Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **543,** 1518–1534 (2025) https://doi.org/10.1093/mnras/staf1541
Advance Access publication 2025 September 13

Calibrating baryonic effects in cosmic shear with external data in the LSST era

Amy Wayland ,★ David Alonso and Matteo Zennaro
Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Accepted 2025 September 10. Received 2025 August 19; in original form 2025 June 13

ABSTRACT

Cosmological constraints derived from weak lensing (WL) surveys are limited by baryonic effects, which suppress the non-linear matter power spectrum on small scales. By combining WL measurements with data from external tracers of the gas around massive structures, it is possible to calibrate baryonic effects and, therefore, obtain more precise cosmological constraints. In this study, we generate mock data for a Stage-IV weak lensing survey such as the Legacy Survey of Space and Time (LSST), X-ray gas fractions, and stacked kinetic Sunyaev–Zel'dovich (kSZ) measurements, to jointly constrain cosmological and astrophysical parameters describing baryonic effects (using the Baryon Correction Model–BCM). First, using WL data alone, we quantify the level to which the BCM parameters will need to be constrained to recover the cosmological constraints obtained under the assumption of perfect knowledge of baryonic feedback. We identify the most relevant baryonic parameters and determine that they must be calibrated to a precision of $\sim$ 10–20 per cent to avoid significant degradation of the fiducial WL constraints. We forecast that long-term X-ray data from $\mathcal{O}(5000)$ clusters should be able to reach this threshold for the parameters that characterize the abundance of hot virialized gas. Constraining the distribution of ejected gas presents a greater challenge, however, but we forecast that long-term kSZ data from a cosmic microwave background-S4-like experiment should achieve the level of precision required for full self-calibration.

Key words: cosmology: large-scale structure of Universe.

1 INTRODUCTION

The distribution of baryonic matter on small scales is strongly affected by complex astrophysical processes, which are collectively referred to as baryonic feedback (van Daalen et al. 2011; Chisari et al. 2019a). These processes, including gas cooling, star formation, and feedback from active galactic nuclei (AGN), can substantially modify the matter distribution within dark matter haloes. On non-linear scales, baryonic feedback leads to a significant suppression of the matter power spectrum. For instance, AGN feedback redistributes baryons to the outer regions of haloes, leading to an overall lower matter density on halo-sized scales. Cosmological simulations indicate that these effects lead to suppression of the power spectrum by $O(10$ per cent) on scales in the range $0.1 \lesssim k \lesssim 10\,h\,\mathrm{Mpc}^{-1}$ (Steinborn et al. 2015; Chisari et al. 2019a; Pakmor et al. 2023; Schaye et al. 2023). In contrast, star formation in the centres of haloes results in a contraction of the matter profile, enhancing the power spectrum on the smallest scales ($k \gtrsim 10\,h\,\mathrm{Mpc}^{-1}$). Baryonic effects lead to a non-trivial modification of the underlying matter distribution that is difficult to model precisely without direct observations (Crain & van de Voort 2023). As a result, the impact of baryonic feedback on the matter power spectrum introduces systematic uncertainties that hinder the precise determination of cosmological parameters from weak lensing surveys (Daalen et al. 2011; Chisari et al. 2019a; Aricò et al. 2023; Bigwood et al. 2024; García-García et al. 2024).

Weak gravitational lensing allows us to directly map the overdensity of matter fluctuations, providing a direct means to trace the history of the growth of structure in the Universe. However, the sensitivity of weak lensing to the matter distribution on small scales is limited by the uncertainties surrounding baryonic feedback, which in turn decreases the precision of cosmological constraints (Amon et al. 2022; Abbott et al. 2023). To mitigate the systematic errors introduced by these effects, many current weak lensing analyses exclude the data from small scales altogether (e.g. Prat et al. 2022; Zacharegkas et al. 2022; Amon et al. 2023; Lange et al. 2023). On the other hand, the Kilo-Degree Survey (KiDS) and Hyper Suprime-Cam (HSC) mitigate uncertainties due to feedback by modelling the small scales with HMCODE (Mead et al. 2021; Abbott et al. 2023; Dalal et al. 2023). Incorporating these non-linear scales is essential to fully benefit from the cosmological potential of upcoming Stage-IV weak lensing surveys, such as the Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory (The LSST Dark Energy Science Collaboration 2018; Ivezić et al. 2019), or the Euclid satellite (Laureijs et al. 2011).

Furthermore, the S_8 tension, which is the discrepancy between the constraints on the clustering amplitude parameter $S_8 = \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$ inferred from some weak lensing surveys and those obtained from the cosmic microwave background (CMB), could be explained by a suppression of the non-linear matter power spectrum (Amon &

★ E-mail: amy.wayland@physics.ox.ac.uk

Efstathiou 2022; Schneider et al. 2022; Aricò et al. 2023; Chen et al. 2023; McCarthy et al. 2023). This suppression may be indicative of either stronger baryonic feedback than predicted by hydrodynamical simulations, or a departure from the standard cosmological model. Recent results from the KiDS collaboration point instead to other causes related to the calibration of redshift distributions and shape measurements (Wright et al. 2025). Independent of the cause, it is essential to improve our understanding of baryonic feedback in order to determine the origin of the S_8 tension and to ensure the robustness of any future cosmological constraints from weak lensing.

External tracers of the large-scale structure, such as X-ray gas fractions and measurements of the kinematic Sunyaev–Zel'dovich (kSZ) effect, can offer valuable complementary information to help us constrain baryonic feedback. X-ray gas fractions, which measure the fraction of baryons in the bound gas component of galaxy clusters as a function of halo mass, are a direct probe of the hot baryon content within haloes. Meanwhile, kSZ measurements, which capture the motion and abundance of ionized gas through the scattering of hot electrons off CMB photons, allow us to trace both the bound and ejected gas components of haloes. These external tracers can help refine our models of baryonic feedback, providing additional constraints on both the cosmological and baryonic parameters (Schneider et al. 2019, 2022; Bigwood et al. 2024; Grandis et al. 2024; To et al. 2024; McCarthy et al. 2025; Sunseri et al. 2025).

At present, it remains uncertain whether external data will be sufficient to constrain baryonic feedback parameters to the level required by the next-generation of cosmic shear surveys (Eifler et al. 2024; Fang et al. 2024). In this work, we use mock LSST-like weak lensing data to determine the level to which each baryonic parameter needs to be calibrated to recover optimal cosmological constraints. We then investigate whether near-term and long-term X-ray and kSZ data have the potential to achieve these calibration requirements. Finally, we combine all three data sets (WL + X-rays + kSZ) to quantify the improvement in cosmological constraints over weak lensing (WL) alone.

The structure of the paper is as follows: In Section 2, we describe the theoretical model for cosmic shear, baryonic effects, X-ray gas fractions, and the stacked kSZ profile. The results of our analysis are presented in Section 3, where we forecast the constraints required on baryonic parameters, before subsequently examining whether weak lensing combined with external tracers can achieve these constraints. Finally, we conclude with a summary of our key findings and their implications for future cosmological surveys in Section 4.

2 COSMIC SHEAR AND BARYONIC EFFECTS

2.1 The cosmic shear signal

Weak gravitational lensing is the deflection of photon trajectories as they propagate along the line of sight from a background source to the observer, caused by the gravitational potential of the intervening large-scale structure (Bartelmann & Schneider 2001; Mandelbaum 2018). This effect in turn leads to coherent distortions in the observed shapes of background galaxies, which is referred to as cosmic shear. As an unbiased tracer of the matter overdensity field, cosmic shear provides a direct probe of the large-scale distribution of matter in the Universe.

We model the cosmic shear signal using the Limber approximation (Limber 1954), under which the cross-correlation of the E-mode cosmic shear signal between two redshift bins, i and j, in harmonic space, is related to the 3D matter power spectrum, $P_{\rm mm}(k, z)$, as

$$C_\ell^{ij} = G_\ell^2 \int \frac{{\rm d}\chi}{\chi^2}\, q_i(\chi) q_j(\chi) P_{\rm mm}\left(k = \frac{\ell + 1/2}{\chi},\, z(\chi)\right), \tag{1}$$

where ℓ is the angular multipole, k is the wavenumber, z is the redshift, χ is the comoving radial distance, and the pre-factor G_ℓ accounts for the difference between the 3D Laplacian of the gravitational potential and the angular Hessian of the associated lensing potential (Kilbinger et al. 2017),

$$G_\ell \equiv \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\frac{1}{(\ell+1/2)^2}. \tag{2}$$

The lensing kernel q_i of the i-th redshift bin is related to the redshift distribution of the sample, $p_i(z)$, via

$$q_i(\chi) = \frac{3}{2} H_0^2 \Omega_{\rm m} \frac{\chi}{a(\chi)} \int_{z(\chi)}^{\infty} {\rm d}z'\, p_i(z') \frac{\chi(z') - \chi}{\chi(z')}, \tag{3}$$

where $a = 1/(1+z)$ is the scale factor.

In this work, all theory calculations are carried out using the Core Cosmology Library (CCL, Chisari et al. 2019b). To accelerate the evaluation of the theory predictions, we use the HALOFIT parametrization to model the dark matter-only non-linear power spectrum (Smith et al. 2003; Takahashi et al. 2012). We include the suppression due to baryonic effects using the baryonification approach as implemented in BACCOEMU (Angulo et al. 2021; Aricò et al. 2021c), with the specific parametrizations given in Section 2.3. We note that we ignore the redshift evolution of the baryonic suppression, and generate synthetic data with one set of baryonic parameters applied to the redshift-dependent non-linear power spectra. While the baryonic parameters should in principle evolve across redshift bins, analyses such as García-García et al. (2024) have shown that current observational data sets lack the sensitivity required to robustly constrain such evolution. For simplicity, and without hindering the generality of our analysis, we adopt the same assumption for our LSST-like synthetic data.

To generate our mock cosmic shear data, we follow the description of the LSST Dark Energy Science Collaboration Science Requirements Document (The LSST Dark Energy Science Collaboration 2018) for the 10-yr weak lensing sample. We assume a number density of sources, redshift distribution, and photometric redshift uncertainty as described in Leonard et al. (2023), and we divide the sample into the same five equal-density redshift bins. The cosmic shear data vector then consists of all the auto- and cross-power spectra between the all redshift bin pairs.

2.2 Systematics affecting the cosmic shear signal

The cosmic shear signal is affected by various sources of systematic uncertainty, which can be categorized into two classes: calibratable and non-calibratable parameters. Tight priors can be imposed on the calibratable systematics through independent external observations or by calibrating the instrument measurements. In contrast, non-calibratable systematics can only be constrained by the data itself.

2.2.1 Intrinsic alignments

The intrinsic alignment (IA) of galaxies with their local environment is a non-calibratable systematic effect that impacts cosmic shear data (Brown et al. 2002). The simplest physical model for intrinsic alignments is the tidal alignment model, more commonly called the non-linear linear alignment model (NLA; Hirata & Seljak 2004).

Consider a nearby structure at the same redshift as a sample of galaxies. The gravitational tidal field is stronger at the end of a galaxy closest to the structure and weaker at the opposite end. This results in a distortion in the shape of the galaxy, causing it to elongate and align with the structure along the longitudinal direction. The NLA model leads to a contribution to the shear power spectrum that is proportional to the matter power spectrum integrated over the redshift distributions of the samples being correlated. The strength of the effect is quantified through a redshift-dependent amplitude, commonly parametrized as

$$A_{\rm IA}(z) = A_{\rm IA,0}\left(\frac{1+z}{1+z_*}\right)^{\eta_{\rm IA}}. \tag{4}$$

Here $A_{\rm IA,0}$ and $\eta_{\rm IA}$ are free parameters that characterize the amplitude and slope of the redshift power-law, respectively. We use $z_* = 0.62$ as in Troxel et al. (2018) and Abbott et al. (2018). Hence, using the NLA model to account for intrinsic alignments introduces two additional free parameters to marginalize over into the likelihood function.

More complex physical process may also contribute to the alignment of galaxies with the large-scale structure, and more general models exist to describe them. These include, for instance, the Tidal Alignment and Tidal Torquing model (Blazek et al. 2019), as well as halo models targeting IAs on small scales (Fortuna et al. 2021). Given the relatively low amplitude of IAs found in current weak lensing samples (Secco et al. 2022; Wright et al. 2025), and the reduced accuracy with which the effect needs to be modelled (Paopiamsap et al. 2024), we limit our analysis to the NLA parametrization for simplicity. Nevertheless, we acknowledge that the NLA model represents one of the simplest alternatives in practice. Other methods have been proposed in the literature to assess the robustness of cosmological constraints derived from weak lensing. One such approach is blue-only analysis, which minimizes the impact of intrinsic alignments (McCullough et al. 2024; Siegel et al. 2025).

2.2.2 *Photometric redshifts and multiplicative bias*

Photometric redshift uncertainties and multiplicative shape measurement biases are important calibratable systematics that affect cosmic shear data (Bonnett et al. 2016; Hildebrandt et al. 2020). The true redshift distributions of the galaxy samples are subject to uncertainty due to the lack of precise redshift measurements in photometric surveys, with the associated errors referred to as photometric redshift uncertainties. The uncertainty in the distribution of the i-th redshift bin can be characterized by a parameter Δz_i that shifts the mean of the redshift bin. This has been shown to capture the impact of redshift uncertainties on cosmic shear data with sufficient accuracy (Bonnett et al. 2016; Ruiz-Zapatero et al. 2023). The true redshift distribution is then given by $p_i(z) = \hat{p}_i(z + \Delta z_i)$, where $\hat{p}_i(z)$ represents the best-guess redshift distribution (Ruiz-Zapatero et al. 2023).

In addition to photometric redshift uncertainties, cosmic shear data is also affected by biases in the measured shape of galaxies arising from the limited resolution and image noise (Miller et al. 2013). This bias is commonly characterized in terms of a multiplicative factor m_i for the i-th sample, known as the multiplicative bias (see e.g. Hildebrandt et al. 2017). Specifically, the measured angular power spectra, $\widetilde{C}_\ell$, are related to the true angular power spectra, C_ℓ, via

$$\widetilde{C}_\ell^{ij} = (1+m_i)(1+m_j)C_\ell^{ij}, \tag{5}$$

where m_i and m_j are the multiplicative biases of the redshift bins i and j, respectively.

