

What is the optimal way to measure the galaxy power spectrum?

Article (Published Version)

Smith, Robert and Marian, Laura (2016) What is the optimal way to measure the galaxy power spectrum? Monthly Notices of the Royal Astronomical Society, 457 (4). pp. 4285-4290. ISSN 0035-8711

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/62241/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

What is the optimal way to measure the galaxy power spectrum?

Robert E. Smith^{1,2★} and Laura Marian^{1★}

¹*Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*

²*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str.1, Postfach 1523, D-85740 Garching, Germany*

Accepted 2016 February 1. Received 2016 February 1; in original form 2015 July 20

ABSTRACT

Measurements of the galaxy power spectrum contain a wealth of cosmological information. In Smith & Marian, we generalized the power spectrum methodology of Feldman et al. to take into account the key tenets of galaxy formation: galaxies form and reside exclusively in dark matter haloes; a given dark matter halo may host galaxies of various luminosities; galaxies inherit the large-scale bias of their host halo. In this paradigm, we derived the optimal weighting scheme for maximizing the signal-to-noise (S/N) on a given band power estimate. For a future all-sky flux-limited galaxy redshift survey of depth $b_J > 22$, we demonstrate that the optimal weighting scheme does indeed provide improved S/N at the level of ~ 20 per cent when compared to Feldman et al. and ~ 60 per cent relative to Percival et al., for scales of the order of $k \sim 0.5 h \text{ Mpc}^{-1}$. Using a Fisher matrix approach, we show the cosmological information yield is also increased relative to these alternate methods – especially the primordial power spectrum amplitude and dark energy equation of state. Caveats: uncertainties in cluster masses, non-linear halo bias and redshift distortions may reduce information gains.

Key words: large-scale structure of Universe.

1 INTRODUCTION

The matter power spectrum is a fundamental tool for constraining the cosmological parameters. It contains detailed information about the large-scale geometrical structure of space-time, as well as the phenomenological properties of dark energy and dark matter. Given a galaxy redshift survey two things are crucial: how to obtain an unbiased and optimal estimate of the information in the matter fluctuations.

State-of-the-art galaxy redshift surveys, such as the Baryon Oscillation Spectroscopic Survey (Anderson et al. 2012, 2014b,a, hereafter BOSS), Galaxy And Mass Assembly (Blake et al. 2013, hereafter GAMA), and WiggleZ (Blake et al. 2011), have all used the approach of Feldman, Kaiser & Peacock (1994, hereafter FKP) to estimate the power spectrum. This assumes that galaxies are a Poisson sampling of the underlying density field. Hence, provided one subtracts an appropriate shot-noise term, and deconvolves for the survey window function, one should obtain an unbiased estimate of the matter power spectrum.

In the last two decades, our understanding of galaxy formation has made rapid progress since the work of FKP and our current best models strongly suggest that galaxies are not related to matter in the way they envisioned (White & Rees 1978; White & Frenk 1991; Kauffmann et al. 1999; Benson et al. 2000; Springel et al. 2005). Furthermore, observational studies have discovered that galaxy clus-

tering depends on various physical properties: e.g. luminosity (Park et al. 1994; Norberg et al. 2001, 2002; Zehavi et al. 2002a, 2005; Swanson et al. 2008; Zehavi et al. 2011a), colour (Brown, Webster & Boyle 2000; Zehavi et al. 2002b, 2005; Swanson et al. 2008; Zehavi et al. 2011b), morphology (Davis & Geller 1976; Guzzo et al. 1997; Norberg et al. 2002), and stellar mass (Li et al. 2006) etc.

Percival, Verde & Peacock (2004, hereafter PVP) attempted to correct the FKP framework to take into account the effects of luminosity-dependent clustering. In a recent paper (Smith & Marian 2015, hereafter SM15), we argued that the approach of PVP, whilst appearing qualitatively reasonable, is in fact at odds with our current understanding of galaxy formation, and so non-optimal. More recent studies by Seljak, Hamaus & Desjacques (2009), Hamaus et al. (2010) and Cai, Bernstein & Sheth (2011) suggested that weighting the galaxy density field by a linear function of halo mass would reduce stochasticity.

In SM15, we developed a new scheme incorporating a number of the key ideas from galaxy formation: galaxies only form in dark matter haloes (White & Rees 1978); haloes can host galaxies of various luminosities; the large-scale bias associated with a given galaxy is largely inherited from the bias of the host dark matter halo.

In this work we demonstrate that our new optimal estimator indeed provides both improved signal-to-noise (hereafter S/N) estimates of the galaxy power spectrum and boosted cosmological information content, when compared with the FKP and PVP approaches.

*E-mail: r.e.smith@sussex.ac.uk (RES); l.marian@sussex.ac.uk (LM)

This letter is broken down as follows: First, in Section 2, we provide a brief overview of the results from SM15. In Section 3, we evaluate the \mathcal{S}/\mathcal{N} expressions for the various weighting schemes and in Section 4 we assess the cosmological information content of the weighted power spectra measured from a putative all-sky galaxy redshift survey. Finally, in Section 5, we conclude.

