

Towards optimal cluster power spectrum analysis

Robert E. Smith^{1,2★} and Laura Marian^{1,3★}

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

²Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK

³Argelander-Institute for Astronomy, Auf dem Hügel 71, D-53121 Bonn, Germany

Accepted 2016 January 6. Received 2016 January 3; in original form 2014 September 3

ABSTRACT

The power spectrum of galaxy clusters is an important probe of the cosmological model. In this paper, we develop a formalism to compute the optimal weights for the estimation of the matter power spectrum from cluster power spectrum measurements. We find a closed-form analytic expression for the optimal weights, which takes into account: the cluster mass, finite survey volume effects, survey masking, and a flux limit. The optimal weights are $w(M, \chi) \propto b(M, \chi) / [1 + \bar{n}_h(\chi) \bar{b}^2(\chi) \bar{P}(k)]$, where $b(M, \chi)$ is the bias of clusters of mass M at radial position $\chi(z)$, $\bar{n}_h(\chi)$ and $\bar{b}^2(\chi)$ are the expected space density and bias squared of all clusters, and $\bar{P}(k)$ is the matter power spectrum at wavenumber k . This result is analogous to that of Percival et al. We compare our optimal weighting scheme with mass weighting and also with the original power spectrum scheme of Feldman et al. We show that our optimal weighting scheme outperforms these approaches for both volume- and flux-limited cluster surveys. Finally, we present a new expression for the Fisher information matrix for cluster power spectrum analysis. Our expression shows that for an optimally weighted cluster survey the cosmological information content is boosted, relative to the standard approach of Tegmark.

Key words: large-scale structure of Universe.

1 INTRODUCTION

The number counts of massive galaxy clusters has long been known to provide strong constraints on the cosmological model, provided one understands how to map from observed mass proxies to a theoretical halo mass (for a recent review, see Allen, Evrard & Mantz 2011). Since the first measurements of the clustering of Abell clusters was performed in the 70's and early 80's, it has been understood that the clustering of clusters contains additional vital information about the cosmological model (Hauser & Peebles 1973; Bahcall & Soneira 1983). In particular, these studies were able to show that the clustering of clusters was stronger than that of the galaxies. This quickly led to the realization that galaxies and clusters could not both be unbiased tracers of the mass distribution (Kaiser 1984). One of the major attractions of the cold dark matter (hereafter, CDM) framework, is that ‘biased’ clustering naturally emerges within it. In Kaiser’s seminal work, he showed that the peaks and troughs of a Gaussian random field were correlated more strongly than the correlation function of the unconstrained field. Under the assumption that the Abell clusters formed out of the high peaks of a Gaussian random field, one would then expect the Abell clusters to be more strongly correlated than galaxies. Further theoretical support comes from the excursion set formalism, which showed that initially overdense patches of a CDM universe would collapse to form dark matter haloes, and that these would, in general, be positively biased with respect to the underlying matter (Cole & Kaiser 1989; Mo & White 1996; Sheth & Tormen 1999). One important consequence of these developments was that in the mid-90s, it was also realized that, if one combined measurements of the clustering of clusters with measurements of their abundances, one could break the degeneracies in cosmological parameters that were inherent in one single method (Mo, Jing & White 1996; Lima & Hu 2004, 2005; Majumdar & Mohr 2004; Oguri 2009; Cunha, Huterer & Doré 2010; Oguri & Takada 2011; Smith & Marian 2011).

Some notable measurements of the clustering of clusters are: in the X-rays, initial measurements of the cluster correlation function for *ROSAT* data were performed by Romer et al. (1994), and were later improved upon by Collins et al. (2000) using the 344 clusters in the REFLEX survey (see also Abadi, Lambas & Muriel 1998, for results from the XBACS survey). In the optical, cluster samples tend to be orders of magnitude larger (for a review of early results, see Bahcall 1988). The APM galaxy survey was able to identify several hundred

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

clusters for which clustering was computed (Dalton et al. 1992; Miller & Batuski 2001; Miller, Nichol & Batuski 2001). In the past decade, the Sloan Digital Sky Survey (SDSS) has produced, by far, the largest homogeneous sample of optical clusters; the MaxBCG sample whose clusters are detected via the ‘red sequence’ cluster detection method in multiband imaging data (Koester et al. 2007). This sample contains 13 823 clusters with velocity dispersions $\gtrsim 400 \text{ km s}^{-1}$ and covers an area of $\gtrsim 7000 \text{ deg}^2$. The cluster correlation functions were explored by Bahcall et al. (2003) and Estrada, Sefusatti & Frieman (2009), and the power spectrum analysis was performed by Hütsi (2010).

In the future, X-ray cluster surveys, such as eROSITA, should produce homogeneous cluster samples with numbers of clusters on the order of $\sim 100\,000$ (Pillepich, Porciani & Reiprich 2012). Deep multiband optical surveys such as the Dark Energy Survey¹ should also produce tens of thousands of high signal-to-noise ratio (hereafter, S/N) clusters. The question then arises: how should one perform an optimal measurement of the clustering of galaxy clusters? In a series of recent theoretical studies (Seljak, Hamaus & Desjacques 2009; Hamaus et al. 2010), it was claimed that if galaxy clusters were weighted by a function with a linear dependence on mass, then the shot-noise on cluster power spectra measurements would be significantly reduced, hence yielding improved cosmological information. A complex study by Cai, Bernstein & Sheth (2011), proposed that the optimal way to reconstruct the mass distribution from a set of clusters was to weight the galaxy clusters by some combination of their mass and bias. In the limit of a low-mass detection threshold for the clusters, these works lead to the conclusion that weighting by mass results in a maximal S/N measurement of the matter density fluctuations. However, the caveat to the above analysis was that neither work directly demonstrated that the S/N for cluster power spectra would be maximized. In this paper, we shall directly perform this task. As we shall show, our analysis generalizes the galaxy power spectrum methods developed by Feldman, Kaiser & Peacock (1994, hereafter FKP) and Percival, Verde & Peacock (2004, hereafter PVP).

This paper is broken down as follows: in Section 2, we overview the specifications of a cluster survey, the construction of the density field of clusters, and its basic statistical properties. In Section 3, we detail the estimators of the two-point correlation function and the power spectrum. In Section 4, we write down the covariance matrix of the power spectrum estimator in the most general form and the Gaussian limit. In Section 5, we provide details of the derivation of the optimal weighting scheme. In Section 6, we compare various weighting schemes with our optimal weights for the cases of volume limited and flux-limited cluster surveys. We also present a new expression for the Fisher information matrix, which may be used for predicting the cosmological information content of optimally weighted measurements of the cluster power spectrum. Finally, in Section 7 we summarize our findings and conclude.

2 SURVEY SPECIFICATIONS AND THE \mathcal{F}_c -FIELD

2.1 A generic cluster survey

Let us begin by defining our fiducial cluster survey; suppose that we have observed N clusters and to the i th cluster we assign a mass M_i , redshift z_i and angular position on the sky $\Omega_i = \Omega(\theta_i, \phi_i)$. The cluster selection function depends on both position and cluster mass and in general, is a complex function of the survey flux limit, and the cluster detection procedure (for an example of the complexities involved in computing this for the eROSITA mission, see Pillepich et al. 2012). However, it may be simplified in the following ways. First, provided the flux-limit is homogeneous across the survey area, the angular and radial parts of the selection function are separable:

$$\Theta(\mathbf{x}|M) = \Theta(\Omega)\Theta(\chi|M), \quad (1)$$

where $\chi = \chi(z)$ is the radial comoving geodesic distance to redshift z . The angular selection function may be written as

$$\Theta(\Omega) = \begin{cases} 1; & [\Omega \in \Omega_\mu] \\ 0; & [\text{otherwise}] \end{cases}, \quad (2)$$

where Ω_μ defines the survey mask. The radial selection function may be written as

$$\Theta(\chi|M) = \begin{cases} 1; & [\chi \in \chi_{\max}(M)] \\ 0; & [\text{otherwise}] \end{cases}, \quad (3)$$

where $\chi_{\max}(M)$ is the maximum comoving geodesic distance out to which a cluster of mass M could have been detected. This last relation may be inverted to obtain the minimum detectable cluster mass at radial position $\chi(z)$ in the survey, denoted as $M_{\min}(\chi)$. Thus, for any function $\mathcal{B}(\chi, M)$ the following relations apply:

$$\int_0^\infty dM \int_0^\infty d\chi \Theta(\chi|M) \mathcal{B}(\chi, M) = \int_0^\infty dM \int_0^{\chi_{\max}(M)} d\chi \mathcal{B}(\chi, M) = \int_0^\infty d\chi \int_{M_{\min}(\chi)}^\infty dM \mathcal{B}(\chi, M). \quad (4)$$

Finally, the survey volume is the integral of the selection function over all space, $V_\mu(M) = \int d\mathbf{x} \Theta(\mathbf{x}|M)$.

¹ www.darkenergysurvey.org

2.2 The cluster delta expansion

In general, the spatial density distribution of clusters, per unit mass, at position $\mathbf{x}(\chi, \boldsymbol{\Omega})$ may be written as a sum over Dirac delta functions:

$$n_c(M, \mathbf{x}) = \sum_{i=1}^N \delta^D(M - M_i) \delta^D(\mathbf{x} - \mathbf{x}_i). \quad (5)$$

If the selection function is inhomogeneous, then the mean density of clusters varies spatially over the survey. Next, in analogy with PVP, we define a field \mathcal{F}_c , which is related to the overdensity of clusters. This can be written as

$$\mathcal{F}_c(\mathbf{x}) = \int dM \frac{w(\mathbf{x}, M)}{\sqrt{A}} \Theta(\mathbf{x}|M) [n_c(\mathbf{x}, M) - \alpha n_s(\mathbf{x}, M)], \quad (6)$$

where $n_s(\mathbf{x}, M)$ represents the number density of clusters in a mock sample that has no intrinsic spatial correlations, and whose density is $1/\alpha$ times that of the true cluster field at that mass. Note that whilst the field $n_s(\mathbf{x}, M)$ has no intrinsic spatial correlations it does possess a spectrum of masses, which is closely related to the mass spectrum of the field $\langle n_c(\mathbf{x}, M) \rangle$. The choice for the normalization parameter A will be given later. The quantity $w(\mathbf{x}, M)$ denotes a weight function that, in general, may depend on both the spatial position and mass of the cluster. It is this quantity that we shall aim to determine in an optimal way.

2.3 Statistical properties of the cluster density field

Determination of the optimal weight function will require statistical analysis on the field \mathcal{F}_c , therefore we will now introduce the necessary tools. As a simple example let us compute the ensemble average value of the field \mathcal{F}_c , which can be written as

$$\langle \mathcal{F}_c(\mathbf{x}) \rangle = \int dM \frac{w(\mathbf{x}, M)}{\sqrt{A}} \Theta(\mathbf{x}|M) [\langle n_c(\mathbf{x}, M) \rangle - \alpha \langle n_s(\mathbf{x}, M) \rangle], \quad (7)$$

where the angled brackets denote an ensemble average in the following sense:

$$\langle B(\{\mathbf{x}_i, M_i\}) \rangle \equiv \int d\mathbf{x}_1, \dots, d\mathbf{x}_N dM_1, \dots, dM_N p(\mathbf{x}_1, \dots, \mathbf{x}_N, M_1, \dots, M_N) B(\{\mathbf{x}_i, M_i\}). \quad (8)$$

In the above equation, $p(\mathbf{x}_1, \dots, \mathbf{x}_N, M_1, \dots, M_N)$ is the N -point joint probability distribution for the N clusters being located at the set of spatial positions $\{\mathbf{x}_i\}$ and having the set of masses $\{M_i\}$. Thus, the first expectation on the right-hand side of equation (7) can be written as

$$\langle n_c(\mathbf{x}, M) \rangle = \int \prod_{i=1}^N \{d\mathbf{x}_i dM_i\} p(\mathbf{x}_1, \dots, \mathbf{x}_N, M_1, \dots, M_N) \sum_{i=1}^N \delta^D(\mathbf{x} - \mathbf{x}_i) \delta^D(M - M_i) = \sum_{i=1}^N p(\mathbf{x}, M) = Np(\mathbf{x}, M). \quad (9)$$

On the first line, we inserted the expansion of the cluster density field from equation (5) and to obtain the second we integrated over the sum of Dirac delta functions. The quantity $p(\mathbf{x}, M)$ can be written in a more transparent way

$$p(\mathbf{x}, M) \equiv p(\mathbf{x})p(M) = \frac{1}{V} \frac{\bar{n}(M)}{N/V} = \frac{\bar{n}(M)}{N}, \quad (10)$$

where $\bar{n}(M)$ is the intrinsic number density of clusters per unit mass – i.e. the dark matter halo mass function.

Turning to the second expectation value, we note that the only difference between $\langle n_c(\mathbf{x}, M) \rangle$ and $\langle n_s(\mathbf{x}, M) \rangle$ is the artificially increased space density of clusters and the absence of any intrinsic clustering. Hence, we also have

$$\alpha \langle n_s(\mathbf{x}, M) \rangle = \bar{n}(M). \quad (11)$$

Putting this all together, we arrive at the result

$$\langle \mathcal{F}_c(\mathbf{x}) \rangle = 0. \quad (12)$$

Hence the \mathcal{F}_c -field, like the overdensity field of matter, is a mean-zero field.

Note that we have neglected to take into account the statistical properties of obtaining the N clusters in the survey volume. In what follows, we shall assume that the survey volumes are sufficiently large that this may be essentially treated as a deterministic quantity. However, it can be taken into account (e.g. see Sheth & Lemson 1999; Smith & Watts 2005; Smith 2009).