We can marginalize over the calibratable parameters using an analytical method based on the Laplace approximation, as proposed in Hadzhiyska et al. (2023) and Ruiz-Zapatero et al. (2023). This approach linearizes the dependence of the theoretical prediction for the data with respect to the calibratable parameters. As a result, the covariance matrix is modified such that the linear combinations of the data most sensitive to variations in these parameters are assigned higher variance. The key advantage of this method is that it avoids introducing two additional free parameters, Δz_i and m_i, per redshift bin i into the likelihood function, thus significantly improving the computational efficiency of running Markov Chain Monte Carlo (MCMC) chains. Furthermore, Ruiz-Zapatero et al. (2023) demonstrated that analytical marginalization recovers the same constraints on the cosmological parameters as directly sampling the full parameter space, even when marginalizing over uncertainties in the full shape of the redshift distribution, instead of shifts in its mean.

To implement the analytical marginalization numerically, we simply update the covariance matrix to

$$\widetilde{\mathsf{C}} \equiv \mathsf{C} + \mathsf{T}\,\mathsf{P}\,\mathsf{T}^T, \tag{6}$$

where P is the covariance matrix of the calibratable systematics, which we assume to be uncorrelated, and the matrix T contains the derivatives of the theoretical prediction t with respect to the set of calibratable parameters v, evaluated at their prior mean values, $\overline{v}$,

$$\mathsf{T} \equiv \left.\frac{\mathrm{d}t}{\mathrm{d}v}\right|_{v=\overline{v}}. \tag{7}$$

In our analysis, we assume a Gaussian prior on both m_i and Δz_i, with 68 per cent uncertainties $\sigma(m_i) = 0.01$ and $\sigma(\Delta z_i) = 0.001(1+\overline{z}_i)$, where $\overline{z}_i$ is the mean of the ith redshift bin. These calibration priors correspond to the LSST requirements as reported in The LSST Dark Energy Science Collaboration (2018).

2.3 The Baryon Correction Model

The other main source of non-calibratable systematic uncertainty for cosmic shear is baryonic feedback. To characterize this effect, we use the Baryon Correction Model (BCM), developed in Schneider & Teyssier (2015). In this model, each halo is decomposed into four components: adiabatically relaxed dark matter, bound gas in hydrodynamical equilibrium, gas ejected from feedback processes, and stars in the central galaxy. This framework allows for a direct connection between the total matter density and the observed distribution of gas and stars in haloes, in turn enabling the physics-based study of baryonic effects on the matter power spectrum.

The BCM modifies the dark matter density profile of haloes to account for the presence of gas and stars via:

$$\rho_{\rm dmo}(r) \rightarrow \rho_{\rm dmb}(r) = \rho_{\rm cdm}(r) + \rho_{\rm gas}(r) + \rho_*(r). \tag{8}$$

Here, $\rho_{\rm dmo}$ represents the dark matter density profile in the absence of baryonic effects, while $\rho_{\rm dmb}$ denotes the dark-matter-baryon profile, which is in turn composed of three components as follows. The cold dark matter profile, $\rho_{\rm cdm}$, is calculated by modifying the standard Navarro–Frenk–White (NFW) profile (Navarro, Frenk & White 1997) to allow for the adiabatic expansion and relaxation of cold dark matter. The total gas profile, $\rho_{\rm gas}(r)$, is composed of both bound and ejected gas components, as described in Sections 2.3.3 and 2.3.4. Baryonic effects, such as AGN feedback, play a significant role in altering the gas distribution around haloes over cosmological scales. The stellar profile, $\rho_*(r)$, is characterized by a power law with an exponential cut-off, which primarily influences the central

region of the halo, rather than larger cosmological scales. Within the framework of the BCM, particles in each halo from a dark-matter only simulation are displaced according to the relative differences between the theoretical baryonic and non-baryonic profiles. As a consequence, the original matter distribution is corrected to account for baryonic effects, while preserving the triaxial structure induced by gravity in the simulation.

In this work, we use the emulator BACCOEMU introduced in Aricò et al. (2021c), which implements the BCM over a wide parameter space, to simultaneously constrain cosmology and baryonic feedback. The model contains seven free baryonic parameters that are connected to the gas and stellar profiles within dark matter haloes. This framework allows us to establish the relationship between the degree of suppression of the power spectrum due to baryonic feedback and the gas and stellar fractions in haloes. In the following sections, we present the mass fractions and density profile models that are subsequently used to calculate the bound gas fraction inferred from X-ray data and the temperature shift of the CMB resulting from the kSZ effect. Specifically, we use the dark matter profile and ejected gas profile of Schneider & Teyssier (2015), the bound gas profile of Aricò et al. (2021a), and the stellar profile of Behroozi, Wechsler & Conroy (2013).

2.3.1 Dark matter

The dark matter profile is assumed to follow the NFW profile of Navarro et al. (1997). Following Baltz, Marshall & Oguri (2009) and Oguri & Hamana (2011), we truncate the NFW profile at radius $r_{\rm tr}$ to prevent the total mass diverging,

$$\rho_{\rm nfw}(x,\, y) = \frac{\rho_0}{x(1+x)^2}\frac{1}{(1+y^2)^2}, \tag{9}$$

where $x \equiv r/r_{\rm s}$ and $y \equiv r/r_{\rm tr}$. Here, $r_{\rm s}$ denotes the scale radius, which is related to the halo virial radius $r_{\rm 200c}$ through $r_{\rm s} \equiv r_{\rm 200c}/c$, where c is the concentration parameter. The truncation radius, $r_{\rm tr}$, marks the outer boundary of the halo. Schneider & Teyssier (2015) evaluated the truncated NFW profile using haloes from DM-only simulations and found that the best-fitting value for the dimensionless truncation parameter $\tau \equiv r_{\rm tr}/r_{\rm s}$ is $\tau = 8c$. Consequently, we adopt $r_{\rm tr} = 8r_{\rm 200c}$ in our analysis.

2.3.2 Central galaxy

We adopt the parametrization from Behroozi et al. (2013) for the mass fraction of stars in the central galaxy, which is based on abundance matching,

$$f_*(M_{\rm 200c}) = \epsilon\left(\frac{\mathcal{M}_1}{M_{\rm 200c}}\right)10^{g(\log_{10}(M_{\rm 200c}/\mathcal{M}_1))-g(0)}, \tag{10}$$

where the function $g(x)$ is defined as

$$g(x) = -\log_{10}(10^{\alpha x}+1) + \frac{\delta(\log_{10}(1+e^x))^\gamma}{1+e^{10^{-x}}}. \tag{11}$$

In this expression, α, γ, δ, and ϵ are redshift-dependent parameters, with their functional form given in Behroozi et al. (2013). The free parameter $\mathcal{M}_1$ is related to the BACCOEMU parameter M_1 via

$$\log_{10}(\mathcal{M}_1) = \log_{10} M_1 + ((a-1)\log_{10} M_{1,a} + z\log_{10} M_{1,z})\nu, \tag{12}$$

where we use the values of the parameters $M_{1,a}$ and $M_{1,z}$ given in Behroozi et al. (2013) and $\nu = \exp(-4a^2)$. Hence, the stellar component depends on a single free parameter, $\log_{10} M_1$, which represents the characteristic mass of the central galaxy of the halo at redshift $z = 0$.

2.3.3 Bound gas

The mass fraction of the bound gas as a function of the halo mass is parametrized as

$$f_{\rm bgas}(M_{\rm 200c}) = \frac{\Omega_{\rm b}/\Omega_{\rm m} - f_*}{(1+(M_{\rm c}/M_{\rm 200c})^\beta)}, \tag{13}$$

where $M_{\rm c}$ and β are free parameters. Notably, $f_{\rm bgas} = \frac{1}{2}(\Omega_{\rm b}/\Omega_{\rm m} - f_*)$ at $M_{\rm c} = M_{\rm 200c}$, which means that $M_{\rm c}$ represents the halo mass at which half of the gas has been ejected from the halo. The parameter β describes the rate at which the gas is depleted towards smaller halo masses.

We adopt a flexible form of the bound gas density profile, as introduced by Aricò et al. (2021a), in which the profile is parametrized as a double power law for $r < r_{\rm out}$,

$$\rho_{\rm bgas}(r < r_{\rm out}) = \frac{y_0}{(1+r/r_{\rm inn})^{\beta_{\rm inn}}}\frac{1}{(1+(r/r_{\rm out})^2)^2}. \tag{14}$$

Here, $\beta_{\rm inn} = 3 - (M_{\rm inn}/M_{\rm 200c})^{\mu_{\rm inn}}$ where $\mu_{\rm inn} = 0.31$ and $M_{\rm inn}$ is a free parameter that characterizes the transition mass of the density profile of the hot gas in haloes. The characteristic scales $r_{\rm inn}$ and $r_{\rm out}$ describe how the slope of the profile evolves at small and large radii, respectively, and are defined as

$$r_{\rm inn} = \theta_{\rm inn} r_{\rm 200c} \text{ and } r_{\rm out} = \theta_{\rm out} r_{\rm 200c}, \tag{15}$$

where $\theta_{\rm inn}$ and $\theta_{\rm out}$ are free parameters of the BACCOEMU model. The normalization constant y_0 in equation (14) ensures that the integrated profile attains the correct mass $f_{\rm bgas} M_{\rm 200c}$ at $r_{\rm 200c}$.

For radii beyond $r_{\rm out}$, we assume that the bound gas profile follows the truncated NFW profile given in equation (9). This assumption is physically motivated by the fact that the bound gas acts as a collisionless fluid in the outer regions of the halo. The normalization constant ρ_0 in equation (9) can be written as $\rho_0 = A y_0$ where the constant A is determined by the requirement that the bound gas density profile is continuous at $r = r_{\rm out}$ and y_0 is the normalization constant in equation (14).

Hence, from equations (13) and (14), we note that the bound gas component is governed by a total of five free parameters: $M_{\rm c}$, β, $\theta_{\rm out}$, $\theta_{\rm inn}$, and $M_{\rm inn}$.

2.3.4 Ejected gas

The ejected gas component represents the remaining gas that has neither been transformed into stars nor is part of the bound gas component, but has instead been mostly expelled from the halo by AGN-driven outflows. Hence, the mass fraction of ejected gas is given by

$$f_{\rm egas}(M_{\rm 200c}) = \Omega_{\rm b}/\Omega_{\rm m} - f_{\rm bgas}(M_{\rm 200c}) - f_*(M_{\rm 200c}), \tag{16}$$

where $\Omega_{\rm b}/\Omega_{\rm m}$ is the total baryon mass fraction, $f_{\rm bgas}$ is the bound gas mass fraction given in equation (13), and f_* is the stellar mass fraction given in equation (10). Assuming that the energy released by the AGN induces shifts in the velocities of the gas particles that follow a Maxwell–Boltzmann distribution, the ejected gas profile takes the form

$$\rho_{\rm egas}(r) = \frac{M_{\rm 200c} f_{\rm egas}}{(2\pi r_{\rm ej}^2)^{3/2}}\exp\left[-\frac{1}{2}\left(\frac{r}{r_{\rm ej}}\right)^2\right], \tag{17}$$

Table 1. Qualitative descriptions of cosmological parameters, baryonic parameters used in the BACCOEMU model, and intrinsic alignment parameters. Note that all masses are in units of $M_\odot$ throughout this work. Here, fiducial value refers to the parameter value used to generate the mock data. We use Gaussian priors, centred on the corresponding fiducial value, for the cosmological and intrinsic alignment parameters. We use the flat BACCOEMU priors for the baryonic parameters. We note that the adopted priors on the cosmological parameters ensure they remain within the valid calibration range of the BACCOEMU emulator.

Parameter	Description	Fiducial Value	Prior
Cosmological			
Ω_m	Density of matter in units of the critical density of the Universe	0.309 67	$\mathcal{N}(0.30967, 0.1)$
σ_8	Cold mass linear mass variance in $8\,h^{-1}$ Mpc spheres	0.8102	$\mathcal{N}(0.8102, 0.1)$
Ω_b	Density of baryons in units of the critical density of the Universe	0.048 97	–
h	Dimensionless Hubble constant	0.6766	$\mathcal{N}(0.6766, 0.1)$
n_s	Scalar spectral index	0.9665	$\mathcal{N}(0.9665, 0.1)$
Σm_ν	Sum of neutrino masses in units of eV	0.06	$\mathcal{N}(0.06, 0.1)$
Baryonic Feedback			
$\log_{10} M_c$	Characteristic halo mass for which half the gas is retained	14.0	U[9.0, 15.0]
$\log_{10} \eta$	Directly proportional to the radius of ejected gas from the halo	−0.3	U[−0.7, 0.7]
$\log_{10} \beta$	Describes the rate at which the depletion of gas increases towards smaller haloes	−0.22	U[−1.0, 0.7]
$\log_{10} M_1$	Characteristic mass of haloes that host a central galaxy at $z = 0$	10.674	U[9.0, 13.0]
$\log_{10} M_{inn}$	Characteristic transition mass of the density profile of hot gas in haloes	13.0	U[9.0, 13.5]
$\log_{10} \theta_{inn}$	Controls the inner radius of the bound gas density profile	−0.86	U[−2.0, 0.5]
$\log_{10} \theta_{out}$	Controls the outer radius of the bound gas density profile	0.25	U[0.0, 0.5]
Intrinsic Alignments			
$A_{IA,0}$	Amplitude of the NLA model for IAs	0.0	$\mathcal{N}(0, 1.0)$
η_{IA}	Slope of the NLA model for IAs	0.0	$\mathcal{N}(0, 1.0)$

where the ejection radius r_{ej} is the maximum radius reached by the expelled gas, which we model as

$$r_{ej} \equiv \eta \times 0.75\, r_{esc}.$$

Here, η is a free parameter and r_{esc} is the escape radius, which can be related to the virial radius of the halo, r_{200c} (Schneider & Teyssier 2015). Consequently, the ejected gas component depends only on a single free parameter, η.

In summary, the Baryon Correction Model, as implemented in BACCOEMU, incorporates a total of seven free baryonic parameters, as outlined in Table 1. It is important to note, however, that the model does not account for potential complexities such as mass-dependent or redshift-dependent variations in these parameters. The impact of introducing redshift dependence is explored in Appendix A. Moreover, the model assumes a universal value of η across all haloes. In practice, this assumption may be oversimplified, as η is likely to exhibit a dependence on the halo mass (e.g. more massive haloes, with deeper gravitational potential wells, should be able to retain the ejected gas more effectively).

