A&A, 686, A190 (2024)
https://doi.org/10.1051/0004-6361/202347876
© The Authors 2024

Astronomy & Astrophysics

Methodological refinement of the submillimeter galaxy magnification bias

I. Cosmological analysis with a single redshift bin

M. M. Cueli[1,2], J. González-Nuevo[3,4], L. Bonavera[3,4], A. Lapi[1,2,5,6], D. Crespo[3,4], and J. M. Casas[3,4]

[1] SISSA, Via Bonomea 265, 34136 Trieste, Italy
e-mail: mmunizcu@sissa.it
[2] IFPU – Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy
[3] Departamento de Fisica, Universidad de Oviedo, C. Federico Garcia Lorca 18, 33007 Oviedo, Spain
[4] Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), C. Independencia 13, 33004 Oviedo, Spain
[5] IRA-INAF, Via Gobetti 101, 40129 Bologna, Italy
[6] INFN-Sezione di Trieste, Via Valerio 2, 34127 Trieste, Italy

Received 4 September 2023 / Accepted 10 March 2024

ABSTRACT

Aims. The main goal of this work is to test the results of a methodological improvement in the measurement of the magnification bias signal on a sample of submillimeter galaxies. In particular, we investigate the constraining power of cosmological parameters within the ΛCDM model. We also discuss important points that can affect the results.

Methods. We measured the angular cross-correlation function between a sample of foreground GAMA II galaxies in a single wide spectroscopic redshift bin of $0.2 < z < 0.8$ and a sample of background submillimeter galaxies from *Herschel*-ATLAS. We focused on the photometric redshift range of $1.2 < z < 4.0$, with an improved methodological framework. Interpreting the weak lensing signal within the halo model formalism and performing a Markov chain Monte Carlo (MCMC) algorithm, we obtained the posterior distribution of both the halo occupation distribution and cosmological parameters within a flat ΛCDM model. Our analysis was also performed with additional galaxy clustering information via a foreground angular auto-correlation function.

Results. We observed an overall remarkable improvement in terms of uncertainties in both the halo occupation distribution and cosmological parameters with respect to previous results. A priori knowledge about β, the logarithmic slope of the background integral number counts, is found to be paramount to derive constraints on σ_8 when using the cross-correlation data alone. Assuming a physically motivated prior distribution for β, we obtain mean values of $\Omega_m = 0.23^{+0.03}_{-0.06}$ and $\sigma_8 = 0.79^{+0.10}_{-0.10}$ and an unconstrained distribution for the Hubble constant. These results are likely to suffer from sampling variance, since one of the fields, G15, appears to have an anomalous behavior with a systematically higher cross-correlation. We find that removing it from the sample yields mean values of $\Omega_m = 0.27^{+0.02}_{-0.04}$ and $\sigma_8 = 0.72^{+0.04}_{-0.04}$ and, for the first time, a (loose) restriction of the Hubble constant is obtained via this observable: $h = 0.79^{+0.13}_{-0.14}$. The addition of the angular auto-correlation of the foreground sample in a joint analysis tightens the constraints, but also reveals a discrepancy between both observables that might be an aggravated consequence of sampling variance or due to the presence of unmodeled aspects on small and intermediate scales.

Key words. gravitation – gravitational lensing: weak – galaxies: high-redshift – dark matter – submillimeter: galaxies

1. Introduction

The submillimeter galaxy magnification bias was recently proposed as a novel approach to constrain cosmology via a weak lensing-induced cross-correlation between a foreground galaxy sample and a background set of submillimeter galaxies (Bonavera et al. 2020, 2021; González-Nuevo et al. 2021). Indeed, the phenomenon of magnification bias (see Bartelmann & Schneider 2001, and references therein) can boost the flux of background sources, while increasing the solid angle they subtend. However, imposing a flux threshold effectively creates a mismatch between the two effects, which can result in an excess of background sources around those in the foreground with respect to the absence of lensing. Although it has traditionally been deemed inferior to shear analyses for the probing of the galaxy-matter cross-correlation, the realization that submillimeter galaxies provide an optimal background sample for magnification bias studies (as shown by the very significant early detections of this cross-correlations in Wang et al. 2011; González-Nuevo et al. 2014) has turned this observable into an emerging independent cosmological probe.

The reason behind their relevance for these studies lies in the fact that submillimeter galaxies are typically located at high ($z \gtrsim 1-1.5$) redshifts (Chapman et al. 2004, 2005; Amblard et al. 2010; Lapi et al. 2011; González-Nuevo et al. 2012; Pearson et al. 2013) are faint in the optical band due to thermal emission from dust and, most importantly, have a steep luminosity function (Granato et al. 2004; Lapi et al. 2006, 2011; Eales et al. 2010). The strength of the magnification bias effect depends strongly on this last feature; indeed, the steeper the number counts, the larger the number of faint sources that may go over the detection threshold and the more likely it is that the dilution effect of magnification bias is overcome by the flux boosting.

The current concordance cosmological model has been shown to successfully reproduce a large number of

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

cosmological observations, namely, the cosmic microwave background (CMB) temperature and polarization spectra (Planck Collaboration XVI 2014; Planck Collaboration XIII 2016; Planck Collaboration VI 2020), baryon acoustic oscillation measurements (Eisenstein et al. 2005; Beutler et al. 2011; Bautista et al. 2021), the primordial abundance of light nuclei (Cyburt et al. 2016; Fields et al. 2020), or the magnitude-redshift relation from type Ia supernovae (Perlmutter et al. 1999; Brout et al. 2022; Scolnic et al. 2022; Riess et al. 2022), among many others. However, in this day and age, the necessity for additional independent cosmological probes is unquestionable. Indeed, regardless of its countless successes, the standard cosmological model is not short of both theoretical and observational challenges (Bull et al. 2016; Di Valentino & Melchiorri 2022; Abdalla et al. 2022; Perivolaropoulos & Skara 2022). Regarding the latter, a special mention should be made to the ubiquitous $\gtrsim 4\sigma$ tension between local measurements of the Hubble constant derived via a distance ladder approach (Riess et al. 2022) and the corresponding values from CMB anisotropy measurements (Planck Collaboration VI 2020). Additional inconsistencies that are worth mentioning are related, for instance, to differences in the value of the structure growth parameter, $S_8 \equiv \sqrt{\Omega_\mathrm{m}/0.3}\sigma_8$, between "direct" approaches and the CMB power spectra (Secco et al. 2022) or to the presence of anomalies in the anisotropies of the CMB (Planck Collaboration VII 2020).

Along this line, the submillimeter galaxy magnification bias has been put forward as a novel and independent cosmological probe that does not seem to suffer from the typical σ_8–Ω_m degeneracy found in other lensing observables. In particular, González-Nuevo et al. (2021) performed a first analysis and correction of the large-scale biases that could contaminate the signal and, as a consequence, the cosmological constraints. Moreover, Bonavera et al. (2021) divided up the foreground sample into four redshift bins to perform a tomographic analysis, which opened up the possibility of studying the time evolution of the dark energy equation of state. Although the Hubble constant could not be constrained, they obtained mean values of $\Omega_\mathrm{m} = 0.33^{+0.08}_{-0.16}$ and $\sigma_8 = 0.87^{+0.13}_{-0.12}$ at 68% credibility within the ΛCDM model.

The present work, which is intended to be released along another companion paper, aims to build upon the aforementioned results, along with a refinement in terms of the methodology. Here, we address the cosmological and halo occupation distribution (HOD) constraints that can be derived using a single broad foreground redshift bin and updated cross-correlation data. The computation of the theoretical model for the signal is also revisited with respect to Bonavera et al. (2020) and González-Nuevo et al. (2021), assessing the importance of the value of the logarithmic slope of the background galaxy number counts and of a numerical correction regarding the computation of the non-linear power spectra. In Bonavera et al. (2024), to be referred to as Paper II, the analysis is extended to a tomographic setup, where the foreground sample is split into different redshift bins. The dependence on the number of redshift bins and their range is discussed along with the possible improvements with respect to the use of a single broad bin.

The study carried out in this work uses the measurement of the angular cross-correlation function between a sample of background submillimeter galaxies from H-ATLAS (Pilbratt et al. 2010; Eales et al. 2010) and a sample of foreground galaxies from GAMA II (Driver et al. 2011; Baldry et al. 2010, 2014; Liske et al. 2015). Assuming a flat ΛCDM cosmology, we performed a Markov chain Monte Carlo (MCMC) analysis to derive the posterior probability distribution of both HOD and cosmological parameters. Additionally, we include the information about the clustering of the foreground sample and discuss the importance of the steepness of the submillimeter galaxy number counts.

This paper has been structured as follows. Section 2 lays out the theoretical background of this work. We discuss the physical origin of the weak lensing-induced foreground-background cross-correlation and how it is computed within the halo model formalism. The methodology followed in the paper is described in Sect. 3: the galaxy samples are detailed, along with the estimation procedure for both the angular cross- and auto-correlation functions and the MCMC setup to sample the posterior probability distribution of the parameters involved in each of the cases we studied. Section 4 presents the results we obtained and Sect. 5 summarizes our conclusions.

2. Theoretical framework

2.1. Galaxy clustering and the cross-correlation induced by magnification bias

As noted in the introduction, the weak lensing-induced foreground-background number cross-correlation is the main observable of this paper. However, a joint analysis together with the clustering of the foreground galaxy sample was also carried out in addition to study the potential tightening of the parameter constraints. Therefore, we proceed to describe the theoretical modeling of both observables below.

Under the well-known Limber (1953) and flat-sky approximations, the foreground angular auto-correlation function is given by

$$w_\mathrm{ff}(\theta) = \int_0^\infty \mathrm{d}z\, \frac{H(z)}{c} \left[\frac{\mathrm{d}N_\mathrm{f}/\mathrm{d}z}{\chi(z)}\right]^2 \int_0^\infty \frac{\mathrm{d}l}{2\pi}\, l\, P_\mathrm{gg}(l/\chi(z), z)\, J_0(l\theta), \tag{1}$$

where P_gg is the galaxy power spectrum, $H(z)$ is the Hubble parameter at redshift z, $\chi(z)$ is the comoving distance at redshift z, $\mathrm{d}N_\mathrm{f}/\mathrm{d}z$ is the normalized foreground source distribution, and J_0 is the zeroth-order Bessel function of the first kind.

Moreover, the phenomenon of magnification bias, central to this work, probes the galaxy-mass correlation via the foreground-background angular cross-correlation function. Indeed, in the presence of lensing, the phenomenon of magnification bias produces fluctuations in the number density of background sources that, in the weak-lensing regime, can be expressed as (Bartelmann & Schneider 2001):

$$\delta n_\mathrm{b}^\mu(\theta) \approx 2(\beta - 1)\,\kappa(\theta),$$

where β is the logarithmic slope of the unlensed background number counts[1] and κ is the effective convergence field. Since the foreground sources trace the underlying matter field, their fluctuations are due to pure clustering, that is,

$$\delta n_\mathrm{f}^c(\varphi) = \int_0^\infty \mathrm{d}z\, \frac{\mathrm{d}N_\mathrm{f}}{\mathrm{d}z}\, \delta_\mathrm{g}(\varphi, z).$$

Therefore, for two galaxy samples with nonoverlapping redshift distributions, the only non-negligible contribution to the

[1] The intrinsic integral number counts of the background sources are assumed to be described by a power law in a neighborhood of the detection limit: $n_\mathrm{b}(>S) = A\, S^{-\beta}$.

foreground-background angular cross-correlation comes from the two above terms: $w_{\rm fb}(\theta) \equiv \langle \delta n_{\rm f}^{\rm c}(\varphi)\, \delta n_{\rm b}^{\mu}(\varphi+\theta)\rangle_{\varphi}$. Once again, using the Limber and flat-sky approximations, this can be expressed as (Cooray & Sheth 2002):

$$w_{\rm fb}(\theta) = 2(\beta-1)\int_0^\infty \frac{{\rm d}z}{\chi^2(z)}\frac{{\rm d}N_{\rm f}}{{\rm d}z}W^{\rm lens}(z) \times \int_0^\infty {\rm d}l\,\frac{l}{2\pi}P_{\rm g\text{-}m}(l/\chi(z),z)J_0(l\theta), \quad (2)$$

where

$$W^{\rm lens}(z) \equiv \frac{3}{2}\frac{1}{c^2}\left[\frac{H(z)}{1+z}\right]^2\int_z^\infty {\rm d}z'\frac{\chi(z)\chi(z'-z)}{\chi(z')}\frac{{\rm d}N_{\rm b}}{{\rm d}z'}.$$

Here, $P_{\rm g\text{-}m}$ is the galaxy-matter cross-power spectrum and ${\rm d}N_{\rm b}/{\rm d}z$ is the normalized background source distribution.

2.2. Modeling of the non-linear power spectra

According to the halo model of structure formation, the non-linear galaxy, matter and galaxy-matter power spectra can be computed analytically if the following ingredients are provided: the halo mass function, $n(M,z)$, the deterministic linear halo bias, $b_1(M,z)$, the halo density profile, $\rho(r)$, the linear matter power spectrum, $P^{\rm lin}(k,z)$, and the HOD. For instance, the galaxy power spectrum can be expressed as:

$$P_{\rm gg}(k,z) = P_{\rm gg}^{\rm 1h}(k,z) + P_{\rm gg}^{\rm 2h}(k,z),$$

where $P_{\rm gg}^{\rm 1h}$ and $P_{\rm gg}^{\rm 2h}$ are known as the one-halo and two-halo terms and account for galaxy correlations within single halos and among different halos, respectively. They can be written as

$$P_{\rm gg}^{\rm 1h}(k,z) = \int {\rm d}M\, n(M,z)\,\frac{\langle N_{\rm c}\rangle_{\rm M}\langle N_{\rm s}\rangle_{\rm M}}{\bar{n}_{\rm g}(z)^2}|u(k|M,z)| + \int {\rm d}M\, n(M,z)\,\frac{\langle N_{\rm s}(N_{\rm s}-1)\rangle_{\rm M}}{\bar{n}_{\rm g}(z)^2}|u(k|M,z)|^2$$

and

$$P_{\rm gg}^{\rm 2h}(k,z) = P^{\rm lin}(k,z)\left[\int {\rm d}M\, n(M,z)\, b_1(M,z)\frac{\langle N\rangle_{\rm M}}{\bar{n}_{\rm g}(z)}|u(k|M,z)|\right]^2,$$

where $\bar{n}_{\rm g}(z)$ is the mean number density of galaxies at redshift z, $u(k|M,z)$ is the normalized Fourier transform of the density profile of a typical halo of mass M at redshift z and $\langle N\rangle_{\rm M}$ is the first moment of the HOD, that is, the mean number of galaxies in a halo of mass M, which has been split into the contributions of central and satellite galaxies, expressed as $\langle N_{\rm c}\rangle_{\rm M}$ and $\langle N_{\rm s}\rangle_{\rm M}$, respectively.