3 CLUSTERING ESTIMATORS

3.1 An estimator for the two-point correlation function

The two-point correlation function of the field \mathcal{F}_c can be computed directly as

$$\begin{aligned} \langle \mathcal{F}_c(\mathbf{x}_1) \mathcal{F}_c(\mathbf{x}_2) \rangle &= \frac{1}{A} \int dM_1 dM_2 w(\mathbf{x}_1, M_1) w(\mathbf{x}_2, M_2) \Theta(\mathbf{x}_1|M_1) \Theta(\mathbf{x}_2|M_2) \left[\langle n_c(\mathbf{x}_1, M_1) n_c(\mathbf{x}_2, M_2) \rangle \right. \\ &\quad \left. - \alpha \langle n_c(\mathbf{x}_1, M_1) n_s(\mathbf{x}_2, M_2) \rangle - \alpha \langle n_c(\mathbf{x}_2, M_2) n_s(\mathbf{x}_1, M_1) \rangle + \alpha^2 \langle n_s(\mathbf{x}_1, M_1) n_s(\mathbf{x}_2, M_2) \rangle \right]. \end{aligned} \quad (13)$$

The calculation of the above terms is detailed in the Appendix A. On substituting the results of equations (A2) and (A3), we write

$$\begin{aligned} \xi_{\mathcal{F}_c}(\mathbf{x}_1, \mathbf{x}_2) &= \frac{1}{A} \int dM_1 dM_2 w(\mathbf{x}_1, M_1) w(\mathbf{x}_2, M_2) \bar{n}(M_1) \bar{n}(M_2) \Theta(\mathbf{x}_1 | M_1) \Theta(\mathbf{x}_2 | M_2) \xi^c(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) \\ &\quad + \frac{(1 + \alpha)}{A} \int dM w^2(\mathbf{x}_1, M) \bar{n}(M) \Theta(\mathbf{x}_1 | M) \delta^D(\mathbf{x}_1 - \mathbf{x}_2), \end{aligned} \quad (14)$$

where we defined $\xi_{\mathcal{F}_c}(\mathbf{x}_1, \mathbf{x}_2) \equiv \langle \mathcal{F}_c(\mathbf{r}_1) \mathcal{F}_c(\mathbf{r}_2) \rangle$. Note that in the second line of the above equation, we have benefited from the identity: $\Theta^2(\mathbf{x} | M) = \Theta(\mathbf{x} | M)$. If we assume that the cluster density field is some local function of the underlying dark matter density (Fry & Gaztanaga 1993; Mo & White 1996; Mo, Jing & White 1997; Smith, Scoccimarro & Sheth 2007), the cross-correlation function of clusters of masses M_1 and M_2 , at leading order, can be written as

$$\xi^c(|\mathbf{x}_1 - \mathbf{x}_2|, M_1, M_2) = b(M_1) b(M_2) \xi(|\mathbf{x}_1 - \mathbf{x}_2|), \quad (15)$$

where $\xi(r)$ is the correlation of the underlying matter fluctuations. On inserting this relation into equation (14), we find

$$\xi_{\mathcal{F}_c}(\mathbf{x}_1, \mathbf{x}_2) = \mathcal{G}_{(1,1)}(\mathbf{x}_1) \mathcal{G}_{(1,1)}(\mathbf{x}_2) \xi(|\mathbf{x}_1 - \mathbf{x}_2|) + (1 + \alpha) \mathcal{G}_{(2,0)}(\mathbf{x}_1) \delta^D(\mathbf{x}_1 - \mathbf{x}_2), \quad (16)$$

where we have defined the weighted selection function

$$\mathcal{G}_{(l,m)}(\mathbf{x}) \equiv \frac{1}{A^{l/2}} \int dM \bar{n}(M) b^m(M) w^l(\mathbf{x}, M) \Theta(\mathbf{x} | M). \quad (17)$$

One possible estimator for the matter correlation function from the \mathcal{F}_c field is therefore

$$\hat{\xi}_{\mathcal{F}_c}(\mathbf{x}) \equiv \int d\mathbf{x}' \mathcal{F}_c(\mathbf{x}') \mathcal{F}_c(\mathbf{x} + \mathbf{x}'); \quad (\mathbf{x} \neq \mathbf{0}). \quad (18)$$

If we now compute the expectation of this estimator we find

$$\langle \hat{\xi}_{\mathcal{F}_c}(\mathbf{x}) \rangle = \xi(\mathbf{x}) \int d\mathbf{x}' \mathcal{G}_{(1,1)}(\mathbf{x}') \mathcal{G}_{(1,1)}(\mathbf{x} + \mathbf{x}'); \quad (\mathbf{x} \neq \mathbf{0}). \quad (19)$$

Although $\hat{\xi}_{\mathcal{F}_c}(\mathbf{x})$ is not an unbiased estimator of the matter correlation function, we may construct one that is

$$\hat{\xi}_0(\mathbf{x}) \equiv \hat{\xi}_{\mathcal{F}_c}(\mathbf{x}) / \Xi_0(\mathbf{x}); \quad \Xi_0(\mathbf{x}) \equiv \int d\mathbf{x}' \mathcal{G}_{(1,1)}(\mathbf{x}') \mathcal{G}_{(1,1)}(\mathbf{x} + \mathbf{x}'). \quad (20)$$

3.2 An estimator for the cluster power spectrum

We may now compute the Fourier space equivalent of the two-point correlation function, the power spectrum. In what follows, we shall adopt the following Fourier transform conventions:

$$\tilde{B}(\mathbf{k}) = \int d\mathbf{x} B(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} \Leftrightarrow B(\mathbf{x}) = \int \frac{d\mathbf{k}}{(2\pi)^3} \tilde{B}(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{x}}.$$

We also define the power spectrum, $P_B(k)$, of any infinite statistically homogeneous random field $\tilde{B}(\mathbf{k})$ to be

$$\langle \tilde{B}(\mathbf{k}) \tilde{B}(\mathbf{k}') \rangle \equiv (2\pi)^3 \delta^D(\mathbf{k} + \mathbf{k}') P_B(k).$$

Note that if the field B were statistically isotropic, the power spectrum would simply be a function of the scalar k . With the above definitions in hand, the covariance of the Fourier modes of the cluster field \mathcal{F}_c can be written as

$$\begin{aligned} \langle \tilde{\mathcal{F}}_c(\mathbf{k}_1) \tilde{\mathcal{F}}_c(\mathbf{k}_2) \rangle &= \frac{1}{A} \int \prod_{i=1}^2 \{ d\mathbf{x}_i dM_i w(\mathbf{x}_i, M_i) \bar{n}(M_i) \Theta(\mathbf{x}_i | M_i) e^{i\mathbf{k}_i \cdot \mathbf{x}_i} \} \xi^c(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) \\ &\quad + (1 + \alpha) \int d\mathbf{x} dM \bar{n}(M) \Theta(\mathbf{x} | M) w^2(\mathbf{x}, M) e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{x}}. \end{aligned} \quad (21)$$

For any infinite homogeneous random field, the two-point correlation function and the power spectrum form a Fourier pair, hence we may write

$$\xi^c(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) = \int \frac{d\mathbf{k}}{(2\pi)^3} P_c(\mathbf{k}, M_1, M_2) e^{-i\mathbf{k} \cdot (\mathbf{x}_1 - \mathbf{x}_2)}.$$

If we assume the linear biasing relation of equation (15), then the cross-power spectrum of clusters of different masses M_1 and M_2 , at leading order, can be written as

$$P_c(\mathbf{k}, M_1, M_2) = b(M_1) b(M_2) P(\mathbf{k}), \quad (22)$$

where $P(\mathbf{k})$ is the matter power spectrum. On using these expressions in equation (21), and considering the case $\mathbf{k}_1 = -\mathbf{k}_2$ we find

$$\langle |\tilde{\mathcal{F}}_c(\mathbf{k})|^2 \rangle = \int \frac{d\mathbf{q}}{(2\pi)^3} P(\mathbf{q}) \left| \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k} - \mathbf{q}) \right|^2 + P_{\text{shot}}, \quad (23)$$

where in the above expression $\tilde{\mathcal{G}}_{(l,m)}(\mathbf{k})$ is the Fourier transform of the weighted survey selection function from equation (17) and we have defined the shot-noise term as

$$P_{\text{shot}} \equiv (1 + \alpha)\tilde{\mathcal{G}}_{(2,0)}(\mathbf{0}). \quad (24)$$

Just as in the case of galaxies (Feldman et al. 1994), the expectation of the square amplitude of the Fourier modes of \mathcal{F}_c is given by the convolution of the matter power spectrum with the modulus square of the Fourier modes of the survey window function, plus a constant shot noise.

In the limit that the survey volume is large, the functions $\tilde{\mathcal{G}}_{(l,m)}(\mathbf{k})$ will be very narrowly peaked around $\mathbf{k} = 0$. Provided the matter power spectrum is a smoothly varying function of scale, the window functions $\tilde{\mathcal{G}}_{(l,m)}(\mathbf{k})$ take on Dirac delta-function-like behaviour. Thus, in the large-survey volume limit, the first term on the right-hand side of equation (23) becomes

$$\int \frac{d\mathbf{q}}{(2\pi)^3} P(\mathbf{q}) \left| \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k} - \mathbf{q}) \right|^2 \approx P(\mathbf{k}) \int \frac{d\mathbf{q}}{(2\pi)^3} \left| \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k} - \mathbf{q}) \right|^2. \quad (25)$$

Let us focus on the integral factor on the right-hand side of the above expression. Transforming the variable $\mathbf{k} - \mathbf{q} \rightarrow \mathbf{q}'$ and using Parseval's theorem, as well as equation (17), we find

$$\int \frac{d\mathbf{q}}{(2\pi)^3} \left| \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k} - \mathbf{q}) \right|^2 = \int \frac{d\mathbf{q}'}{(2\pi)^3} \left| \tilde{\mathcal{G}}_{(1,1)}(\mathbf{q}') \right|^2 = \int d\mathbf{x} [\mathcal{G}_{(1,1)}(\mathbf{x})]^2 = \frac{1}{A} \int d\mathbf{x} \left[\int dM \bar{n}(M) b(M) w(\mathbf{x}, M) \Theta(\mathbf{x}|M) \right]^2. \quad (26)$$

Recall that we have not yet specified the parameter A , let us now define it to be

$$A \equiv \int d\mathbf{x} [\mathcal{G}_{(1,1)}(\mathbf{x})]^2 = \int d\mathbf{x} \left[\int dM \bar{n}(M) b(M) w(\mathbf{x}, M) \Theta(\mathbf{x}|M) \right]^2. \quad (27)$$

With this choice of normalization, equation (26) is simply unity. Hence, for the case of large homogeneous survey volumes, an unbiased estimator for the dark matter power spectrum is

$$\hat{P}(\mathbf{k}) \approx |\tilde{\mathcal{F}}_c(\mathbf{k})|^2 - P_{\text{shot}}. \quad (28)$$

The above estimator is for the power spectrum at a particular mode, whereas we are more interested in a band-power estimate of the power spectrum. Thus, our final estimator is

$$\hat{P}(k_i) = \frac{1}{V_i} \int_{V_i} d\mathbf{k} \hat{P}(\mathbf{k}) = \frac{1}{V_i} \int_{V_i} d\mathbf{k} |\tilde{\mathcal{F}}_c(\mathbf{k})|^2 - P_{\text{shot}}, \quad (29)$$

where in the above we have summed all modes over a shell in k -space of thickness Δk , i.e.

$$V_i \equiv \int_{V_i} d\mathbf{k} = 4\pi \int_{k_i - \Delta k/2}^{k_i + \Delta k/2} k^2 dk = 4\pi k_i^2 \Delta k \left[1 + \frac{1}{12} \left(\frac{\Delta k}{k_i} \right)^2 \right].$$

If the survey window function possesses small-scale structure, then the matter power spectrum can only safely be recovered by deconvolution of the window function $\left| \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}) \right|^2$, or alternatively one must convolve theory predictions with the window function. Otherwise, equation (28) is a biased estimator.

4 STATISTICAL FLUCTUATIONS IN THE CLUSTER POWER SPECTRUM

In order to obtain the optimal estimator, we need to know how the S/N varies with the shape of the weight function $w(\mathbf{x}, M)$. Hence, we need to understand the noise properties of our power spectrum estimator.

4.1 The covariance of the cluster power spectrum estimator

In general, the covariance matrix of the band-power spectrum estimator can be written as

$$\text{Cov} \left[\hat{P}(k_i), \hat{P}(k_j) \right] \equiv \left\langle \hat{P}(k_i) \hat{P}(k_j) \right\rangle - \langle \hat{P}(k_i) \rangle \langle \hat{P}(k_j) \rangle = \frac{1}{V_i} \int_{V_i} d\mathbf{k}_1 \frac{1}{V_j} \int_{V_j} d\mathbf{k}_2 \text{Cov} \left[\hat{P}(\mathbf{k}_1), \hat{P}(\mathbf{k}_2) \right].$$

In Appendix B1, we show that the latter term can be written as

$$\text{Cov} \left[\hat{P}(\mathbf{k}_1), \hat{P}(\mathbf{k}_2) \right] \approx \text{Cov} \left[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \right] = \left\langle |\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2 |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \right\rangle - \left\langle |\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2 \right\rangle \left\langle |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \right\rangle \quad (30)$$

As detailed in Appendix B2, the covariance can be written as

$$\begin{aligned} \text{Cov} \left[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \right] &= \frac{1}{A^2} \int d\mathbf{k}_3 d\mathbf{k}_4 \delta^D(\mathbf{k}_1 + \mathbf{k}_3) \delta^D(\mathbf{k}_2 + \mathbf{k}_4) \int d\mathbf{x}_1, \dots, d\mathbf{x}_4 e^{i\mathbf{k}_1 \cdot \mathbf{x}_1 + \dots + i\mathbf{k}_4 \cdot \mathbf{x}_4} \\ &\quad \times [\langle \mathcal{F}_c(\mathbf{x}_1) \dots \mathcal{F}_c(\mathbf{x}_4) \rangle - \langle \mathcal{F}_c(\mathbf{x}_1) \mathcal{F}_c(\mathbf{x}_3) \rangle \langle \mathcal{F}_c(\mathbf{x}_2) \mathcal{F}_c(\mathbf{x}_4) \rangle]. \end{aligned} \quad (31)$$

The last term in the equation above can be computed with equation (B23). In Appendix B3, we provide a general relation for the covariance of the \mathcal{F}_c power spectrum in terms of Fourier-space quantities. These results generalize the expressions for the covariance matrix of the cluster power spectrum presented in (Smith 2009), extending that work to the case of finite survey geometry and an arbitrary weighting scheme that depends on mass and position. In Appendix B4, we derive the covariance matrix in the presence of the following approximations: the underlying matter density is Gaussian; the sampling fluctuations for a given realization of the cluster field are small; the survey volume is large and homogeneous. These assumptions lead to a simplified form of the general equation (B25)

$$\text{Cov}[|\mathcal{F}_c(k_i)|^2, |\mathcal{F}_c(k_j)|^2] \approx \frac{2(2\pi)^3}{V_i} \bar{P}^2(k_i) \int d\mathbf{x} \left\{ [\mathcal{G}_{(1,1)}(\mathbf{x})]^2 + \frac{(1+\alpha)}{\bar{P}(k_i)} \mathcal{G}_{(2,0)}(\mathbf{x}) \right\}^2 \delta_{i,j}^K. \quad (32)$$

We shall use this result for the covariance of the cluster power spectrum estimator to derive the optimal weights in the next section. In future work, it will be interesting to explore how the optimal weights change as each of these assumptions is gradually relaxed.

At this point, it is also worth comparing our analysis with that of FKP and PVP. Our expressions for the correlation function equation (20) and the power spectrum equation (28) are analogous to those found in PVP for the luminosity dependence of the galaxy bias. However, our derivation is different, particularly in the way in which we treat the statistical properties of the cluster field, which we believe to be more general and transparent. Thus, our method establishes a clear formalism for evaluating how the optimal weights change when the above-mentioned assumptions are relaxed, which will be very useful to future studies of optimal power spectra estimators. Although our choice of the normalization parameter A is different from PVP, it is of no relevance until we derive $w(\mathbf{r}, M)$. Indeed, in the next section we will show that, once the optimal weights are selected, their and our choice for the field \mathcal{F}_c turn out to be equivalent.

Comparing to FKP, we also note that our derivation of the statistical fluctuations in the power and the approximations that we make in order to obtain a diagonal covariance matrix, are more rigorous and transparent than in the study of FKP. For instance, the original FKP derivation makes use of Parseval's theorem to transform from a band-power k -space integral to an integral over the entirety of real space – that is they effectively go from equation (B30) to equation (32) using Parseval's theorem. Owing to the fact that the band-power averages do not extend over the whole of k -space, indeed they may be limited to lie only in a very narrow shell, the application of Parseval's theorem appears incorrect. As a minor point, their derivation also misses a factor of 2, which arises due to the Hermitian nature of the Fourier modes. This however plays no role in the derivation of the optimal weights.

