Characterizing unknown systematics in large scale structure surveys

Nishant Agarwal,[a] **Shirley Ho,**[a] **Adam D. Myers,**[b] **Hee-Jong Seo,**[c] **Ashley J. Ross,**[d] **Neta Bahcall,**[e] **Jonathan Brinkmann,**[f] **Daniel J. Eisenstein,**[g] **Demitri Muna,**[h] **Nathalie Palanque-Delabrouille,**[i] **Isabelle Pâris,**[j] **Patrick Petitjean,**[k] **Donald P. Schneider,**[l,m] **Alina Streblyanska,**[n,o] **Benjamin A. Weaver,**[p] **and Christophe Yèche**[i]

[a]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University,
Pittsburgh, PA 15213, USA

[b]Department of Physics and Astronomy, University of Wyoming,
Laramie, WY 82071, USA

[c]Berkeley Center for Cosmological Physics, LBL and Department of Physics, University of California,
Berkeley, CA 94720, USA

[d]Institute of Cosmology and Gravitation, University of Portsmouth,
Portsmouth, PO1 3FX, UK

[e]Princeton University Observatory,
Peyton Hall, Princeton, NJ 08544, USA

[f]Apache Point Observatory,
P.O. Box 59, Sunspot, NM 88349, USA

[g]Harvard-Smithsonian Center for Astrophysics,
60 Garden St., Cambridge, MA 02138, USA

[h]Department of Astronomy, Ohio State University,
Columbus, OH 43210, USA

[i]CEA, Centre de Saclay, Irfu/SPP,
F-91191 Gif-sur-Yvette, France

[j]Departamento de Astronomía, Universidad de Chile,
Casilla 36-D, Santiago, Chile

[k]Université Paris 6 et CNRS, Institut d'Astrophysique de Paris,
98bis blvd. Arago, 75014 Paris, France

[l]Department of Astronomy and Astrophysics, Pennsylvania State University,
University Park, PA 16802, USA

[m]Institute for Gravitation and the Cosmos, Pennsylvania State University,
University Park, PA 16802, USA

[n]Instituto de Astrofisica de Canarias (IAC),
E-38200 La Laguna, Tenerife, Spain

^oDept. Astrofisica, Universidad de La Laguna (ULL),
E-38206 La Laguna, Tenerife, Spain
^pCenter for Cosmology and Particle Physics, New York University,
New York, NY 10003, USA

E-mail: nishanta@andrew.cmu.edu

Abstract. Photometric large scale structure (LSS) surveys probe the largest volumes in the Universe, but are inevitably limited by systematic uncertainties. Imperfect photometric calibration leads to biases in our measurements of the density fields of LSS tracers such as galaxies and quasars, and as a result in cosmological parameter estimation. Earlier studies have proposed using cross-correlations between different redshift slices or cross-correlations between different surveys to reduce the effects of such systematics. In this paper we develop a method to characterize unknown systematics. We demonstrate that while we do not have sufficient information to correct for unknown systematics in the data, we can obtain an estimate of their magnitude. We define a parameter to estimate contamination from unknown systematics using cross-correlations between different redshift slices and propose discarding bins in the angular power spectrum that lie outside a certain contamination tolerance level. We show that this method improves estimates of the bias using simulated data and further apply it to photometric luminous red galaxies in the Sloan Digital Sky Survey as a case study.

Contents

1 Introduction

Large scale structure (LSS) offers a complementary probe to the cosmic microwave background (CMB) to do precision cosmology. Owing to the large volume with full three-dimensional (3D) information that is available in LSS surveys, they offer the promise to provide competitive, or even better, constraints on cosmological parameters compared to CMB-based experiments, in the near future. For example, it is possible that the next level of improved constraints on primordial non-Gaussianity would come from LSS [1–8].

Ideally, in order to maximize volume, one would like a full-sky spectroscopic LSS survey to very high redshift. The Baryon Oscillation Spectroscopic Survey (BOSS) [9] currently provides the largest volume of spectroscopic measurements [10], and next generation experiments such as the Dark Energy Spectroscopic Instrument (DESI) [11] and Euclid [12] will increase the observed volume by an order of magnitude. However, redshift determination in photometric surveys is simpler and they invariably probe larger volumes. A large amount of Sloan Digital Sky Survey (SDSS) [13, 14] data is based on photometry [15, 16]; with future surveys such as the Dark Energy Survey (DES) [17] and the Large Synoptic Survey Telescope (LSST) [18] the amount of photometric data will virtually explode.

A major challenge in LSS photometric surveys, however, is understanding and accounting for various systematics in the data [15, 19–24]. Without properly correcting for sample systematics in the observed number density of objects, we cannot obtain an accurate measurement of the angular power spectrum of the LSS tracer, and in turn cannot extract cosmological information from it. Potential sources of contamination in the data include stellar obscuration, sky brightness, seeing variations, dust extinction, color offsets, and magnitude errors. While it has been possible to construct maps for these systematic fields and

account for them in the data, what we will focus on in this paper is the effects of *unknown systematics*.[1]

We propose a method to use cross-correlations between different redshift slices to estimate contamination from unknown systematics in the auto-power spectra. Cosmologically, the cross-power between two LSS samples that have no overlap in redshift should be negligibly small.[2] In practice, however, a real-observation detects a non-zero signal, either due to an overlap in redshift or due to common systematics [15, 21]. We can, in fact, even estimate the cross-power theoretically using the overlap of the redshift distribution between different redshift slices. Subtracting the known systematic fields from the difference between the measured and theoretically expected cross-power yields the contribution from unknown systematics.

We introduce an "unknown contamination coefficient", which we denote $\mathcal{U}_\ell^{\alpha,\beta}$, between any two redshift slices α and β, ℓ being the multipole moment. This parameter can be thought of as approximately a product of the ratio of power due to unknown systematics to the observed power in each of the two redshift slices. We drop all ℓ bins in the auto-power spectra of both redshift slices α and β for which $\mathcal{U}_\ell^{\alpha,\beta}$ is greater than some contamination threshold. We propose that only the remaining bins should be trusted to not be contaminated significantly with unknown systematics and consequently be used for cosmological parameter estimation.

The plan of this paper is as follows. We begin with a review of the theory and measurement of the angular power spectrum, for use later in the paper, in section 2. In section 3, we describe our method to estimate the effects of unknown systematics and exclude significantly contaminated ℓ bins. We first test the method on simulated angular power spectra contaminated with unknown systematics and demonstrate that it leads to improved bias estimates, in section 4. In section 5, we then account for unknown systematics in the SDSS-III Data Release Eight (SDSS-III DR8) luminous red galaxies (LRGs) sample [15, 19, 29], as an example of application to real data. We conclude with a discussion in section 6.

2 Review of the angular power spectrum

The power spectrum of the matter distribution in the Universe has two important features on scales relevant to current sky surveys. The first is a turn-over at $k \sim 0.01h$ Mpc^{-1}, which provides a measure of the size of the horizon at matter-radiation equality [30]. The second is oscillations caused by acoustic waves in the baryon-photon plasma in the early Universe, the scale of which is set by the sound horizon at hydrogen recombination at $z \sim 1000$ [31–37]. Together these features offer the possibility of directly measuring the angular diameter distance as a function of redshift and thus constraining standard cosmological parameters [38–46].

Practically, it is simpler to carry out an imaging survey compared to a spectroscopic survey in a given amount of telescope time. We can therefore utilize large photometric surveys to more precisely measure the angular power spectrum, and compare these measurements

---

[1]We define unknown systematics as any remaining systematics in the data after we have corrected for known systematics using given templates and assuming some model for the corrections. In addition to truly unaccounted sources of contamination in the data, these may include systematics due to inaccurate maps and/or an incomplete modeling of known systematic fields.

[2]Here we are not taking into account effects from lensing that could lead to a magnification bias for high redshift objects, as these effects will be small on the linear scales considered in this work [25–28]. We are also assuming a negligible real-space contribution to the correlation signal on very large scales.

to a theoretical 3D power spectrum by using the spectroscopic redshift distribution of the sample. In the following subsections we briefly review the theory and computation of the angular power spectrum.

2.1 From the full sample distribution to the angular power spectrum

The overdensity of dark matter halos is related to the underlying matter overdensity through the bias factor b_1 (assuming a linear bias), $\delta_{\mathrm{halo}} = b_1 \delta_{\mathrm{matter}}$. The halo power spectrum for halos of mass M at redshift z is thus given by $P_{\mathrm{halo}}(M, k, z) = b_1^2(M, z) P_{\mathrm{matter}}(k, z)$.

The theoretical angular power spectrum can be calculated by projecting the full 3D power spectrum on the sky. Using the full Bessel integration on the largest scales and accounting for redshift space distortions as described, for example, in [47], the angular power spectrum is given by

$$C_\ell = C_\ell^{\mathrm{gg}} + C_\ell^{\mathrm{gv}} + C_\ell^{\mathrm{vv}} + a. \tag{2.1}$$

The superscripts g and v denote galaxy and velocity terms, respectively, and a is an extra (constant) shot noise-like term that can be added to obtain a better fit to the non-linear power spectrum [48]. The three contributions to the angular power spectrum above are given by the integrals [47]

$$C_\ell^{\mathrm{gg}} = \frac{2}{\pi} \int \mathrm{d}\ln k \ k^3 P_{\mathrm{matter}}(k, 0) W_\ell^2(k), \tag{2.2}$$

$$C_\ell^{\mathrm{gv}} = \frac{4}{\pi} \int \mathrm{d}\ln k \ k^3 P_{\mathrm{matter}}(k, 0) W_\ell(k) W_\ell^r(k), \tag{2.3}$$

$$C_\ell^{\mathrm{vv}} = \frac{2}{\pi} \int \mathrm{d}\ln k \ k^3 P_{\mathrm{matter}}(k, 0) \left[W_\ell^r(k) \right]^2, \tag{2.4}$$

where the window functions can be calculated using

$$W_\ell(k) = \int \mathrm{d}z \ b_1 \frac{D(z)}{D(0)} \frac{\mathrm{d}N}{\mathrm{d}z} j_\ell(kr), \tag{2.5}$$

$$W_\ell^r(k) = \int \mathrm{d}z \ \Omega_m^{0.56}(z) \frac{D(z)}{D(0)} \frac{\mathrm{d}N}{\mathrm{d}z} \left[\frac{2\ell^2 + 2\ell - 1}{(2\ell - 1)(2\ell + 3)} j_\ell(kr) \right.$$
$$\left. - \frac{\ell(\ell - 1)}{(2\ell - 1)(2\ell + 1)} j_{\ell-2}(kr) - \frac{(\ell + 1)(\ell + 2)}{(2\ell + 1)(2\ell + 3)} j_{\ell+2}(kr) \right]. \tag{2.6}$$

Here $\mathrm{d}N/\mathrm{d}z$ is the redshift distribution normalized to unity, $r(z)$ is the comoving distance, and $j_\ell(kr)$ is the ℓth order spherical Bessel function. In the flat sky (large ℓ) limit, one can switch to the Limber approximation [49].

