All-sky angular power spectra from cleaned WISE×SuperCOSMOS galaxy number counts

H.S. Xavier,[a,1] **M.V. Costa-Duarte,**[a] **A. Balaguera-Antolínez**[b,c] **and M. Bilicki**[d,e]

[a]Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo,
Rua do Matão, 1226, Cidade Universitária, São Paulo, SP, Brazil

[b]Instituto de Astrofísica de Canarias, s/n, E-38205, La Laguna, Tenerife, Spain

[c]Departamento de Astrofísica, Universidad de La Laguna,
E-38206, La Laguna, Tenerife, Spain

[d]Leiden Observatory, Leiden University,
Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands

[e]Center for Theoretical Physics, Polish Academy of Sciences,
al. Lotników 32/46, 02-668, Warsaw, Poland

E-mail: hsxavier@if.usp.br, mvcduarte@astro.iag.usp.br, balaguera@iac.es,
bilicki@strw.leidenuniv.nl

Abstract. Aiming to extract cosmological information from linear scales of the WISE×SuperCOSMOS photometric redshift catalog, we perform a characterization of the systematic effects associated with stellar content, evidencing the presence of contamination and obscuration. We create an integrated model for these effects (which together we call "usurper contamination"), devise a method to remove both of them simultaneously and show its functionality by applying it to a set of mock catalogs. When administered to WISE×SuperCOSMOS data, our method shows to improve the measurements of angular power spectra on scales $\ell \lesssim 15$ and the extraction of cosmological parameters therefrom, even though a significant excess of power remains at these scales. When ignoring scales $\ell < 15$, we still find strong indications of systematics, albeit these can be localized in the southern equatorial hemisphere. An independent analysis of the northern hemisphere at $\ell \geq 15$ agrees with the ΛCDM model with parameters from the Planck satellite and gives $\Omega_c = 0.254 \pm 0.020$ and $\Omega_b < 0.065$ at 95% confidence limit when combined with priors on H_0, A_s and n_s.

[1]Corresponding author.

Contents

1 Introduction

Galaxy and quasar surveys are key observational tools in cosmology. Through them we can probe the matter distribution in the Universe and, therefore, test predictions of cosmological models and constrain their parameters [1–9]. A few interesting questions that have been tackled using wide-angle surveys are the homogeneity and isotropy of the Universe [10–12], the shape of the distribution function of galaxy density [13], the behavior of gravity and the growth of structure [14, 15], the accelerated expansion of the Universe and the properties of

dark energy [16, 17], the amount of matter in the Universe [18], the existence of primordial non-Gaussianities [19] and the mass of neutrinos [20, 21].

To improve constraints on the relevant cosmological parameters — especially on the equation of state of dark energy, the mass of neutrinos, and on non-Gaussianities and other inflation signatures — deeper and wider surveys are desirable [22–24]. Moreover, these surveys are required in order to improve our knowledge in various other interesting topics. For instance, the cross-correlation between galaxy distribution and gravitational lensing of the cosmic microwave background (CMB) — which serves as a consistency check for our cosmological model, besides helping improve constraints on halo bias and on the amplitude of matter perturbations — peaks at large angular scales [25]. This is also the case for the integrated Sachs-Wolfe effect [26], probed by the cross-correlation of CMB temperature maps with galaxy distribution, whose signal comes from the evolution of gravitational potential along the CMB photon path. Another topic that requires almost full-sky galaxy surveys is the measurement of our own peculiar velocity with respect to all other galaxies [27–29]. Large galaxy surveys can also help us improve our understanding of galaxy bias, that could be enhanced at the largest scales if we assume that galaxies form at density peaks (see e.g. [30]).

This type of studies are among the main aims of the largest forthcoming surveys. The Large Synoptic Survey Telescope[1] (LSST) [31] will cover $18,000$ deg^2 up to redshifts $z \lesssim 3$, while the Euclid space telescope[2] will observe 15,000 deg^2 and detect galaxies up to $z \lesssim 2$ [32]. The huge volume probed by these surveys will lead to such a high number of observed galaxies and quasars that spectroscopic measurements will be performed only for a small fraction of them, while for most sources we will have to rely on photometric redshifts (photo-zs hereafter) and photometric classification. These techniques are, however, plagued with high contamination levels, large systematic effects and uncertainties, which must be controlled and understood in order to fully mine the relevant datasets.

An existing testbed for such surveys is the WISE×SuperCOSMOS catalog [33] (WSC hereafter), a full-sky photometric galaxy dataset reaching $z \lesssim 0.4$. Our goal in this paper is to estimate cosmological parameters from this sample under the standard cosmological model ΛCDM. In this process, we characterize the systematic effects associated with stellar density (which, although more prominent in the WSC, will also be present in LSST and Euclid) and devise a method for mitigating such effects, tested on simulations and real data. These efforts were aimed at improving the survey's reach toward the largest scales ($\ell \lesssim 20$), commonly distrusted in cosmological analyses (e.g. [34, 35]). Finally, we also verified the impact of different assumptions about photo-z properties on the measured parameters.

This work adds to previous WSC analyses that have investigated its observational and cosmological properties, such as hemispherical anisotropies [36], Minkowski functionals [37], cross-correlation with CMB lensing [38, 39] and temperature [40]. It also contributes with a new strategy for mitigating systematics to those previously proposed [41–46].

The basic technique we employ here — the measurement of angular power spectra in tomographic redshift shells — has been applied before to other photometric galaxy catalogs such as the MegaZ [47] and the 2MASS Photometric Redshift datasets (2MPZ) [34], to photometric quasar samples [42] and spectroscopic galaxy data [35].

The outline of this paper is as follows: in section 2 we describe the datasets used in this work, namely the WSC, the Sloan Digital Sky Survey, and the Gaia Data Release 2 catalogs. In sections 3.1 and 3.2 we expose the stellar contamination and obscuration effects existent

---

[1]http://www.lsst.org/.

[2]http://www.euclid-ec.org.

in WSC. A method of modeling and mitigating these effects is presented in section 3.3. The WSC data had their angular power spectra measured and cosmological parameters inferred using the methodology described in section 4. Using simulations, we validate our treatment of systematic effects and our measurement methodology in section 5. Our results — WSC angular power spectra and inferred cosmological parameters — are shown in section 6. Section 7 presents tests for robustness and systematics, and in particular, section 7.7 shows evidence for a systematic effect represented by a tension in the cosmological parameters constrained from the angular clustering measured from different equatorial hemispheres. Finally, we conclude and summarize our work in section 8.

2 The dataset

The main dataset used in this work is the WISE×SuperCOSMOS (WSC) galaxy catalog [33], described in section 2.1. To estimate the stellar contamination and obscuration in WSC, we used the Sloan Digital Sky Survey (SDSS) [48] and Gaia DR2 [49, 50] photometric catalogs as tracers of stars. These are described in sections 2.2 and 2.3, respectively.

2.1 WISE×SuperCOSMOS

The WSC catalog is a cross-match between two parent full-sky catalogs: the AllWISE release [51] from the Wide-field Infrared Survey Explorer (WISE) [52], a mid-infrared space survey in four bands $W1$–$W4$ (3.4, 4.6, 12 and 22μm); and the SuperCOSMOS Sky Survey [53], a program of automated digitalization of optical photographic plates in the B, R and I filters, taken by the United Kingdom Schmidt Telescope (UKST, in the southern hemisphere) and the Palomar Observatory Sky Survey-II (POSS-II, in the northern hemisphere).

To build the WSC catalog, only the bands $W1$, $W2$, B and R were used. Following the magnitude limits of the parent surveys, all WSC sources have magnitudes $W1 < 17$, $B < 21$ and $R < 19.5$. To preferentially select galaxies over quasars and stars, WSC only includes extended sources according to SuperCOSMOS morphological classification `meanClass`$= 1$. It should be noted that the SuperCOSMOS morphological classification is not as accurate as most recent optical surveys; its angular resolution is approximately $2''$ [54], and the quality of the imaging technique is lower (photographic plates vs. CCDs).

We also point out that the WSC star/galaxy separation is not based only on Super-COSMOS morphology but also on WISE colours [33]. WSC sources have $W1 > 13.8$ (a cut aimed at removing bright stars), $R - W2 < 7.6 - 4(W1 - W2)$ and $W1 - W2 < 0.9$ (two color-cuts aimed at removing quasars). In this work, we further enforced: the requirement $W1 - W2 > 0.2$ for all sources to reduce stellar contamination and maintain a constant galaxy selection function across the sky (in opposition to a cut dependent on Galactic latitude [33]); maximum Galactic extinction, $E(B - V) < 0.10$; removal of sources in highly contaminated regions such as near the Galactic plane and bulge, around the Magellanic Clouds and in regions presenting unusually high densities according to a lognormal distribution, all accomplished with the WSC final mask [33]; removal of sources in regions where Gaia stellar density is greater than 7 times its average (this increases the masked region around the bulge; see section 2.3); and removal of a SuperCOSMOS tile with bad photometry.

The photo-zs for all the WSC sources were obtained with the artificial neural network code ANNz [56],[3] trained on a WSC cross-match with the complete and deep Galaxy And

---

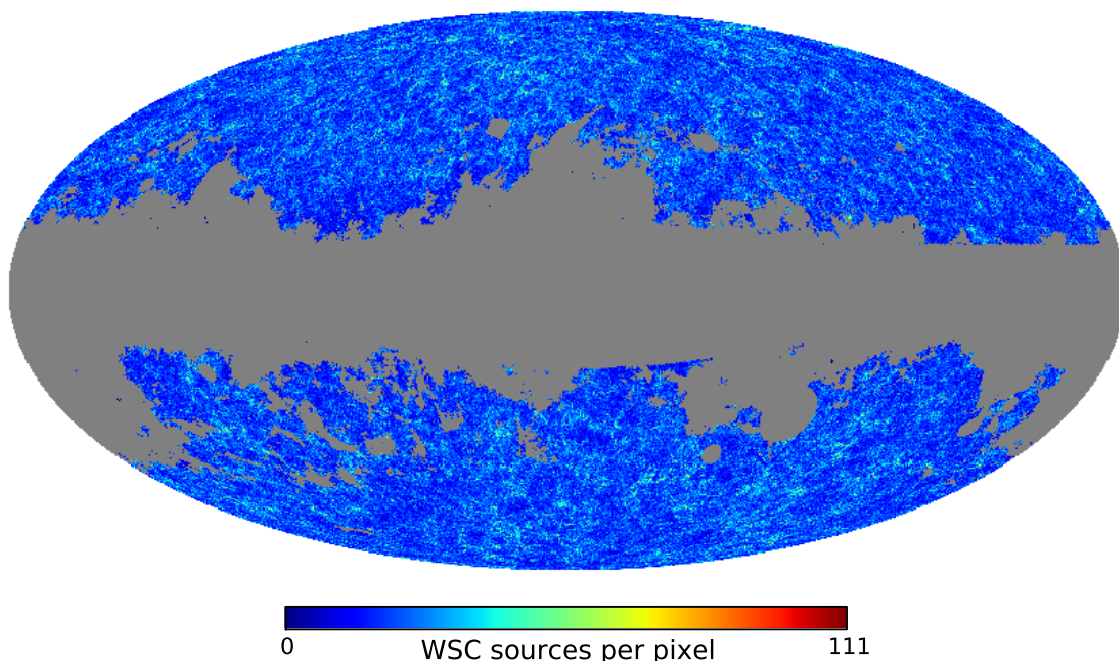[3]http://www.homepages.ucl.ac.uk/~ucapola/annz.html.

Figure 1. Sky map of number counts of WSC sources (mostly galaxies) in the photo-z range $0.10 < z < 0.35$ in Galactic coordinates, after applying all the cuts described in section 2, using Healpix [55] resolution parameter $N_{\text{side}} = 256$. The gray regions were masked out.

Mass Assembly (GAMA)-II spectroscopic dataset [57]. The adequacy of GAMA for photo-z training in WSC has been discussed and demonstrated in [33]. As explained there, these photo-zs were further corrected for asymmetries between their northern and southern distributions, likely caused by differences between the POSS-II and UKST pass-bands. That paper contains also a comprehensive analysis of photo-z properties based on comparisons with several external spectroscopic datasets. A projection of all sources in the photometric redshift (photo-z) range $0.10 < z < 0.35$ is presented in figure 1, and figure 2 shows the photo-z distribution of WSC galaxies. In this work we binned the WSC sources in photo-z shells of width $\Delta z = 0.05$, covering the range $0.10 < z < 0.35$. However, for reasons presented in section 3, most of our analysis ignores the first bin ($0.10 < z < 0.15$).

It was shown in [38] that the WSC photo-z errors can be modeled by a Gaussian with zero mean (i.e. no bias) and standard deviation $\sigma(z_{\text{p}}) = 0.02 + 0.08 z_{\text{p}}$. Also, that an even better model would be a generalized Lorentzian, where the probability density of a galaxy with photo-z z_{p} to have true redshift z_{s} is:

$$p(z_{\text{s}}|z_{\text{p}}) = \frac{\sqrt{2\pi a(z_{\text{p}})}\Gamma\left[a(z_{\text{p}}) - \frac{1}{2}\right]}{\Gamma\left[a(z_{\text{p}})\right]} \left[1 + \frac{(z_{\text{s}} - z_{\text{p}})^2}{2a(z_{\text{p}})s^2(z_{\text{p}})}\right]^{-a(z_{\text{p}})}, \tag{2.1}$$

where $\Gamma(x)$ denotes the Gamma function. The WSC photo-zs follows $a(z) = 3 - 4z$ and $s(z) = 0.02 + 0.04z$.
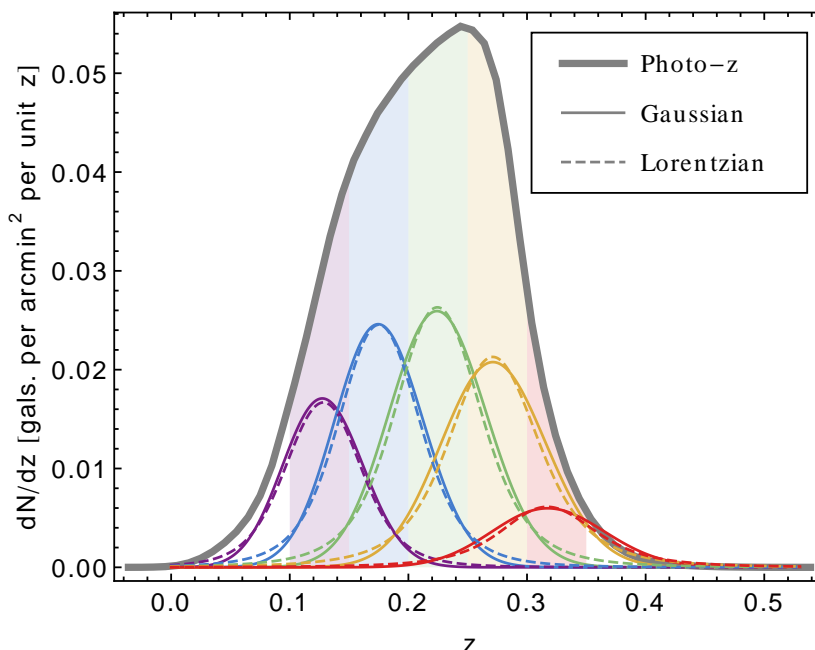
Figure 2. Redshift distribution of WSC sources, given in number of galaxies per arcmin^2, per unit redshift. The thick gray line represents their distribution in terms of their photo-zs, after all cuts described in section 2 have been applied. The thin, colored lines represent their estimated spectroscopic redshift distributions inside each photo-z bin (represented as colored bands), assuming Gaussian (solid lines) and generalized Lorentzian (dashed lines) photo-z errors, both computed with eq. (2.2).

The spectroscopic redshift (spec-z) distribution $\mathrm{d}N^i/\mathrm{d}z_\mathrm{s}$ of the galaxies in our photo-z bin i can be estimated from a convolution of $p(z_\mathrm{s}|z_\mathrm{p})$ and the photo-z distribution $\mathrm{d}N/\mathrm{d}z_\mathrm{p}$:

$$\frac{\mathrm{d}N^i}{\mathrm{d}z_\mathrm{s}} = \int_{z^i_\mathrm{min}}^{z^i_\mathrm{max}} p(z_\mathrm{s}|z_\mathrm{p})\frac{\mathrm{d}N}{\mathrm{d}z_\mathrm{p}}\mathrm{d}z_\mathrm{p}, \tag{2.2}$$

where the ith bin in photo-z ranges from z^i_min to z^i_max. The resulting distributions in the different redshift bins are shown in figure 2. We verified that the expected spec-z distributions in each bin are very well fitted by a Gaussian (generalized Lorentzian) when the photo-z errors are Gaussian (generalized Lorentzian). Also, one can see in figure 2 that the mean values of the spec-zs in every bin are practically independent of the distribution assumed for the photo-z errors, whereas the FWHM may depend on it. Thus, when computing theoretical angular power spectra C_ℓs to fit the data, we decided to keep the means of the spec-z distributions fixed and let the widths vary as nuisance parameters.