2.4 Probes of baryons: kSZ and X-ray mass fractions

Here we will consider two gas probes that have been often used in the past to quantify the impact of baryonic feedback in the matter power spectrum: X-ray constraints on the bound gas fraction and measurements of the kinematic Sunyaev–Zel'dovich effect (e.g. Schneider et al. 2019; Bigwood et al. 2024; Grandis et al. 2024; Hadzhiyska et al. 2024).

2.4.1 X-ray observations

The gas bound within galaxy clusters is primarily composed of ionized hydrogen and metals, which emit photons in the X-ray regime. Hence, observations of X-ray radiation emanating from clusters enable the study of the bound gas component. The resulting measurements of the bound gas fraction can thus be used to constrain the BCM parameters M_c and β, governing the mass dependence of the bound fraction in equation (13). To summarize the gas fractions measured in X-ray data, we parametrize the total mass of bound gas, M_{bgas}, as in Grandis et al. (2024),

$$\frac{M_{bgas}}{M_{bgas}^{piv}} = A_{bgas} \left(\frac{M_{500c}}{M_{piv}} \right)^{B_{bgas}}, \tag{18}$$

where M_{500c} is the halo mass contained within a radius r_{500c}, enclosing a total mass density corresponding to 500 times the critical density, M_{piv} is the median mass value of the cluster sample, and M_{bgas}^{piv} is the mass normalization at the pivot mass. The free parameters A_{bgas} and B_{bgas} describe the normalization at the pivot mass and the power-law index of the gas mass to halo mass relation, respectively. Theoretical predictions for these parameters can be derived in terms of the bound gas fraction in a given halo, f_{bgas}, which in X-ray studies is defined as

$$f_{bgas} \equiv \frac{M_{bgas}}{M_{500c}}. \tag{19}$$

Using the scaling relations from Grandis et al. (2024), the theoretical predictions for A_{bgas} and B_{bgas} are given by

$$A_{bgas} = \frac{M_{piv}}{M_{bgas}^{piv}}\, f_{bgas}\big|_{M_{500c}=M_{piv}}, \tag{20}$$

$$B_{bgas} = 1 + \left. \frac{d \ln f_{bgas}}{d \ln M_{500c}} \right|_{M_{500c}=M_{piv}}. \tag{21}$$

In this study, we use the parametrization of the bound gas fraction implemented in BACCOEMU, as given in equation (13). To generate our X-ray mock data for A_{bgas} and B_{bgas}, we use the pivot masses and redshifts from Chiu et al. (2018, 2022) and Akino et al. (2022), with $M_{bgas}^{piv} = 10^{13} M_\odot$. Since observations are mostly sensitive to the X-ray signal originating from within r_{500c}, X-ray cluster studies typically adopt the halo mass definition M_{500c}. Hence, we use the concentration–mass relation from Ishiyama et al. (2021) to convert from M_{500c} to M_{200c} to calculate the bound gas and stellar fractions, f_{bgas} and f_*, using equations (13) and (10).

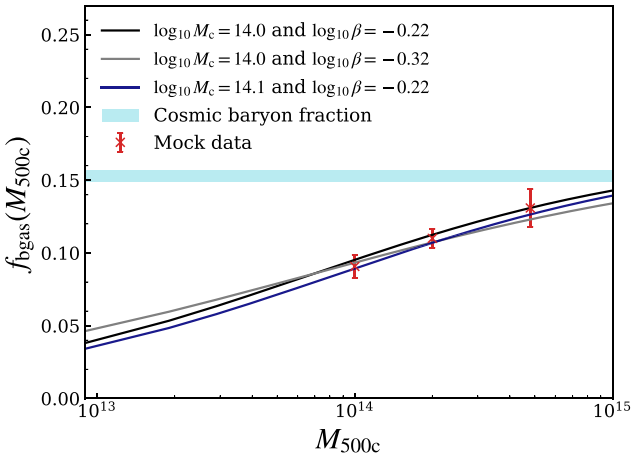

Figure 1. The bound gas fraction and its associated error from mock measurements of $A_{\rm bgas}$ and $B_{\rm bgas}$ based on the pivot masses and redshifts from recent X-ray studies, with $M_{\rm bgas}^{\rm piv} = 10^{13}{\rm M}_\odot$. The error bars are derived from currently available measurements summarized in Grandis et al. (2024). The black curve shows the BACCOEMU parametrization for the bound gas fraction (equation 13) for the fiducial cosmology used to generate the mock measurements. The blue and grey curves show the sensitivity of the model to variations in $\log_{10} M_{\rm c}$ and $\log_{10}\beta$, respectively. The horizontal band shows the mean cosmological value of the bound gas fraction, defined as $f_{\rm bgas} = \Omega_{\rm b}/\Omega_{\rm m} - f_*$. The variation in the value of $f_{\rm bgas}$ across the band reflects differences in the stellar fraction, f_*, associated with each individual mock data point.

We present the predictions for the bound gas fraction, $f_{\rm bgas}$, derived from the BACCOEMU parametrization in equation (13), as a function of halo mass, M_{500c}, in Fig. 1. To generate the mock data points, shown in red, we use a fixed set of BCM parameters to characterize the gas ($\log_{10} M_{\rm c} = 14.0$, $\log_{10}\eta = -0.30$, $\log_{10}\beta = -0.22$, $\log_{10} M_1 = 10.674$, $\log_{10}\theta_{\rm inn} = -0.86$, $\log_{10}\theta_{\rm out} = 0.25$, $\log_{10} M_{\rm inn} = 13.0$). Here, all masses are in units of ${\rm M}_\odot$. The blue and grey lines show the sensitivity of the model to variations in $\log_{10} M_{\rm c}$ and $\log_{10}\beta$, respectively. This illustrates the ability of X-ray gas fractions to constrain the bound gas abundance. Moreover, Fig. 1 shows that smaller haloes exhibit a lower gas fraction compared to larger ones. This deficiency in gas is primarily attributed to the stronger influence of AGN feedback on smaller haloes, which are less capable of retaining gas due to their weaker gravitational field. The gas fraction approaches the mean cosmological value, $f_{\rm bgas} = \Omega_{\rm b}/\Omega_{\rm m} - f_*$, for haloes of larger mass, in agreement with hydrodynamical simulations (e.g. Sorini et al. 2022; Ayromlou, Nelson & Pillepich 2023).

Subsequently, we use the mock data points for the bound gas fraction presented in Fig. 1 to generate mock measurements for $A_{\rm bgas}$ and $B_{\rm bgas}$ via equations (20) and (21), respectively.

2.4.2 *kSZ stacking observations*

The kSZ effect is a secondary anisotropy of the CMB in which a Doppler shift is imparted on CMB photons via inverse Compton scattering, which is induced by the bulk motion of ionized gas in galaxy clusters (Sunyaev & Zeldovich 1970, 1972). In turn, this shifts the temperature of the CMB by

$$\frac{\Delta T_{\rm kSZ}}{T_{\rm CMB}} = -\sigma_{\rm T}\int {\rm d}l\, e^{-\tau(z)} n_{\rm e}\frac{v_{\rm e}\cdot\hat{n}}{c}, \tag{22}$$

where $\sigma_{\rm T}$ is the Thomson cross-section, $\tau(z)$ is the optical depth to Thomson scattering along the line of sight at redshift z, $n_{\rm e}$ is the physical number density of free electrons, $v_{\rm e}$ is the peculiar velocity, and c is the speed of light. The integral $\int {\rm d}l = \int {\rm d}\chi/(1+z)$ is taken along the line of sight in the direction given by the unit vector $\hat{n}$. Since the kSZ signal is proportional to the integrated electron momentum along the line of sight, it can provide a direct measurement of the electron density, provided that the velocity field is known.

We generate mock data for stacked kSZ measurements, in which the CMB map is correlated with the positions of galaxies weighted by their reconstructed velocities. The result may be interpreted as being proportional to the gas density profile. We assume measurements similar to those made by the Atacama Cosmology Telescope (ACT) DR5 and Planck (Amodeo et al. 2021; Schaan et al. 2021) for the constant stellar mass (CMASS) galaxy sample from the Baryon Oscillation Spectroscopic Survey (Ahn et al. 2014), with an average halo mass of $M_{200c} = 3\times10^{13}\,{\rm M}_\odot$ and an average redshift of $z = 0.55$. Since the optical depth is below the percent level (Adam et al. 2016), we can approximate the galaxy sample as optically thin, $e^{-\tau(z)} \approx 1$. In this limit, we can recast equation (22) in terms of the optical depth of the galaxy cluster, $\tau_{\rm gal}$, as

$$\frac{\Delta T_{\rm kSZ}}{T_{\rm CMB}} = -\tau_{\rm gal}\frac{v_{\rm e,r}}{c}, \tag{23}$$

where $v_{\rm e,r}$ is the rms radial velocity of the free electrons, which we approximate as $v_{\rm e,r} = 1.06\times10^{-3}c$, following Schaan et al. (2021) and Amodeo et al. (2021). Assuming spherical symmetry, the optical depth $\tau_{\rm gal}$ measured at an angular separation θ away from the centre of the cluster is given by

$$\tau_{\rm gal}(\theta) = 2\sigma_{\rm T}\int_0^{r_{\rm 200c}} {\rm d}l\, n_{\rm e}\left(\sqrt{d_{\rm A}(z)^2\theta^2 + l^2}\right), \tag{24}$$

where $d_{\rm A}(z)$ is the angular diameter distance, and the electron number density profile, $n_{\rm e}(r)$, is related to the total gas density profile, $\rho_{\rm gas}(r) = \rho_{\rm bgas}(r) + \rho_{\rm egas}(r)$, via

$$n_{\rm e}(r) = \frac{X_{\rm H}+1}{2m_{\rm amu}}\rho_{\rm gas}(r). \tag{25}$$

Here, $X_{\rm H} = 0.76$ is the hydrogen mass fraction and $m_{\rm amu}$ is the atomic mass unit.

It is important to note the potential limitations of this interpretation of the stacked kSZ profile. Specifically, we do not incorporate the 2-halo term, and the velocity reconstruction method has inherent shortcomings. Consequently, our interpretation of the constraining power of the kSZ is likely optimistic, and any real constraints derived from this analysis may be less stringent.

Since the kSZ signal has the same frequency spectrum as the primary CMB, the CMB anisotropies themselves contaminate the kSZ map. To limit the estimator variance caused by large-scale CMB fluctuations, the standard stacked kSZ estimators apply a compensated aperture photometry filter (CAP) to the CMB map (Alonso et al. 2016; Schaan et al. 2016, 2021). This method involves considering two concentric circles with radii $\theta_{\rm d}$ and $\sqrt{2}\theta_{\rm d}$, chosen such that the area enclosed by the inner disc and the outer ring are equal. Since the CMB fluctuations have a typical angular size much larger than $\theta_{\rm d}$, the CMB fluctuations will be approximately constant over the aperture. Therefore, the net CMB flux will be zero when the flux over the inner disc is subtracted from that over the outer ring, leaving only the kSZ flux. Mathematically, we can model the CAP filter using a window function,

$$W_{\theta_{\rm d}} = \begin{cases} 1 & \text{for } 0 < \theta < \theta_{\rm d}, \\ -1 & \text{for } \theta_{\rm d} < \theta < \sqrt{2}\theta_{\rm d}, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

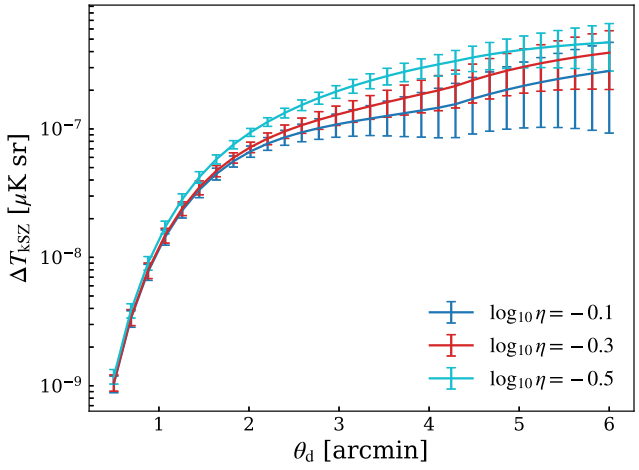

Figure 2. The CMB temperature shift due to the kSZ effect as a function of the aperture radius $\theta_{\rm d}$ of the CAP filter and for different feedback strengths. The model profile is convolved with the ACT DR5 f90 beam profile with FWHM = 2.1 arcmin. The error bars are derived from the baseline noise level of the Simons Observatory.

Hence, the observed kSZ signal is a convolution of equation (23) with the CAP filter in equation (26) and the CMB beam function. For the latter, we approximate the beam as a Gaussian with a full width at half-maximum (FWHM) of 2.1 arcmin, corresponding to a diffraction-limited telescope with a ~ 6 m diameter aperture at 98 GHz (Schaan et al. 2021). To generate mock measurements of the stacked kSZ profile, we use 30 bins for $\theta_{\rm d}$, ranging from $\theta_{\rm d} = 0.5$ to 6.0 arcmin.

From equation (25), we observe that the kSZ signal is directly sensitive to the total gas density profile, with the bound and ejected gas profiles modelled by equations (14) and (17), respectively. Thus, tracing the kSZ signal provides a means to constrain the baryonic feedback parameters that describe both the bound and ejected gas profiles. In contrast to X-ray data alone, which is mostly sensitive to the bound gas component, kSZ data allows us to constrain the parameter η, which governs the physical extent of the ejected gas. However, it is important to note that, in practice, our mock measurements of the kSZ profile cannot provide individual constraints on the bound gas parameters, $M_{\rm c}$ and β. This limitation arises because our kSZ data probes only a single halo mass and redshift, leading to a degeneracy between $M_{\rm c}$ and β that cannot be broken without additional data.

In Fig. 2, we illustrate the kSZ signal as a function of the CAP aperture radius for a halo of fixed mass. In this figure, we explore the impact of varying the strength of baryonic feedback by modifying the parameter $\log_{10}\eta$. We find that an increase in $\log_{10}\eta$ decreases in the kSZ signal at small aperture radii $\theta_{\rm d}$, as a larger fraction of the gas is ejected beyond the virial radius. Independent of the feedback strength, the kSZ signal asymptotically approaches the same value at large $\theta_{\rm d}$, since the total mass enclosed within the halo remains constant.