2.2 Angular power spectrum estimation

The angular power spectrum is usually calculated using the optimal quadratic estimator (OQE) method described in [47, 50–52]. We start by parameterizing the power spectrum with N step functions in ℓ, $\tilde{C}_\ell^i$,

$$C_\ell = \sum_i p_i \tilde{C}_\ell^i, \tag{2.7}$$

where the p_i are the parameters that determine the power spectrum. We form quadratic combinations of the data,

$$q_i = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{\mathcal{C}}_i\boldsymbol{\mathcal{C}}^{-1}\boldsymbol{\mathcal{C}}_i\boldsymbol{x}, \tag{2.8}$$

where $\boldsymbol{x}$ is a vector of pixelized galaxy overdensities, $\boldsymbol{\mathcal{C}}$ is the covariance matrix of the data, and $\boldsymbol{\mathcal{C}}_i$ is the derivative of the covariance matrix with respect to p_i. The covariance matrix requires a prior power spectrum to account for cosmic variance; we estimate the prior by computing an estimate of the power spectrum with a flat prior and then iterating once. We also construct the Fisher matrix,

$$F_{ij} = \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\mathcal{C}}_i\boldsymbol{\mathcal{C}}^{-1}\boldsymbol{\mathcal{C}}_j\boldsymbol{\mathcal{C}}^{-1}\right]. \tag{2.9}$$

The power spectrum can then be estimated, $\hat{\boldsymbol{p}} = \boldsymbol{F}^{-1}\boldsymbol{q}$, with covariance matrix $\boldsymbol{F}^{-1}$.

2.3 Markov-Chain Monte-Carlo procedure

Having discussed the theoretical and observed angular power spectra, we now describe our method to obtain the bias b_1 and the non-linear fitting parameter a in each redshift slice.

We use a Markov-Chain Monte-Carlo (MCMC) approach to explore the available parameter space using a modified version of the widely-used package `CosmoMC` [53]. We calculate the linear matter power spectrum using the `CAMB` code [54] included in the `CosmoMC` package, and apply the HaloFit prescription [55] to account for non-linear effects on the matter power spectrum. The resulting matter power spectrum is used in the equations of section 2.1 to calculate the theoretical angular power spectrum, which is then used in conjunction with the photometric power spectrum outlined in section 2.2 to calculate the likelihood (assumed Gaussian), that is the input to the MCMC procedure,

$$\chi^2 = (\boldsymbol{d} - \boldsymbol{t})^T \cdot \boldsymbol{\mathcal{C}}^{-1} \cdot (\boldsymbol{d} - \boldsymbol{t}). \tag{2.10}$$

Here $\boldsymbol{d}$ is the data C_ℓ vector, $\boldsymbol{t}$ is the theory C_ℓ vector convolved with the full survey window function, and $\boldsymbol{\mathcal{C}}$ is, as before, the covariance matrix. Although the OQE used to calculate the data vector is designed to compute nearly anti-correlated power spectra across different multipole bins, it does retain a small contribution ($\lesssim 5\%$) from other multipole bins. We therefore convolve the theoretical spectrum with the full window function before calculating the likelihood. This step is especially important, for example, in models with non-zero primordial non-Gaussianity, since the power spectrum rises dramatically at low ℓ in these models [56–58]. Maximizing the likelihood in the full parameter space in our MCMC analysis provides constraints on the bias and the non-linear fitting parameter in each redshift slice.

For our analysis in sections 4 and 5 we use standard cosmological data, including the WMAP nine-year CMB data [59, 60] and the "Union 2" supernova data set that includes 557 supernovae [61], as our baseline model. The LSS data that we use is the observations of LRGs in the SDSS-III DR8 sample [15, 19, 29]. As mentioned earlier, we only use photometric angular power spectra of LRGs. The underlying redshift distributions, however, are calculated using the BOSS spectroscopic redshifts of the same sample.

3 Characterizing unknown systematics

Systematics limit the amount of information that can be extracted from any survey [20, 21]. The observed LSS tracer density field is likely to be contaminated with residual systematics, the most dominant sources being stellar contamination, sky brightness, and image

quality/seeing variations. While it is possible (to some extent) to remove the effects of such *known* systematics, how do we account for *unknown* systematics? In this section we will describe a method to use cross-correlations between different redshift slices to selectively exclude certain ℓ bins that appear to have large unknown contamination. We propose that only the remaining bins should be used for cosmological parameter estimation.

3.1 The method

In the absence of strong evidence of non-linear effects of systematics on the observed density field, we adopt a linear relationship between systematics and the observed density [15, 19]. More general multiplicative errors have been considered in [20, 23, 62–64]. We find that systematics in the SDSS galaxy data that we consider are well-modeled with additive errors and adopt this simple parameterization here. We therefore write the following expression,

$$\delta_{g,\mathrm{obs}}^{\alpha}(\ell, m) = \delta_{g,\mathrm{true}}^{\alpha}(\ell, m) + \sum_{i=1}^{N_{\mathrm{sys}}} \epsilon_i^{\alpha}(\ell)\delta_i(\ell, m) + u^{\alpha}(\ell, m), \tag{3.1}$$

where $\delta_{g,\mathrm{obs}}^{\alpha}(\ell, m)$ and $\delta_{g,\mathrm{true}}^{\alpha}(\ell, m)$ are the observed and true tracer density fields in the αth redshift slice, $\delta_i(\ell, m)$ is the contribution of the ith systematic (with a total of N_{sys} systematics) to the density map, and $\epsilon_i^{\alpha}(\ell)$ (assumed independent of m) is a weight factor that characterizes the effect of the ith systematic. We further parameterize the effect of any unknown systematics as $u^{\alpha}(\ell, m)$. The above equation holds in each ℓ bin, with the observed and true angular auto-power spectrum in each redshift slice defined as $C_{\ell,\mathrm{obs}}^{\alpha,\alpha} \equiv \langle \delta_{g,\mathrm{obs}}^{\alpha}(\ell, m)\delta_{g,\mathrm{obs}}^{\alpha}(\ell, m)\rangle$ and $C_{\ell,\mathrm{true}}^{\alpha,\alpha} \equiv \langle \delta_{g,\mathrm{true}}^{\alpha}(\ell, m)\delta_{g,\mathrm{true}}^{\alpha}(\ell, m)\rangle$. Here we have assumed isotropy as there is no evidence for anisotropy in the LSS power spectrum [65].

Let us assume that the true density field is not correlated with any of the systematics, so that $\langle \delta_{g,\mathrm{true}}^{\alpha}(\ell, m)\delta_i(\ell, m)\rangle = 0$ and $\langle \delta_{g,\mathrm{true}}^{\alpha}(\ell, m)u^{\beta}(\ell, m)\rangle = 0$. We can measure the observed auto-power spectra $C_{\ell,\mathrm{obs}}^{\alpha,\alpha}$, the cross-correlations between the observed density map and systematics $\langle \delta_{g,\mathrm{obs}}^{\alpha}(\ell, m)\delta_i(\ell, m)\rangle$, and the correlations amongst various systematics $\langle \delta_i(\ell, m)\delta_j(\ell, m)\rangle$. We can now write the following set of N_{sys} equations in each ℓ bin of the αth redshift slice,

$$\langle \delta_{g,\mathrm{obs}}^{\alpha}(\ell, m)\delta_j(\ell, m)\rangle$$
$$= \sum_{i=1}^{N_{\mathrm{sys}}} \epsilon_i^{\alpha}(\ell)\langle \delta_i(\ell, m)\delta_j(\ell, m)\rangle + \langle u^{\alpha}(\ell, m)\delta_j(\ell, m)\rangle, \quad j = 1, \ldots, N_{\mathrm{sys}}. \tag{3.2}$$

The unknown quantities in the above equations are the weights $\epsilon_i^{\alpha}(\ell)$ and the correlations $\langle u^{\alpha}(\ell, m)\,\delta_j(\ell, m)\rangle$. We do not have a sufficient number of equations to solve for all of these quantities (we have N_{sys} equations and $2N_{\mathrm{sys}}$ unknowns). We investigated the use of cross-correlations between different redshift slices to solve for the remaining unknowns, but this approach failed, as we will now demonstrate.

We can write the following equation connecting the true and observed angular auto- or cross-power spectra,

$$C_{\ell,\mathrm{true}}^{\alpha,\beta} = C_{\ell,\mathrm{obs}}^{\alpha,\beta} - \sum_{i,j=1}^{N_{\mathrm{sys}}} \epsilon_i^{\alpha}(\ell)\epsilon_j^{\beta}(\ell)\langle \delta_i(\ell, m)\delta_j(\ell, m)\rangle - U_{\ell}^{\alpha,\beta}, \tag{3.3}$$

where $U_\ell^{\alpha,\beta} = \sum_{i=1}^{N_{\rm sys}} \left(\epsilon_i^\alpha(\ell) \langle \delta_i(\ell,m) u^\beta(\ell,m) \rangle + \epsilon_i^\beta(\ell) \langle \delta_i(\ell,m) u^\alpha(\ell,m) \rangle \right) + \langle u^\alpha(\ell,m) u^\beta(\ell,m) \rangle$.
Although we do not know a priori the true auto-power spectra $C_{\ell,\rm true}^{\alpha,\alpha}$, we can estimate the
true cross-power spectra $C_{\ell,\rm true}^{\alpha,\beta}$ ($\alpha \neq \beta$) theoretically using the equations given in section 2.1.
Here we assume that the true cross-power closely follows background cosmology, although,
strictly speaking, this is correct only if the overlap in the redshift distribution between the
two slices is small. We need the bias in each of the two redshift slices α and β — let us
assume that we know this — and the cross-redshift distribution, which we can calculate as
the overlap of that in the two redshift slices. Then the only remaining unknowns in these
equations are the weights $\epsilon_i^\alpha(\ell)$, the correlations $\langle u^\alpha(\ell,m) \delta_j(\ell,m) \rangle$, and additionally the
correlations $\langle u^\alpha(\ell,m) u^\beta(\ell,m) \rangle$. One can easily check that we still lack a sufficient number
of equations to solve for all unknown quantities. Therefore we can *not* correct for unknown
systematics in the data. We *can*, however, estimate their contribution and decide to drop
those ℓ bins that are dominated by unknown systematics.