2 OPTIMAL POWER SPECTRUM ESTIMATION

In the original work of FKP, the starting concept is that galaxies are simply an independent point sampling of the underlying galaxy field. Hence,

$$n_g(\mathbf{r}) = \sum_{i=1}^{N_g} \delta^D(\mathbf{r} - \mathbf{r}_i), \quad (1)$$

where N_g is the number of galaxies, and \mathbf{r}_i is the position of the i th galaxy in the survey. From this field one then may construct an effective galaxy over-density field:

$$\mathcal{F}_g^{\text{FKP}}(\mathbf{r}) = \Theta(\mathbf{r})w(\mathbf{r})[n_g(\mathbf{r}) - \alpha n_s(\mathbf{r})], \quad (2)$$

where $\Theta(\mathbf{r})$ is a survey mask function, which is 1 if the galaxy lies inside the survey volume and 0 otherwise, α is a scaling factor for the spatially random galaxy field $n_s(\mathbf{r})$ and $w(\mathbf{r})$ is an optimal weight function that depends on \mathbf{r} . If we now follow the FKP logic and compute the power spectrum of the $\mathcal{F}_g^{\text{FKP}}$ field, one finds that it is related to the galaxy power spectrum $P_g(k)$ through the relation:

$$\left\langle \left| \mathcal{F}_g^{\text{FKP}}(\mathbf{k}) \right|^2 \right\rangle = \int \frac{d^3k'}{(2\pi)^3} P_g(k') |\tilde{\mathcal{G}}^{\text{FKP}}(\mathbf{k} - \mathbf{k}')|^2 + P_{\text{shot}}^{\text{FKP}}, \quad (3)$$

where $\tilde{\mathcal{G}}^{\text{FKP}}(\mathbf{k})$ is the weighted version of the Fourier transform of the survey mask function $\Theta(\mathbf{r})$, and $P_{\text{shot}}^{\text{FKP}}$ is an effective shot-noise correction. If one subtracts the shot noise and deconvolves for the survey window function, then one may obtain an estimate of $P_g(k)$. It is important to realize that the above procedure is only correct under the assumption that the galaxy power spectrum does not depend on any observable, e.g. galaxy luminosity, colour, spectral type, host halo mass, etc. and that shot noise is as was given by FKP. If these assumptions are wrong, then the functions $P_g(k)$, $\tilde{\mathcal{G}}^{\text{FKP}}(\mathbf{k})$, and $P_{\text{shot}}^{\text{FKP}}$ will all pick up these dependences, resulting in a biased and sub-optimal reconstruction of the ‘true’ power spectrum.¹ As noted earlier, current observational evidence indicates that clustering strength does depend on the sample selection. Hence, FKP must be biased and sub-optimal (PVP also came to a similar conclusions).

We now summarize the SM15 formalism, designed to account for a number of these effects. Consider a large survey volume containing N_g galaxies that are constrained to be distributed inside N_h dark matter haloes. Thus the i th dark matter halo of mass M_i and position of the centre of mass \mathbf{x}_i , will have $N_g(M_i)$ galaxies. The j th galaxy will have a position vector \mathbf{r}_j relative to the centre of the halo and a luminosity L_j . For this more complicated distribution, equation (1) can be generalized to:

$$n_g(\mathbf{r}, L, \mathbf{x}, M) = \sum_{i=1}^{N_h} \delta^D(\mathbf{x} - \mathbf{x}_i) \delta^D(M - M_i) \times \sum_{j=1}^{N_g(M_i)} \delta^D(\mathbf{r} - \mathbf{r}_j - \mathbf{x}_i) \delta^D(L - L_j), \quad (4)$$

where the four Dirac delta functions, going from right to left, are sampling: the luminosity of each galaxy in a given halo; the spatial location of a given galaxy relative to the halo centre; the halo mass from the distribution of masses; and the halo centre in the survey volume.

In direct analogy with FKP’s equation (2), we define an effective galaxy over-density field:

$$\mathcal{F}_g(\mathbf{r}) = \int dL \int d^3x \int dM \Theta(\mathbf{r}|L) \frac{w(\mathbf{r}, L, \mathbf{x}, M)}{\sqrt{A}} \times [n_g(\mathbf{r}, L, \mathbf{x}, M) - \alpha n_s(\mathbf{r}, L, \mathbf{x}, M)], \quad (5)$$

where $\Theta(\mathbf{r}|L)$ is the luminosity-dependent survey geometry function; A is a normalization constant; α is a scaling factor for the random halo catalogue; w is a general weight function. The function n_s is the same as n_g , except the spatial locations of the halo centres have been randomized and the number density has been scaled up by a factor of $1/\alpha$. As shown by SM15, in the large-scale limit, when the distribution of galaxies in haloes adopts a Dirac delta function like behaviour, the \mathcal{F}_g power spectrum can be written:

