculated to be

$$\sigma_{\kappa_0}^2 \equiv \langle(\hat{\kappa}_0-\kappa_0)^2\rangle = \frac{1}{2\pi}\left[\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{|\kappa_{\rm t}(\boldsymbol{L})|^2}{N_{\kappa}(\boldsymbol{L})}\right]^{-1}. \tag{23}$$

[4] https://github.com/dhanson/quicklens

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

In order to use the filter, an estimate of the cluster angular scale is needed. We use the SZ mass proxy $M_{\rm SZ}$ of each cluster, as provided in the MMF3 catalogue, as the filter mass, $M^{\rm fid}$, from which we derive an SZ angular size $\theta_{\rm s}^{\rm fid}$:

$$M^{\rm fid} = 500\frac{4\pi}{3}\left[d_A(z)c_{500}\theta_{\rm s}^{\rm fid}\right]^2\rho_{\rm c}(z). \tag{24}$$

Since $\kappa_0\theta_{\rm s}^2 \propto M_{500}$, we can write an estimator of the cluster's mass as

$$\hat{M}_{500} = M^{\rm fid}\left(\frac{\hat{\kappa}_0}{\kappa_0^{\rm fid}}\right), \tag{25}$$

where $\kappa_0^{\rm fid}$ is the template's convergence at its scale radius $\theta_{\rm s}^{\rm fid}$. The standard deviation of this estimator is simply

$$\sigma_{M_{500}} = M^{\rm fid}\left(\frac{\sigma_{\kappa_0}}{\kappa_0^{\rm fid}}\right). \tag{26}$$

If the filter matches exactly the cluster's true profile, $\hat{M}_{500}$ is an unbiased estimator of the cluster's true mass, M_{500}. However, as we now discuss, $\hat{M}_{500}$ is biased at linear order in any mismatch between the template and the true profile. We can avoid such linear bias by instead working with the signal-to-noise on κ_0 (or, equivalently, on M_{500}), defining the observable

$$p_{\rm obs} \equiv \hat{\kappa}_0/\sigma_{\kappa_0} = \hat{M}_{500}/\sigma_{M_{500}}. \tag{27}$$

4.3.3 *Biases from template errors*

We consider three possible deviations of the template $\kappa_{\rm t}$ from the true cluster convergence, which will bias the estimator to some extent: mismatch between the filter fiducial angular scale and the true cluster angular scale, actual mismatch between the filter profile and the true profile, and misplacing of the filter position with respect to the cluster's true position (which we refer to as miscentring).

Let us first consider the effect of mismatch between the filter fiducial angular scale and the true cluster angular scale. Consider a cluster of true size $\theta_{\rm s}$, but filtered with $\theta_{\rm s}^{\rm fid}$. The expected value of the signal-to-noise is

$$\langle p_{\rm obs}\rangle = \sqrt{2\pi}\kappa_0\left[\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{|\kappa_{\rm t}(\boldsymbol{L};\theta_{\rm s}^{\rm fid})|^2}{N_\kappa(\boldsymbol{L})}\right]^{-1/2} \times\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{\kappa_{\rm t}(\boldsymbol{L};\theta_{\rm s})\kappa_{\rm t}^*(\boldsymbol{L};\theta_{\rm s}^{\rm fid})}{N_\kappa(\boldsymbol{L})}, \tag{28}$$

where κ_0 is the true value of the convergence at the true scale radius. Recalling that the template convergence $\kappa_{\rm t}(\boldsymbol{\theta};\theta_{\rm s})$ depends only on $\theta/\theta_{\rm s}$, we can write

$$\kappa_{\rm t}(\boldsymbol{\theta};\theta_{\rm s}) = f(\theta/\theta_{\rm s}), \tag{29}$$

where the function f is circularly symmetric. The Fourier transform $\kappa_{\rm t}(\boldsymbol{L};\theta_{\rm s})$ is therefore related to the 2D Fourier transform of f through

$$\kappa_{\rm t}(\boldsymbol{L};\theta_{\rm s}) = \theta_{\rm s}^2 f(\theta_{\rm s}\boldsymbol{L}). \tag{30}$$

Substituting into equation (28), we have

$$\langle p_{\rm obs}\rangle = \sqrt{2\pi}\kappa_0\theta_{\rm s}^2\left[\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{f^2(\theta_{\rm s}^{\rm fid}\boldsymbol{L})}{N_\kappa(\boldsymbol{L})}\right]^{-1/2} \times\int\frac{d^2\boldsymbol{L}}{2\pi}\frac{f(\theta_{\rm s}\boldsymbol{L})f(\theta_{s}^{\rm fid}\boldsymbol{L})}{N_\kappa(\boldsymbol{L})}. \tag{31}$$

The dependence of $\langle p_{\rm obs}\rangle$ on $\theta_{\rm s}^{\rm fid}$ can be assessed by writing $\theta_{\rm s}^{\rm fid} = \theta_{\rm s} + \delta\theta_{\rm s}$ and series expanding in $\delta\theta_{\rm s}$. For lens reconstruction with the *Planck* data, the reconstruction noise is large on typical cluster scales, i.e. for $\theta_{\rm s}|\boldsymbol{L}| \gtrsim 1$, so $\delta\theta_{\rm s}|\boldsymbol{L}| \ll 1$ for the modes that dominate the integrals in equation (31) and the series expansion is expected to converge rapidly. It is straightforward to show that there is no linear dependence of $\langle p_{\rm obs}\rangle$ on $\delta\theta_{\rm s}$, so the bias from using the incorrect cluster angular size in the filtering is only second order in the size error.

Table 1. Results of our numerical test of the dependence of the expected value of the matched filter output on the filter angular scale $\theta_{\rm s}^{\rm fid}$. The cluster considered here has a mass $M_{500} = 0.5\times10^{15}\,{\rm M}_\odot$, is at redshift $z = 0.2$, and has our truncated NFW profile as its convergence. The templates with which we filter its convergence have functionally the same profile and are also placed at redshift $z = 0.2$, but have a number of different masses $M^{\rm fid}$, which, at fixed redshift, correspond uniquely to a set of different filter angular scales $\theta_{\rm s}^{\rm fid}$. The first column shows the filter masses, $M^{\rm fid}$, that we have considered; the second column shows the corresponding values of $\langle p_{\rm obs}\rangle$; and the third column, the corresponding values of the expected value of the filter mass estimate, $\langle\hat{M}_{500}\rangle$. It can be seen that $\langle p_{\rm obs}\rangle$ is very insensitive to $M^{\rm fid}$ (and, therefore, to $\theta_{\rm s}^{\rm fid}$) if $M^{\rm fid}$ is reasonably close to the cluster's true mass, whereas $\langle\hat{M}_{500}\rangle$ has a significant dependence on $M^{\rm fid}$.

$M^{\rm fid}$ $10^{15}\,{\rm M}_\odot$	$\langle p_{\rm obs}\rangle$	$\langle\hat{M}_{500}\rangle$ $10^{15}\,{\rm M}_\odot$
0.01	0.2023	0.25
0.10	0.2136	0.36
0.35	0.2164	0.46
0.50	0.2165	0.50
1.00	0.2159	0.58
10.00	0.2055	0.98

We can check numerically that this bias on $p_{\rm obs}$ from adopting an incorrect filter size is indeed small. In particular, we consider a cluster with a mass $M_{500} = 0.5\times10^{15}\,{\rm M}_\odot$ at redshift $z = 0.2$ and with our truncated NFW profile as its true convergence profile. We compute $\langle p_{\rm obs}\rangle$ with equation (31), taking the template to be the true convergence profile but scaled at a number of different angular scales $\theta_{\rm s}^{\rm fid}$. For the variance $N_\kappa(\boldsymbol{L})$ we use that for a *TT* quadratic estimator reconstruction for an idealized *Planck*-like experiment with a Gaussian beam of full width at half-maximum equal to 5 arcmin and noise level of 45 μK arcmin. The cluster masses associated with each of the filter angular scales considered, $M^{\rm fid}$, computed with equation (24), and their corresponding $\langle p_{\rm obs}\rangle$ values, are shown in Table 1. It can be seen that value of the filter angular scale $\theta_{\rm s}^{\rm fid}$, parametrized in Table 1 in terms of the filter mass proxy $M^{\rm fid}$, has little impact on $\langle p_{\rm obs}\rangle$ if it is within a reasonable range of the cluster's true angular scale $\theta_{\rm s}$. Since our filter mass proxies for each cluster are their SZ mass proxies, as provided in the MMF3 catalogue, and given that we expect them to be about $1-b\approx0.7$ times the cluster true masses, we conclude that the choice of filter angular scale has a negligible impact on our analysis. This would have not been the case if we had chosen to use the filter mass estimates directly, as was done in the CMB lensing calibration in Planck 2015 Results XXIV (2016). Indeed, the third column of Table 1 shows the expected values of the cluster mass estimates at each filter angular scale, $\langle\hat{M}_{500}\rangle$. It can be observed that a significant bias in the mass estimates appears if the filter angular scale $\theta_{\rm s}^{\rm fid}$ is different from the truth, $\theta_{\rm s}$ – in our test, that corresponding to $M_{500} = 0.5\times10^{15}\,{\rm M}_\odot$. Indeed, $\hat{M}_{500}$ is biased at linear order, while $p_{\rm obs}$ is biased only at second order. We therefore use $p_{\rm obs}$ as our lensing mass observable.

Let us now consider the more general case of mismatch between the filter profile and the mean true cluster profile at given cluster

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

mass and redshift. Following an analogous argument to that for the bias from mismatch in angular size, we argue that profile mismatch biases $\hat{M}_{500}$ to linear order, but $p_{\rm obs}$ only to second order. If the mismatch is small, the bias on $p_{\rm obs}$ will therefore also be small.

It should also be noted that, even if the mean cluster profile at given mass M_{500} and redshift z matches our model, the profile of a given real cluster with mass M_{500} and redshift z can be thought of as a 'noisy' realization of such a mean profile, with the scatter arising from triaxiality and large-scale structure correlated with the cluster. This intrinsic scatter is not modelled in our matched filtering process, which only includes variance arising from the lensing reconstruction noise and from uncorrelated large-scale structure, so is still present in our $p_{\rm obs}$ measurements. As we explain in Section 6, we do, however, account for intrinsic scatter in the mass measurements in our likelihood.

Finally, let us consider the possibility of miscentring of the filter with respect to the cluster's true position. We centre the filter at the mid-point of each reduced cluster temperature field, corresponding to the SZ-estimated position of the cluster centre as provided in the *Planck* MMF3 catalogue. These positions are, however, not known perfectly. The *Planck* MMF3 catalogue includes an estimate of the 95 per cent confidence interval of the positional uncertainty of each cluster due to the *Planck* beam. Its mean and median values over the cluster sample are both around 2.4 arcmin, which implies a typical offset of about 1 arcmin. In addition, there is another source of miscentring arising from the offset between the SZ-estimated centre and the centre of mass of each cluster, which is where we ought to place our template. This offset is difficult to estimate, but we expect it to be smaller than the offset due to the beam. Indeed, using hydrodynamical simulations, Gupta et al. (2017) find that for about 80 per cent of the clusters the typical value of this offset is of about $0.04R_{500}$. For a typical cluster in our sample, this translates into an offset of about 0.25 arcmin, which, added in quadrature to the offset due to the beam, does not increase it significantly. We can therefore expect a typical miscentring angle, d, of about 1 arcmin. Assuming that the cluster's mean true profile exactly matches the filter, the filter and the true mean profile are related by

$$\kappa_{\rm t}^{\rm true}(\boldsymbol{L}) = \kappa_{\rm t}(\boldsymbol{L};\theta_{\rm s})e^{i\boldsymbol{L}\cdot\boldsymbol{d}}, \tag{32}$$

where $\boldsymbol{d}$ is the centring error. It is straightforward to show that the bias this induces in $p_{\rm obs}$ is only second order in d (see Appendix A for details). We expect the leading-order bias to be negative, following the $O(\boldsymbol{L}\cdot\boldsymbol{d})^2$ term in the expansion of $\exp(i\boldsymbol{L}\cdot\boldsymbol{d})$, with a value of around 3 per cent for $d = 1$ arcmin in the case in which the filter variance $N_\kappa(\boldsymbol{L})$ is appropriate for a TT quadratic estimator reconstruction for a *Planck*-like experiment. In this scenario, the next-order non-vanishing contribution to the bias is $O(d^4)$ and is at the 10^{-4} level, and thus negligible.