In order to achieve this goal, we first obtain a zeroth order estimate of the weights
$\epsilon_i^\alpha(\ell)$ by solving eq. (3.2) in each ℓ bin under the assumption that $u^\alpha(\ell,m) = 0$. We then
obtain the systematics-corrected auto-power spectrum in each redshift slice (again under the
assumption that $u^\alpha(\ell,m) = 0$) using

$$C_{\ell,\rm true}^{\alpha,\alpha} = C_{\ell,\rm obs}^{\alpha,\alpha} - \sum_{i,j=1}^{N_{\rm sys}} \epsilon_i^\alpha(\ell) \epsilon_j^\alpha(\ell) \langle \delta_i(\ell,m) \delta_j(\ell,m) \rangle. \tag{3.4}$$

We now perform an MCMC procedure using the above auto-power spectra to fit for the
background cosmology, the bias, and the non-linear fitting parameter a (introduced in eq.
(2.1)) in each redshift slice. With these parameters in hand, we obtain a theoretical estimate
of $C_{\ell,\rm true}^{\alpha,\beta}$ ($\alpha \neq \beta$) as described earlier. We also include the a parameter in the cross-power
— ideally this should be zero, so we add the maximum value of $\sqrt{|a^\alpha|\,|a^\beta|}$ (where a^α and a^β
correspond to values of a for redshift slices α and β) only in the cross-power of neighboring
redshift slices. Finally, using eq. (3.3) we solve for $U_\ell^{\alpha,\beta}$ ($\alpha \neq \beta$).[3]

We now define an unknown contamination coefficient as

$$\mathcal{U}_\ell^{\alpha,\beta} = \frac{\left(U_\ell^{\alpha,\beta} \right)^2}{C_{\ell,\rm obs}^{\alpha,\alpha} C_{\ell,\rm obs}^{\beta,\beta}}, \quad \alpha \neq \beta, \tag{3.5}$$

and compare it to the quantity

$$n^2 \frac{\sigma\left(C_{\ell,\rm obs}^{\alpha,\alpha} \right) \sigma\left(C_{\ell,\rm obs}^{\beta,\beta} \right)}{C_{\ell,\rm obs}^{\alpha,\alpha} C_{\ell,\rm obs}^{\beta,\beta}}, \tag{3.6}$$

where n refers to the sigma tolerance level. We drop all ℓ bins in the auto-power of each
redshift slice α and β for which $\mathcal{U}_\ell^{\alpha,\beta}$ is greater than the quantity in eq. (3.6).[4] For example,
we can choose $n = 1$ and perform a 1σ-cut on unknown systematics. The motivation for this
definition is that it provides some idea of unknown contamination in units of the error in the

---

[3]When calculating $U_\ell^{\alpha,\beta}$, we do not convolve the theoretical cross-power with the survey window function.
[4]Clipping the cross-power at $n\sigma$ will in general introduce a bias in the variance of each auto-power spectrum.
We checked that this effect is small for two Gaussian random distributions, introducing a percent level bias
for a 2σ-cut. We ignore this effect here.

observed auto-power. This is not a precise cut, however, as there is an error associated with the cross-power itself as well. Nevertheless, it does serve as a uniform measure to quantify the level of unknown contamination. After performing this cut, we use the remaining bins in the auto-power spectra to fit for the background cosmology, the bias and the a parameter in each redshift slice. If the new bias values lie within one sigma of the previous estimates of the bias then these are the final values to use. If not, then we repeat this analysis until the bias lies within one sigma of the previous iteration.

Summary. To summarize, we use the following algorithm to characterize unknown systematics —

1. Correct the observed auto-power spectra for known systematics, assuming no unknown contamination, using eq. (3.2) with $u^\alpha(\ell) = 0$ and eq. (3.4).

2. Perform an MCMC analysis over background cosmological parameters, the bias, and the non-linear fitting parameter a in each redshift slice.

3. Use best-fit values of the bias and a parameters to obtain theoretical cross-power spectra. Compare the resulting true cross-power with the observed cross-power using eq. (3.3) to obtain $U_\ell^{\alpha,\beta}$, and subsequently obtain $\mathcal{U}_\ell^{\alpha,\beta}$ defined in eq. (3.5).

4. Perform an $n\sigma$-cut on unknown systematics by excluding all ℓ bins in each auto-power spectrum for which $\mathcal{U}_\ell^{\alpha,\beta}$ with any other redshift slice is greater than the quantity in eq. (3.6).

5. Using remaining bins in the auto-power spectra, perform an MCMC analysis over background cosmological parameters, the bias, and the parameter a in each redshift slice, to obtain new estimates of various parameters.

We repeat steps 3, 4, and 5 until the new bias values in each redshift slice lie within one sigma of the previous iteration.

3.2 Estimating the covariance matrix

In the above discussion we have not yet defined the errors on the angular auto-power spectra and the full covariance matrix between different redshift slices that will be needed for the MCMC fitting. In the absence of a complete understanding of mock systematic fields, it is unlikely that we can obtain optimal errors on systematic corrections within the scope of this analysis. We therefore adopt a simplistic model to estimate the covariance of the systematics-corrected power spectra, similar to [15, 19], which we describe below.

Assuming Gaussianity of the fields involved, the covariance matrix has a simple structure of non-zero diagonal elements, with entries for multipoles $\ell^\alpha = \ell^\beta$ for redshift slices α and β. These covariances are given by

$$\sigma^2\big(C_\ell^{\alpha,\alpha}\big) = \frac{2}{f_{\mathrm{sky}}(2\ell + 1)}\left(C_{\ell,\mathrm{smooth}}^{\alpha,\alpha} + N_{\mathrm{shot},\alpha}\right)^2, \tag{3.7}$$

$$\sigma^2\big(C_\ell^{\alpha,\alpha}C_\ell^{\beta,\beta}\big) = \frac{2}{f_{\mathrm{sky}}(2\ell + 1)}\left(C_{\ell,\mathrm{smooth}}^{\alpha,\beta}\right)^2, \tag{3.8}$$

where $\sigma^2\big(C_\ell^{\alpha,\alpha}\big)$ is the (diagonal) variance in the auto-power of redshift slice α and $\sigma^2\big(C_\ell^{\alpha,\alpha} C_\ell^{\beta,\beta}\big)$ is the (diagonal) cross-covariance between redshift slices α and β. Note that the cross-covariance between two redshift slices is different from the variance in the cross-power between

the two redshift slices, which is given by eqs. (4.1) and (4.2) in the next section. Here f_{sky} is the fraction of the sky observed and the shot noise is given by $N_{\text{shot}} = f_{\text{sky}} \times 4\pi/N_{\text{sample}}$, where N_{sample} is the effective number of objects observed (e.g. galaxies, after weighting by the probability that an observed object is a galaxy). The shot noise term only contributes to the variance of the auto-power spectra. The subscript 'smooth' denotes a smooth fit to the observed auto- or cross-power spectra. Also, the above equations assume a unit width for each ℓ bin, i.e. $\Delta\ell = 1$.

We modify the above set of equations by adding the correctional power due to systematics [15, 19],

$$\sigma^2\big(C_\ell^{\alpha,\alpha}\big) = a_{\text{fac}}^2 \frac{2}{f_{\text{sky}} \sum_{\ell=\ell_{\min}}^{\ell_{\max}-1}(2\ell+1)} \left(\sqrt{\left(C_{\ell,\text{smooth}}^{\alpha,\alpha}\right)^2 + \left(\Delta C_\ell^{\alpha,\alpha}\right)^2} + N_{\text{shot},\alpha} \right)^2, \quad (3.9)$$

$$\sigma^2\big(C_\ell^{\alpha,\alpha} C_\ell^{\beta,\beta}\big) = a_{\text{fac}}^2 \frac{2}{f_{\text{sky}} \sum_{\ell=\ell_{\min}}^{\ell_{\max}-1}(2\ell+1)} \left(\left(C_{\ell,\text{smooth}}^{\alpha,\beta}\right)^2 + \left(\Delta C_\ell^{\alpha,\beta}\right)^2 \right), \quad (3.10)$$

where $\Delta C_\ell^{\alpha,\alpha} = \sum_{i,j=1}^{N_{\text{sys}}} \epsilon_i^\alpha(\ell)\epsilon_j^\alpha(\ell)\langle\delta_i(\ell,m)\delta_j(\ell,m)\rangle$ and $\Delta C_\ell^{\alpha,\beta} = \sum_{i,j=1}^{N_{\text{sys}}} \epsilon_i^\alpha(\ell)\epsilon_j^\beta(\ell)\langle\delta_i(\ell,m)$ $\delta_j(\ell,m)\rangle$. We have also taken into account the fact that $\Delta\ell \neq 1$, with the total number of modes in each ℓ bin coming from the sum of modes in $\ell_{\min} \leq \ell < \ell_{\max}$. Further, we have boosted the diagonal error with an empirical factor of a_{fac} to account for the fact that the Gaussian approximation is not perfect and neighboring ℓ bins contribute to the diagonal error as well. For example, in our LRG analysis in section 5 we choose $a_{\text{fac}} = 1.1$ [15].

Finally, we calculate the off-diagonal elements of the covariance matrix by preserving the structure of the OQE covariance matrix,

$$\sigma^2\big(C_{\ell,\ell'}^{\alpha,\alpha}\big) = \frac{\sigma^2\big(C_{\ell,\ell',\text{OQE}}\big)}{\sqrt{\sigma^2\big(C_{\ell,\text{OQE}}\big)\sigma^2\big(C_{\ell',\text{OQE}}\big)}} \sqrt{\sigma^2\big(C_\ell^{\alpha,\alpha}\big)\sigma^2\big(C_{\ell'}^{\beta,\beta}\big)}, \quad (3.11)$$

$$\sigma^2\big(C_\ell^{\alpha,\alpha} C_{\ell'}^{\beta,\beta}\big) = \frac{\sigma^2\big(C_{\ell,\ell',\text{OQE}}\big)}{\sqrt{\sigma^2\big(C_{\ell,\text{OQE}}\big)\sigma^2\big(C_{\ell',\text{OQE}}\big)}} \sqrt{\sigma^2\big(C_\ell^{\alpha,\alpha} C_\ell^{\beta,\beta}\big)\sigma^2\big(C_{\ell'}^{\alpha,\alpha} C_{\ell'}^{\beta,\beta}\big)}. \quad (3.12)$$

This is a valid approximation since all redshift slices usually have similar OQE covariance structures (arising from the use of a common mask) [15, 16]. One can then choose any redshift slice to generate the ratio in the above equations.

We will now apply the method described here to simulated (section 4) and real (section 5) angular power spectra in the following sections.