It is worth emphasizing the limitations of our treatment of the kSZ data in this analysis. On the one hand, we neglect a number of potentially important sources of uncertainty that could degrade the constraining power of kSZ data. First, the amplitude of the stacked kSZ signal depends on the quality of the reconstructed galaxy velocities. Uncertainties in the amplitude of these velocities (e.g. due to uncertainties in galaxy bias, or inaccuracies in the reconstruction algorithm), should be propagated when using the kSZ measurements to extract information about the gas distribution. Secondly, we will interpret the stacked kSZ measurements as being solely sensitive to the electron density profile of haloes of a given mass. This neglects important additional contributions, including the '2-halo' contribution from surrounding structures, the impact of miscentring and satellite galaxies, and additional correlations between the velocity and density fields. Specifically, the presence of satellite galaxies and the effects of miscentring can introduce a bias in the measured kSZ amplitude. To illustrate, McCarthy et al. (2025) demonstrate that accounting for satellite contributions can enhance the kSZ signal by approximately 20 per cent. On the other hand, we have only considered a single galaxy sample, probing a single halo mass, at a single effective redshift. kSZ measurements are able to probe a range of redshifts and masses (Hadzhiyska et al. 2024; Ried Guachalla et al. 2025), which could allow us to constrain $\log_{10} M_{\rm c}$ and β separately, as well as the potential mass/redshift dependence of η. Our analysis should, nevertheless, provide an indication of whether the sensitivity of current and future CMB observations will be enough to constrain feedback parameters at the rough level required of Stage-IV weak lensing data.

2.5 Likelihood

Throughout this study, we adopt a *Planck* cosmology (Planck Collaboration VI 2020) with parameter values $\{\Omega_{\rm c}, \Omega_{\rm b}, h, n_{\rm s}, \sigma_8, \Sigma m_\nu\} = \{0.2607,\ 0.048\,97,\ 0.6766,\ 0.9665,\ 0.8102, 0.06\}$ as the fiducial model to generate the mock data. We assume that the mock data, d, follows a Gaussian likelihood with the χ^2 statistic for the set of parameters, $\vec{\theta}$, given by

$$\chi^2 = -2\log p(d|\vec{\theta}) = (d - t(\vec{\theta}))^{\rm T}\mathsf{C}^{-1}(d - t(\vec{\theta})) + K, \tag{27}$$

where t is the theoretical prediction for the data, C is the covariance matrix of the data, and K is a normalization constant. To sample the posterior distribution, we use `Cobaya` (Torrado & Lewis 2021), which implements the Metropolis–Hastings MCMC method (Metropolis et al. 1953). We simultaneously marginalize over five cosmological parameters ($\Omega_{\rm m}, \sigma_8, h, n_{\rm s}, \Sigma m_\nu$), two intrinsic alignment parameters ($A_{\rm IA,0}$, $\eta_{\rm IA}$), and the seven baryonic parameters implemented in BACCOEMU ($\log_{10} M_{\rm c}$, $\log_{10}\eta$, $\log_{10}\beta$, $\log_{10} M_1$, $\log_{10}\theta_{\rm inn}$, $\log_{10}\theta_{\rm out}$, $\log_{10} M_{\rm inn}$). The calibratable systematics, Δz_i and m_i, are marginalized over using the analytical approximation outlined in Section 2.2.2, which avoids introducing an additional two free parameters for each redshift bin into the likelihood function. A summary of the free parameters and their priors is presented in Table 1.

We justify our selection of parameters as follows. Although the primordial spectral index, $n_{\rm s}$, is tightly constrained by measurements of the CMB, we marginalize over it in this analysis to incorporate complementary constraints from other cosmological probes. In contrast, we fix the cosmological baryon fraction, $\Omega_{\rm b}$, given its precise determination from both the CMB and Big Bang Nucleosynthesis (BBN). This approach allows for an independent cross-validation of the CMB-derived baryon fraction through BBN, making additional constraints from weak lensing non-essential in this context.

We use CCL (Chisari et al. 2019b) to generate LSST-like weak lensing data for multipoles up to $\ell_{\rm max} = 2000$. To account for the finite width of the ℓ bins, we use a linear spacing of $\Delta\ell = 10$ up to a given value $\ell_{\rm linear}$. Beyond this value, we use logarithmic sampling with ten ℓ values per decade. The value of $\ell_{\rm linear}$ is chosen such that the separation between adjacent ℓ values beyond $\ell_{\rm linear}$ using logarithmic sampling is greater than or equal to $\Delta\ell_{\rm linear} = 10$. We construct the covariance matrix for the LSST cosmic shear power

spectra using the Knox formula,

$$\mathrm{Cov}(C_\ell^{ij}, C_{\ell'}^{km}) = \frac{\delta_{\ell\ell'}}{f_{\rm sky}(2\ell+1)}\left(C_\ell^{ik}C_\ell^{jm} + C_\ell^{im}C_\ell^{jk}\right), \tag{28}$$

where $f_{\rm sky} = 0.4$ is the fraction of the sky covered by the weak lensing survey. We also include the correction $\mathsf{T}\,\mathsf{P}\,\mathsf{T}^T$ from marginalizing analytically over photometric redshift uncertainties and multiplicative shape biases (see equation 6).

It is important to note that, in principle, the shear data set for the LSST 10-yr sample described in Section 2.1 is sensitive to redshifts up to $z = 4.0$, and physical wavenumbers up to $k = 10\,\mathrm{Mpc}^{-1}$, whereas the BACCOEMU emulator is calibrated only up to redshift $z = 1.5$ and wavenumbers of $k = 5\,h\,\mathrm{Mpc}^{-1}$, which corresponds to $k \sim 3.5\,\mathrm{Mpc}^{-1}$ in our fiducial cosmology. However, the contribution from shape noise to the power spectrum covariance limits the sensitivity to the smallest angular scales, and the lensing kernels peak half-way between the observer and the source, suppressing the contribution from structure at high redshifs. This allows us to rely on relatively simple extrapolation schemes for *Baccoemu*. These are described in Appendix B, where we also assess their validity for the data set simulated here.

For the X-ray mock data, we assume that the different cluster samples exhibit minimal to no overlap, and therefore we treat the measurements as mutually independent. As a result, the covariance matrix for the X-ray mock data takes a diagonal form with the errors for the $A_{\rm bgas}$ and $B_{\rm bgas}$ measurements in each cluster sample on the leading diagonal. We use the errors obtained from near-term X-ray samples, as given in Chiu et al. (2018, 2022) and Akino et al. (2022). We also consider potential futuristic data based on the Extended Roentgen Survey with an Imaging Telescope Array (eROSITA) survey, as discussed in Section 3.3.

The covariance for the kSZ mock data can be calculated analytically, as described in Alonso et al. (2016)

$$\mathrm{Cov}(\Delta T_{\rm kSZ}(\theta_1), \Delta T_{\rm kSZ}(\theta_2)) = 2\pi\theta_1^2\theta_2^2 \times \int \mathrm{d}\ell\,\ell\,C_N(\ell)\widetilde{W}(\ell|\theta_1)\widetilde{W}(\ell|\theta_2), \tag{29}$$

where $C_N(\ell)$ is the noise power spectrum which includes contributions from both the CMB and instrument noise, $C_N(\ell) = C_\ell^{\rm CMB} + N_\ell$, and $\widetilde{W}$ is the Fourier transform of the CAP filter given in equation (26),

$$\widetilde{W}(\ell|\theta_{\rm d}) = \frac{2J_1(\ell\theta_{\rm d}) - \sqrt{2}J_1(\sqrt{2}\ell\theta_{\rm d})}{\ell\theta_{\rm d}}, \tag{30}$$

where $J_1(x)$ is the first-order cylindrical Bessel function. As described in Section 3.3, we will consider future kSZ measurements based on the sensitivities of the Simons Observatory and CMB-S4 experiments (Ade et al. 2019; Schiappucci et al. 2025).

3 RESULTS

Using the data and theory pipeline discussed in the previous sections, we now perform an inference analysis using the likelihood and MCMC sampling method described in Section 2.5. We first quantify the level of information loss in cosmic shear analyses due to lack of understanding of baryonic effects. Then, we determine the level to which different baryonic parameters must be calibrated to avoid this information loss, and the ability of X-ray and kSZ data to achieve this requirement. Finally, we examine the cosmological constraints that may be obtained by combining cosmic shear with forecasts of X-ray and kSZ data as external calibrators.

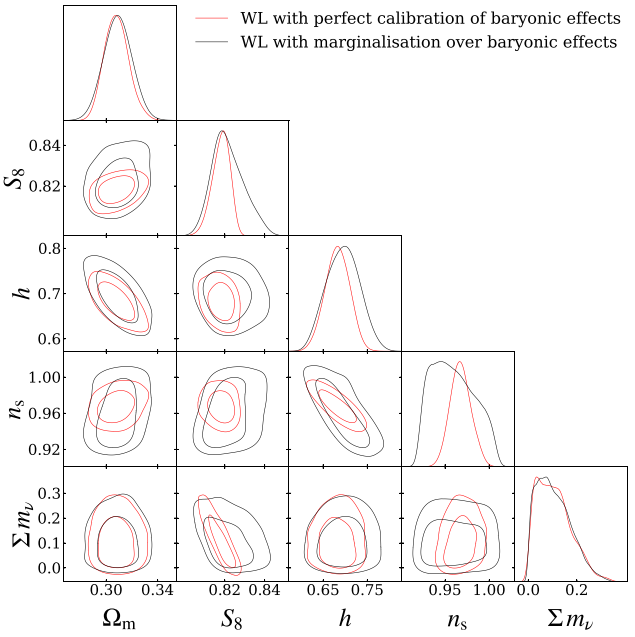

Figure 3. The marginalized posteriors for $\Omega_{\rm m}$, $S_8 = \sigma_8\sqrt{\Omega_{\rm m}/0.3}$, h, $n_{\rm s}$, and Σm_ν obtained from LSST-like weak lensing data up to multipoles of $\ell = 2000$. We compare the posteriors obtained under the marginalization over baryonic effects to the case where the baryonic parameters are kept fixed at the values used to generate the mock data. The inner and outer contours show the 95 per cent and 68 per cent confidence levels, respectively. We marginalize over intrinsic alignments, photometric redshift uncertainties, and multiplicative shape biases in both cases.

3.1 Information loss due to baryonic effects

As a first result, we quantify the extent to which having freedom over the baryonic parameters degrades the cosmological constraints for LSST-like cosmic shear data. We use the complete mock data setup to multipoles $\ell = 2000$ in order to forecast the constraints achievable with the full LSST data set in the future. Consequently, we do not evaluate the impact of applying scale cuts in this analysis.

In Fig. 3, the posterior constraints obtained assuming perfect knowledge of baryonic effects (red) are compared to those obtained after marginalizing over baryonic feedback parameters (black). The numerical constraints are listed in the first two rows of Table 2. We find that the errors on S_8, the parameter best determined by current weak lensing surveys, increase by a factor of 1.9 when marginalizing over baryonic parameters. Future weak lensing data will be able to make precision measurements of other cosmological parameters, which are also affected by uncertainties on the impact of baryons. In particular, we find that the errors on the Hubble parameter, h, and the scalar spectral index, $n_{\rm s}$, grow by factors of 1.4 and 1.8, respectively. This highlights the importance of self-calibrating baryonic effects in order to fully benefit from the cosmological power of the upcoming Stage-IV weak lensing data. Furthermore, we find that, interestingly, baryonic effects do not affect the constraints on neutrino masses that are obtained from weak lensing data alone. This may change in the presence of additional cosmological data (e.g. from CMB primary anisotropies, or by combining weak lensing and galaxy clustering), but we leave this study for future work. We note that the loss of constraining power in S_8 comes mainly from σ_8 rather than from $\Omega_{\rm m}$. This is because the amplitude of the lensing power spectrum is influenced by both the intrinsic amplitude of the matter power spectrum and the distance–redshift relation. Baryonic effects affect

Table 2. The constraints on S_8, Ω_m, h, and n_s obtained from different combinations of large-scale structure tracers. We report the mean marginal value of each parameter and their associated errors given by the 95 per cent confidence level. Here, WL refers to LSST-like weak lensing data. We marginalize over intrinsic alignments, photometric redshift uncertainties, and multiplicative shape biases for the weak lensing data. In all cases other than the first row, we marginalize over the parameters describing baryonic effects. Note that we do not report the constraints on the neutrino mass, as weak lensing is not able to place tight constraints on this parameter.

Tracer(s)	S_8	Ω_m	h	n_s
WL only with baryons fixed	$0.8181^{+0.0081}_{-0.0090}$	$0.308^{+0.019}_{-0.018}$	$0.682^{+0.051}_{-0.055}$	$0.968^{+0.024}_{-0.022}$
WL only	$0.8220^{+0.0170}_{-0.0140}$	$0.309^{+0.022}_{-0.021}$	$0.695^{+0.073}_{-0.071}$	$0.959^{+0.043}_{-0.038}$
WL + near-term X-ray	$0.8210^{+0.0150}_{-0.0130}$	$0.309^{+0.019}_{-0.018}$	$0.693^{+0.071}_{-0.068}$	$0.962^{+0.041}_{-0.039}$
WL + near-term kSZ	$0.8200^{+0.0090}_{-0.0100}$	$0.309^{+0.015}_{-0.016}$	$0.678^{+0.049}_{-0.040}$	$0.971^{+0.035}_{-0.039}$
WL + near-term X-ray + near-term kSZ	$0.8196^{+0.0094}_{-0.0100}$	$0.308^{+0.015}_{-0.016}$	$0.683^{+0.054}_{-0.052}$	$0.970^{+0.035}_{-0.037}$
WL + long-term X-ray	$0.8210^{+0.0130}_{-0.0130}$	$0.309^{+0.014}_{-0.015}$	$0.688^{+0.066}_{-0.061}$	$0.963^{+0.042}_{-0.040}$
WL + long-term kSZ	$0.8196^{+0.0087}_{-0.0098}$	$0.308^{+0.014}_{-0.016}$	$0.683^{+0.054}_{-0.034}$	$0.968^{+0.034}_{-0.037}$
WL + long-term X-ray + long-term kSZ	$0.8196^{+0.0088}_{-0.0096}$	$0.309^{+0.011}_{-0.012}$	$0.679^{+0.032}_{-0.027}$	$0.972^{+0.024}_{-0.024}$
WL + near-term X-ray + near-term kSZ with $n_{gal} \times 10$	$0.8196^{+0.0078}_{-0.0086}$	$0.309^{+0.014}_{-0.014}$	$0.672^{+0.040}_{-0.028}$	$0.974^{+0.027}_{-0.032}$

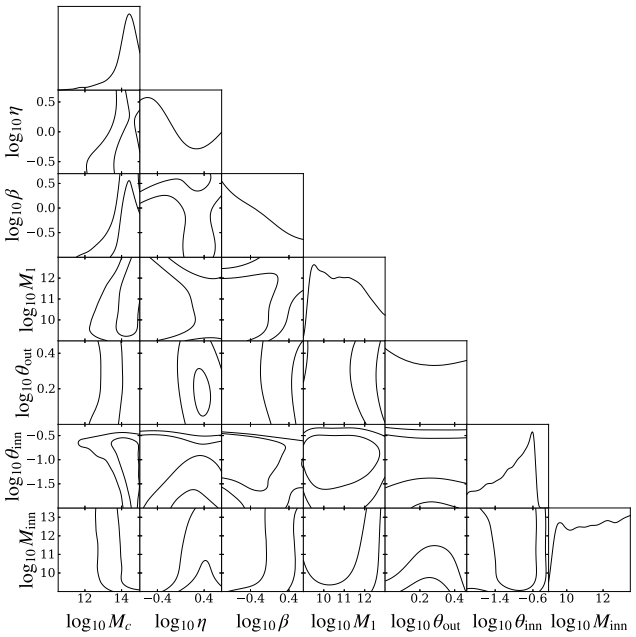

Figure 4. The marginalized posteriors on the baryonic parameters for LSST-like cosmic shear data alone. The inner and outer contours show the 68 per cent and 95 per cent confidence levels, respectively. We marginalize over intrinsic alignments, photometric redshift uncertainties, and multiplicative shape biases.

the amplitude of the matter power spectrum, and not the distance–redshift relation, which results in a direct impact on σ_8 but not Ω_m.