5 AN OPTIMAL WEIGHTING SCHEME

5.1 Optimal weights as a functional problem

Our aim is to find the optimal weighting scheme that will maximize the S/N ratio on a given band power estimate of the cluster power spectrum. To begin, note that maximizing the S/N ratio is equivalent to minimizing its inverse, the noise-over-signal ratio (N/S). This can be expressed as

$$F[w(\mathbf{x}, M)] \equiv \frac{\sigma_p^2(k_i)}{\bar{P}^2(k_i)} = \frac{2(2\pi)^3}{V_i} \int d\mathbf{x} \left\{ [\mathcal{G}_{(1,1)}(\mathbf{x})]^2 + c \mathcal{G}_{(2,0)}(\mathbf{x}) \right\}^2, \quad (33)$$

where we have defined the constant

$$c \equiv (1 + \alpha) / \bar{P}(k_i). \quad (34)$$

In the above expression, we have written the quantity $F[w]$ as a functional of the weights $w(\mathbf{x}, M)$. The standard way for finding the optimal weights, is to perform the functional variation of F with respect to the weights w . Operationally, the functional variation of $F[w]$ may be defined as

$$\delta F[w(\mathbf{x}, M)] = F[w(\mathbf{x}, M) + \delta w(\mathbf{x}, M)] - F[w(\mathbf{x}, M)] = \int d\mathbf{x} dM \left\{ \frac{\delta F}{\delta w(\mathbf{x}, M)} \right\} \delta w(\mathbf{x}, M).$$

The extremization condition is that the functional derivative is stationary for the optimal weights:

$$\frac{\delta F}{\delta w(\mathbf{x}, M)} = 0.$$

From equations (17) and (27), we now note that $F[w]$ is the ratio of two weight-dependent functionals, since the \mathcal{G} functions contain weights not only by definition, but also through the normalization constant A :

$$F[w] = \mathcal{N}[w] / \mathcal{D}[w], \quad (35)$$

where we have defined the numerator and denominator functionals:

$$\mathcal{N}[w] \equiv \int d\mathbf{x} \left\{ [\bar{\mathcal{G}}_{(1,1)}(\mathbf{x})]^2 + c \bar{\mathcal{G}}_{(2,0)}(\mathbf{x}) \right\}^2; \quad (36)$$

$$\mathcal{D}[w] \equiv A^2[w] = \left\{ \int d\mathbf{x} [\bar{\mathcal{G}}_{(1,1)}(\mathbf{x})]^2 \right\}^2. \quad (37)$$

In the above, we have ignored the constant $2(2\pi)^3/V_i$ from equation (33), since it does not play any part in the minimization process, and we have introduced a scaled version of the survey window function which is independent of the normalization constant:

$$\bar{\mathcal{G}}_{(l,m)} \equiv A^{l/2} \mathcal{G}_{(l,m)}. \quad (38)$$

To minimize $F[w]$, we must solve the following functional problem:

$$\frac{1}{\mathcal{D}[w]} \left[\delta \mathcal{N}[w] - \frac{\mathcal{N}[w]}{\mathcal{D}[w]} \delta \mathcal{D}[w] \right] = 0 \iff \delta \mathcal{N}[w] - F[w] \delta \mathcal{D}[w] = 0. \quad (39)$$

In Appendix C, we compute the variations of \mathcal{N} and \mathcal{D} with a perturbation δw . On replacing the results of equations (C4) and (C5) in equation (39) and dropping all constant terms, we arrive at the general optimal weight equation:

$$\left\{ [\bar{\mathcal{G}}_{(1,1)}(\mathbf{x})]^2 + c \bar{\mathcal{G}}_{(2,0)}(\mathbf{x}) \right\} \left\{ \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) b(M) + c w(\mathbf{x}, M) \right\} = \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) b(M). \quad (40)$$

5.2 The optimal weights

We are now in a position to derive the optimal weights. Consider equation (40), we examine the scaling with mass and position of the functions on both sides of the equation. Since the right-hand side is proportional to a bias term, it follows that the left-hand side must also obey this proportionality. The first bracket on the left-hand side does not have any mass dependence. We therefore infer that the weights' dependence on position and mass must be separable:

$$w(\mathbf{x}, M) = b(M) \tilde{w}(\mathbf{x}). \quad (41)$$

As an immediate consequence of this separability, the functions $\bar{\mathcal{G}}$ can be written as

$$\bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) = \tilde{w}(\mathbf{x}) \int dM \bar{n}(M) b^2(M) \Theta(\mathbf{x}|M);$$

$$\bar{\mathcal{G}}_{(2,0)}(\mathbf{x}) = [\tilde{w}(\mathbf{x})]^2 \int dM \bar{n}(M) b^2(M) \Theta(\mathbf{x}|M) = \tilde{w}(\mathbf{x}) \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}).$$

Replacing these relations and equation (41) in equation (40), we arrive after a little work at the following solution for the space-dependent part of the optimal weights:

$$\tilde{w}(\mathbf{x}) = 1 / \left[c + \int dM \bar{n}(M) b^2(M) \Theta(\mathbf{x}|M) \right]. \quad (42)$$

On putting together equations (41) and (42) and substituting back the constant from equation (34), we write the final expression for the optimal weights for achieving maximal S/N on a given band-power estimate of the cluster power spectrum:

$$w(\mathbf{x}, M) = \frac{b(M)}{(1 + \alpha) + \int dM \bar{n}(M) b^2(M) \Theta(\mathbf{x}|M) \bar{P}_i}. \quad (43)$$

As in the case of the original FKP weights, we see that the answer is somewhat circular, in that in order to estimate the cluster power spectrum optimally, we already need to have some reasonably good estimate of the underlying matter power spectrum. Indeed, in order to fully implement our scheme we are also required to have knowledge of the functions $b(M)$ and $\bar{n}(M)$. These two functions are theoretically very well known for dark matter haloes. They may also be measured directly from the data – albeit with noise. One also should have very good understanding of the selection function $\Theta(\mathbf{x}|M)$. The parameter α is determined directly from the density of the random cluster sample.

We note that equation (43) is virtually identical to the result found by PVP for the case of luminosity-dependent galaxy bias. However, they provided no analytic proof for their result, but simply proposed a conjecture for the general weight solution and showed that it satisfied their weight equation. Our derivation, on the other hand, is more elegant and easily verified. Finally, as was pointed out earlier, once the optimal weights are determined, our choice of field \mathcal{F}_c and theirs, are virtually equivalent. Their inverse weighting of the cluster field by the bias, therefore appears to be an unnecessary step.

Lastly, we generalize equation (43) to the case where both $b(M)$ and $\bar{n}(M)$ are time dependent quantities. This may be achieved by simply making them a function of conformal time η or equivalently χ , and also accounting for the presence of the growth factor $D(\chi)$ in equation (15). If one propagates these transformations through the entire derivation, then one finds that the weights may be written as

$$w_{\text{OPT}}(\mathbf{x}(\chi, \boldsymbol{\Omega}), M) = \frac{D(\chi) b(M, \chi)}{(1 + \alpha) + \int dM \bar{n}(M, \chi) D^2(\chi) b^2(M, \chi) \Theta(\mathbf{x}|M) \bar{P}_i}. \quad (44)$$

6 CASE STUDIES

We shall now examine how the S/N ratios vary as a function of minimum cluster mass, for both volume- and flux-limited samples of clusters. We shall compare the optimal weighting scheme derived in the previous section with the standard FKP weighting and also the mass weighting advocated by Seljak et al. (2009). The optimal weighting scheme proposed by Hamaus et al. (2010) is based on the idea of shot-noise

minimization in the measurements of halo density fluctuations. The complex study of Cai et al. (2011) derives, for various scenarios and among other things, an optimal weight for estimating the matter density fluctuations from measurements of the halo density fluctuations. These last two weighting schemes are complicated to implement in our calculations here. However, in the limit where the haloes included in the estimates have sufficiently low masses (i.e. mass-detection threshold for haloes is low), both latter weighting schemes converge towards mass weighting.

The standard FKP weighting is

$$w_{\text{FKP}}(\mathbf{x}(\chi, \boldsymbol{\Omega}), M) = w_{\text{FKP}}(\chi) = \frac{1}{(1 + \alpha) + \bar{n}_h(\chi)P_i}. \quad (45)$$

We shall denote the mass weighting of the clusters (Seljak et al. 2009; Hamaus et al. 2010) combined with FKP's space weighting as

$$w_{\text{M+FKP}}(\mathbf{x}(\chi, \boldsymbol{\Omega}), M) = w_{\text{M+FKP}}(\chi, M) = M w_{\text{FKP}}(\chi). \quad (46)$$

The general expressions for the S/N corresponding to these weights are shown in Appendix D.

6.1 Volume-limited samples

If the survey is volume limited, the minimum detectable mass is a constant throughout the survey, i.e. $M_{\min}(\chi) = M_{\min}$. Thus, the radial selection function is fully separable:

$$\Theta^{\text{VL}}(\chi|M) = \Theta^{\text{VL}}(\chi)\Theta^{\text{VL}}(M), \quad (47)$$

where

$$\Theta^{\text{VL}}(\chi) = \begin{cases} 1; [\chi \leq \chi_{\max}] \\ 0; [\text{otherwise}] \end{cases}; \Theta^{\text{VL}}(M) = \begin{cases} 1; [M \geq M_{\min}] \\ 0; [\text{otherwise}] \end{cases}.$$

We also note that for a volume-limited sample, the weight function possesses no spatial dependence and so we are free to take the weights simply as

$$w_{\text{FKP}}(\mathbf{x}, M) = 1; w_{\text{OPT}}(\mathbf{x}, M) = b(M); w_{\text{M+FKP}}(\mathbf{x}, M) = M. \quad (48)$$

With the above weights, equations (D2), (D3), (D4) transform into

$$\begin{aligned} \left(\frac{S}{N}\right)_{\text{FKP}}(k_i) &= \sqrt{\frac{V_\mu V_i}{2(2\pi)^3}} \left[\frac{\bar{b}^2 \bar{n}_h \bar{P}_i}{(1 + \alpha) + \bar{b}^2 \bar{n}_h \bar{P}_i} \right], \\ \left(\frac{S}{N}\right)_{\text{OPT}}(k_i) &= \sqrt{\frac{V_\mu V_i}{2(2\pi)^3}} \left[\frac{\bar{b}^2 \bar{n}_h \bar{P}_i}{(1 + \alpha) + \bar{b}^2 \bar{n}_h \bar{P}_i} \right], \\ \left(\frac{S}{N}\right)_{\text{M+FKP}}(k_i) &= \sqrt{\frac{V_\mu V_i}{2(2\pi)^3}} \left[\frac{r_M^2 \bar{b}_M^{-2} \bar{n}_h \bar{P}_i}{(1 + \alpha) + r_M^2 \bar{b}_M^{-2} \bar{n}_h \bar{P}_i} \right], \end{aligned} \quad (49)$$

where we defined $r_M^2 \equiv \langle M \rangle^2 / \langle M^2 \rangle$. All quantities on the right-hand side of these equations are the same as defined by equation (D1), only evaluated for a constant mass threshold M_{\min} . The S/N values obtained from the three weighting schemes may be more easily compared if we take the ratio of the optimal and mass weighting scheme with respect to the original FKP weights. Whereupon,

$$\frac{(S/N)_{\text{OPT}}}{(S/N)_{\text{FKP}}} = \frac{1 + (1 + \alpha)/\bar{n}_h \bar{P}_i \bar{b}^2}{1 + (1 + \alpha)/\bar{n}_h \bar{P}_i \bar{b}^2}; \frac{(S/N)_{\text{M+FKP}}}{(S/N)_{\text{FKP}}} = \frac{1 + (1 + \alpha)/\bar{n}_h \bar{P}_i \bar{b}^2}{1 + (1 + \alpha)/\bar{n}_h \bar{P}_i \bar{b}_M^{-2} r_M^2}. \quad (50)$$

There are two limiting cases of interest:

(i) $\bar{n}_h \bar{P}_i \gg 1$: in this limit the two terms in equation (50) become

$$\frac{(S/N)_{\text{OPT}}}{(S/N)_{\text{FKP}}} \approx 1 + \frac{(1 + \alpha)}{\bar{n}_h P_i} \left[\frac{1}{\bar{b}^2} - \frac{1}{\bar{b}^2} \right] \geq 1; \frac{(S/N)_{\text{M+FKP}}}{(S/N)_{\text{FKP}}} \approx 1 + \frac{(1 + \alpha)}{(\bar{n}_h P_i)} \left[\frac{1}{\bar{b}^2} - \frac{1}{\bar{b}_M^{-2} r_M^2} \right]. \quad (51)$$

(ii) $\bar{n}_h \bar{P}_i \ll 1$: in this limit the expressions in equation (50) become

$$\frac{(S/N)_{\text{OPT}}}{(S/N)_{\text{FKP}}} \approx \frac{\bar{b}^2}{\bar{b}^2} \geq 1; \frac{(S/N)_{\text{M+FKP}}}{(S/N)_{\text{FKP}}} \approx \frac{1}{r_M^2} \left(\frac{\bar{b}_M}{\bar{b}} \right)^2. \quad (52)$$

The first inequalities in equations (51) and (52) both follow from the fact that $\bar{b}^2 \leq \bar{b}_M^2$ (for a proof of this relation see Appendix E). In order to determine whether the second expressions in equations (51) and (52) are greater or less than unity, one must examine the product of the quantities $\bar{b}^2/\bar{b}_M^2 \leq 1$ and $\langle M^2 \rangle/\langle M \rangle^2 \geq 1$. Unfortunately, this is not so easy to determine and we therefore turn to numerical evaluation of the expressions.

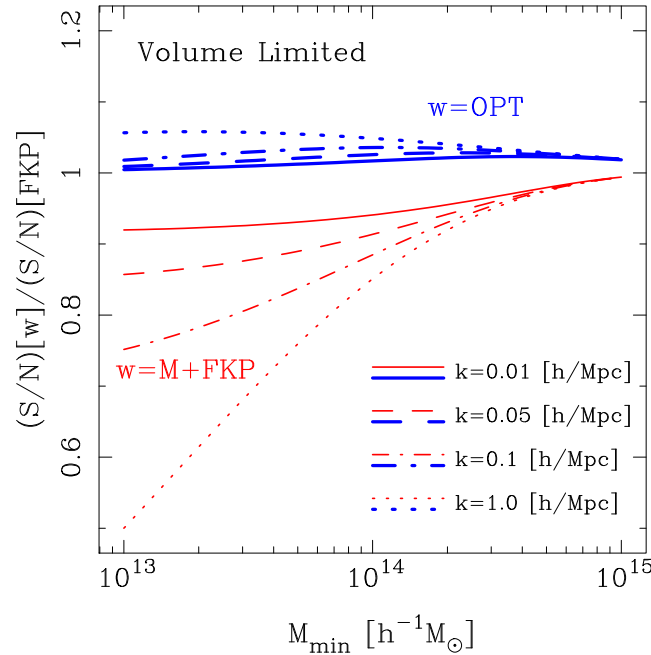


Figure 1. Signal-to-noise ratios for the optimal and mass-weighted cluster power spectrum relative to the FKP signal to noise, as a function of the minimum detectable mass in a volume-limited cluster survey. Thick blue lines and thin red lines denote the optimal and mass-weighting schemes, respectively. The solid, dash, dot–dash and dotted lines show the results for band-power wavenumbers $k = \{0.01, 0.05, 0.1, 1.0\} h \text{Mpc}^{-1}$.

In Fig. 1, we present the evolution of the S/N for the optimal and mass weighting schemes, relative to the S/N obtained from the FKP weighting, as a function of M_{\min} . The blue and red lines denote the optimal and mass-weighting schemes, respectively, and results are presented for several k -band powers. Here, we have used the LCDM model as a particular example and have adopted the bias and dark matter halo mass function formula of Sheth & Tormen (1999) to evaluate equation (50). We notice the following: the optimal weighting scheme does indeed maximize the S/N; the mass-weighting scheme is inferior to the optimal and FKP weighting schemes; the overall gains of the optimal weighting scheme relative to the FKP scheme appear modest ~ 5 per cent.