4 Simulation

In this section we generate mock angular power spectra contaminated with known and unknown systematics in two redshift slices, and obtain estimates of the bias by discarding heavily contaminated bins. We choose redshift slices $0.50 \leq z \leq 0.55$ and $0.55 \leq z \leq 0.60$, labeled LRG2s and LRG3s ('s' standing for 'simulation'), with redshift distributions identical to those for LRGs in SDSS-III DR8 [15, 19]. We generate auto- and cross-power spectra using the equations in section 2.1 with bias values of 2.0 and 2.2 and the non-linear fitting parameter a set to zero in the two redshift slices, assuming a WMAP9 + SN ΛCDM cosmology. We then add mock systematic fields for stellar contamination, sky brightness, and

seeing variations to both, the auto- and cross-power spectra. In order to generate the mock systematic fields, we use actual maps of these systematics and their cross-correlations with the observed galaxy density given in [15] and slightly exaggerate the effect of stars. Having maps of the systematics and their cross-correlations with galaxies one can calculate the various $\epsilon_i^\alpha(\ell)$ coefficients using eq. (3.2), with $u^\alpha(\ell) = 0$, and add in these systematics to obtain the observed angular power spectra from the theoretical power spectra. We subsequently correct for only sky brightness and seeing variations, leaving stars as an unknown systematic in the density fields.[5]

In order to correct for known systematics (i.e. sky and seeing) we first solve eq. (3.2) for the weights $\epsilon_i^\alpha(\ell)$ in each ℓ bin, under the assumption that $u^\alpha(\ell) = 0$. We then use eq. (3.4) to obtain the corrected auto-power spectra in each redshift slice. In fig. 1 we present the auto- and cross-power spectra in the two redshift slices before and after adding systematics. We also show the power spectra corrected for known systematics for comparison. Eqs. (3.9) - (3.12) further provide the full covariance matrix for the MCMC analysis, where for the smooth power spectra we use the observed spectra that include all three systematics and for the off-diagonal elements we use the structure of the OQE covariance matrix from one of the redshift slices in DR8.

To compute the errors in the cross-power (which we do not need in our analysis but we do show in fig. 1), we use a modified version of the following equation that assumes Gaussian errors (see, e.g., [66, 67]),

$$\sigma^2\big(C_\ell^{\alpha,\beta}\big) = \frac{1}{f_{\text{sky}}(2\ell+1)}\bigg(\Big(C_{\ell,\text{smooth}}^{\alpha,\beta}\Big)^2 + \Big(C_{\ell,\text{smooth}}^{\alpha,\alpha} + N_{\text{shot},\alpha}\Big)\Big(C_{\ell,\text{smooth}}^{\beta,\beta} + N_{\text{shot},\beta}\Big) \bigg), \qquad (4.1)$$

to include the correctional power due to systematics,

$$\sigma^2\big(C_\ell^{\alpha,\beta}\big) = a_{\text{fac}}^2 \frac{1}{f_{\text{sky}}\sum_{\ell=\ell_{\min}}^{\ell_{\max}-1}(2\ell+1)}\bigg(\Big(C_{\ell,\text{smooth}}^{\alpha,\beta}\Big)^2 + \Big(\Delta C_\ell^{\alpha,\beta}\Big)^2$$
$$+ \bigg(\sqrt{\Big(C_{\ell,\text{smooth}}^{\alpha,\alpha}\Big)^2 + \Big(\Delta C_\ell^{\alpha,\alpha}\Big)^2} + N_{\text{shot},\alpha}\bigg)$$
$$\times \bigg(\sqrt{\Big(C_{\ell,\text{smooth}}^{\beta,\beta}\Big)^2 + \Big(\Delta C_\ell^{\beta,\beta}\Big)^2} + N_{\text{shot},\beta}\bigg) \bigg). \qquad (4.2)$$

This is the variance in the cross-power between redshift slices α and β, which includes an extra contribution from the auto-power, since the auto-power effectively acts like an extra source of noise when computing the error in the cross-power.

Before we proceed with the MCMC analysis, in fig. 2 we show the amount of unknown contamination (i.e. the contamination due to stars) in the auto- and cross-power spectra relative to the errors in the observed angular power spectrum in the two redshift slices. The quantity $U_\ell^{\alpha,\beta}$ defined in eq. (3.3) is simply calculated as the difference between the corrected data points and the corresponding theoretical angular power spectra in fig. 1. Then fig. 2 shows that in redshift slice LRG2s, bins at $\ell = 8.5, 16.5, 55, 75, 85, 105$ are contaminated

---

[5]We checked that on correcting for all three systematics we reproduce the theoretical power spectra as expected.
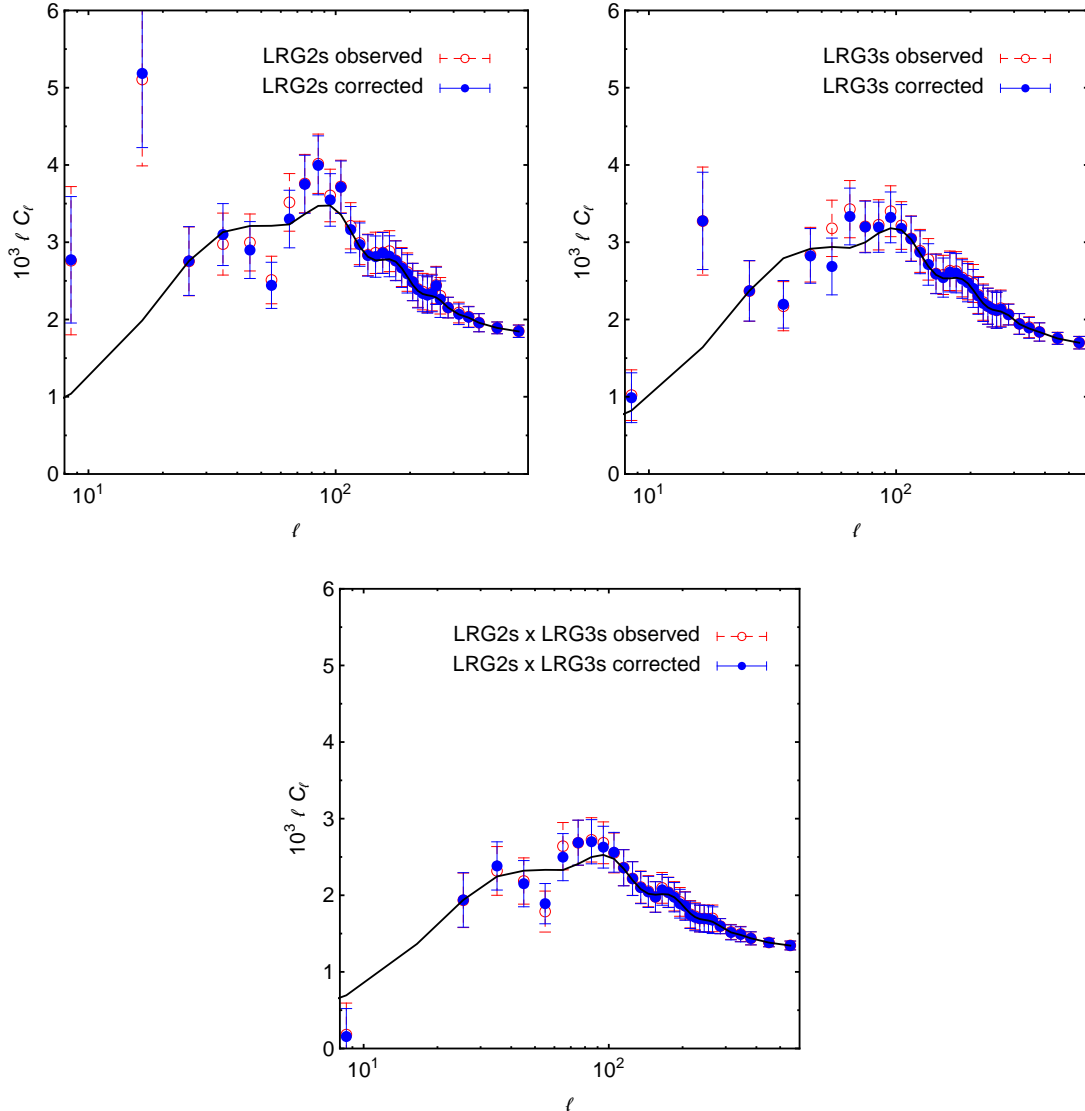
Figure 1. The auto- (top panels) and cross-power (bottom panel) spectra in simulated data for two redshift slices. The solid black curves are the original theoretical power spectra for bias values of 2.0 and 2.2 in redshift slices LRG2s and LRG3s, respectively. Red open circles with dashed (1σ) error bars are mock observed data points in which we have added three systematics, while blue filled circles with solid (1σ) error bars are data points corrected for only two systematics. The corrected cross-power is calculated analogous to the auto-power, using eq. (3.3) with $U_\ell^{\alpha,\beta} = 0$.

at more than one sigma, while in redshift slice LRG3s, bins at $\ell = 16.5$, 35, 65 are similarly contaminated. The cross-power, however, is expected to pick up contaminations in bins at $\ell = 16.5$, 55. We will see below that this is in fact what we find in the MCMC analysis.

Now we perform the MCMC analysis described in section 2.3. We choose a low-ℓ cutoff for the angular power spectrum in each redshift slice at $\ell_{\min} = 30$ since we expect lower multipoles to mostly be dominated by systematics. We also choose a high-ℓ cutoff, $\ell_{\max}$,
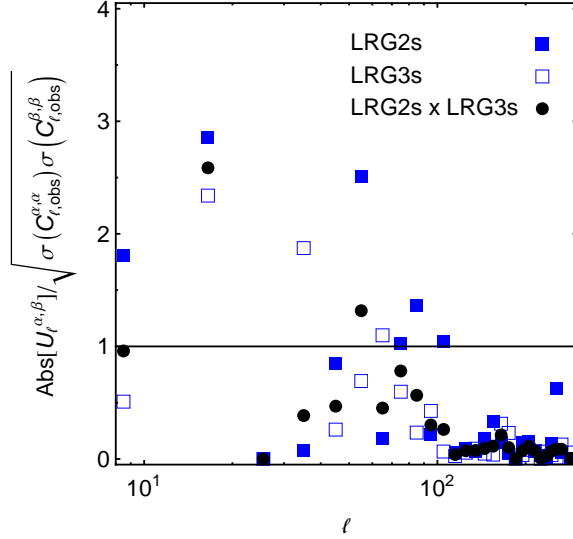
Figure 2. The absolute value of the unknown contamination $U_\ell^{\alpha,\beta}$ defined in eq. (3.3) compared to the errors in the observed angular power spectra in the two redshift slices for simulated LRGs. Squares show the contamination in the auto-power spectra, i.e. $\text{Abs}[U_\ell^{\alpha,\alpha}]/\sigma\left(C_{\ell,\text{obs}}^{\alpha,\alpha}\right)$ while circles show the contamination in the cross-power. The solid line is the 1σ-line.

corresponding to $k = 0.1h\ \text{Mpc}^{-1}$ (determined using ΛCDM cosmology, see table 1) to avoid the strongly non-linear regime of the matter power spectrum. In addition to the bias and non-linear fitting parameter in each redshift slice, we also vary over the standard cosmological parameters $\left\{\Omega_b h^2, \Omega_{\text{DM}} h^2, \theta, \tau, n_s, \log A_s, A_{\text{SZ}}\right\}$. Here $\Omega_b h^2$ is the physical baryon density, $\Omega_{\text{DM}} h^2$ is the physical dark matter density, θ is the ratio of the sound horizon to the angular diameter distance at decoupling, τ is the reionization optical depth, n_s is the scalar spectral index, A_s is the amplitude of the primordial scalar curvature perturbations at $k = 0.05\ \text{Mpc}^{-1}$, and A_{SZ} represents a Sunyaev-Zeldovich template normalization. We use flat priors for all parameters. This analysis yields the bias and a parameters shown in the sixth and seventh columns of table 1.