2.2 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS, [48]) is one of the largest astronomical surveys in operation. Here we use its 14$^{\text{th}}$ Data Release (DR14) [58]. The survey has already covered $\sim 1/3$ of the northern hemisphere, imaging in the five $ugriz$ broad bands. The effective photometric limit of the survey is $r = 22.2$ (95% completeness for point sources). In addition, its spectroscopic counterpart is complete down to $r = 17.77$ for the Main Galaxy Sample (MGS) and deeper for other sub-projects, such as the Baryon Oscillation Spectroscopic Survey (BOSS) [16].
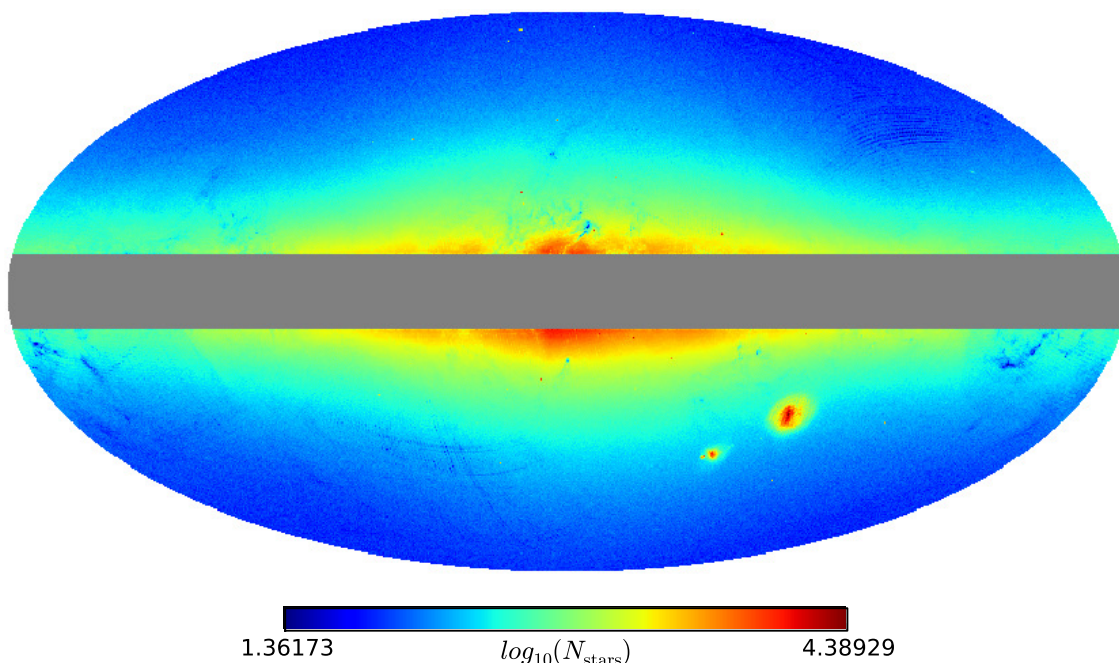
Figure 3. Sky map of number counts (sources per pixel, using $N_{\text{side}} = 256$, in logarithmic scale) at Galactic latitudes $|b| > 10°$ of Gaia DR2 sources at $G < 20$ magnitude limit, after removing those with both parallax and proper motion measurements consistent with zero.

Despite containing a wealth of cosmological information, the SDSS data was only used here to verify the existence of stellar contamination and obscuration in the WSC catalog, as described in section 3. The SDSS star-galaxy separation classifies extended objects as galaxies and those point-like as stars. This classification is based on a difference between two types of magnitudes, namely sources with psfMag − cModelMag > 0.145 are classified as galaxies and otherwise as stars.[4] The PSF of 1.24 arcsec (full width at half maximum, FWHM) allows for a reliable star-galaxy separation up to $r < 21.5$ [59] and accurate astrometric positions (< 0.1 arcsec per coordinate) at $r < 20.5$ [60].

2.3 Gaia

Gaia is a space mission of the European Space Agency (ESA) aimed at measuring the three-dimensional positions and velocities of about 10^9 Milky Way stars [49]. It uses two optical telescopes of 0.7m^2 collecting area equipped with astrometry, photometry and spectrometry instruments. The astrometry is performed in the range $330 − 1050$ nm (G-band hereafter). Gaia Data Release 2 (DR2, [50]) contains 1.3×10^9 point sources with measured parallaxes and proper motions at ~ 1 milliarcsecond (mas) and ~ 1 mas per year (mas y^{-1}) precision, respectively, up to a G band magnitude of 20.

---

[4]The two types of magnitudes psfMag and cModelMag represent magnitudes obtained using the point spread function (PSF) model and the best linear combination of exponential and de Vaucouleurs profiles, respectively. Note that the 'stars' according to this definition will include any point sources, also extragalactic ones such as quasars.

In order to create a representative full-sky stellar density map using the Gaia DR2 source catalog, we had to ensure minimal observational effects (i.e. anisotropic completeness) and minimal contamination by galaxies and quasars. To achieve the first goal we applied a magnitude cut $G < 20$; as the limiting magnitudes in the regions of interest (away from the Galactic plane) are $G \gtrsim 20.8$ [61], we expect uniform sky completeness under this selection criterion. The second goal was achieved by taking advantage of the excellent astrometric measurements of Gaia: we assumed that the majority of observed stars should have either a non-zero parallax or a non-zero total proper motion. Thus, we enforced, for all the selected sources, a 5σ detection of at least one of these quantities. We computed the total proper motion as $\mu = \sqrt{\mu_\delta^2 + \mu_{\alpha'}^2}$, where $\mu_{\alpha'} \equiv \mu_\alpha \cos\delta$, μ_α is the proper motion in the direction of right ascension α and μ_δ is the proper motion in the direction of declination δ. The uncertainty on μ was computed with the usual error propagation, taking into account the covariance between μ_δ and $\mu_{\alpha'}$ measurement errors.

In order to avoid regions of high stellar density, here we only selected sources at Galactic latitudes $|b| > 10°$. A map of the $282,045,470$ selected Gaia sources, in Galactic coordinates, is presented in figure 3.

3 Characterizing and cleaning the WSC catalog

3.1 Stellar contamination

As described in [33], the WSC catalog is affected by considerable stellar contamination, especially near the Galactic plane. In order to demonstrate this contamination, we follow the strategy applied in [33] of cross-matching WSC with SDSS photometric sources, taking advantage of the latter's better star-galaxy separation. Our final analysis, however, is based on a stellar density template derived from Gaia's observations (see section 3.3).

We cross-matched the WSC sample and the photometric SDSS DR14, linking the nearest objects in the latter (within 2 arcsec) to objects in the former. In total, our cross-matched sample consists of $4,282,564$ sources which includes both stars and galaxies. The result of this cross-match is shown in figure 4. From the bottom panel, it is quite clear that stellar contamination exists and increases towards the Galactic plane and bulge. Also, there are hints that the WSC galaxies are partly obscured by stars, since the top panel shows a deficit of galaxies in the same regions where stars creep in. Overall, 96% percent of the matched sources are galaxies and 4% are stars, according to SDSS.

Figure 5 shows how the amount of stellar contamination changes with redshift. We can see that, at high redshifts, the contamination (i.e. stellar fraction) increases significantly. This increase (and the one at low redshift) are mostly caused by a drop in the number of observed galaxies and not by an increase in the number of misclassified stars (see the top panel).

The average dependence with Galactic latitude b of the number of stars contaminating WSC (according to SDSS classification) is shown in figure 6. We see that both Galactic hemispheres, as well as different redshift ranges, follow very similar trends. For instance, the binned redshift curves differ from the full range one mostly by a constant factor. Despite these similarities, subtle slope changes and hemispherical asymmetries are noticeable.

Figure 6 also shows the result of an exponential fit to the average number of stars per pixel n_{stars} in the range $0.10 < z < 0.35$ as a function of Galactic latitude b:

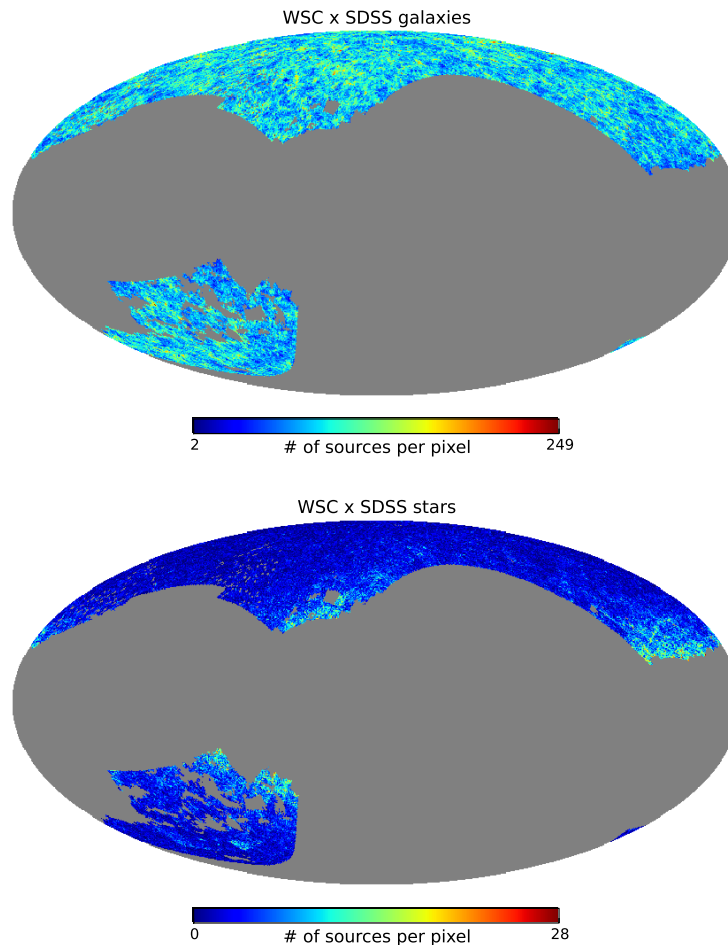$$n_{\text{stars}}(b) = \bar{n} e^{-a|b|} + c, \tag{3.1}$$

Figure 4. Number counts map in Galactic coordinates of the cross-match between SDSS and our WSC sources, selected according to section 2.1. In this particular case we used $N_{\text{side}} = 128$ to improve visualization. The top (bottom) panel shows the cross-match for the SDSS sources classified as galaxies (stars).

where $\bar{n} = 7.14$, $a = 0.06$ and $c = 0.35$; in the plot, the fit was scaled by 0.71 to ease visualization. We can see that the fit represents the stellar density reasonably well.

3.2 Stellar obscuration

As shown in figure 4 and already noted in previous works with SDSS data [41], the density of observed galaxies anti-correlates with the stellar density: regions with high concentration of stars tend to present lower galaxy counts. We call this phenomenon *stellar obscuration*. This is expected to affect mostly photometric datasets, such as WSC, due to issues such as blending and source misidentification. Spectroscopic surveys can avoid stellar contamination given that spectroscopy can easily separate stars from galaxies, but even they suffer from such obscuration, as shown in [41].

The approach used in [41] to estimate stellar obscuration was to correlate the observed number of galaxies with the density of foreground stars: as galaxies are, on average, homogeneously distributed on the sky, these two should be uncorrelated in the absence of systematic
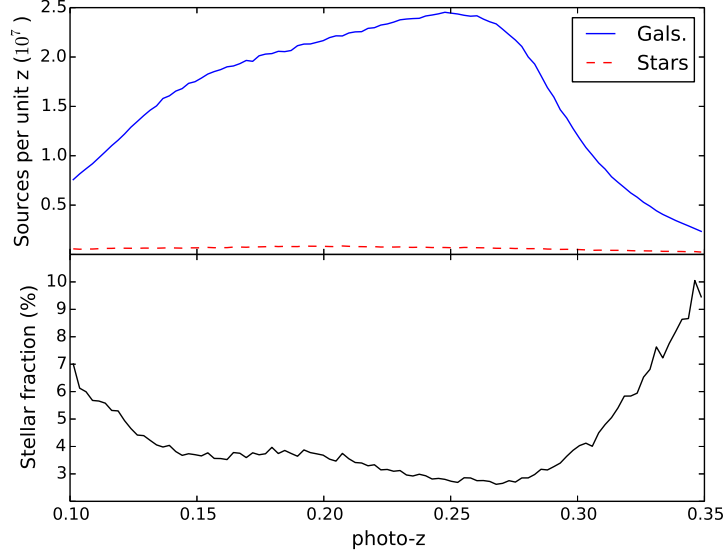
Figure 5. *Top:* total number (per unit redshift) of WSC sources matched with SDSS galaxies (blue solid line) and stars (red dashed line). *Bottom:* fraction of matched sources that are stars according to SDSS, as a function of WSC photo-z.

effects. This approach is feasible in the spectroscopic SDSS case because the galaxy sample there is free from stellar contamination. For WSC sources that include a large amount of stars disguised as galaxies, however, this approach is not possible: correlating WSC sources with stellar density would mix contributions from obscuration and contamination.

An alternative approach is to look at the standard deviation of the counts of galaxies inside pixels as a function of stellar density. The basic idea is that stellar obscuration will not only modulate the mean observed density $\bar{n}_{\rm obs}$ but also the cosmological fluctuations around the mean, since the former multiplies the latter. Let us formalize this statement through a model for the observed projected density of sources $n_{\rm obs}^i(\boldsymbol{\theta})$ at angular position $\boldsymbol{\theta}$, for a particular redshift bin i, that takes into account both contamination and obscuration:

$$n_{\rm obs}^i(\boldsymbol{\theta}) = W(\boldsymbol{\theta})\{[1 - \alpha_i S(\boldsymbol{\theta})]n_{\rm g}^i(\boldsymbol{\theta}) + \beta_i S(\boldsymbol{\theta}) + \epsilon_{\rm g,i}(\boldsymbol{\theta}) + \epsilon_{\rm s,i}(\boldsymbol{\theta})\}. \tag{3.2}$$

In the equation above, $W(\boldsymbol{\theta})$ is the survey binary window function, $S(\boldsymbol{\theta})$ is a template for the expected density of stars, $n_{\rm g}^i(\boldsymbol{\theta})$ is the real galaxy density, and $\epsilon_{\rm g,i}(\boldsymbol{\theta})$ and $\epsilon_{\rm s,i}(\boldsymbol{\theta})$ are Poisson fluctuations in the number of observed galaxies and stars, respectively. The non-negative parameters α_i and β_i control the amount of obscuration and contamination in the ith bin, respectively. The noise terms obey the statistical properties $\langle\epsilon_{\rm g,i}(\boldsymbol{\theta})\rangle = \langle\epsilon_{\rm s,i}(\boldsymbol{\theta})\rangle = 0$, $\langle\epsilon_{\rm g,i}(\boldsymbol{\theta})\epsilon_{\rm s,i}(\boldsymbol{\theta})\rangle = 0$, $\langle\epsilon_{\rm g,i}^2(\boldsymbol{\theta})\rangle = W(\boldsymbol{\theta})[1 - \alpha_i S(\boldsymbol{\theta})]\bar{n}_{\rm g,i}$ and $\langle\epsilon_{\rm s,i}^2(\boldsymbol{\theta})\rangle = W(\boldsymbol{\theta})\beta_i S(\boldsymbol{\theta})$, such that the variance of the total noise is equal to the expected number of observed sources, $\langle[\epsilon_{\rm g,i}(\boldsymbol{\theta}) + \epsilon_{\rm s,i}(\boldsymbol{\theta})]^2\rangle = \langle n_{\rm obs}^i(\boldsymbol{\theta})\rangle$.

Our model for the observed sources (eq. (3.2)) makes three important simplifications: first, the number of observed galaxies is linearly suppressed by the stellar density; second, both contamination and obscuration depend on the same template; third, the template $S(\boldsymbol{\theta})$ is independent of redshift, although the final obscuration and contamination are modulated by bin-dependent parameters α_i and β_i. The fact that both obscuration and contamination
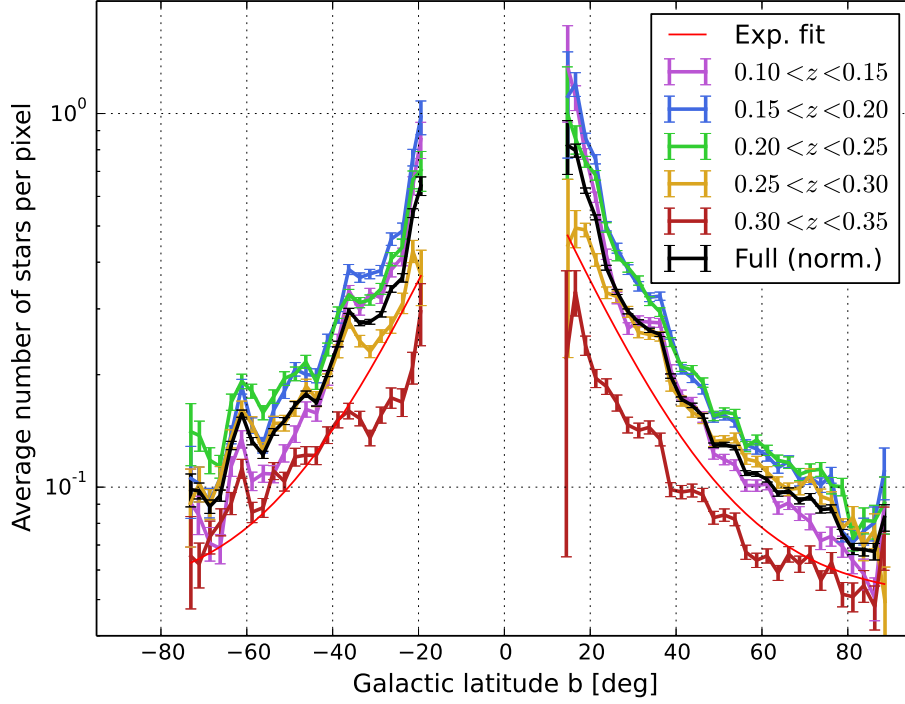
Figure 6. Average number of WSC sources matched to SDSS stars per pixel ($N_{\mathrm{side}} = 256$) as a function of Galactic latitude b. Each thick colored curve (with error bars) represents a different photo-z bin, and the black curve represents the full range $0.10 < z < 0.35$, normalized by 5 to ease comparison with the other curves. The error bars represent Poisson noise inside the b bins used to build the plot. The thin red line shows the result of an exponential fit (eq. (3.1)) to the black line, scaled by 0.71 to unclutter the plot.

follow the same template leads to a replacement of galaxies for stars, and thus we call this combined effect "usurper contamination".