$$\left\langle \left| \tilde{\mathcal{F}}_g(\mathbf{k}) \right|^2 \right\rangle \approx \int \frac{d^3q}{(2\pi)^3} P(q) \left| \tilde{\mathcal{G}}_{(1,1)}^{(1)}(\mathbf{k} - \mathbf{q}) \right|^2 + P_{\text{shot}}, \quad (6)$$

where $P(q)$ is the true matter power spectrum, which is convolved with the effective survey window function $\tilde{\mathcal{G}}_{(1,1)}^{(1)}(\mathbf{k})$, and P_{shot} is a new effective shot noise. Note, in arriving at the above result, whilst we have assumed a linear bias model, the matter power spectrum is the fully non-linear spectrum. The set of survey window functions that are required to evaluate these expressions can be written in general:

$$\mathcal{G}_{(l,m)}^{(n)}(\mathbf{r}) \equiv A^{-nl/2} \int dM \bar{n}(M, \chi) b^m(M, \chi) N_g^{(n)}(M) \times \left[\int dL \Theta(\mathbf{r}|L) \Phi(L|M) w^l(\mathbf{r}, L, \mathbf{x}, M) \right]^n, \quad (7)$$

where in the above $\bar{n}(M, \chi)$ and $b(M, \chi)$ are the mass function and large-scale linear bias of haloes of mass M at radial position χ from the observer (χ here is also acting as coordinate time); $N_g^{(n)}(M)$ gives the n th factorial moment of the halo occupation distribution (hereafter HOD); $\Phi(L|M)$ gives the conditional probability density that a galaxy hosted in a halo of mass M has a luminosity L . Using these functions, the effective shot noise term can be written:

$$P_{\text{shot}} \equiv (1 + \alpha) \left[\int \frac{d^3q}{(2\pi)^3} \tilde{\mathcal{G}}_{(1,0)}^{(2)}(\mathbf{q}) + \tilde{\mathcal{G}}_{(2,0)}^{(1)}(\mathbf{0}) \right]. \quad (8)$$

We also introduce the normalization-free window functions $\bar{\mathcal{G}}_{(l,m)}^{(n)} = A^{nl/2} \mathcal{G}_{(l,m)}^{(n)}$, which enables us to write: $A \equiv \int d^3r |\bar{\mathcal{G}}_{(1,1)}^{(1)}(\mathbf{r})|^2$. We thus see that, similar to the FKP approach, in order to recover the matter power spectrum, one must subtract the effective shot-noise term and deconvolve for the square of the effective survey window function $\tilde{\mathcal{G}}_{(1,1)}^{(1)}(\mathbf{k})$.

In the large-scale limit and under the assumption that the matter density field is Gaussian distributed, SM15 also showed that the

¹ This bias is lessened if the survey is volume limited, but the bias persists in the form of a modification to the shot noise.

S/\mathcal{N} can, for an arbitrary weight w , be written in general as

$$\begin{aligned} \left(\frac{S}{\mathcal{N}}\right)^2 &= \frac{V(k)}{2(2\pi)^3} \left[\int d^3r \left| \bar{\mathcal{G}}_{(1,1)}^{(1)}(\mathbf{r}) \right|^2 \right]^2 \\ &\times \left\{ \int d^3r \left(\left[\bar{\mathcal{G}}_{(1,1)}^{(1)}(\mathbf{r}) \right]^2 \right. \right. \\ &\left. \left. + \frac{(1+\alpha)}{P(k)} \left[\bar{\mathcal{G}}_{(1,0)}^{(2)}(\mathbf{r}) + \bar{\mathcal{G}}_{(2,0)}^{(1)}(\mathbf{r}) \right] \right)^2 \right\}^{-1}, \end{aligned} \quad (9)$$

where $V(k) = 4\pi k_i^2 \Delta k [1 + (\Delta k/k_i)^2/12]$ is the volume of the i th k -space shell in which $P(k)$ is estimated.

3 COMPARISON OF WEIGHTING SCHEMES

The failure of the **FKP** scheme to characterize the true clustering strengths of galaxies means that it is a biased and sub-optimal estimator. We will now show explicitly, under the assumption that the **SM15** description of the galaxy population is the correct one, that both the **FKP** and **PVP** weighting schemes do indeed lead to sub-optimal measurements of $P(k)$. The weighting schemes are:

(i) The **FKP** weights: these depend only on the position of the galaxy in the survey:

$$w_{\text{FKP}}(\mathbf{r}) = 1 / [1 + \bar{n}_{\text{g}}(\mathbf{r})P(k)], \quad (10)$$

where $\bar{n}_{\text{g}}(\mathbf{r})$ is the mean number density of galaxies.