6.2 Flux-limited samples

For flux-limited samples, the S/N for the three weighting schemes is given by the general equations (D2)–(D4). Further analytic developments are non-trivial, and so we numerically evaluate them for the particular case of LCDM. To accomplish this, we require knowledge of the minimum detectable mass as a function of χ . As mentioned in Section 2, $M_{\min}(\chi)$, in general, is a complicated function of the survey flux-limit and the cluster identification algorithm. For simplicity, we shall assume that this can be written as

$$M_{\min}(\chi) = M_{\min}(0)e^{\gamma z(\chi)}. \quad (53)$$

If we adopt the value $\gamma = 1$, the above functional form roughly matches the cluster selection as a function of redshift, which one finds for weak lensing detected cluster surveys (Marian & Bernstein 2006). We evaluate the above S/N ratios as a function of $M_{\min}(0)$, with $M_{\min}(0)$ varying in the range $[10^{13}, 10^{15}] h^{-1} M_{\odot}$.

Fig. 2 is the analogue of Fig. 1 for a flux-limited cluster survey. The blue and red lines denote the optimal and mass weighting schemes, respectively, and results are presented for several k -band powers. Again, we have used the LCDM model as an example and have adopted the bias and dark matter halo mass functions from Sheth & Tormen (1999) to evaluate equations (D2), (D3) and (D4). Several features may be noted: the optimal weighting scheme always maximizes the signal-to-noise ratio; the mass-weighting scheme is inferior to both the optimal and FKP schemes for all scales; the overall S/N gains for the optimal weighting are very modestly (~ 3 per cent) improved over the FKP approach.

6.3 The fisher matrix

As a final corollary to this section, we explore the cosmological information content of an optimally weighted cluster power spectrum analysis. Following Tegmark (1997), the Fisher information matrix for a power spectrum analysis can be defined as

$$\mathcal{F}_{\alpha\beta} = \sum_{i,j} \frac{\partial \bar{P}_i}{\partial \alpha} C_{ij}^{-1} \frac{\partial \bar{P}_j}{\partial \beta} = \sum_{i,j} \frac{\partial \log \bar{P}_i}{\partial \alpha} \bar{P}_i C_{ij}^{-1} \bar{P}_j \frac{\partial \log \bar{P}_j}{\partial \beta}, \quad (54)$$

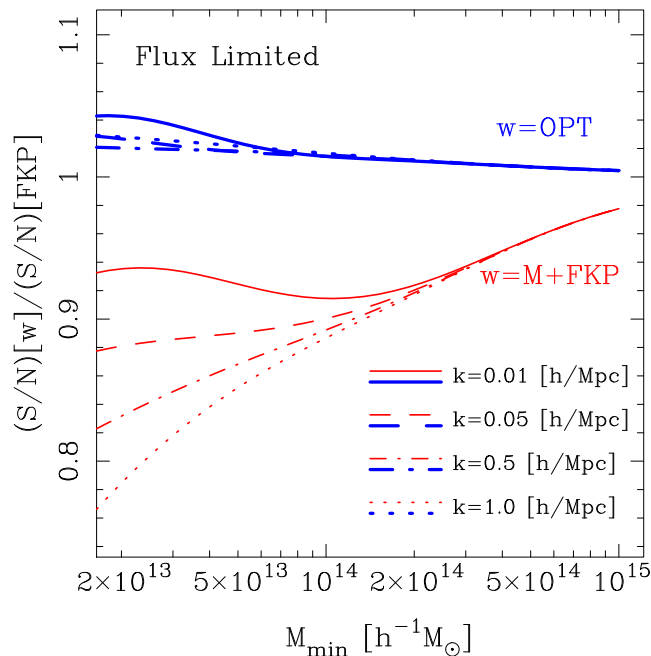


Figure 2. Same as Fig. 1 except this time for a flux-limited survey.

where $\partial/\partial\alpha \equiv \partial/\partial\theta_\alpha$ denote partial derivatives with respect to the cosmological parameters θ_α , and where here we have neglected the information content in the covariance matrix (Tegmark, Taylor & Heavens 1997). On inserting our earlier expression for the covariance matrix, as given by equation (32), into the above expression, the Fisher matrix becomes

$$\mathcal{F}_{\alpha\beta} = \sum_{i,j} \frac{\partial \log \bar{P}_i}{\partial \alpha} \frac{\delta_{ij}^K}{\sigma_{\bar{P}}^2(k_i)} \bar{P}_j \frac{\partial \log \bar{P}_j}{\partial \beta} = \sum_i \frac{\partial \log \bar{P}_i}{\partial \alpha} \frac{\partial \log \bar{P}_i}{\partial \beta} \left(\frac{S}{N} \right)_{\text{OPT}}^2(k_i). \quad (55)$$

With the S/N for the optimal weights given by equation (D3), the Fisher matrix in the continuum limit of Fourier modes can be written as

$$\mathcal{F}_{\alpha\beta} = \frac{1}{2} \int \frac{d\mathbf{k}}{(2\pi)^3} \frac{\partial \log P(k)}{\partial \alpha} \frac{\partial \log P(k)}{\partial \beta} V_{\text{eff}}(k), \quad (56)$$

where the effective survey volume has the new form

$$V_{\text{eff}}(k) = \int_0^\infty d\chi \chi^2 \left[\frac{\bar{n}_h(\chi) \bar{b}^2(\chi) P(k)}{(1 + \alpha) + \bar{n}_h(\chi) \bar{b}^2(\chi) P(k)} \right]^2. \quad (57)$$

In Fig. 3, we show the effective survey volume as a function of wavenumber, for a flux-limited survey of similar type to that described in Section 6.2. Here, we consider the three cases where the minimum detectable mass normalization parameter from, cf. equation (53), has the values 5.0×10^{13} , 1.0×10^{14} , $5.0 \times 10^{14} h^{-1} M_\odot$, respectively. In all cases, the optimal weighting increases V_{eff} . Thus, we conclude that the cosmological information extractable from an optimal weighted cluster survey will exceed that from the suboptimal strategies, such as FKP or M+FKP weighting. equation (56) may thus be used as the starting point for exploring the cosmological information content of optimally weighted cluster power spectra (for a review of the Fisher matrix approach, see Heavens 2009).

7 CONCLUSIONS

In this paper, we have developed a formalism to compute weights maximizing the S/N when estimating the cluster power spectrum, and used it to make inferences about the information in the matter power spectrum. Our derivation generalizes the original approach of FKP for galaxies, and is analogous to the derivation by PVP for examining the impact of luminosity dependent galaxy bias on the weights. Our derivation provides for the first time a completely analytic proof for the optimal weight equation, and it also corrects several errors that were found in these earlier works.

In Section 2, we described the generic properties of a cluster survey, taking into account finite survey geometry, arbitrary weighting of position and mass, and a flux limit; we also introduced our statistical treatment for describing the cluster density field using delta function expansions.

In Section 3, we presented estimators for recovering the matter clustering from the computation of the two-point correlation function and the power spectrum of the cluster field. We demonstrated that in order to extract information from the matter clustering one must deconvolve the survey window function. For large homogeneous survey volumes, provided an appropriate choice for the normalization of the cluster field is taken, this estimate was shown to be an unbiased estimator of the dark matter power spectrum, modulo a shot-noise correction.

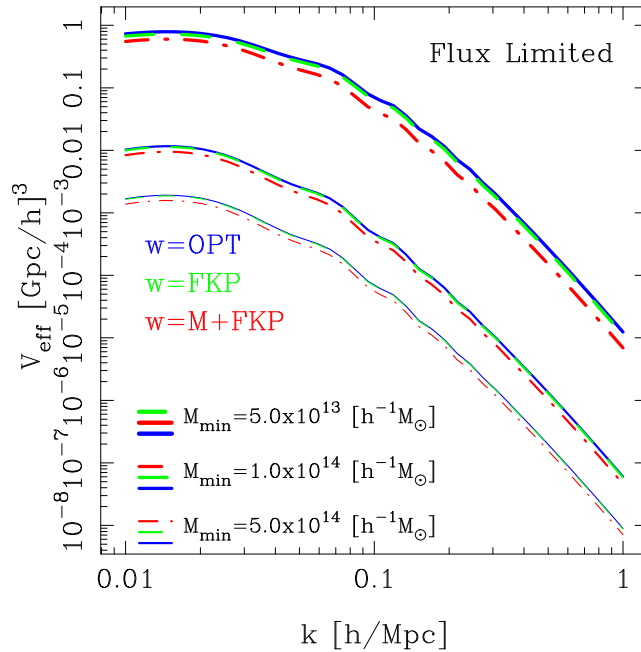


Figure 3. Effective survey volume as a function of the power spectrum wavenumber probed. Lines of decreasing thickness correspond to a minimum detectable mass $M_{\min}(0)$, of 5.0×10^{13} , 1.0×10^{14} , $5.0 \times 10^{14} h^{-1} M_{\odot}$, respectively. The solid, dashed and dot-dashed lines correspond to V_{eff} for optimal, **FKP** and **M+FKP** weighting, respectively.

In Section 4, we explored the statistical fluctuations in the power spectrum of clusters. We derived general expressions for the covariance of the cluster sample, including all non-Gaussian terms arising from the non-linear evolution of matter fluctuations and discreteness effects. This generalized the result of Smith (2009). We then proved the necessary conditions for the covariance matrix to be diagonal.

In Section 5, we have provided an analytic derivation of the optimal weights for a cluster power spectrum analysis. We show in general terms that the optimal weights are separable functions of mass and space.

In Section 6, we presented a comparison of the optimal weighting scheme with the original **FKP** scheme and with a mass-weighting scheme. The latter was advocated by Seljak et al. (2009), later as a linear function of mass by Hamaus et al. (2010), in the context of reducing stochasticity in halo fields. In the limit of a low-mass cluster detection threshold, the study of Cai et al. (2011) also found mass weighting to be optimal for matter density field reconstruction. For the case of both volume- and flux-limited cluster surveys the optimal weighting scheme outperforms both alternate weighting schemes. The gains over the **FKP** method are very modest ~ 5 per cent. Mass weighting performs significantly worse than the optimal scheme, with a relative loss in S/N of ~ 10 per cent on large scales, and of ~ 20 per cent on intermediate scales. Whilst the mass-dependent weighting may be useful for reconstructing the matter field from a cluster sample, we recommend that it should not be used to extract cosmological information from cluster power spectrum analysis. We also presented a new expression for the Fisher information matrix, for an optimally weighted cluster power spectrum measurement.

In this paper, we have derived the optimal weights for measuring the cluster power spectrum under certain conditions, if these conditions are relaxed then the weights are no longer optimal. It will be interesting in future work to explore whether a more general weighting scheme can be derived for the more realistic situations where a non-diagonal covariance matrix is considered. We also expect that the weights that we have derived for the power spectrum should also be used to obtain optimal measurements of the cluster correlation function. However, we have not yet demonstrated this explicitly.

ACKNOWLEDGEMENTS

We thank Ravi Sheth and Simon White for useful discussions. RES acknowledges support from ERC Advanced grant 246797 GALFORMOD. LM thanks MPA for its kind hospitality while this work was being performed. LM was partly supported by the DFG through the grant MA 4967/1-2.

REFERENCES

- Abadi M. G., Lambas D. G., Muriel H., 1998, *ApJ*, 507, 526
 Allen S. W., Evrard A. E., Mantz A. B., 2011, *ARA&A*, 49, 409
 Bahcall N. A., 1988, *ARA&A*, 26, 631
 Bahcall N. A., Soneira R. M., 1983, *ApJ*, 270, 20
 Bahcall N. A., Dong F., Hao L., Bode P., Annis J., Gunn J. E., Schneider D. P., 2003, *ApJ*, 599, 814
 Cai Y.-C., Bernstein G., Sheth R. K., 2011, *MNRAS*, 412, 995

- Cole S., Kaiser N., 1989, MNRAS, 237, 1127
 Collins C. A. et al., 2000, MNRAS, 319, 939
 Cunha C., Huterer D., Doré O., 2010, Phys. Rev. D, 82, 023004
 Dalton G. B., Efstathiou G., Maddox S. J., Sutherland W. J., 1992, ApJ, 390, L1
 Estrada J., Sefusatti E., Frieman J. A., 2009, ApJ, 692, 265
 Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
 Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447
 Hamaus N., Seljak U., Desjacques V., Smith R. E., Baldauf T., 2010, Phys. Rev. D, 82, 043515
 Hauser M. G., Peebles P. J. E., 1973, ApJ, 185, 757
 Heavens A., 2009, preprint (arXiv:0906.0664)
 Hütsi G., 2010, MNRAS, 401, 2477
 Kaiser N., 1984, ApJ, 284, L9
 Koester B. P. et al., 2007, ApJ, 660, 239
 Lima M., Hu W., 2004, Phys. Rev. D, 70, 043504
 Lima M., Hu W., 2005, Phys. Rev. D, 72, 043006
 Majumdar S., Mohr J. J., 2004, ApJ, 613, 41
 Marian L., Bernstein G. M., 2006, Phys. Rev. D, 73, 123525
 Miller C. J., Batuski D. J., 2001, ApJ, 551, 635
 Miller C. J., Nichol R. C., Batuski D. J., 2001, ApJ, 555, 68
 Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
 Mo H. J., Jing Y. P., White S. D. M., 1996, MNRAS, 282, 1096
 Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189
 Oguri M., 2009, Phys. Rev. Lett., 102, 211301
 Oguri M., Takada M., 2011, Phys. Rev. D, 83, 023008
 Percival W. J., Verde L., Peacock J. A., 2004, MNRAS, 347, 645
 Pillepich A., Porciani C., Reiprich T. H., 2012, MNRAS, 422, 44
 Romer A. K., Collins C. A., Böhringer H., Cruddace R. G., Ebeling H., MacGillivray H. T., Voges W., 1994, Nature, 372, 75
 Seljak U., Hamaus N., Desjacques V., 2009, Phys. Rev. Lett., 103, 091303
 Sheth R. K., Lemson G., 1999, MNRAS, 304, 767
 Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
 Smith R. E., 2009, MNRAS, 400, 851
 Smith R. E., Marian L., 2011, MNRAS, 418, 729
 Smith R. E., Watts P. I. R., 2005, MNRAS, 360, 203
 Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512
 Tegmark M., 1997, Phys. Rev. Lett., 79, 3806
 Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22

APPENDIX A: COMPUTING THE TWO-POINT CORRELATIONS

Throughout this section, we shall use the shorthand notation $\langle n_c n'_c \rangle \equiv \langle n_c(\mathbf{x}, M) n_c(\mathbf{x}', M') \rangle$, and similar for $\langle n_s n'_s \rangle$. The evaluation of these terms follows closely that of $\langle \mathcal{F}_c \rangle$. Proceeding with the first term, we write

$$\langle n_c n'_c \rangle = \int \prod_{l=1}^N \{d\mathbf{x}_l dM_l\} p(\mathbf{x}_1, \dots, \mathbf{x}_N, M_1, \dots, M_N) \sum_{i,j} \delta^D(\mathbf{x} - \mathbf{x}_i) \delta^D(M - M_i) \delta^D(\mathbf{x}' - \mathbf{x}_j) \delta^D(M' - M_j).$$