To estimate α_i from the observed variance of n^i_{obs}, we first write $\boldsymbol{\theta}$ as a pair of generalized coordinates $\boldsymbol{\theta} = (s, \lambda)$, where s specifies an isocontour of $S(\boldsymbol{\theta})$ while by varying λ we move along this isocontour. We then compute the mean of n^i_{obs} for a fixed s:

$$\bar{n}^i_{\mathrm{obs}}(s) = \frac{1}{A(s)} \int n^i_{\mathrm{obs}}(s, \lambda) \mathrm{d}\lambda = [1 - \alpha_i S(s)] \bar{n}_{\mathrm{g},i}(s) + \beta_i S(s), \tag{3.3}$$

where

$$A(s) \equiv \int W(s, \lambda) \mathrm{d}\lambda, \tag{3.4}$$

and

$$\bar{n}_{\mathrm{g},i}(s) \equiv \frac{1}{A(s)} \int W(s, \lambda) n^i_{\mathrm{g}}(s, \lambda) \mathrm{d}\lambda. \tag{3.5}$$

Note that due to our choice of coordinates, the value of $S(\boldsymbol{\theta})$ only depends on s. Also, we assumed that when averaging the noise $\epsilon_{\mathrm{g},i} + \epsilon_{\mathrm{s},i}$ along λ, we may approximate the result to zero. Now we can define the difference $\delta n_{\mathrm{obs},i}(s, \lambda) \equiv n^i_{\mathrm{obs}}(s, \lambda) - W(s, \lambda) \bar{n}^i_{\mathrm{obs}}(s)$ and compute

the variance of source number counts for an isocontour s of $S(\boldsymbol{\theta})$:

$$\sigma_{\mathrm{obs},i}^2(s) = \frac{1}{A(s)} \int \delta_{\mathrm{obs},i}^2(s,\lambda)\mathrm{d}\lambda \simeq [1 - \alpha_i S(s)]^2 v_{\mathrm{g}}^i + \bar{n}_{\mathrm{obs}}^i(s), \qquad (3.6)$$

where $v_{\mathrm{g}}^i = \langle [n_{\mathrm{g}}^i(s,\lambda) - \bar{n}_{\mathrm{g},i}(s)]^2 \rangle$ is the true galaxy variance (i.e. it does not include systematic effects nor shot-noise). Eq. (3.6) assumes that the mean source number counts in s, $\bar{n}_{\mathrm{obs}}^i(s)$, approximates the expected value $\langle n_{\mathrm{obs}}^i(s,\lambda) \rangle$, and that v_{g}^i is independent of the angular position. By placing all directly measurable quantities on the left hand side, we get a function of s that might be fitted to extract the obscuration parameter α_i:

$$\sigma_{\mathrm{obs},i}^2(s) - \bar{n}_{\mathrm{obs}}^i(s) = [1 - \alpha_i S(s)]^2 v_{\mathrm{g}}^i. \qquad (3.7)$$

In the absence of obscuration, the quantity on the left would be the variance due to cosmological fluctuations (and thus we call it $\sigma_{\mathrm{cosmo}}^2$) and it should be constant all over the sky. We point out that, in practice, the estimation of $\sigma_{\mathrm{cosmo}}^2$ is performed in a discrete way [the integral in eq. (3.6), for instance, is actually a sum over pixels], and the isocontours are approximated by bands of similar $S(\boldsymbol{\theta})$ values.

Figure 7 shows σ_{cosmo} as a function of Galactic latitude b for each photo-z bin. It is completely evident that the amplitude of the observed fluctuations is not constant and decreases towards the Galactic plane, quantitatively proving the existence of obscuration in the WSC data. This figure also shows that our obscuration model, given by the square root of eq. (3.7) (depicted as a red line) describes really well the overall behavior of this effect (except for the first bin). Here it is important to emphasize that only α_i and v_{g}^i were fitted to σ_{cosmo}, while $S(s)$ was held fixed according to eq. (3.1), estimated from contamination. Thus, figure 7 also shows that our assumption that both obscuration and contamination follow the same template $S(s)$ is reasonable.

The error bars in figure 7 were computed by error propagation from the uncertainties on $\bar{n}_{\mathrm{obs}}^i(s)$ and $\sigma_{\mathrm{obs},i}(s)$, which are given by $\sigma_{\mathrm{obs},i}(s)/\sqrt{N(s)}$ and $\sigma_{\mathrm{obs},i}(s)/\sqrt{2[N(s)-1]}$, respectively [$N(s)$ is the number of pixels taken into account in each isocontour]. We interpret the fact that the scatter of the data points is larger than the one expected from the error bars as an indication that v_{g}^i actually varies from one isocontour to another due to large-scale density fluctuations that are the size of the isocontours. It might also indicate that other factors are in play (e.g. Galactic extinction, seeing and bad photographic plates) or that the template $S(\boldsymbol{\theta})$ can be improved.

3.3 Cleaning the WSC projected galaxy distribution

We can note from figure 6 that the stellar contamination is not perfectly symmetric across the Galactic plane. Moreover, figure 3 shows that the stellar density (and likely contamination) is larger near the Galactic bulge, as expected. Aiming at improving the removal of contamination and obscuration from WSC projected galaxy distribution, we built a stellar density template $S(\boldsymbol{\theta})$ from the Gaia DR2 map presented in figure 3, which should accurately represent the stellar distribution — and thus contamination and obscuration — all over the sky.

A few unusually bright pixels — with excessive object counts due to the presence of globular clusters and nearby dwarf galaxies — can be seen in figure 3. To remove these count peaks, we computed, for all the pixels, the median of the number counts in neighboring pixels; if the central pixel value was more than twice the median, we replaced the former by the latter. Then, to attenuate the Poisson noise (visible in figure 6), we applied Gaussian
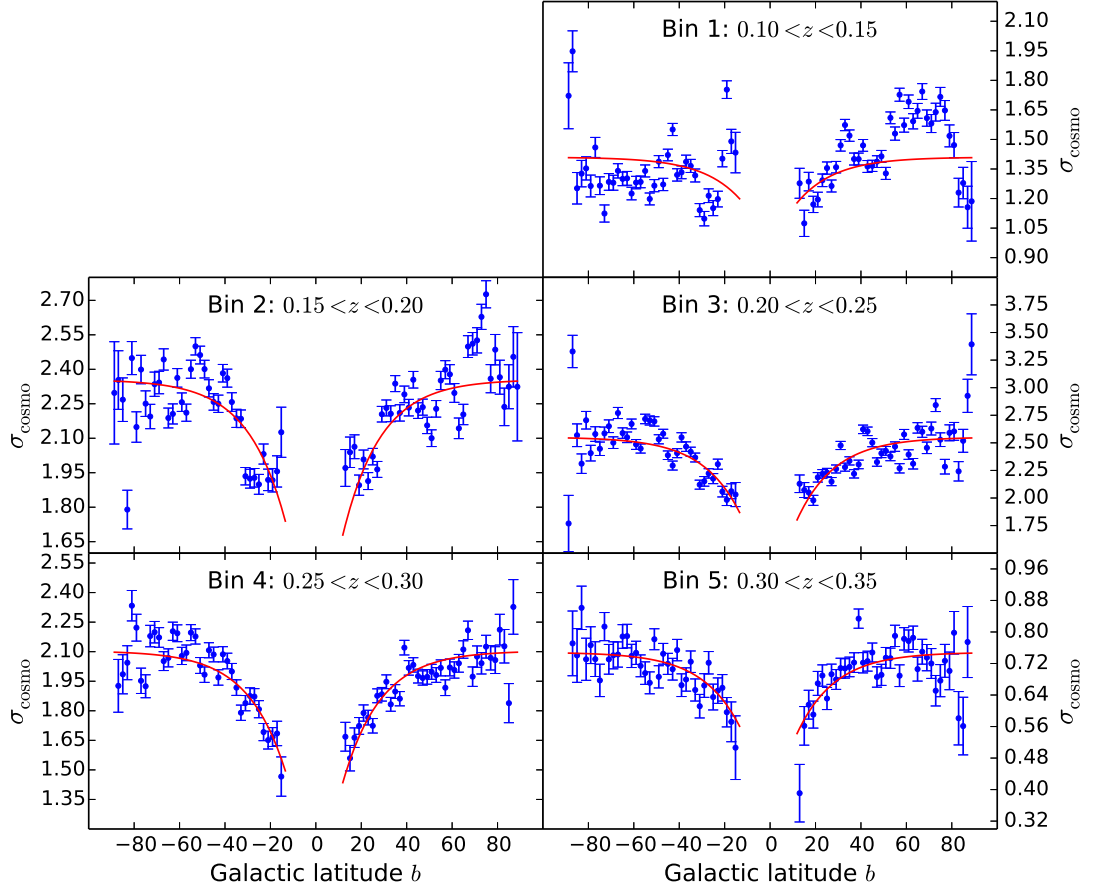
Figure 7. Each panel shows, for a different redshift bin, the cosmological contribution to the standard deviation of WSC number counts in pixels ($N_{\mathrm{side}} = 256$), computed inside bands of fixed Galactic latitude b (blue data points). The red curve is the fitted model given by the square root of eq. (3.7), with stellar template $S(\boldsymbol{\theta})$ given by eq. (3.1) (fixed according the contamination estimated in section 3.1).

smoothing with standard deviation $\sigma_{\mathrm{G}} = 34.4$ arcmin in order to retain all the structure in our template up to the scales of interest. Finally, we applied the same mask as in the data (see section 2.1) and normalized the template by its mean value (the template's absolute scale does not affect the data treatment). The final result is shown in figure 8.

Using the template from figure 8, we estimated α_i for all redshift bins with eq. (3.7). Figure 9 shows the data variance in template isocontours as a function of the template's value and our fit. Again, our model describes reasonably well the obscuration effect, apart from the first bin.

We next removed the contamination using a slightly modified version of the approach proposed in [46, 62], i.e. we computed the correlation of the (weighted) data with the con-
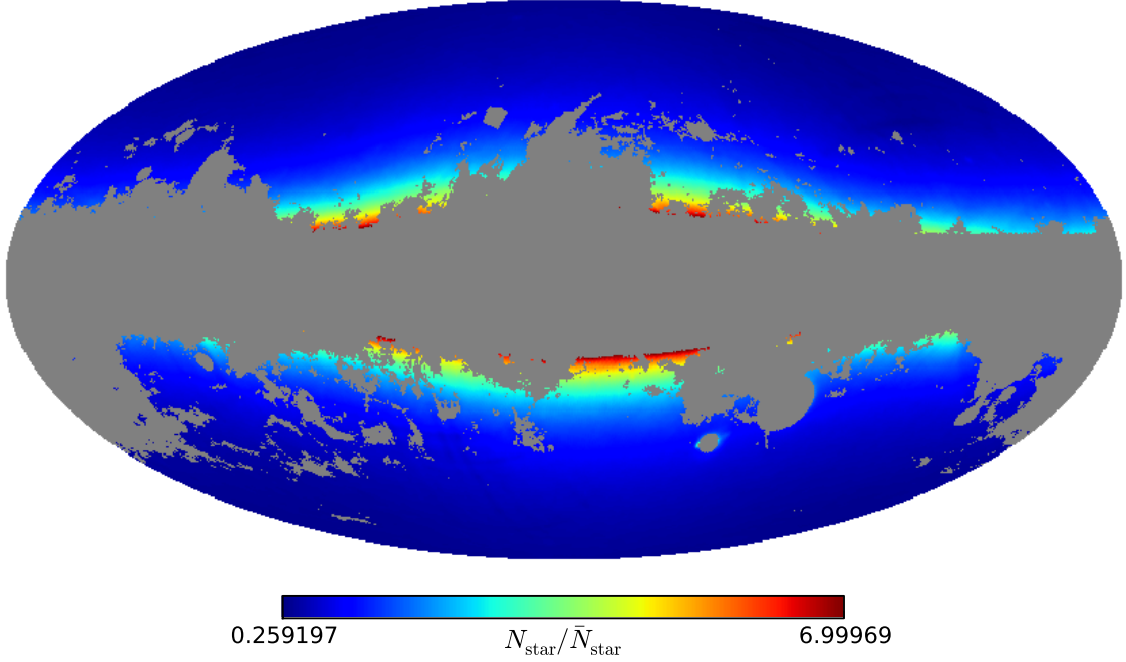
Figure 8. Stellar density template, derived from Gaia DR2 sources selected as described in section 2.3, used to estimate contamination and obscuration in WSC projected galaxy distribution.

tamination template. The weighted number counts are defined as:

$$n_{\rm w}^i(\boldsymbol{\theta}) \equiv \frac{n_{\rm obs}^i(\boldsymbol{\theta})}{1 - \hat{\alpha}_i S(\boldsymbol{\theta})} \simeq W(\boldsymbol{\theta}) \left[n_{\rm g}^i(\boldsymbol{\theta}) + \beta_i \frac{S(\boldsymbol{\theta})}{1 - \hat{\alpha}_i S(\boldsymbol{\theta})} + \frac{\epsilon_{{\rm g},i}(\boldsymbol{\theta}) + \epsilon_{{\rm s},i}(\boldsymbol{\theta})}{1 - \hat{\alpha}_i S(\boldsymbol{\theta})} \right]. \qquad (3.8)$$

In the equation above, $\hat{\alpha}_i$ is our estimate for the obscuration parameter, and the first term inside the square brackets shows we assumed $\hat{\alpha}_i = \alpha_i$. Deviations from this assumption were studied in [44] and they mostly lead to a small re-scaling of the angular power spectrum; we also verified this effect in our simulations (see section 5). We call the term inside the brackets that multiplies β_i the *weighted template*.

Now we can compute the average weighted density $\bar{n}_{\rm w}^i$ in the redshift bin i over the survey footprint $W(\boldsymbol{\theta})$ and do the same for the weighted template:

$$\bar{S}_{\rm w}^i \equiv \frac{1}{A} \int \frac{S(\boldsymbol{\theta})}{1 - \hat{\alpha}_i S(\boldsymbol{\theta})} {\rm d}^2\theta, \qquad (3.9)$$

where A is the unmasked area of the survey. Assuming that the Poisson noise and the true galaxy overdensity $\delta n_{\rm g}^i(\boldsymbol{\theta}) \equiv n_{\rm g}^i(\boldsymbol{\theta}) - \bar{n}_{{\rm g},i}$ do not correlate with our weighted template difference $\delta S_{\rm w}^i(\boldsymbol{\theta}) \equiv S_{\rm w}^i(\boldsymbol{\theta}) - \bar{S}_{\rm w}^i$, β_i can be estimated by:

$$\beta_i \simeq \frac{\int \delta n_{\rm w}^i(\boldsymbol{\theta}) \delta S_{\rm w}^i(\boldsymbol{\theta}) {\rm d}^2\theta}{\int [\delta S_{\rm w}^i(\boldsymbol{\theta})]^2 {\rm d}^2\theta}, \qquad (3.10)$$

where $\delta n_{\rm w}^i(\boldsymbol{\theta}) \equiv n_{\rm w}^i(\boldsymbol{\theta}) - \bar{n}_{\rm w}^i$. In practice, we may expect a small random correlation between $\delta n_{\rm g}^i(\boldsymbol{\theta})$ and $\delta S_{\rm w}^i(\boldsymbol{\theta})$, which was investigated and corrected for in [46]. We nevertheless ignore

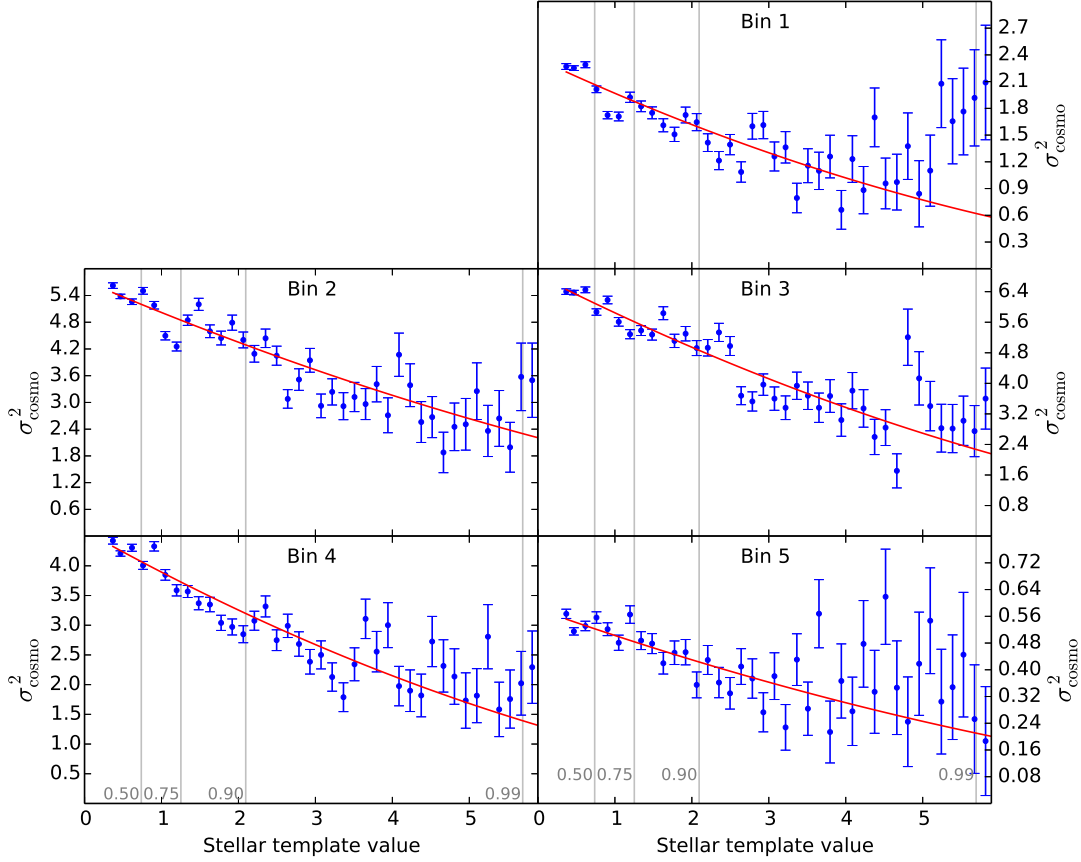**Figure 9**. Variance attributed to cosmological fluctuations σ^2_{cosmo} as a function of the stellar density template value $S(s)$ (figure 8), for each redshift bin. The blue data points show the values measured from data, and the red curve is the obscuration model fit (eq. (3.7)). The vertical gray lines with labels show the survey area fraction with a smaller $S(s)$.

this bias as it does not affect our analysis significantly (see section 5). Lastly, the true mean galaxy density $\bar{n}_{\text{g},i}$ is estimated from the data as:

$$\hat{n}_{\text{g},i} = \frac{\int [n^i_{\text{obs}}(\boldsymbol{\theta}) - \hat{\beta}_i S(\boldsymbol{\theta})] \mathrm{d}^2\theta}{\int W(\boldsymbol{\theta})[1 - \hat{\alpha}_i S(\boldsymbol{\theta})] \mathrm{d}^2\theta}. \tag{3.11}$$

The estimated parameters α_i, β_i and $\bar{n}_{\text{g},i}$, are presented in table 1, where we see that both obscuration and contamination are clearly present in all bins. Once more, the first bin shows higher values of systematics, and the estimated error bars are much larger as well, specially for the obscuration parameter. All these difficulties presented by the first bin indicate that its systematics are not as successfully modeled as for the other bins. On top of that, this redshift range is well covered by data with better constrained photo-zs, larger sky coverage, and apparently less affected by systematics — namely the 2MPZ catalog [34, 63]. Thus, we decided to ignore the first redshift bin in the subsequent analysis.