Label	z_{mid}	Input b_1	Input a	l_{max}	b_1	$10^6 a$	b_1 (1$^{\text{st}}$ it.)	$10^6 a$ (1$^{\text{st}}$ it.)
LRG2s	0.525	2.0	0	140	$1.91^{+0.09}_{-0.09}$	$2.97^{+2.71}_{-2.72}$	$2.03^{+0.08}_{-0.09}$	$-0.04^{+2.97}_{-2.92}$
LRG3s	0.575	2.2	0	151	$2.11^{+0.10}_{-0.09}$	$2.02^{+2.33}_{-2.36}$	$2.15^{+0.09}_{-0.10}$	$1.31^{+2.44}_{-2.44}$

Table 1. The best-fit Gaussian bias and non-linear fitting parameter (with 1σ errors) in the two simulated redshift slices for LRGs, using WMAP9 + SN + simulated LRG data. In the sixth and seventh columns we have used all available ℓ bins in $30 \leq \ell \leq \ell_{\text{max}}$, while in the next two columns we have used only those bins in $10 \leq \ell \leq \ell_{\text{max}}$ that satisfy a 1σ-cut on $\mathcal{U}_\ell^{\alpha,\beta}$.

Next we use the best-fit bias and a parameters to obtain the theoretical cross-power spectrum between the two redshift slices. As noted earlier, we calculate the cross-redshift distribution needed here as an overlap of the distribution in the two slices. We also add $\sqrt{|a^\alpha|\,|a^\beta|}$ to the theoretical cross-power, to take into account any non-linear contributions

in the region with which we are concerned. Comparing this 'true' power spectrum with the observed cross-power and using the weights $\epsilon_i^\alpha(\ell)$ obtained earlier, we calculate $U_\ell^{\alpha,\beta}$ using eq. (3.3) in each ℓ bin. Finally we determine the unknown contamination coefficient $\mathcal{U}_\ell^{\alpha,\beta}$ using eq. (3.5). We show the absolute value of this quantity with the 1σ-cut that we apply in fig. 3 and note that bins with a contamination greater than one sigma do indeed correspond to those shown earlier in the cross-power in fig. 2.
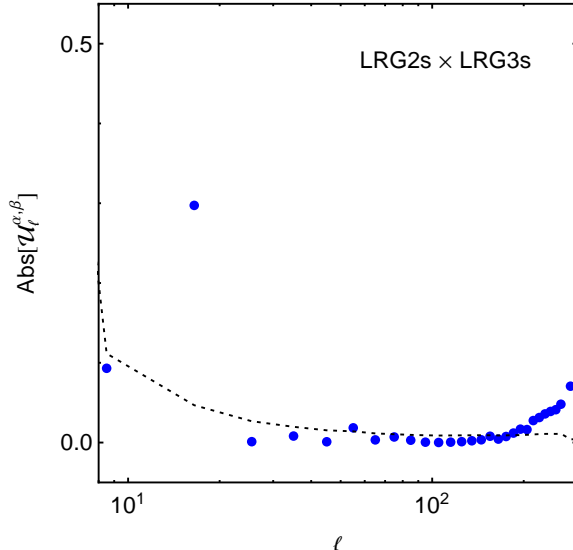
Figure 3. The absolute value of the unknown contamination coefficient $\mathcal{U}_\ell^{\alpha,\beta}$ defined in eq. (3.5) for simulated LRGs (filled circles). The dotted line shows the absolute value of the 1σ-cut — we drop all bins that lie above this cut. The upturn in the value of $\mathcal{U}_\ell^{\alpha,\beta}$ at high ℓ comes from the cross-power being dominated by the non-linear fitting parameters on these scales.

We now drop all ℓ bins in the auto-power of each of the two redshift slices, for which the unknown contamination coefficient lies outside the 1σ-cut. Using only the remaining bins in $10 \leq \ell \leq \ell_{\max}$ and the corresponding rows and columns of the full covariance matrix obtained earlier, we perform an MCMC analysis on the standard cosmological parameters, the bias, and the a parameter in each redshift slice. This exercise yields the bias and a values denoted as '1$^{\text{st}}$ it.' (First iteration) in table 1. The bias in the redshift slice LRG2s lies slightly outside the one sigma bound of the previous estimate. When we used these new estimates of the bias to determine which bins to drop, however, we ended up with the same bins in $10 \leq \ell \leq \ell_{\max}$ as the first iteration. These values of the bias are therefore the final values obtained on applying our method. As one can see, estimates of the bias are now closer to the original values of 2.0 and 2.2. We also find that this fit leads to a reduction in χ^2 and thus represents an improved fit to the data.

In fig. 4 we present the angular power spectrum corrected for known systematics and mark the bins that we drop in each redshift slice. We also show theoretical curves obtained using best-fit values of the bias and the a parameter before and after dropping bins. According to the method presented here we exclude bins from each of the two redshift slices that have a significantly contaminated cross-correlation, as we do not know which of the two redshift slices is responsible for this contamination. For this reason, there are some ℓ bins in the

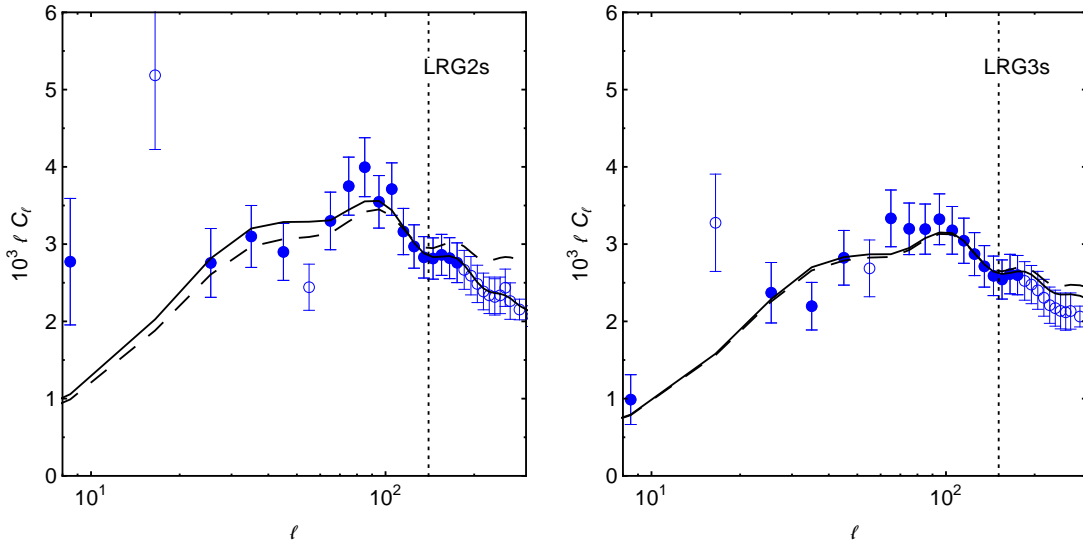auto-power spectra of each redshift slice that may not appear to be contaminated but are still dropped.

Figure 4. The angular power spectrum in the two simulated redshift slices of LRGs. Open circles with (1σ) error bars represent data points that are dropped due to large unknown systematics, as determined using the cross-power spectra. Filled circles, on the other hand, are data points that are not dominated with unknown systematics. The vertical dotted line shows $\ell_{\max}$. The dashed curve is the theoretical power spectrum obtained using the best-fit bias parameters that correspond to the MCMC analysis which uses all data points in $30 \le \ell \le \ell_{\max}$. The solid curve represents the MCMC analysis which uses only filled circles in $10 \le \ell \le \ell_{\max}$. In both theoretical curves we have added in the corresponding best-fit a parameter to all C_ℓs; this causes the upturn at large ℓ.

It is important to note that the theoretically calculated 'true' cross-power spectrum depends on the assumed background cosmological model. There is a possibility therefore that the bins that are discarded could be showing a real cosmological signal of interest. In order to assess this effect, we repeat the analysis of this section with a different background cosmology. In particular, we consider the case of non-zero primordial non-Gaussianity. As was first reported in [56], the halo bias includes a scale-dependent term given by

$$\Delta b(M, k, z, f_{\mathrm{NL}}) = 3f_{\mathrm{NL}}[b_1(M, z) - p]\delta_c \frac{\Omega_m H_0^2}{k^2 T(k) D(z)} \tag{4.3}$$

in the presence of primordial (local) non-Gaussianity, usually parameterized by the parameter f_{NL}. Here, $\delta_c \approx 1.686$ denotes the critical density for spherical collapse, Ω_m is the present day matter density, H_0 is the Hubble constant, $T(k)$ is the matter transfer function normalized to unity as $k \to 0$, and $D(z)$ is the linear growth function normalized to $(1 + z)^{-1}$ in the matter-dominated era. The parameter p ranges from unity for objects that populate all halos equally to 1.6 for objects that populate only recently merged halos [58]. For LRGs we can set p to unity. We generate auto- and cross-power spectra as before, assuming a WMAP9 + SN ΛCDM cosmology, but in addition setting $f_{\mathrm{NL}} = 25$. We then add in three mock systematic fields and correct for two of them, leaving the third as an unknown systematic.

We subsequently perform an MCMC analysis assuming that $f_{\rm NL} = 0$, expecting to find more low-ℓ bins to be contaminated with unknown systematics, since the effect of non-zero $f_{\rm NL}$ is more predominant on large scales. Since the effect of systematics is large compared to the actual cosmological signal of a small non-zero $f_{\rm NL}$, however, we find results very similar to what we found earlier, with the method dropping the same bins as before. We do find, however, slightly higher estimates of the Gaussian bias since $f_{\rm NL} > 0$ increases the power on large scales. With more precise data in the future it may be useful to implement the method presented here for each parameter space being explored, for instance including $f_{\rm NL}$ in the MCMC analysis that determines which bins to discard and using the resulting best-fit value of $f_{\rm NL}$ to calculate the true cross-power.