Regarding the galaxy-matter cross-power spectrum, the halo model prescription is expressed as:

$$P_{\rm g\text{-}m}(k,z) = P_{\rm g\text{-}m}^{\rm 1h}(k,z) + P_{\rm g\text{-}m}^{\rm 2h}(k,z),$$

where

$$P_{\rm g\text{-}m}^{\rm 1h}(k,z) = \int_0^\infty {\rm d}M\, M\frac{n(M,z)}{\bar{\rho}_0}\frac{\langle N_{\rm c}\rangle_{\rm M}}{\bar{n}_{\rm g}(z)}|u(k|M,z)| + \int_0^\infty {\rm d}M\, M\frac{n(M,z)}{\bar{\rho}_0}\frac{\langle N_{\rm s}\rangle_{M}}{\bar{n}_{\rm g}(z)}|u(k|M,z)|^2$$

and

$$P_{\rm g\text{-}m}^{\rm 2h}(k,z) = P(k,z)\left[\int_0^\infty {\rm d}M\, M\frac{n(M,z)}{\bar{\rho}_0}b_1(M,z)u(k|M,z)\right] \times \left[\int_0^\infty {\rm d}M\frac{n(M,z)}{\bar{n}_{\rm g}(z)}b_1(M,z)\left(\langle N_{\rm c}\rangle_{\rm M} + \langle N_{\rm s}\rangle_{\rm M}\, u(k|M,z)\right)\right].$$

Although a detailed description about the ingredients of the model is given in Appendix A, it suffices to say here that we have chosen the Sheth and Tormen model for the halo mass function and the corresponding linear halo bias derived via the peak-background split (Sheth & Tormen 1999), the NFW model for the halo density profile (Navarro et al. 1997), the usual power-law primordial power spectrum, and the three-parameter HOD model of Zehavi et al. (2005). For this, we have:

$$\langle N\rangle_{\rm M} = \langle N_{\rm c}\rangle_{\rm M} + \langle N_{\rm s}\rangle_{\rm M} = \left[1+\left(\frac{M}{M_1}\right)^{\alpha}\right]\Theta(M-M_{\rm min}). \quad (3)$$

Therefore, within the halo model prescription, the galaxy and galaxy-matter power spectra depend both on cosmology and the parameters of the HOD. The corresponding angular auto- and cross-correlation functions inherit this dependence via Eqs. (1) and (2), where an additional cosmological influence is present, as well as the additional parameter β in the case of the latter, as we discuss in Sect. 4.1.1.

These two last comments should be made before proceeding regarding the computation of the model with respect to previous works. Firstly, given the large number of integrals that have to be carried out for each angular value, we have resorted to a mean-redshift approximation in which the outermost integrals in Eqs. (1) and (2) are not computed directly for all the redshift distribution, but through evaluation at the mean redshift of the sample due to the reduction of computational time by a factor of 10, so that:

$$w_{\rm fb}(\theta) \approx 2(\beta-1)\frac{W^{\rm lens}(\bar{z})}{\chi^2(\bar{z})}\int_0^\infty {\rm d}l\,\frac{l}{2\pi}P_{\rm g\text{-}dm}(l/\chi(\bar{z}),\bar{z})J_0(l\theta).$$

The validity of this approximation ought to be proven in the first MCMC run, so that all subsequent analyses can be safely computed with it, which speeds up the computations dramatically.

Secondly, a crucial point should be raised regarding the computation of the two-halo term of the galaxy-matter cross-power spectrum within the halo model. As discussed by Mead et al. (2020) and Mead & Verde (2021), the evaluation of the corresponding integral poses a numerical problem, since typical halo mass functions assign a large fraction of mass to low-mass halos; this causes convergence of the integral to be extremely slow and to bias the large-scale behavior, which no longer reflects the typical linear regime. Although a more detailed analysis is made in Appendix B, where the correction procedure is explained, we stress here that overlooking this issue can induce a very strong bias on the cosmological results from a halo modeling of the signal. We have implemented the above correction for all cases in this paper.

3. Data and methodology

3.1. The foreground and background galaxy samples

The foreground and background galaxy samples have been extracted from the GAMA II (Driver et al. 2011; Baldry et al. 2010, 2014; Liske et al. 2015) and H-ATLAS (Pilbratt et al. 2010; Eales et al. 2010) surveys, respectively. Their common

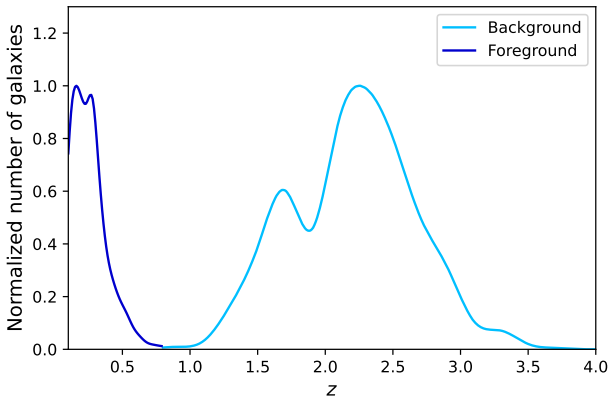

Fig. 1. Normalized redshift distribution of the background (light blue) and foreground (dark blue) samples of galaxies.

area covered three regions on the celestial equator at 9, 12 and 14.5 h (named G09, G12 and G15) and part of the south galactic pole (SGP) region, which amounts to a total of ∼207 deg^2. These are the same samples used in González-Nuevo et al. (2017, 2021) and Bonavera et al. (2020), where a more detailed discussion of the selection procedure can be found.

In essence, the foreground sample is made up of GAMA II sources in the common region with H-ATLAS with spectroscopic redshifts in the range $0.2 < z < 0.8$, resulting in ∼130 000 galaxies with a median redshift of 0.28 and surveyed down to a magnitude of ∼19.8 in the r band. Figure 1 (in dark blue) depicts the associated redshift distribution.

The background sample is made up of ∼37 000 H-ATLAS sources in the common area, obtained via a photometric redshift selection of $1.2 < z < 4.0$ to ensure that there is no overlap with the foreground galaxies. The redshift estimation procedure, which is thoroughly described in González-Nuevo et al. (2017) and Bonavera et al. (2019), consists of a χ^2 fit of the photometry to the spectral energy distribution of SMM J2135−0102 ("the Cosmic Eyelash"; Ivison et al. 2010; Swinbank et al. 2010), a gravitationally-lensed submillimeter galaxy at $z = 2.3$ that was shown to provide the best overall fit to H-ATLAS data (Lapi et al. 2011; González-Nuevo et al. 2012; Ivison et al. 2016). The redshift distribution of the background sample, taking into account the effect of random errors, is depicted in light blue in Fig. 1.

3.2. Measurements and methodological aspects

The scanning strategy employed by the H-ATLAS survey resulted in slightly overlapping rhomboidal shapes (or "tiles") in most fields, each of them covering about 16 deg^2. An average using this natural division, along with a subdivision into "minitiles" (i.e., one fourth of a tile) was analyzed in the detailed analysis of González-Nuevo et al. (2021). Indeed, a cross-correlation measurement averaged over all minitiles was concluded to be the most robust method, washing out large-scale inhomogeneities and needing only the so-called integral constraint (IC) correction. The use of minitiles was, in fact, the default strategy in previous related works (Bonavera et al. 2020, 2021; Cueli et al. 2021, 2022). However, as detailed in Appendix C, this approach could bias the data and, ultimately, the cosmological parameter constraints.

Therefore, and as customary in galaxy clustering studies, we now explore the possibility of performing a sole measurement of the cross-correlation function using all the available area, a methodology that should be free of IC biases given the scales probed. Although the minitiles strategy is no longer used for the measurement itself, an analogous subdivision of the whole area into minimal subregions is still necessary to assign meaningful uncertainties via internal covariance estimation. To define the subregions, we drew inspiration from TreeCorr (Jarvis 2015), a popular software package for computing two-point correlation functions. TreeCorr uses a k-means clustering algorithm to partition the data into subregions known as "patches", which are similar to our minitiles. Specifically, we adopted the k-means algorithm provided by the SciPy library, which aims to minimize the sum of the squared distances between data points and their assigned patch centroid. We determined the number of patches by imposing a minimum area for each of them and introduced an additional step by repeating the procedure ten times with different random initial centroids and selecting the case yielding the most consistent number of data points across patches. Figure 2 shows the distribution of foreground (top panel) and background (bottom panel) sample galaxies for the G15 field. A choice of (approximately equal-area) patches for the computation of the covariance matrix (as explained below) is also depicted.

The angular auto-correlation function of the foreground sample is measured over the available area through the standard Landy & Szalay (1993) estimator:

$$\hat{w}_{\rm auto}(\theta) = \frac{D_{\rm f}D_{\rm f}(\theta) - 2D_{\rm f}R_{\rm f}(\theta) + R_{\rm f}R_{\rm f}(\theta)}{R_{\rm f}R_{\rm f}(\theta)},$$

where $D_{\rm f}D_{\rm f}(\theta)$, $D_{\rm f}R_{\rm f}(\theta)$, and $R_{\rm f}R_{\rm f}(\theta)$ denote the normalized foreground-foreground, foreground-random and random-random pair counts at an angular separation of θ, computed in practice using equally spaced logarithmic bins. The random catalog is generated from mock random positions for ten times the number of foreground sources. The measurements are shown in black in the upper part of Fig. 3.

In turn, the foreground-background angular cross-correlation is computed over the available area via the natural modification of the above estimator (Herranz 2001):

$$\hat{w}_{\rm cross}(\theta) = \frac{D_{\rm f}D_{\rm b}(\theta) - D_{\rm f}R_{\rm b}(\theta) - D_{\rm b}R_{\rm f}(\theta) + R_{\rm f}R_{\rm b}(\theta)}{R_{\rm f}R_{\rm b}(\theta)},$$

where $D_{\rm f}D_{\rm b}(\theta)$, $D_{\rm f}R_{\rm b}(\theta)$, $D_{\rm b}R_{\rm f}(\theta)$ and $R_{\rm f}R_{\rm b}(\theta)$ denote the normalized foreground-background, foreground-random, background-random and random-random pair counts at an angular separation of θ.

The cross-correlation measurements are also depicted in black in Fig. 3, where they are compared with the ones used by González-Nuevo et al. (2021), shown in gray[2]. As expected, the cross-correlation signal is much weaker than the auto-correlation. The new data reach larger angular scales and have smaller statistical uncertainties due to the different and more efficient measurement methodology. As concluded in Bonavera et al. (2020), these aspects are expected to lead to an improvement in the cosmological constraining power. It should be pointed out that unlike the data from González-Nuevo et al. (2021), the cross-correlation signal at large angular scales does not seem to die off particularly steeply. This behavior (and its consequences on the cosmological estimates) are analyzed in Sect. 4.1.2.

The covariance matrix is estimated for both observables through a Bootstrap method, which involves dividing the whole

[2] It should be noted that the measurements from González-Nuevo et al. (2021) were performed using the average-over-minitiles strategy.

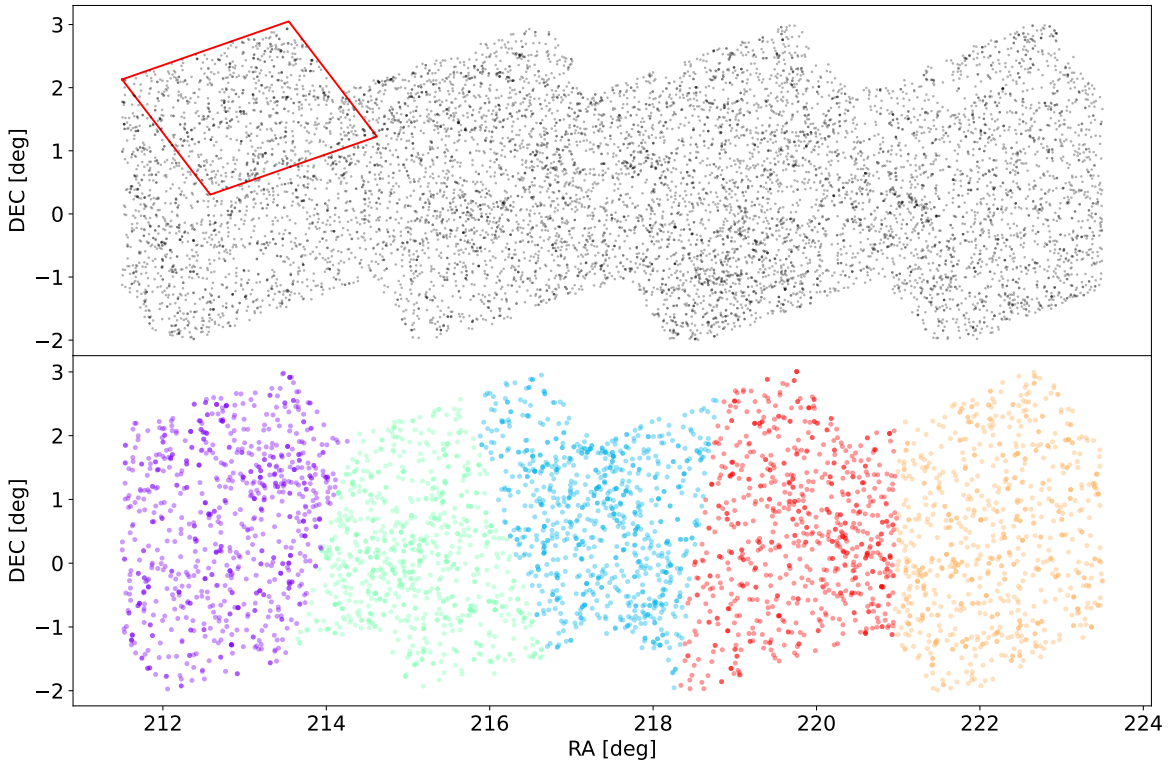

Fig. 2. Angular distribution of foreground (top panel) and background (bottom panel) sample galaxies in the G15 field. In the bottom panel, different colours indicate the definition of the patches used for the Bootstrap error estimation of the angular cross- and auto-correlations. The red square in the top panel indicates the typical shape and size of a minitile, used in previous works to divide the sample into subregions (see text for more details). The number density has been artificially reduced in both panels for loading and visualization purposes.

common area into N subregions, resampling N_r of them with replacement and repeating the process N_b times. In essence, a random integer from 0 to N is assigned to each subregion with the condition that all of them add up to N_r, effectively constructing a Bootstrap sample from the existing data on which one performs a measurement. This procedure is repeated N_b times, each one with different assignments of random integers to each subregion. The covariance matrix is then computed via:

$$\mathrm{Cov}(\theta_i, \theta_j) = \frac{1}{N_\mathrm{b} - 1} \sum_{k=1}^{N_\mathrm{b}} \Big[\hat{w}_k(\theta_i) - \bar{\hat{w}}(\theta_i)\Big]\Big[\hat{w}_k(\theta_j) - \bar{\hat{w}}(\theta_j)\Big], \quad (4)$$

where $\hat{w}_k$ denotes the measured correlation function from the kth Bootstrap sample and $\bar{\hat{w}}$ is the corresponding average value over all Bootstrap samples.

Regarding the choice of N_r, that is, the number of subregions to be drawn with replacement for each Bootstrap sample, we follow the conclusions of Norberg et al. (2009) and let $N_\mathrm{r} = 3N$, for which they obtained a very good agreement between the Bootstrap errors and those derived with an external estimate. To reach a compromise between the largest scales probed and the fact that we have 13 data points, we chose $N = 20$, that is, we divided up each independent field into five patches. The procedure was repeated $N_\mathrm{b} = 10\,000$ times. Lastly, it should be noted that our internal approach does not take super-sample covariance (Lacasa & Kunz 2017) into account, since our dominant source of uncertainty is currently the sampling variance between the different fields used for the measurements (see Sect. 4.1.2).