Bin	Photo-z range	Mean spec-z	α	β	$\bar{n}_{\mathrm{g}}$
1	$0.10 < z < 0.15$	0.1280	0.0902(156)	0.421(39)	3.223(48)
2	$0.15 < z < 0.20$	0.1753	0.0681(75)	0.259(28)	5.499(41)
3	$0.20 < z < 0.25$	0.2248	0.0735(56)	0.136(27)	6.601(33)
4	$0.25 < z < 0.30$	0.2719	0.0818(50)	0.095(25)	5.826(23)
5	$0.30 < z < 0.35$	0.3180	0.0674(80)	0.051(13)	1.748(08)

Table 1. Redshift bin number, photo-z range and estimated values for the mean spectroscopic redshift and obscuration α, contamination β and true mean galaxy density $\bar{n}_{\mathrm{g}}$ parameters. The uncertainties were estimated from mock realizations described in section 5.1.

Using the estimated α_i, β_i and $\bar{n}_{\mathrm{g},i}$, we can solve eq. (3.2) for $n_{\mathrm{g}}^i(\boldsymbol{\theta})$ inside the survey footprint and estimate the true galaxy density contrast $\delta_{\mathrm{g}}^i(\boldsymbol{\theta}) \equiv n_{\mathrm{g}}^i(\boldsymbol{\theta})/\bar{n}_{\mathrm{g},i} - 1$:

$$\hat{\delta}_{\mathrm{g}}^i(\boldsymbol{\theta}) = W(\boldsymbol{\theta}) \left\{ \frac{n_{\mathrm{obs}}^i(\boldsymbol{\theta}) - \hat{\beta}_i S(\boldsymbol{\theta})}{\hat{n}_{\mathrm{g},i}[1 - \hat{\alpha}_i S(\boldsymbol{\theta})]} - 1 \right\}. \tag{3.12}$$

These are shown in figure 10. From eq. (3.12) we see that, even if our template $S(\boldsymbol{\theta})$ and our parameter estimates all perfectly match the data, the first term inside the curly brackets might be negative due to Poisson fluctuations in $n_{\mathrm{obs}}^i(\boldsymbol{\theta})$ [for instance, $n_{\mathrm{obs}}^i(\boldsymbol{\theta})$ might be zero in a given pixel].

4 Methodology for cosmological analysis

4.1 Angular power spectrum estimation

We measured the auto and cross angular power spectra C_ℓ^{ij} of the WSC density contrast maps using the `NaMaster`[5] code [64], which is based on a pseudo-C_ℓ estimator proposed in [65], and includes further improvements [46, 66]. `NaMaster` first computes the so-called pseudo power spectra D_ℓ^{ij}, which measure the variance of the coefficients $a_{\ell m}^i$ used to expand the *masked* density contrast maps $\delta_{\mathrm{g}}^i(\boldsymbol{\theta})$ in spherical harmonics $Y_{\ell m}(\boldsymbol{\theta})$:

$$D_\ell^{ij} \equiv \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} a_{\ell m}^i a_{\ell m}^{j*}. \tag{4.1}$$

The coefficients $a_{\ell m}$ represent the spherical harmonic transform of the product of the mask $W(\boldsymbol{\theta})$ and the estimate of the underlying galaxy number density contrast, according to eq. (3.12). Therefore, when transformed to harmonic space, this product translates into a convolution of the underlying (i.e., full sky) power spectrum $F_{\ell'}^{ij}$ and a mixing matrix $R_{\ell\ell'}$ [66]:

$$D_\ell^{ij} = \sum_{\ell'} R_{\ell\ell'} F_{\ell'}^{ij}. \tag{4.2}$$

The mixing matrix — determined from the power spectrum of the mask (see e.g. equation 6 of [34]) — is shown in figure 11, where we present on the left hand side a 2D color map of the full $R_{\ell\ell'}$, while on the left panel we present particular examples of the $R_{\ell\ell'}$ for three
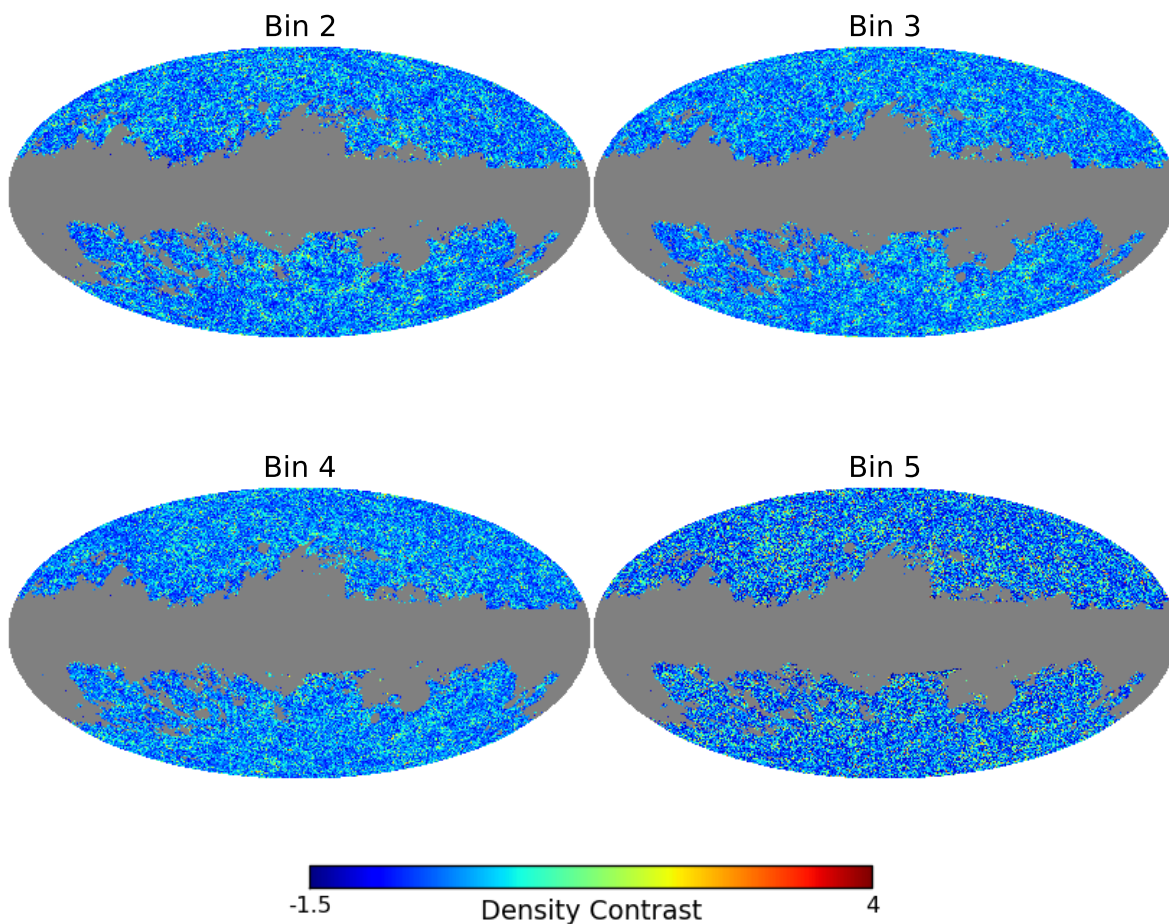
---

[5]http://github.com/LSSTDESC/NaMaster.

Figure 10. WSC density contrast $\delta_{\mathrm{g}}(\boldsymbol{\theta})$ maps for photo-z bins 2–5 ($0.15 < z < 0.35$) after correcting for stellar obscuration and contamination. Due to the contamination template subtraction and Poisson fluctuations in the number counts maps, some pixels end up with $\delta_{\mathrm{g}}(\boldsymbol{\theta}) < -1$.
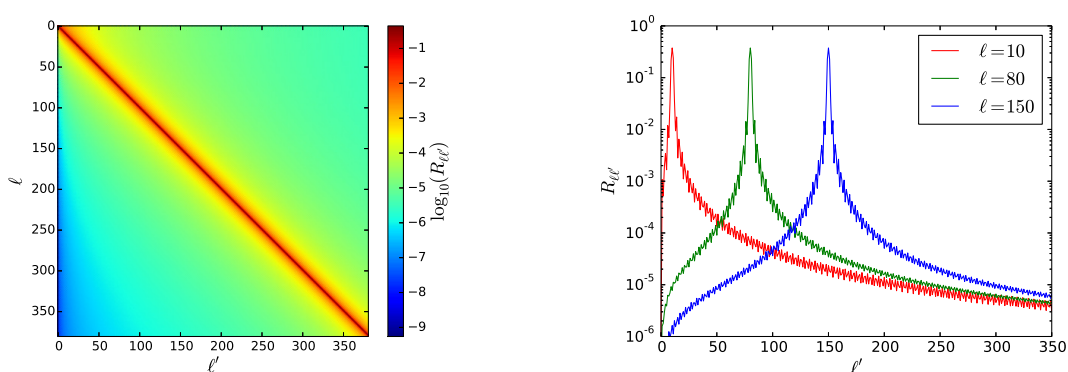

Figure 11. *Left panel:* a color plot of the logarithm of the mixing matrix $R_{\ell\ell'}$ computed for our WSC mask, up to $\ell = 380$. *Right panel:* plot of $R_{\ell\ell'}$ for ℓ fixed at 10 (red), 80 (green) and 150 (blue).

values of ℓ'. Solving for $F_{\ell'}^{ij}$ by inverting the mixing matrix, our final estimate of the true, full-sky angular power spectra C_ℓ^{ij} is given by:

$$C_\ell^{ij} = \frac{F_\ell^{ij} - N^{ij}}{w_\ell^2}, \tag{4.3}$$

where the denominator removes the effect of the pixelization (w_ℓ is known as the pixel window function, computed with the Healpix package for our mask resolution) and N^{ij} is the shot-noise contribution, given by:

$$N^{ij} = \frac{\delta_{\mathrm{K}}^{ij}}{\bar{n}_{\mathrm{g},i}^2} \int \frac{n_{\mathrm{obs}}^i(\boldsymbol{\theta})}{[1 - \hat{\alpha}_i S(\boldsymbol{\theta})]^2} \mathrm{d}^2\theta \tag{4.4}$$

with δ_{K}^{ij} the Kronecker delta. Table 2 shows that thanks to the large mean number density of galaxies in most redshift bins, the shot-noise term does not dominate on the scales we are interested in. There are two aspects of eq. (4.4) worth pointing out. First, that these are the full number counts $n_{\mathrm{obs}}^i(\boldsymbol{\theta})$ (including contamination by stars) that contribute to the shot-noise. Second, since we weight the data by a factor $[1 - \hat{\alpha}_i S(\boldsymbol{\theta})]$, the noise is also amplified.

We will present our measurements of the power spectrum in the form of averages in bins of width $\Delta\ell = 6$, and we consider angular scales from $\ell_0 = 3$.

4.2 Estimation of cosmological parameters

In order to set constraints on cosmological parameters from the measured angular power spectra, we used the Markov Chain Monte Carlo (MCMC) code `Montepython`[6] [67, 68]. We have adopted a flat ΛCDM cosmological model with the following free parameters: $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$, where H_0 is the Hubble constant; Ω_{b}, the baryon density parameter; Ω_{c}, the cold dark matter density parameter; $\ln(10^{10}A_{\mathrm{s}})$, where A_{s} is the amplitude of the primordial power spectrum; n_{s}, the spectral index of the primordial power spectrum; the galaxy biases b_i in each of the four redshift bins i; and the widths $\sigma_{\mathrm{z},i}$ of the ith bin's spec-z selection function as nuisance parameters; in total 13 parameters. We also computed the amplitude of matter density fluctuations σ_8 as a derived parameter. As priors we used independent Gaussian distributions for parameters poorly constrained by our data: h, $\ln(10^{10}A_{\mathrm{s}})$ and n_{s}, with means and standard deviations according to Planck [69]. We point out that our narrow prior on $\ln(10^{10}A_{\mathrm{s}})$ is translated into prior constraints on σ_8 as well.

The theoretical C_ℓ^{ij}s used in the MCMC were computed by `CLASS`[7] [70, 71], including contributions from redshift space distortions and non-linear structure growth [72, 73]. Together with the Planck priors, we applied the extra restriction $\Omega_{\mathrm{b}} > 0.0065$, to avoid a crash in `CLASS` caused by reionization happening at too late times. To describe the spec-z distribution of galaxies in each photo-z bin, we used their Gaussian approximation; the impact of assuming a generalized Lorentzian distribution for the spec-zs is discussed in section 7.4.

To reduce the number of possible sources of systematic deviations between measured and true parameters, we restricted our analysis to linear scales. More specifically, for each C_ℓ^{ij} we only kept the multipoles for which the contribution from the non-linear part of the matter power spectrum was smaller than 10%. Table 2 shows this maximum ℓ for each angular power spectrum.

---

[6]http://baudren.github.io/montepython.html.
[7]http://class-code.net.

Bin i	Bin j	$\ell_{\max}$	ℓ_{shot}
2	2	70	362
2	3	80	—
2	4	89	—
2	5	103	—
3	3	93	367
3	4	104	—
3	5	117	—
4	4	117	316
4	5	129	—
5	5	140	180

Table 2. Maximum multipole (i.e. linear scales) $\ell_{\max}$ included in our MCMC estimation of the posterior, and the multipole at which the shot noise starts to dominate, ℓ_{shot}, for each C_ℓ^{ij}.

The binned C_ℓ^{ij}s were combined into a single data vector with 169 entries. We assumed a Gaussian likelihood, with the covariance (see figure 12) extracted from 2500 lognormal mock catalogs of the WSC, described in section 5.1. The C_ℓ^{ij}s are measured from these mocks, following the same procedure as used for the data (described in sections 3.3 and 4.1). The inverse of the covariance matrix estimated from the simulations has been corrected for the finite sample used to estimate it, following [74].

The sampling of the posterior was performed with 10 Markov chains of 100,000 entries, each one starting from a different point in the parameter space. The convergence of the chains was verified using the Gelman-Rubin diagnostic R, and all parameters presented $R < 1.15$.

5   Validation of the analysis

To make sure that all the steps of our analysis work properly, we used the Full-sky Lognormal Astro-fields Simulation Kit (FLASK[8]) [75] to create hundreds of mock WSC catalogs including the obscuration and contamination effects described in section 3. These mocks are described in detail in section 5.1. The results obtained with our methodology are shown in section 5.2.