5   A case study: SDSS-III LRGs

Let us now apply the method discussed in sections 3 and 4 to actual data from SDSS-I, II and III [68–71]. We use LRGs from the SDSS-III DR8 sample described in [15, 19]. We refer the reader to these papers for details and highlight some of the main properties of the sample here. The data set spans $\sim 11{,}000$ square degrees of the sky and probes a volume of $\sim 3h^{-3}$ Gpc^3. We focus on the approximately stellar mass-limited CMASS sample of luminous galaxies, that follows the CMASS galaxy selection detailed in [72]. Photometric redshifts and the probability that an object is a galaxy are obtained using a training sample of 112,778 BOSS CMASS spectra. The final photometric redshift catalog consists of a total of 872,921 luminous galaxies in the redshift range $0.45 \leq z \leq 0.65$, divided into four photometric bins (labeled LRG1 through LRG4) — $z = 0.45-0.50$, $0.50-0.55$, $0.55-0.60$, and $0.60-0.65$, with the effective number of galaxies in each bin being 214971, 258736, 248895, and 150319, respectively. The calculation of the angular power spectra in the four redshift bins uses the OQE method outlined in section 2.2.

We first take the output of the OQE and correct for dominant known systematics, which include stellar contamination, sky brightness, and seeing variations. The maps of these systematics and their cross-correlations with the observed galaxy density field were given in [15], so we will not repeat them here. As before, we correct for these systematics assuming that there are no unknown systematics, and also obtain the full covariance matrix that is used in the MCMC analysis.

We perform the MCMC analysis described in section 2.3, choosing a low-ℓ cutoff for the angular power spectrum in each redshift slice at $\ell_{\rm min} = 30$ and a high-ℓ cutoff, $\ell_{\rm max}$, corresponding to $k = 0.1h$ Mpc^{-1} (see table 2). In addition to the bias and non-linear fitting parameter in each redshift slice, we also vary over the standard cosmological parameters $\left\{\Omega_b h^2, \Omega_{\rm DM} h^2, \theta, \tau, n_s, \log A_s, A_{\rm SZ}\right\}$. The resulting bias and a parameters obtained are given in the fourth and fifth columns of table 2. Using the best-fit bias and a parameters we calculate the theoretical cross-power spectra between the four redshift slices, adding $\sqrt{|a^\alpha|\,|a^\beta|}$ to the cross-power between consecutive redshift slices. Comparing these 'true' power spectra with the observed cross-power and using the weights $\epsilon_i^\alpha(\ell)$ obtained earlier (under the assumption of no unknown systematics), we calculate the unknown contamination coefficient $\mathcal{U}_\ell^{\alpha,\beta}$. We show the absolute value of this quantity along with the 3σ-cut that we apply in fig. 5. We choose a more conservative 3σ-cut as we are now dealing with real data.

We now drop all ℓ bins in the auto-power of each redshift slice, whose unknown contamination in the cross-power with some other redshift slice lies outside the corresponding 3σ-cut. Again, using only the remaining bins in $10 \leq \ell \leq \ell_{\rm max}$ and the corresponding rows

Label	$z_{\rm mid}$	$l_{\rm max}$	b_1	$10^6 a$	b_1 (1$^{\rm st}$ it.)	$10^6 a$ (1$^{\rm st}$ it.)
LRG1	0.475	128	$1.89^{+0.09}_{-0.10}$	$4.93^{+4.38}_{-4.27}$	$1.96^{+0.14}_{-0.14}$	$2.86^{+5.78}_{-5.79}$
LRG2	0.525	140	$1.88^{+0.09}_{-0.10}$	$6.32^{+2.90}_{-2.88}$	$1.97^{+0.11}_{-0.10}$	$3.83^{+3.32}_{-3.30}$
LRG3	0.575	151	$2.16^{+0.10}_{-0.09}$	$0.74^{+2.48}_{-2.45}$	$2.09^{+0.12}_{-0.13}$	$2.14^{+3.10}_{-3.08}$
LRG4	0.625	162	$2.26^{+0.11}_{-0.11}$	$1.40^{+2.31}_{-2.32}$	$2.26^{+0.11}_{-0.12}$	$1.44^{+2.35}_{-2.35}$

Table 2. The best-fit Gaussian bias and non-linear fitting parameter (with 1σ errors) in the four redshift slices for LRGs, using WMAP9 + SN + DR8 (LRG) data. In the fourth and fifth columns we have used all available ℓ bins in $30 \leq \ell \leq \ell_{\rm max}$, while in the next two columns we have used only those bins in $10 \leq \ell \leq \ell_{\rm max}$ that satisfy a 3σ-cut on $\mathcal{U}_\ell^{\alpha,\beta}$.

and columns of the full covariance matrix obtained earlier, we perform an MCMC analysis on the standard cosmological parameters, the bias, and the a parameter in each redshift slice. This yields the bias and a values given in the last two columns of table 2. Values of the bias lie within one sigma of the previous bias estimates, so these are the final values to use.

In fig. 6 we present the measured angular power spectrum in each redshift slice after correcting for known systematics and mark the bins that are dropped based on the method discussed above. We also display theoretical curves[6] obtained using best-fit values for the bias and a before and after dropping bins. In each redshift slice there are some bins that do not appear to be contaminated but are still dropped as their cross-power with another redshift slice is significantly contaminated.

To see how this method changes the estimates of other cosmological parameters, we examine a specific example of primordial non-Gaussianity. We add the term given in eq. (4.3) to the bias b_1 and introduce an extra parameter $f_{\rm NL}$ to the MCMC analysis. We do *not* include $f_{\rm NL}$ in the MCMC method to determine which bins to drop, but rather only include it as a parameter to fit *after* determining which bins to drop assuming $f_{\rm NL} = 0$. In order to constrain $f_{\rm NL}$ we set the (Gaussian) bias b_1 and the non-linear fitting parameter a to their corresponding best-fit values in each redshift slice, and perform an MCMC analysis over the standard cosmological parameters and $f_{\rm NL}$. Using all ℓ bins in $30 \leq \ell \leq \ell_{\rm max}$ we find that $f_{\rm NL} = -5^{+62}_{-62}$ (1σ error), using instead all bins in $10 \leq \ell \leq \ell_{\rm max}$ we obtain a tighter constraint of $f_{\rm NL} = -77^{+47}_{-46}$ (1σ error), and finally using only those bins in $10 \leq \ell \leq \ell_{\rm max}$ that satisfy the 3σ-cut we obtain $f_{\rm NL} = -17^{+68}_{-68}$ (1σ error). For data at higher redshifts (such as quasars) we expect the variation in parameter estimation to be even more significant, and therefore propose that one should use a method such as that discussed in this paper to exclude significantly contaminated bins in photometric LSS surveys (for instance, see [16]).

6   Discussion

Photometric surveys are currently limited by systematics. In order to correctly use data from current and upcoming surveys, it is important to understand and account for various sources of contamination that affect photometric samples. This issue has motivated works such as [21, 22, 24]. In [21], the authors demonstrated that auto-correlations of quasars in the SDSS DR6 sample [73, 74] are unfit to constrain primordial non-Gaussianity. They constructed templates