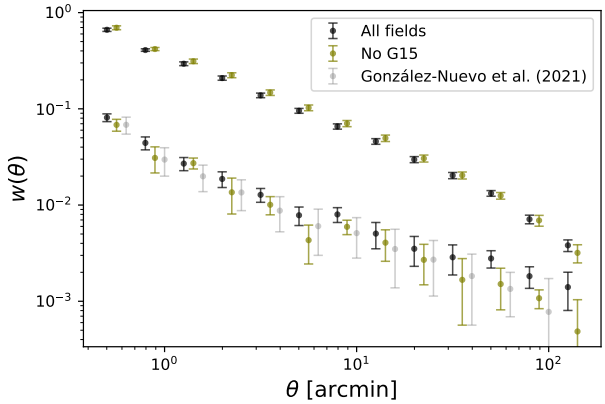

Fig. 3. Measurements of the foreground auto-correlation function and the foreground-background cross-correlation function (in black) compared to the cross-correlation function excluding the G15 region (in olive green). The cross-correlation data from González-Nuevo et al. (2021) are shown in gray.

3.3. Parameter estimation

For the purposes of this paper, the free parameters we consider here are: M_min, M_1, α, Ω_m, σ_8, h and β. A flat ΛCDM cosmology is assumed throughout the paper, with Ω_b and n_s fixed to the latest *Planck* values (Planck Collaboration VI 2020). For the estimation procedure, a Bayesian statistical approach is followed, for which the sampling of the posterior distributions was carried out via an MCMC algorithm using the open-source *emcee* software package (Foreman-Mackey et al. 2013), a

Table 1. Parameter prior distributions and summarized posterior results from the MCMC runs on the cross-correlation function with uniform and Gaussian priors on the β parameter.

	Uniform β				Gaussian β			
Parameter	Prior	Mean	Mode	68% CI	Prior	Mean	Mode	68% CI
α	$\mathcal{U}[0.00, 1.50]$	0.83	0.64	[0.45, 1.24]	$\mathcal{U}[0.00, 1.50]$	0.72	0.64	[0.32, 1.04]
$\log M_{\rm min}$	$\mathcal{U}[10.00, 16.00]$	11.67	11.73	[11.48, 11.94]	$\mathcal{U}[10.00, 16.00]$	11.47	11.54	[11.34, 11.69]
$\log M_1$	$\mathcal{U}[10.00, 16.00]$	13.41	13.57	[12.60, 14.23]	$\mathcal{U}[10.00, 16.00]$	13.04	12.94	[12.02, 13.71]
$\Omega_{\rm m}$	$\mathcal{U}[0.10, 1.00]$	0.21	0.20	[0.16, 0.24]	$\mathcal{U}[0.10, 1.00]$	0.23	0.21	[0.17, 0.26]
σ_8	$\mathcal{U}[0.60, 1.20]$	0.95	–	[0.87, 1.20]	$\mathcal{U}[0.60, 1.20]$	0.79	0.79	[0.69, 0.89]
h	$\mathcal{U}[0.50, 1.00]$	0.70	0.65	[0.50, 0.75]	$\mathcal{U}[0.50, 1.00]$	0.72	–	[0.50, 0.80]
β	$\mathcal{U}[1.50, 3.50]$	2.46	2.16	[1.81, 2.86]	$\mathcal{N}[2.90, 0.04]$	2.90	2.91	[2.86, 2.94]

Python-based implementation of the Goodman and Weare affine invariant MCMC ensemble sampler (Goodman & Weare 2010).

Two main cases are distinguished in the analysis. The first one deals only with the angular cross-correlation function and assesses the parameter constraints that can be derived from its observation alone. The corresponding log-likelihood function can be described as a multivariate Gaussian, that is:

$$\log \mathcal{L}_{\rm cross}(\theta_1, \ldots, \theta_m) = -\frac{1}{2}\Big[m \log(2\pi) + \log|C_{\rm cross}| + \boldsymbol{\varepsilon}_{\rm cross}^{\rm T} C_{\rm cross}^{-1} \boldsymbol{\varepsilon}_{\rm cross}\Big],$$

where $\boldsymbol{\varepsilon}_{\rm cross} \equiv [\varepsilon_{\rm cross}(\theta_1), \ldots, \varepsilon_{\rm cross}(\theta_m)]$,

$$\varepsilon_{\rm cross}(\theta_i) \equiv w_{\rm fb}(\theta_i) - \hat{w}_{\rm cross}(\theta_i) \qquad \forall i \in \{1, \ldots, m\}$$

and $C_{\rm cross}$ is the covariance matrix associated to the cross-correlation measurements, computed according to Eq. (4).

The second main case deals with a joint analysis of the foreground-background cross-correlation and the foreground-foreground auto-correlation. The corresponding log-likelihood will be expressed as:

$$\log \mathcal{L}_{\rm auto+cross}(\theta_1, \ldots, \theta_m) = -\frac{1}{2}\Big[m \log(2\pi) + \log|C| + \boldsymbol{\varepsilon}^{\rm T} C^{-1} \boldsymbol{\varepsilon}\Big],$$

where $\boldsymbol{\varepsilon} \equiv [\varepsilon_{\rm cross}(\theta_1), \ldots, \varepsilon_{\rm cross}(\theta_m), \varepsilon_{\rm auto}(\theta_1), \ldots, \varepsilon_{\rm auto}(\theta_m)]$,

$$\varepsilon_{\rm auto}(\theta_i) \equiv w_{\rm ff}(\theta_i) - \hat{w}_{\rm auto}(\theta_i) \qquad \forall i \in \{1, \ldots, m\}$$

and C is the full covariance matrix associated to the cross- and auto-correlation measurements, again computed according to Eq. (4).

4. Results

4.1. Cross-correlation analysis

4.1.1. The β parameter

The preliminary cosmological and astrophysical results derived in Bonavera et al. (2020) and González-Nuevo et al. (2021) had assumed a fixed value of $\beta = 3$, on account of the results found by Lapi et al. (2011), where the observed number counts of high-redshift *Herschel* submillimeter galaxies had been successfully reproduced using an updated version of the galaxy formation model by Lapi et al. (2006). Indeed, high-redshift submillimeter galaxies were interpreted as massive protospheroids and the logarithmic slope of their intrinsic (unlensed) number counts at 350 μm was predicted to be near 3 at the 3σ detection limit. However, the exact value of β that should be used is not straightforward, since the behavior of the counts around the detection limit needs to be taken into account and the minimum flux that can be statistically boosted above the threshold cannot be predicted. Nonetheless, an analysis of the predicted integral number counts according to the above model allowed us to derive an average of the β values in a sensible neighborhood of the detection limit. In essence, this translates into the possibility of considering a plausible prior for β consisting of a Gaussian distribution with mean equal to 2.90 and standard deviation equal to 0.04.

Notwithstanding this analysis, we assess the importance of prior information on β by studying four different cases in this first subsection assuming several choices for the β parameter: a uniform prior distribution between 1.5 and 3.5, the aforementioned Gaussian prior distribution with mean 2.9 and standard deviation 0.04, and fixed values of 2.2 and 3.0. The last two cases are considered for different reasons: the first to assess large deviations from usual values and the second to make a comparison with previous results.

However, as noted at the end of Sect. 2, the validity of the mean-redshift approximation needs to be tested before it can be adopted for all MCMC runs. We started by carrying out an analysis with the most general case (uniform prior distribution on β) to assess if there could be any important deviation from the full model. The resulting corner plot is depicted in Fig. D.1 and the summarized statistical results are shown in Table D.1. As can be clearly seen, only minor differences are present and we can safely adopt it for the rest of the paper. Moreover, this approximation is even more accurate in the tomographic setup of Paper II, where the model is evaluated within narrower redshift bins.

With regard to the four main MCMC runs of this first subsection, the corresponding full corner plots are shown in Fig. D.2 in red, green, blue, and cyan for the cases of a uniform, Gaussian, and fixed 2.2. and 3.0 values for the β priors, respectively, and the numerical results are summarized[3] in Tables 1 and D.2. For visual purposes, the marginalized posterior distributions of all parameters are depicted in Fig. 4.

A preliminary global view shows, as expected, that the case with the weakest priors (uniform distribution on β) yields the least stringent constraints on most parameters. However, the marginalized distributions of $M_{\rm min}$, M_1, and σ_8 have a distinctive feature which sheds light onto the influence of the β parameter;

[3] Throughout the paper, all halo masses are expressed in $M_\odot\, h^{-1}$.

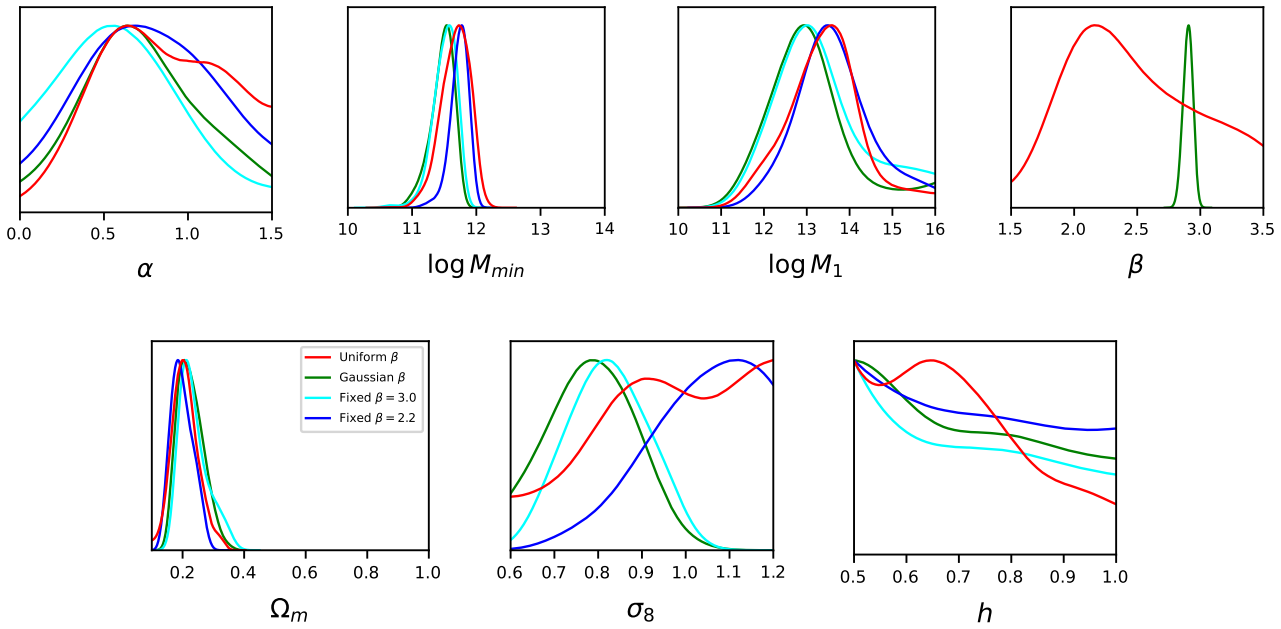

Fig. 4. Marginalized posterior distribution of the HOD, β (top panels) and cosmological (bottom panels) parameters for the different cases discussed in Sect. 4.1.1, i.e., a uniform prior (in red), a Gaussian prior (in green), and fixed values on β of 2.2. and 3.0 (in blue and cyan, respectively).

indeed, two different peak values for these parameters appear according to the two seemingly plausible values of β, which are around 2.2 and a 3.0. In particular, for σ_8, the interplay with the β parameter becomes clear on the σ_8–β plane, where larger values of the former imply smaller values of the latter.

Regarding the constraints on the HOD parameters, the least informative case (uniform prior on β) yields mean values of $\alpha = 0.83$, $\log M_{\rm min} = 11.67$ and $\log M_1 = 13.41$, with 68% credible intervals of [0.45, 1.24], [11.48, 11.94], and [12.60, 14.23]. As we discuss below, even in this case, which has an additional degree of freedom with respect to González-Nuevo et al. (2021), a remarkable improvement takes place concerning the determination of α, which had not even been constrained, and M_1, for which mostly lower limits had been derived up to now. This is a direct consequence of the reduction of the error bars on small scales with respect to the measurement strategy in previous works. Nevertheless, the uncertainty in $M_{\rm min}$ is comparable to that of González-Nuevo et al. (2021) due to the range of β values that are allowed in the present work. Indeed (and interestingly), the posterior distribution of β has a clear maximum at 2.20, but displays a long tail toward the high end. However, in the case of a narrow Gaussian prior around 2.90, the posterior does not deviate from it in any noticeable way. This fact, which in principle could be indicative of a wrong assumption for the Gaussian β prior, will be commented further in the next subsection.

When the value of β is fixed, either to 3.0. or 2.2, the constraints are clearly tightened, yielding mean values of $\log M_{\rm min} = 11.67^{+0.27}_{-0.19}$ and $11.47^{+0.22}_{-0.13}$ and $\log M_1 = 13.41^{+0.82}_{-0.81}$ and $13.04^{+0.67}_{-1.02}$, respectively. The difference between both cases explains the higher uncertainties discussed above. Indeed, as made clear in Fig. D.2 by the red $\beta - \log M_{\rm min}$ and $\beta - \log M_1$ contours and the marginalized posterior distributions of $M_{\rm min}$ and M_1 for all cases, a lower (higher) value of β is related to a higher (lower) value of both $M_{\rm min}$ and M_1. This is a consequence of the interplay between halo masses and the logarithmic slope of the background number counts: the small-scale (1-halo) behavior of the cross-correlation caused by larger HOD masses can be counterbalanced by smaller β values, which reduce the normalization of the signal.

Concerning cosmology, the case with the uniform prior on β yields an unconstrained posterior distribution for the σ_8 parameter. However, when β is fixed to 3.0 and 2.2, mean values of $\sigma_8 = 0.82^{+0.10}_{-0.10}$ and $1.02^{+0.18}_{-0.04}$ are obtained at 68% credibility, respectively. With regard to the matter density parameter, it is the most robust in terms of dependence on the value of β. Compared to typical cosmological probes, low values of $\Omega_{\rm m}$ are obtained in all four cases, with a mean of $0.21^{+0.03}_{-0.05}$ for the first (uniform prior on β), $0.23^{+0.03}_{-0.06}$ for the second (Gaussian prior), $0.24^{+0.02}_{-0.06}$ for the third ($\beta = 3.0$), and $0.20^{+0.03}_{-0.05}$ for the last ($\beta = 2.2$). The Hubble constant, however, cannot be constrained at the moment in any case, as in previous studies (Bonavera et al. 2020, 2021; González-Nuevo et al. 2021). The reason behind this is twofold: firstly, the sensitivity of the angular cross-correlation function to the h parameter is concentrated on the largest scales, where the uncertainties are still large. Secondly, given the degeneracy between h and $\Omega_{\rm m}$, which have opposite effects, and the fact that the angular cross-correlation is much more sensitive to the latter parameter, constraining the former seems to prove challenging at the current stage.

Nonetheless, for comparison purposes, Fig. 5 depicts the results for the $\beta = 3.0$ case from this paper to those of González-Nuevo et al. (2021), which had used the same β value. The results show a remarkable overall improvement in terms of parameter uncertainties, mainly due to the reduction of error bars and to the possibility of reaching larger angular scales with the new methodology. The differences in the specific posterior distributions can be explained by two different facts[4]. Firstly, and as discussed in Sect. 3.2, the data from González-Nuevo et al. (2021), which employed a different measurement methodology, might have been biased due to the relative arbitrariness of the so-called integral constraint correction on large scales. Secondly, and as will be highlighted in the next subsection, the steeper

[4] It should also be noted that tighter prior distributions were imposed for the HOD parameters in González-Nuevo et al. (2021).