5.1   Mock WSC maps

To create lognormal realizations that reproduce the basic statistical properties of the data, FLASK may use as input: a set of auto- and cross-C_ℓs for all the redshift slices being simulated; the mean density of sources $\bar{n}_{\mathrm{g},i}$; and the angular completeness $f(\boldsymbol{\theta})$, which includes the mask, obscuration and/or angular variations in density. FLASK output is a set of Healpix maps that include Poisson noise and that reproduce all one- and two-point statistics as required, including cross-correlations between redshift bins. It can also create realizations of stellar densities by assuming $C_\ell^{\mathrm{star}} = 0$, setting $f(\boldsymbol{\theta})$ to our template and Poisson sampling from it.
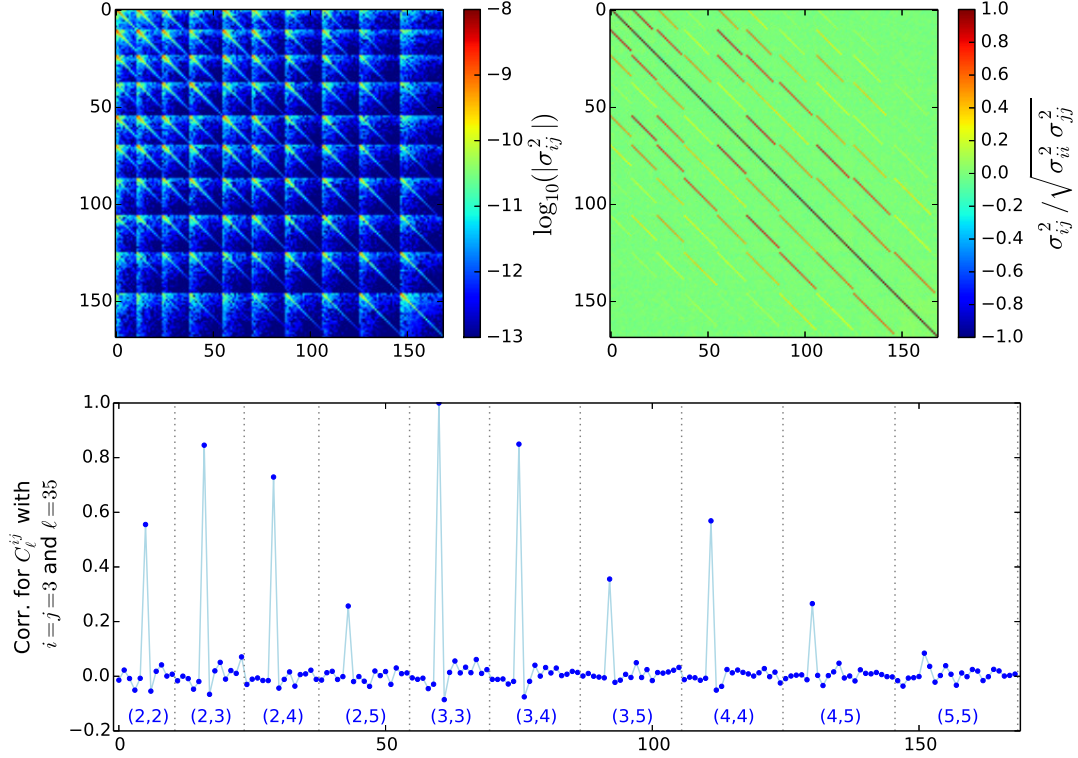
---

[8]http://www.astro.iag.usp.br/~flask.

Figure 12. *Top panels:* logarithm of the absolute value of our data vector's covariance matrix (left) and the data vector's correlation matrix (right). The data vector is composed of 10 binned auto- and cross-C_ℓs limited to linear scales, concatenated in the following order: $C_\ell^{22}, C_\ell^{23}, \ldots, C_\ell^{25}, C_\ell^{33}, C_\ell^{34}, \ldots, C_\ell^{55}$. The left-hand plot was clipped at -13 to improve visualization (its largest values are unclipped). The covariances between C_ℓ^{ij} and C_ℓ^{kq} appear as sub-matrices. *Bottom panel:* 60th row of the correlation matrix, corresponding to C_ℓ^{ij} with $i = j = 3$ and $\ell = 35$. The sectors corresponding to the pair of redshift bins i and j are separated by vertical dotted lines and labeled by (i, j). We note that different (binned) ℓs are practically uncorrelated, and that the power spectra present stronger covariances when they involve nearby bins.

As the expected densities (angular-position dependent), we used $\bar{n}_{\mathrm{g},i}[1 - \alpha_i S(\boldsymbol{\theta})]$ for galaxies and $\beta_i S(\boldsymbol{\theta})$ for stars, where $S(\boldsymbol{\theta})$ is our template (figure 8) and the remaining parameters are set according to table 1. As input galaxy C_ℓs we used smooth fits to the measured cleaned C_ℓs, presented later in figure 16, up to $\ell_{\max} = 640$. The maps of galaxies and stars, for each bin, were summed into a single map of sources, to which we applied our analysis. Figure 13 shows an example of such a realization.

5.2 Recovery of mock properties

In this section we test our methodology to account for the systematic effects and the likelihood machinery on our mock catalogs.
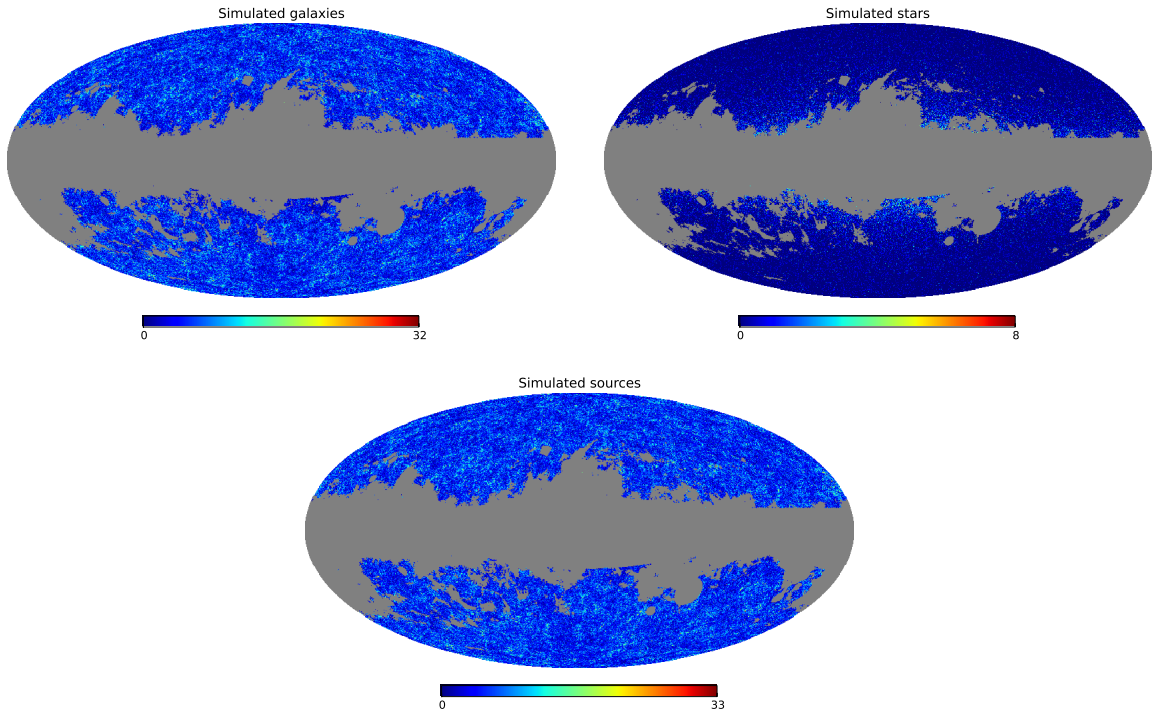
Figure 13. *Top left panel:* a lognormal realization of WSC galaxies following C_ℓ^{ij}, $\bar{n}_{\mathrm{g},2}$ and α_2 [it includes obscuration according to the stellar density template $S(\boldsymbol{\theta})$]. *Top right panel:* a realization of a stellar density map, made by Poisson sampling $S(\boldsymbol{\theta})$ scaled by β_2. *Bottom panel:* the combination of both maps in the top panels, representing all sources detected in the photo-z bin 2 ($0.15 < z < 0.20$).

5.2.1 Obscuration and contamination parameters

The first step in our method is to estimate the obscuration α_i and contamination β_i parameters for all bins i, along with the mean galaxy density $\bar{n}_{\mathrm{g},i}$. Table 3 shows the mean and standard deviation of these parameters, recovered from 2500 WSC mock realizations. It shows that our method traces quite well the true values in each of the bins, being able to detect obscuration and contamination at no less than 5.7σ and 3.9σ, respectively. It does, however, produce a detectable average bias of $\sim 3\%$ on α and $\sim 7\%$ on β (i.e. overcorrecting, on average, both effects). This bias is likely caused by our simplifying assumption that the cosmological variance $\sigma^2_{\mathrm{cosmo}}(s)$ for counts-in-pixels, computed from a pixel set, does not depend on the template's isocontour (eq. (3.6)), which happens to vary both in shape and in area. As density fluctuations are spatially correlated, variances computed in a smaller and compact set will, in general, be smaller. Since, as we show next, this bias does not significantly affect our results, we considered the method's performance adequate despite of its presence.

5.2.2 Angular power spectra

In order to verify the impact of our cleaning method on the C_ℓs, we processed 750 simulations in three different ways:[9] (i) we ignored the obscuration and contamination present in them (equivalent to assuming $\alpha = \beta = 0$ even though they are not); (ii) we only corrected for

---

[9]Due to the large amount of time required to measure 10 C_ℓs in three different ways from each simulation, we did not use all 2500 at this step.

Bin	α_0	$\bar{\alpha}$	σ_α	% bias	β_0	$\bar{\beta}$	σ_β	% bias	n_{g0}	$\bar{n}_g$	σ_{n_g}	% bias
1	0.0902	0.0934	0.0157	3.4	0.421	0.435	0.039	3.2	3.223	3.22	0.049	-0.1
2	0.0681	0.0705	0.0077	3.5	0.259	0.273	0.029	5.3	5.499	5.498	0.042	-0.0
3	0.0735	0.0755	0.0056	2.6	0.136	0.15	0.028	9.0	6.601	6.601	0.033	-0.0
4	0.0818	0.0836	0.005	2.2	0.095	0.106	0.024	10.5	5.826	5.826	0.023	0.0
5	0.0674	0.0703	0.0081	4.0	0.051	0.056	0.013	8.8	1.748	1.748	0.008	0.0

Table 3. Comparison of true obscuration, contamination and mean galaxy density parameters and those recovered from 2500 simulations. The columns are: the redshift bin and the true value x_0, the mean value from the simulations $\bar{x}$, the standard deviation of the simulations σ_x, and the fractional bias ($\bar{x}/x_0 - 1$) in percent for each parameter x.

contamination (equivalent to assuming $\alpha = 0$); and (*iii*) we applied the full cleaning method described in section 3.3. We then measured the angular power spectra using our standard procedure (section 4.1). Figure 14 shows the average of 750 estimated C_ℓs obtained after applying the three different cleaning processes in comparison to the true C_ℓ in the case of the cross-correlation between redshift bins 3 and 4, as an example. The general characteristics of the results are the same for all bin pairs. We see that, in our simulations, the typical effect of contamination is to increase the power on the largest scales ($\ell \lesssim 20$), showing at $\ell \lesssim 10$ deviations with respect to the true spectrum larger than the 1σ error bars. This is clearly a systematic effect we need to correct for in the data. Similarly, we see that correcting only for contamination gives an estimate of the angular power spectrum which is biased low. This bias, also reported in [44], amounts to $\sim 15\%$ and reaches the size of the error bars on small scales. After correcting for obscuration, this bias is eliminated (although there seems to be a small positive and constant bias of $\sim 1\%$, possibly caused by an overestimated α). This shows that when measuring cosmological or astrophysical parameters controlling the amplitude of the angular clustering signal (e.g. galaxy bias and primordial spectrum amplitude), correcting for obscuration is a key step to obtain accurate results. This comparison shows that, assuming our modeling of the systematic is correct, its application significantly improves the accuracy of the estimated power spectrum, leaving only possible residual bias which amounts to no more than 2% of the error bars.

5.2.3   Cosmological parameters

Once we have quantified that our cleaning method properly corrects the measured power spectrum, we use the MCMC methodology to recover the input cosmological and nuisance parameters from cleaned C_ℓs. The input C_ℓs for the FLASK code were computed by CAMB sources[10] [76], including not only contributions from redshift space distortions and non-linearities, but also from a minimal neutrino configuration and all available effects (e.g. gravitational lensing and general relativistic corrections). Furthermore, the computation of the input C_ℓs did not use the Limber approximation and adopted higher precision than the C_ℓs sampled by the MCMC. In this way, we verified that any potential deviations in the recovered parameters are not caused by approximation schemes or by the omission of small physical effects. The input C_ℓs were computed assuming a flat ΛCDM model and Gaussian spec-z selection function, and the simulations included the effects of obscuration and contamination. The procedure followed here mimics that applied to the real data.

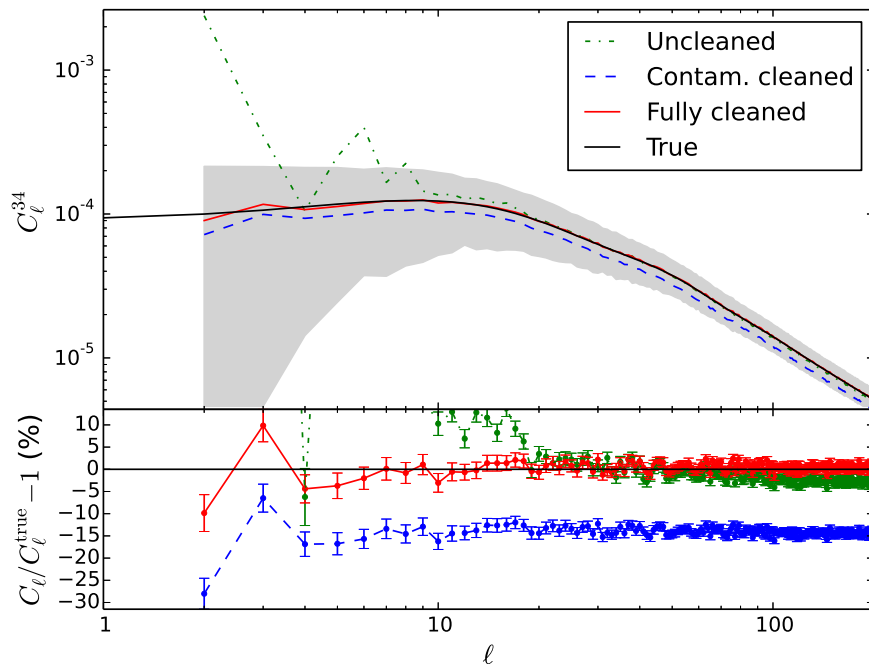
[10]http://camb.info/sources.

Figure 14. *Top panel:* cross-C_ℓ from simulations for the redshift bins 3 and 4. The smooth black curve represents the simulation input C_ℓ, while the remaining lines represent the average over 750 C_ℓs extracted from simulated maps that include obscuration and contamination. The green dot-dashed line shows the results of ignoring both systematic effects; the blue dashed line shows the results when only contamination is corrected for; and the solid red line shows the results when both effects are accounted for. The gray shaded area represents the typical uncertainty over one single C_ℓ measurement, and is centered on the input C_ℓ. *Bottom panel:* the fractional difference between the average of the 750 estimated C_ℓ^{34}s and its true value, for all the cleaning strategies. The error bars show the uncertainty of the mean.

Figure 15 shows that all marginal posterior distributions for all the parameters measured from two independent simulations are compatible with their input values, thus validating our pipeline. We also point out that, while the covariance matrix used in our likelihood was estimated from simulations based on smooth fits to the measured C_ℓ^{ij}s, the simulated data used in this test was generated from a different set of C_ℓ^{ij}s, thus attesting that our covariance matrix does not bias the results. The reader might notice that posterior distribution for e.g. $\sigma_{z,3}$ from one simulation is displaced from its true value. In this case, we point out that we have 13 independent parameters and, in fact, we should expect a few to deviate from their expected values by chance. The second simulation shows that this is indeed the case. Moreover, we see that the ability to pinpoint Ω_b is affected by the specific realization of noise and cosmological fluctuations.

6 Results

In figure 16 we show the binned angular power spectra measured from the WSC data, both cleaned and uncleaned for contamination and obscuration. We see that the effect of clean-
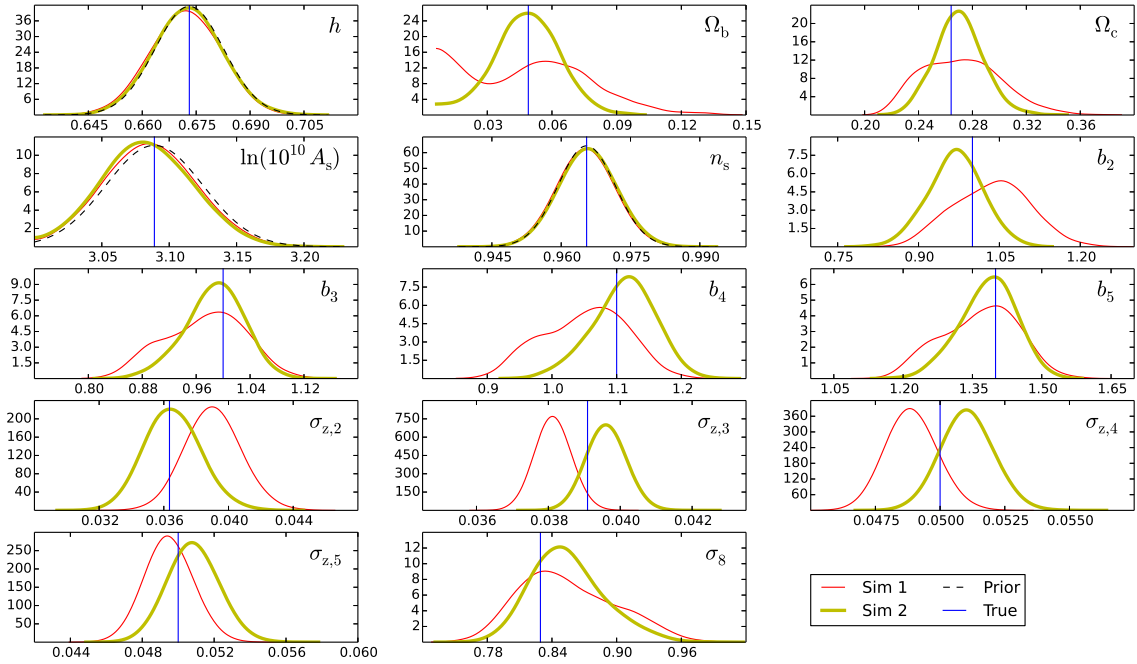
Figure 15. Performance of two example simulations in recovering input parameters. The solid (thin red and thick ochre) curves show the marginal posterior distributions for each cosmological or nuisance parameter given a set of C_ℓ^{ij}s extracted from two independent realizations of source count maps. The dashed black curves show the Gaussian priors used in some parameters [h, $\ln(10^{10}A_{\rm s})$ and $n_{\rm s}$], and the blue vertical lines depict the true values used in the simulations.

ing the WSC is reasonably small on most scales, indicating that galaxy catalogs containing systematics can still provide cosmological information. Despite the smallness of the corrections, we point out two relevant aspects of them: first, applying them significantly reduces the power on the largest scales ($\ell \lesssim 10$) in many cases, and as we showed in section 5.2, this is caused by contamination removal; second, some cleaned C_ℓs show an almost constant-factor increase in power at all scales, caused by obscuration correction, as we have previously discussed (see figure 14).