---

[6]The theoretical curves shown in fig. 6 differ from those used in calculating the likelihood by the effect of the survey window function.
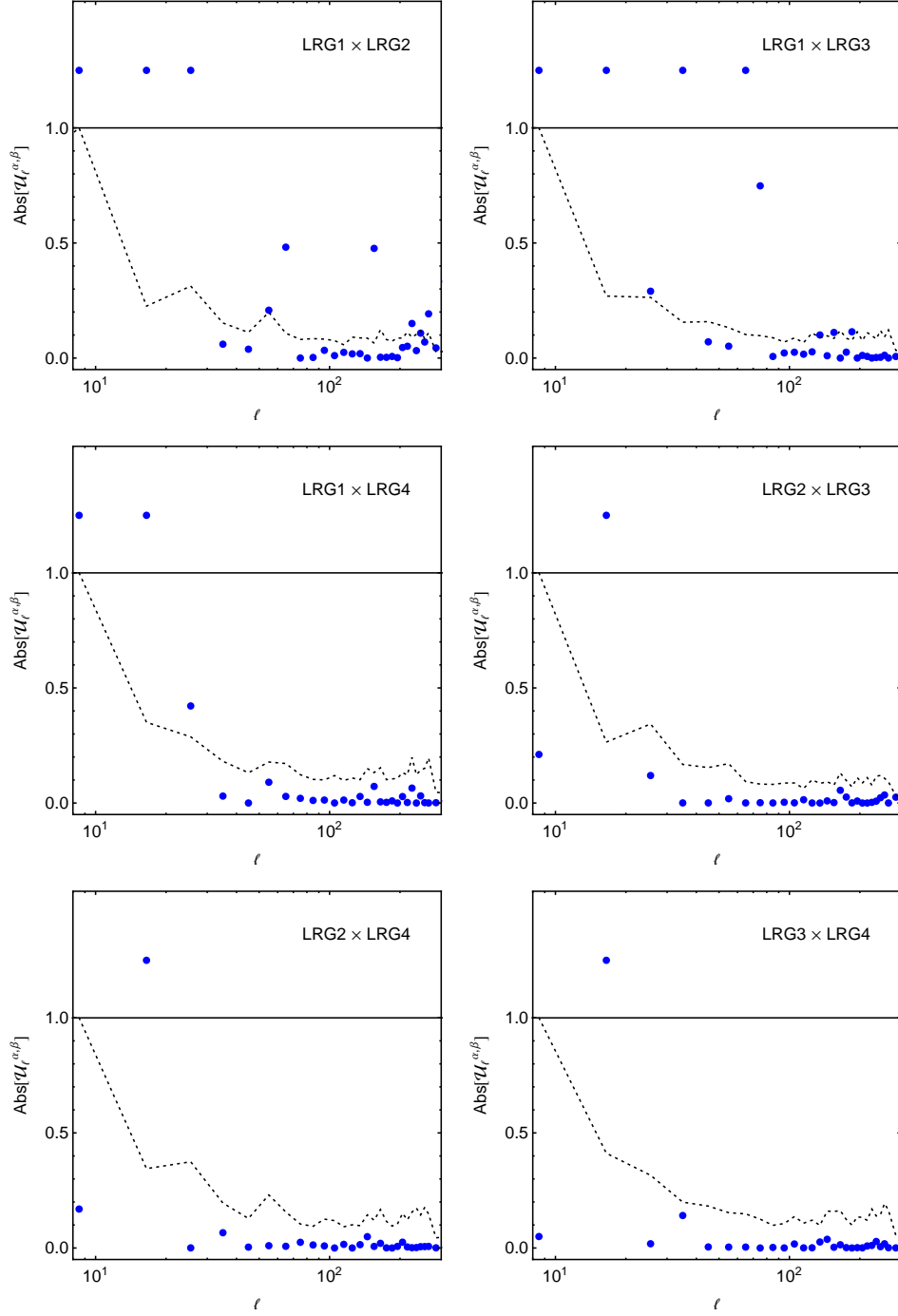
Figure 5. The absolute value of the unknown contamination coefficient $\mathcal{U}_\ell^{\alpha,\beta}$ defined in eq. (3.5) for SDSS-III DR8 LRGs (filled circles). The dotted line shows the absolute value of the 3σ-cut — we drop all bins in the corresponding auto-power spectra that lie above this cut. The points that lie above the solid line have $\text{Abs}[\mathcal{U}_\ell^{\alpha,\beta}] > 1$.
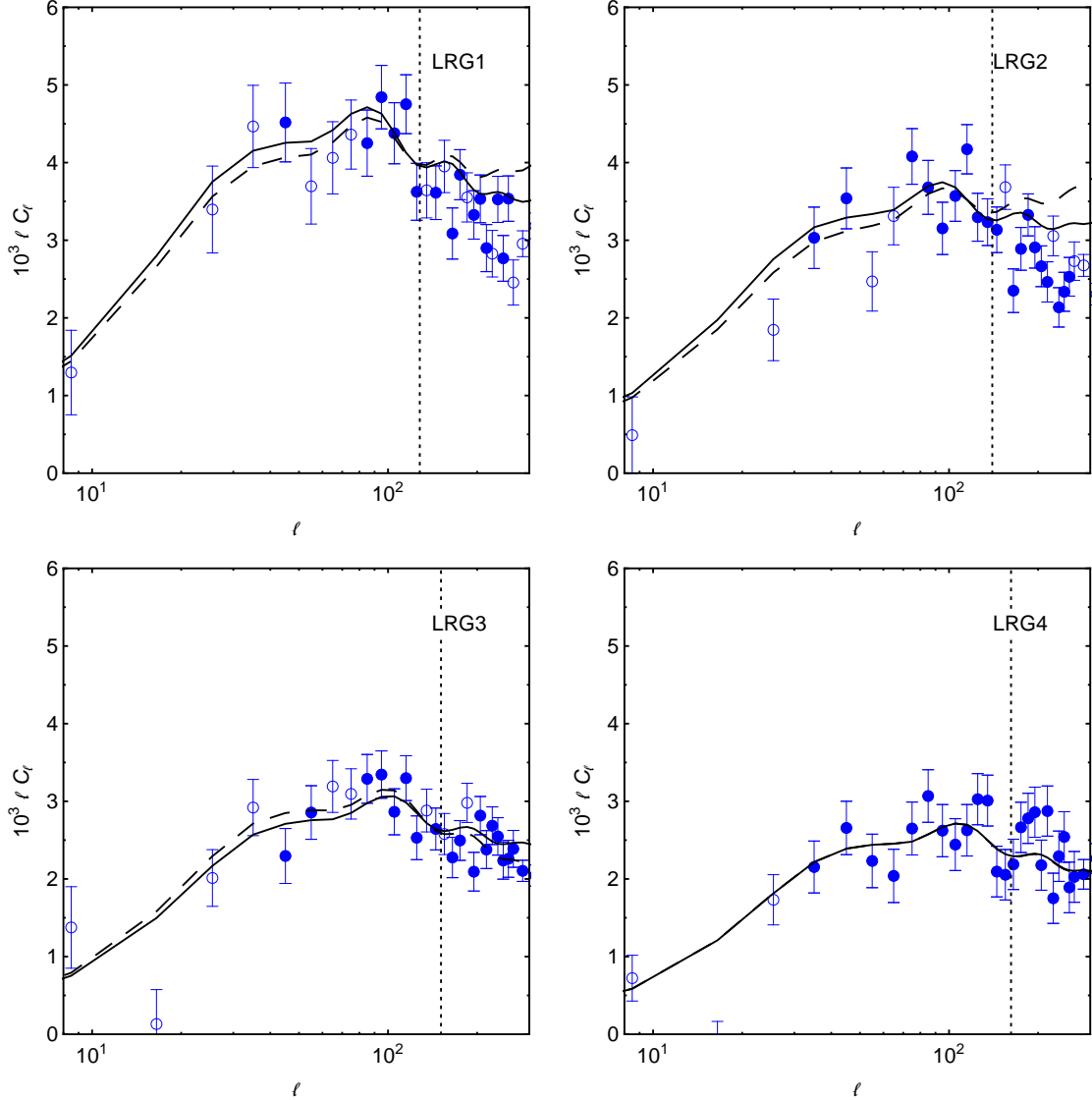
Figure 6. The angular power spectrum in the four redshift slices of LRGs. Open circles with (1σ) error bars represent data points that are dropped due to large unknown systematics, as determined using the cross-power spectra. Filled circles, on the other hand, are data points that are not dominated with unknown systematics. The vertical dotted line shows $\ell_{\max}$. The dashed curve is the theoretical power spectrum obtained using the best-fit bias parameters that correspond to the MCMC analysis which uses all data points in $30 \leq \ell \leq \ell_{\max}$. The solid curve represents the MCMC analysis which uses only filled circles in $10 \leq \ell \leq \ell_{\max}$. In both theoretical curves we have added in the corresponding best-fit non-linear fitting parameter a to all C_ℓs; this causes the upturn at large ℓ.

for various potential systematic effects and mode-projected them from the angular cross-power spectra. Although this improved the significance of the cross-correlation measurement, they concluded that other systematics are still contaminating the SDSS photometric quasar sample. In [24], the authors used improved sky masks and mode projection to further reduce contamination levels, and in [22], the authors obtained improved cosmological parameter

estimates by using cross-correlations between different surveys.

In this paper we have adopted a different approach and have developed a method to exclude bins in the angular power spectrum that are significantly contaminated by unknown systematics. For this purpose, we use cross-correlations between different redshift slices and define an unknown contamination parameter to estimate the contribution from unknown systematics in the auto-power spectra. This allows one to selectively drop bins that lie outside a specified contamination tolerance. We apply our method on simulated data for LRGs in the SDSS to demonstrate that it does improve estimates of the bias in each redshift slice. We further apply the method to real LRGs in the SDSS-III DR8 sample as a case study. This also allows a comparison, as an example of cosmological parameter estimation, of constraints on primordial non-Gaussianity before and after applying our method.

In the analysis presented in this paper, we assume a certain model for non-linear structure formation — in particular, we use the HaloFit prescription and model non-linearities with an extra a parameter. Although we cut the analysis at $k = 0.1h$ Mpc^{-1}, we expect some amount of unknown contamination to be attributed to a lack of theoretical understanding of the matter power spectrum in the non-linear regime. We would also like to emphasize that an imperfect modeling of known systematic fields, for instance a more complex relationship compared to the simple linear one we use between systematics and the observed density, is also expected to contribute to unknown systematics. Further, as discussed in section 4 the true cross-power depends on the assumed background cosmological model; one could thus implement the method presented here for each cosmological parameter space being studied.

It is also worthwhile to note that the method discussed here provides an upper estimate of the errors since it drops bins from *both* redshift slices whose cross-correlation is significantly contaminated. It may be possible to use different combinations of the auto- and cross-power spectra and additionally cross-correlations between photometric and spectroscopic measurements to derive less aggressive algorithms. Further, it would be useful to generalize such a method for spectroscopic surveys. These issues will be dealt with in future work.

Acknowledgments

Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

References

[1] C. Fedeli, L. Moscardini, and S. Matarrese, *The clustering of galaxy clusters in cosmological models with non-Gaussian initial conditions: Predictions for future surveys*, *Mon.Not.Roy.Astron.Soc.* **397** (2009) 1125–1137, [arXiv:0904.3248].

[2] C. Carbone, O. Mena, and L. Verde, *Cosmological parameters degeneracies and non-Gaussian halo bias*, *JCAP* **1007** (2010) 020, [arXiv:1003.0456].

[3] C. Cunha, D. Huterer, and O. Doré, *Primordial non-Gaussianity from the covariance of galaxy cluster counts*, *Phys.Rev.* **D82** (2010) 023004, [arXiv:1003.2416].

[4] C. Fedeli, C. Carbone, L. Moscardini, and A. Cimatti, *The clustering of galaxies and galaxy clusters: constraints on primordial non-Gaussianity from future wide-field surveys*, *Mon.Not.Roy.Astron.Soc.* **414** (2011) 1545–1559, [arXiv:1012.2305].

[5] T. Giannantonio, C. Porciani, J. Carron, A. Amara, and A. Pillepich, *Constraining primordial non-Gaussianity with future galaxy surveys*, *Mon.Not.Roy.Astron.Soc.* **422** (2012) 2854–2877, [arXiv:1109.0958].

[6] A. Pillepich, C. Porciani, and T. H. Reiprich, *The X-ray cluster survey with eROSITA: Forecasts for cosmology, cluster physics, and primordial non-Gaussianity*, *Mon.Not.Roy.Astron.Soc.* **422** (2012) 44–69, [arXiv:1111.6587].

[7] Z. Huang, L. Verde, and F. Vernizzi, *Constraining inflation with future galaxy redshift surveys*, *JCAP* **1204** (2012) 005, [arXiv:1201.5955].

[8] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, *The effective field theory of cosmological large scale structures*, *JHEP* **1209** (2012) 082, [arXiv:1206.2926].

[9] K. S. Dawson et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *Astron.J.* **145** (2013) 10, [arXiv:1208.0022].

[10] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, 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** (2012), no. 4 3435–3467, [arXiv:1203.6594].

[11] D. Schelgel et al., *The BigBOSS experiment*, arXiv:1106.1706.

[12] A. Cimatti, R. Laureijs, B. Leibundgut, S. Lilly, R. Nichol, et al., *Euclid assessment study report for the ESA Cosmic Visions*, arXiv:0912.0914.

[13] D. G. York et al., *The Sloan Digital Sky Survey: Technical summary*, *Astron.J.* **120** (2000) 1579–1587, [astro-ph/0006396].

[14] D. J. Eisenstein et al., *SDSS-III: Massive spectroscopic surveys of the distant Universe, the milky way galaxy, and extra-solar planetary systems*, *Astron.J.* **142** (2011) 72, [arXiv:1101.1529].

[15] S. Ho, A. Cuesta, H.-J. Seo, R. de Putter, A. J. Ross, et al., *Clustering of Sloan Digital Sky Survey III photometric luminous galaxies: The measurement, systematics and cosmological implications*, *Astrophys.J.* **761** (2012) 14, [arXiv:1201.2137].

[16] S. Ho, N. Agarwal, A. D. Myers, R. Lyons, A. Disbrow, et al., *Sloan Digital Sky Survey III photometric quasar clustering: Probing the initial conditions of the Universe using the largest volume*, arXiv:1311.2597.

[17] *The Dark Energy Survey*, http://www.darkenergysurvey.org/.

[18] P. A. Abell et al., *LSST Science Book, Version 2.0*, arXiv:0912.0201.

[19] A. J. Ross, S. Ho, A. J. Cuesta, R. Tojeiro, W. J. Percival, et al., *Ameliorating systematic uncertainties in the angular clustering of galaxies: A study using SDSS-III*, Mon.Not.Roy.Astron.Soc. **417** (2011) 1350, [arXiv:1105.2320].