(ii) The **PVP** weights: these depend explicitly on the luminosity dependence of the galaxy bias and also the position in the survey:

$$w_{\text{PVP}}(\mathbf{r}, L) = b(L) / [1 + \bar{n}_{\text{g}}(\mathbf{r})\bar{b}_L^2(\mathbf{r})P(k)], \quad (11)$$

where the luminosity-dependent galaxy bias is $b(L) \equiv \int dM \bar{n}(M)b(M)N_{\text{g}}^{(1)}(M)\Phi(L|M)/\Phi(L)$. $\bar{b}_L^2(\mathbf{r}) \equiv \int_{L_{\text{min}}(\mathbf{r})} dL b^2(L)\Phi(L)/\bar{n}_{\text{g}}(\mathbf{r})$, is the average square of the luminosity bias. The galaxy luminosity function is given by $\Phi(L) \equiv \int dM \bar{n}(M)N_{\text{g}}^{(1)}(M)\Phi(L|M)$, and $N_{\text{g}}^{(1)}(M)$ was introduced after equation (7).

(iii) Optimal weights: in the large-scale limit, these weights depend only on the galaxy's spatial position and its host halo mass, and not explicitly on its luminosity. The weights are

$$w_{\text{OPT}}(\mathbf{r}, M) = b(M) / [1 + R(M)\mathcal{S}(\mathbf{r}, M)][1 + \bar{n}_{\text{eff}}(\mathbf{r})P(k)], \quad (12)$$

where $R(M) \equiv N_{\text{g}}^{(2)}(M)/N_{\text{g}}^{(1)}(M)$ is the ratio of the second and first factorial moments of the halo occupation distribution and we introduced the effective number density of galaxies: $\bar{n}_{\text{eff}}(\mathbf{r}) \equiv \int dM \bar{n}(M)b^2(M)N_{\text{g}}^{(1)}(M)\mathcal{S}(\mathbf{r}, M) / [1 + R(M)\mathcal{S}(\mathbf{r}, M)]$. We defined $\mathcal{S}(\mathbf{r}, M) \equiv \int_{L_{\text{min}}(\mathbf{r})}^{\infty} dL \Phi(L|M)$ as the fraction of galaxies hosted by haloes of mass M that are observable at a spatial position \mathbf{r} , with $L_{\text{min}}(\mathbf{r})$ the minimum luminosity that a galaxy could have and still be observable given the survey flux-limit. Explicitly, $L_{\text{min}}(\mathbf{r}) = 10^{-\frac{2}{5}(m_{\text{lim}} - 25 - M_{\odot})} h^{-2} [d_L(\mathbf{r})/1 h^{-1} \text{Mpc}]^{-2} [L_{\odot}]$, where m_{lim} is the apparent magnitude limit of the survey, M_{\odot} is the absolute magnitude of the sun, h is the dimensionless Hubble parameter and $d_L(\mathbf{r}) = (1+z)\chi(z)$ is the luminosity distance in flat cosmological models. Note that $\mathcal{S}(\mathbf{0}, M) = 1$ and $\mathcal{S}(\infty, M) = 0$. For more details on the S/\mathcal{N} expressions for the three weights considered, we refer the interested reader to **SM15**.

We now show the S/\mathcal{N} on the galaxy power spectrum corresponding to the **FKP**, **PVP** and **SM15** methods for weighting the

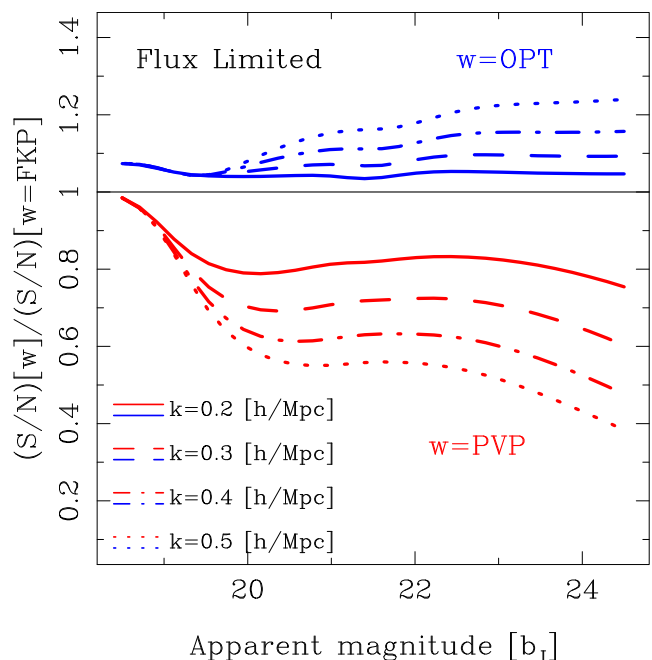


Figure 1. S/\mathcal{N} level for the optimal (blue lines, ratio >1) and the **PVP** (red lines, ratio <1) weighting schemes relative to the **FKP** one, as a function of survey b_j flux-limit. The solid-dashed, dot-dashed and dotted line styles denote results for $k \in \{0.2, 0.3, 0.4, 0.5\} h \text{ Mpc}^{-1}$, respectively. The optimal scheme of **SM15** clearly maximizes the S/\mathcal{N} ratio for the future survey cases considered.