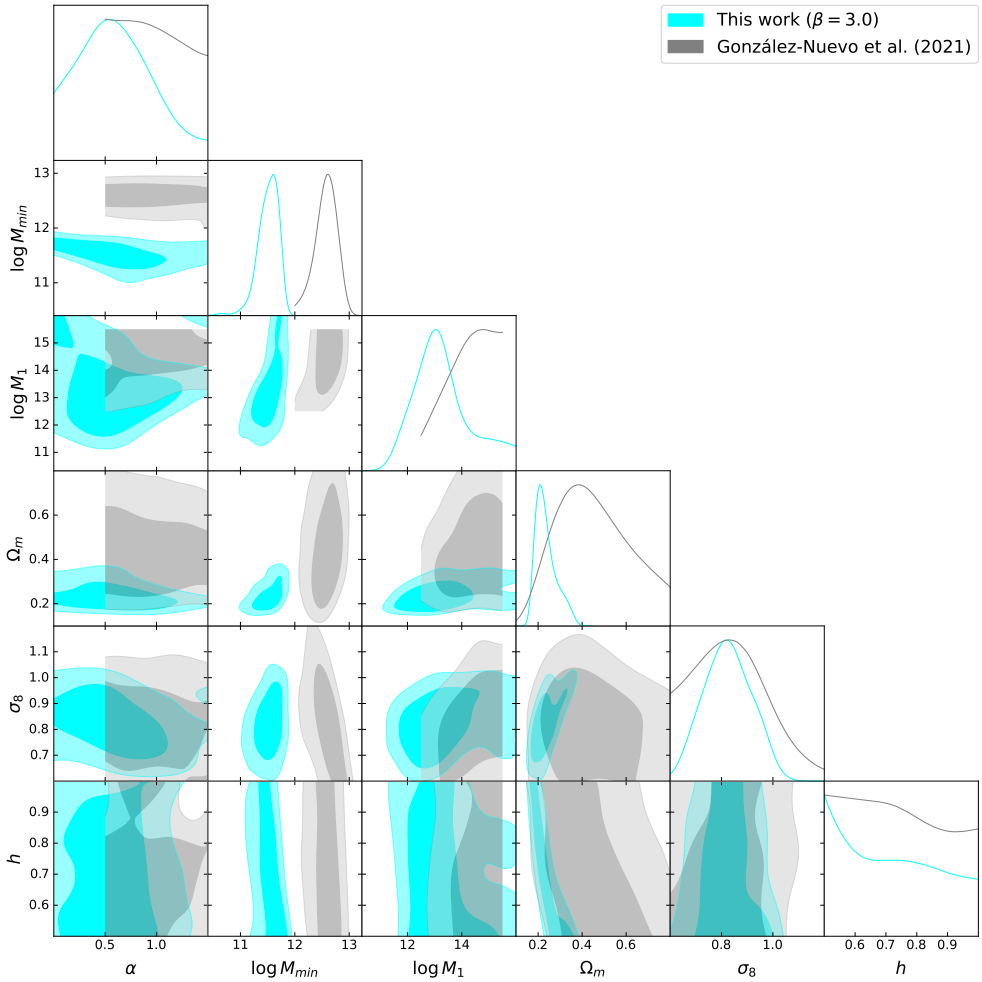

Fig. 5. Marginalized posterior distributions and probability contours for the MCMC run on the cross-correlation function with a fixed value of $\beta = 3.0$ (in cyan), compared to the results from González-Nuevo et al. (2021) depicted in gray.

ones fall above 20 arcmin in the data from González-Nuevo et al. (2021) can only be accounted for with larger values of Ω_m.

Figure 6 shows the posterior sampling of the cross-correlation function for all four cases in red, green, blue and cyan, respectively, along with the best fit (solid black line) and the data. The model explains the data correctly on most scales, but it tends to slightly overestimate the signal between 10 and 40 arcmin and to underestimate it above 60 arcmin, especially when β has larger values (2.9 and 3.0). Moreover, the tight small-scale uncertainties of the data seem to induce a preference toward a steeper fall above 60 arcmin than the data suggest. Indeed (as commented in Sect. 3.2), our current data do not die off particularly steeply and a sensitivity analysis shows that this large-scale behavior forces particularly low values of Ω_m. This issue will be studied in detail in the next subsection.

The first conclusion to be drawn up to now is that a priori knowledge on the β parameter is a necessary condition for constraining cosmology using the submillimeter galaxy magnification bias with a single redshift bin. Indeed, although the matter density parameter is barely affected by its value, too loose a prior on β erases the possibility to constrain σ_8 as a consequence of the large degeneracy and interplay between both parameters. Furthermore, a wrong assumption of a fixed value of β can substantially bias the constraint on σ_8, with larger values of the former linked to lower values of the latter. In particular, it should be stressed once again that previous works on the submillimeter galaxy magnification bias assumed a value of $\beta = 3.0$.

4.1.2. Sampling variance and the large-scale cross-correlation

The moderate fall in the large-scale behavior of the cross-correlation function seems to suggest a more thorough analysis. Given the fact that we have four spatially separated regions in the sky, we decided to measure the angular cross-correlation function in each of the different fields independently so as to examine if there are substantial differences among them. Figure 7 shows the corresponding results for the G09, G12, G15 and SGP regions (in red, blue, green and pink, respectively) along with the overall data (in black).

Although the cross-correlation signal shows a relatively stable profile in the case where all regions are combined, the data display non-negligible variations among the fields that are nonetheless mostly contained within the error bars. However the G15 regions represents a clear exception, with a systematically higher signal on all scales. Regarding possible explanations of this phenomenon, it should first be noted that the selection criteria of both the foreground and background samples are uniform across all four regions and the redshift distributions barely differ. Moreover, stellar masses derived via stellar population fitting of

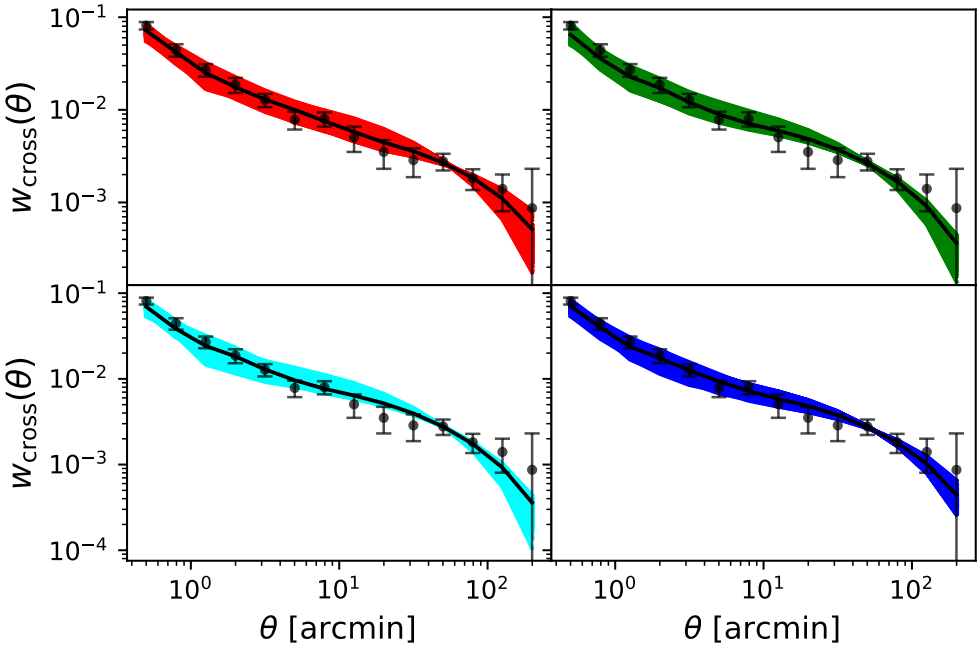

Fig. 6. Posterior-sampled angular cross-correlation function for the different cases of Sect. 4.1.1, that is, uniform prior on β (in red), Gaussian prior on β (in green) and fixed β values of 3.0 and 2.2 (cyan and blue, respectively). The cross-correlation data are shown in black.

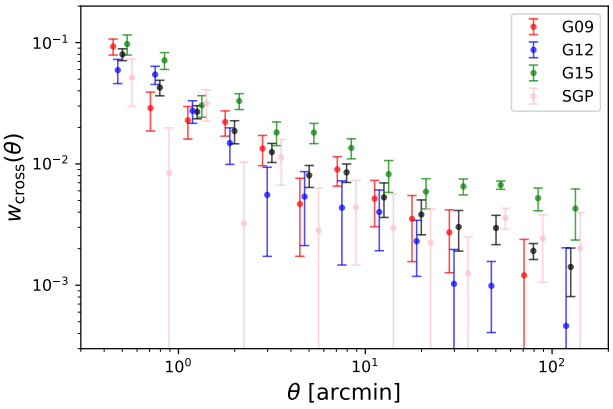

Fig. 7. Measurements of the foreground-background angular cross-correlation function in each independent field separately (G09, in red, G12, in blue, G15, in green and SGP, in pink), compared to the global measurement combining all fields (in black).

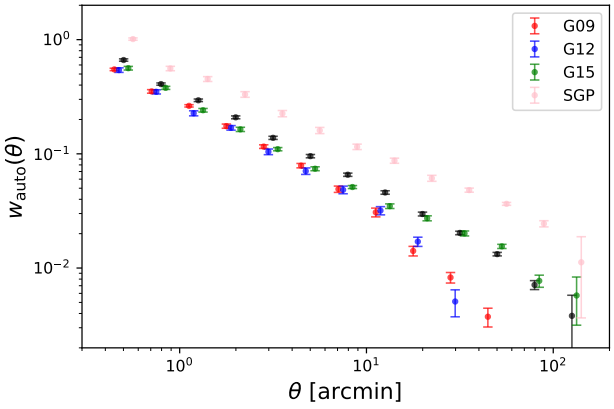

Fig. 8. Measurements of the foreground angular auto-correlation function in each independent field separately (G09, in red, G12, in blue, G15, in green and SGP, in pink), compared to the global measurement combining all fields (in black).

the SEDs as found in the GAMA catalogs show a remarkable uniformity among the fields.

Another possibility would be failing to correctly consider the selection function of the samples in the construction of the random catalogs. This is automatically taken into account for the foreground sample, since we made use of the random catalogs generated by the GAMA team and available on the database that reproduce the underlying selection function[5]. As for the background sample, the spatial variation in the instrumental noise is considered via H-ATLAS noise maps, which are imprinted in the random catalog. Even if this didn't account for the whole selection function of the background sample, the differences among regions are already present in the foreground sample itself, so they should not be sourced by a missing consideration in the selection function.

Indeed, this variability is observed even more strongly in the angular auto-correlation function of the foreground sample, depicted in Fig. 8 with the same colors as the cross-correlation. The foreground galaxies in the G15 region do not seem to cluster particularly strongly (which would explain a higher cross-correlation signal), but the large variation among fields, even at small scales, seems to point to sampling variance as the underlying cause. Although this issue will be studied further in future work with a larger data set, we performed a preliminary analysis by excluding the seemingly anomalous G15 region.

Figure 3 shows a comparison between the cross-correlation measurements taking all four regions into account (in black) and the corresponding ones resulting from excluding the G15 region (in olive green). Indeed, the removal of G15 induces a steeper fall above 40 arcmin, more in keeping with the qualitative behavior

[5] Using a purely Poissonian random catalog was shown to introduce almost negligible variations in the signal.

of the data from González-Nuevo et al. (2021). Although this region was also included in the cross-correlation computation in González-Nuevo et al. (2021), we believe that the different measurement strategy (involving splitting each field into 16 subregions and averaging the cross-correlation signal) might have helped mitigate the differences.

In order to assess the influence of the (unexpectedly) moderate fall in the large-scale data, we considered excluding the G15 region from the computation of the angular cross-correlation function. We performed two MCMC runs, namely, with a uniform and a Gaussian prior on β, and compared the results with those from the previous subsection. The full cornerplots are shown in Figs. D.3 and D.4, respectively.

The results on the HOD parameters behave as expected for both cases. Indeed, the exclusion of the G15 region reduces the cross-correlation signal even at small scales, which implies smaller galaxy halo masses. As regards cosmology, important differences appear with respect to using all fields. Figure 9 summarizes the deviations for Ω_m and σ_8. For both cases, Ω_m is the most affected parameter, with the same qualitative behavior: the exclusion of the G15 region implies higher values of Ω_m due to the steeper fall at the large-scale end. In particular, for a uniform (Gaussian) prior distribution on β, the mean value goes from $\Omega_m = 0.21^{+0.03}_{-0.05}$ ($0.23^{+0.03}_{-0.06}$) using all data and regions to $\Omega_m = 0.29^{+0.03}_{-0.06}$ ($0.27^{+0.03}_{-0.04}$) when the G15 region is excluded. As regards σ_8, the interplay with the β parameter still doesn't allow a clear determination when no information about the latter is given, but an upper limit of $\sigma_8 < 0.81$ is obtained at 68% credibility. However, for the Gaussian prior on β, the distribution is tightened, with the mean value going from $\sigma_8 = 0.79^{+0.10}_{-0.10}$ using all regions to $\sigma_8 = 0.72^{+0.04}_{-0.04}$ when the G15 region is excluded. Interestingly, the posterior distribution of the Hubble constant, although wide, displays (for the first time in submillimeter galaxy magnification bias studies) a clear maximum and we derived a mean value of $h = 0.79^{+0.13}_{-0.14}$. It should be noticed that, although the observed anticorrelation between Ω_m and h would be expected to imply smaller values of the Hubble constant with respect to the previous case (given the larger Ω_m after removing the G15 region), such values are not favored. This seems to be a consequence of the weaker overall signal compared to the four-field case. Indeed, even if the anticorrelation between the two parameters is still present, too small a value for the Hubble constant would not allow Ω_m (which cannot be arbitrarily high due to the interplay with σ_8 and the HOD parameters) to explain the weaker signal[6]. Lastly, and as opposed to the previous case, the marginalized distribution of β for the uniform-prior case now displays a clear maximum at 3, well above the value of 2.2. Therefore, the physically motivated $\beta \sim 2.9$ is recovered by excluding the G15 region, effectively solving the apparent inconsistency described in the previous subsection.

The conclusion at this point is that sampling variance can induce a bias in the cosmological parameter constraints derived from an analysis of the submillimeter galaxy magnification bias. In our case, an excess of cross-correlation seems to be present in the G15 region. When excluded, the behavior resembles the qualitative fall of the signal from González-Nuevo et al. (2021), which averaged over smaller subregions and was thus less likely to be affected by large-scale inhomogeneities. Further work is needed along this line to confirm this hypothesis by enlarging the galaxy samples.

[6] A sensitivity analysis shows that, unlike for the Hubble constant, variations of Ω_m still affect the signal at small and intermediate scales and thus imply a non-trivial interplay with the HOD parameters.

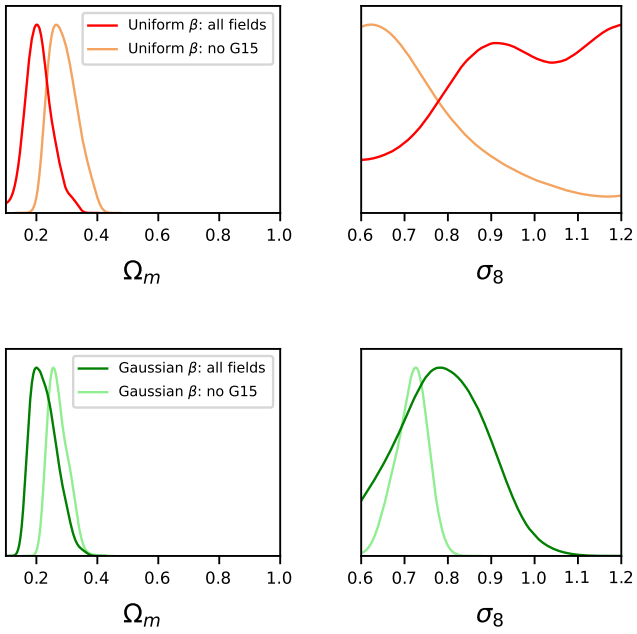

Fig. 9. Marginalized posterior distribution for the Ω_m and σ_8 parameters. Top panel: results for the MCMC runs with a uniform prior on β. Bottom panel: results for the MCMC runs with a Gaussian prior on β.