We do not explicitly correct for these small, second-order residual biases in our lensing measurements. We do, however, include an effective CMB lensing mass bias parameter in our likelihood analysis, as explained in Section 6, which we marginalize over to account for the biases.

5 PITFALLS OF SIMPLE ESTIMATION OF THE SZ MASS BIAS

As described in Planck 2015 Results XXIV (2016), the mass bias $1-b$ is introduced in the *Planck* cluster-count analysis in order to account for a possible bias in the X-ray-derived masses used to calibrate the SZ–mass scaling relations. It enters the analysis by multiplying the true cluster mass, M_{500}, wherever M_{500} appears in the scaling relations.

As proposed in Planck 2015 Results XXIV (2016), a possible way to estimate the mass bias is to construct the following estimator of $1/(1-b)$:

$$\frac{\hat{1}}{1-b} = \frac{M_{\rm lens}}{M_{Y_z}}, \tag{33}$$

where $M_{\rm lens}$ is a lensing estimate of the cluster mass and M_{Y_z} is the corresponding scaling-relation-derived mass. Averaging over an ensemble of clusters, this estimator is an unbiased estimator of $1/(1-b)$ if the following conditions are met: (i) the $M_{\rm lens}$ measurements are unbiased estimates of M_{500}; (ii) the M_{Y_z} measurements are unbiased estimates of the mean SZ mass at a given M_{500}, which is what we can define as $(1-b)M_{500}$; and (iii) the scatter on the mean SZ mass at a given M_{500} can be neglected, that is, the M_{Y_z} measurements are precisely $(1-b)M_{500}$. Indeed, if these conditions are satisfied,

$$\left\langle \frac{\hat{1}}{1-b} \right\rangle = \left\langle \frac{M_{\rm lens}}{M_{Y_z}} \right\rangle = \frac{\langle M_{\rm lens}\rangle}{M_{Y_z}} = \frac{M_{500}}{(1-b)M_{500}} = \frac{1}{1-b}, \tag{34}$$

where angular brackets denote averaging over the cluster sample with some appropriate weighting to minimize the variance. It should be noted that even with such weighting, this estimator is not in general the optimal estimator of the mass bias.

The conditions needed for this estimator of $1/(1-b)$ to be unbiased are not fully met for the *Planck* CMB lensing calibration sample, yielding, in general, an incorrect estimate of the mass bias. A detailed quantification of the size of the error, and the extent to which the underlying assumptions are invalid, is beyond the scope of this paper. However, here we offer a brief, qualitative description of the unsuitability of two of the three assumptions that are necessary for the estimator to be accurate.

First, as already noted in Section 4, it is known that, even in the limit of an unbiased lensing convergence reconstruction around a cluster, the matched filtering process generally gives biased mass estimates due to, e.g. mismatch between the true and template convergence profiles and miscentring. If unaccounted for, as in the calibration presented in Planck 2015 Results XXIV (2016), these biases in the lensing mass estimates propagate directly into $1/(1-b)$.

Secondly, at a given M_{500}, the SZ mass estimates in the sample, M_{Y_z}, are biased high compared to the true mean SZ mass, $(1-b)M_{500}$. This is due to clusters in our calibration sample being selected through a cut-off on the SZ signal-to-noise, q (plus the additional neglect of six clusters for which there are no redshift measurements available, although this should have very little impact on the modelling of the selection of our sample). This selection effect, which has already been studied in the literature (e.g. Mantz et al. 2010; Allen et al. 2011; Nagarajan et al. 2018), can be understood as follows. Since M_{Y_z} are noisy realizations of the mean SZ mass at a given M_{500}, close to the selection cut-off the mean of the SZ masses that get included in the sample for a given M_{500} is necessarily larger than the true population mean $(1-b)M_{500}$. As a consequence, M_{Y_z} becomes a biased estimator of $(1-b)M_{500}$ close to the cut-off for our sample, even if before selection it was unbiased. We therefore expect the simple estimator (33) to underestimate $1/(1-b)$, i.e. $1-b$ to be biased high.

This selection effect also has some impact on the lensing mass estimates, despite these not intervening in the selection of the

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

sample. This is because the intrinsic scatter in lensing masses is expected to be correlated with that in the SZ masses to a certain extent. Indeed, for a given cluster, both the SZ and the lensing mass estimates are obtained from quantities integrated along the line of sight. Thus, it is reasonable to expect some correlation in their intrinsic scatter, e.g. due to cluster triaxiality. Correlations between SZ and galaxy weak lensing masses have indeed been observed in realistic simulations of galaxy clusters, with a reported intrinsic correlation coefficient as high as 0.8 (Shirasaki, Nagai & Lau 2016). Similarly, we can expect the scatter in the CMB lensing mass of a cluster to correlate with that in its SZ mass, biasing the mean lensing mass in the sample high if the correlation is positive. Nevertheless, we expect this to be a small effect for current CMB lensing mass estimates since the measurement errors dominate over the intrinsic scatter.

6 JOINT BAYESIAN ANALYSIS OF CLUSTER SZ AND MASS MEASUREMENTS

In this section, we develop our Bayesian model to constrain cosmological parameters, notably $\Omega_{\rm m}$ and σ_8, and the SZ mass bias (along with several other nuisance parameters) from joint analysis of the SZ data and cluster mass estimates from CMB lensing. The central object is the likelihood, $\mathcal{L}$, giving the probability of the data given the model parameters. The data we use are as follows: (i) the total number of clusters in the MMF3 cosmology sample, N; (ii) the cluster locations on the celestial sphere, $\{\hat{\boldsymbol{n}}_i\}$; and (iii) the cluster redshifts, $\{z_{{\rm obs},\,i}\}$, as measured by the Planck collaboration in follow-up observations, SZ signal-to-noise ratios, $\{q_{{\rm obs},\,i}\}$, as measured by the Planck collaboration through their MMF3 pipeline, and CMB lensing signal-to-noise ratios, $\{p_{{\rm obs},\,i}\}$, as measured in this work (see Section 4). We collectively refer to $z_{{\rm obs},\,i}$, $q_{{\rm obs},\,i}$, and $p_{{\rm obs},\,i}$ as the 'mass data point' D_i of cluster i. Since redshift measurements are not available for six of the clusters in the MMF3 cosmology sample, the mass data point of each such cluster reduces to $q_{{\rm obs},\,i}$.

The likelihood can be written as the probability density function of N, $\underline{\hat{\boldsymbol{n}}}$, and $\underline{D}$, $P\,(N, \underline{\hat{\boldsymbol{n}}}, \underline{D})$, where $\underline{\hat{\boldsymbol{n}}}$ is the vector whose i-th component is $\hat{\boldsymbol{n}}_i$, and $\underline{D}$ is the vector whose i-component is D_i. This probability density function can be decomposed as

$$P\,(N, \underline{\hat{\boldsymbol{n}}}, \underline{D}) = P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})P(\underline{\hat{\boldsymbol{n}}}|N)P(N). \tag{35}$$

The first term, $P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$, which we shall refer to as $\mathcal{L}_1$, is the probability of obtaining the cluster mass data points $\underline{D}$ given that N clusters have been included in our sample (the MMF3 cosmology sample) and that their sky locations have been found to be $\underline{\hat{\boldsymbol{n}}}$. The dependence on sky location is important, since foregrounds and the *Planck* instrumental noise vary significantly across the sky. The second term, $P(\underline{\hat{\boldsymbol{n}}}|N)$, which we shall denote with $\mathcal{L}_2$, is the probability that, given N clusters have been included in our sample, they are located at the sky locations $\underline{\hat{\boldsymbol{n}}}$. Finally, the third term, $P(N)$, which we shall refer to as $\mathcal{L}_3$, is the probability of including a total of N clusters in our cluster sample.

Our likelihood is a natural way to extend the SZ count formalism in order to incorporate the CMB lensing mass measurements to allow for self-calibration of the SZ mass bias. In the rest of this section, we develop the three factors in equation (35), making clear the parameters on which they depend. In Appendix B we show how, in general, a likelihood like ours is equivalent to a Poisson count likelihood in z–$q_{\rm obs}$–$p_{\rm obs}$ space (in a suitable limit) and how, in particular, our likelihood can be reduced by marginalization over $p_{\rm obs}$ to a likelihood similar to the *Planck* SZ count likelihood.

6.1 $\mathcal{L}_1$: mass data likelihood

In order to construct $\mathcal{L}_1 = P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$, we assume that each cluster in our sample is statistically independent of the others. This assumption was also made in Planck 2015 Results XXIV (2016), where it was claimed that the impact on the results of the correlations between the different clusters in the sample due to large-scale clustering is weak. As a consequence, we can write $P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$ as a product of the probability density functions of the mass data point of each cluster in the sample:

$$P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}}) = \prod_{i=1}^{N} P(D_i|{\rm in}, \hat{\boldsymbol{n}}_i). \tag{36}$$

Recall that D_i is the mass data point of each cluster, $D_i = (q_{{\rm obs},\,i}, p_{{\rm obs},\,i}, z_{{\rm obs},\,i})$ for the clusters with known redshift, and $D_i = (q_{{\rm obs},\,i})$ for the clusters with unknown redshift. The condition in $P(\underline{D}|N, \underline{\hat{\boldsymbol{n}}})$ that N clusters are included in the sample is translated into each $P(D_i|{\rm in}, \hat{\boldsymbol{n}}_i)$ as the condition that each of the clusters is included in the sample. We denote this condition with 'in' in equation (36) and throughout.

We take $P(D_i|{\rm in}, \hat{\boldsymbol{n}}_i)$ to have the same functional form for all the clusters with known redshift. The same applies for the clusters with unknown redshift. In this case, $P(D_i|{\rm in}, \hat{\boldsymbol{n}}_i)$ is obtained by marginalizing the probability density function of the mass data point of a cluster with known redshift over the corresponding $p_{{\rm obs},\,i}$ and $z_{{\rm obs},\,i}$.

In the following we describe how we construct $P(D_i|{\rm in}, \hat{\boldsymbol{n}}_i)$ for a cluster with known redshift.[5] To avoid clutter in the notation, hereafter we will drop the cluster index, i. First, we note that

$$P(D|{\rm in}, \hat{\boldsymbol{n}}) = \frac{P({\rm in}|D, \hat{\boldsymbol{n}})P(D|\hat{\boldsymbol{n}})}{P({\rm in}|\hat{\boldsymbol{n}})}. \tag{37}$$

Given the selection criterion applied in the construction of the MMF3 cosmology sample, $q_{\rm obs} > 6$, $P({\rm in}|D, \hat{\boldsymbol{n}})$ is simply a step function at $q_{\rm obs} = 6$. In order to obtain the other two terms in equation (37), we adopt a hierarchical model to link each cluster mass data point, $D = (q_{\rm obs}, p_{\rm obs}, z_{\rm obs})$, to the true cluster mass, M_{500}, and redshift, z, and then assume a probability distribution for M_{500} and z (which is what theory can predict).