The double sum can be broken up into the terms where $(i \neq j)$ and the terms where $(i = j)$, whereupon

$$\begin{aligned}
 \langle n_c n'_c \rangle &= \int \prod_{l=1}^N \{d\mathbf{x}_l dM_l\} p(\mathbf{x}_1, \dots, \mathbf{x}_N, M_1, \dots, M_N) \left[\sum_{i \neq j} \delta^D(\mathbf{x} - \mathbf{x}_i) \delta^D(M - M_i) \delta^D(\mathbf{x}' - \mathbf{x}_j) \delta^D(M' - M_j) \right. \\
 &\quad \left. + \sum_{i=j} \delta^D(\mathbf{x} - \mathbf{x}_i) \delta^D(\mathbf{x}' - \mathbf{x}_i) \delta^D(M - M_i) \delta^D(M' - M_i) \right] \\
 &= \sum_{i \neq j} p(\mathbf{x}, \mathbf{x}', M, M') + \sum_{i=j} p(\mathbf{x}, M) \delta^D(\mathbf{x} - \mathbf{x}') \delta^D(M - M').
 \end{aligned} \tag{A1}$$

In order to proceed further, we need to specify the joint probability density distribution for obtaining clusters at positions \mathbf{x} and \mathbf{x}' and with masses M and M' . This we do through the introduction of correlation functions:

$$p(\mathbf{x}, \mathbf{x}', M, M') \equiv p(\mathbf{x}, M) p(\mathbf{x}', M') [1 + \xi^c(\mathbf{x}, \mathbf{x}', M, M')] = \frac{\bar{n}(M) \bar{n}(M')}{N^2} [1 + \xi^c(\mathbf{x}, \mathbf{x}', M, M')],$$

where $\xi^c(\mathbf{x}, \mathbf{x}', M, M')$ is the two-point cross-correlation function of clusters with masses M and M' . On use of the above definition in equation (A1), we find

$$\langle n_c n'_c \rangle \approx \bar{n}(M) \bar{n}(M') [1 + \xi^c(\mathbf{x}, \mathbf{x}', M, M')] + \bar{n}(M) \delta^D(\mathbf{x} - \mathbf{x}') \delta^D(M - M'), \tag{A2}$$

where in arriving at the last equality we have assumed that $N \gg 1$ and therefore $N - 1 \approx N$. We may now write down directly the remaining expectation values that enter equation (13), whereupon:

$$\begin{aligned} \langle n_c n'_s \rangle &= \alpha^{-1} \bar{n}(M) \bar{n}(M') ; \\ \langle n_s n'_s \rangle &= \alpha^{-2} \bar{n}(M) \bar{n}(M') + \alpha^{-1} \bar{n}(M) \delta^D(\mathbf{x} - \mathbf{x}') \delta^D(M - M') . \end{aligned} \quad (\text{A3})$$

APPENDIX B: STATISTICAL FLUCTUATIONS IN THE \mathcal{F}_c POWER SPECTRUM

B1 General expression of the covariance of the power spectrum estimator

The covariance of the matter power spectrum estimator given by equation (28) is formed of four terms:

$$\text{Cov}[\hat{P}(\mathbf{k}_1), \hat{P}(\mathbf{k}_2)] = \text{Cov}[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2] + \text{Var}[P_{\text{shot}}] - \text{Cov}[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, P_{\text{shot}}] - \text{Cov}[|\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2, P_{\text{shot}}] , \quad (\text{B1})$$

where

$$\begin{aligned} \text{Cov}[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2] &\equiv \langle |\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2 |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \rangle - \langle |\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2 \rangle \langle |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \rangle ; \\ \text{Var}[P_{\text{shot}}] &\equiv \langle [P_{\text{shot}}]^2 \rangle - \langle P_{\text{shot}} \rangle^2 ; \\ \text{Cov}[|\tilde{\mathcal{F}}_c(\mathbf{k}_i)|^2, P_{\text{shot}}] &\equiv \langle |\tilde{\mathcal{F}}_c(\mathbf{k}_i)|^2 P_{\text{shot}} \rangle - \langle |\tilde{\mathcal{F}}_c(\mathbf{k}_i)|^2 \rangle \langle P_{\text{shot}} \rangle ; \quad i \in (1, 2) . \end{aligned} \quad (\text{B2})$$

Assuming that the total number density of clusters is a deterministic quantity, all of the covariance terms involving P_{shot} vanish. Thus, we are left with the task of determining the covariance of the modulus square of the $\tilde{\mathcal{F}}_c$ field, i.e. equation (30).

B2 Derivation of the covariance of the power spectrum estimator

For a succinct presentation of the covariance calculation, we make use of the following shorthand notation:

$$\bar{n}_i \equiv \bar{n}(M_i) ; w_i \equiv w(\mathbf{x}_i, M_i) ; b_i \equiv b(M_i) ; \Theta_i \equiv \Theta(\mathbf{x}_i | M_i) ; n_{c,i} \equiv n_c(\mathbf{x}_i, M_i) ; n_{s,i} \equiv n_s(\mathbf{x}_i, M_i) ; \delta_{ij}^D \equiv \delta^D(\mathbf{x}_i - \mathbf{x}_j) \delta^D(M_i - M_j) . \quad (\text{B3})$$

The covariance matrix of the power in the field \mathcal{F}_c can be written as

$$\text{Cov}[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2] = \int d\mathbf{k}_3 d\mathbf{k}_4 \delta^D(\mathbf{k}_1 + \mathbf{k}_3) \delta^D(\mathbf{k}_2 + \mathbf{k}_4) \left[\langle \tilde{\mathcal{F}}_c(\mathbf{k}_1) \dots \tilde{\mathcal{F}}_c(\mathbf{k}_4) \rangle - \langle \tilde{\mathcal{F}}_c(\mathbf{k}_1) \tilde{\mathcal{F}}_c(\mathbf{k}_3) \rangle \langle \tilde{\mathcal{F}}_c(\mathbf{k}_2) \tilde{\mathcal{F}}_c(\mathbf{k}_4) \rangle \right] \quad (\text{B4})$$

In order to proceed further, we see that we must compute the four-point correlation function of the Fourier modes $\tilde{\mathcal{F}}_c(\mathbf{k})$. This is equivalent to specifying the four-point correlation function of the field \mathcal{F}_c :

$$\langle \tilde{\mathcal{F}}_c(\mathbf{k}_1) \dots \tilde{\mathcal{F}}_c(\mathbf{k}_4) \rangle = \int d\mathbf{x}_1 \dots d\mathbf{x}_4 \langle \mathcal{F}_c(\mathbf{x}_1) \dots \mathcal{F}_c(\mathbf{x}_4) \rangle e^{ik_1 \cdot \mathbf{x}_1 + \dots + k_4 \cdot \mathbf{x}_4} . \quad (\text{B5})$$

Substituting equation (6) into the expression for the four-point correlation function, we find:

$$\langle \mathcal{F}_c(\mathbf{x}_1) \dots \mathcal{F}_c(\mathbf{x}_4) \rangle = \frac{1}{A^2} \int \prod_{i=1}^4 \{ dM_i w(\mathbf{x}_i, M_i) \Theta(\mathbf{x}_i | M_i) \} \langle [n_c(\mathbf{x}_1, M_1) - \alpha n_s(\mathbf{x}_1, M_1)] \dots [n_c(\mathbf{x}_4, M_4) - \alpha n_s(\mathbf{x}_4, M_4)] \rangle . \quad (\text{B6})$$

Expanding the term in angled braces on the right-hand side gives

$$\begin{aligned} \langle [n_{c,1} - \alpha n_{s,1}] \dots [n_{c,4} - \alpha n_{s,4}] \rangle &= \langle n_{c,1} \dots n_{c,4} \rangle - \alpha \left[\langle n_{c,1} n_{c,2} n_{c,3} n_{s,4} \rangle + 3\text{cyc} \right] + \alpha^2 \left[\langle n_{c,1} n_{c,2} n_{s,3} n_{s,4} \rangle + 5\text{perm} \right] \\ &\quad - \alpha^3 \left[\langle n_{c,1} n_{s,2} n_{s,3} n_{s,4} \rangle + 3\text{cyc} \right] + \alpha^4 \langle n_{s,1} \dots n_{s,4} \rangle . \end{aligned} \quad (\text{B7})$$

We focus on the first term in curly brackets on the right-hand side, and insert our delta function expansion for the number density field:

$$\begin{aligned} \langle n_{c,1} \dots n_{c,4} \rangle &= \left\langle \sum_{i_1, i_2, i_3, i_4=1}^N \delta^D(\mathbf{x}_1 - \mathbf{x}'_{i_1}) \dots \delta^D(\mathbf{x}_4 - \mathbf{x}'_{i_4}) \delta^D(M_1 - M'_{i_1}) \dots \delta^D(M_4 - M'_{i_4}) \right\rangle \\ &= \int \prod_{p=1}^N \{ d\mathbf{x}'_p dM'_p \} p(\mathbf{x}'_1, \dots, \mathbf{x}'_N, M'_1, \dots, M'_N) \{ \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4 + \Gamma_5 \} , \end{aligned} \quad (\text{B8})$$

where in the above we have defined the following terms:

$$\begin{aligned}
 \Gamma_1 &\equiv \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \delta^D(\mathbf{x}_1 - \mathbf{x}'_{i_1}) \dots \delta^D(\mathbf{x}_4 - \mathbf{x}'_{i_4}) \delta^D(M_1 - M'_{i_1}) \dots \delta^D(M_4 - M'_{i_4}); \\
 \Gamma_2 &\equiv \sum_{i_1 \neq i_2 \neq i_3 = i_4} \delta^D(\mathbf{x}_1 - \mathbf{x}'_{i_1}) \delta^D(\mathbf{x}_2 - \mathbf{x}'_{i_2}) \delta^D(M_1 - M'_{i_1}) \delta^D(M_2 - M'_{i_2}) \prod_{p=3}^4 \{ \delta^D(\mathbf{x}_p - \mathbf{x}'_{i_3}) \delta^D(M_p - M'_{i_3}) \} + 5 \text{ perms}; \\
 \Gamma_3 &\equiv \sum_{i_1 = i_2 \neq i_3 = i_4} \prod_{p=1}^2 \{ \delta^D(\mathbf{x}_p - \mathbf{x}'_{i_1}) \delta^D(M_p - M'_{i_1}) \} \prod_{q=3}^4 \{ \delta^D(\mathbf{x}_q - \mathbf{x}'_{i_3}) \delta^D(M_q - M'_{i_3}) \} + 2 \text{ perms}; \\
 \Gamma_4 &\equiv \sum_{i_1 \neq i_2 = i_3 = i_4} \delta^D(\mathbf{x}_1 - \mathbf{x}'_{i_1}) \delta^D(M_1 - M'_{i_1}) \prod_{p=2}^4 \{ \delta^D(\mathbf{x}_p - \mathbf{x}'_{i_2}) \delta^D(M_p - M'_{i_2}) \} + 3 \text{ perms}; \\
 \Gamma_5 &\equiv \sum_{i_1 = i_2 = i_3 = i_4} \prod_{p=1}^4 \{ \delta^D(\mathbf{x}_p - \mathbf{x}'_{i_1}) \delta^D(M_p - M'_{i_1}) \}. \tag{B9}
 \end{aligned}$$

We now compute the contributions from each of the five Γ terms.

(i) *Computing $\langle \Gamma_1 \rangle$.* On performing the integrations over the delta functions, we have

$$\begin{aligned}
 \langle \Gamma_1 \rangle &= \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} p(\mathbf{x}_1, \dots, \mathbf{x}_4, M_1, \dots, M_4) \\
 &= N(N-1)(N-2)(N-3)p(\mathbf{x}_1, M_1) \dots p(\mathbf{x}_4, M_4) \left[1 + \xi_{12}^c + \xi_{13}^c + \xi_{14}^c + \xi_{23}^c + \xi_{24}^c + \xi_{34}^c \right. \\
 &\quad \left. + \zeta_{123}^c + \zeta_{234}^c + \zeta_{341}^c + \zeta_{412}^c + \xi_{12}^c \xi_{34}^c + \xi_{13}^c \xi_{24}^c + \xi_{14}^c \xi_{23}^c + \eta_{1234}^c \right] \\
 &\approx \prod_{p=1}^4 \{ \bar{n}_p \} \left[1 + \xi_{12}^c + \xi_{13}^c + \xi_{14}^c + \xi_{23}^c + \xi_{24}^c + \xi_{34}^c + \zeta_{123}^c + \zeta_{234}^c + \zeta_{341}^c + \zeta_{412}^c + \xi_{12}^c \xi_{34}^c + \xi_{13}^c \xi_{24}^c + \xi_{14}^c \xi_{23}^c + \eta_{1234}^c \right], \tag{B10}
 \end{aligned}$$

where the last equality holds in the limit that $N \gg 1$. The terms ζ^c and η^c are the three- and four-point correlation functions of the haloes, respectively, and in the above we have also made use of the short-hand notation:

$$\xi_{ij}^c \equiv \xi^c(\mathbf{r}_i, \mathbf{r}_j, M_i, M_j); \zeta_{ijk}^c \equiv \zeta^c(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, M_i, M_j, M_k); \eta_{ijkl}^c \equiv \eta^c(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_l, M_i, M_j, M_k, M_l). \tag{B11}$$

(ii) *Computing $\langle \Gamma_2 \rangle$.* On integration over the delta functions and using equation (B3), we obtain

$$\begin{aligned}
 \langle \Gamma_2 \rangle &\equiv \sum_{i_1 \neq i_2 \neq i_3 = i_4} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, M_1, M_2, M_3) \delta_{34}^D + \sum_{i_1 \neq i_2 = i_3 \neq i_4} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4, M_1, M_2, M_4) \delta_{23}^D + \sum_{i_1 = i_2 \neq i_3 \neq i_4} p(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, M_1, M_3, M_4) \delta_{12}^D \\
 &\quad + \sum_{i_1 = i_3 \neq i_2 \neq i_4} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4, M_1, M_2, M_4) \delta_{13}^D + \sum_{i_1 = i_4 \neq i_2 \neq i_3} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, M_1, M_2, M_3) \delta_{14}^D + \sum_{i_2 = i_4 \neq i_1 \neq i_3} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, M_1, M_2, M_3) \delta_{24}^D.
 \end{aligned}$$

On performing the summations and taking the limit of large numbers, the above expression reduces to

$$\begin{aligned}
 \langle \Gamma_2 \rangle &= \prod_{p \in \{1,3,4\}} \{ \bar{n}_p \} [1 + \xi_{13}^c + \xi_{14}^c + \xi_{34}^c + \zeta_{134}^c] \delta_{12}^D + \prod_{p \in \{1,2,4\}} \{ \bar{n}_p \} [1 + \xi_{12}^c + \xi_{24}^c + \xi_{24}^c + \zeta_{124}^c] \delta_{23}^D + \prod_{p \in \{1,2,3\}} \{ \bar{n}_p \} [1 + \xi_{12}^c + \xi_{13}^c + \xi_{23}^c + \zeta_{123}^c] \delta_{14}^D \\
 &\quad + \prod_{p \in \{1,2,4\}} \{ \bar{n}_p \} [1 + \xi_{12}^c + \xi_{14}^c + \xi_{24}^c + \zeta_{124}^c] \delta_{23}^D + \prod_{p \in \{1,2,3\}} \{ \bar{n}_p \} [1 + \xi_{12}^c + \xi_{23}^c + \xi_{31}^c + \zeta_{123}^c] \delta_{24}^D \\
 &\quad + \prod_{p \in \{1,2,3\}} \{ \bar{n}_p \} [1 + \xi_{12}^c + \xi_{23}^c + \xi_{31}^c + \zeta_{123}^c] \delta_{34}^D. \tag{B12}
 \end{aligned}$$

(iii) *Computing $\langle \Gamma_3 \rangle$.* On performing the integrations over the delta functions, we have

$$\begin{aligned}
 \langle \Gamma_3 \rangle &\equiv \sum_{i_1 = i_2 \neq i_3 = i_4} p(\mathbf{x}_1, \mathbf{x}_3, M_1, M_3) \delta_{12}^D \delta_{34}^D + \sum_{i_1 = i_3 \neq i_2 = i_4} p(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) \delta_{13}^D \delta_{24}^D + \sum_{i_1 = i_4 \neq i_2 = i_3} p(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) \delta_{14}^D \delta_{23}^D \\
 &= \bar{n}_1 \bar{n}_3 [1 + \xi_{13}^c] \delta_{12}^D \delta_{34}^D + \bar{n}_1 \bar{n}_2 [1 + \xi_{12}^c] \left[\delta_{13}^D \delta_{24}^D + \delta_{14}^D \delta_{23}^D \right], \tag{B13}
 \end{aligned}$$

where to arrive at the second equality, we performed the summations in the limit of a large number of clusters.