It is also interesting to study the level to which future, more sensitive, cosmic shear data will be able to self-calibrate baryonic effects, building on the work conducted in Preston, Amon & Efstathiou (2024). Fig. 4 shows the marginalized posteriors on the baryonic parameters obtained from this analysis. The only baryonic parameter for which meaningful constraints can be obtained is $\log_{10} M_c$. We note that the upper bound on M_c is due to the upper limit of the BACCOEMU prior. This parameter determines the abundance of bound gas in haloes (and hence also the fraction of ejected gas), and thus controls the amplitude of the baryonic power suppression. This result is in agreement with the results found in the literature when analysing current cosmic shear data sets (Aricò et al. 2023; García-García et al. 2024), and it is interesting to observe that the enhanced sensitivity of LSST cosmic shear data will not allow it to significantly constrain other baryonic parameters on its own. As we show in the next section, although cosmic shear is only able to constrain $\log_{10} M_c$, other baryonic parameters play a significant role in significantly degrading cosmological constraints. We note, however, that higher-order statistics, such as the bispectrum, are expected to enhance the WL-only constraints on baryonic parameters (Foreman et al. 2020; Aricò et al. 2021b).

3.2 Calibration requirements for baryonic effects

As a next step, we identify the baryonic parameters that most significantly contribute to the degradation of the cosmological constraints, and estimate the level to which they must be calibrated to minimize this degradation. To achieve this, we perform a series of MCMC analyses in which one baryonic parameter is varied at a time, keeping all other parameters fixed. This approach allows us to isolate the individual impact of each baryonic parameter on the cosmological constraints. Note that this approach will not capture the internal degeneracies between different baryonic parameters, which usually lead to further degradation in the cosmological constraints. Nevertheless, it will allow us to obtain a rough estimate of the calibration requirements.

We focus on the constraints obtained on S_8, h, and n_s, as these are the parameters most affected by uncertainties on baryonic effects (see Fig. 3). In Fig. 5, we present the errors on these parameters, as a function of the calibration prior in each of the baryonic parameters. Our results indicate that variations in $\log_{10} M_c$ and $\log_{10} \eta$ most substantially degrade the constraints on S_8, while the influence of the remaining baryonic parameters is negligible. Quantitatively, the error on S_8 grows roughly linearly with the uncertainty on $\log_{10} M_c$ and $\log_{10} \eta$, and both parameters must be calibrated at the level of $\sigma(\log_{10} M_c) \sim \sigma(\log_{10} \eta) \lesssim 0.1$ in order to avoid degrading the constraints on S_8 by more than ~ 10 per cent. When expressed in terms of the fractional uncertainty on the value of M_c and η themselves, this corresponds to a $\sim 10 - 20$ per cent calibration requirement.[1]

[1]The fractional uncertainty in a given quantity corresponds to the uncertainty in its natural logarithm, as given by $\Delta \ln(x) = \Delta x/x$. When expressed

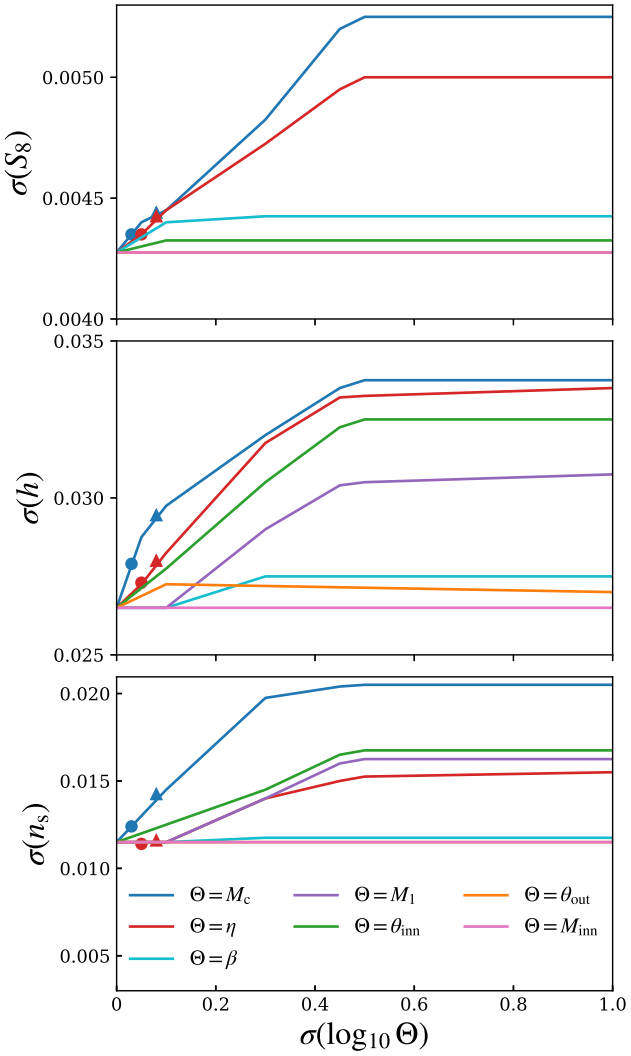

Figure 5. The errors on S_8, h, and n_s, denoted as $\sigma(S_8)$, $\sigma(h)$, and $\sigma(n_s)$, respectively, as a function of the error on each baryonic parameter, denoted as $\sigma(\log_{10}\Theta)$ where $\Theta = \{M_c, \eta, \beta, M_1, \theta_{inn}, \theta_{out}, M_{inn}\}$. Here, the triangular and circular points represent the level to which we can constrain each parameter using near-term and long-term data from external tracers of the gas distribution in and around haloes, respectively. These constraints come from X-ray and kSZ data for M_c and η, respectively.

The two bottom panels of Fig. 5 show the constraints on h and n_s. In both cases, as for S_8, the errors on the cosmological parameters are most affected by the uncertainties in $\log_{10} M_c$, with $\log_{10}\eta$ also playing a significant role (although less so in the case of n_s). Interestingly, however, we find that marginalizing over the inner scale radius, $\log_{10}\theta_{inn}$, and the stellar parameter, $\log_{10} M_1$, can significantly degrade the constraints on both h and n_s, even if this does not propagate to S_8. However, as we will see in the next section, neither of the external gas probes considered here is sufficiently sensitive to these parameters.

in terms of base-10 logarithms, this relation becomes $\Delta x/x = \ln(10) \times \Delta \log_{10}(x) \approx 2.3\, \Delta \log_{10}(x)$.

3.3 External calibrators for baryonic effects

We can use external tracers of the gas distribution in and around haloes to self-calibrate the baryonic parameters, which in turn allows for an improvement in the cosmological constraints extracted from weak lensing analyses. Here, we explore the combination of X-ray gas fraction estimates and stacked kSZ measurements, and determine whether this data will be able to match the calibration requirements described in the previous section. We will consider two types of data sets. First, we will use mock observations compatible with data that already exists or will become available in the near future, certainly within the lifetime of LSST and other Stage-IV weak lensing surveys. We will also consider more futuristic data, based on larger X-ray samples or more sensitive CMB observations. We will refer to these mock data sets as the 'near-term' and 'long-term' external calibrators.

For the near-term X-ray data set, we generate mock gas fraction measurements assuming a sample similar to the currently available data used in Grandis et al. (2024), corresponding to a total of 540 galaxy clusters from the South Pole Telescope (SPT), the Subaru Hyper Suprime-Cam combined with the XMM-Newton Ultimate Extra-galactic Survey (HSC-XXL), and the eROSITA Final Equatorial-Depth Survey (eFEDS). We use the measurement uncertainties of Chiu et al. (2018, 2022) and Akino et al. (2022), on which the Grandis et al. (2024) data is based. For near-term kSZ measurements, we assume noise levels that are compatible with what could be achieved by the Simons Observatory (SO; Ade et al. 2019). Specifically, we use the official temperature noise curves presented in Ade et al. (2019) for the 'baseline' noise level of the SO Large Aperture Telescope (LAT), assuming a 2.1 arcmin FWHM Gaussian beam. The resulting covariance matrix is estimated using equation (29), including contributions from both instrumental noise and primary CMB fluctuations. For the long-term X-ray data set, we construct mock observations based on the full eROSITA sample, which includes 5259 galaxy clusters (Ghirardini et al. 2024). As this represents approximately an order of magnitude increase in cluster count compared to the near-term mock data set, we estimate the covariance matrix for the long-term X-ray measurements by scaling the uncertainties of the near-term data set by a factor of $\sqrt{10}$. For the long-term kSZ data set, we use the noise power spectrum for the CMB-S4 experiment to generate the covariance matrix. Specifically, we assume the noise parameters listed in table 1 of Schiappucci et al. (2025) for the 90 GHz channel. However, we note that the CMB-S4 mission has unfortunately been discontinued since conducting this work. At present, the next phase of CMB observations remains uncertain. One possibility is the CMB-HD or future extensions of the Simons Observatory (Sehgal et al. 2019).

With these mock data sets at hand, we run MCMC chains assuming fixed cosmological parameters and without weak lensing data. We do so for each data set (X-ray or kSZ) separately, in order to explore their individual constraining power. The results are shown in Figs 6 and 7 for the X-ray and kSZ data sets, respectively. The light blue and dark blue contours represent the marginalized posteriors obtained from the near-term and long-term data sets, respectively.

We find that X-ray data primarily constrain $\log_{10} M_c$ and $\log_{10}\beta$, whereas $\log_{10}\eta$ remains prior-dominated. In contrast to the X-ray signal, which traces only the bound gas, the kSZ signal is sensitive to the total gas profile. This broader sensitivity allows for constraints on $\log_{10}\eta$. However, the constraints on $\log_{10} M_c$ and $\log_{10}\beta$ from kSZ data are highly degenerate, and neither parameter can be constrained to sufficient precision. This is because the kSZ mock data set assumed here probes only a single halo mass scale and redshift, which leads to a degeneracy between both parameters, since both control the

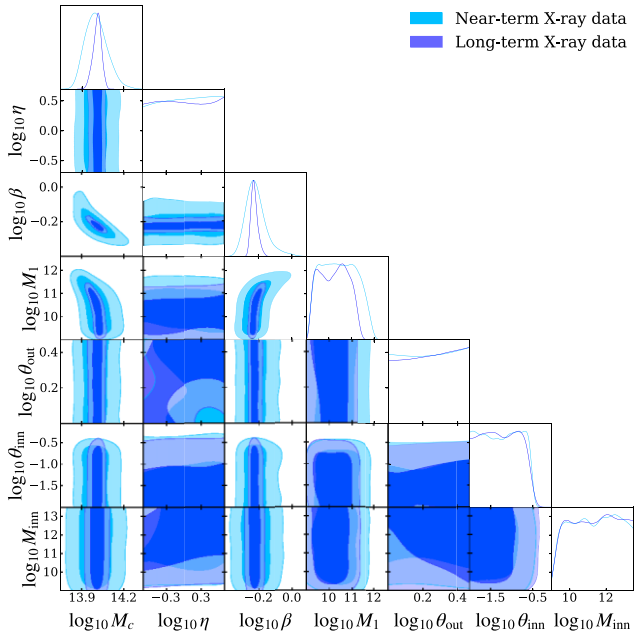

Figure 6. The marginalized posteriors on baryonic parameters obtained from mock measurements of the X-ray gas fraction for near-term data and long-term data, with the cosmological parameters kept fixed at the fiducial values used to generate the LSST-like cosmic shear data. We base the mock measurements for near-term data on 540 clusters from a combination of currently available SPT, HSC-XXL, and eROSITA data. The mock measurements for long-term data are derived assuming a sample of 5259 clusters from the full eROSITA data set.

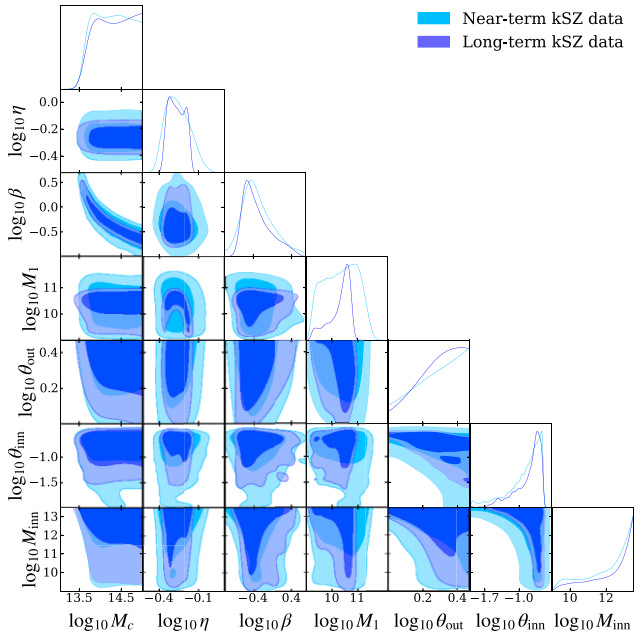

Figure 7. The marginalized posteriors on baryonic parameters obtained from mock measurements of the stacked kSZ profile for near-term data and long-term data, with the cosmological parameters kept fixed at the fiducial values used to generate the LSST-like cosmic shear data. We base the mock measurements for near-term and long-term data on the baseline noise level of the SO LAT and the CMB-S4 experiment, respectively.

abundance of bound gas. Hence, this prevents them from being individually constrained. This degeneracy is evident in the elongation of the posterior contours in the $\log_{10} M_{\rm c} - \log_{10}\beta$ panel in Fig. 7.