Our hierarchical model has two layers between M_{500} and z and the mass data point, D. First, $q_{\rm obs}$, $p_{\rm obs}$, and $z_{\rm obs}$ are thought of as noisy realizations of their 'true' values, $q_{\rm t}$, $p_{\rm t}$, and $z_{\rm t}$, respectively. These are the values of these quantities that would be obtained after averaging over all sources of 'observational noise'. We refer to this layer as 'observational scatter', and specify what we mean exactly by observational noise below. Secondly, $q_{\rm t}$, $p_{\rm t}$, and $z_{\rm t}$ are understood as noisy realizations of some mean $\bar{q}\,(M_{500}, z, \hat{\boldsymbol{n}})$, $\bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}})$, and $\bar{z}\,(M_{500}, z, \hat{\boldsymbol{n}})$, respectively, which are specified at given M_{500}, z, and cluster sky location, $\hat{\boldsymbol{n}}$. We refer to this layer as 'intrinsic scatter', and specify its physical sources below.[6] With this hierarchical model

[5] For generality, we develop the likelihood formalism allowing for scatter in the redshift estimates, although in our implementation with the *Planck* clusters we can ignore such scatter for those clusters with redshift information.

[6] More precisely, as explained below, $\ln \bar{q}\,(M_{500}, z, \hat{\boldsymbol{n}})$ and $\ln \bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}})$ are the means of $\ln q_{\rm t}$ and $\ln p_{\rm t}$, respectively.

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

in mind, we can write the probability density function followed by the mass data point of one single cluster as

$$P(D|\hat{\boldsymbol{n}}) = P(q_{\rm obs}, p_{\rm obs}, z_{\rm obs}|\hat{\boldsymbol{n}}) = \int {\rm d}q_{\rm t}{\rm d}p_{\rm t}{\rm d}z_{\rm t}{\rm d}M_{500}{\rm d}z \, [P(q_{\rm obs}, p_{\rm obs}, z_{\rm obs}|q_{\rm t}, p_{\rm t}, z_{\rm t}, \hat{\boldsymbol{n}}) \times P(q_{\rm t}, p_{\rm t}, z_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}})P(M_{500}, z|\hat{\boldsymbol{n}})] \,, \quad (38)$$

where the first factor of the integrand accounts for the observational scatter, the second one accounts for the intrinsic scatter, and the last one is the unconditioned probability density function followed by M_{500} and z.

Since the selection criterion applied in the construction of the MMF3 cosmology sample depends exclusively on the value of $q_{\rm obs}$, and given the functional form of the intrinsic and observational scatters, which will be described below, the remaining term on the right of equation (37), $P({\rm in}|\hat{\boldsymbol{n}})$, can be written using a simplified version of the hierarchical model. Indeed, we can write

$$P({\rm in}|\hat{\boldsymbol{n}}) = \int {\rm d}q_{\rm obs}{\rm d}q_{\rm t}{\rm d}M_{500}{\rm d}z \, [P({\rm in}|q_{\rm obs}, \hat{\boldsymbol{n}})P(q_{\rm obs}|q_{\rm t}, \hat{\boldsymbol{n}}) \times P(q_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}})P(M_{500}, z|\hat{\boldsymbol{n}})] \,. \quad (39)$$

Here, as we will see, the probability distributions governing the intrinsic and observational scatters on the q variables, $P(q_{\rm t}|M_{500}, z, \hat{\boldsymbol{n}})$ and $P(q_{\rm obs}|q_{\rm t}, \hat{\boldsymbol{n}})$, respectively, can be readily obtained from those of the full hierarchical model by marginalizing over the corresponding p and z variables. Also, as in equation (37), $P({\rm in}|q_{\rm obs}, \hat{\boldsymbol{n}})$ is a step function with the step located at $q_{\rm obs} = 6$.

In the following, we describe how we compute the mean signal-to-noise at given M_{500} and z for the SZ, CMB lensing, and redshift measurements, and the specific models we adopt for the intrinsic scatter, the observational scatter, and for the unconditioned probability density function of M_{500} and z.

6.1.1 *Mean quantities: $\bar{q}$, $\bar{p}$, and $\bar{z}$*

Let us consider a cluster with some given values of M_{500} and z, and at sky location $\hat{\boldsymbol{n}}$. As stated above, we assume that these three variables specify some mean values of the SZ signal-to-noise, the CMB lensing mass signal-to-noise, and the redshift of each cluster, $\bar{q}\,(M_{500}, z, \hat{\boldsymbol{n}})$, $\bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}})$, and $\bar{z}\,(M_{500}, z, \hat{\boldsymbol{n}})$.

First, let us consider the mean CMB lensing mass signal-to-noise. As detailed in Section 4, for each cluster we filter the noisy reconstruction of the lensing convergence around its location with a truncated NFW profile in order to obtain an estimate of its lensing mass and the signal-to-noise on this.

In our hierarchical model, we assume that clusters of some M_{500} and z have, on averaging over cluster shape and correlated and uncorrelated large-scale structure, projected mass distributions and so lensing convergences described by the truncated NFW profile specified in Section 4 at the true redshift, z, but at a biased mass. Matched filtering the reconstructed convergence with an assumed profile to estimate the mass can introduce further biases due to miscentring and profile mismatch. We assume that the net effect of all sources of bias is to give a mean lensing mass signal-to-noise for clusters of true mass M_{500}, redshift z, and sky location $\hat{\boldsymbol{n}}$ equal to that for a truncated NFW cluster at redshift z and mass $(1 - b_{\rm CMBlens})M_{500}$ filtered at this same (biased) mass scale and redshift:

$$\bar{p}\,(M_{500}, z, \hat{\boldsymbol{n}}) = \frac{(1 - b_{\rm CMBlens})M_{500}}{\sigma_{M_{500}}\,[(1 - b_{\rm CMBlens})M_{500}, z, \hat{\boldsymbol{n}}]}. \quad (40)$$

Here, $\sigma_{M_{500}}\,[(1 - b_{\rm CMBlens})M_{500}, z]$ is the matched filter noise, given by equation (26). This introduces a dependence of $\mathcal{L}_1$ on the sky location of each cluster, since the lensing reconstruction noise varies across the sky. We assume that the lensing mass bias $b_{\rm CMBlens}$ is constant across the sample. At the resolution and noise levels of *Planck*, $\bar{p}$ scales roughly linearly with $(1 - b_{\rm CMBlens})M_{500}$.

For the mean SZ signal-to-noise, $\bar{q}$, since we use the measured SZ signal-to-noise ratios $q_{\rm obs}$ as provided in the *Planck* MMF3 catalogue, we compute $\bar{q}$ at given cluster mass, redshift, and sky location in exactly the same way as the *Planck* team (see Planck 2013 Results XXIX 2014, Planck 2015 Results XXIV 2016, and Planck 2015 Results XXVII 2016). That is, we write the mean SZ signal-to-noise at M_{500}, z, and sky location $\hat{\boldsymbol{n}}$ as

$$\bar{q}(M_{500}, z, \hat{\boldsymbol{n}}) = \frac{\bar{Y}_{\rm SZ}\,[(1 - b_{\rm SZ})M_{500}, z]}{\sigma_{\rm f}\,[\theta_{500}((1 - b_{\rm SZ})M_{500}, z), \hat{\boldsymbol{n}}]}. \quad (41)$$

Here, $\sigma_{\rm f}(\theta_{500}, \hat{\boldsymbol{n}})$ is the noise of the multifrequency matched filter used to detect the clusters evaluated at the SZ angular scale of the cluster, θ_{500}, where

$$\theta_{500} = \theta_\star \left(\frac{h}{0.7}\right)^{-2/3} \left(\frac{(1 - b_{\rm SZ})M_{500}}{3 \times 10^{14}\,{\rm M}_\odot}\right)^{1/3} E^{-2/3}(z) \times \left(\frac{d_A(z)}{500\,{\rm Mpc}}\right)^{-1}. \quad (42)$$

The constant $\theta_\star = 6.997\,{\rm arcmin}$, $E(z) = H(z)/H_0$, where $H(z)$ is the Hubble parameter and the Hubble constant $H_0 = 100h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, and $d_A(z)$ is the angular diameter distance to redshift z. The SZ mass bias parameter $b_{\rm SZ}$ is introduced to account for any differences between the X-ray mass estimates that are used in the calibration of the SZ scaling relations and the true masses, as discussed in Section 1. The mean integrated Comptonization parameter in equation (41), $\bar{Y}_{\rm SZ}$, is given by the scaling relation (Planck 2015 Results XXIV 2016)

$$E^{-\beta}(z)\left(\frac{d_A^2(z)\bar{Y}_{\rm SZ}}{10^{-4}\,{\rm Mpc}^2}\right) = Y_\star \left(\frac{h}{0.7}\right)^{-2+\alpha} \left(\frac{(1 - b_{\rm SZ})M_{500}}{6 \times 10^{14}\,{\rm M}_\odot}\right)^{\alpha}, \quad (43)$$

where $Y_\star$, α, and β are parameters that need to be calibrated. We compute $\sigma_{\rm f}(\theta_{500}, \hat{\boldsymbol{n}})$ by interpolating over the tabulated values used in the likelihood analyses in Planck 2015 Results XXIV (2016); their likelihood code is freely available as part of the COSMOMC package.[7] As for the CMB lensing case, $\sigma_{\rm f}(\theta_{500}, \hat{\boldsymbol{n}})$ introduces an additional dependence of $\mathcal{L}_1$ on the sky location of each cluster.

Finally, as in Planck 2015 Results XXIV (2016), we take the mean redshift at M_{500}, z, and $\hat{\boldsymbol{n}}$ to be simply z, i.e.

$$\bar{z}\,(M_{500}, z, \hat{\boldsymbol{n}}) = z. \quad (44)$$

6.1.2 *Intrinsic scatter*

The first layer in our hierarchical model is what we refer to as intrinsic scatter. For the SZ and CMB lensing measurements, by intrinsic scatter we mean all sources of statistical scatter on the signal-to-noise measurements that are not incorporated in the noise budget of the matched filtering process. For the CMB lensing measurements, these include deviations of the cluster convergence profile from its assumed mean profile at M_{500} and z, deviations which often exhibit a triaxial nature, and contributions to the

[7] https://cosmologist.info/cosmomc/

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

observed convergence profile from correlated large-scale structure. For the SZ measurements, intrinsic scatter includes deviations from the mean Comptonization profile, also often of a triaxial nature. Finally, as in Planck 2015 Results XXIV (2016), we assume that there is no intrinsic scatter on the redshift measurements.

In Planck 2015 Results XXIV (2016), a lognormal model for the intrinsic scatter of the SZ signal-to-noise was adopted. Intrinsic scatter on lensing-derived cluster masses has also been shown to be approximately lognormal (see e.g. Becker & Kravtsov 2011), and the SZ and CMB lensing measurements are also expected to be intrinsically correlated, since they are both obtained from quantities integrated along the line of sight (see e.g. Shirasaki et al. 2016). We therefore adopt the following probability distribution for the intrinsic scatter:

$$P(q_{\rm t},\, p_{\rm t},\, z_{\rm t}|M_{500},\, z,\, \hat{\boldsymbol{n}}) = \frac{1}{q_{\rm t}p_{\rm t}}\, P(\ln q_{\rm t},\, \ln p_{\rm t}|M_{500},\, z,\, \hat{\boldsymbol{n}})\delta(z_{\rm t} - z), \tag{45}$$

where the delta function reflects the absence of intrinsic scatter on the redshift measurements, and where $P(\ln q_{\rm t},\, \ln p_{\rm t}|M_{500},\, z,\, \hat{\boldsymbol{n}})$ is a bivariate Gaussian with means $\ln \bar{q}(M_{500},\, z,\, \hat{\boldsymbol{n}})$ and $\ln \bar{p}(M_{500},\, z,\, \hat{\boldsymbol{n}})$, respectively, and a covariance matrix that we parametrize with its two associated standard deviations, $\sigma_{\rm SZ}$ and $\sigma_{\rm CMBlens}$, respectively, and a correlation coefficient, $r_{\rm CMBlens\text{-}SZ}$. We take $\sigma_{\rm SZ}$, $\sigma_{\rm CMBlens}$, and $r_{\rm CMBlens\text{-}SZ}$ to be constants.