Figure 16 also shows the best-fit model to the cleaned data (black lines), as well as C_ℓs obtained using cosmological parameters given by Planck [69] and nuisance parameters fitted "by eye" as a rough reference (green lines). By comparing them we see that WSC C_ℓs present steeper slopes than the fiducial cosmology on the linear scales, and that these slope differences do not seem to be restricted to the largest scales only ($\ell \lesssim 20$).

A comparison of the best-fit χ^2 obtained under different data processing strategies is presented in table 4. In every case, the best-fit C_ℓs and χ^2 were obtained for that particular dataset by fitting all 13 parameters to it. We see that correcting the WSC count maps for obscuration and contamination improves the fit, but the model remains not very representative of the data. If we ignore multipoles $\ell < 15$, the fit improves significantly and reaches a p-value level that could be considered adequate (larger than 5%). This may indicate that the largest scales are likely still contaminated even after our corrections, possibly due to a mismatch between the real contamination and our template. If we push the multipole cut to higher values, the agreement does not improve significantly, suggesting that the best-fit is not constrained by the largest scales after this point.
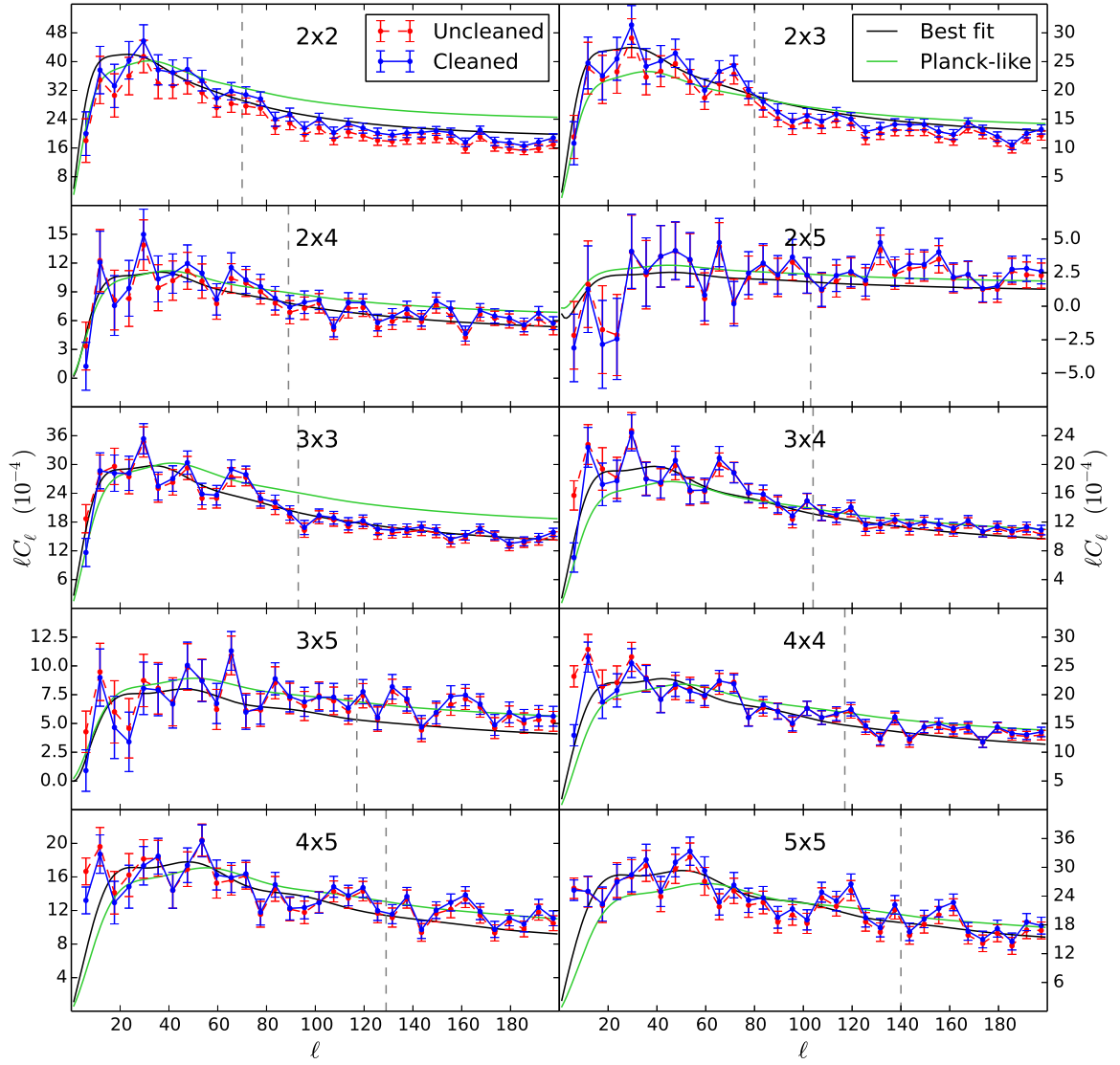
Figure 16. Binned auto and cross angular power spectra of the WSC sources. The two correlated redshift bins are indicated in each plot as $i \times j$. The power spectra were multiplied by the effective ℓ to improve readability. The red data points and dashed lines represent the C_ℓs extracted directly from the data, without correcting for contamination and obscuration. The blue data points and lines represent the C_ℓs taking into account such corrections. The black curves show the best fit model to the cleaned data. The green curves show a ΛCDM model with Planck values [69] and nuisance parameters set by hand, as a guiding reference. The vertical dashed gray lines mark the linear limit up to which the data points were fitted.

Figure 17 shows the 1D and 2D posterior distributions for the cosmological parameters that were not directly restricted by Planck priors, for the first three approaches listed in table 4, and compares them with the values according to Planck. In any case, the posteriors disagree with Planck values by more than 2σ. The most discrepant parameter is Ω_c, which WSC data push to lower values. This is caused by the steep slope of WSC C_ℓs, as shown in figure 16: lower cold matter density leads to less small-scale clustering, tilting the power

– 24 –

Approach	χ^2	d.o.f.	p-value
Uncleaned	268	156	5.87×10^{-8}
Cleaned	237	156	2.92×10^{-5}
$\ell \geq 15$	163	136	0.058
$\ell \geq 21$	151	126	0.063
North uncleaned	398	156	$< 1 \times 10^{-16}$
North cleaned	376	156	$< 1 \times 10^{-16}$
North $\ell \geq 15$	148	136	0.23
South uncleaned	248	156	3.77×10^{-6}
South cleaned	185	156	0.054
South $\ell \geq 15$	157	136	0.10

Table 4. Summary statistics showing the performance of different processing strategies and best-fit models used to fit the angular power spectra of the WSC data. The first column names the strategy adopted. The second displays the best-fit χ^2, computed with respect to the C_ℓs that best-fit that particular processing strategy. The third column contains the number of degrees of freedom, while the last column displays the p-value. All the approaches shown in the table that are restricted to $\ell \geq 15$ (or $\ell \geq 21$) were also subjected to the cleaning process.

spectrum and making it redder. Figure 17 also shows that the power reduction at the largest scales produced by our cleaning method significantly shifts the posterior closer to the Planck values without increasing its size. This demonstrates, using real data, the importance and advantages of our cleaning method. The removal of the largest scales ($\ell < 15$) moves the posterior even closer to the Planck values but increases its width. This suggests that the stellar density template may be improved, or that the largest scales are affected by extra effects besides stellar obscuration and contamination. In either case, it is clear that identifying and taking these effects into account might be a better approach than ignoring the largest scales altogether, as the latter strategy would lead to worse precision.

7 Robustness tests and search for further systematics

To search for other potential sources of systematic effects in the WSC data, we applied a series of modifications to our analysis and checked their impact on the measured C_ℓs, using the full-sky cleaned C_ℓs with $\ell \geq 15$ as our starting point. One of these tests, already mentioned in section 6, consisted in ignoring scales $\ell < 21$, with the only impact on the results being that of making the posterior distributions wider.

7.1 Different stellar maps and masks

When considering possible residual systematics related to stellar density, we first raised two hypotheses, viz. (*i*) that problematic regions near the Galactic bulge could be biasing our results, and (*ii*) that the Gaussian smoothing applied to the Gaia map could have erased Galactic structures at intermediate scales ($\ell \gtrsim 30$) that were not removed by our fiducial template. In order to verify these claims, we created the template $S(\boldsymbol{\theta})$ shown in the left panel of figure 18, consisting in an unsmoothed map with regions of stellar density three times the mean masked, and applied it in the removal of obscuration and contamination effects.
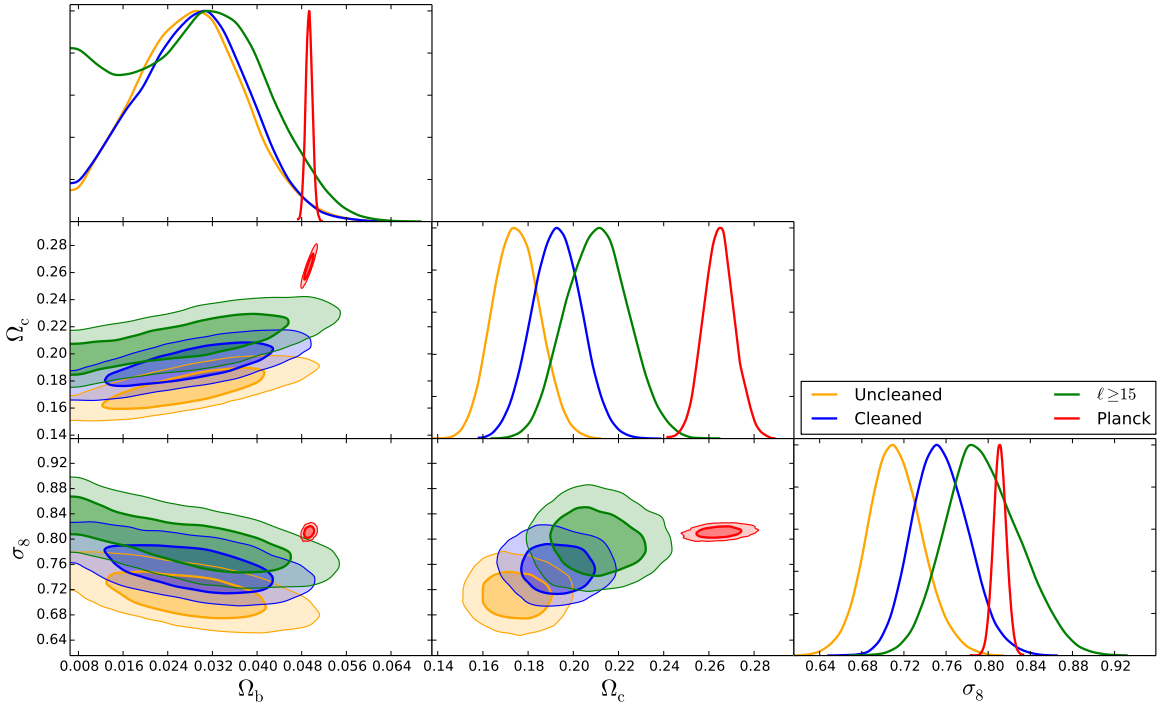
Figure 17. Matrix of posterior plots for the cosmological parameters $\Omega_{\rm b}$, $\Omega_{\rm c}$ and σ_8. The diagonal shows the marginal 1D posteriors for each parameter (normalized to their maximum), while the off-diagonal plots show the 2D posteriors, with contours at 68% and 95% confidence levels. The orange and blue posteriors represent the results for the $3 \leq \ell \leq \ell_{\max}$ range for the uncleaned and cleaned data, respectively, while the green posteriors show the results for the cleaned data for the $15 \leq \ell \leq \ell_{\max}$ range ($\ell_{\max}$ is the linear limit given by table 2). The plots also show in red the posteriors from Planck baseline.

Another possibility we considered was that the map obtained from Gaia was not representing the contamination in the WSC. Hence, and based on the cross-match with SDSS presented in section 3.1, we created a template with the simple exponential model given by eq. (3.1) (central panel in figure 18). Finally, a third template was created, in which the model fitted to the matched SDSS stars had an exponential disk (with an extra longitudinal dependence) and a Gaussian bulge (figure 18, right panel). Note that despite the fact that this approach is based on a direct estimate of the contamination, it relies on an extrapolation from the SDSS footprint to the whole sky.

In all cases, the cleaning process reduced the power on the largest scales and only significantly affected multipoles $\ell \lesssim 20$. The two templates based on Gaia led to very similar outcomes and, for most C_ℓ^{ij}s, improved the results (in terms of large scale power reduction) more than those based on the SDSS cross-match.

7.2 Uncleaned data with $\ell \geq 15$

Although we have shown in sections 5.2.2 and 5.2.3 that our cleaning method does not bias the results (and, in particular, does not tilt the C_ℓ^{ij}s), this demonstration assumed that our stellar density template had the correct shape. To test whether our cleaning method could be biasing the results due to a mismatch between the template and the true stellar density

Figure 18. Alternative stellar density templates used to clean the WSC counts maps during the robustness tests. From left to right: a Gaia map without smoothing and with a larger mask; a latitudinal exponential model; and a "disk & bulge" model fitted to the cross-match between WSC sources and SDSS stars. The color scales are independent for each map and cover the full template's range.

distribution, we estimated the posterior distribution of the cosmological parameters from the uncleaned data, while ignoring the largest scales ($\ell < 15$).

Once more, the posteriors remained highly compatible with those for the cleaned $\ell \geq 15$ case shown in figure 17. This testifies that our cleaning method only affects the large-scale modes $\ell < 15$ and is not responsible for the disagreement between WSC and Planck.

7.3 Mixing matrix deconvolution and H_0 prior

As stated in section 4.1, we estimated the angular power spectra of the unmasked sky by inverting eq. (4.2), that is, we deconvolved the mixing matrix from the data by multiplying the observed pseudo-C_ℓs D_ℓ^{ij} by the inverse $R_{\ell'\ell}^{-1}$ of the mixing matrix. As this process combines data from different scales, it could transfer to smaller scales the power coming from systematic effects on the largest scales. An alternative analysis that should not suffer from this effect would be to compare the data pseudo-C_ℓs (after removing shot-noise and the pixel window function) with theoretical C_ℓs convolved with the mixing matrix $R_{\ell'\ell}$. If no anomalous power transfer happens during the mixing matrix deconvolution in the first approach, both methods should yield the same results. We ran the simple test of comparing the ratio between deconvolved data C_ℓs and fiducial C_ℓs with the ratio between data pseudo-C_ℓs and the same (but convolved) fiducial C_ℓs. Both ratios agreed to very good extent, demonstrating the robustness of our analysis against forward/backward modeling schemes.

Another analysis choice impact we decided to test was that of the H_0 prior. It is known that the Hubble constant inferred from Planck ($H_0 = 67.31 \pm 0.96 \, \mathrm{km\,s^{-1}\,Mpc^{-1}}$) is in tension with direct measurements using type Ia supernovae ($H_0 = 73.24 \pm 1.74 \, \mathrm{km\,s^{-1}\,Mpc^{-1}}$) [77]. We verified if the Planck H_0 prior could be responsible for the observed tension between Ω_c estimated from WSC and Planck data by replacing it with direct measurements prior. Although this change displaced the WSC posteriors, it actually increased the tension on Ω_c slightly.

7.4 Redshift distributions

As mentioned in section 2.1, the WSC photo-z errors can also be described by a generalized Lorentzian distribution, which results (from eq. (2.2)) in a spec-z distribution in each photo-z bin that can also be well modeled by the same distribution. To test whether the Gaussianity assumption used in our analysis could bias the results, we employed generalized Lorentzian functions as spec-z distributions, and used it to generate a model of C_ℓs and sample the posterior distribution of the constrained parameters.

Generalized Lorentzian functions have an extra parameter a that controls the importance of its tails (see eq. (2.1)). We ran the MCMC under two configurations: (i) we let

a_i for each i redshift bin vary freely between 1.5 and 4.4; and (ii) we fixed a_i to the values estimated in [38]. In both cases, the posterior on the cosmological parameters remained quite similar to the one in figure 17.