galaxy distribution. As a concrete example we consider a flux-limited, full-sky galaxy redshift survey spanning the redshift range $z = 0.3-0.9$. In order to evaluate the above expressions, we need to specify several model ingredients. For the evolution of $\bar{n}(M)$ and $b(M)$, we use the models of Sheth & Tormen (1999). For the conditional probability distribution $\Phi(L|M)$ and the first factorial moment of the HOD $N_{\text{g}}^{(1)}(M)$, we use the Conditional Luminosity Function model of Yang, Mo & van den Bosch (2003). For the second factorial moment, we use the model: $N_{\text{g}}^{(2)}(M) = \beta(M)[N_{\text{g}}^{(1)}(M)]^2$, where from fitting to semi-analytic models of galaxy formation $\beta^{1/2}(M) = 1/2 \log_{10}(M/10^{11} h^{-1} M_{\odot})$ for the case that $M < 10^{13} h^{-1} M_{\odot}$ and unity otherwise (Cooray & Sheth 2002). From these ingredients, all required variables may be computed.

Fig. 1 shows the S/\mathcal{N} for the **SM15** (blue lines) and **PVP** (red lines) schemes ratioed with the S/\mathcal{N} for the **FKP** scheme, respectively. The results are presented as a function of limiting b_j magnitude and for various k -mode bins. Clearly, the optimal scheme of **SM15** does indeed lead to the largest S/\mathcal{N} : $\gtrsim 5$ per cent improvement over **FKP** at $k \sim 0.2 h \text{ Mpc}^{-1}$, and $\gtrsim 20$ per cent improvement at $k \sim 0.5 h \text{ Mpc}^{-1}$ for surveys with depth $b_j \gtrsim 22$. Interestingly, the scheme of **PVP** leads to the least optimal set of estimates, being ~ 20 per cent lower than **FKP** at $k = 0.2 h \text{ Mpc}^{-1}$ and ~ 40 per cent lower by $k = 0.5 h \text{ Mpc}^{-1}$, again for surveys with $b_j \gtrsim 22$.

4 FORECASTING COSMOLOGICAL INFORMATION

The ability of a set of power spectrum band-power estimates to constrain the cosmological parameters θ_{α} , can be forecasted through construction of the Fisher information matrix (Tegmark, Taylor &

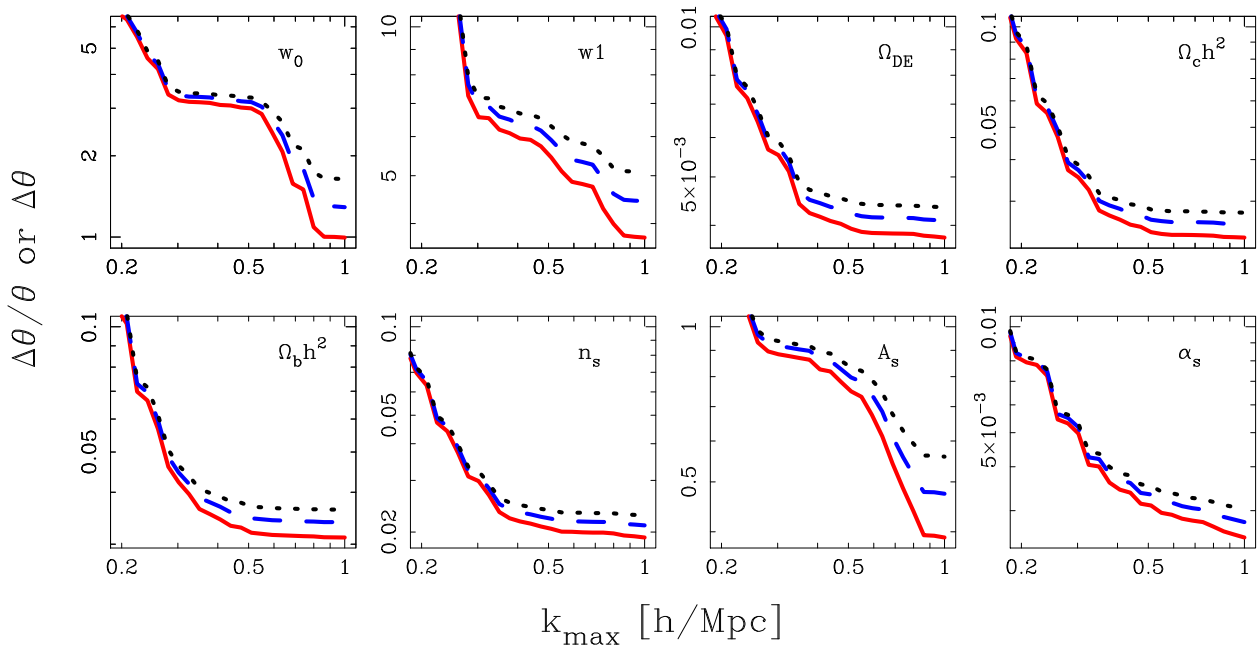


Figure 2. Forecasted 1D marginalized errors and relative errors on cosmological parameters as a function of the maximum wavenumber considered in the power spectrum estimates from a full-sky galaxy clustering survey of depth $b_J \sim 22$. The solid red, dashed blue and black dotted lines represent the **SM15**, **FKP** and **PVP** weighting schemes, respectively. The panels, going clockwise from the top left show the results for the eight cosmological parameters considered. The largest potential information gains to be had from optimal weighting are in the measurements of $\{w_0, w_a, A_s\}$. Note that we have not properly taken into account the growth evolution of structure, and used power spectrum derivatives suitable for only a single redshift.