6.1.3 *Observational scatter*

The second layer in our hierarchical model is what we call observational scatter. For the CMB lensing and SZ measurements, it accounts for the scatter caused by what is treated as observational noise in the matched filtering processes that yield the values of $q_{\rm obs}$ and $p_{\rm obs}$ of each cluster. For example, for the CMB lensing masses, observational scatter arises from the reconstruction noise and large-scale structure that is uncorrelated with the cluster. We assume that the observational scatter is Gaussian for both SZ and lensing measurements. This assumption was also made for the SZ measurements in Planck 2015 Results XXIV (2016), and should be also a good approximation for the CMB lensing measurements, since they are obtained by matched filtering a reconstructed convergence map, which involves a sum over many reconstructed modes. We also assume that there is no correlation between the observational scatter of the CMB lensing measurements and that of the SZ measurements. Finally, as in Planck 2015 Results XXIV (2016), we assume that there is no observational scatter on $z_{\rm t}$.

We can therefore write the probability distribution function accounting for the observational scatter as

$$P(q_{\rm obs},\, p_{\rm obs},\, z_{\rm obs}|q_{\rm t},\, p_{\rm t},\, z_{\rm t},\, \hat{\boldsymbol{n}}) = P(q_{\rm obs}|q_{\rm t})P(p_{\rm obs}|p_{\rm t})\delta(z_{\rm obs} - z_{\rm t}), \tag{46}$$

where $P(q_{\rm obs}|q_{\rm t})$ and $P(p_{\rm obs}|p_{\rm t})$ are Gaussians with means $q_{\rm t}$ and $p_{\rm t}$, respectively, and unit standard deviations. The delta function in equation (46) accounts for the absence of observational scatter on the redshift measurements.

6.1.4 *Unconditioned distribution of* M_{500} *and* z

The remaining factor of the integrand in equation (38) is $P(M_{500},\, z|\hat{\boldsymbol{n}})$. This is the probability density function followed by M_{500} and z with no conditions imposed, other than the sky location $\hat{\boldsymbol{n}}$. It does not depend on $\hat{\boldsymbol{n}}$ and is simply proportional to the halo mass function, $d^2N/({\rm d}V{\rm d}M_{500})$, times the volume element, ${\rm d}V/{\rm d}z$:

$$P(M_{500},\, z|\hat{\boldsymbol{n}}) \propto \frac{d^2N}{{\rm d}V\,{\rm dM}_{500}}(M_{500},\, z)\frac{{\rm d}V}{{\rm d}z}(z). \tag{47}$$

As in Planck 2015 Results XXIV (2016), we use the halo mass function from Tinker et al. (2008).

6.1.5 *Probability of inclusion, or renormalization*

Finally, let us revisit the probability for a cluster to be included in the sample, $P({\rm in}|\hat{\boldsymbol{n}})$, as written down in equation (39). From Sections 6.1.2 and 6.1.3, it should now be apparent that such a decomposition is possible, and that $P(q_{\rm obs}|q_{\rm t},\, \hat{\boldsymbol{n}})$ is just $P(q_{\rm obs}|q_{\rm t})$ in equation (46), and $P(q_{\rm t}|M_{500},\, z,\, \hat{\boldsymbol{n}})$ is just $P(q_{\rm t},\, p_{\rm t},\, z_{\rm t}|M_{500},\, z,\, \hat{\boldsymbol{n}})$ in equation (45), marginalized over $p_{\rm t}$ and $z_{\rm t}$.

6.2 $\mathcal{L}_2$: sky location likelihood

The second factor of our total likelihood, $\mathcal{L}_2$ (see equation 35), is $P(\underline{\hat{\boldsymbol{n}}}|N)$, the probability that, given N clusters have been included in our sample (the MMF3 cosmology sample), their sky locations are $\underline{\hat{\boldsymbol{n}}}$. Since we assume that clusters are statistically independent from each other, it is simply given by

$$P(\underline{\hat{\boldsymbol{n}}}|N) = N!\prod_{i=1}^{N}\frac{1}{\bar{N}}\frac{{\rm d}\bar{N}}{{\rm d}\Omega}(\hat{\boldsymbol{n}}_i), \tag{48}$$

where ${\rm d}\bar{N}/{\rm d}\Omega$ is the mean number of clusters in the sample per solid angle, and $\bar{N}$ is the mean total number of clusters in the sample, i.e.

$$\bar{N} = \int {\rm d}\hat{\boldsymbol{n}}\frac{{\rm d}\bar{N}}{{\rm d}\Omega}(\hat{\boldsymbol{n}}), \tag{49}$$

where the integration is performed across the region of the sky left unmasked in the construction of the MMF3 cosmology sample. We include the $N!$ factor in equation (48) to reflect the fact that the ordering of the elements in $\underline{\hat{\boldsymbol{n}}}$ does not matter.[8] The mean number of clusters in the sample per solid angle can be written as

$$\frac{{\rm d}\bar{N}}{{\rm d}\Omega}(\hat{\boldsymbol{n}}) = \int {\rm d}q_{\rm obs}{\rm d}M_{500}{\rm d}z\; P\left(q_{\rm obs}|M_{500},\, z,\, \hat{\boldsymbol{n}}\right)\frac{d^2N}{{\rm d}V\,{\rm d}M_{500}}\frac{d^2V}{{\rm d}z{\rm d}\Omega}. \tag{50}$$

Here, $q_{\rm obs}$ is integrated from 6 to infinity (to meet the inclusion selection), $d^2N/({\rm d}V{\rm d}M_{500})$ is the halo mass function, and $d^2V/({\rm d}z{\rm d}\Omega)$ is the volume element (see Section 6.1.4). The probability density

$$P\left(q_{\rm obs}|M_{500},\, z,\, \hat{\boldsymbol{n}}\right) = \int {\rm d}q_{\rm t}\; P\left(q_{\rm obs}|q_{\rm t}\right)\, P\left(q_{\rm t}|M_{500},\, z,\, \hat{\boldsymbol{n}}\right), \tag{51}$$

which involves the (marginalized) probability density functions for the observational and intrinsic scatter introduced in Section 6.1.

6.3 $\mathcal{L}_3$: Poisson likelihood for the total number of clusters

Finally, the third factor in our total likelihood, $\mathcal{L}_3$ (see equation 35), is $P(N)$, the probability of including a total of N clusters in our sample (the MMF3 cosmology sample). Since we assume that clusters are statistically independent from each other, $P(N)$ is simply a Poisson distribution with expected value $\bar{N}$, where $\bar{N}$ is given by equation (49).

[8]The integration measure over the data for N clusters is then $(1/N!)\prod_{i=1}^{N}{\rm d}D_i{\rm d}\hat{\boldsymbol{n}}_i$.

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

Table 2. Summary of the cluster parameters of our model.

Parameter	Definition
$1-b_{\rm SZ}$	SZ mass bias
$1-b_{\rm CMBlens}$	CMB lensing mass bias
$\sigma_{\rm SZ}$	SZ intrinsic scatter
$\sigma_{\rm CMBlens}$	CMB lensing intrinsic scatter
$r_{\rm CMBlens\text{-}SZ}$	SZ–CMB lensing intrinsic correlation coefficient
$Y_\star$	Scaling relation normalization
α	Scaling relation mass exponent
β	Scaling relation $E(z)$ exponent

Table 3. Means and standard deviations of the Gaussian priors adopted in our likelihood analysis of the real cluster data. Parameters not listed have broad flat priors.

Parameter	Mean	Standard deviation
$1-b_{\rm CMBlens}$	0.93	0.05
$\sigma_{\rm CMBlens}$	0.20	0.05
$r_{\rm CMBlens\text{-}SZ}$	0.83	0.10
α	1.79	0.08
$\log Y_\star$	−0.19	0.02
$100\theta_{\rm MC}$	1.040 93	0.000 30

6.4 Model parameters and priors

6.4.1 *Cluster parameters*

For convenience, we summarize below the set of cluster model parameters on which our likelihood, $\mathcal{L}$, depends:

$$\boldsymbol{p}_{\rm cl} = \{1-b_{\rm SZ}, 1-b_{\rm CMBlens}, \sigma_{\rm SZ}, \sigma_{\rm CMBlens}, r_{\rm CMBlens-SZ}, Y_\star, \alpha, \beta\}\,. \tag{52}$$

A list of these parameters can also be found in Table 2, along with their definitions.

We adopt priors on these cluster parameters in our likelihood analysis as follows. For the SZ parameters, we fix $\beta = 0.66$ (i.e. self-similar evolution), following the main analysis of Planck 2015 Results XXIV (2016). Although the redshift evolution of the SZ scaling relation is not well constrained observationally, Planck 2015 Results XXIV (2016) showed that the main impact of allowing β to vary within a broad Gaussian prior, $\beta = 0.66 \pm 0.50$, was to broaden constraints along the $\Omega_{\rm m}$–σ_8 degeneracy direction, rather than in the perpendicular direction that responds to the absolute cluster mass scale (see their appendix A.3). We expect similar effects in our analysis, with only a small impact on constraints on the mass bias parameter $1-b_{\rm SZ}$ of primary interest. We also fix the intrinsic scatter in the SZ signal-to-noise to $\sigma_{\rm SZ} = 0.173$, the central value of the empirically derived prior adopted in Planck 2015 Results XXIV (2016), for reasons of computational efficiency.[9] For the remaining parameters in the SZ scaling relation, α and $\log Y_\star$, we adopt the Gaussian priors reported in Table 3, following Planck 2015 Results XXIV (2016). Finally, for the SZ mass bias we adopt a flat positivity prior $1-b_{\rm SZ} \geq 0$.

For the CMB lensing mass parameters, we impose Gaussian priors on $1-b_{\rm CMBlens}$, $\sigma_{\rm CMBlens}$, and $r_{\rm CMBlens\text{-}SZ}$ with means and standard deviations given in Table 3. In Zubeldia & Challinor (in prep.), we study the intrinsic bias and scatter of our CMB lensing observable using N-body simulations [as was done, e.g. in Becker & Kravtsov (2011) for galaxy weak lensing]. Our prior on $1-b_{\rm CMBlens}$ is taken to be consistent with the values we find in Zubeldia & Challinor (in prep.) in addition to the bias expected from miscentring. Similarly, our prior on $\sigma_{\rm CMBlens}$ is taken to be consistent with the values we report in Zubeldia & Challinor (in prep.). Finally, our prior for the correlation between the intrinsic scatter on SZ and CMB lensing mass signal-to-noise ratios, $r_{\rm CMBlens\text{-}SZ}$, is taken to be consistent with the values of the SZ–lensing mass intrinsic correlation reported in Shirasaki et al. (2016).

[9]The constraints on this parameter from Planck 2015 Results XXIV (2016) for the CCCP and WtG calibrations are prior-driven, the central values being a fraction of a σ away from the prior central value. In addition, we find from their corresponding MCMC chains that $\sigma_{\rm SZ}$ has small correlation with our main parameter of interest, $\sigma_8(\Omega_{\rm m}/0.33)^{0.25}$, with a correlation coefficient of about 0.2 for both calibrations (see Section 8 for the motivation behind this parametrization). We therefore conclude that letting this parameter vary and imposing on it the same prior as in Planck 2015 Results XXIV (2016) would have little impact on our results.