We also investigated the scenario in which ignoring the photo-z correction due to the north-south hemisphere asymmetry (used in our main analysis; see [33] for details) could bias our results: using the photo-z computed directly by ANNz, we performed again the redshift binning, the cleaning and the measurement of the C_ℓ. As a result, no significant change was observed.

7.5 Ignoring cross-C_ℓs

Our main analysis extracted cosmological information not only from the power spectra in each redshift bin but also from the cross-correlation between different bins. These cross-C_ℓs are more sensitive to the spec-z distribution of the galaxies in each bin than the auto-C_ℓs, since the overlap in the true redshift distributions is one of the main sources of signal in their case. To test if the cross-C_ℓs could be biasing our results, we repeated the procedure to estimate cosmological parameters using only the auto-C_ℓs. Although we noticed no significant displacement in the mean values, the width of the 1D posterior distributions are larger, showing that the cross-C_ℓs do contain extra information.

7.6 Different color and magnitude cuts

The WSC catalog originally had a variable color cut dependent on the distance from the Galactic center, which we removed by applying a more stringent constant cut all over the sky (see section 2.1). We computed the WSC C_ℓs using this original color cut and verified that it results in more power on the largest scales than when using our cut $W1 - W2 > 0.2$, without much change on smaller scales. This effect is expected since a position-dependent color cut creates a variable galaxy selection function, likely making their expected number and bias also position-dependent. In other words, the homogeneity in the number counts discussed by [33] is achieved by the original WSC color cut not by removing stars at low latitudes but by removing both stars and galaxies.

Another potential source of systematics is angular variations of the survey depth, caused by changes in observing conditions (in SuperCOSMOS) or total exposure times (number of observations in WISE). To test for the presence of such features, we applied more conservative magnitude cuts, going from the WSC fiducial limits $W1 < 17$, $B < 21$ and $R < 19.5$ to $W1 < 16.5$, $B < 20.7$ and $R < 19.1$. Apart from a constant factor increase in power in all C_ℓs, expected as a consequence of the larger bias usually associated with brighter galaxies, no significant change was seen in the angular power spectra.

7.7 Analysis by hemispheres

In the full-sky analysis above we have found that, although our corrections for stellar systematics do shift the WSC posterior in the direction of Planck (with respect to the no-correction case), there is still considerable discrepancy between WSC and Planck parameters. As there is no significant evidence from other LSS studies at similar redshifts that this should be physical [34, 35, 78, 79], here we analyze a possible observational effect related to the WSC catalog. Namely, the underlying SuperCOSMOS data are based on measurements from two telescopes which had different effective passbands [53, 80]. The division between the two 'hemispheres' is at $\delta_{1950} = 2.5°$. As noted in [33], despite color-dependent corrections applied in SuperCOSMOS to match north and south [80], within the WSC flux limits (section 2.1)
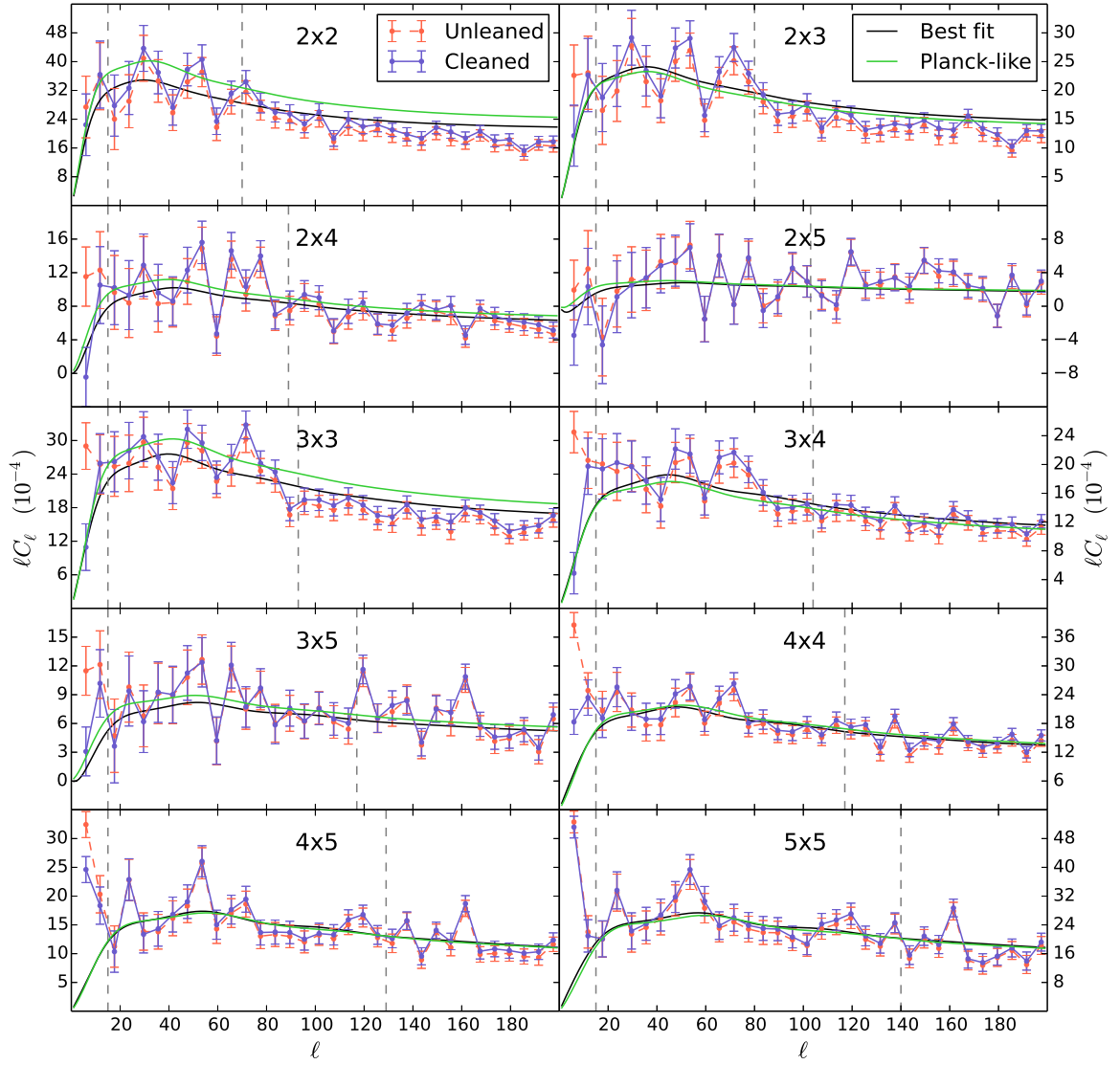
Figure 19. Similar as figure 16, but for data restricted to the northern Galactic hemisphere. The cleaning process reduces power at $\ell < 9$ and re-scales the C_ℓs. The best-fit C_ℓs to the northern data present slopes that are compatible with Planck.

there is still residual difference in mean galaxy surface density, the north one being up to 4% larger than in the south. As this might be a sign of a genuine systematic in the data, we decided to perform the same analysis as for the full sky also for the two hemispheres. Similar tests were done for 2MPZ [34], which also incorporates SuperCOSMOS as one of the input samples, but no significant differences between resulting cosmological parameters were found. We note that the second input sample included in WSC — WISE — does present non-uniformities in the data but not of hemispherical nature, but rather related to the satellite's polar orbit and Moon avoidance maneuvers.[11] Since these non-uniformities are much more involved, we did not investigate their possible impact on our results.

---

[11]See http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec3_4.html.

Figure 20. Similar as figure 17, but all WSC posteriors employ scales $\ell \geq 15$ and use cleaned data. The cyan and pink posteriors were obtained from the northern and southern hemispheres, respectively, while the green posteriors show the results for the full sky.

Both hemispheres present similar behavior as the full-sky analysis with respect to the impact of our cleaning process and removal of multipoles $\ell < 15$. When including all scales with $\ell \geq 3$ and ignoring contamination, both hemispheres display a bad χ^2 and a clear tension with the results from Planck. The cleaning process improves the fit and reduces the tension with Planck by lowering the power on the largest scales ($\ell < 9$). Figure 19 shows the northern C_ℓs before and after cleaning. The southern hemisphere — which already presented less excess of power on these scales — is better improved than the northern hemisphere and actually reaches an acceptable χ^2, although both remain inconsistent with Planck. When we remove the scales with $\ell < 15$, the χ^2 becomes acceptable for both hemispheres, but only the northern hemisphere agrees with Planck. This may indicate that different systematics dominate each hemisphere. The χ^2 values are shown in table 4 and the posterior distributions for each hemisphere under the $\ell \geq 15$ cut are shown and compared with the full-sky posteriors in figure 20. The values of all 13 parameters associated with these posteriors are presented in table 5.

8 Conclusions and summary

In this work we have tested a method for estimating and correcting for stellar obscuration and contamination in galaxy redshift catalogs, aiming at extracting cosmological information therefrom. In particular, we have implemented it on the WISE×SuperCOSMOS (WSC) photo-z dataset [33]. We first exposed in section 3 the presence of both systematic effects in

	Full-sky $\ell \geq 15$		North $\ell \geq 15$		South $\ell \geq 15$	
Parameter	Best-fit	Mean $\pm\sigma$	Best-fit	Mean $\pm\sigma$	Best-fit	Mean $\pm\sigma$
h	0.6703	0.6729(99)	0.6758	0.6729(99)	0.6758	0.6723(97)
$\Omega_{\rm b}$	0.0321	< 0.0477	0.0340	< 0.0648	0.0076	< 0.0397
$\Omega_{\rm c}$	0.214	0.211(14)	0.251	0.254(20)	0.178	0.184(14)
$\ln(10^{10}A_{\rm s})$	3.103	3.088(36)	3.092	3.086(36)	3.078	3.088(36)
$n_{\rm s}$	0.9678	0.9652(64)	0.9687	0.9657(63)	0.9638	0.9653(65)
b_2	1.432	1.424(75)	1.274	1.322(80)	1.394	1.455(82)
b_3	1.284	1.274(63)	1.195	1.229(68)	1.251	1.314(67)
b_4	1.149	1.144(55)	1.107	1.134(61)	1.104	1.152(58)
b_5	1.264	1.266(64)	1.254	1.274(75)	1.210	1.238(70)
$\sigma_{\rm z,2}$	0.0575	0.0581(22)	0.0587	0.0598(36)	0.05591	0.0557(29)
$\sigma_{\rm z,3}$	0.04970	0.04969(90)	0.0495	0.0498(15)	0.0496	0.0500(12)
$\sigma_{\rm z,4}$	0.04401	0.04400(73)	0.0449	0.0453(11)	0.0436	0.0439(10)
$\sigma_{\rm z,5}$	0.0353	0.0354(13)	0.0384	0.0379(18)	0.0344	0.0333(17)
σ_8	0.7915	0.796(35)	0.869	0.854(43)	0.792	0.763(38)

Table 5. Best-fit values of the parameters of a ΛCDM model and the mean and standard deviation of their marginal posteriors for the full-sky, northern and southern datasets. All datasets shown here were cleaned and fit from $\ell \geq 15$ to linear scales. Since the posterior of $\Omega_{\rm b}$ reaches the CLASS limit $\Omega_{\rm b} = 0.0065$, we quote its 95% confidence upper limit instead of the mean.

WSC, a feature that will be common — albeit with a smaller amplitude — in every photometric survey, including LSST, J-PAS and Euclid. In that section we also proposed a model for this feature that ties both systematic effects together with a single stellar density template (eq. (3.2)), and showed that such an assumption is supported by the data (figures 7 and 9).

The fact that both effects — obscuration and contamination — are linked by the template demands a strategy other than simple correlation [46, 62] to quantify them. We proposed the usage of variance measurements in bands of constant stellar density as a way of determining obscuration (eq. (3.7)), which later can be used to estimate the contamination and real galaxy density (eqs. (3.10) and (3.11)). We verified that this strategy is able to detect and quantify all these parameters with less than 10% bias (table 3). Although figures 14 and 15 show that the effect of any residual bias is negligible in the rest of our analysis, we propose that this cleaning method can be improved if we take into account that galaxy densities are spatially correlated.

It is worth pointing out that one may disentangle the effects of obscuration and contamination directly on the power spectra — since they result in different signatures (see section 5.2.2) — by treating them as nuisance parameters and fitting them together with the cosmology. In this process, it is relevant to not only analyse the data C_ℓs but also the cross power spectra between data and stellar density template. The contamination level may be better constrained by it, whereas the measurement of obscuration might benefit from a cross-C_ℓ between the template and the observed density difference squared, $\delta n_{\rm obs}^2$. Testing this approach should be part of a future work.

All of our methodology was validated by applying it to simulations created by the FLASK code [75] and comparing the measured power spectrum and the constrained parameters with their true input values (see section 5). That analysis demonstrated that applying our cleaning method to data containing obscuration and contamination significantly improved the cosmological parameter inference. This method was also applied to real WSC data, and table 4 and figure 17 show that it improved the power spectra fit at $\ell < 15$ and increased the agreement between the cosmology derived from WSC power spectra and from the Planck satellite.

Despite the improvement obtained both in the full-sky and in separate hemispheres by removing contamination and obscuration, the constraints derived from the angular clustering of WSC still disagree with those derived from the Planck mission if we include the largest scales. In fact, the minimum χ^2 shows that the standard model is a bad fit to the data in the full-sky and northern hemisphere case (see table 4) even for parameter values significantly different from Planck. The fit is significantly improved if we ignore multipoles $\ell < 15$ (reaching a p-value of $\sim 6\%$ for the full-sky case, which could be considered adequate), proving that much of the deviation between data and model happens at the largest scales. The reasons for this excess of power is still unclear and should be further investigated. Possible explanations are that (i) stellar density templates, used to remove contamination, must be improved to better represent WSC true distribution; and that (ii) important systematic effects not related to stars — possibly Galactic extinction and photometric/observational errors — need to be taken into account. We do point out, however, that we have tested three other templates (figure reffig:alt-templates) and they have led to similar (or more biased) results.

As the scales $\ell \sim 15$ are close to the power spectrum turn-over at redshifts probed by WSC, we could claim that the excess of power at scales larger than this comes from physical effects not considered by our cosmological model, e.g.: by non-Gaussian initial conditions caused by certain inflationary models [81]; by the fact that galaxy forms at matter density peaks, which could lead to a scale-dependent bias [30]; or by alternative dark energy models that produce different redshift space distortions [82]. These effects, however, are expected to be small and probably should wait for the investigation of other possible systematics sources. It is worth pointing out that an excess of power is constantly seen on the largest scales of galaxy surveys [34, 35, 47], and due to its correlation with stellar density and other reasons, these modes are mostly discarded. This is unfortunate as the physical effects mentioned above remain untested.

We have tested our data for several potential sources of systematics using the full-sky data limited to $\ell \geq 15$ as a reference. In section 7 we verified that our results are not affected by photo-z error distributions, cleaning strategy, color and magnitude cuts, photo-z estimations, mixing matrix deconvolution, the use of cross-C_ℓs and the restriction to $\ell \geq 21$. We did find, however, a hemispherical pattern in the data. Figure 20 shows that the independent analyses of the two equatorial hemispheres lead to incompatible cosmological parameters: while the analysis restricted to $\ell \geq 15$ and to the northern hemisphere agrees with Planck (with $\Omega_c = 0.254^{+0.018}_{-0.022}$, $\sigma_8 = 0.854^{+0.042}_{-0.045}$ and $\Omega_b < 0.065$ at 95% confidence limit when combined with Planck priors on H_0, A_s and n_s), the same analysis in the southern hemisphere does not. This difference in extracted cosmological parameters is traced back to different slopes in the measured power spectra, with the southern and full-sky C_ℓs being steeper than the northern one. Considering it is unlikely that such a large difference has a cosmological origin — especially since cosmological parameters have been extracted from the southern hemisphere before and shown to agree with Planck [34, 79] — this indicates that systematics affecting scales $\ell \gtrsim 15$ are located in the southern part of WSC data.

Investigating what is the actual source of these systematics is beyond the scope of the present work. One possible explanation is residual miscalibration of SuperCOSMOS data in the south. Namely, the input photographic material for that sample was based on observations from two telescopes (POSS-II in the north and UKST in the south) with effectively different passbands. Despite the efforts to match the passbands using specific color terms, anchored to SDSS as the calibrator [80], differences may still exist. As the WSC sample is selected using maximum apparent magnitude cuts, in particular in the SuperCOSMOS B and R bands, this may propagate into inconsistent sample selections in the two hemispheres. Indeed, already in [33], a difference in effective source densities between north and south was reported.

However, the most plausible explanation for the north-south difference in large-scale power comes from residual uncertainties in photometric zero points. Where direct SDSS calibration was lacking, plate zero points were given an initial estimate from mean optical-2MASS color, and these values were then perturbed in an attempt to optimize agreement in plate overlap zones. It seems likely that this process will yield zero points that are imperfect at the level of a few hundredths of a magnitude [83], imprinting a pattern of $5 \times 5 \deg^2$ tiles on the galaxy distribution with an rms variation in surface density of several percent. This is the right order of magnitude to cause the offsets that we measure (J. Peacock, private communication). The second input dataset of WSC — WISE — is unlikely to be at the origin of the hemispherical differences as any non-uniformity in that survey is related mostly to telescope's polar orbit and Moon avoidance strategy. Thus, our present findings seem to suggest that the observed asymmetry is due to residual systematics in the UKST data.

Acknowledgments

References

[1] A.J.S. Hamilton, M. Tegmark and N. Padmanabhan, *Linear redshift distortions and power in the PSCZ survey*, *Mon. Not. Roy. Astron. Soc.* **317** (2000) L23 [astro-ph/0004334] [inSPIRE].