[20] D. Huterer, C. E. Cunha, and W. Fang, *Calibration errors unleashed: Effects on cosmological parameters and requirements for large-scale structure surveys*, Mon.Not.Roy.Astron.Soc. **432** (2013) 2945–2961, [arXiv:1211.1015].

[21] A. R. Pullen and C. M. Hirata, *Systematic effects in large-scale angular power spectra of photometric quasars and implications for constraining primordial non-Gaussianity*, Publications of the Astronomical Society of the Pacific **125** (2013) 705–718, [arXiv:1212.4500].

[22] T. Giannantonio, A. J. Ross, W. J. Percival, R. Crittenden, D. Bacher, et al., *Improved primordial non-Gaussianity constraints from measurements of galaxy clustering and the integrated Sachs-Wolfe effect*, Phys.Rev. **D89** (2014) 023511, [arXiv:1303.1349].

[23] C. Hernández-Monteagudo, A. Ross, A. Cuesta, R. Génova-Santos, F. Prada, et al., *The SDSS-III Baryonic Oscillation Spectroscopic Survey: Constraints on the integrated Sachs-Wolfe effect*, arXiv:1303.4302.

[24] 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].

[25] R. Scranton et al., *Detection of cosmic magnification with the Sloan Digital Sky Survey*, Astrophys.J. **633** (2005) 589–602, [astro-ph/0504510].

[26] G. Bernstein and D. Huterer, *Catastrophic photometric redshift errors: Weak lensing survey requirements*, Mon.Not.Roy.Astron.Soc. **401** (2010) 1399, [arXiv:0902.2782].

[27] M. McQuinn and M. White, *On using angular cross-correlations to determine source redshift distributions*, Mon.Not.Roy.Astron.Soc. **433** (2013) 2857, [arXiv:1302.0857].

[28] R. de Putter, O. Doré, and S. Das, *Using cross-correlations to calibrate lensing source redshift distributions: Improving cosmological constraints from upcoming weak lensing surveys*, Astrophys.J. **780** (2014) 185, [arXiv:1306.0534].

[29] H. Aihara et al., *The eighth data release of the Sloan Digital Sky Survey: First data from SDSS-III*, Astrophys.J.Suppl. **193** (2011) 29, [arXiv:1101.1559].

[30] S. Dodelson, *Modern cosmology*, Academic Press, Elsevier Science (2003).

[31] P. J. E. Peebles and J. T. Yu, *Primeval adiabatic perturbation in an expanding universe*, Astrophys.J. **162** (1970) 815–836.

[32] R. A. Sunyaev and Y. B. Zeldovich, *The Interaction of matter and radiation in the hot model of the Universe, II*, Astrophys.Space Sci. **7** (1970) 20–30.

[33] J. R. Bond and G. Efstathiou, *Cosmic background radiation anisotropies in Universes dominated by nonbaryonic dark matter*, Astrophys.J. **285** (1984) L45–L48.

[34] J. A. Holtzman, *Microwave background anisotropies and large scale structure in Universes with cold dark matter, baryons, radiation, and massive and massless neutrinos*, Astrophys.J.Suppl. **71** (1989) 1–24.

[35] W. Hu and N. Sugiyama, *Small scale cosmological perturbations: An analytic approach*, Astrophys.J. **471** (1996) 542–570, [astro-ph/9510117].

[36] W. Hu and M. J. White, *Acoustic signatures in the cosmic microwave background*, Astrophys.J. **471** (1996) 30–51, [astro-ph/9602019].

[37] D. J. Eisenstein and W. Hu, *Baryonic features in the matter transfer function*, Astrophys.J. **496** (1998) 605, [astro-ph/9709112].

[38] D. J. Eisenstein, W. Hu, and M. Tegmark, *Cosmic complementarity: H_0 and Ω_m from combining CMB experiments and redshift surveys*, *Astrophys.J.* **504** (1998) L57–L61, [astro-ph/9805239].

[39] C. Blake and K. Glazebrook, *Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler*, *Astrophys.J.* **594** (2003) 665–673, [astro-ph/0301632].

[40] W. Hu and Z. Haiman, *Redshifting rings of power*, *Phys.Rev.* **D68** (2003) 063004, [astro-ph/0306053].

[41] E. V. Linder, *Baryon oscillations as a cosmological probe*, *Phys.Rev.* **D68** (2003) 083504, [astro-ph/0304001].

[42] T. Matsubara and A. S. Szalay, *Apparent clustering of intermediate-redshift galaxies as a probe of dark energy*, *Phys.Rev.Lett.* **90** (2003) 021302, [astro-ph/0208087].

[43] H.-J. Seo and D. J. Eisenstein, *Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys*, *Astrophys.J.* **598** (2003) 720–740, [astro-ph/0307460].

[44] T. Matsubara, *Correlation function in deep redshift space as a cosmological probe*, *Astrophys.J.* **615** (2004) 573–585, [astro-ph/0408349].

[45] M. J. White, *Baryon oscillations*, *Astropart.Phys.* **24** (2005) 334–344, [astro-ph/0507307].

[46] D. Dolney, B. Jain, and M. Takada, *Baryon oscillations and dark-energy constraints from imaging surveys*, *Mon.Not.Roy.Astron.Soc.* **366** (2006) 884–898, [astro-ph/0409445].

[47] N. Padmanabhan et al., *The clustering of luminous red galaxies in the Sloan Digital Sky Survey imaging data*, *Mon.Not.Roy.Astron.Soc.* **378** (2007) 852–872, [astro-ph/0605302].

[48] R. de Putter, O. Mena, E. Giusarma, S. Ho, A. Cuesta, et al., *New neutrino mass bounds from Sloan Digital Sky Survey III Data Release 8 photometric luminous galaxies*, *Astrophys.J.* **761** (2012) 12, [arXiv:1201.1909].

[49] D. N. Limber, *The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field. II*, *Astrophys.J.* **119** (1954) 655.

[50] U. Seljak, *Weak lensing reconstruction and power spectrum estimation: Minimum variance methods*, *Astrophys.J.* **506** (1997) 64, [astro-ph/9711124].

[51] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay, *Measuring the galaxy power spectrum with future redshift surveys*, *Astrophys.J.* **499** (1998) 555–576, [astro-ph/9708020].

[52] N. Padmanabhan, U. Seljak, and U. L. Pen, *Mining weak lensing surveys*, *New Astron.* **8** (2003) 581–603, [astro-ph/0210478].

[53] A. Lewis and S. Bridle, *Cosmological parameters from CMB and other data: A Monte-Carlo approach*, *Phys.Rev.* **D66** (2002) 103511, [astro-ph/0205436].

[54] A. Lewis, A. Challinor, and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473–476, [astro-ph/9911177].

[55] R. E. Smith et al., *Stable clustering, the halo model and nonlinear cosmological power spectra*, *Mon.Not.Roy.Astron.Soc.* **341** (2003) 1311, [astro-ph/0207664].

[56] 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.* **D77** (2008) 123514, [arXiv:0710.4560].

[57] S. Matarrese and L. Verde, *The effect of primordial non-Gaussianity on halo bias*, *Astrophys.J.* **677** (2008) L77–L80, [arXiv:0801.4826].

[58] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, *Constraints on local primordial non-Gaussianity from large scale structure*, *JCAP* **0808** (2008) 031, [arXiv:0805.3580].

[59] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological parameter results*, *Astrophys.J.Suppl.* **208** (2012) 19, [arXiv:1212.5226].

[60] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Final maps and results*, *Astrophys.J.Suppl.* **208** (2012) 20, [arXiv:1212.5225].

[61] R. Amanullah, C. Lidman, D. Rubin, G. Aldering, P. Astier, et al., *Spectra and light curves of six type Ia supernovae at $0.511 < z < 1.12$ and the Union2 compilation*, *Astrophys.J.* **716** (2010) 712–738, [arXiv:1004.1711].

[62] C. Gordon, W. Hu, D. Huterer, and T. M. Crawford, *Spontaneous isotropy breaking: A mechanism for CMB multipole alignments*, *Phys.Rev.* **D72** (2005) 103002, [astro-ph/0509301].

[63] D. Huterer, M. Takada, G. Bernstein, and B. Jain, *Systematic errors in future weak lensing surveys: Requirements and prospects for self-calibration*, *Mon.Not.Roy.Astron.Soc.* **366** (2006) 101–114, [astro-ph/0506030].

[64] L. Anderson et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations in the Data Release 10 and 11 galaxy samples*, arXiv:1312.4877.

[65] A. R. Pullen and C. M. Hirata, *Non-detection of a statistically anisotropic power spectrum in large-scale structure*, *JCAP* **1005** (2010) 027, [arXiv:1003.0673].

[66] M. White, Y.-S. Song, and W. J. Percival, *Forecasting cosmological constraints from redshift surveys*, *Mon.Not.Roy.Astron.Soc.* **397** (2008) 1348–1354, [arXiv:0810.1518].

[67] A. J. Ross, W. J. Percival, M. Crocce, A. Cabre, and E. Gaztanaga, *Measuring redshift-space distortions using photometric surveys*, *Mon.Not.Roy.Astron.Soc.* **415** (2011) 2193, [arXiv:1102.0968].

[68] J. E. Gunn et al., *The Sloan Digital Sky Survey photometric camera*, *Astron.J.* **116** (1998) 3040, [astro-ph/9809085].

[69] J. E. Gunn et al., *The 2.5 m telescope of the Sloan Digital Sky Survey*, *Astron.J.* **131** (2006) 2332–2359, [astro-ph/0602326].

[70] A. S. Bolton et al., *Spectral classification and redshift measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey*, *Astron.J.* **144** (2012) 144, [arXiv:1207.7326].

[71] S. Smee, J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, et al., *The multi-object, fiber-fed spectrographs for SDSS and the Baryon Oscillation Spectroscopic Survey*, *Astron.J.* **126** (2013) 32, [arXiv:1208.2233].

[72] M. White, M. Blanton, A. Bolton, D. Schlegel, J. Tinker, et al., *The clustering of massive galaxies at $z \sim 0.5$ from the first semester of BOSS data*, *Astrophys.J.* **728** (2011) 126, [arXiv:1010.4915].

[73] J. K. Adelman-McCarthy et al., *The sixth data release of the Sloan Digital Sky Survey*, *Astrophys.J.Suppl.* **175** (2008) 297–313, [arXiv:0707.3413].

[74] G. T. Richards, A. D. Myers, A. G. Gray, R. N. Riegel, R. C. Nichol, et al., *Efficient photometric selection of quasars from the Sloan Digital Sky Survey. II. $\sim 1,000,000$ quasars from Data Release 6*, *Astrophys.J.Suppl.* **180** (2009) 67–83, [arXiv:0809.3952].