6.4.2 *Cosmological parameters*

We also stress the dependence of our likelihood on cosmological parameters, so we can jointly constrain cosmology and cluster physics. The strongest dependences are on $\Omega_{\rm m}$, σ_8, and H_0 through the halo mass function, the volume element, and the mean values of the SZ and CMB lensing signal-to-noise ratios at given M_{500}, z, and sky location. There are only weak dependences on the baryon density, $\Omega_{\rm b}h^2$, and spectral index of the primordial curvature perturbations, $n_{\rm s}$, through the shape of the matter power spectrum around cluster scales that enters the halo mass function. In our analysis, we fix these two parameters to the values $\Omega_{\rm b}h^2 = 0.0223$ and $n_{\rm s} = 0.9667$ determined by *Planck* (Planck 2015 Results XIII 2016). We also fix $\Omega_K = 0$ and the summed neutrino mass to the minimal value, $\sum m_\nu = 0.06\,{\rm eV}$.

We impose broad, flat priors on $\Omega_{\rm m}$, σ_8, and h. However, since h is poorly constrained by the cluster data, we further impose a Gaussian prior on the CMB acoustic scale parameter, $\theta_{\rm MC}$ (see Table 3). This parameter is an analytic approximation to the ratio of the sound horizon at CMB last scattering to the angular diameter distance – see Kosowsky, Milosavljevic & Jimenez (2002), where $\theta_{\rm MC}$ is their $\mathcal{A}$ – which at fixed baryon density depends only on $\Omega_{\rm m}$ and H_0 in flat ΛCDM models. The acoustic scale parameter is very well measured by *Planck* from the acoustic peak locations, and is almost model independent. Our prior is from the analysis reported in Planck 2015 Results XIII (2016). We stress that this prior is *geometric* and uses no information on the amplitude of the fluctuations from the CMB power spectrum.

7 LIKELIHOOD VALIDATION

We test the implementation of our likelihood by analysing a set of simulated data, $\{N, \underline{\hat{\boldsymbol{n}}}, \underline{D}\}$. We produce these data following our model assumptions at fixed values of cosmological and cluster model parameters, and then explore the corresponding posterior distribution with Markov chain Monte Carlo (MCMC).

7.1 Simulated data

In order to generate our set of simulated data, we assume a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.315$, $\sigma_8 = 0.811$, $h = 0.674$,

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

Table 4. Marginalized constraints (mean and 68 per cent confidence limits) on the parameters that we allow to vary in our analysis of the simulated data, along with their corresponding input values.

Parameter	Constraints (68 per cent uncert.)	Input value
$\Omega_{\rm m}$	0.31 ± 0.02	0.315
σ_8	0.77 ± 0.04	0.811
h	0.67 ± 0.02	0.674
$1 - b_{\rm SZ}$	0.72 ± 0.1	0.62
$1 - b_{\rm CMBlens}$	0.91 ± 0.05	0.92
$\sigma_{\rm CMBlens}$	0.23 ± 0.05	0.22
$r_{\rm CMBlens\text{-}SZ}$	0.75 ± 0.09	0.77
α	1.84 ± 0.06	1.79
$Y_\star$	0.65 ± 0.03	0.646

and $\Sigma m_\nu = 0.06\,$eV. We set the cluster model parameters close to the mean values of the Gaussian priors in Table 3 along with $\beta = 0.66$, $\sigma_{\rm SZ} = 0.173$ (as fixed by our priors), and $1 - b_{\rm SZ} = 0.62$. These fiducial values are listed in the final column of Table 4.

The first step in producing our simulated data set is to obtain N, the total number of clusters that will be in our sample. We obtain it by drawing one sample from a Poisson distribution with mean value $\bar{N}$ given by equation (49), evaluated at our assumed values of cosmological and model parameters. We obtain a total of $N = 416$ clusters.

We then generate the set of sky locations of these N clusters, $\hat{\underline{\boldsymbol{n}}}$. To do this, in order to produce a catalogue that is statistically as close as possible to the MMF3 catalogue we use the SZ-matched filter noise estimates across the sky produced by the Planck collaboration and that were used in the construction of the MMF3 catalogue. In these, the SZ-matched filter noise is estimated for a set of filter angular scales in a set of patches that cover the whole sky. For each patch, interpolating over these tabulated values and using equation (49), where we restrict the integration over the sky to the extent of the patch, we compute the expected number of clusters falling within the patch. For patch i, this is $\bar{N}_i$, and the probability for a detected cluster to be in that patch is simply $P_i = \bar{N}_i/\bar{N}$. We then assign a number of clusters to each patch by obtaining N samples from a multinomial distribution with probabilities P_i.

Finally, we produce a mass data point, D_i, for each of the N clusters. We generate each of them in the following way. First, we obtain values of M_{500} and z by drawing a sample from $P(M_{500}, z|\hat{\boldsymbol{n}})$ (see equation 47) using rejection sampling. We then produce values of $q_{\rm t}$, $p_{\rm t}$, and $z_{\rm t}$ by drawing one sample from equation (45), with the values of M_{500} and z obtained in the previous step and $\hat{\boldsymbol{n}}$ as conditioning information. In this step, for each cluster, $\bar{q}$ is computed with the SZ-matched filter noise estimate of the sky patch in which the cluster is located. On the other hand, $\bar{p}$ is computed with the empirical average CMB lensing matched filter noise of all the clusters in the MMF3 cosmology sample falling within the cluster's sky patch. This is for the sake of simplicity, since in our CMB lensing mass pipeline we have produced such noise estimates only at the MMF3 cluster locations. To avoid inconsistencies in the analysis of the *simulated* data, we use these averaged CMB lensing noise estimates in the likelihood in this case (but not in the analysis of the real cluster sample). Finally, we generate values of $q_{\rm obs}$, $p_{\rm obs}$, and $z_{\rm obs}$ by drawing one sample from equation (46), with the values of $q_{\rm t}$, $p_{\rm t}$, and $z_{\rm t}$ obtained in the previous step as conditioning information. If $q_{\rm obs} > 6$, we set $(q_{\rm obs}, p_{\rm obs}, z_{\rm obs})$ as the mass data point corresponding to the cluster at sky location $\hat{\boldsymbol{n}}$. If not, we repeat the process again until $q_{\rm obs} > 6$.

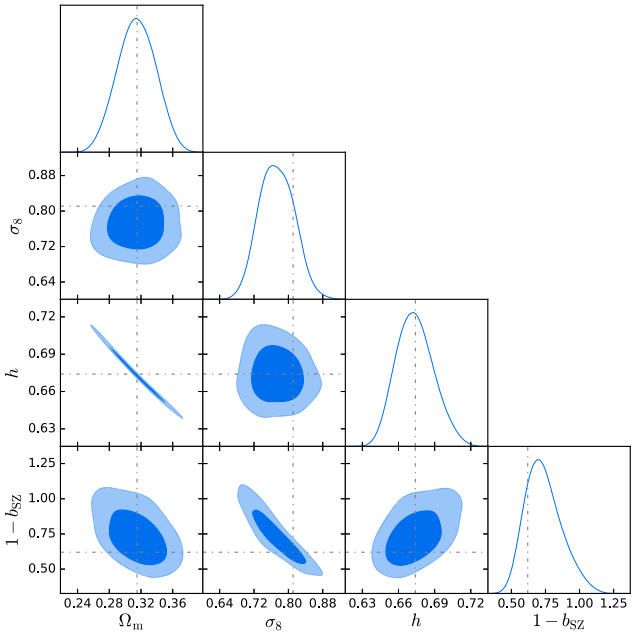


Figure 1. Constraints (68 and 95 per cent confidence regions) on $\Omega_{\rm m}$, σ_8, h, and $1 - b_{\rm SZ}$ for our simulated cluster sample. One-dimensional, marginalized posterior distributions are also shown. The input values of the parameters are shown as dash–dotted lines.

7.2 Parameter constraints

In order to validate the implementation of our likelihood, $\mathcal{L}$, we apply it to the simulated data set, described in Section 7.1, and explore the corresponding posterior distribution with an MCMC. We allow the same set of cosmological and cluster model parameters to vary as in our analysis of the real data, $\boldsymbol{p} = \{\Omega_{\rm m}, \sigma_8, h, b_{\rm SZ}, b_{\rm CMBlens}, \sigma_{\rm CMBlens}, r_{\rm CMBlens-SZ}, \alpha, Y_\star\}$, and impose similar priors on them. For those parameters on which we impose Gaussian priors when analysing the real data (see Table 3), we retain the widths of the priors but, where necessary, recentre the means on the input values used to generate the simulations. As for the real data, we impose broad, flat priors on $\Omega_{\rm m}$, σ_8, h, and $1 - b_{\rm SZ}$, while $\sigma_{\rm SZ}$ and β are fixed to their input values.

We explore our posterior using the `emcee` package, which performs affine-invariant MCMC sampling.[10] The two-dimensional marginalized constraints that we obtain for $\Omega_{\rm m}$, σ_8, h, and $1 - b_{\rm SZ}$ are shown in Fig. 1, and the one-dimensional marginalized constraints for all the parameters varied in the analysis are given in Table 4. It can be seen that these constraints are consistent with the input parameter values used to construct the simulated data. The two-dimensional marginalized constraints on the remaining parameters are not shown, but are similarly consistent with their true values. This validates our likelihood implementation and, as long as our Bayesian model remains a good description of the real data, our analysis. Of course, as our validation is limited to one simulation, we cannot test for systematic biases at a level below the statistical errors expected from our real data.

[10] http://dfm.io/emcee/current/

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

8 RESULTS AND DISCUSSION

We explore the posterior distribution for the real cluster data with an MCMC using the `emcee` package, as for our tests on simulated data. The two-dimensional marginalized constraints on all the parameters we allow to vary are shown in Fig. 2, and the two-dimensional marginalized constraints on Ω_m, σ_8, and $1-b_{SZ}$ are shown in Fig. 3, along with the corresponding constraints on the cosmological parameters from the *Planck* 2018 measurements of the CMB angular power spectra and CMB lensing (the TT,TE,EE+lowE+lensing likelihood; Planck 2018 Results VI 2018). In addition, the one-dimensional marginalized constraints on $\sigma_8(\Omega_m/0.33)^{0.25}$ and on our cluster parameters are given in Table 5.

We note that the constraints on Ω_m, σ_8, and h are dependent on our choice of priors on α (Gaussian prior) and β (delta function prior). As argued in Section 6.4, allowing β to vary would widen the constraints in the Ω_m–σ_8 plane along the long degeneracy axis. Through the very strong degeneracy between Ω_m and h due to our prior on θ_{MC}, the constraints on h would also widen accordingly. Planck 2015 Results XXIV (2016) find that leaving α free also widens the constraints along the long degeneracy axis in the Ω_m–σ_8 plane; we expect a similar effect in our analysis, with constraints on h also widening accordingly. We therefore decide to quote our cosmological constraints in terms of $\sigma_8(\Omega_m/0.33)^{0.25}$, a parameter that runs perpendicular to the long degeneracy axis in the Ω_m–σ_8 plane, and which therefore is more immune to the choice of priors on α and β. Purely for reference, we find $\Omega_m = 0.33 \pm 0.02$, $\sigma_8 = 0.76 \pm 0.04$, and $h = 0.66 \pm 0.01$, stressing that they are dependent on the choice of priors on α and β.