The 68 per cent constraints on $\log_{10} M_{\rm c}$, $\log_{10}\beta$, and $\log_{10}\eta$ are listed in Table 3. We include $\log_{10}\beta$ to highlight the sensitivity of the X-ray data to this parameter, despite its negligible effect on the degradation of cosmological constraints. We find that the posterior uncertainty on $\log_{10} M_{\rm c}$ improves by a factor of ~ 2.5 between the near-term and long-term X-ray mock data sets. The constraint on $\log_{10} M_{\rm c}$ derived from the long-term data set is sufficient to achieve the calibration requirement found in the previous section of $\sigma(\log_{10} M_{\rm c}) \lesssim 0.1$. In turn, the constraints on $\log_{10}\eta$ improve by a factor ~ 1.6 from the near-term to the long-term kSZ samples. We find that the near-term mock data set approaches the calibration requirement on $\log_{10}\eta$, while the long-term mock data set satisfies this threshold. Hence, a joint analysis of WL with the long-term kSZ data set ensures that cosmological constraints are not significantly degraded by marginalization over η, whereas combining WL with the long-term X-ray data set prevents degradation arising from uncertainties in $M_{\rm c}$. As mentioned above, these kSZ constraints could be significantly improved by assuming a more ambitious galaxy sample, covering a range of redshifts and halo masses. Note, however, that uncertainties in velocity reconstruction degrade our ability to use the total amplitude of the stacked kSZ measurements, which would affect the derived constraints on baryonic parameters (Schaan et al. 2021). We leave a more detailed study for future work.

Interestingly, none of the data sets are able to obtain sufficiently precise constraints on the other baryonic parameters, particularly $\log_{10}\theta_{\rm inn}$ or $\log_{10} M_1$, which, as we saw, can degrade the cosmological constraints on h and $n_{\rm s}$. This is not surprising in the case of $\log_{10} M_1$, since neither probe is directly sensitive to the stellar component. However, it is interesting that, although kSZ measurements are sensitive to the scale dependence of the gas profile, they are not sufficient to constrain $\log_{10}\theta_{\rm inn}$ with sufficient precision. Other external probes, such as density profiles derived from X-ray data (Grandis et al. 2024), or the cross-correlation between X-ray maps and other large-scale structure tracers (Ferreira et al. 2024; La Posta et al. 2024) could help constrain this parameter. Interestingly, we find that long-term kSZ data offers an improvement in the constraint on the stellar parameter, $\log_{10} M_1$. This sensitivity likely arises from the dependence of the baryon fraction of a given halo on the stellar mass fraction. An alternative explanation could involve the scale dependence of the kSZ signal at small radii, where stellar contributions become non-negligible. However, we exclude this possibility based on an analysis of the ratio of the kSZ profile as M_1 is varied relative to keeping M_1 fixed at its fiducial value. We find that the ratio remains effectively constant across all aperture radii examined, indicating that the constraining power is not driven by scale-dependent changes in the kSZ profile.

To summarize, based on these results (summarized in Table 3), and the requirements presented in the previous sections, we can draw the following conclusions:

(i) Near-term X-ray data, based on the currently available data reported in Grandis et al. (2024), has the potential to constrain $\log_{10} M_{\rm c}$ to the level required for calibration, with long-term data further tightening the constraints. Additionally, $\log_{10}\beta$ is already tightly constrained with near-term data, contributing to reduced uncertainties in the derived cosmological parameters.

(ii) Near-term kSZ data, based on the baseline noise level of the SO LAT, is expected to meet the calibration requirement for $\log_{10}\eta$. Long-term CMB-S4 data will offer additional constraining power

Table 3. The level to which we can constrain the baryonic parameters $\log_{10} M_c$, $\log_{10} \beta$, and $\log_{10} \eta$ using X-ray gas fractions and stacked kSZ profiles as external tracers of the large-scale structure.

Tracer	$\sigma(\log_{10} M_c)$	$\sigma(\log_{10} \beta)$	$\sigma(\log_{10} \eta)$
Near-term X-ray	0.08	0.06	prior-dominated
Long-term X-ray	0.03	0.02	prior-dominated
Near-term kSZ	prior-dominated	prior-dominated	0.08
Long-term kSZ	prior-dominated	prior-dominated	0.05

on $\log_{10} \eta$, further enhancing the overall cosmological parameter constraints. In both cases, these forecast constraints may be improved through the use of more ambitious galaxy samples, as well as refinements to velocity reconstruction.

(iii) Neither data set is able to calibrate $\log_{10} M_1$ and $\log_{10} \theta_{\rm inn}$. Other external probes may be needed to break their degeneracy with cosmological parameters (in particular h and n_s). For instance, measurements of the thermal Sunyaev–Zel'dovich (tSZ) effect have the capability to constrain $\theta_{\rm inn}$ (e.g. Aricò & Angulo 2024).

As a further investigation, we explore the effects of improvements to the galaxy catalogue. We combine the weak lensing data set with the near-term X-ray and kSZ data sets, with the galaxy number density, $n_{\rm gal}$, in the near-term kSZ data set increased by a factor of ten. Our results, displayed in the final row of Table 2, show that the error on S_8 decreases by a factor of 1.2 relative to the WL + near-term X-ray + near-term kSZ data set. Similar improvements are observed for h and n_s, with their respective errors decreasing by factors of 1.6 and 1.2.

It is important to highlight that our results are optimistic in the sense that we have not accounted for systematic uncertainties in the modelling of the external tracers. In particular, a more realistic model for the stacked kSZ profile might discard compensated aperture photometry scales below ~ 1 arcmin. Such a scale cut may be necessary to avoid small-scale modelling uncertainties (e.g. the impact of satellite galaxies). Moreover, we have interpreted the kSZ data as representing a sample of central galaxies at a fixed halo mass, consistent with the approach taken in recent kSZ analyses (Bigwood et al. 2024). To model this, we introduce $\log_{10} M_{\rm halo}$ as a free parameter with a Gaussian prior centred at the value used to generate our previous stacked kSZ measurements, $\log_{10}(M_{\rm halo}/{\rm M}_\odot) = \log_{10}(3 \times 10^{13})$. We investigate the extent to which these limitations affect the constraining power of the kSZ mock measurements for the near-term data set. Our results indicate that these systematics do not significantly degrade the constraints on η, with similar conclusions holding for the long-term data set. In particular, we find that the application of a scale cut and, removing scales $\theta_{\rm d} < 1$ arcmin, degrades the constraints on $\log_{10} \eta$ by ~ 5 per cent. Likewise, marginalizing over the sample halo mass, assuming a 10 per cent prior uncertainty, leads to a similar increase in the posterior uncertainties on $\log_{10} \eta$. The combined impact of these systematics leads to a reduction in constraining power by a factor of ~ 1.1. It is also important to note that stacked kSZ measurements are sensitive only to baryonic effects on halo masses corresponding to the host haloes of the galaxies that have been used in their measurements (Lucie-Smith et al. 2025). Consequently, incorporating data from other baryonic probes, such as X-rays and the tSZ effect, is valuable for a more comprehensive analysis.

On the X-ray side, our analysis does not include possible selection and mass measurement biases affecting X-ray clusters. We investigate this by introducing a free parameter f to represent the discrepancy between the true and observed cluster masses, with

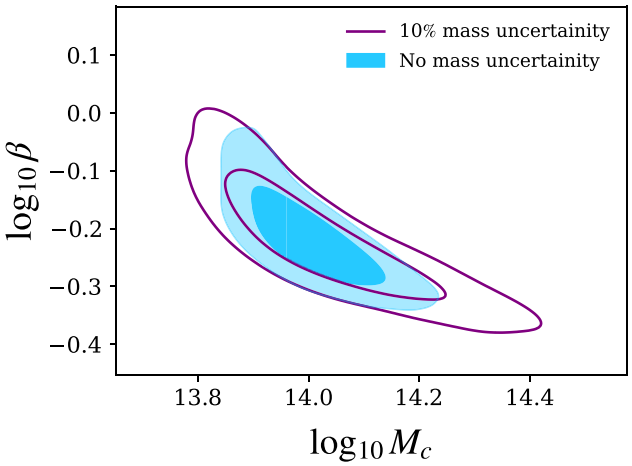

Figure 8. The constraints on $\log_{10} M_c$ and $\log_{10} \beta$ from mock measurements of the X-ray gas fraction, based on near-term data for 540 clusters from a combination of SPT, HSC-XXL, and eROSITA. Solid blue contours represent constraints in the optimistic case without marginalization over hydrostatic bias, while purple contours include marginalization over a hydrostatic bias parameter with mass uncertainties of ~ 10 per cent.

$M_i = f M_i^{\rm true}$. This is effectively equivalent to accounting for a potential hydrostatic mass bias of the form $f = 1 - b_{\rm H}$ (Salvati, Douspis & Aghanim 2018). We then include the logarithm of the measured masses of each of the cluster samples used as additional elements of the data vector, with Gaussian calibration uncertainties of $\sigma(\log_{10}(M_i/M_\odot)) = 0.05$ (corresponding to a ~ 10 per cent uncertainty in the mass measurements). Our results, presented in Fig. 8, demonstrate that incorporating a ~ 10 per cent mass measurement uncertainty degrades the constraining power of the near-term X-ray data set on $\log_{10} M_c$ by a factor of ~ 1.5. Thus, a tight control over selection effects and mass inference uncertainties is essential to obtain sufficiently precise constraints over baryonic parameters. Nevertheless, these results suggest that both kSZ and X-ray measurements remain powerful as complementary probes, even when these systematics are considered.

3.4 Cosmological constraints from cosmic shear with self-calibrated baryons

Having explored the ability of X-ray gas fractions and stacked kSZ measurements to calibrate baryonic parameters, we now derive the constraints that an LSST-like cosmic shear data set could obtain on cosmological parameters when analysed in combination with these two external tracers in order to calibrate baryonic effects in a self-consistent likelihood.

In Fig. 9, we present an updated version of Fig. 3, now including the marginalized posteriors obtained from the combined WL + X-ray + kSZ data vector. The light blue and dark blue contours in Fig. 9 show the improvement in the recovery of cosmological constraints

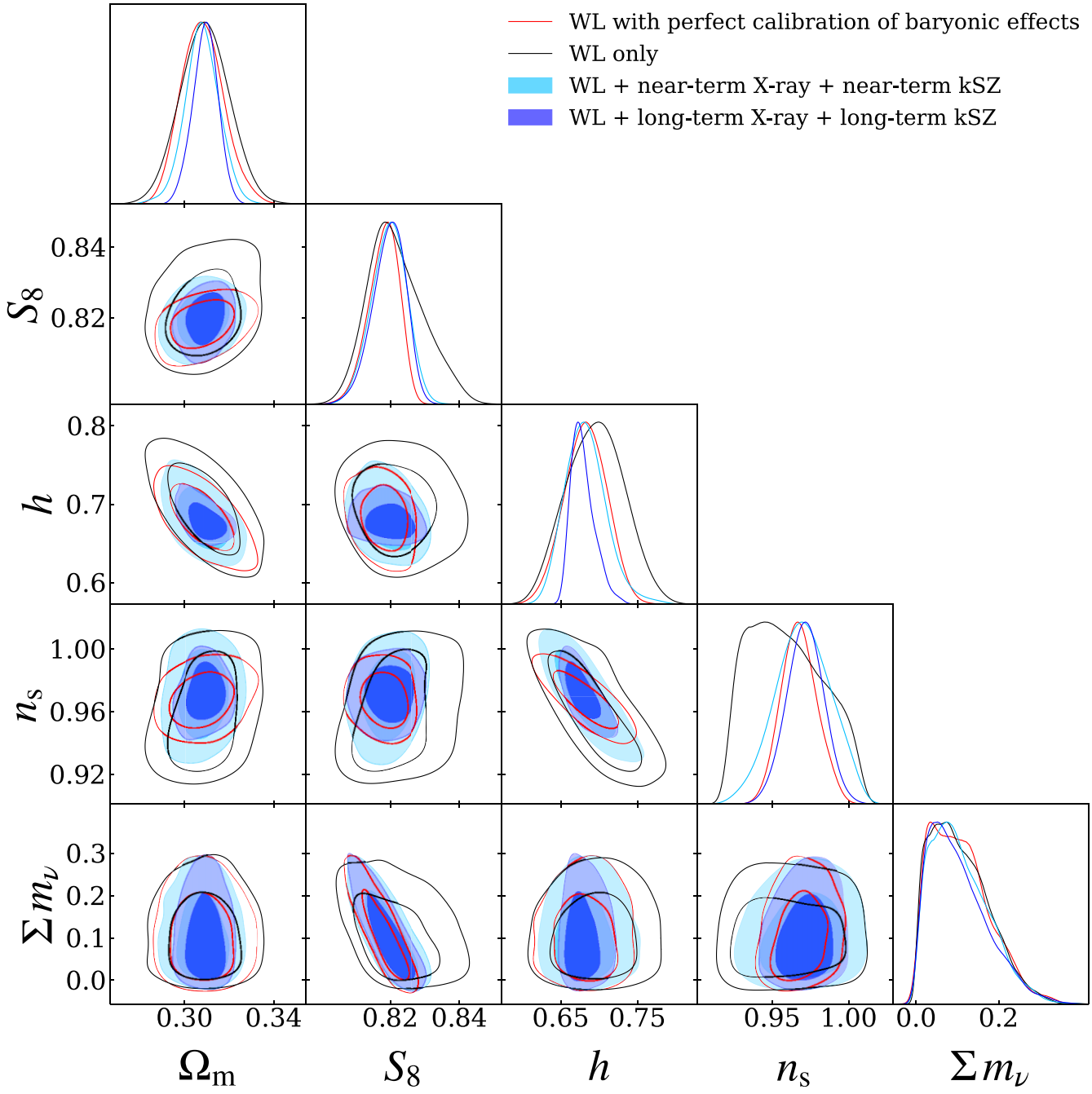

Figure 9. The marginalized posteriors for cosmological parameters for LSST-like weak lensing data only for fixed baryonic parameters (red), LSST-like weak lensing data only with marginalization over baryonic parameters (black), and a joint analysis of LSST-like weak lensing data with near-term (light blue) and long-term (dark blue) external data. The inner and outer contours show the 68 per cent and 95 per cent confidence levels, respectively. We marginalize over intrinsic alignments, photometric redshift uncertainties, and multiplicative shape biases for the weak lensing data. The near-term X-ray and kSZ mock data is based on measurements from SPT, HSC-XXL, and eROSITA and the SO baseline noise curve, respectively. The long-term X-ray and kSZ mock data is based on future forecasts of eROSITA cluster samples and the S4-CMB experiment noise curve, respectively.

when self-calibrating baryonic effects by including external data, using near-term and long-term measurements, respectively. We find that the inclusion of long-term external data sets significantly tightens the constraints on all cosmological parameters when marginalizing over baryonic effects (although, as noted in Section 3.1, the neutrino mass sum constraints from cosmic shear alone remain unaffected by baryonic effects).