(iv) *Computing $\langle \Gamma_4 \rangle$.* We perform the integrations over the delta functions to obtain

$$\begin{aligned}
 \langle \Gamma_4 \rangle &\equiv \sum_{i_1 = i_2 = i_3 \neq i_4} p(\mathbf{x}_1, \mathbf{x}_4, M_1, M_4) \delta_{12}^D \delta_{13}^D + \sum_{i_1 = i_2 = i_4 \neq i_3} p(\mathbf{x}_1, \mathbf{x}_3, M_1, M_3) \delta_{12}^D \delta_{14}^D \\
 &\quad + \sum_{i_1 = i_3 = i_4 \neq i_2} p(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) \delta_{13}^D \delta_{14}^D + \sum_{i_2 = i_3 = i_4 \neq i_1} p(\mathbf{x}_1, \mathbf{x}_2, M_1, M_2) \delta_{23}^D \delta_{24}^D.
 \end{aligned}$$

Summing and taking the limit of large numbers of clusters we find

$$\langle \Gamma_4 \rangle = \bar{n}_1 \bar{n}_4 [1 + \xi_{14}^c] \delta_{12}^D \delta_{13}^D + \bar{n}_1 \bar{n}_3 [1 + \xi_{13}^c] \delta_{12}^D \delta_{14}^D + \bar{n}_1 \bar{n}_2 [1 + \xi_{12}^c] \{ \delta_{13}^D \delta_{14}^D + \delta_{23}^D \delta_{24}^D \}. \quad (\text{B14})$$

(v) *Computing* $\langle \Gamma_5 \rangle$. Integrating over the delta functions leads to

$$\langle \Gamma_5 \rangle \equiv \sum_{i_1=i_2=i_3=i_4} p(\mathbf{x}_1, M_1) \delta_{12}^D \delta_{13}^D \delta_{14}^D = \bar{n}_1 \delta_{12}^D \delta_{13}^D \delta_{14}^D. \quad (\text{B15})$$

Putting together the terms equations (B10)–(B15), we arrive at the expression

$$\begin{aligned} \langle n_{c,1} \dots n_{c,4} \rangle &= \prod_{p=1}^4 \{ \bar{n}_p \} \left\{ \left[1 + \xi_{12}^c + \xi_{13}^c + \xi_{14}^c + \xi_{23}^c + \xi_{24}^c + \xi_{34}^c + \zeta_{123}^c + \zeta_{234}^c + \zeta_{341}^c + \zeta_{412}^c + \xi_{12}^c \xi_{34}^c + \xi_{13}^c \xi_{24}^c + \xi_{14}^c \xi_{23}^c + \eta_{1234}^c \right] \right. \\ &+ \left[1 + \xi_{13}^c + \xi_{14}^c + \xi_{34}^c + \zeta_{134}^c \right] \frac{\delta_{12}^D}{\bar{n}_2} + \left[1 + \xi_{12}^c + \xi_{24}^c + \xi_{41}^c + \zeta_{124}^c \right] \frac{\delta_{23}^D}{\bar{n}_3} + \left[1 + \xi_{12}^c + \xi_{13}^c + \xi_{23}^c + \zeta_{123}^c \right] \frac{\delta_{14}^D}{\bar{n}_4} \\ &+ \left[1 + \xi_{12}^c + \xi_{14}^c + \xi_{24}^c + \zeta_{124}^c \right] \frac{\delta_{13}^D}{\bar{n}_3} + \left[1 + \xi_{12}^c + \xi_{23}^c + \xi_{31}^c + \zeta_{123}^c \right] \frac{\delta_{r,24}^D}{\bar{n}_4} + \left[1 + \xi_{12}^c + \xi_{23}^c + \xi_{31}^c + \zeta_{123}^c \right] \frac{\delta_{34}^D}{\bar{n}_4} \\ &+ \left[1 + \xi_{13}^c \right] \frac{\delta_{12}^D \delta_{34}^D}{\bar{n}_2 \bar{n}_4} + \left[1 + \xi_{12}^c \right] \frac{\delta_{13}^D \delta_{24}^D}{\bar{n}_3 \bar{n}_4} + \left[1 + \xi_{12}^c \right] \frac{\delta_{14}^D \delta_{23}^D}{\bar{n}_3 \bar{n}_4} + \left[1 + \xi_{14}^c \right] \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} \\ &+ \left. \left[1 + \xi_{13}^c \right] \frac{\delta_{12}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_4} + \left[1 + \xi_{12}^c \right] \frac{\delta_{13}^D \delta_{14}^D + \delta_{23}^D \delta_{24}^D}{\bar{n}_3 \bar{n}_4} + \frac{\delta_{12}^D \delta_{13}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_3 \bar{n}_4} \right\}. \quad (\text{B16}) \end{aligned}$$

In the above ξ_{ij}^c , ζ_{ijkl}^c and η_{ijkl}^c represent the true three- and four-point connected correlation functions of galaxy clusters, respectively. Under the assumption of linear biasing these may be written in terms of the connected correlation functions of the matter as

$$\xi_{ij}^c \equiv \xi_c(\mathbf{x}_i, \mathbf{x}_j, M_i, M_j) = b(M_i) b(M_j) \xi_{ij} = b_i b_j \xi_{ij};$$

$$\zeta_{ijk}^c \equiv \zeta_c(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, M_i, M_j, M_k) = b(M_i) b(M_j) b(M_k) \zeta_{ijk} = b_i b_j b_k \zeta_{ijk};$$

$$\eta_{1234}^c \equiv \eta_c(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, M_1, M_2, M_3, M_4) = b(M_1) b(M_2) b(M_3) b(M_4) \eta_{1234} = b_1 b_2 b_3 b_4 \eta_{1234}.$$

Similar to the result of equation (B16), we can also write down the terms entering equation (B7). Thus we obtain

$$\begin{aligned} \langle n_{c,1} n_{c,2} n_{c,3} n_{s,4} \rangle &\equiv \alpha^{-1} \prod_{p=1}^4 \{ \bar{n}_p \} \left\{ \left[1 + \xi_{12}^c + \xi_{13}^c + \xi_{23}^c + \zeta_{123}^c \right] + \left[1 + \xi_{13}^c \right] \frac{\delta_{12}^D}{\bar{n}_2} + \left[1 + \xi_{12}^c \right] \frac{\delta_{23}^D + \delta_{13}^D}{\bar{n}_3} + \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} \right\}; \\ \langle n_{c,1} n_{c,2} n_{s,3} n_{s,4} \rangle &\equiv \alpha^{-2} \prod_{p=1}^4 \{ \bar{n}_p \} \left\{ \left[1 + \xi_{12}^c \right] + \frac{\delta_{12}^D}{\bar{n}_2} + \alpha \left[1 + \xi_{12}^c \right] \frac{\delta_{34}^D}{\bar{n}_4} + \alpha \frac{\delta_{12}^D \delta_{34}^D}{\bar{n}_2 \bar{n}_4} \right\}; \\ \langle n_{c,1} n_{s,2} n_{s,3} n_{s,4} \rangle &\equiv \alpha^{-3} \prod_{p=1}^4 \{ \bar{n}_p \} \left\{ 1 + \alpha \left[\frac{\delta_{23}^D}{\bar{n}_3} + \frac{\delta_{24}^D}{\bar{n}_4} + \frac{\delta_{34}^D}{\bar{n}_4} \right] + \alpha^2 \frac{\delta_{23}^D \delta_{24}^D}{\bar{n}_3 \bar{n}_4} \right\}; \\ \langle n_{s,1} n_{s,2} n_{s,3} n_{s,4} \rangle &\equiv \alpha^{-4} \prod_{p=1}^4 \{ \bar{n}_p \} \left\{ 1 + \alpha \left[\frac{\delta_{12}^D}{\bar{n}_2} + \frac{\delta_{13}^D + \delta_{23}^D}{\bar{n}_3} + \frac{\delta_{14}^D + \delta_{24}^D + \delta_{34}^D}{\bar{n}_4} \right] + \alpha^2 \left[\frac{\delta_{12}^D \delta_{14}^D + \delta_{12}^D \delta_{34}^D}{\bar{n}_2 \bar{n}_4} + \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} \right. \right. \\ &+ \left. \left. \frac{\delta_{13}^D \delta_{24}^D + \delta_{14}^D \delta_{23}^D + \delta_{13}^D \delta_{14}^D + \delta_{23}^D \delta_{24}^D}{\bar{n}_3 \bar{n}_4} \right] + \alpha^3 \frac{\delta_{12}^D \delta_{13}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_3 \bar{n}_4} \right\}. \quad (\text{B17}) \end{aligned}$$

Hence, for the second term in equation (B7) we have

$$\begin{aligned} \langle n_{c,1} n_{c,2} n_{c,3} n_{s,4} \rangle + 3\text{perms} &= \alpha^{-1} \prod_{p=1}^4 \{ \bar{n}_p \} \left[4 + 2 \left(\xi_{12}^c + \xi_{13}^c + \xi_{23}^c + \xi_{14}^c + \xi_{24}^c + \xi_{34}^c \right) + \zeta_{123}^c + \zeta_{124}^c + \zeta_{134}^c + \zeta_{234}^c \right] \\ &+ \left[1 + \xi_{13}^c \right] \frac{\delta_{12}^D}{\bar{n}_2} + \left[1 + \xi_{12}^c \right] \frac{\delta_{23}^D}{\bar{n}_3} + \left[1 + \xi_{12}^c \right] \frac{\delta_{31}^D}{\bar{n}_3} + \left[1 + \xi_{14}^c \right] \frac{\delta_{12}^D}{\bar{n}_2} + \left[1 + \xi_{12}^c \right] \frac{\delta_{24}^D}{\bar{n}_4} \\ &+ \left[1 + \xi_{12}^c \right] \frac{\delta_{41}^D}{\bar{n}_4} + \left[1 + \xi_{14}^c \right] \frac{\delta_{13}^D}{\bar{n}_3} + \left[1 + \xi_{13}^c \right] \frac{\delta_{14}^D}{\bar{n}_4} + \left[1 + \xi_{13}^c \right] \frac{\delta_{34}^D}{\bar{n}_4} + \left[1 + \xi_{24}^c \right] \frac{\delta_{23}^D}{\bar{n}_3} \\ &+ \left[1 + \xi_{23}^c \right] \frac{\delta_{24}^D}{\bar{n}_4} + \left[1 + \xi_{23}^c \right] \frac{\delta_{34}^D}{\bar{n}_4} + \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} + \frac{\delta_{12}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_4} + \frac{\delta_{13}^D \delta_{14}^D + \delta_{23}^D \delta_{24}^D}{\bar{n}_3 \bar{n}_4}. \quad (\text{B18}) \end{aligned}$$

For the third term in equation (B7), we obtain

$$\begin{aligned} \langle n_{c,1} n_{c,2} n_{s,3} n_{s,4} \rangle + 5 \text{ perms} &= \alpha^{-2} \prod_{p=1}^4 \{\bar{n}_p\} \left\{ \left[6 + \xi_{12}^c + \xi_{13}^c + \xi_{14}^c + \xi_{23}^c + \xi_{24}^c + \xi_{34}^c \right] + \frac{\delta_{12}^D}{\bar{n}_2} + \frac{\delta_{13}^D + \delta_{23}^D}{\bar{n}_3} + \frac{\delta_{14}^D + \delta_{24}^D + \delta_{34}^D}{\bar{n}_4} \right. \\ &\quad + \alpha \left\{ \left[1 + \xi_{12}^c \right] \frac{\delta_{34}^D}{\bar{n}_4} + \left[1 + \xi_{13}^c \right] \frac{\delta_{24}^D}{\bar{n}_4} + \left[1 + \xi_{23}^c \right] \frac{\delta_{14}^D}{\bar{n}_4} + \left[1 + \xi_{14}^c \right] \frac{\delta_{23}^D}{\bar{n}_3} + \left[1 + \xi_{24}^c \right] \frac{\delta_{13}^D}{\bar{n}_3} \right. \\ &\quad \left. \left. + \left[1 + \xi_{34}^c \right] \frac{\delta_{12}^D}{\bar{n}_2} + 2 \frac{\delta_{12}^D \delta_{34}^D}{\bar{n}_2 \bar{n}_4} + 2 \frac{\delta_{13}^D \delta_{24}^D + \delta_{14}^D \delta_{23}^D}{\bar{n}_3 \bar{n}_4} \right\} \right\}. \end{aligned} \quad (\text{B19})$$

For the fourth term in equation (B7), we write

$$\langle n_{c,1} n_{s,2} n_{s,3} n_{s,4} \rangle + 3 \text{ perms} = \alpha^{-3} \prod_{p=1}^4 \{\bar{n}_p\} \left\{ 4 + 2\alpha \left[\frac{\delta_{12}^D}{\bar{n}_2} + \frac{\delta_{13}^D + \delta_{23}^D}{\bar{n}_3} + \frac{\delta_{14}^D + \delta_{24}^D + \delta_{34}^D}{\bar{n}_4} \right] + \alpha^2 \left[\frac{\delta_{13}^D \delta_{14}^D + \delta_{23}^D \delta_{24}^D}{\bar{n}_3 \bar{n}_4} + \frac{\delta_{12}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_4} + \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} \right] \right\}. \quad (\text{B20})$$

On collecting all of the terms and after a little algebra, we find that equation (B6) can be written as

$$\begin{aligned} \langle \mathcal{F}_c(\mathbf{x}_1) \dots \mathcal{F}_c(\mathbf{x}_4) \rangle &= \frac{1}{A^2} \prod_{p=1}^4 \left(\int dM_p w_p \bar{n}_p \Theta_p \right) \left\{ \eta_{1234}^c + \zeta_{134}^c \frac{\delta_{12}^D}{\bar{n}_2} + \zeta_{124}^c \frac{\delta_{23}^D}{\bar{n}_3} + \zeta_{123}^c \frac{\delta_{14}^D}{\bar{n}_4} + \zeta_{124}^c \frac{\delta_{13}^D}{\bar{n}_3} + \zeta_{123}^c \frac{\delta_{24}^D}{\bar{n}_4} \right. \\ &\quad + \zeta_{123}^c \frac{\delta_{34}^D}{\bar{n}_4} + \left[\xi_{12}^c + \frac{(1+\alpha)}{\bar{n}_2} \delta_{12}^D \right] \left[\xi_{34}^c + \frac{(1+\alpha)}{\bar{n}_4} \delta_{34}^D \right] + \left[\xi_{13}^c + \frac{(1+\alpha)}{\bar{n}_3} \delta_{13}^D \right] \left[\xi_{24}^c + \frac{(1+\alpha)}{\bar{n}_4} \delta_{24}^D \right] + \left[\xi_{14}^c + \frac{(1+\alpha)}{\bar{n}_4} \delta_{14}^D \right] \\ &\quad \left. \times \left[\xi_{23}^c + \frac{(1+\alpha)}{\bar{n}_3} \delta_{23}^D \right] + \xi_{12}^c \frac{(\delta_{13}^D + \delta_{23}^D)(\delta_{14}^D + \delta_{24}^D)}{\bar{n}_3 \bar{n}_4} + \xi_{13}^c \frac{\delta_{12}^D (\delta_{14}^D + \delta_{34}^D)}{\bar{n}_2 \bar{n}_4} + \xi_{14}^c \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} + \frac{(1+\alpha^3) \delta_{12}^D \delta_{13}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_3 \bar{n}_4} \right\}. \end{aligned} \quad (\text{B21})$$