4.2. Cross- and auto-correlation joint analysis

Inspired by studies of galaxy-galaxy lensing, where the clustering of galaxies is added to tighten constraints (Abbott et al. 2018, 2022), we performed a joint analysis for both the angular cross- and foreground auto-correlation functions. To assess whether galaxy clustering could help tighten the constraints, we analyzed our base case: a Gaussian prior for β. The full cornerplot with the results is shown in Fig. D.5 in purple, where the cross-correlation-only case is also depicted in green for comparison purposes. Table D.5 presents the corresponding summarized statistical results.

The constraints on all parameters are extremely tightened after the incorporation of the auto-correlation function. Regarding the HOD, we obtained mean values of $\alpha = 0.92^{+0.04}_{-0.05}$, $\log M_{\rm min} = 11.54^{+0.09}_{-0.11}$ and $\log M_1 = 12.41^{+0.25}_{-0.17}$. Interestingly, the distribution of the matter density parameter is displaced toward higher values, with a mean value of $\Omega_m = 0.36^{+0.01}_{-0.02}$. The normalization parameter of the power spectrum is reduced to lower values with respect to the sole cross-correlation measurement, with a mean of $\sigma_8 = 0.72^{+0.04}_{-0.03}$. Regarding the Hubble constant, we obtained an extremely saturated distribution toward low values, possibly related to the apparent inconsistency between both observables.

Indeed, the sampling of the posterior distribution, shown in Fig. 10, reveals a poor joint fit to the data, mainly due to the underestimation of the auto-correlation signal below 20 arcmin. A similar issue was found in the joint clustering and galaxy-galaxy lensing analysis of Abbott et al. (2022), where an internal inconsistency between the tangential shear and the auto-correlation function of the redMaGiC lens sample was apparent. Although sampling variance would again be a plausible explanation in our case and, possibly, the most relevant one at the current stage, other reasons could be behind this discrepancy. Indeed, as shown by hydrodynamical simulations, baryonic feedback can impact the galaxy and halo distributions differently (van Daalen et al. 2014, 2020; Renneby et al. 2020),

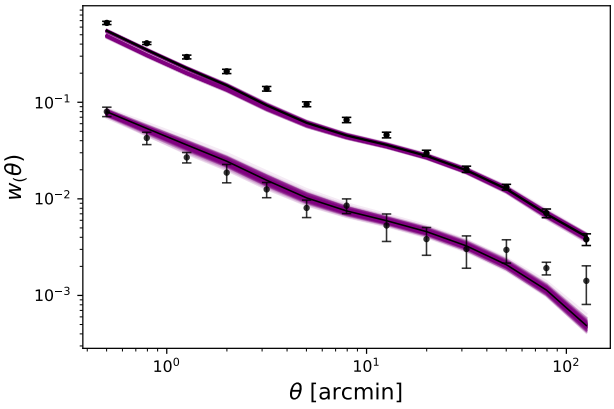

Fig. 10. Posterior-sampled angular auto- and cross-correlation functions from the joint analysis of both observables. The data are shown in black.

which can effectively reduce the lensing signal in the one-halo regime while having no impact on clustering (Amon et al. 2023). Moreover, assembly bias (i.e., the dependency of halo occupation and clustering on secondary properties other than halo mass) can suppress the lensing signal on small and even intermediate scales ($\lesssim$5−10 Mpc h^{-1}), as shown by Gao et al. (2005), Wechsler et al. (2006), Leauthaud et al. (2017), and Amon et al. (2023), among others. This matter goes beyond the scope of this paper and will be pursued in future work.

However, slightly different HOD values were found in previous related works (Bonavera et al. 2020, 2021) from a separated analysis of the auto- and cross-correlation functions. Since this could be related to the above discussion, we decided to run a preliminary joint test allowing a different HOD model for each observable to analyze if the galaxy-halo connection could respond differently to each physical situation. The results are shown in Fig. D.6 and Table D.6. The overall fit, shown in Fig. D.7, is better than the previous case, but the cross-correlation is now overestimated below 30 arcmin. The galaxy-halo connection, although compatible at 2σ, differs slightly between the auto- and cross-correlation models, since $\log M_{\rm min}^{\rm auto} = 11.58^{+0.12}_{-0.22}$, $\log M_{\rm min}^{\rm cross} = 11.83^{+0.08}_{-0.08}$, $\log M_1^{\rm auto} = 12.33^{+0.24}_{-0.47}$ (at 68% credibility), and $\log M_1^{\rm cross} > 12.34$ (at 95% credibility). In turn, this implies a very high value of σ_8, whose mean value is constrained to be $\sigma_8 = 1.10^{+0.05}_{-0.04}$. We found a matter density parameter that is perfectly consistent with the previous cases ($\Omega_{\rm m} = 0.34^{+0.05}_{-0.05}$) and, interestingly, the Hubble constant is no longer unconstrained, with a mean value of $h = 0.69^{+0.06}_{-0.17}$. These results should of course be taken with a pinch of salt and further work will be needed regarding the enlargement of the galaxy sample to shed light on this apparent inconsistency.

5. Summary and conclusions

This paper has addressed the cosmological constraints resulting from the refinement of the methodology of the submillimeter galaxy magnification bias. With an improved strategy with respect to previous works, we have measured the weak-lensing induced angular cross-correlation function between a sample of H-ATLAS submillimeter galaxies with photometric redshifts $1.2 < z < 4.0$ and a sample of GAMA II galaxies with spectroscopic redshifts $0.2 < z < 0.8$ using a single wide foreground bin. Additional aspects on the modeling side have been addressed, as the relevance of the logarithmic slope of the background number counts and the explicit numerical computation of the two-halo terms of the halo model power spectra. Additionally, we have measured the angular auto-correlation function of the GAMA II galaxies to assess the possibility of improving the constraints on cosmology through additional information about the foreground sample. By means of a halo model interpretation of the galaxy-matter and galaxy-galaxy power spectra, we carried out a Bayesian analysis through an MCMC algorithm to obtain posterior probability distributions about the HOD and the cosmological parameters $\Omega_{\rm m}$, σ_8 and h of a flat ΛCDM model.

We began the analysis using only the cross-correlation data and an important point was immediately raised: the value of β (i.e., the logarithmic slope of the background (unlensed) integral number counts) can have a large influence on the constraints of some of the parameters, namely $M_{\rm min}$, M_1 and, most importantly, σ_8. Although the matter density parameter, $\Omega_{\rm m}$, was barely affected by this issue, we conclude that a priori information on the value of β is of paramount importance to derive unbiased constraints using the cross-correlation data alone. Indeed, prior studies relied on a fixed value of $\beta = 3$, but the degeneracy and interplay between this parameter and σ_8 imply that larger values of the former are related to lower values of the latter. Since the choice of a uniform prior distribution for β does not produce any constraint on σ_8, an analysis of the (predicted) intrinsic integral number counts allowed us to restrict the prior distribution of β on physically-motivated grounds: $\beta = 2.90 \pm 0.04$. For this case, we obtained mean values of $\Omega_{\rm m} = 0.23^{+0.03}_{-0.06}$ and $\sigma_8 = 0.79^{+0.10}_{-0.10}$ at 68% credibility and no deviation whatsoever from its prior distribution for β.

However, inspecting the sampling of the posterior distribution and comparing the current cross-correlation data with the ones from González-Nuevo et al. (2021) raised suspicion that there could be an intrinsic bias in the signal. Indeed, the moderate large-scale fall seems to be induced by sampling variance; more precisely, by the seemingly anomalous behavior of the G15 region, where the signal is notably stronger than the rest at all scales. To test how this could influence the results, we performed additional MCMC runs by excluding the G15 region from the measurement of the signal. The two cases that we studied (uniform and Gaussian prior on β) yielded qualitatively similar results, confirming that nonnegligible differences arose with respect to the previous case. Although the influence of β on the constraining power over the σ_8 parameter remained, the exclusion of the G15 region yielded a mean value of $\Omega_{\rm m} = 0.29^{+0.03}_{-0.06}$ for the matter density parameter. When a Gaussian prior was assigned to β, we obtained mean values of $\Omega_{\rm m} = 0.27^{+0.02}_{-0.04}$ and $\sigma_8 = 0.72^{+0.04}_{-0.04}$, larger and smaller than the results of the previous paragraph, where all regions had been used. Interestingly, in this case we obtained the first (albeit loose) constraint on then the Hubble constant using the submillimeter galaxy magnification bias, for which results up to now had only consisted in one-sided limits. A mean value of $h = 0.79^{+0.13}_{-0.14}$ was obtained at 68% credibility.

When a joint analysis of the angular cross- and foreground auto-correlation functions is carried out, parameter constraints are tightened with respect to the previous case, but the fit is worsened due to a clear underestimation of the auto-correlation signal that is reminiscent of the internal inconsistency found in Abbott et al. (2022) between tangential shear and autocorrelation measurements in one of their lens samples. Although sampling variance is still likely the most plausible explanation, other phenomena, such as baryonic feedback or assembly bias acting on small to intermediate scales, may be behind the apparent discrepancy between both observables.

Along with the inclusion of these smaller-scale effects, the overall behavior of the G15 region (and, more generally, the issue of sampling variance) will be a subject for future studies, since there is no physical indication that it should be discarded for our analyses. The addition of more independent regions coming from other surveys will allow us to estimate the significance of this deviation of the cross-correlation signal by performing a statistical analysis.

Acknowledgements. L.B., J.G.N., J.M.C. and D.C. acknowledge the PID2021-125630NB-I00 project funded by MCIN/AEI/10.13039/501100011033/FEDER, UE. L.B. also acknowledges the CNS2022-135748 project funded by MCIN/AEI/10.13039/501100011033 and by the EU "NextGenerationEU/PRTR". J.M.C. also acknowledges financial support from the SV-PA-21-AYUD/2021/51301 project. We deeply acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and support. In particular the projects "SIS22_lapi", "SIS23_lapi" in the framework "Convenzione triennale SISSA-CINECA". The *Herschel*-ATLAS is a project with *Herschel*, which is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. The H-ATLAS website is `http://www.h-atlas.org`. GAMA is a joint European Australasian project based around a spectroscopic campaign using the Anglo Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including GALEX MIS, VST KIDS, VISTA VIKING, WISE, *Herschel*-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is: `http://www.gama-survey.org`. This research has made use of the python packages `ipython` (Pérez & Granger 2007), `matplotlib` (Hunter 2007), and `Scipy` (Jones et al. 2001).

References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, Phys. Rev. D, 105, 023520
Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. 2022, J. High Energy Astrophys., 34, 49
Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697
Amblard, A., Cooray, A., Serra, P., et al. 2010, A&A, 518, L9
Amon, A., Robertson, N. C., Miyatake, H., et al. 2023, MNRAS, 518, 477
Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, MNRAS, 404, 86
Baldry, I. K., Alpaslan, M., Bauer, A. E., et al. 2014, MNRAS, 441, 2440
Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291
Bautista, J. E., Paviot, R., Vargas Magaña, M., et al. 2021, MNRAS, 500, 736
Beutler, F., Blake, C., Colless, M., et al. 2011, MNRAS, 416, 3017
Bonavera, L., González-Nuevo, J., Suárez Gómez, S. L., et al. 2019, JCAP, 2019, 021
Bonavera, L., González-Nuevo, J., Cueli, M. M., et al. 2020, A&A, 639, A128
Bonavera, L., Cueli, M. M., González-Nuevo, J., et al. 2021, A&A, 656, A99
Bonavera, L., Cueli, M. M., González-Nuevo, J., Casas, J. M., & Crespo, D. 2024, A&A, 686, A191
Brout, D., Scolnic, D., Popovic, B., et al. 2022, ApJ, 938, 110
Bull, P., Akrami, Y., Adamek, J., et al. 2016, Phys. Dark Univ., 12, 56
Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559
Chapman, S. C., Smail, I., Blain, A. W., et al. 2004, ApJ, 614, 671
Chapman, S. C., Blain, A. W., Smail, I., et al. 2005, ApJ, 622, 772
Cooray, A., & Sheth, R. K. 2002, Phys. Rep., 372, 1
Cueli, M. M., Bonavera, L., González-Nuevo, J., et al. 2021, A&A, 645, A126
Cueli, M. M., Bonavera, L., González-Nuevo, J., et al. 2022, A&A, 662, A44
Cyburt, R. H., Fields, B. D., Olive, K. A., & Yeh, T.-H. 2016, Rev. Mod. Phys., 88, 015004
Di Valentino, E., & Melchiorri, A. 2022, ApJ, 931, L18
Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971
Eales, S., Dunne, L., Clements, D., et al. 2010, PASP, 122, 499
Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
Fields, B. D., Olive, K. A., Yeh, T.-H., & Young, C. 2020, JCAP, 2020, 010
Foreman-Mackey, D., Hogg, D. W., Lang, D., et al. 2013, PASP, 125, 306
Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66
González-Nuevo, J., Lapi, A., Fleuren, S., et al. 2012, ApJ, 749, 65
González-Nuevo, J., Lapi, A., Negrello, M., et al. 2014, MNRAS, 442, 2680
González-Nuevo, J., Lapi, A., Bonavera, L., et al. 2017, JCAP, 2017, 024
González-Nuevo, J., Cueli, M. M., Bonavera, L., et al. 2021, A&A, 646, A152
Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Granato, G. L., De Zotti, G., Silva, L., et al. 2004, ApJ, 600, 580
Herranz, D. 2001, in Cosmological Physics with Gravitational Lensing, eds. J. Tran Thanh Van, Y. Mellier, & M. Moniez, 197
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Ivison, R. J., Swinbank, A. M., Swinyard, B., et al. 2010, A&A, 518, L35
Ivison, R. J., Lewis, A. J. R., Weiss, A., et al. 2016, ApJ, 832, 78
Jarvis, M. 2015, Astrophysics Source Code Library [record ascl:`1508.007`]
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python, `http://www.scipy.org`
Kitayama, T., & Suto, Y. 1996, ApJ, 469, 480
Lacasa, F., & Kunz, M. 2017, A&A, 604, A104
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lapi, A., Shankar, F., Mao, J., et al. 2006, ApJ, 650, 42
Lapi, A., González-Nuevo, J., Fan, L., et al. 2011, ApJ, 742, 24
Leauthaud, A., Saito, S., Hilbert, S., et al. 2017, MNRAS, 467, 3024
Limber, D. N. 1953, ApJ, 117, 134
Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, MNRAS, 452, 2087
Mead, A. J., & Verde, L. 2021, MNRAS, 503, 3095
Mead, A. J., Tröster, T., Heymans, C., et al. 2020, A&A, 641, A130
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19
Pearson, E. A., Eales, S., Dunne, L., et al. 2013, MNRAS, 435, 2753
Pérez, F., & Granger, B. E. 2007, Comput. Sci. Eng., 9, 21
Perivolaropoulos, L., & Skara, F. 2022, New Astron. Rev., 95, 101659
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1
Planck Collaboration XVI. 2014, A&A, 571, A16
Planck Collaboration XIII. 2016, A&A, 594, A13
Planck Collaboration VI. 2020, A&A, 641, A6
Planck Collaboration VII. 2020, A&A, 641, A7
Renneby, M., Henriques, B. M. B., Hilbert, S., et al. 2020, MNRAS, 498, 5804
Riess, A. G., Yuan, W., Macri, L. M., et al. 2022, ApJ, 934, L7
Schmidt, F. 2016, Phys. Rev. D, 93, 063512
Scolnic, D., Brout, D., Carr, A., et al. 2022, ApJ, 938, 113
Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515
Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119
Swinbank, A. M., Smail, I., Longmore, S., et al. 2010, Nature, 464, 733
van Daalen, M. P., Schaye, J., McCarthy, I. G., Booth, C. M., & Dalla Vecchia, C. 2014, MNRAS, 440, 2997
van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, MNRAS, 491, 2424
Wang, L., Cooray, A., Farrah, D., et al. 2011, MNRAS, 414, 596
Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71
Weinberg, N. N., & Kamionkowski, M. 2003, MNRAS, 341, 251
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1

Appendix A: Ingredients of the halo model

This section describes in detail the computation of the halo model prescription for the galaxy and galaxy-matter power spectra. The halo mass function is expressed as:

$$n(M,z) = \frac{\rho_0}{M^2} f_{\rm ST}(\nu)\left|\frac{d\log\nu}{d\log M}\right|,$$

where

$$f_{\rm ST}(\nu) = A\sqrt{\frac{a\nu}{2\pi}}\left[1+\left(\frac{1}{a\nu}\right)^p\right]e^{-a\nu/2}$$

is the Sheth and Tormen model (Sheth & Tormen 1999), for which $A = 0.33$, $a = 0.75$, and $p = 0.3$. The ν parameter is defined as

$$\nu(M,z) \equiv \left[\frac{\hat{\delta}_c(z)}{\sigma(M,z)}\right]^2,$$

where $\hat{\delta}_c(z)$ is linear critical overdensity at redshift z for a region to collapse into a halo at this redshift (computed via the fit of Kitayama & Suto 1996) and $\sigma(M,z) \equiv D(z)\sigma(M,0)$, where $D(z)$ is the linear growth factor for a ΛCDM universe (normalized at $z = 0$) and $\sigma(M,0)$ is the square root of the mass variance of the filtered linear overdensity field at $z = 0$.