Heavens 1997). For a continuum limit of Fourier modes the Fisher matrix can be expressed as (Tegmark 1997):

$$\mathcal{F}_{\alpha\beta} = \int \frac{d^3k}{V(k)} \frac{\partial \log P(k)}{\partial \theta_\alpha} \frac{\partial \log P(k)}{\partial \theta_\beta} \left(\frac{\mathcal{S}}{\mathcal{N}} \right)^2(k). \quad (13)$$

Thus, in order to compute the Fisher matrix, one needs to specify the \mathcal{S}/\mathcal{N} , and the derivatives of the power spectra with respect to the cosmological parameters. The former were computed in the previous section, and we estimate the latter at a single redshift. Therefore our forecasts will be pessimistic, since we do not fully take into account the information in the growth of structure, but here we are only interested in the relative differences between the three weighting schemes.

For our fiducial model, we adopt a flat, dark-energy dominated cosmological model, characterized by eight parameters: $\theta_\alpha \in \{w_0, w_1, \Omega_{\text{DE}}, \Omega_c h^2, \Omega_b h^2, A_s, n_s, \alpha_s\}$. The first two characterize the equation of state for dark energy: $w(a) = p_w/\rho_w = w_0 + (1-a)w_1$; Ω_{DE} is the dark energy density parameter; $\Omega_c h^2$ and $\Omega_b h^2$ are the physical densities in cold dark matter and baryons, respectively; and A_s, n_s , and α_s denote the amplitude, spectral index, and running of the primordial scalar power spectrum, respectively. We adopt the values $\theta_\alpha = \{-1, 0, 0.69, 0.12, 0.02, 2.15 \times 10^{-9}, 0.96, 0\}$, consistent with Planck data (Planck Collaboration XVI 2014). The power spectrum derivatives we compute through finite differencing matter power spectra from CAMB (Lewis, Challinor & Lasenby 2000).

Fig. 2 shows the forecasted 1D marginalized errors on the parameters, as a function of the maximum wavenumber k_{max} entering the integral of equation (13). The panels show the fractional error, or if the fiducial value is zero, the error. Clearly, the smallest errors are obtained when one implements the optimal weighting scheme of **SM15** (red solid lines), followed by **FKP** (blue dashed line) and

then **PVP** (black dotted lines). We notice that the constraints on (A_s, w_0, w_1) show the most significant improvements from the optimal weighting.

Fig. 3 shows the forecasted 2D marginalized errors on various parameter combinations. The line styles are the same as in Fig. 2. Again, the optimal weighting of **SM15** performs best and the parameters (A_s, w_0, w_1) appear to be the most affected by the new scheme.

5 DISCUSSION AND CONCLUSIONS

In this paper, we presented an overview of the optimal power spectrum estimation scheme of **SM15**. We argued that the **FKP** scheme was biased and sub-optimal since it does not take into account variations of clustering with the galaxy sample. We argued that the **SM15** framework, which encodes several key concepts from the theory of galaxy formation, is able to describe these variations. We evaluated the \mathcal{S}/\mathcal{N} resulting from the **FKP**, **PVP**, and **SM15** weighting schemes for the case of an all-sky galaxy survey. The **SM15** weighting scheme was indeed found to be the most efficient estimator. We then turned to the issue of cosmological information and using the Fisher matrix approach showed that the **SM15** scheme also produced the smallest errors on cosmological parameters. In particular, the parameters governing the amplitude of the primordial power spectrum and the evolution of the dark energy equation of state were noticeably improved.

It is interesting to note that studies that have looked into minimizing the stochasticity in the halo density field (Seljak et al. 2009; Hamaus et al. 2010), or reconstructing the mass density field (Cai et al. 2011), have demonstrated that optimal results can be achieved through weighting galaxy groups by some linear function of halo mass. In contrast, others have argued that better results can be