As it can be seen in Table 5, our constraint on $\sigma_8(\Omega_m/0.33)^{0.25}$ from the *Planck* MMF3 cosmology sample of clusters with CMB lensing calibration of the cluster masses is consistent with that derived in the ΛCDM model from the *Planck* 2018 CMB angular power spectra (and lensing), $\sigma_8(\Omega_m/0.33)^{0.25} = 0.8020 \pm 0.0085$. Consistency can also be seen in Fig. 3. We therefore find no evidence for tension between the abundance of galaxy clusters measured at lower redshifts and the predictions for the ΛCDM model with parameters (mostly) calibrated at CMB decoupling. An alternative view of this consistency is to compare our constraint on the SZ mass bias parameter, $1-b_{SZ} = 0.71 \pm 0.10$ (see Table 6), with that obtained in ΛCDM from joint analysis of the CMB angular power spectra and cluster counts without further mass calibration. The latter was reported as $1-b_{SZ} = 0.62 \pm 0.04$ by Planck 2018 Results VI (2018), showing good consistency.

Our CMB-lensing constraints on the SZ mass bias are also consistent with those from lensing of background galaxies by WtGs (von der Linden et al. 2014) and from the CCCP (Hoekstra et al. 2015); see Table 6. These mass calibrations were used as priors in the analysis of Planck 2015 Results XXIV (2016). They are also consistent with the weak lensing constraints of subsamples of *Planck* clusters by Penna-Lima et al. (2017) and Sereno et al. (2017), with the constraint from the combined analysis of *Planck* clusters and the tSZ power spectrum and bispectrum of Hurier & Lacasa (2017), and with the HSC galaxy weak lensing mass calibration of eight ACTPol clusters of Miyatake et al. (2018) and of five *Planck* clusters of Medezinski et al. (2018) (see also Table 6). We note also that our constraint on $1-b_{SZ}$ is of comparable precision to current constraints from galaxy lensing. While the mass constraints on individual low-redshift clusters are weaker from current CMB lensing data than from galaxy lensing, the former approach can be applied to the full cluster sample.

Our constraint on the SZ mass bias, along with those from galaxy lensing, is in mild disagreement with that obtained from CMB lensing in Planck 2015 Results XXIV (2016). Reasons for this discrepancy were discussed in Section 5. We note that our constraint is, however, of similar precision: we find $1/(1-b_{SZ}) = 1.41 \pm 0.20$ from our samples of b_{SZ}.

There are some degeneracies between parameters that are worth commenting on. The strongest degeneracy is that between Ω_m and h (see Figs 2 and 3). This degeneracy is due to the prior we impose on θ_{MC}, which, at fixed baryon density, constrains $\Omega_m h^3$ to be approximately constant. More interesting, regarding the cluster counts, is the degeneracy between $1-b_{SZ}$ and σ_8. As expected, there is a negative correlation between these parameters, which arises mostly through the dependence of our likelihood on the halo mass function. The physical interpretation of this degeneracy is straightforward: a given set of cluster SZ measurements can be explained with some given values of $1-b_{SZ}$ and σ_8, or with a smaller value of $1-b_{SZ}$ (making the mass larger at a given SZ cluster signal) and a larger value of σ_8 (to enhance the number of more massive clusters to match the observed number). Finally, there is also an anticorrelation between Ω_m and σ_8, arising primarily from the mass function: the increase in the number density of massive clusters with increasing σ_8 can be offset by the overall reduction in the number density of clusters with decreasing matter density.

Finally, we note that our CMB lensing signal-to-noise p_{obs} measurements are interesting per se in that they represent a significant detection of cluster CMB lensing with the *Planck* data. Since the observational scatter is much larger than the intrinsic scatter and than the scatter associated with the spread of M_{500} and z across the sample, our p_{obs} measurements roughly follow a Gaussian distribution with some mean and standard deviation of unity. If no cluster lensing signal were present, they would follow a Gaussian distribution with zero mean and unit standard deviation. We find $\langle p_{obs} \rangle = 0.234 \pm 0.052$, where angular brackets denote averaging over the cluster sample, thus detecting the CMB cluster lensing signal at 4.5σ significance. For comparison, Baxter et al. (2015) detect the CMB lensing cluster signal of 513 SZ-selected SPT clusters with SPT data at 3.1σ significance, Baxter et al. (2017) detect that of 3697 high-redshift DES clusters with SPT data at 8.1σ significance, and Raghunathan et al. (2019) detect that of 4003 and 1741 DES clusters with SPTpol data at 8.7σ and 6.7σ significance, respectively. In addition, Madhavacheril et al. (2015) detect the CMB lensing signal of 12 000 optically selected CMASS galaxies using ACT data at 3.2σ significance.

9 CONCLUSION

In this paper, we have presented constraints on the ΛCDM-model parameters Ω_m, σ_8, and H_0, and a CMB lensing calibration of the SZ mass bias, $1-b_{SZ}$, obtained from 439 SZ-selected galaxy clusters from the *Planck* MMF3 cosmology sample. Our analysis revisits the *Planck* SZ count analysis with CMB lensing mass calibration presented in Planck 2015 Results XXIV (2016). The analysis there used cluster counts in the SZ signal-to-noise and redshift plane in order to constrain cosmological (and cluster model) parameters, imposing a prior on $1-b_{SZ}$ derived from a CMB lensing mass calibration also presented in Planck 2015 Results XXIV (2016). Such calibration found no evidence for a SZ mass bias, in contrast to galaxy weak lensing calibrations on subsets of the cluster sample (e.g. von der Linden et al. 2014; Hoekstra et al. 2015) that favour $1-b_{SZ} < 1$ at more than 2σ. Although the statistical significance

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

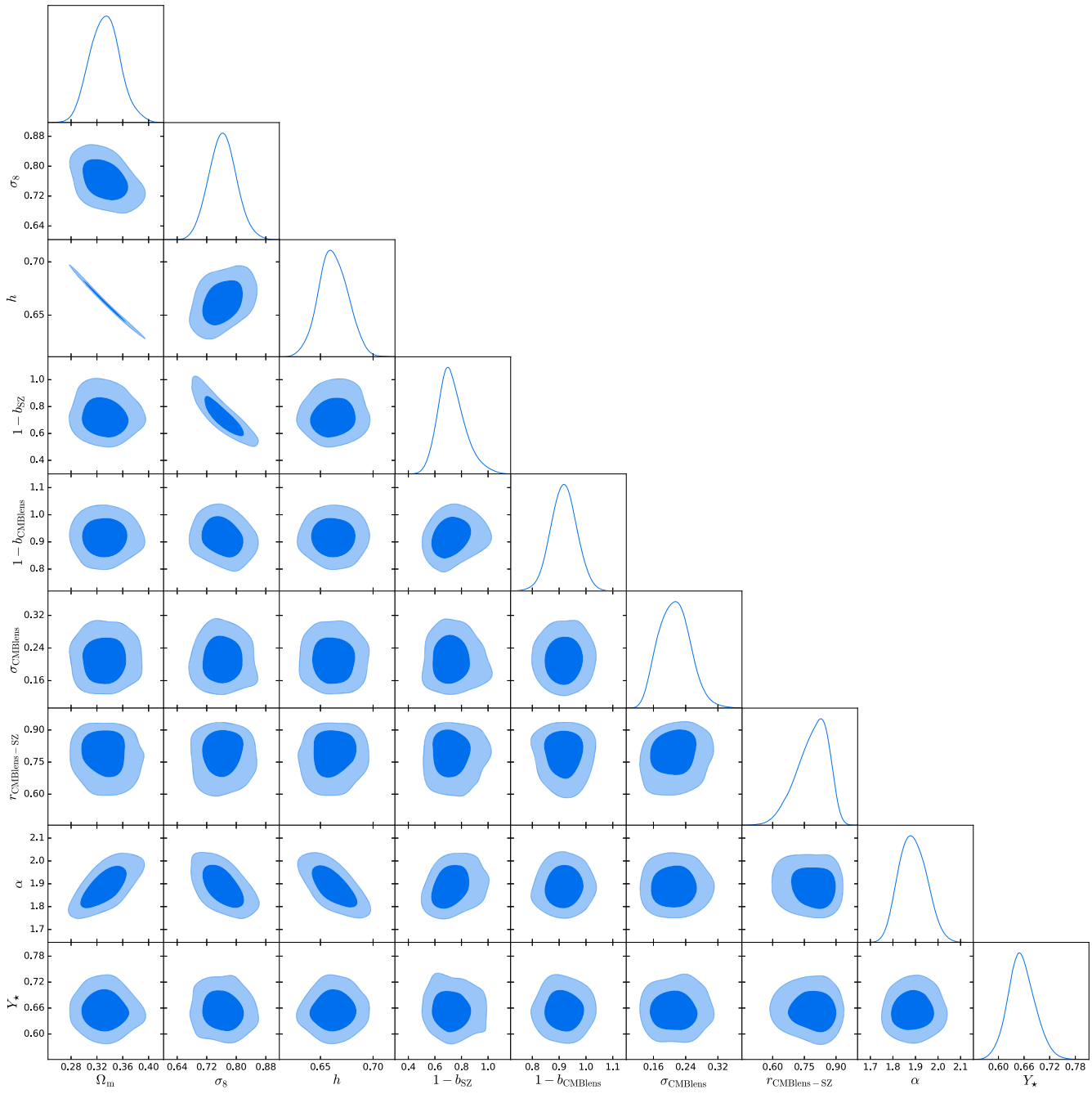


Figure 2. Constraints (68 and 95 per cent confidence regions) on the cosmological and cluster model parameters that we allow to vary in our analysis of the *Planck* MMF3 cosmology sample of clusters with CMB lensing mass calibration and a prior on θ_{MC}. One-dimensional, marginalized posterior distributions are also shown.

of the difference in mass bias measured from CMB and galaxy lensing is relatively weak, given the large measurement errors for CMB lensing, adopting the CMB lensing calibration of Planck 2015 Results XXIV (2016) exacerbates tension between cluster constraints in the σ_8–Ω_m plane and those derived from the primary CMB anisotropies in ΛCDM.

We argued in Section 5 that there are several effects that may have led to bias in this previous mass calibration from CMB lensing. We have remeasured the cluster masses via CMB lensing, and included the signal-to-noises of these and the SZ measurements, along with the cluster redshifts, in a Bayesian analysis that naturally takes account of all significant effects that likely biased the analysis in Planck 2015 Results XXIV (2016). This approach allows us to constrain jointly the cosmological parameters and the SZ mass bias (and other cluster model parameters) in an unbiased way, as demonstrated through simulated data in Section 7.

With our likelihood, and including priors on some of the cluster model parameters informed by results from numerical simulations (but, notably, with no prior on the SZ mass bias $1-b_{SZ}$), we obtain constraints on Ω_m, σ_8, and H_0 that are consistent with the constraints in the ΛCDM model derived from the *Planck* measurements of the CMB power spectra (see Figs 2 and 3, and Table 5). We mea-

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

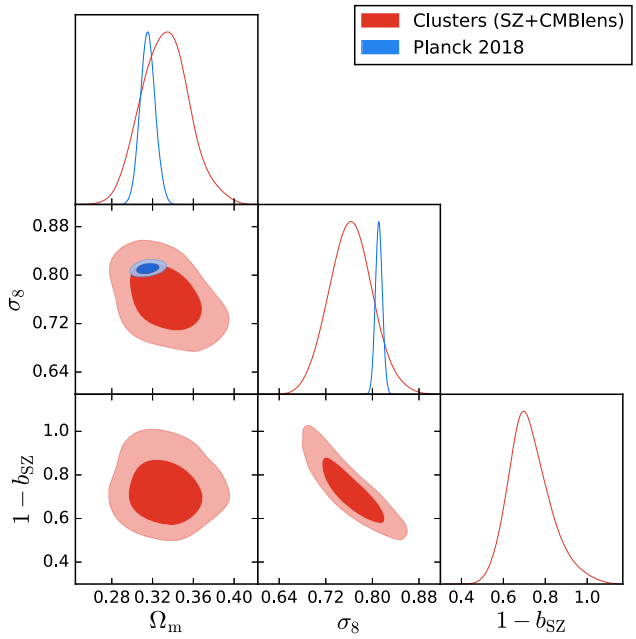


Figure 3. Constraints on the cosmological parameters $\Omega_{\rm m}$ and σ_8, and on the SZ mass bias, $1-b_{\rm SZ}$, from our analysis of the *Planck* MMF3 cosmology sample of clusters (red). We stress that the joint constraints on $\Omega_{\rm m}$ and σ_8 along their long degeneracy axis are strongly dependent on the priors on α and β, whereas the perpendicular direction (the approximate combination $\sigma_8\Omega_{\rm m}^{0.25}$) is more immune to these priors. The one-dimensional constraints on $\Omega_{\rm m}$ and σ_8 are similarly strongly affected. Also shown (in blue) are the constraints from the *Planck* 2018 measurements of the CMB angular power spectra and CMB lensing, using the data combination TT,TE,EE+lowE+lensing (Planck 2018 Results VI 2018). Both constraints assume a flat ΛCDM cosmology.