The linear deterministic halo bias is computed from the above halo mass function via the peak-background split (Sheth & Tormen 1999), yielding

$$b_1(M,z) = 1 + \frac{a\nu - 1}{\hat{\delta}_c(z)} + \frac{2p/\hat{\delta}_c(z)}{1+(a\nu)^p}.$$

Moreover, the mean number of galaxies in a halo of mass M, $\langle N\rangle_M$, follows the model described in Eq. (3) and the mean number density of galaxies at redshift, z, is computed via:

$$\bar{n}_g(z) = \int_0^\infty dM\, n(M,z)\,\langle N\rangle_M.$$

The matter transfer function used to compute the linear matter power spectrum is computed via the analytical formula of Eisenstein & Hu (1998), which takes baryonic effects into account. This approach has been favored over a numerical one based on Boltzmann codes due to computation time, since no significant differences were found for our purposes.

Lastly, the halo density profile is taken to follow the two-parameter NFW model (Navarro et al. 1997), that is,

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}.$$

The halo is effectively truncated at a comoving virial radius, r_h, which yields a relation between ρ_s and the mass M of the virialized halo (or, equivalently, its mean density, ρ_h):

$$\rho_s = \frac{\rho_h}{3}\frac{c}{\ln(1+c) - c/1+c},$$

where $c \equiv r_h/r_s$ is the halo concentration parameter. Since the mass of the halo satisfies

$$M = \frac{4}{3}\pi\, r_h^3(z)\,\rho_h(z),$$

the choice of both a typical density for a collapsed halo, ρ_h, and of a prescription for the mean mass-concentration relation, $c(M,z)$, completely specifies the values of r_s and ρ_s; thus, the NFW profile of a typical halo of mass, M, at redshift, z. We have chosen a virial overdensity criterion, that is,

$$\rho_h(z) = \Delta_{\rm vir}(z)\,\bar{\rho}_0,$$

where $\Delta_{\rm vir}(z)$ has been computed through the fit by Weinberg & Kamionkowski (2003). Regarding the mass-concentration relation, we have used that of Bullock et al. (2001). With all this in mind, the normalized Fourier transform of a typical halo of mass, M, at redshift, z, is expressed as:

$$u(k|M,z) = \frac{1}{\ln(1+c) - c/1+c}\Big[\sin kr_s[{\rm Si}([1+c]kr_s) - {\rm Si}(kr_s)] - \frac{\sin ckr_s}{[1+c]kr_s} + \cos kr_s[{\rm Ci}([1+c]kr_s) - {\rm Ci}(kr_s)]\Big],$$

where Si and Ci are the sine and cosine integrals, respectively.

Appendix B: Numerical issue about the computation of the two-halo term

As pointed out by Mead et al. (2020) and Mead & Verde (2021), the evaluation of the two-halo terms of the power spectra involving the matter field presents a numerical problem. The halo model considers that all matter is bound up into distinct halos, which is translated into the following condition for the halo mass function:

$$\int_0^\infty dM\,\frac{n(M,z)}{\bar{\rho}_0}\,M = 1,$$

which, in the case of the Sheth & Tormen model, fixes the normalization parameter A in terms of p. However, an additional condition involving the linear halo bias must be enforced in order for the halo model to reproduce the correct large-scale behavior of the power spectra. This condition is expressed as:

$$\int_0^\infty dM\,\frac{n(M,z)}{\bar{\rho}_0}\,b_1(M,z)\,M = 1$$

and presents a very relevant numerical problem. Indeed, defining

$$I(M_a,M_b,k,z) \equiv \int_{M_a}^{M_b} dM\,\frac{n(M,z)}{\bar{\rho}_0}\,b_1(M,z)\,M\,u(k|M,z), \quad \text{(B.1)}$$

the halo model requires that $I(M_a\to 0, M_b\to\infty, k\to 0, z)\to 1$. However, choosing sensible limits of $M_a = 10^8 M_\odot/h$ and $M_b = 10^8 M_\odot/h$, a regular *Planck* cosmology yields a value of ≈ 0.7. The problem stems from the fact that typical models for the halo mass function assign a large fraction of the total matter to low-mass halos, causing convergence of the above integral to be extremely slow as the lower limit of the integral becomes smaller.

To work around the problem, Schmidt (2016) suggests modifying the halo mass function under the assumption that all mass below M_a is contained in halos of a mass that is exactly M_a. This amounts to making the substitution:

$$n(M,z) \Longrightarrow n(M,z)\Theta(M-M_a) + \frac{A(M_a)}{b_1(M_a)M_a/\bar{\rho}_0}\delta_{\rm D}(M-M_a)$$

in (B.1), where

$$A(M_a) \equiv 1 - \int_{M_a}^\infty dM\frac{n(M,z)}{\bar{\rho}_0}\,b_1(M,z)\,M.$$

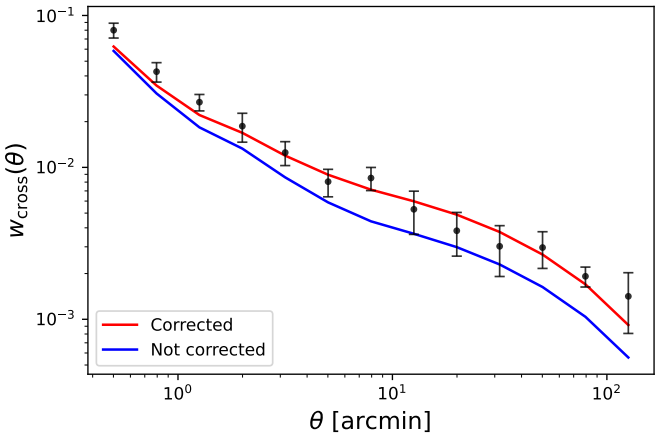

Fig. B.1. Effect on the cross-correlation function of the numerical correction in the computation of the two-halo term of the galaxy-matter power spectrum.

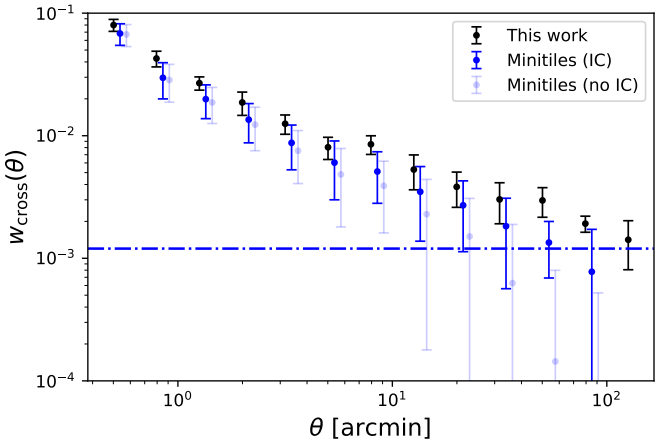

Fig. C.1. Angular cross-correlation data as measured in this work (in black) and using the minitiles strategy from González-Nuevo et al. (2021) (in blue). The light blue points represent the measurements without the IC correction, which is depicted by the dot-dashed line.

If this is not taken into account, an extremely strong bias is induced in any cross-correlation involving the matter field. For instance, Fig. B.1 shows the effect of not including this correction in the angular cross-correlation function. The discrepancy increases with the angular scale, reaching about 60% above 5 arcmin and compromising any reliable constraint on cosmology. We stress that this should be addressed in any computation of halo model cross-correlations involving matter.

Appendix C: Integral constraint correction

Given a certain patch, the number of detected galaxies in it will certainly be higher or lower than what we would have in a fair sample of the Universe, thus affecting our estimates in the random catalog of the patch. Averaging over a large number of patches, as done in previous works (Bonavera et al. 2020, 2021; Cueli et al. 2021, 2022), tends to introduce an artificial weakening of the observed clustering signal. This is because sources sufficiently close to edges of the corresponding field are less likely to have pairs at large distances. This effectively causes the estimated cross-correlation to be biased low by a constant (Adelberger et al. 2005), so that:

$$\hat{w}_{\rm ideal}(\theta) = \hat{w}(\theta) + \mathrm{IC}.$$

Although there are theoretical approaches to estimate the IC for a particular scanning strategy (see e.g., Adelberger et al. 2005), in practice, it is commonly estimated numerically using random-random counts. Specifically, the IC can be estimated for the cross-correlation using the formula:

$$\mathrm{IC} = \frac{\sum_i \mathrm{R_f R_b}(\theta_i) w_{\rm ideal}(\theta_i)}{\sum_i \mathrm{R_f R_b}(\theta_i)}, \tag{C.1}$$

where $w_{\rm ideal}(\theta)$ is an assumed model for the cross-correlation function. An equivalent expression is used for the auto-correlation.

However, the obvious caveat is that the ideal model for the cross-correlation function is not known. Two approaches appear to be reasonable in this situation. Thus, inspired by the usual procedure of the angular auto-correlation estimation, we can first assume a power-law model of $w_{\rm ideal}(\theta) = A\theta^{-\gamma}$. The procedure, however, is highly dependent on the limiting angular scales used for the fit. Although we can derive a wide-range estimate of IC$\sim 7 - 18 \times 10^{-4}$, we prefer to abandon this method in favor of a more precise alternative that does still make use of this value.

The second approach implies assuming the halo model as the ideal cross-correlation. An additional problem stems from the fact that the correction would depend on the HOD and, most importantly, on cosmology. This is crucial, because the largest angular scales are the most important for constraining cosmological parameters but, at the same time, they are the most affected by the adopted IC correction. Therefore, we performed the following analysis to try and refine the procedure: firstly, we performed maximum likelihood estimation searches for the HOD parameters using only the measurements in the one-halo regime and random uniform cosmological parameters. With the obtained HOD parameters, we computed the IC distribution for 100 random sets of cosmological parameters. We then discarded those cosmologies resulting in IC values outside the range $7 - 18 \times 10^{-4}$, as obtained above. The final estimate we obtained was IC=$(11 \pm 3) \times 10^{-4}$. To test the sensitivity of these results to the assumed cosmological parameter distribution, we repeated the analysis using Gaussian random cosmological parameter sets based on Planck (Planck Collaboration VI 2020) with dispersion values of ~ 0.05. The results we obtained were IC=$(13 \pm 2) \times 10^{-4}$. After considering all cases, we chose a value of

$$\mathrm{IC} = (12 \pm 3) \times 10^{-4}$$

as the most appropriate IC estimate for our study with the minitiles strategy.

Figure C.1 provides a comparison of the cross-correlation function estimated using different approaches. Light blue circles represent the minitiles results before the IC correction, which show a sharp decline above 10 arcmin (where the blue dashed line, representing the IC correction, becomes relevant). Blue circles correspond to the minitiles results after applying the IC correction, while the black ones are estimated using the new approach without any further correction. The uncertainties are derived from the covariance matrix, as described in the main text.

The minitiles results are consistent with the new results within the uncertainties up to 30 arcmin, but they appear to be slightly underestimated across all angular scales. Above 30 arcmin, which are the most important for cosmological constraints, the new results are clearly higher than those of

the minitiles. When we consider higher IC values, such as IC=15 − 20 × 10^{-4}, the two measurements become compatible even at the largest angular separations. We conclude that the IC correction procedure is not at all straightforward, since it depends on relatively arbitrary choices, such as the angular scales over which we may fit the data or the cosmologies sampled, which might be restricted but depend on a range dictated by the power-law fit. Therefore, we decided to avoid this methodology given the availability of a better alternative, which we chose for the purposes of this work.