In Table 2, we present the marginal mean values and the corresponding 95 per cent confidence levels for the cosmological parameters S_8, Ω_m, h, and n_s evaluated across a range of tracer combinations. Our analysis demonstrates that incorporating LSST-like weak lensing data with X-ray and kSZ observations significantly mitigates the impact of baryonic effects on cosmological constraints. Specifically, the inclusion of near-term X-ray and kSZ data reduces

the degradation factor from 1.9 to 1.2 for S_8, from 1.9 to 1.6 for n_s, and from 1.4 to 1.0 for h. Furthermore, the uncertainty on Ω_m is reduced by ~ 20 per cent. Hence, we find that near-term data is already sufficient to restore the constraint on h and Ω_m to the level achieved under the assumption of perfect knowledge of baryonic feedback. Additional gains are achieved when forecasts for long-term X-ray and kSZ measurements are incorporated. This further reduces the degradation factor to 1.1 for S_8 and n_s, while the constraints on both h and Ω_m are enhanced by a factor of 1.7 and 1.6, respectively with respect to the case of perfect calibration of baryonic effects with WL data alone. Long-term external calibrators thus have the potential to fully restore the cosmological constraining power of cosmic shear data in the presence of baryonic effects. The residual degradation in n_s and S_8 is likely due to the impact of the secondary BCM parameters θ_{inn} and M_1, controlling the baryonic suppression factor on the smallest scales, as we noted in Section 3.2. As discussed in Section 3.3, the kSZ and X-ray observations considered here are not able to constrain these parameters, and other baryonic probes, such as tSZ and X-ray cross-correlations (La Posta et al. 2024), or X-ray electron profile measurements (Grandis et al. 2024), may be used to address this shortcoming. Moreover, other cosmological probes (e.g. CMB anisotropies) can place significantly tighter constraints on n_s. These results highlight the value of integrating complementary probes of the large-scale structure in enhancing the precision of cosmological measurements.

The improvement we observe in the constraints on Ω_m and h arises from the sensitivity of the kSZ measurements to cosmology via the baryon fraction controlling its normalization, and the distance–redshift relation through the projected shape of the gas density profile. This additional constraining power may only be possible within the specific BCM parametrization used here, and thus such constraints on Ω_m and h may not be achievable when employing a more general or less restrictive model. To verify that our main result, regarding the ability of external calibrators to restore the constraining power of cosmic shear in the presence of baryonic effects, is not affected by this, we repeat our analysis fixing the cosmological parameters used to predict the X-ray and kSZ measurements, while keeping them free for the cosmic shear power spectra. In this case, the sensitivity to cosmology of these tracers is cancelled, and they serve only to self-calibrate the baryonic parameters. Our results, presented in Fig. 10, show that the improvement in the uncertainties on Ω_m and h for the long-term data set in Fig. 9 disappears, and the resulting constraints converge to those obtained with cosmic shear data assuming a perfect knowledge of baryonic effects. Meanwhile, the residual degradation in the case of S_8 and n_s observed earlier, remains in place (and in fact grows in the case of n_s).

Finally, we note that, while combining weak lensing data with a single external tracer also enhances parameter estimates, the improvements are less substantial than those obtained from a joint analysis involving multiple tracers. Thus, it will be vital to combine complementary information from different probes of gas to fully self-calibrate the impact of baryons on weak lensing observables.

4 CONCLUSIONS

Baryonic feedback is the rearrangement of the matter density within and around dark matter haloes, driven by complex hydrodynamic processes. This phenomenon is one of the most significant sources of uncertainty in current weak lensing analyses, limiting the use of small-scale data for cosmological inference and potentially contributing to the tension between the S_8 measurements made by some weak lensing experiments and CMB data. Incorporating external tracers of the gas distribution in and around haloes provides a data-driven means to constrain the parameters that describe baryonic feedback and mitigate its impact on cosmological constraints.

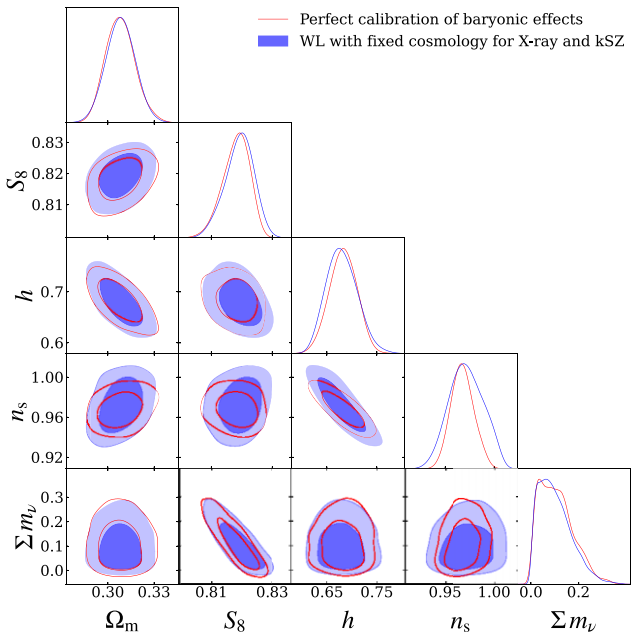

Figure 10. The marginalised posteriors for cosmological parameters derived from LSST-like weak lensing data. Results are shown for two cases: using weak lensing data alone with fixed baryonic parameters, and from a joint analysis combining weak lensing with long-term external data sets (i.e. X-ray and kSZ). In the external data sets, the cosmological parameters are fixed to the fiducial values used to generate the mock weak lensing data.

In this work, we present forecasts quantifying the precision required on baryonic parameters to recover the cosmological information that would be obtained from weak lensing data under the assumption of perfect knowledge of baryonic effects. Baryonic feedback is modelled using the BACCOEMU emulator, which implements the baryon correction model, describing the gas content around haloes after being redistributed via feedback processes. We generate LSST-like weak lensing data up to multipoles $\ell_{max} = 2000$ and examine the impact of marginalizing over baryonic parameters to account for uncertainties arising from an incomplete understanding of feedback processes.

Our analysis identifies $\log_{10} M_c$ and $\log_{10} \eta$ as the baryonic parameters most responsible for the degradation of the cosmological constraints. The parameter $\log_{10} M_c$ describe the abundance of bound, virialized gas, while $\log_{10} \eta$ characterizes the ejected gas profile. We also find that the parameters characterizing the fractional contribution from the stellar component, as well as the small-scale shape of the gas density profile, can also have a significant effect on the final constraints for future Stage-IV WL data. By marginalizing over each baryonic parameter individually within a likelihood framework, we find that the baryonic parameters must be constrained at the level of $\sigma(\log_{10} M_c) \sim \sigma(\log_{10} \eta) \lesssim 0.1$ to prevent degradation of cosmological constraints due to uncertainties associated with baryonic feedback.

Subsequently, we explore the potential of external tracers of the gas distribution to achieve the required calibration precision. Specifically, we consider measurements of the stacked kSZ profile and the bound gas fraction from X-ray observations of galaxy clusters. We find that

X-ray gas fractions based on near-term measurements already satisfy the calibration requirement on $\log_{10} M_c$ for cosmological constraints to be recovered, with long-term data expected to further tighten these constraints. Similarly, mock kSZ measurements based on near-term data achieve the necessary precision in $\log_{10} \eta$, while forecasts for upcoming experiments indicate that long-term kSZ observations will further enhance the constraints.

The joint analysis combining LSST-like weak lensing data with X-ray and kSZ mock observations yields improved cosmological constraints compared to weak lensing alone, driven by the self-consistent calibration of baryonic parameters. To illustrate this, the combination with near-term external data decreases the factor by which the error on S_8 increases from 1.9 to 1.2, with long-term external data further decreasing this factor to 1.1 (with the constraints on other parameters similarly improved). However, these results cannot be achieved by any individual gas probe. This demonstrates the potential of multitracer analyses to recover cosmological information degraded by baryonic effects. Future work could extend this framework by including additional tracers of the large-scale structure, such as the tSZ effect (Tröster et al. 2022; Aricò & Angulo 2024), fast radio bursts (Madhavacheril et al. 2019; Nicola et al. 2022; Reischke et al. 2023, 2024; Wang et al. 2025), and X-ray cross-correlations (Lau et al. 2023; Ferreira et al. 2024; La Posta et al. 2024; Lau et al. 2025).

The analysis presented here suffers from a number of shortcomings. Our description of the kSZ stacking observable is relatively simplistic, and has ignored a number of potential sources of modelling uncertainty (e.g. velocity reconstruction uncertainties, the contribution of satellites and correlated structures, as well as subdominant mass–velocity correlations), as described in Section 2.4.2. Nevertheless, we have shown that our results are largely insensitive to uncertainties in the kSZ measurements on small scales, or to moderate uncertainties in the halo mass of the galaxy sample used for kSZ stacking. We have also used a relatively limited kSZ data set, which will be improved by current and future data, by probing a range of redshifts and stellar masses. We have also not propagated uncertainties in the X-ray gas fraction measurements in our fiducial analysis, caused by selection effects or simplifying assumptions such as hydrostatic equilibrium. As we show in Section 3.3, these may have a critical impact on the ability of X-ray data to deliver the precision on baryonic parameters predicted here. The model used to describe baryonic feedback could also be easily generalized. For instance, more general BCM-like parametrizations have been proposed in the literature. Moreover, we have assumed that the BCM parameters do not have any redshift dependence (although we demonstrate in Appendix A that allowing for redshift dependence in the most relevant baryonic parameters, M_c and η, does not significantly degrade the cosmological constraints obtained here). We have also focused our analysis on the Λcold dark matter framework, and studying the impact of external gas probes on the constraints achievable on extended models (e.g. considering an evolving dark energy equation of state) could be of interest given recent results from the latest measurements of type-Ia supernovae and baryon acoustic oscillations (Vincenzi et al. 2024; DESI Collaboration 2025). Finally, our work has focused on quantifying the amount of information loss (i.e. the degradation in precision) incurred through uncertainties on baryonic effects, while ignoring the bias in the inferred cosmological parameters (i.e. the degradation in accuracy) that mismodelling this contribution could lead to. External gas probes could be used to detect inconsistencies in the physical model that would lead to these biases, although the potential to do so will depend on the sensitivity of these probes, as well as the impact of their own observational and modelling systematics (Kovač et al. 2025). Further work will be needed to address these questions.

ACKNOWLEDGEMENTS

We would like to thank Raul Angulo and Boryana Hadzhiyska for useful comments on earlier versions of this manuscript. AW was supported by a Science and Technology Facilities Council studentship. DA and MZ acknowledge support from the Beecroft Trust. We made extensive use of computational resources at the University of Oxford Department of Physics, funded by the John Fell Oxford University Press Research Fund.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding authors.

REFERENCES

Abbott T. M. C. et al., 2018, Phys. Rev. D, 98, 043526
Abbott T. M. C. et al., 2023, Open J. Astrophys., 6, 2305.17173
Adam R. et al., 2016, A&A, 596, A108
Ade P. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 056
Ahn C. P. et al., 2014, ApJS , 211, 17
Akino D. et al., 2022, PASJ, 74, 175
Alonso D., Louis T., Bull P., Ferreira P. G., 2016, Phys. Rev. D, 94, 043522
Amodeo S. et al., 2021, Phys. Rev. D, 103, 063514
Amon A., Efstathiou G., 2022, MNRAS, 516, 5355
Amon A. et al., 2022, Phys. Rev. D, 105, 023514
Amon A. et al., 2023, MNRAS, 518, 477
Angulo R. E., Zennaro M., Contreras S., Aricò G., Pellejero-Ibañez M., Stücker J., 2021, MNRAS, 507, 5869
Aricò G., Angulo R. E., 2024, A&A, 690, A188
Aricò G., Angulo R. E., Hernández-Monteagudo C., Contreras S., Zennaro M., 2021a, MNRAS, 503, 3596
Aricò G., Angulo R. E., Hernández-Monteagudo C., Contreras S., Zennaro M., 2021b, MNRAS, 503, 3596
Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021c, MNRAS, 506, 4070
Aricò G., Angulo R. E., Zennaro M., Contreras S., Chen A., Hernández-Monteagudo C., 2023, A&A, 678, A109
Ayromlou M., Nelson D., Pillepich A., 2023, MNRAS, 524, 5391
Baltz E. A., Marshall P., Oguri M., 2009, J. Cosmol. Astropart. Phys., 2009, 015
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bigwood L. et al., 2024, MNRAS, 534, 655
Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, Phys. Rev. D, 100, 103506
Bonnett C. et al., 2016, Phys. Rev. D, 94, 042005
Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002, MNRAS, 333, 501
Chen A. et al., 2023, MNRAS, 518, 5340
Chisari N. E. et al., 2019a, Open J. Astrophys., 2, 4
Chisari N. E. et al., 2019b, ApJS, 242, 2
Chiu I. et al., 2018, MNRAS, 478, 3072
Chiu I.-N. et al., 2022, A&A, 661, A11
Crain R. A., van de Voort F., 2023, ARA&A , 61, 473
DESI Collaboration, 2025, preprint (arXiv:2503.14738)
van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649
Dalal R. et al., 2023, Phys. Rev. D, 108, 123519
Eifler T. et al., 2024, preprint (arXiv:2411.04088)
Fang X. et al., 2024, MNRAS, 527, 9581
Ferreira T., Alonso D., Garcia-Garcia C., Chisari N. E., 2024, Phys. Rev. Lett., 133, 051001