Table 5. Marginalized constraints (mean and 68 per cent confidence limits) on $\sigma_8(\Omega_{\rm m}/0.33)^{0.25}$ and on the cluster parameters that we allow to vary in our analysis of the *Planck* MMF3 cosmology sample of clusters.

Parameter	Constraints (68 per cent uncert.)
$\sigma_8(\Omega_{\rm m}/0.33)^{0.25}$	0.765 ± 0.035
$1-b_{\rm SZ}$	0.71 ± 0.1
$1-b_{\rm CMBlens}$	0.92 ± 0.05
$\sigma_{\rm CMBlens}$	0.21 ± 0.04
$r_{\rm CMBlens\text{-}SZ}$	$0.80^{+0.06}_{-0.09}$
α	1.89 ± 0.07
$Y_\star$	0.65 ± 0.03

sure a significant SZ mass bias, $1-b_{\rm SZ}=0.71\pm0.10$, consistent with measurements from galaxy weak lensing. We therefore find no evidence in our analysis of tensions between ΛCDM model parameters derived from the *Planck* cluster sample and the primary CMB anisotropies.

Our work is further evidence of the growing power of CMB lensing to calibrate the overall cluster mass scale in SZ count analyses. It will be a particularly powerful approach for future high-resolution CMB experiments, such as CMB-S4, which should detect around 10^5 clusters through their SZ signatures (Abazajian et al. 2016). Indeed, CMB lensing allows one to estimate cluster masses from CMB data alone (plus, in principle, redshift measurements of each cluster), which implies that the whole cluster sample can be used in the calibration analysis, rather than just a small subsample,

Table 6. Constraints (mean and 68 per cent confidence region) on $1-b_{\rm SZ}$ from the data sets indicated. These include weak lensing of background galaxies for subsets of *Planck* clusters from Weighting the Giants (WtG; von der Linden et al. 2014), the CCCP (Hoekstra et al. 2015), Sereno et al. (2017, the 'cosmological subsample'), Penna-Lima et al. (2017), and Medezinski et al. (2018). CMB-derived constraints are for the CMB lensing calibration presented in Planck 2015 Results XXIV (2016), the *Planck* 2015 SZ counts together with the *Planck* 2018 TT,TE,EE+lowE likelihood (Planck 2018 Results VI 2018), and this work. The constraint from the combined analysis of *Planck* clusters, the tSZ power spectrum, and the tSZ bispectrum of Hurier & Lacasa (2017) is also included.

Data set	$1-b_{\rm SZ}$
WtG	0.688 ± 0.072
CCCP	0.780 ± 0.092
Sereno et al. (2017)[a]	0.66 ± 0.10
Penna-Lima et al. (2017)	0.73 ± 0.10
Medezinski et al. (2018)	0.80 ± 0.14
Hurier & Lacasa (2017)	0.71 ± 0.07
Planck 2015 CMB lensing[b]	$1.01^{+0.24}_{-0.16}$
Planck 2015 SZ + *Planck* 2018 CMB	0.62 ± 0.04
This work	0.71 ± 0.10

[a]In our notation, Sereno et al. (2017) actually constrain $\ln(1-b_{\rm SZ}) = -0.40\pm0.14$.
[b]The CMB lensing measurement in Planck 2015 Results XXIV (2016) is actually $1/(1-b_{\rm SZ})=0.99\pm0.19$.

which is currently the case for galaxy lensing calibrations. It is also not affected by the uncertainties in the photometric redshifts of the background galaxies, which are a limiting factor in galaxy lensing analyses, and, furthermore, allows for the determination of masses of high-redshift clusters, where galaxy lensing mass reconstructions perform badly due to the dearth of background galaxies. For future CMB experiments such as CMB-S4, the CMB lensing signal will be such that, in SZ count analyses, the SZ–mass scaling relations will be able to be calibrated completely with CMB lensing masses alone to subpercent accuracy (Louis & Alonso 2017). In order to achieve this, however, work will be required in order to understand better the biases, intrinsic scatter, and correlations in the cluster observables, constraining them more accurately through numerical simulations and taking into account their dependence on mass and redshift.

ACKNOWLEDGEMENTS

The authors would like to thank James G. Bartlett, Jean-Baptiste Melin, Sebastian Bocquet, and William Handley for useful discussions, and Anna Bonaldi and Richard Battye for providing us with the *Planck* SZ filter noise estimates.

IZ is supported by the Isaac Newton Studentship from the University of Cambridge.

Our MCMC analyses were performed 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).

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

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

Arnaud M., Pratt G. W., Piffaretti R., Böhringer H., Croston J. H., Pointecouteau E., 2010, A&A, 517, A92
Baxter E. J. et al., 2015, ApJ, 806, 247
Baxter E. J. et al., 2017, *MNRAS*, 476, 2688
Becker M. R., Kravtsov A. V., 2011, ApJ, 740, 25
Benoit-Lévy A., Déchelette T., Benabed K., Cardoso J.-F., Hanson D., Prunet S., 2013, A&A, 555, A37
Bleem L. E. et al., 2015, ApJS, 216, 27
Bocquet S. et al., 2018, *ApJ*, 878, 55
de Haan T. et al., 2016, ApJ, 832, 95
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gruetjen H. F., Fergusson J. R., Liguori M., Shellard E. P. S., 2017, Phys. Rev. D, 95, 043532
Gupta N., Saro A., Mohr J. J., Dolag K., Liu J., 2017, MNRAS, 469, 3069
Hanson D., Challinor A., Efstathiou G., Bielewicz P., 2011, Phys. Rev. D, 83, 043005
Hasselfield M. et al., 2013, J. Cosmol. Astropart. Phys., 7, 008
Hoekstra H., Herbonnet R., Muzzin A., Babul A., Mahdavi A., Viola M., Cacciato M., 2015, MNRAS, 449, 685
Hu W., 2001, ApJ, 557, L79
Hu W., Okamoto T., 2002, ApJ, 574, 566
Hu W., DeDeo S., Vale C., 2007, New J. Phys., 9, 441
Hurier G., Lacasa F., 2017, A&A, 604, A71
Kesden M., Cooray A., Kamionkowski M., 2003, Phys. Rev. D, 67, 123507
Kosowsky A., Milosavljevic M., Jimenez R., 2002, Phys. Rev. D, 66, 063007
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Lewis A., Challinor A., Hanson D., 2011, J. Cosmol. Astropart. Phys., 2011, 018
Louis T., Alonso D., 2017, Phys. Rev. D, 95, 043517
Madhavacheril M. et al., 2015, Phys. Rev. Lett., 114, 151302
Mantz A., Allen S. W., Ebeling H., Rapetti D., Drlica-Wagner A., 2010, MNRAS, 406, 1773
Medezinski E. et al., 2018, PASJ, 70, S28
Melin J.-B., Bartlett J. G., 2015, A&A, 578, A21
Meneghetti M., Rasia E., Merten J., Bellagamba F., Ettori S., Mazzotta P., Dolag K., Marri S., 2010, A&A, 514, A93
, , Miyatake H. et al., 2018, *ApJ*, 875, 63,
Muñoz-Cuartas J. C., Macciò A. V., Gottlöber S., Dutton A. A., 2011, MNRAS, 411, 584
Nagai D., Kravtsov A. V., Vikhlinin A., 2007, ApJ, 668, 1
Nagarajan A. et al., 2018, MNRAS, 488, 1759
Namikawa T., Hanson D., Takahashi R., 2013, MNRAS, 431, 609
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Penna-Lima M., Bartlett J. G., Rozo E., Melin J.-B., Merten J., Evrard A. E., Postman M., Rykoff E., 2017, A&A, 604, A89
Piffaretti R., Valdarnini R., 2008, A&A, 491, 71
Planck 2013 Results XXIX, 2014, A&A, 571, A29
Planck 2015 Results VIII, 2016, A&A, 594, A8
Planck 2015 Results XIII, 2016, A&A, 594, A13
Planck 2015 Results XXIV, 2016, A&A, 594, A24
Planck 2015 Results XXVII, 2016, A&A, 594, A27
Planck 2018 Results VI, 2018, preprint (arXiv:1807.06209)
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. Cosmol. Astropart. Phys., 8, 30,
Raghunathan S. et al., 2019, ApJ, 872, 170
Remazeilles M., Delabrouille J., Cardoso J.-F., 2011, MNRAS, 410, 2481
Rephaeli Y., 1995, ARA&A, 33, 541
Salvati L., Douspis M., Aghanim N., 2018, A&A, 614, A13
Seljak U., Zaldarriaga M., 2000, ApJ, 538, 57
Sereno M., Covone G., Izzo L., Ettori S., Coupon J., Lieu M., 2017, MNRAS, 472, 1946
Shirasaki M., Nagai D., Lau E. T., 2016, MNRAS, 460, 3913
Staniszewski Z. et al., 2009, ApJ, 701, 32
Sunyaev R. A., Zeldovich Y. B., 1972, Comments Astrophys. Space Phys., 4, 173
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
von der Linden A. et al., 2014, MNRAS, 439, 2
Yoo J., Zaldarriaga M., Hernquist L., 2010, Phys. Rev. D, 81, 123006

APPENDIX A: EFFECT OF MISCENTRING ON CLUSTER MASS MEASUREMENTS

In this appendix, we provide further details of the bias caused by miscentring in our lensing cluster mass estimates. We also propose a method to account for miscentring bias more carefully in a Bayesian analysis of cluster data.