The last factor on the right-hand side of equation (B4) can be written as

$$\langle \mathcal{F}_c(\mathbf{x}_1) \mathcal{F}_c(\mathbf{x}_3) \rangle \langle \mathcal{F}_c(\mathbf{x}_2) \mathcal{F}_c(\mathbf{x}_4) \rangle = \prod_{p=1}^4 \left(\int dM_p w_p \bar{n}_p \Theta_p \right) \left[\xi_{13}^c + \frac{(1+\alpha) \delta_{13}^D}{\bar{n}_3} \right] \left[\xi_{24}^c + \frac{(1+\alpha) \delta_{24}^D}{\bar{n}_4} \right]. \quad (\text{B22})$$

Subtracting equation (B22) from equation (B21), the argument of the bracket on the right-hand side of equation (B4) can be expressed as

$$\begin{aligned} \langle \mathcal{F}_c(\mathbf{x}_1) \dots \mathcal{F}_c(\mathbf{x}_4) \rangle - \langle \mathcal{F}_c(\mathbf{x}_1) \mathcal{F}_c(\mathbf{x}_3) \rangle \langle \mathcal{F}_c(\mathbf{x}_2) \mathcal{F}_c(\mathbf{x}_4) \rangle &= \frac{1}{A^2} \prod_{p=1}^4 \left(\int dM_p w_p \bar{n}_p \Theta_p \right) \left\{ \eta_{1234}^c + \zeta_{134}^c \frac{\delta_{12}^D}{\bar{n}_2} + \zeta_{124}^c \frac{\delta_{23}^D}{\bar{n}_3} + \zeta_{123}^c \frac{\delta_{14}^D}{\bar{n}_4} \right. \\ &\quad + \zeta_{124}^c \frac{\delta_{13}^D}{\bar{n}_3} + \zeta_{123}^c \frac{\delta_{24}^D}{\bar{n}_4} + \zeta_{123}^c \frac{\delta_{34}^D}{\bar{n}_4} + \left[\xi_{12}^c + \frac{(1+\alpha)}{\bar{n}_2} \delta_{12}^D \right] \left[\xi_{34}^c + \frac{(1+\alpha)}{\bar{n}_4} \delta_{34}^D \right] + \left[\xi_{14}^c + \frac{(1+\alpha)}{\bar{n}_4} \delta_{14}^D \right] \left[\xi_{23}^c + \frac{(1+\alpha)}{\bar{n}_3} \delta_{23}^D \right] \\ &\quad \left. + \xi_{12}^c \frac{(\delta_{13}^D + \delta_{23}^D)(\delta_{14}^D + \delta_{24}^D)}{\bar{n}_3 \bar{n}_4} + \xi_{13}^c \frac{\delta_{12}^D (\delta_{14}^D + \delta_{34}^D)}{\bar{n}_2 \bar{n}_4} + \xi_{14}^c \frac{\delta_{12}^D \delta_{13}^D}{\bar{n}_2 \bar{n}_3} + \frac{(1+\alpha^3) \delta_{12}^D \delta_{13}^D \delta_{14}^D}{\bar{n}_2 \bar{n}_3 \bar{n}_4} \right\}. \end{aligned} \quad (\text{B23})$$

B3 General expression for the covariance matrix of the power spectrum of \mathcal{F}_c

Consider equations (31) and (B23), we may Fourier transform all of the space-dependent terms. For the case of the n -point correlation functions, these are Fourier dual with the n -point multispectra

$$\begin{aligned} \xi_{12} &\equiv \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} (2\pi)^3 \delta^D(\mathbf{q}_1 + \mathbf{q}_2) P(\mathbf{q}_1, \mathbf{q}_2) \exp[-i\mathbf{q}_1 \cdot \mathbf{x}_1 - i\mathbf{q}_2 \cdot \mathbf{x}_2]; \\ \zeta_{123} &\equiv \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} \frac{d\mathbf{q}_3}{(2\pi)^3} (2\pi)^3 \delta^D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) B(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \exp[-i\mathbf{q}_1 \cdot \mathbf{x}_1 - i\mathbf{q}_2 \cdot \mathbf{x}_2 - i\mathbf{q}_3 \cdot \mathbf{x}_3]; \\ \eta_{1234} &\equiv \int \frac{d\mathbf{q}_1}{(2\pi)^3} \dots \frac{d\mathbf{q}_4}{(2\pi)^3} (2\pi)^3 \delta^D(\mathbf{q}_1 + \dots + \mathbf{q}_4) T(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) \exp[-i\mathbf{q}_1 \cdot \mathbf{x}_1 \dots - i\mathbf{q}_4 \cdot \mathbf{x}_4], \end{aligned} \quad (\text{B24})$$

where B and T are bispectrum and trispectrum, respectively. Note that, owing to the Dirac delta function in the above expressions the bispectrum and trispectrum are in fact functions of two and three k -vectors, respectively. Using these relations in equations (31) and (B23), we find that the covariance matrix may be written in general as

$$\begin{aligned} \text{Cov}[|\mathcal{F}_c(\mathbf{k}_1)|^2, |\mathcal{F}_c(\mathbf{k}_2)|^2] &= \int \prod_{p=1}^3 \left(\frac{d\mathbf{q}_p}{(2\pi)^3} \right) T(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 - \mathbf{q}_2) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_1 - \mathbf{q}_3) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \\ &\quad + \left| \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 + \mathbf{q}_1) + (1+\alpha) \tilde{\mathcal{G}}_{(2,0)}(\mathbf{k}_1 + \mathbf{k}_2) \right|^2 \end{aligned}$$

$$\begin{aligned}
& + \left| \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1) + (1 + \alpha) \tilde{\mathcal{G}}_{(2,0)}(\mathbf{k}_1 - \mathbf{k}_2) \right|^2 \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} B(\mathbf{q}_1, \mathbf{q}_2) \tilde{\mathcal{G}}_{(2,1)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_1 - \mathbf{q}_2) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2) \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} B(\mathbf{q}_1, \mathbf{q}_2) \tilde{\mathcal{G}}_{(2,1)}(\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(+\mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2) \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} B(\mathbf{q}_1, \mathbf{q}_2) \tilde{\mathcal{G}}_{(2,1)}(-\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2) \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} B(\mathbf{q}_1, \mathbf{q}_2) \tilde{\mathcal{G}}_{(2,1)}(-\mathbf{k}_2 - \mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2) \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} B(\mathbf{q}_1, \mathbf{q}_2) \tilde{\mathcal{G}}_{(2,1)}(-\mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_1 + \mathbf{q}_1 + \mathbf{q}_2) \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} \frac{d\mathbf{q}_2}{(2\pi)^3} B(\mathbf{q}_1, \mathbf{q}_2) \tilde{\mathcal{G}}_{(2,1)}(-\mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1 + \mathbf{q}_2) \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \left\{ \left| \tilde{\mathcal{G}}_{(2,1)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1) \right|^2 + \left| \tilde{\mathcal{G}}_{(2,1)}(\mathbf{q}_1) \right|^2 + \left| \tilde{\mathcal{G}}_{(2,1)}(\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{q}_1) \right|^2 \right\} \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \left\{ \tilde{\mathcal{G}}_{(3,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_1 + \mathbf{q}_1) + \tilde{\mathcal{G}}_{(3,1)}(\mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1) + \right. \\
& \quad \left. \tilde{\mathcal{G}}_{(3,1)}(-\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 + \mathbf{q}_1) + \tilde{\mathcal{G}}_{(3,1)}(-\mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 + \mathbf{q}_1) \right\} + (1 + \alpha^3) \tilde{\mathcal{G}}_{(4,0)}(0). \tag{B25}
\end{aligned}$$

B4 The Gaussian approximation of the covariance matrix of the power spectrum of \mathcal{F}_c

Under the assumption that the matter density field is Gaussianly distributed, all connected correlation functions beyond two-point ones vanish (e.g. ζ and η from equation B24). The general expression given by equation (B25) simplifies to

$$\begin{aligned}
\text{Cov} \left[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \right] & = \left| \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 + \mathbf{q}_1) + (1 + \alpha) \tilde{\mathcal{G}}_{(2,0)}(\mathbf{k}_1 + \mathbf{k}_2) \right|^2 \\
& + \left| \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1) + (1 + \alpha) \tilde{\mathcal{G}}_{(2,0)}(\mathbf{k}_1 - \mathbf{k}_2) \right|^2 \\
& + \int \frac{d\mathbf{q}_1}{(2\pi)^3} P(\mathbf{q}_1) \left\{ \left| \tilde{\mathcal{G}}_{(2,1)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{q}_1) \right|^2 + \left| \tilde{\mathcal{G}}_{(2,1)}(\mathbf{q}_1) \right|^2 + \left| \tilde{\mathcal{G}}_{(2,1)}(\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{q}_1) \right|^2 \right. \\
& + \tilde{\mathcal{G}}_{(3,1)}(\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_1 + \mathbf{q}_1) + \tilde{\mathcal{G}}_{(3,1)}(\mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(-\mathbf{k}_2 + \mathbf{q}_1) \\
& \left. + \tilde{\mathcal{G}}_{(3,1)}(-\mathbf{k}_1 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_1 + \mathbf{q}_1) + \tilde{\mathcal{G}}_{(3,1)}(-\mathbf{k}_2 - \mathbf{q}_1) \tilde{\mathcal{G}}_{(1,1)}(\mathbf{k}_2 + \mathbf{q}_1) \right\} + (1 + \alpha^3) \tilde{\mathcal{G}}_{(4,0)}(0). \tag{B26}
\end{aligned}$$

To further proceed with the power spectrum covariance, let us introduce the functions:

$$\mathcal{Q}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{x}) \equiv \mathcal{G}_{(i_1, j_1)}(\mathbf{x}) \mathcal{G}_{(i_2, j_2)}(\mathbf{x}); \text{ and } \mathcal{Q}_{(j)}^{(i)}(\mathbf{x}) \equiv \mathcal{G}_{(i, j)}(\mathbf{x}). \tag{B27}$$

The Fourier transform of $\mathcal{Q}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{x})$ is obtained through the convolution theorem:

$$\tilde{\mathcal{Q}}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{k}) \equiv \int \frac{d\mathbf{q}}{(2\pi)^3} \tilde{\mathcal{G}}_{(i_1, j_1)}(\mathbf{q}) \tilde{\mathcal{G}}_{(i_2, j_2)}(\mathbf{k} - \mathbf{q}). \tag{B28}$$

Consider the case where the survey volume is large and hence the $\tilde{\mathcal{G}}_{j,k}(\mathbf{k})$ functions are all very narrowly peaked around $\mathbf{k} = 0$. With this assumption and using equations (B27) and (B28), we express equation (B26) more compactly as

$$\begin{aligned}
\text{Cov} \left[|\tilde{\mathcal{F}}_c(\mathbf{k}_1)|^2, |\tilde{\mathcal{F}}_c(\mathbf{k}_2)|^2 \right] & \approx \left| P(\mathbf{k}_1) \tilde{\mathcal{Q}}_{(1,1)}^{(1,1)}(\mathbf{k}_1 + \mathbf{k}_2) + (1 + \alpha) \tilde{\mathcal{Q}}_{(0)}^{(2)}(\mathbf{k}_1 + \mathbf{k}_2) \right|^2 + \left| P(\mathbf{k}_1) \tilde{\mathcal{Q}}_{(1,1)}^{(1,1)}(\mathbf{k}_1 - \mathbf{k}_2) + (1 + \alpha) \tilde{\mathcal{Q}}_{(0)}^{(2)}(\mathbf{k}_1 - \mathbf{k}_2) \right|^2 \\
& + \tilde{\mathcal{Q}}_{(1,1)}^{(2,2)}(\mathbf{0}) \left[P(\mathbf{k}_1 + \mathbf{k}_2) + P(\mathbf{k}_1 - \mathbf{k}_2) \right] + \tilde{\mathcal{Q}}_{(1,1)}^{(3,1)}(\mathbf{0}) \left[P(\mathbf{k}_1) + P(\mathbf{k}_2) + P(-\mathbf{k}_1) + P(-\mathbf{k}_2) \right] + (1 + \alpha^3) \tilde{\mathcal{Q}}_{(0)}^{(4)}(\mathbf{0}). \tag{B29}
\end{aligned}$$

Notice that the equation above is invariant under $\mathbf{k}_i \rightarrow -\mathbf{k}_i$, $i \in (1, 2)$. We shall use this property next, when evaluating the bin-averaged covariance of the \mathcal{F}_c power spectra:

$$\begin{aligned} \text{Cov} \left[|\tilde{\mathcal{F}}_c(k_i)|^2, |\tilde{\mathcal{F}}_c(k_j)|^2 \right] &\approx 2\bar{P}^2(k_i) \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \tilde{\mathcal{Q}}_{(1,1)}^{(1,1)}(\mathbf{k}_1 + \mathbf{k}_2) \tilde{\mathcal{Q}}_{(1,1)}^{(1,1)}(-\mathbf{k}_1 - \mathbf{k}_2) + 4(1 + \alpha)\bar{P}(k_i) \\ &\quad \times \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \tilde{\mathcal{Q}}_{(1,1)}^{(1,1)}(\mathbf{k}_1 + \mathbf{k}_2) \tilde{\mathcal{Q}}_{(0)}^{(2)}(-\mathbf{k}_1 - \mathbf{k}_2) + 2(1 + \alpha)^2 \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \tilde{\mathcal{Q}}_{(0)}^{(2)}(\mathbf{k}_1 + \mathbf{k}_2) \tilde{\mathcal{Q}}_{(0)}^{(2)}(-\mathbf{k}_1 - \mathbf{k}_2) \\ &\quad + 2\tilde{\mathcal{Q}}_{(1,1)}^{(2,2)}(\mathbf{0})\bar{P}[k_i, k_j] + 2\tilde{\mathcal{Q}}_{(1,1)}^{(3,1)}(\mathbf{0}) [\bar{P}(k_i) + \bar{P}(k_j)] + (1 + \alpha^3)\tilde{\mathcal{Q}}_{(0)}^{(4)}(\mathbf{0}), \end{aligned} \quad (\text{B30})$$

where we defined

$$\bar{P}[k_i, k_j] \equiv \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} P(\mathbf{k}_1 + \mathbf{k}_2).$$

To get equation (B30), we expanded the modulus-squared terms in equation (B29) and we assumed that the k -space shells are sufficiently narrow that the power spectrum can be considered constant over the shell.