[2] 2DFGRS collaboration, *The 2dF galaxy redshift survey: the power spectrum and the matter content of the universe*, *Mon. Not. Roy. Astron. Soc.* **327** (2001) 1297 [astro-ph/0105252] [inSPIRE].

[3] M. Tegmark, A.J.S. Hamilton and Y.-Z. Xu, *The power spectrum of galaxies in the 2dF 100k redshift survey*, *Mon. Not. Roy. Astron. Soc.* **335** (2002) 887 [astro-ph/0111575] [inSPIRE].

[4] 2DFGRS collaboration, *The 2dF galaxy redshift survey: power-spectrum analysis of the final dataset and cosmological implications*, *Mon. Not. Roy. Astron. Soc.* **362** (2005) 505 [astro-ph/0501174] [inSPIRE].

[5] SDSS collaboration, *Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies*, *Astrophys. J.* **633** (2005) 560 [astro-ph/0501171] [inSPIRE].

[6] W.J. Percival et al., *The shape of the SDSS DR5 galaxy power spectrum*, *Astrophys. J.* **657** (2007) 645 [astro-ph/0608636] [inSPIRE].

[7] A.G. Sánchez et al., *The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: cosmological constraints from the full shape of the clustering wedges*, *Mon. Not. Roy. Astron. Soc.* **433** (2013) 1202 [arXiv:1303.4396] [inSPIRE].

[8] M. Yoon, D. Huterer, C. Gibelyou, A. Kovács and I. Szapudi, *Dipolar modulation in number counts of WISE-2MASS sources*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) L60 [arXiv:1406.1187] [inSPIRE].

[9] D. Alonso, A.I. Salvador, F.J. Sánchez, M. Bilicki, J. García-Bellido and E. Sánchez, *Homogeneity and isotropy in the two micron all sky survey photometric redshift catalogue*, *Mon. Not. Roy. Astron. Soc.* **449** (2015) 670 [arXiv:1412.5151] [inSPIRE].

[10] P. Laurent et al., *A $14\,h^{-3}\,Gpc^3$ study of cosmic homogeneity using BOSS DR12 quasar sample*, *JCAP* **11** (2016) 060 [arXiv:1602.09010] [inSPIRE].

[11] C.A.P. Bengaly, A. Bernui, J.S. Alcaniz, H.S. Xavier and C.P. Novaes, *Is there evidence for anomalous dipole anisotropy in the large-scale structure?*, *Mon. Not. Roy. Astron. Soc.* **464** (2017) 768 [arXiv:1606.06751] [INSPIRE].

[12] F. Avila, C.P. Novaes, A. Bernui and E. de Carvalho, *The scale of homogeneity in the local universe with the ALFALFA catalogue*, *JCAP* **12** (2018) 041 [arXiv:1806.04541] [INSPIRE].

[13] DES collaboration, *Testing the lognormality of the galaxy and weak lensing convergence distributions from dark energy survey maps*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 1444 [arXiv:1605.02036] [INSPIRE].

[14] C. Blake et al., *The WiggleZ dark energy survey: the growth rate of cosmic structure since redshift $z = 0.9$*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) 2876 [arXiv:1104.2948] [INSPIRE].

[15] BOSS collaboration, *The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: testing gravity with redshift-space distortions using the power spectrum multipoles*, *Mon. Not. Roy. Astron. Soc.* **443** (2014) 1065 [arXiv:1312.4611] [INSPIRE].

[16] L. Anderson et al., *The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample*, *Mon. Not. Roy. Astron. Soc.* **427** (2013) 3435 [arXiv:1203.6594] [INSPIRE].

[17] E. de Carvalho, A. Bernui, G.C. Carvalho, C.P. Novaes and H.S. Xavier, *Angular baryon acoustic oscillation measure at $z = 2.225$ from the SDSS quasar survey*, *JCAP* **04** (2018) 064 [arXiv:1709.00113] [INSPIRE].

[18] J.A. Peacock et al., *A measurement of the cosmological mass density from clustering in the 2dF galaxy redshift survey*, *Nature* **410** (2001) 169 [astro-ph/0103143] [INSPIRE].

[19] A.J. Ross et al., *The clustering of galaxies in SDSS-III DR9 baryon oscillation spectroscopic survey: constraints on primordial non-Gaussianity*, *Mon. Not. Roy. Astron. Soc.* **428** (2013) 1116 [arXiv:1208.1491] [INSPIRE].

[20] S. Riemer-Sorensen et al., *The WiggleZ dark energy survey: cosmological neutrino mass constraint from blue high-redshift galaxies*, *Phys. Rev.* **D 85** (2012) 081101 [arXiv:1112.4940] [INSPIRE].

[21] A. Loureiro et al., *On the upper bound of neutrino masses from combined cosmological observations and particle physics experiments*, arXiv:1811.02578 [INSPIRE].

[22] M. Takada, E. Komatsu and T. Futamase, *Cosmology with high-redshift galaxy survey: neutrino mass and inflation*, *Phys. Rev.* **D 73** (2006) 083520 [astro-ph/0512374] [INSPIRE].

[23] C. Carbone, L. Verde, Y. Wang and A. Cimatti, *Neutrino constraints from future nearly all-sky spectroscopic galaxy surveys*, *JCAP* **03** (2011) 030 [arXiv:1012.2868] [INSPIRE].

[24] R. de Putter and O. Doré, *Designing an inflation galaxy survey: how to measure $\sigma(f_{NL}) \sim 1$ using scale-dependent galaxy bias*, *Phys. Rev.* **D 95** (2017) 123513 [arXiv:1412.3854] [INSPIRE].

[25] A. Lewis and A. Challinor, *Weak gravitational lensing of the CMB*, *Phys. Rept.* **429** (2006) 1 [astro-ph/0601594] [INSPIRE].

[26] PLANCK collaboration, *Planck 2013 results. XIX. The integrated Sachs-Wolfe effect*, *Astron. Astrophys.* **571** (2014) A19 [arXiv:1303.5079] [INSPIRE].

[27] C. Blake and J. Wall, *Detection of the velocity dipole in the radio galaxies of the NRAO VLA sky survey*, *Nature* **416** (2002) 150 [astro-ph/0203385] [INSPIRE].

[28] P. Erdogdu et al., *The dipole anisotropy of the 2 micron all-sky redshift survey*, *Mon. Not. Roy. Astron. Soc.* **368** (2006) 1515 [astro-ph/0507166] [INSPIRE].

[29] M. Bilicki, M. Chodorowski, G.A. Mamon and T. Jarrett, *Is the 2MASS clustering dipole convergent?*, *Astrophys. J.* **741** (2011) 31 [arXiv:1102.4356] [INSPIRE].

[30] R. Durrer, A. Gabrielli, M. Joyce and F. Sylos Labini, *Bias and the power spectrum beyond the turn-over*, *Astrophys. J.* **585** (2003) L1 [astro-ph/0211653] [INSPIRE].

[31] LSST SCIENCE and LSST PROJECT collaborations, *LSST science book, version* 2.0, arXiv:0912.0201 [INSPIRE].

[32] L. Amendola et al., *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **21** (2018) 2 [arXiv:1606.00180] [INSPIRE].

[33] M. Bilicki et al., *WISE × SuperCOSMOS photometric redshift catalog:* 20 *million galaxies over* 3π *steradians*, arXiv:1607.01182 [INSPIRE].

[34] A. Balaguera-Antolínez, M. Bilicki, E. Branchini and A. Postiglione, *Extracting cosmological information from the angular power spectrum of the* 2*MASS photometric redshift catalogue*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) 1050 [arXiv:1711.04583] [INSPIRE].

[35] A. Loureiro et al., *Cosmological measurements from angular power spectra analysis of BOSS DR*12 *tomography*, *Mon. Not. Roy. Astron. Soc.* **485** (2019) 326 [arXiv:1809.07204] [INSPIRE].

[36] C.A.P. Bengaly, C.P. Novaes, H.S. Xavier, M. Bilicki, A. Bernui and J.S. Alcaniz, *The dipole anisotropy of WISE × SuperCOSMOS number counts*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) L106 [arXiv:1707.08091] [INSPIRE].

[37] C.P. Novaes, A. Bernui, H.S. Xavier and G.A. Marques, *Tomographic local* 2*D analyses of the WISE × SuperCOSMOS all-sky galaxy catalogue*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 3253 [arXiv:1805.04078] [INSPIRE].

[38] J.A. Peacock and M. Bilicki, *Wide-area tomography of CMB lensing and the growth of cosmological density fluctuations*, *Mon. Not. Roy. Astron. Soc.* **481** (2018) 1133 [arXiv:1805.11525] [INSPIRE].

[39] S. Raghunathan, F. Bianchini and C.L. Reichardt, *Imprints of gravitational lensing in the Planck cosmic microwave background data at the location of WISE × SCOS galaxies*, *Phys. Rev.* **D 98** (2018) 043506 [arXiv:1710.09770] [INSPIRE].

[40] B. Stölzner, A. Cuoco, J. Lesgourgues and M. Bilicki, *Updated tomographic analysis of the integrated Sachs-Wolfe effect and implications for dark energy*, *Phys. Rev.* **D 97** (2018) 063506 [arXiv:1710.03238] [INSPIRE].

[41] BOSS collaboration, *The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: analysis of potential systematics*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 564 [arXiv:1203.6499] [INSPIRE].

[42] B. Leistedt, H.V. Peiris, D.J. Mortlock, A. Benoit-Lévy and A. Pontzen, *Estimating the large-scale angular power spectrum in the presence of systematics: a case study of Sloan Digital Sky Survey quasars*, *Mon. Not. Roy. Astron. Soc.* **435** (2013) 1857 [arXiv:1306.0005] [INSPIRE].

[43] B. Leistedt and H.V. Peiris, *Exploiting the full potential of photometric quasar surveys: optimal power spectra through blind mitigation of systematics*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 2 [arXiv:1404.6530] [INSPIRE].

[44] D.L. Shafer and D. Huterer, *Multiplicative errors in the galaxy power spectrum: self-calibration of unknown photometric systematics for precision cosmology*, *Mon. Not. Roy. Astron. Soc.* **447** (2015) 2961 [arXiv:1410.0035] [INSPIRE].

[45] F. Elsner, B. Leistedt and H.V. Peiris, *Unbiased methods for removing systematics from galaxy clustering measurements*, *Mon. Not. Roy. Astron. Soc.* **456** (2016) 2095 [arXiv:1509.08933] [INSPIRE].

[46] F. Elsner, B. Leistedt and H.V. Peiris, *Unbiased pseudo-C_ℓ power spectrum estimation with mode projection*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 1847 [arXiv:1609.03577] [INSPIRE].

[47] S.A. Thomas, F.B. Abdalla and O. Lahav, *Excess clustering on large scales in the MegaZ DR7 photometric redshift survey*, *Phys. Rev. Lett.* **106** (2011) 241301 [arXiv:1012.2272] [INSPIRE].

[48] SDSS collaboration, *The Sloan Digital Sky Survey: technical summary*, *Astron. J.* **120** (2000) 1579 [astro-ph/0006396] [INSPIRE].

[49] GAIA collaboration, *The Gaia mission*, *Astron. Astrophys.* **595** (2016) A1 [arXiv:1609.04153] [INSPIRE].

[50] GAIA collaboration, *Gaia data release 2*, *Astron. Astrophys.* **616** (2018) A1 [arXiv:1804.09365] [INSPIRE].

[51] R.M. Cutri et al., *Explanatory supplement to the AllWISE data release products*, tech. rep., November 2013.

[52] E.L. Wright et al., *The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance*, *Astron. J.* **140** (2010) 1868 [arXiv:1008.0031] [INSPIRE].

[53] N. Hambly et al., *The supercosmos sky survey. Paper I: introduction and description*, *Mon. Not. Roy. Astron. Soc.* **326** (2001) 1279 [astro-ph/0108286] [INSPIRE].

[54] N. Hambly, M. Irwin and H. MacGillivray, *The supercosmos sky survey. Paper II: image detection, parameterisation, classification and photometry*, *Mon. Not. Roy. Astron. Soc.* **326** (2001) 1295 [astro-ph/0108290] [INSPIRE].

[55] K.M. Górski et al., *HEALPix — a framework for high resolution discretization and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [astro-ph/0409513] [INSPIRE].

[56] A.A. Collister and O. Lahav, *ANNz: estimating photometric redshifts using artificial neural networks*, *Publ. Astron. Soc. Pac.* **116** (2004) 345 [astro-ph/0311058] [INSPIRE].

[57] S.P. Driver et al., *GAMA: towards a physical understanding of galaxy formation*, *Astron. Geophys.* **50** (2009) 5.12 [arXiv:0910.5123] [INSPIRE].

[58] SDSS collaboration, *The fourteenth data release of the Sloan Digital Sky Survey: first spectroscopic data from the extended baryon oscillation spectroscopic survey and from the second phase of the apache point observatory galactic evolution experiment*, *Astrophys. J. Suppl.* **235** (2018) 42 [arXiv:1707.09322] [INSPIRE].

[59] R.H. Lupton et al., *SDSS imaging pipelines*, *Proc. SPIE Int. Soc. Opt. Eng.* **4836** (2002) 350 [INSPIRE].

[60] J.R. Pier et al., *Astrometric calibration of the Sloan Digital Sky Survey*, *Astron. J.* **125** (2003) 1559 [astro-ph/0211375] [INSPIRE].

[61] F. Arenou et al., *Gaia data release 2*, *Astron. Astrophys.* **616** (2018) A17 [arXiv:1804.09375].

[62] G.B. Rybicki and W.H. Press, *Interpolation, realization and reconstruction of noisy, irregularly sampled data*, *Astrophys. J.* **398** (1992) 169 [INSPIRE].

[63] M. Bilicki, T.H. Jarrett, J.A. Peacock, M.E. Cluver and L. Steward, *2MASS photometric redshift catalog: a comprehensive three-dimensional census of the whole sky*, arXiv:1311.5246 [INSPIRE].

[64] LSST DARK ENERGY SCIENCE collaboration, *A unified pseudo-C_ℓ framework*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 4127 [arXiv:1809.09603] [INSPIRE].

[65] P.J.E. Peebles, *Statistical analysis of catalogs of extragalactic objects. I. Theory*, *Astrophys. J.* **185** (1973) 413.

[66] E. Hivon, K.M. Górski, C.B. Netterfield, B.P. Crill, S. Prunet and F. Hansen, *Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets*, *Astrophys. J.* **567** (2002) 2 [astro-ph/0105302] [INSPIRE].

[67] B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, *Conservative constraints on early cosmology: an illustration of the Monte Python cosmological parameter inference code*, *JCAP* **02** (2013) 001 [arXiv:1210.7183] [INSPIRE].

[68] T. Brinckmann and J. Lesgourgues, *Monte Python* 3: *boosted MCMC sampler and other features*, arXiv:1804.07261 [INSPIRE].

[69] PLANCK collaboration, *Planck* 2015 *results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [arXiv:1502.01589] [INSPIRE].

[70] D. Blas, J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) II: approximation schemes*, *JCAP* **07** (2011) 034 [arXiv:1104.2933] [INSPIRE].

[71] E. Di Dio, F. Montanari, J. Lesgourgues and R. Durrer, *The CLASSgal code for relativistic cosmological large scale structure*, *JCAP* **11** (2013) 044 [arXiv:1307.1459] [INSPIRE].

[72] VIRGO CONSORTIUM collaboration, *Stable clustering, the halo model and nonlinear cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1311 [astro-ph/0207664] [INSPIRE].

[73] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, *Revising the halofit model for the nonlinear matter power spectrum*, *Astrophys. J.* **761** (2012) 152 [arXiv:1208.2701] [INSPIRE].

[74] J. Hartlap, P. Simon and P. Schneider, *Why your model parameter confidences might be too optimistic: unbiased estimation of the inverse covariance matrix*, *Astron. Astrophys.* (2006) [astro-ph/0608064] [INSPIRE].

[75] H.S. Xavier, F.B. Abdalla and B. Joachimi, *Improving lognormal models for cosmological fields*, *Mon. Not. Roy. Astron. Soc.* **459** (2016) 3693 [arXiv:1602.08503] [INSPIRE].

[76] A. Challinor and A. Lewis, *The linear power spectrum of observed source number counts*, *Phys. Rev.* **D 84** (2011) 043516 [arXiv:1105.5292] [INSPIRE].

[77] A.G. Riess et al., *A* 2.4% *determination of the local value of the Hubble constant*, *Astrophys. J.* **826** (2016) 56 [arXiv:1604.01424] [INSPIRE].

[78] H. Gil-Marín et al., *The power spectrum and bispectrum of SDSS DR11 BOSS galaxies — II. Cosmological interpretation*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 1914 [arXiv:1408.0027] [INSPIRE].

[79] DES collaboration, *Dark energy survey year 1 results: cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev.* **D 98** (2018) 043526 [arXiv:1708.01530] [INSPIRE].

[80] J.A. Peacock et al., *The SuperCOSMOS all-sky galaxy catalogue*, *Mon. Not. Roy. Astron. Soc.* **462** (2016) 2085 [arXiv:1607.01189] [INSPIRE].

[81] N. Dalal, O. Doré, D. Huterer and A. Shirokov, *The imprints of primordial non-Gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev.* **D 77** (2008) 123514 [arXiv:0710.4560] [INSPIRE].

[82] M. Takada, *Can a galaxy redshift survey measure dark energy clustering?*, *Phys. Rev.* **D 74** (2006) 043505 [astro-ph/0606533] [INSPIRE].

[83] M.F. Skrutskie et al., *The two micron all sky survey (*2MASS*)*, *Astron. J.* **131** (2006) 1163 [INSPIRE].