A1 Miscentring bias

As discussed in Section 4.3, we can build an estimator of a cluster's mass, $\hat{M}_{500}$, or, equivalently, of its signal-to-noise, $p_{\rm obs}$, by matched filtering the cluster's reconstructed convergence profile with an appropriate template. This estimator is unbiased if the template matches exactly the true cluster's profile. This condition also requires that the template is placed exactly at the centre of the true profile. Here we consider the case in which the template is placed at a cluster's estimated centre, $\boldsymbol{x}_{\rm est}$, which is offset from the true centre, $\boldsymbol{x}_{\rm true}$, by a certain angle $\boldsymbol{d}$, i.e. $\boldsymbol{x}_{\rm true} = \boldsymbol{x}_{\rm est} + \boldsymbol{d}$. Assuming that otherwise the template is exactly the same as the true cluster's profile, the expected value of $p_{\rm obs}$ can be written as

$$\langle p_{\rm obs}\rangle(\boldsymbol{d}) = \mathcal{N}\int \frac{d^2\boldsymbol{L}}{2\pi}\frac{|\kappa(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\mathrm{e}^{-\mathrm{i}\boldsymbol{L}\cdot\boldsymbol{d}}, \tag{A1}$$

where κ is the true profile and $\mathcal{N}$ is an appropriate normalization. Expanding the phase term in a power series, only the even powers of $\boldsymbol{d}$ contribute since $\kappa(\boldsymbol{L}) = \kappa^*(-\boldsymbol{L})$ because the convergence is real. If we further assume circular symmetry for the convergence profile, so that $\kappa(\boldsymbol{L})$ only depends on $L = |\boldsymbol{L}|$, we can write the fractional bias due to miscentring by $\boldsymbol{d}$ as

$$\frac{\langle p_{\rm obs}\rangle(\boldsymbol{d}) - \langle p_{\rm obs}\rangle(\boldsymbol{0})}{\langle p_{\rm obs}\rangle(\boldsymbol{0})} = \sum_{n\ \rm even}(-1)^{n/2} f^{(n)} d^n. \tag{A2}$$

Here, the $f^{(n)}$ are coefficients quantifying the fractional bias at each order, which can be written as

$$f^{(n)} = \frac{1}{2^n[(n/2)!]^2}\left(\int dL\, L\frac{|\kappa(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}\right)^{-1}\int dL\, L^{n+1}\frac{|\kappa(\boldsymbol{L})|^2}{N_\kappa(\boldsymbol{L})}. \tag{A3}$$

We see that the leading-order bias is second order in the miscentring angle and is negative (while the next-to-leading-order term is positive). We note that the fractional bias in $p_{\rm obs}$ is also the fractional bias in $\hat{M}_{500}$, since the latter is related to $p_{\rm obs}$ through a factor unaffected by miscentring.

To estimate the size of the effect of miscentring, we can write the n-th order contribution to the bias as

$$(-1)^{n/2} f^{(n)} d^n = (-1)^{n/2}\left(\frac{d}{\lambda^{(n)}}\right)^n, \tag{A4}$$

where $\lambda^{(n)}$ is a typical angle given by $\lambda^{(n)} = (f^{(n)})^{-1/n}$. For a *TT* quadratic estimator reconstruction in an idealized *Planck*-like experiment with a Gaussian beam of FWHM = 5 arcmin and temperature noise levels of 45 μK arcmin, we find $\lambda^{(2)} \approx 6$ arcmin and $\lambda^{(4)} \approx 7$ arcmin. For a cluster with $M_{500} = 0.5 \times 10^{15}\,\mathrm{M}_\odot$ at $z = 0.3$ and a typical miscentring error of around 1 arcmin, the second-order fractional bias is around 10^{-2} (i.e. at the few per cent level), while the fourth-order bias is around 10^{-4}.

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

A2 Accounting for miscentring in a likelihood analysis

As explained in Section 6, in our likelihood analysis we model the miscentring bias with an effective lensing mass bias parameter, which we can marginalize over. Here we describe a more rigorous way to deal with miscentring bias in a likelihood analysis, although we leave implementation of this approach to future work.

If we knew the centring offset $\boldsymbol{d}$, for a cluster with true lensing signal-to-noise $p_{\rm t}$ the average of the observed signal-to-noise over other sources of observational scatter is

$$\langle p_{\rm obs}\rangle(\boldsymbol{d}) \approx p_{\rm t}\left(1+\sum_{n\ \rm even}(-1)^{n/2} f^{(n)} d^n\right). \tag{A5}$$

Here, the bias coefficients $f^{(n)}$ should be evaluated at the appropriate cluster parameters (M_{500}, z, $\hat{\boldsymbol{n}}$, $1-b_{\rm CMBlens}$) in the hierarchical model. However, they are independent of $\boldsymbol{d}$, which proves useful for efficient marginalization over $\boldsymbol{d}$.

Obviously, $\boldsymbol{d}$ is not known precisely, but experiments such as *Planck* do provide an estimate of the uncertainty in each reported cluster position. For a given cluster, let us assume that its position on the sky is described by a certain two-dimensional probability distribution, $P(\boldsymbol{d})$. This could be a two-dimensional Gaussian, for example. What we are interested in is the probability distribution followed by $p_{\rm obs}$, $P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}})$, after the uncertainty due to miscentring has been taken into account. This is given by

$$P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}}) = \int d^2\boldsymbol{d}\, P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}}, \boldsymbol{d}) P(\boldsymbol{d}), \tag{A6}$$

where $P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}}, \boldsymbol{d})$ is the conditional distribution for $p_{\rm obs}$ given the centring error $\boldsymbol{d}$. Retaining the Gaussian approximation for the other observational sources of scatter, this conditional distribution is simply a Gaussian with mean given by equation (A5) and unit variance.

The series expansion in equation (A5) allows $P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}})$ to be evaluated efficiently. We note that if indeed we take $P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}}, \boldsymbol{d})$ and $P(\boldsymbol{d})$ both to be Gaussians, and if we correct for miscentring at least to the lowest non-vanishing order (quadratic), then $P(p_{\rm obs}|p_{\rm t}, M_{500}, z, \hat{\boldsymbol{n}})$ will no longer be Gaussian and will depend (weakly) on M_{500}, z, and $\hat{\boldsymbol{n}}$ through the $f^{(n)}$. Thus, the method outlined here allows us to account for miscentring uncertainty in a rigorous way at the expense of complicating the statistics of our mass observable. We note that a similar miscentring argument should apply to the SZ signal-to-noise measurements, which should have common miscentring errors with the CMB lensing signal-to-noise measurements. Accounting properly for miscentring would therefore imply marginalizing over $\boldsymbol{d}$ simultaneously in both observables.

APPENDIX B: RELATING OUR LIKELIHOOD TO A POISSON COUNT LIKELIHOOD

In this appendix, we demonstrate how the likelihood used in this paper, which is described in detail in Section 6, is mathematically equivalent to a Poisson count likelihood in the $q_{\rm obs}$–$p_{\rm obs}$–z–$\hat{\boldsymbol{n}}$ space for cells of sufficiently small size. We also show how it can be reduced to a likelihood similar to the one used in the *Planck* SZ count analysis (Planck 2015 Results XXIV 2016) by suitable marginalization.

Let us consider an M-dimensional data space for each galaxy cluster; in our case, $M=5$. We continue to assume that clusters can be treated independently of each other (i.e. effects of spatial clustering are negligible), so that our cluster sample can be considered a realization of a Poisson process in this data space. One way to describe the sample statistically is via a Poisson count approach.

In this approach, we first divide the M-dimensional (continuous) data space into a set of $N_{\rm bin}$ cells, covering the whole region of data space where there is a non-zero probability for a cluster to be located. We assume that each cell has equal volume Ω. Since the clusters are assumed to be statistically independent from each other, the number of clusters with data falling within cell I, n_I, follows a Poisson distribution, and the joint probability distribution for the sample, now described by the cell occupation numbers $\boldsymbol{n}=(n_1,\ldots,n_{N_{\rm bin}})$, is simply given by the product of the Poisson distributions of each cell,

$$P(\boldsymbol{n}) = \prod_I \frac{\bar{n}_I^{n_I}\mathrm{e}^{-\bar{n}_I}}{n_I!}. \tag{B1}$$

Here, $\bar{n}_I$ is the expected number of clusters in cell I. We can write $\bar{n}_I = P_I\bar{N}$, where P_I is the probability for a single cluster to lie in cell I given that it passes the selection criterion to be in the sample, and $\bar{N}=\sum_I \bar{n}_I$ is the expected total number of clusters in the sample. We can then rewrite the likelihood as

$$P(\boldsymbol{n}) = \frac{\bar{N}^N\mathrm{e}^{-\bar{N}}}{N!}N!\prod_I\frac{P_I^{n_I}}{n_I!}, \tag{B2}$$

where $N=\sum_I n_I$ is the total number of clusters in the sample. The first term in equation (B2) is a Poisson distribution for N, which we shall refer to as $P(N)$. The second term is a multinomial distribution for $\boldsymbol{n}$ with N trials, that is, the probability distribution followed by $\boldsymbol{n}$ given a fixed N. We shall refer to it as $P(\boldsymbol{n}|N)$. We have therefore factored a product of Poisson distributions across the cells into a Poisson distribution for the total number of clusters and a distribution for the particular configuration of cluster data points at a fixed total number of clusters,

$$P(\boldsymbol{n}) = P(\boldsymbol{n}|N)P(N). \tag{B3}$$

Let us now consider how this Poisson count likelihood relates to the unbinned likelihood used in the rest of this paper. Let us denote points in the M-dimensional data space by $\boldsymbol{x}$ and the data corresponding to the i-th cluster as $\boldsymbol{x}_i$. Also, let $P(\boldsymbol{x}_i|\mathrm{in})$ be the probability density for a cluster to have data $\boldsymbol{x}_i$ given that it passes the sample-selection criterion. A given realization of the cluster sample consists of specifying N and the set $\underline{D}'=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$. As the cluster labels are arbitrary, there are $N!$ permutations that give equivalent samples of clusters. Each permutation gives the same cell occupation numbers $\boldsymbol{n}$, which are got from counting the number of clusters n_I that have their data $\boldsymbol{x}_i$ within cell I. Now consider the limit in which the cells that are small enough that the variation of $P(\boldsymbol{x}|\mathrm{in})$ across the cell can be ignored. If some central value of $\boldsymbol{x}$ in the I-th cell is denoted by $\boldsymbol{x}_I$, for a given sample of N clusters, we have

$$\begin{aligned}\prod_{I=1}^{N_{\rm bin}} P_I^{n_I} &\approx \prod_{I=1}^{N_{\rm bin}}[P(\boldsymbol{x}_I|\mathrm{in})\Omega]^{n_I}\\ &\approx \prod_{i=1}^{N}[P(\boldsymbol{x}_i|\mathrm{in})\Omega]\\ &= P(\underline{D}'|N)\Omega^N.\end{aligned} \tag{B4}$$

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019

Thus, in this limit, we can rewrite the cells likelihood, equation (B2), as

$$P(\boldsymbol{n}) \propto P(\underline{D}'|N)P(N), \tag{B5}$$

where the proportionality factor, which accounts for the change in measure between the two data descriptions, is cosmology independent.

Equation (B5) has the same form as our likelihood, presented in Section 6. There, we separated the angular positions of the N clusters, $\underline{\hat{\boldsymbol{n}}}$, from the other cluster observables, $\underline{D}$, so that $\underline{D}' = (\underline{D}, \underline{\hat{\boldsymbol{n}}})$. Indeed, equation (35) can be written as

$$P\,(N, \underline{\hat{\boldsymbol{n}}}, \underline{D}) = P(\underline{D}, \underline{\hat{\boldsymbol{n}}}|N)P(N) = P(\underline{D}'|N)P(N), \tag{B6}$$

where $P(\underline{D}, \underline{\hat{\boldsymbol{n}}}|N)$ is just the product of the mass data likelihood, $\mathcal{L}_1$ (see Section 6.1), and the sky location likelihood, $\mathcal{L}_2$ (see Section 6.2). Our likelihood is therefore equivalent to a Poisson count likelihood in the D–$\hat{\boldsymbol{n}}$ space (or, more explicitly, the $q_{\rm obs}$–$p_{\rm obs}$–z–$\hat{\boldsymbol{n}}$ space), with cells of sufficiently small size. This also implies that if we were to marginalize our likelihood over all the $p_{\rm obs}$ and $\hat{\boldsymbol{n}}$ variables, we would obtain a likelihood equivalent to a Poisson count likelihood in the $q_{\rm obs}$–z plane (with cells of sufficiently small size). This is similar to the *Planck* SZ count likelihood used in Planck 2015 Results XXIV (2016), except there cells with a fixed finite size are used.

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

Downloaded from https://academic.oup.com/mnras/article-abstract/489/1/401/5543966 by University of Cambridge user on 05 September 2019