Appendix D: Additional plots and tables

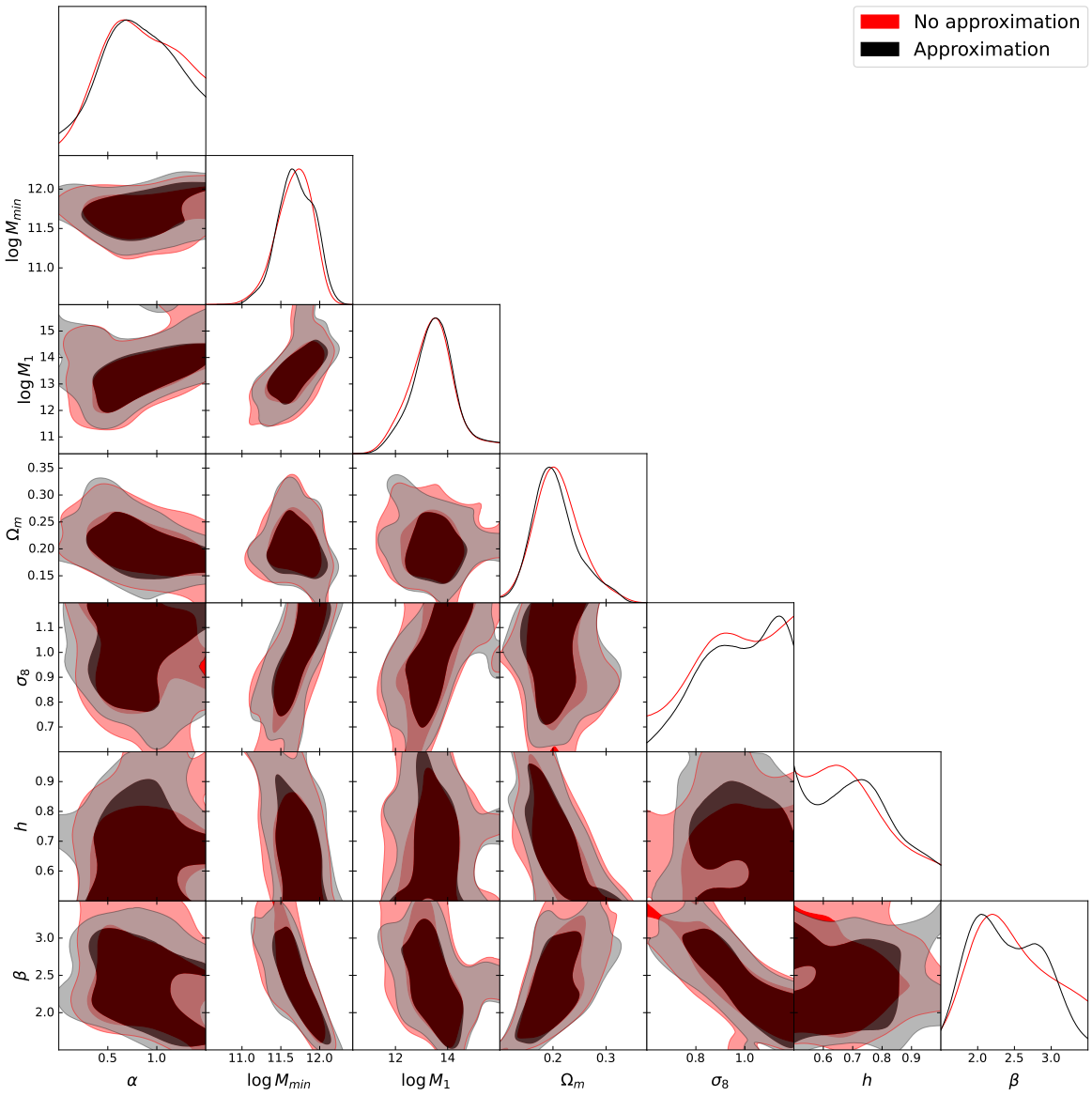

Fig. D.1. Marginalized posterior distributions and probability contours for the MCMC runs on the cross-correlation function with a uniform β prior. The results using the mean-redshift approximation are shown in red, while the ones with the full model are depicted in black.

Table D.1. Parameter prior distributions and summarized posterior results from the MCMC runs on the cross-correlation function with a uniform β prior using the mean redshift approximation and the full model.

	Approximation				No approximation			
Parameter	Prior	Mean	Mode	68% CI	Prior	Mean	Mode	68% CI
α	$\mathcal{U}[0.00, 1.50]$	0.83	0.64	$[0.45, 1.24]$	$\mathcal{U}[0.00, 1.50]$	0.81	0.71	$[0.56, 1.50]$
$\log M_{\rm min}$	$\mathcal{U}[10.00, 16.00]$	11.67	11.73	$[11.48, 11.94]$	$\mathcal{U}[10.00, 16.00]$	11.70	11.66	$[11.46, 11.95]$
$\log M_1$	$\mathcal{U}[10.00, 16.00]$	13.41	13.57	$[12.60, 14.23]$	$\mathcal{U}[10.00, 16.00]$	13.50	13.56	$[12.68, 14.29]$
Ω_m	$\mathcal{U}[0.10, 1.00]$	0.21	0.20	$[0.16, 0.24]$	$\mathcal{U}[0.10, 1.00]$	0.21	0.19	$[0.16, 0.24]$
σ_8	$\mathcal{U}[0.60, 1.20]$	0.95	–	$[0.87, 1.20]$	$\mathcal{U}[0.60, 1.20]$	0.97	–	$[0.89, 1.20]$
h	$\mathcal{U}[0.50, 1.00]$	0.70	0.65	$[0.50, 0.75]$	$\mathcal{U}[0.50, 1.00]$	0.71	–	$[0.50, 0.77]$
β	$\mathcal{U}[1.50, 3.50]$	2.46	2.16	$[1.81, 2.86]$	$\mathcal{U}[1.50, 3.50]$	2.40	2.09	$[1.82, 2.84]$

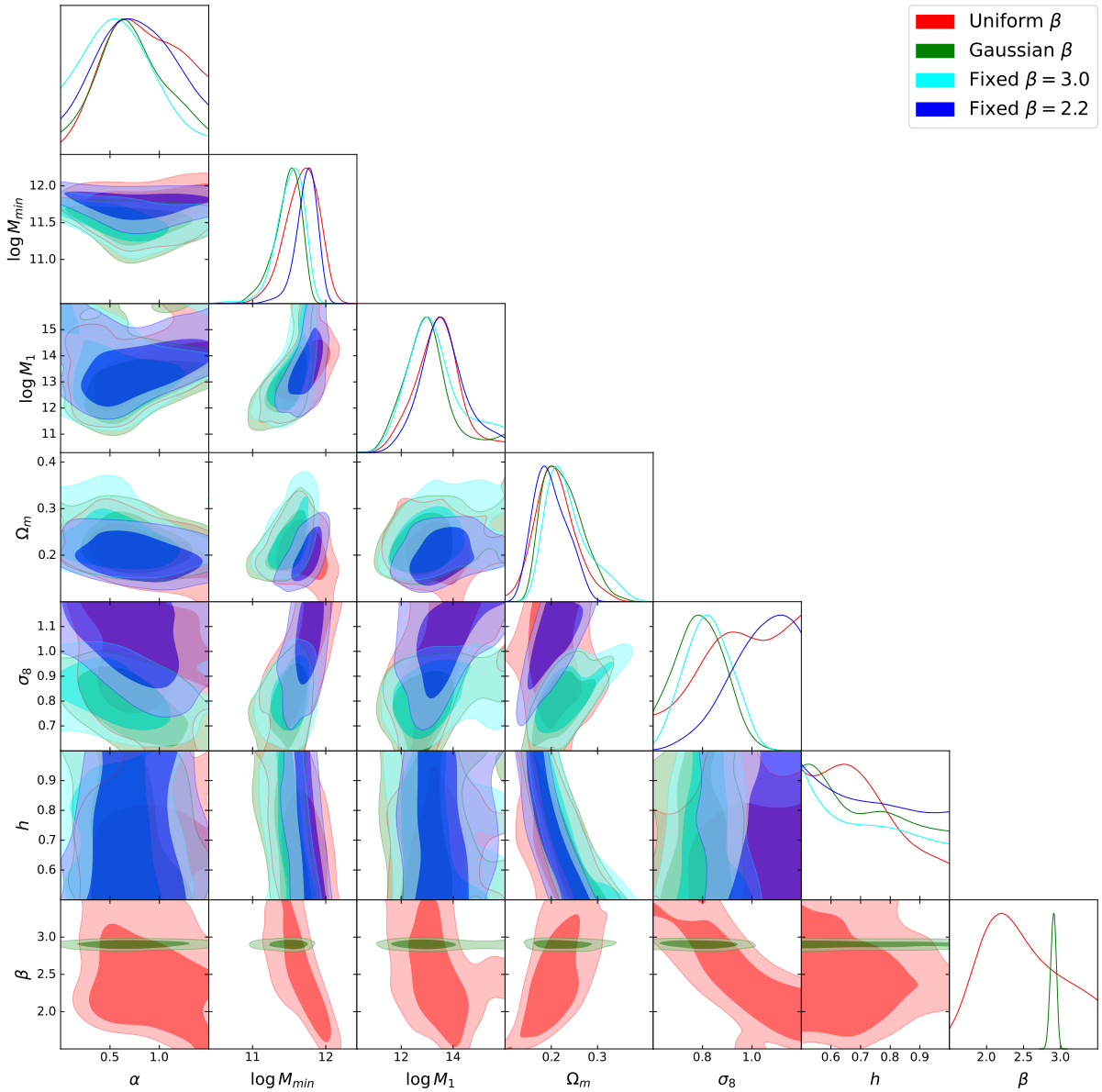

Fig. D.2. Marginalized posterior distributions and probability contours for the MCMC runs on the cross-correlation function for different choices of prior distributions on the β parameter.

Table D.2. Parameter prior distributions and summarized posterior results from the MCMC runs on the cross-correlation function with fixed values of the β parameter: 3.0 and 2.2.

	Fixed $\beta = 3.0$				Fixed $\beta = 2.2$			
Parameter	Prior	Mean	Mode	68% CI	Prior	Mean	Mode	68% CI
α	$\mathcal{U}[0.00, 1.50]$	0.61	0.54	[0.18, 0.90]	$\mathcal{U}[0.50, 1.50]$	0.76	0.69	[0.37, 0.89]
$\log M_{\rm min}$	$\mathcal{U}[10.00, 16.00]$	11.51	11.58	[11.37, 11.73]	$\mathcal{U}[10.00, 16.00]$	11.74	11.77	[11.62, 11.90]
$\log M_1$	$\mathcal{U}[10.00, 16.00]$	13.28	13.02	[12.00, 13.99]	$\mathcal{U}[10.00, 16.00]$	13.49	13.62	[12.66, 14.37]
Ω_m	$\mathcal{U}[0.10, 1.00]$	0.24	0.21	[0.18, 0.26]	$\mathcal{U}[0.10, 1.00]$	0.20	0.19	[0.15, 0.23]
σ_8	$\mathcal{U}[0.60, 1.20]$	0.82	0.82	[0.72, 0.92]	$\mathcal{U}[0.60, 1.20]$	1.02	1.12	[0.98, 1.20]
h	$\mathcal{U}[0.50, 1.00]$	0.72	–	[0.50, 0.81]	$\mathcal{U}[0.50, 1.00]$	0.73	–	[0.50, 1.00]

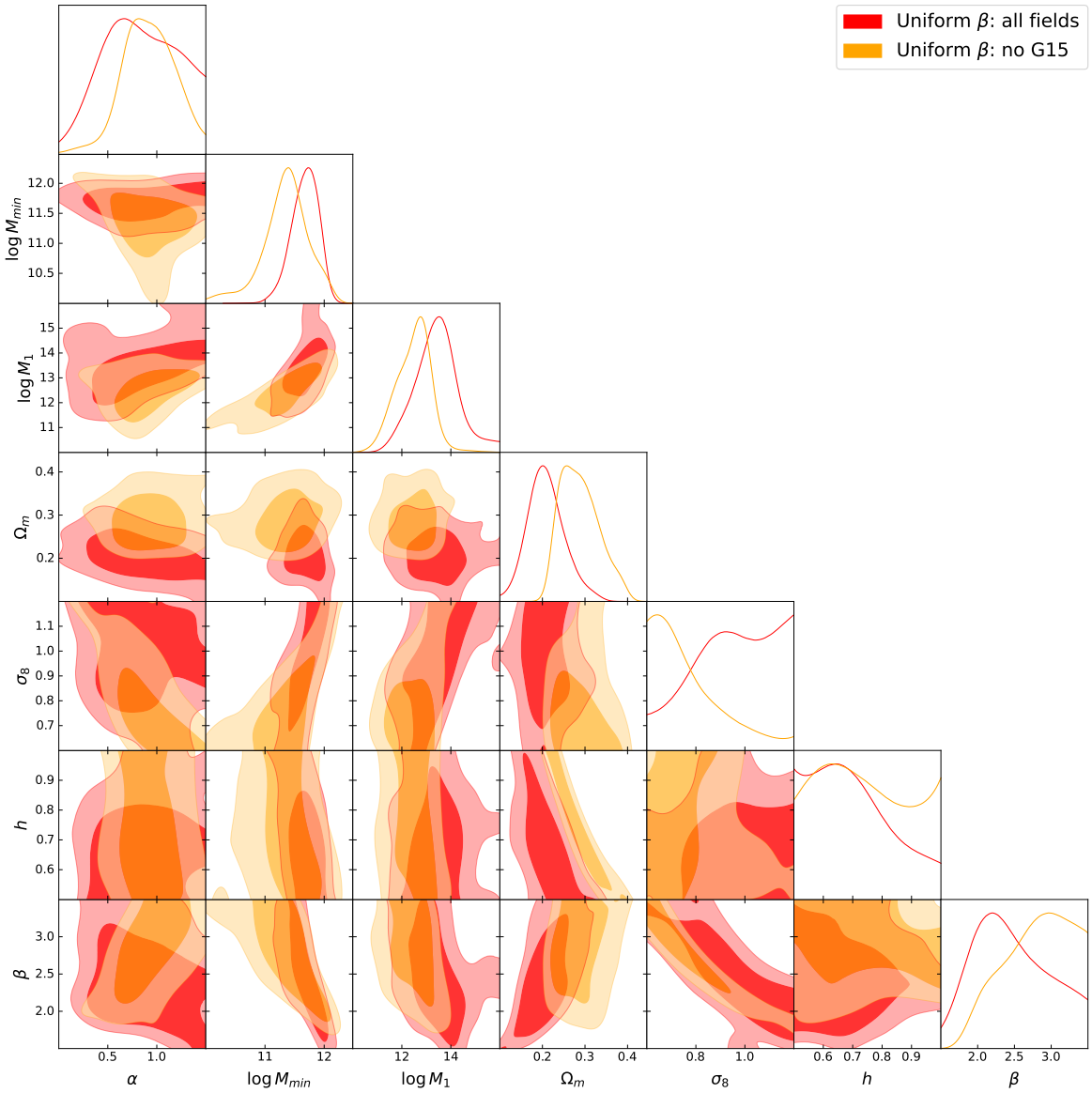

Fig. D.3. Marginalized posterior distributions and probability contours for the MCMC runs on the cross-correlation function with a uniform β prior. The results using all four fields are shown in red, while those where the G15 field has been excluded are depicted in orange.

Table D.3. Parameter prior distributions and summarized posterior results from the MCMC run on the cross-correlation function with a uniform prior on β and where the G15 regions was excluded from the measurement.