We next show how to compute the k -space-shell averages of products of the $\tilde{\mathcal{Q}}$ functions. Consider the following integral:

$$\begin{aligned} \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \tilde{\mathcal{Q}}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{k}_1 + \mathbf{k}_2) \tilde{\mathcal{Q}}_{(m_1, m_2)}^{(l_1, l_2)}(-\mathbf{k}_1 - \mathbf{k}_2) &= \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \int d\mathbf{x}_1 d\mathbf{x}_2 \mathcal{Q}_{(j_1 \dots j_n)}^{(i_1 \dots i_n)}(\mathbf{x}_1) \mathcal{Q}_{(m_1 \dots m_n)}^{(l_1 \dots l_n)}(\mathbf{x}_2) e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2)} \\ &= \int d\mathbf{x}_1 d\mathbf{x}_2 \bar{j}_0(k_i |\mathbf{x}_1 - \mathbf{x}_2|) \bar{j}_0(k_j |\mathbf{x}_1 - \mathbf{x}_2|) \mathcal{Q}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{x}_1) \mathcal{Q}_{(m_1, m_2)}^{(l_1, l_2)}(\mathbf{x}_2) \\ &= \int d\mathbf{x}_{21} \bar{j}_0(k_i x_{21}) \bar{j}_0(k_j x_{21}) \Xi_{(j_1, j_2 | m_1, m_2)}^{(i_1, i_2 | l_1, l_2)}(\mathbf{x}_{21}). \end{aligned} \quad (\text{B31})$$

To get the second line in the above equation, we have defined the shell-averaged spherical Bessel function as

$$\bar{j}_0(k_i x) \equiv \frac{1}{V_i} \int_{k_i - \Delta k/2}^{k_i + \Delta k/2} dk k^2 4\pi j_0(kx),$$

and then integrated the exponential functions over the angles $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_2$. The third line in the above equation resulted from the change of variables $\mathbf{x}_{21} = \mathbf{x}_2 - \mathbf{x}_1$, and from defining the correlation function of the weighted survey window function as

$$\Xi_{(j_1, j_2 | m_1, m_2)}^{(i_1, i_2 | l_1, l_2)}(\mathbf{x}_{21}) \equiv \int \frac{d^2 \hat{\mathbf{x}}_{21}}{4\pi} \int d\mathbf{x} \mathcal{Q}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{x}) \mathcal{Q}_{(m_1, m_2)}^{(l_1, l_2)}(\mathbf{x}_{21} + \mathbf{x}). \quad (\text{B32})$$

In the limit that the survey volume is large, the weighted survey window correlation function is very slowly varying over nearly all length-scales of interest, and so can be approximated by its value at zero-lag. Hence, we write

$$\begin{aligned} \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \tilde{\mathcal{Q}}_{(j_1, j_2)}^{(i_1, i_2)}(\mathbf{k}_1 + \mathbf{k}_2) \tilde{\mathcal{Q}}_{(m_1, m_2)}^{(l_1, l_2)}(-\mathbf{k}_1 - \mathbf{k}_2) &\approx \Xi_{(j_1, j_2 | m_1, m_2)}^{(i_1, i_2 | l_1, l_2)}(\mathbf{0}) \int_{V_i} \frac{d\mathbf{k}_1}{V_i} \int_{V_j} \frac{d\mathbf{k}_2}{V_j} \int_0^\infty d\mathbf{x}_{21} 4\pi x_{21}^2 j_0(x_{21} k_1) j_0(x_{21} k_2) \\ &= \frac{(2\pi)^3}{V_i} \Xi_{(j_1, j_2 | m_1, m_2)}^{(i_1, i_2 | l_1, l_2)}(\mathbf{0}) \delta_{i,j}^K, \end{aligned} \quad (\text{B33})$$

where for the second line we used the orthogonality relation of the spherical Bessel functions

$$\int_0^\infty dx x^2 j_\alpha(ux) j_\alpha(vx) = \frac{\pi}{2u^2} \delta^D(u - v).$$

Using the result in equation (B33), we write the covariance matrix of the power spectra of the field \mathcal{F}_c from equation (B30) as

$$\begin{aligned} \text{Cov} \left[|\mathcal{F}_c(k_i)|^2, |\mathcal{F}_c(k_j)|^2 \right] &\approx \frac{2(2\pi)^3}{V_i} \left[\bar{P}^2(k_i) \Xi_{(1,1|1,1)}^{(1,1|1,1)}(\mathbf{0}) + 2(1 + \alpha)\bar{P}(k_i) \Xi_{(1,1|0)}^{(1,1|2)}(\mathbf{0}) + (1 + \alpha)^2 \Xi_{(0|0)}^{(2|2)}(\mathbf{0}) \right] \delta_{i,j}^K \\ &\quad + 2\tilde{\mathcal{Q}}_{(1,1)}^{(2,2)}(\mathbf{0})\bar{P}[k_i, k_j] + 2\tilde{\mathcal{Q}}_{(1,1)}^{(3,1)}(\mathbf{0}) [\bar{P}(k_i) + \bar{P}(k_j)] + (1 + \alpha^3)\tilde{\mathcal{Q}}_{(0)}^{(4)}(\mathbf{0}), \end{aligned} \quad (\text{B34})$$

where the correlation function $\Xi_{(j_1, j_2 | m_1, m_2)}^{(i_1, i_2 | l_1, l_2)}$ of the weighted survey windows is defined by equation (B32). Finally, on taking the limit that $\bar{n}_h V_\mu \gg 1$, the last three terms on the right-hand side of the above equation will be subdominant (Smith 2009). For the remaining terms, we need the following expressions:

$$\begin{aligned} \Xi_{(1,1|1,1)}^{(1,1|1,1)}(\mathbf{0}) &= \int d\mathbf{x} \left[\mathcal{Q}_{(1,1)}^{(1,1)}(\mathbf{x}) \right]^2 = \int d\mathbf{x} [\mathcal{G}_{(1,1)}(\mathbf{x})]^4; \\ \Xi_{(1,1|0)}^{(1,1|2)}(\mathbf{0}) &= \int d\mathbf{x} \mathcal{Q}_{(1,1)}^{(1,1)}(\mathbf{x}) \mathcal{Q}_{(0)}^{(2)}(\mathbf{x}) = \int d\mathbf{x} [\mathcal{G}_{(1,1)}(\mathbf{x})]^2 \mathcal{G}_{(2,0)}(\mathbf{x}); \\ \Xi_{(0|0)}^{(2|2)}(\mathbf{0}) &= \int d\mathbf{x} \left[\mathcal{Q}_{(0)}^{(2)}(\mathbf{x}) \right]^2 = \int d\mathbf{x} [\mathcal{G}_{(2,0)}(\mathbf{x})]^2. \end{aligned} \quad (\text{B35})$$

Replacing equation (B35) in equation (B34), we arrive at the final expression for the Gaussian covariance matrix of the power spectrum of \mathcal{F}_c , given in the main text of the paper by equation (32).

APPENDIX C: FUNCTIONAL DERIVATIVES

To compute $\delta\mathcal{N}[w]$ and $\delta\mathcal{D}[w]$ entering in the minimization equation (39), we first write down the functional derivatives of the relevant $\bar{\mathcal{G}}$ functions. For a small variation in the path of w , we have

$$\begin{aligned}\bar{\mathcal{G}}_{(1,1)}[w + \delta w] &= \int dM \bar{n}(M) b(M) \Theta(\mathbf{x}|M) [w(\mathbf{x}, M) + \delta w(\mathbf{x}, M)] = \bar{\mathcal{G}}_{(1,1)}[w] + \delta\bar{\mathcal{G}}_{(1,1)}[w]; \\ \delta\bar{\mathcal{G}}_{(1,1)}[w] &\equiv \int dM \bar{n}(M) b(M) \Theta(\mathbf{x}|M) \delta w(\mathbf{x}, M); \end{aligned} \quad (\text{C1})$$

$$\begin{aligned}\bar{\mathcal{G}}_{(2,0)}[w + \delta w] &= \int dM \bar{n}(M) \Theta(\mathbf{x}|M) [w + \delta w]^2 = \bar{\mathcal{G}}_{(2,0)}[w] + \delta\bar{\mathcal{G}}_{(2,0)}[w]; \\ \delta\bar{\mathcal{G}}_{(2,0)}[w] &\equiv 2 \int dM \bar{n}(M) \Theta(\mathbf{x}|M) w(\mathbf{x}, M) \delta w(\mathbf{x}, M). \end{aligned} \quad (\text{C2})$$

Note that the above calculation is linear in δw , i.e. we neglect all terms containing powers higher than 1 in δw . We also need the functional derivative of the normalization constant A defined by equation (27):

$$\begin{aligned}A[w + \delta w] &= \int d\mathbf{x} \{ \bar{\mathcal{G}}_{(1,1)}[w + \delta w] \}^2 = \int d\mathbf{x} \{ \bar{\mathcal{G}}_{(1,1)}[w] + \delta\bar{\mathcal{G}}_{(1,1)}[w] \}^2 = A[w] + 2 \int d\mathbf{x} \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) \delta\bar{\mathcal{G}}_{(1,1)}[w] = A[w] + \delta A[w]; \\ \delta A[w] &\equiv 2 \int d\mathbf{x} \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) \int dM \bar{n}(M) b(M) \Theta(\mathbf{x}|M) \delta w(\mathbf{x}, M). \end{aligned} \quad (\text{C3})$$

We now proceed to evaluation the functional derivative of \mathcal{N} from equation (36):

$$\delta\mathcal{N}[w] = 2 \int d\mathbf{x} \left[[\bar{\mathcal{G}}_{(1,1)}(\mathbf{x})]^2 + c \bar{\mathcal{G}}_{(2,0)}(\mathbf{x}) \right] \left[2 \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) \delta\bar{\mathcal{G}}_{(1,1)}[w] + c \delta\bar{\mathcal{G}}_{(2,0)}[w] \right].$$

Using the results of equations (C1) and (C2), we obtain

$$\delta\mathcal{N}[w] = 4 \int d\mathbf{x} dM \left\{ \left[[\bar{\mathcal{G}}_{(1,1)}(\mathbf{x})]^2 + c \bar{\mathcal{G}}_{(2,0)}(\mathbf{x}) \right] \bar{n}(M) \Theta(\mathbf{x}|M) \left[\bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) b(M) + c w(\mathbf{x}, M) \right] \right\} \delta w(\mathbf{x}, M). \quad (\text{C4})$$

Looking at equations (37) and (C3), we further write down

$$\delta\mathcal{D}[w] = 4A[w] \int d\mathbf{x} dM \left\{ \bar{\mathcal{G}}_{(1,1)}(\mathbf{x}) \bar{n}(M) b(M) \Theta(\mathbf{x}|M) \right\} \delta w(\mathbf{x}, M). \quad (\text{C5})$$

APPENDIX D: SIGNAL-TO-NOISE EXPRESSIONS

Before deriving the S/N for FKP, mass, and optimal weights, it will be very useful to introduce the following quantities:

$$\begin{aligned}\bar{n}_h(\chi) &\equiv \int_0^\infty dM \bar{n}(M, \chi) \Theta(\chi|M) = \int_{M_{\min}(\chi)}^\infty dM \bar{n}(M, \chi); \\ \bar{\rho}_h(\chi) &\equiv \int_0^\infty dM M \bar{n}(M, \chi) \Theta(\chi|M) = \int_{M_{\min}(\chi)}^\infty dM M \bar{n}(M, \chi); \\ \bar{b}^j(\chi) &\equiv \frac{D^j(\chi)}{\bar{n}_h(\chi)} \int_0^\infty dM \bar{n}(M, \chi) b^j(M, \chi) \Theta(\chi|M) = \frac{D^j(\chi)}{\bar{n}_h(\chi)} \int_{M_{\min}(\chi)}^\infty dM \bar{n}(M, \chi) b^j(M, \chi); \text{ and } \bar{b}(\chi) \equiv \bar{b}^1(\chi); \\ \bar{b}_M^j(\chi) &\equiv \frac{D^j(\chi)}{\bar{\rho}_h(\chi)} \int_0^\infty dM M \bar{n}(M, \chi) b^j(M, \chi) \Theta(\chi|M) = \frac{D^j(\chi)}{\bar{\rho}_h(\chi)} \int_{M_{\min}(\chi)}^\infty dM M \bar{n}(M, \chi) b^j(M, \chi); \text{ and } \bar{b}_M(\chi) \equiv \bar{b}_M^1(\chi); \\ \langle M^j(\chi) \rangle &\equiv \frac{1}{\bar{n}_h(\chi)} \int_0^\infty dM M^j \bar{n}(M, \chi) \Theta(\chi|M) = \frac{1}{\bar{n}_h(\chi)} \int_{M_{\min}(\chi)}^\infty dM M^j \bar{n}(M, \chi); \text{ and } \langle M(\chi) \rangle \equiv \langle M^1(\chi) \rangle. \end{aligned} \quad (\text{D1})$$

We now proceed to the calculation of the S/N for the matter power spectrum using the optimal weights given by equation (44), as well as w_{FKP} and $w_{\text{M+FKP}}$. Using equation (33), one can show the following general expressions:

$$\left(\frac{S}{N} \right)_{\text{FKP}}^2(k_i) = \frac{\Omega_\mu V_i}{2(2\pi)^3} \left\{ \int_0^\infty d\chi \chi^2 \left[\frac{\bar{b}(\chi) \bar{n}_h(\chi) \bar{P}_i}{(1 + \alpha) + \bar{n}_h(\chi) \bar{P}_i} \right]^2 \right\}^2 / \int_0^\infty d\chi \chi^2 \left\{ \bar{n}_h(\chi) \bar{P}_i \frac{[(1 + \alpha) + \bar{b}^2(\chi) \bar{n}_h(\chi) \bar{P}_i]}{[(1 + \alpha) + \bar{n}_h(\chi) \bar{P}_i]^2} \right\}; \quad (\text{D2})$$

$$\left(\frac{S}{N}\right)_{\text{OPT}}^2(k_i) = \frac{\Omega_\mu V_i}{2(2\pi)^3} \int_0^\infty d\chi \chi^2 \left[\frac{\bar{n}_h(\chi) \bar{b}^2(\chi) \bar{P}_i}{(1+\alpha) + \bar{n}_h(\chi) \bar{b}^2(\chi) \bar{P}_i} \right]^2; \quad (\text{D3})$$

$$\left(\frac{S}{N}\right)_{\text{M+FKP}}^2(k_i) = \frac{\Omega_\mu V_i}{2(2\pi)^3} \left\{ \int_0^\infty d\chi \chi^2 \left[\frac{\bar{b}_M(\chi) \bar{\rho}_h(\chi) \bar{P}_i}{(1+\alpha) + \bar{n}_h(\chi) \bar{P}_i} \right]^2 \right\}^2 / \int_0^\infty d\chi \chi^2 \left\{ \frac{\bar{P}_i (1+\alpha) \langle M^2(\chi) \rangle \bar{n}_h(\chi) + [\bar{b}_M(\chi) \bar{\rho}_h(\chi)]^2 \bar{P}_i}{[(1+\alpha) + \bar{n}_h(\chi) \bar{P}_i]^2} \right\}^2. \quad (\text{D4})$$

APPENDIX E: PROOF THAT $\bar{b}^2 \leq \bar{b}^2$

To begin, the Cauchy–Schwarz inequality for two functions f and g states that

$$|\langle f|g \rangle|^2 \leq \langle f|f \rangle \langle g|g \rangle, \quad (\text{E1})$$

where in the above we are using the following notation:

$$\langle f|g \rangle \equiv \int dx q(x) f(x) g(x), \quad (\text{E2})$$

and with $q(x)$ an arbitrary positive definite weighting function. Thus, if we take $f = b(M)$, $g = 1$ and $q = \bar{n}(M)$, then we have the following inequality:

$$\left[\int_{M_{\min}}^\infty dM \bar{n}(M) b(M) \right]^2 \leq \left[\int_{M_{\min}}^\infty dM \bar{n}(M) b^2(M) \right] \left[\int_{M_{\min}}^\infty dM \bar{n}(M) \right]. \quad (\text{E3})$$

On dividing both sides of this inequality through by \bar{n}_h^2 we arrive at the stated result.

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