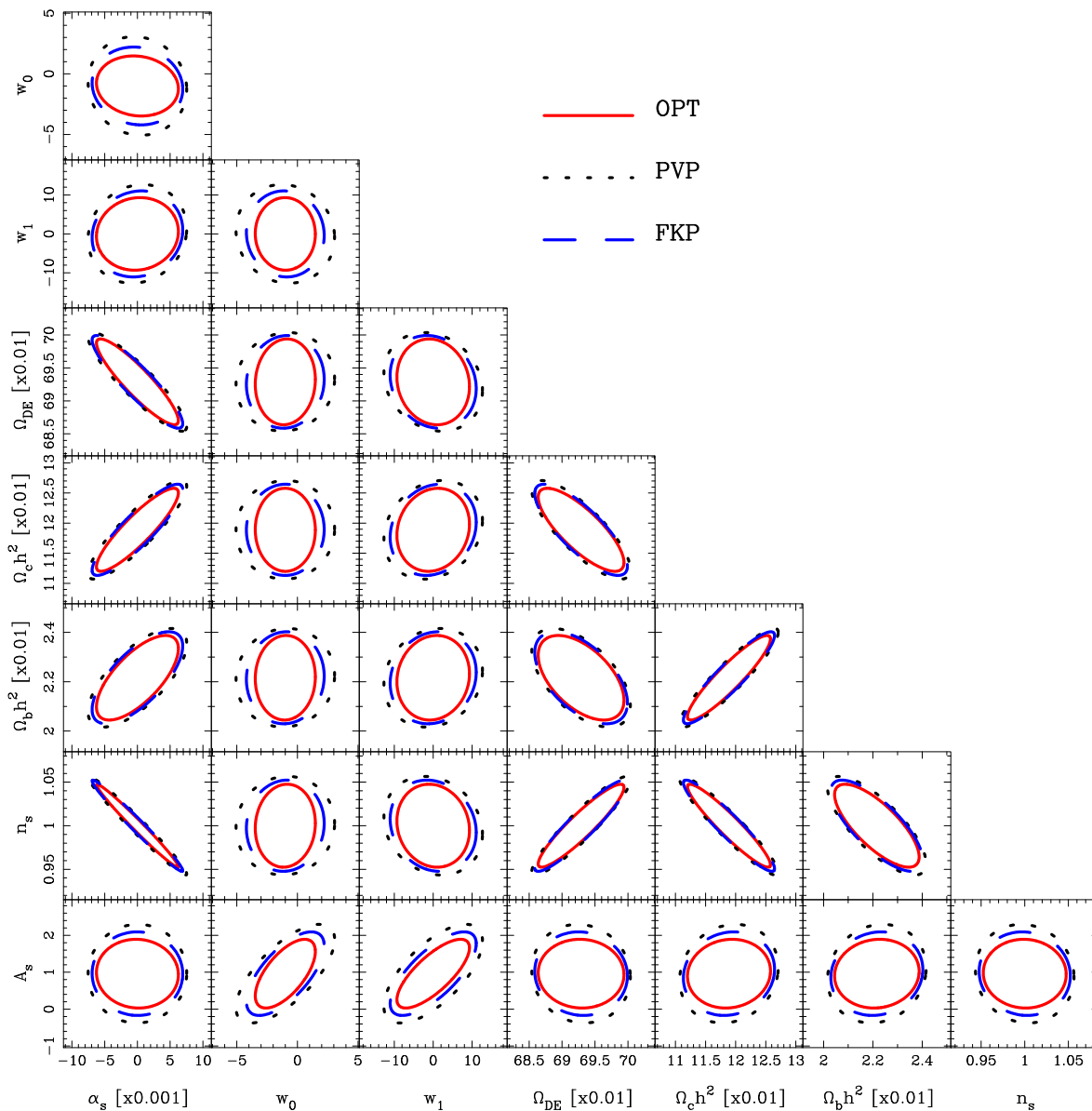


Figure 3. Forecasted 2D marginalized errors on cosmological parameters for the eight cosmological parameters considered. The maximum wavenumber was set to $k = 0.5 h \text{ Mpc}^{-1}$ and the flux-limit was taken to be $b_l = 22$. Note that we have not properly taken into account the growth evolution of structure, and have used only power spectrum derivatives suitable for a single redshift. Nevertheless, it can be clearly seen that the optimal weighting scheme provides the tightest constraints on parameters.

achieved by ‘clipping’ the density field (Simpson et al. 2011). They argue that saturation of the dense regions enables tree-level calculations to be pushed to smaller scales, thus, gaining information. In the limit of large numbers of galaxies per halo, the mass dependence of the SM15 weights is $w \propto b(M)/N_g^{(1)}(M)$ per galaxy, hence, dense regions are weighted by $b(M)$, and in the limit of small numbers we also have $w \propto b(M)$. Clearly, the SM15 weighting scheme does not follow the linear functional form of the optimal density field schemes nor the clipping. In any case, it is not obvious that the weighting schemes should be similar. We would simply argue that if one wishes to optimize galaxy power spectra measurements, then one should follow the SM15 scheme. We also point out that uncertainties in the estimation of cluster masses may likely reduce some of the potential gains. However, for SM15, since the weights depend on $b(M)$ which is a slower function of mass than M alone (at least

over the scales $0.1 < M/M_* < 10$). A scatter in M , therefore does not move the weight as much as one might expect, since as was shown in Seljak & Warren (2004), $b(M) \sim \text{const}$ for $M/M_* < 0.1$.

In this study, we have incorporated a number of key ideas from galaxy formation theory, however there are other important effects that we have not yet taken into account. In particular, we have neglected redshift space distortions. In the FKP formalism, this was examined by Yamamoto (2003), who found that the mathematical structure of the optimal weights was unchanged, except for the fact that the number density and power spectrum that enter are now both functions of the galaxies position in redshift space. In addition, the more pernicious non-linear distortions (see e.g. Okumura, Seljak & Desjacques 2012), may be minimized through use of optimal Finger-of-God compression algorithms (Reid, Spergel & Bode 2009). Based on this, one might conjecture that the optimal weights

of *SM15* may not be too significantly modified in redshift space and so any changes to our results would be relatively small. We intend to explore this in a future work.