Parameter	Prior	Mean	Mode	68% CI
α	$\mathcal{U}[0.00, 1.50]$	0.91	0.86	[0.65, 1.19]
$\log M_{\rm min}$	$\mathcal{U}[10.00, 16.00]$	11.35	11.38	[11.01, 11.75]
$\log M_1$	$\mathcal{U}[10.00, 16.00]$	12.49	12.72	[11.86, 13.24]
Ω_m	$\mathcal{U}[0.10, 1.00]$	0.29	0.27	[0.23, 0.32]
σ_8	$\mathcal{U}[0.60, 1.20]$	0.77	–	[0.60, 0.81]
h	$\mathcal{U}[0.50, 1.00]$	0.74	–	[0.50, 1.00]
β	$\mathcal{U}[1.50, 3.50]$	2.78	3.00	[2.58, 3.50]

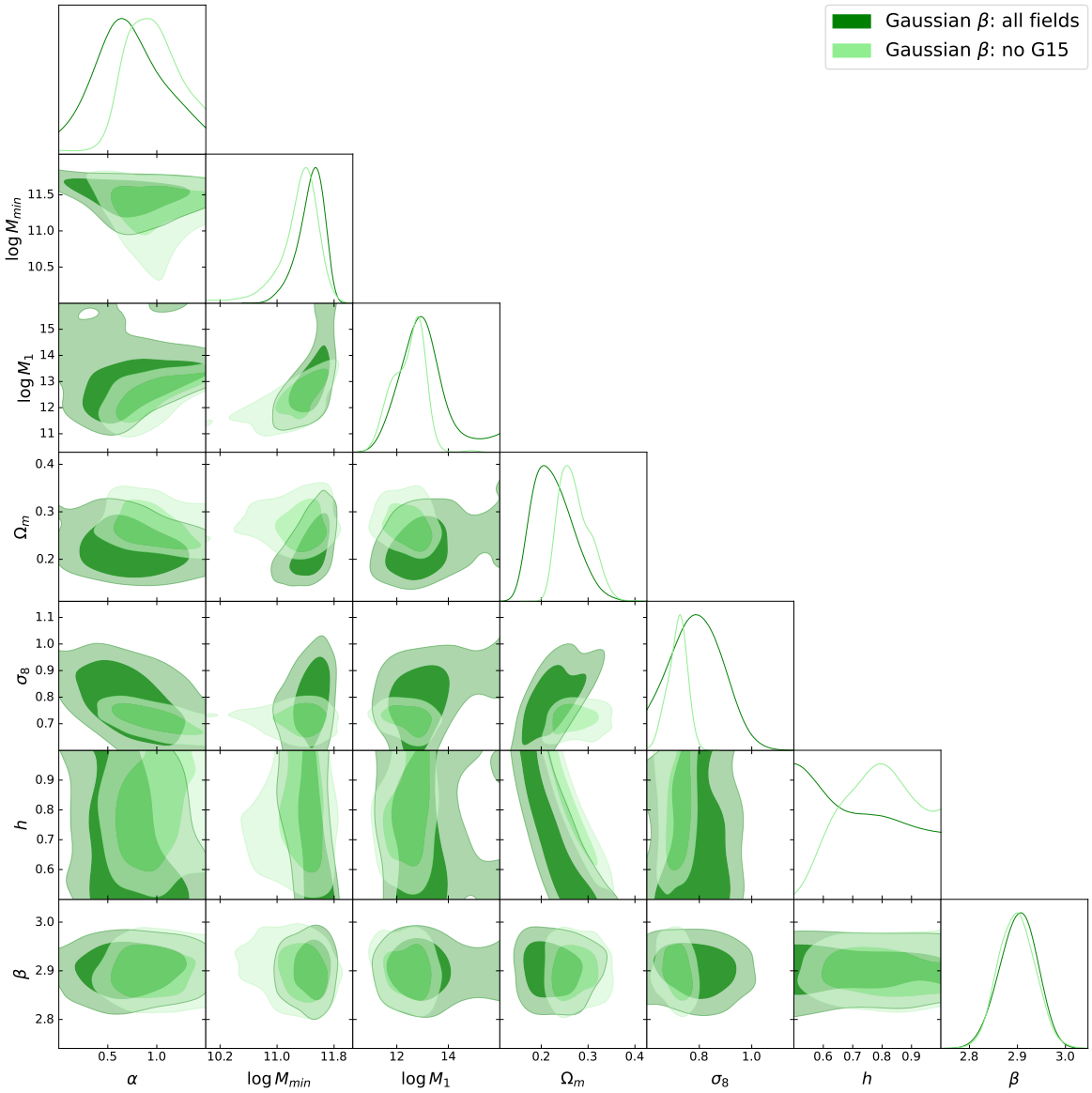

Fig. D.4. Marginalized posterior distributions and probability contours for the MCMC runs on the cross-correlation function with a Gaussian β prior. The results using all fields are shown in dark green, while those where the G15 field has been excluded are depicted in light green.

Table D.4. Parameter prior distributions and summarized posterior results from the MCMC run on the cross-correlation function with a Gaussian prior on β and where the G15 regions was excluded from the measurement.

Parameter	Prior	Mean	Mode	68% CI
α	$\mathcal{U}[0.00, 1.50]$	0.94	0.91	$[0.63, 1.19]$
$\log M_{\rm min}$	$\mathcal{U}[10.00, 16.00]$	11.31	11.41	$[11.15, 11.62]$
$\log M_1$	$\mathcal{U}[10.00, 16.00]$	12.46	12.82	$[11.93, 13.18]$
Ω_m	$\mathcal{U}[0.10, 1.00]$	0.27	0.26	$[0.23, 0.29]$
σ_8	$\mathcal{U}[0.60, 1.20]$	0.72	0.73	$[0.68, 0.76]$
h	$\mathcal{U}[0.50, 1.00]$	0.79	0.80	$[0.65, 0.92]$
β	$\mathcal{U}[1.50, 3.50]$	2.90	2.90	$[2.86, 2.94]$

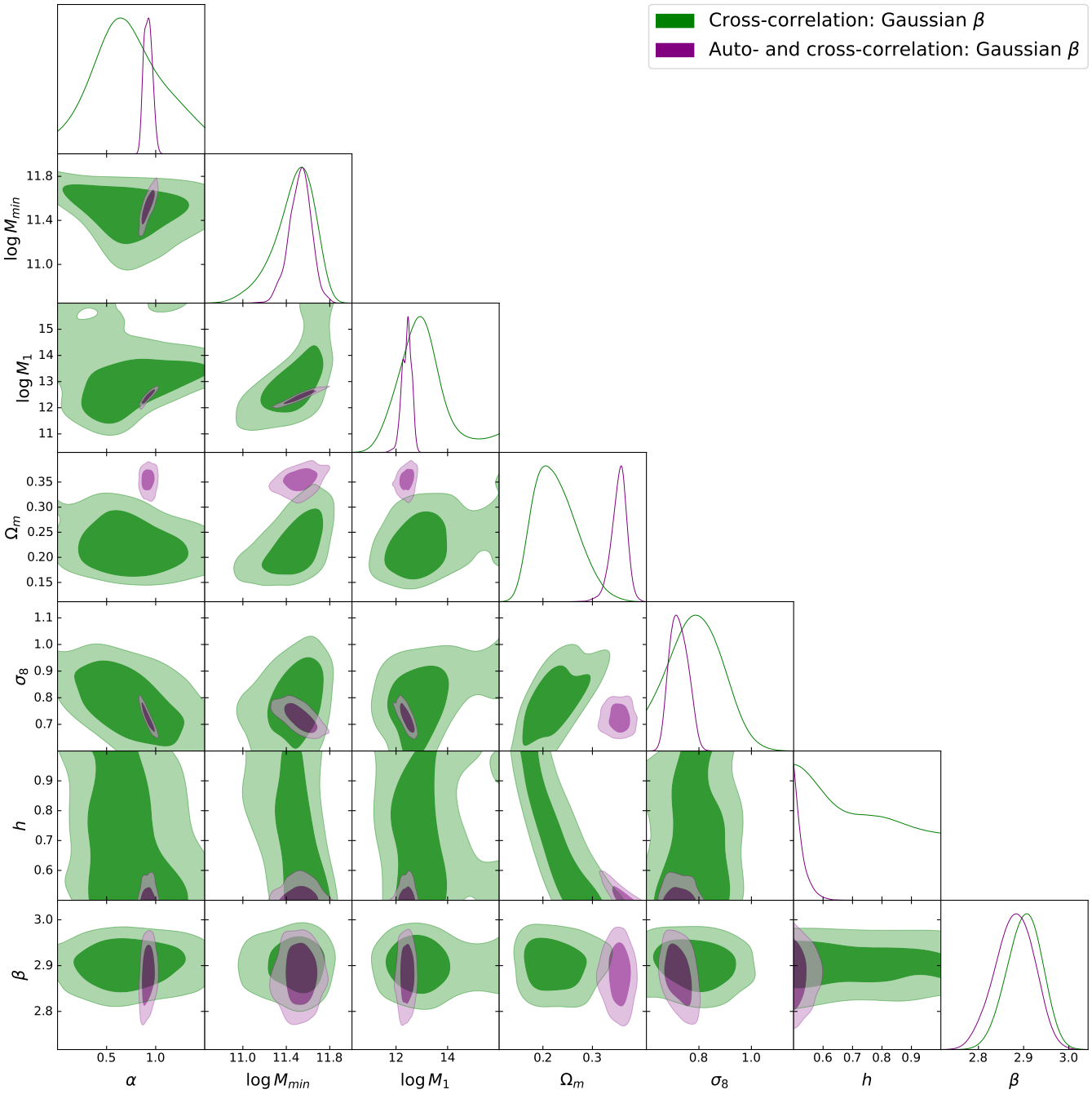

Fig. D.5. Marginalized posterior distributions and probability contours for the joint MCMC run on the cross- and auto-correlation functions in purple, compared with the corresponding ones using only the cross-correlation function in green. Both cases were run with a Gaussian prior on β.

Table D.5. Parameter prior distributions and summarized posterior results from the joint MCMC run on the auto- and cross-correlation functions with a Gaussian prior on β.

Parameter	Prior	Mean	Mode	68% CI
α	$\mathcal{U}[0.00, 1.50]$	0.92	0.92	$[0.87, 0.96]$
$\log M_{\rm min}$	$\mathcal{U}[10.00, 16.00]$	11.53	11.54	$[11.43, 11.63]$
$\log M_1$	$\mathcal{U}[10.00, 16.00]$	12.41	12.46	$[12.24, 12.64]$
Ω_m	$\mathcal{U}[0.10, 1.00]$	0.36	0.36	$[0.34, 0.37]$
σ_8	$\mathcal{U}[0.60, 1.20]$	0.72	0.71	$[0.69, 0.76]$
h	$\mathcal{U}[0.50, 1.00]$	0.54	–	$[0.50, 0.55]$
β	$\mathcal{N}[2.90, 0.04]$	2.88	2.88	$[2.84, 2.93]$

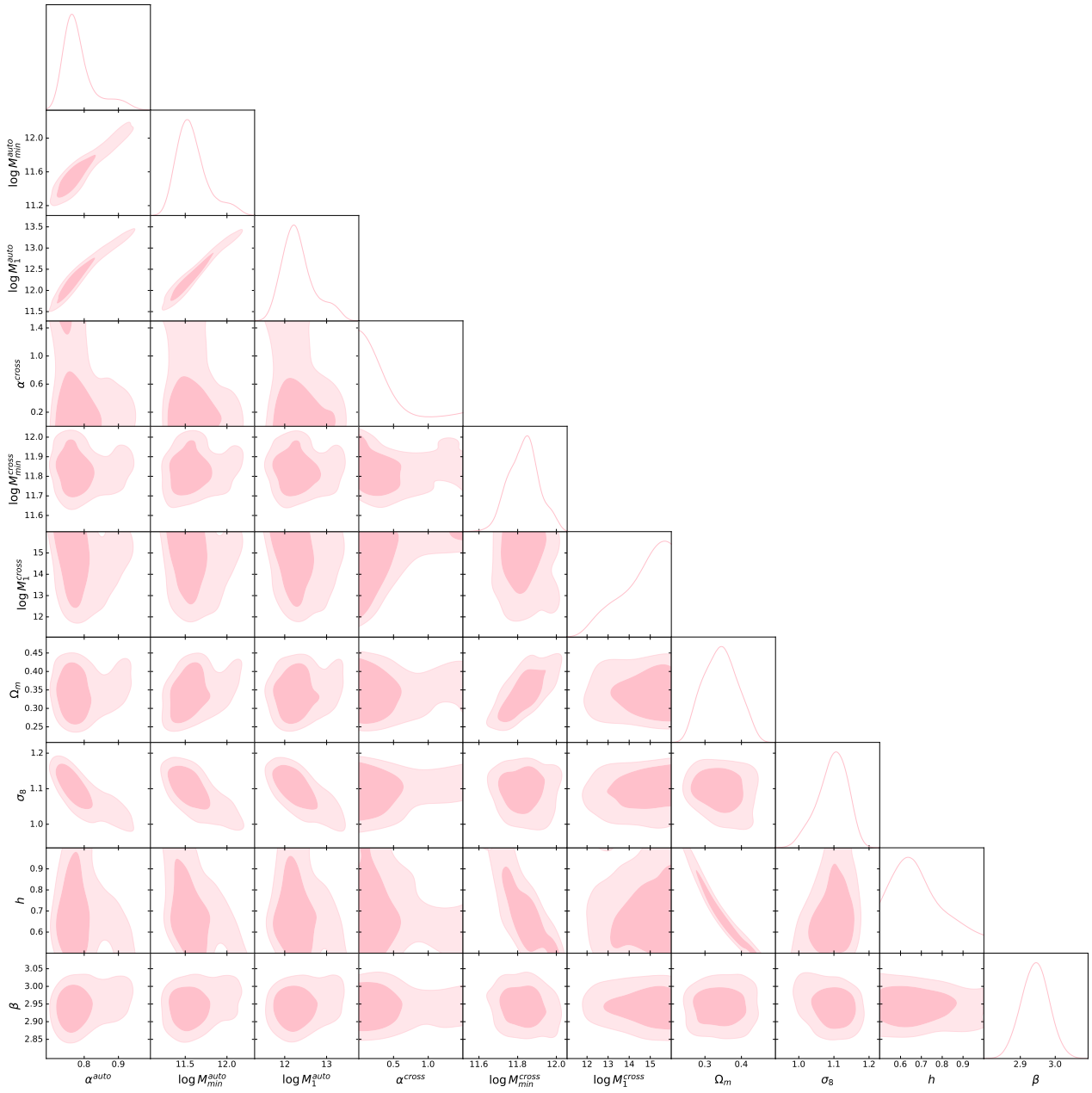

Fig. D.6. Marginalized posterior distributions and probability contours for the MCMC run on the cross-and auto-correlation functions assuming different HOD models for each observable.

Table D.6. Parameter prior distributions and summarized posterior results from the joint MCMC run on the auto- and cross-correlation functions assuming different HOD models for each observable.

Parameter	Prior	Mean	Mode	68% CI
$\alpha^{\rm auto}$	$\mathcal{U}[0.00, 1.50]$	0.79	0.76	$[0.73, 0.80]$
$\log M_{\rm min}^{\rm auto}$	$\mathcal{U}[10.00, 16.00]$	11.58	11.52	$[11.36, 11.70]$
$\log M_{\rm 1}^{\rm auto}$	$\mathcal{U}[10.00, 16.00]$	12.33	12.22	$[11.86, 12.57]$
$\alpha^{\rm cross}$	$\mathcal{U}[0.00, 1.50]$	0.41	–	$[0.00, 0.43]$
$\log M_{\rm min}^{\rm cross}$	$\mathcal{U}[10.00, 16.00]$	11.83	11.85	$[11.75, 11.91]$
$\log M_{\rm 1}^{\rm cross}$	$\mathcal{U}[10.00, 16.00]$	14.50	–	$[14.08, 16.00]$
Ω_m	$\mathcal{U}[0.10, 1.00]$	0.34	0.35	$[0.29, 0.39]$
σ_8	$\mathcal{U}[0.60, 1.20]$	1.10	1.11	$[1.06, 1.15]$
h	$\mathcal{U}[0.50, 1.00]$	0.69	0.64	$[0.52, 0.75]$
β	$\mathcal{N}[2.90, 0.04]$	2.94	2.95	$[2.90, 2.98]$

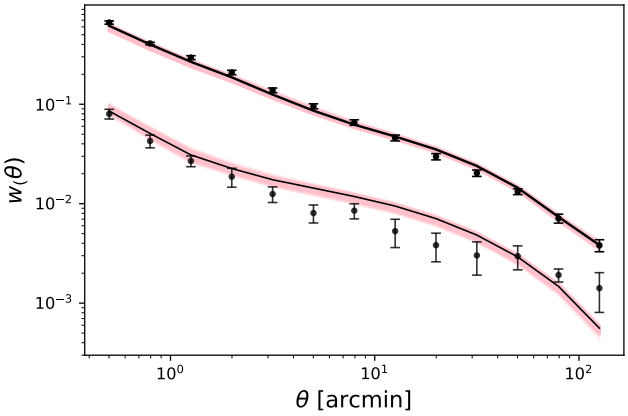

Fig. D.7. Posterior-sampled angular auto- and cross-correlation functions from the joint analysis with different HOD parameters. The data are shown in black.