Foreman S., Coulton W., Villaescusa-Navarro F., Barreira A., 2020, MNRAS, 498, 2887
Fortuna M. C. et al., 2021, A&A, 654, A76
García-García C., Zennaro M., Aricò G., Alonso D., Angulo R. E., 2024, J. Cosmol. Astropart. Phys., 08, 024
Ghirardini V. et al., 2024, A&A, 689, A298
Grandis S., Aricò G., Schneider A., Linke L., 2024, MNRAS, 528, 4379
Hadzhiyska B., Wolz K., Azzoni S., Alonso D., García-García C., Ruiz-Zapatero J., Slosar A., 2023, Open J. Astrophys., 6, 23
Hadzhiyska B. et al., 2024, preprint (arXiv:2407.07152)
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hildebrandt H. et al., 2020, A&A, 633, A69
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Ishiyama T. et al., 2021, MNRAS, 506, 4210
Ivezić v. et al., 2019, ApJ, 873, 111
Kilbinger M. et al., 2017, MNRAS, 472, 2126
Kovač M. et al., 2025, preprint (arXiv:2507.07991)
La Posta A., Alonso D., Chisari N. E., Ferreira T., García-García C., 2024, Phys. Rev. D, 112, 043525
Lange J. U., Hearin A. P., Leauthaud A., van den Bosch F. C., Xhakaj E., Guo H., Wechsler R. H., DeRose J., 2023, MNRAS, 520, 5373
Lau E. T., Bogdán Á., Chadayammuri U., Nagai D., Kraft R. P., Cappelluti N., 2023, MNRAS, 518, 1496
Lau E. T. et al., 2025, ApJ, 984, 190
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Leonard C. D. et al., 2023, Open J. Astrophys., 6, 8
Limber D. N., 1954, ApJ, 119, 655
Lucie-Smith L. et al., 2025, preprint (arXiv:2505.18258)
Madhavacheril M. S., Battaglia N., Smith K. M., Sievers J. L., 2019, Phys. Rev. D, 100, 103532
Mandelbaum R., 2018, ARA&A, 56, 393
McCarthy I. G. et al., 2023, MNRAS, 526, 5494
McCarthy I. G. et al., 2025, MNRAS, 540, 143
McCullough J. et al., 2024, preprint (arXiv:2410.22272)
Mead A., Brieden S., Tröster T., Heymans C., 2021, MNRAS, 502, 1401
Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem. Phys., 21, 1087
Miller L. et al., 2013, MNRAS, 429, 2858
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nicola A. et al., 2022, J. Cosmol. Astropart. Phys., 2022, 046
Oguri M., Hamana T., 2011, MNRAS, 414, 1851
Pakmor R. et al., 2023, MNRAS, 524, 2539
Paopiamsap A., Porqueres N., Alonso D., Harnois-Deraps J., Leonard C. D., 2024, Open J. Astrophys., 7, 34
Planck Collaboration VI, 2020, A&A, 641, A6
Prat J. et al., 2022, Phys. Rev. D, 105, 083528
Preston C., Amon A., Efstathiou G., 2024, MNRAS, 533, 621
Reischke R., Neumann D., Bertmann K. A., Hagstotz S., Hildebrandt H., 2023, preprint (arXiv:2309.09766)
Reischke R., Kovač M., Nicola A., Hagstotz S., Schneider A., 2024, Open J. Astrophys., 8, 127
Ried Guachalla B. et al., 2025, preprint (arXiv:2503.19870)
Ruiz-Zapatero J., Hadzhiyska B., Alonso D., Ferreira P. G., García-García C., Mootoovaloo A., 2023, MNRAS , 522, 5037
Salvati L., Douspis M., Aghanim N., 2018, A&A, 614, A13
Schaan E. et al., 2016, Phys. Rev. D, 93, 082002
Schaan E. et al., 2021, Phys. Rev. D, 103, 063513
Schaye J. et al., 2023, MNRAS, 526, 4978
Schiappucci E. et al., 2025, Phys. Rev. D, 111, 063541
Schneider A., Teyssier R., 2015, J. Cosmol. Astropart. Phys., 2015, 049
Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, J. Cosmol. Astropart. Phys., 2019, 020
Schneider A., Giri S. K., Amodeo S., Refregier A., 2022, MNRAS, 514, 3802
Secco L. F. et al., 2022, Phys. Rev. D, 105, 023515
Sehgal N. et al., 2019, Bull. Am. Astron. Soc., 51, 1
Siegel J. et al., 2025, preprint (arXiv:2507.11530)
Smith R. E. et al., 2003, MNRAS, 341, 1311
Sorini D., Davé R., Cui W., Appleby S., 2022, MNRAS, 516, 883
Steinborn L. K., Dolag K., Hirschmann M., Prieto M. A., Remus R.-S., 2015, MNRAS, 448, 1504
Sunseri J., Amon A., Dunkley J., Battaglia N., Ferraro S., Hadzhiyska B., Ried Guachalla B., Schaan E., 2025, preprint (arXiv:2505.20413)
Sunyaev R. A., Zeldovich Y. B., 1970, Astrophys. Space Sci., 7, 3
Sunyaev R. A., Zeldovich Y. B., 1972, Comments Astrophys. Space Phys., 4, 173
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
The LSST Dark Energy Science Collaboration, 2018, preprint (arXiv:1809.01669)
To C.-H., Pandey S., Krause E., Dalal N., Anbajagane D., Weinberg D. H., 2024, J. Cosmol. Astropart. Phys., 2024, 037
Torrado J., Lewis A., 2021, J. Cosmol. Astropart. Phys., 2021, 057
Tröster T. et al., 2022, A&A, 660, A27
Troxel M. A. et al., 2018, Phys. Rev. D, 98, 043528
Vincenzi M. et al., 2024, ApJ, 975, 86
Wang H. et al., 2025, preprint (arXiv:2506.08932)
Wright A. H. et al., 2025, preprint (arXiv:2503.19441)
Zacharegkas G. et al., 2022, MNRAS, 509, 3119

APPENDIX A: REDSHIFT DEPENDENCE OF THE BARYONIC PARAMETERS

As discussed in Section 2, the vanilla BACCOEMU model does not incorporate potential redshift-dependent variations in the baryonic parameters. To assess the impact of this limitation, we examine the effect of introducing redshift dependence in the two most influential baryonic parameters, M_c and η, on cosmological parameter constraints. Specifically, we adopt the following parametrizations to

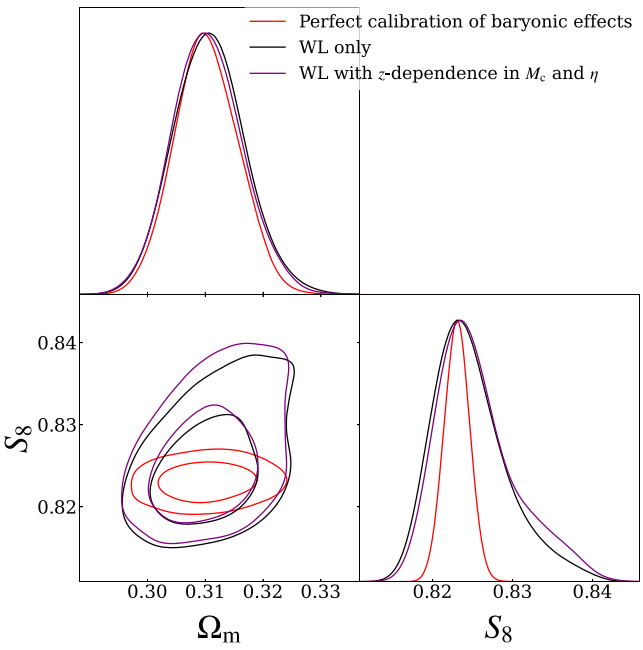

Figure A1. The marginalized posteriors of cosmological parameters derived from LSST-like weak lensing data up to multipoles $\ell = 2000$ for three cases. The red contours represent the ideal case, corresponding to perfectly calibrated baryonic effects. The black and purple contours correspond to constraints obtained by marginalizing over baryonic parameters, without and with redshift dependence in M_c and η, respectively. In each case, the inner and outer contours denote the 95 per cent and 68 per cent confidence levels, respectively. All analyses consistently marginalize over intrinsic alignments, photometric redshift uncertainties, and multiplicative shape biases.

model their redshift evolution

$$\log_{10} \widetilde{M}_{\rm c}(a) = \log_{10} M_{\rm c} + (1-a)\log_{10} M_{{\rm c},z} \tag{A1}$$

and

$$\log_{10} \widetilde{\eta}\,(a) = \log_{10} \eta + (1-a)\log_{10} \eta_z, \tag{A2}$$

where $M_{\rm c}$ and η are the original free parameters of the BACCOEMU model, and $M_{{\rm c},z}$ and η_z are additional free parameters introduced to capture potential redshift dependence.

Fig. A1 demonstrates the impact of incorporating redshift dependence in $M_{\rm c}$ and η for a simplified two-parameter cosmological model. Our results indicate that introducing redshift evolution in in $M_{\rm c}$ and η does not significantly degrade the cosmological constraints on $\Omega_{\rm m}$ and S_8 derived from weak lensing measurements.

APPENDIX B: IMPLEMENTATION OF BACCOEMU

B1 Extrapolating in z

Future cosmic shear surveys are anticipated to probe redshifts up to $z = 4.0$, however, BACCOEMU is calibrated for redshifts up to $z = 1.5$. Consequently, it is necessary to extrapolate the non-linear matter power spectrum beyond the calibration range of BACCOEMU in order to compute the angular power spectra for our LSST-like data at higher redshifts. To achieve this, we conducted BACCOEMU simulations at redshifts $z = 1.5$ and $z = 2.0$ using the same cosmological and baryonic parameters as the mock data. This allows us to compare two extrapolation methods: (a) assuming a constant boost factor, $S_k(z = 1.5)$, for $z > 1.5$, and (b) assuming a linear extrapolation of the boost factor for $z > 1.5$. Our results are presented in Fig. B1, which shows that the statistical uncertainty between the two extrapolation methods varies at most by $10^{-4}\sigma$. Hence, it is valid to assume constant extrapolation of the boost factor, $S_k(z = 1.5)$, for $z > 1.5$. We find that this result is consistent across all redshift bin correlations.

B2 Extrapolating in k

Similarly, BACCOEMU is calibrated up to wavenumbers of $k = 5\,h\,{\rm Mpc}^{-1}$, which for our fiducial cosmology, corresponds to $k \sim 3\,{\rm Mpc}^{-1}$. However, we use wavenumbers up to $k = 10\,{\rm Mpc}^{-1}$ to generate our mock data. Hence, we assess the validity of the built-in extrapolation of the boost factor, $S(k)$, in BACCOEMU for $k > 3\,{\rm Mpc}^{-1}$ by comparing it to three different approximations in the $k > 3\,{\rm Mpc}^{-1}$ regime. The approximations we considered were: (a) assuming the angular power spectrum follows the dark matter only spectrum for all k, (b) assuming the angular power spectrum follows the dark matter only spectrum for $k > 3\,{\rm Mpc}^{-1}$ (i.e. $S(k > 3\,{\rm Mpc}^{-1}) = 1$), and (c) assuming the boost function remains constant at $S(k = 3\,{\rm Mpc}^{-1})$ for $k > 3\,{\rm Mpc}^{-1}$. Our results are presented in Fig. B2. We observe that the difference between the built-in BACCOEMU extrapolation in the regime and the dark matter only power spectrum in the regime $k > 3\,{\rm Mpc}^{-1}$ is at most 0.1σ. Hence, it is valid to use BACCOEMU in the regime $k > 3\,{\rm Mpc}^{-1}$. We found that this conclusion holds consistently across all redshift bin correlations.

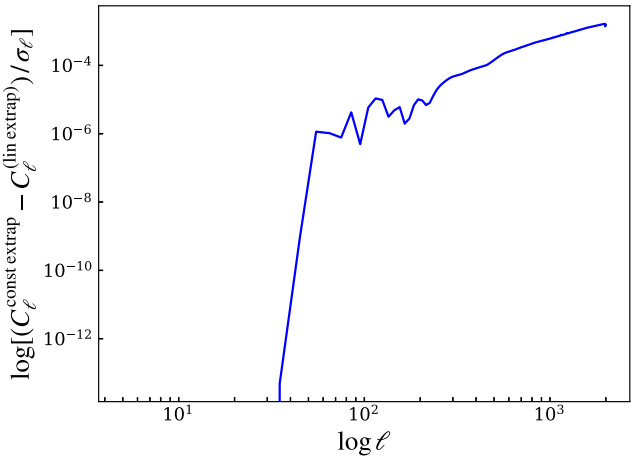

Figure B1. The difference between the angular power spectra for constant and linear extrapolation methods in the $z > 1.5$ regime for the autocorrelation of the fifth redshift bin. Here, σ_ℓ represents the square root of the diagonal of the covariance matrix for the angular power spectrum for the constant extrapolation approach.

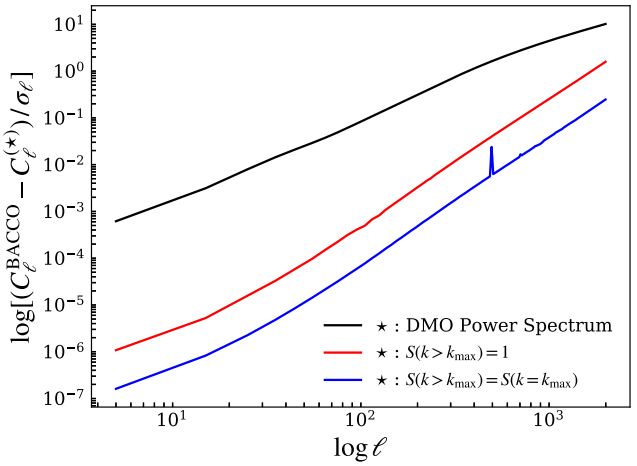

Figure B2. The angular power spectra for the different approximations in the regime $k > 3.0$, relative to the BACCOEMU extrapolated angular power spectra for the cross-correlation between the first and third redshift bins. Here, σ_ℓ represents the square root of the diagonal of the covariance matrix for the BACCOEMU extrapolated angular power spectrum.

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