We have also made the assumption that halo biasing is linear – but note that this does not mean that the galaxy biasing is linear in our approach, since there is still the halo internal structure. A more advanced estimator is needed to take into account non-linearity and non-locality in the halo bias relation (Smith, Scoccimarro & Sheth 2007; Baldauf et al. 2012; Chan, Scoccimarro & Sheth 2012). However, it is necessary that any improved estimator must converge to the *SM15* estimator on large scales. Owing to the fact that linear halo biasing is only valid on scales $k < 0.1 h \text{Mpc}^{-1}$ (Smith et al. 2007; Angulo et al. 2008), it may be that our weighting is sub-optimal on smaller scales and the information gains that we find are more modest.

Another effect that we have so far neglected is halo-exclusion. As was pointed out in Smith et al. (2007) and later by Smith, Desjacques & Marian (2011) and Baldauf et al. (2013), this has the effect of making shot-noise terms sub-Poissonian on large scales. Again, this may need to be incorporated in the estimation procedure. However, we might expect this to be a second-order correction. Incorporating these effects will be the subject of future study.

ACKNOWLEDGEMENTS

We thank the anonymous referee for useful discussion.

REFERENCES

- Anderson L. et al., 2012, *MNRAS*, 427, 3435
 Anderson L. et al., 2014a, *MNRAS*, 439, 83
 Anderson L. et al., 2014b, *MNRAS*, 441, 24
 Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, *MNRAS*, 383, 755
 Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, *Phys. Rev. D*, 86, 083540
 Baldauf T., Seljak U., Smith R. E., Hamaus N., Desjacques V., 2013, *Phys. Rev. D*, 88, 083507
 Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, *MNRAS*, 311, 793
 Blake C. et al., 2011, *MNRAS*, 415, 2876
 Blake C. et al., 2013, *MNRAS*, 436, 3089
 Brown M. J. I., Webster R. L., Boyle B. J., 2000, *MNRAS*, 317, 782
 Cai Y.-C., Bernstein G., Sheth R. K., 2011, *MNRAS*, 412, 995
 Chan K. C., Scoccimarro R., Sheth R. K., 2012, *Phys. Rev. D*, 85, 083509
 Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
 Davis M., Geller M. J., 1976, *ApJ*, 208, 13
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23 (FKP)
 Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, *ApJ*, 489, 37
 Hamaus N., Seljak U., Desjacques V., Smith R. E., Baldauf T., 2010, *Phys. Rev. D*, 82, 043515
 Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, *MNRAS*, 303, 188
 Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
 Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, *MNRAS*, 368, 21
 Norberg P. et al., 2001, *MNRAS*, 328, 64
 Norberg P. et al., 2002, *MNRAS*, 332, 827
 Okumura T., Seljak U., Desjacques V., 2012, *J. Cosmol. Astropart. Phys.*, 11, 014
 Park C., Vogeley M. S., Geller M. J., Huchra J. P., 1994, *ApJ*, 431, 569
 Percival W. J., Verde L., Peacock J. A., 2004, *MNRAS*, 347, 645 (PVP)
 Planck Collaboration XVI, 2014, *A&A*, 571, A16
 Reid B. A., Spergel D. N., Bode P., 2009, *ApJ*, 702, 249
 Seljak U., Warren M. S., 2004, *MNRAS*, 355, 129
 Seljak U., Hamaus N., Desjacques V., 2009, *Phys. Rev. Lett.*, 103, 091303
 Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
 Simpson F., James J. B., Heavens A. F., Heymans C., 2011, *Phys. Rev. Lett.*, 107, 271301
 Smith R. E., Marian L., 2015, *MNRAS*, 454, 1266 (SM15)
 Smith R. E., Scoccimarro R., Sheth R. K., 2007, *Phys. Rev. D*, 75, 063512
 Smith R. E., Desjacques V., Marian L., 2011, *Phys. Rev. D*, 83, 043526
 Springel V. et al., 2005, *Nature*, 435, 629
 Swanson M. E. C., Tegmark M., Blanton M., Zehavi I., 2008, *MNRAS*, 385, 1635
 Tegmark M., 1997, *Phys. Rev. Lett.*, 79, 3806
 Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22
 White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52
 White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
 Yamamoto K., 2003, *ApJ*, 595, 577
 Yang X., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
 Zehavi I. et al., 2002a, *ApJ*, 571, 172
 SDSS Collaboration Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Mo H. J., Strauss M. A., 2002b, *ApJ*, 571, 172
 Zehavi I. et al., 2005, *ApJ*, 630, 1
 Zehavi I. et al., 2011a, *ApJ*, 736, 59
 SDSS Collaboration Zehavi I., Zheng Z., Weinberg D. H., Blanton M. R., 2011b, *ApJ*, 736, 59

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.