

Unbiased contaminant removal for 3D galaxy power spectrum measurements

B. Kalus,^{1★} W. J. Percival,¹ D. J. Bacon¹ and L. Samushia^{1,2,3}

¹*Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth PO1 3FX, UK*

²*Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*

³*National Abastumani Astrophysical Observatory, Ilia State University, 2A Kazbegi Ave., GE-1060 Tbilisi, Georgia*

Accepted 2016 August 6. Received 2016 August 4; in original form 2016 July 8

ABSTRACT

We assess and develop techniques to remove contaminants when calculating the 3D galaxy power spectrum. We separate the process into three separate stages: (i) removing the contaminant signal, (ii) estimating the uncontaminated cosmological power spectrum and (iii) debiasing the resulting estimates. For (i), we show that removing the best-fitting contaminant (*mode subtraction*) and setting the contaminated components of the covariance to be infinite (*mode deprojection*) are mathematically equivalent. For (ii), performing a quadratic maximum likelihood (QML) estimate after *mode deprojection* gives an optimal unbiased solution, although it requires the manipulation of large N_{mode}^2 matrices (N_{mode} being the total number of modes), which is unfeasible for recent 3D galaxy surveys. Measuring a binned average of the modes for (ii) as proposed by Feldman, Kaiser & Peacock (FKP) is faster and simpler, but is sub-optimal and gives rise to a biased solution. We present a method to debias the resulting FKP measurements that does not require any large matrix calculations. We argue that the sub-optimality of the FKP estimator compared with the QML estimator, caused by contaminants, is less severe than that commonly ignored due to the survey window.

Key words: methods: statistical – large-scale structure of Universe.

1 INTRODUCTION

Galaxy surveys provide a rich store of information about the nature of the Universe, allowing us to constrain cosmological models with baryon acoustic oscillations, gravitational models with redshift-space distortions and inflationary models with primordial non-Gaussianity. A basic statistic containing large-scale structure information is the galaxy power spectrum $P(k)$, which is the two-point function of the Fourier transformed density field. Future large-scale structure surveys, such as the Dark Energy Spectroscopic Instrument survey (Schlegel et al. 2011; Levi et al. 2013), *Euclid* (Laureijs et al. 2011)¹ and the Square Kilometre Array,² will probe larger volumes, therefore allowing us to measure more Fourier modes of the galaxy density field.

The observed galaxy field can be contaminated with fluctuations of non-cosmological origin, such as variations due to the galactic extinction and the stellar density. Often the contaminants are not known exactly (e.g. we may know the shape of the spurious mode but may not know its exact amplitude) which makes their exact removal impossible. These modes have the potential to strongly bias

cosmological constraints derived from the clustering measurements, so we need to correct or suppress these misleading modes in a responsible way.

We now introduce the basic mathematical problem that we wish to solve and introduce the main methods of removing contaminants discussed in the literature. We assume that we have measured the galaxy density field as real numbers in configuration space, which we (fast) Fourier transform to obtain a Hermitian density field $F(\mathbf{k})$. Furthermore, we assume that the contamination can be described by another Hermitian field $f(\mathbf{k})$, such that the true density field is given by

$$D(\mathbf{k}) = F(\mathbf{k}) - \varepsilon_{\text{true}} f(\mathbf{k}), \quad (1)$$

with $\varepsilon_{\text{true}}$ unknown. In the cases with multiple contaminants (which we label with capital Latin indices), we extend equation (1) to

$$D(\mathbf{k}) = F(\mathbf{k}) - \sum_A \varepsilon_A^{(\text{true})} f_A(\mathbf{k}). \quad (2)$$

Furthermore, we assume that $F(\mathbf{k})$ and $f(\mathbf{k})$ are uncorrelated, which is a valid assumption for most sources of systematics since they originate from our Galaxy or due to telescope effects. Large-scale surveys will reduce the current sample variance limitation on the power spectrum on scales where the systematic errors have a significant

* E-mail: benedict.kalus@port.ac.uk

¹ www.euclid-ec.org

² www.skatelescope.org

impact. As a consequence, having control of these systematics is a key requirement to provide accurate cosmological measurements.

In order to investigate techniques for estimating the power spectrum in the presence of contaminants, we separate the process into three separate stages: (i) removing the contaminant signal, (ii) estimating the uncontaminated cosmological power spectrum and (iii) debiasing the resulting estimates. Two techniques are in common usage for removing the contaminant signal. (i) The first is *mode subtraction* (cf. Sections 4 and 5), where contaminants are removed by fitting the amplitude of the contaminant field $f(\mathbf{k})$ to the data and simply subtracted off from $F(\mathbf{k})$. The second is *mode deprojection* (Rybicki & Press 1992), which is based on assigning infinitely large covariances to contaminated modes, thus removing them from any analysis. In our nomenclature, a mode is a linear combination of Fourier modes rather than a single \mathbf{k} -mode. This is reflected in the naming of *mode subtraction* and *mode deprojection*. This choice of names shall distinguish the *mode subtraction* technique from a third technique for removing the contaminant signal, called *template subtraction*, where the observed power spectra are corrected using best-fitting amplitudes derived via cross-correlations between the data and the templates. Elsner, Leistedt & Peiris (2016) have shown that this method provides a biased estimate of the power and we will not consider it further in this paper. For (ii), the power spectrum $P(k)$ is commonly estimated by the FKP estimator (Feldman et al. 1994), which is an approximation to the quadratic maximum likelihood (QML) estimator (Tegmark et al. 1998). As well as being optimal, the QML estimator has the advantage of producing unbiased power spectrum estimates. However, when applying this methodology to data with N_{mode} modes, one has to calculate, for each bin, an $N_{\text{mode}} \times N_{\text{mode}}$ matrix, and then, after binning the data into N_{bin} bins, an overall $N_{\text{bin}} \times N_{\text{bin}}$ normalization matrix, which makes the application of this methodology unfeasible for future surveys with increased number of modes N_{mode} . In this work, we suggest a modified FKP-style *mode subtraction* approach. We show that this technique can be made unbiased and, on a mode-by-mode basis, is mathematically identical to *mode deprojection*. The FKP estimator with debiased *mode subtraction* is not optimal in that it discards more information than the full QML estimator, but we expect that, in realistic cases, this loss of information will be small.

The outline of this paper is as follows. We provide an introduction to power spectrum estimation in Section 2, introducing the QML and FKP estimators. We introduce the systematics removal techniques, *mode deprojection* and *mode subtraction*, in Sections 3 and 4, respectively, and we show that before normalization their resulting power spectra are the same. These are extended to multiple contaminants in Appendices A and B, respectively. We introduce a new normalization factor in Section 5 for a single contaminant and compare it to the normalization of the QML estimator of Tegmark et al. (1998). This derivation is extended to allow for a non-diagonal covariance in Appendix C1. We show that we can apply our methodology also to multiple contaminants in Section 6 and test the different methods on simulations in Section 7. We conclude in Section 8.

2 POWER SPECTRUM ESTIMATORS

In this section, we review two basic power spectrum estimators: the QML estimator (Tegmark et al. 1998) and the simplified FKP estimator (Feldman et al. 1994), to which QML reduces in the limit of uncorrelated modes with equal noise per mode in each bin. Even without considering any contaminants, the FKP estimator is easier to implement and is used in most recent analyses of large-

scale structure, while the QML estimator is optimal but difficult to implement especially on smaller scales.

The QML estimator (Tegmark et al. 1998) is given by

$$\widehat{P}(k_i) = \sum_j \mathbf{N}_{ij}^{-1} \mathbf{p}_j, \quad (3)$$

where the power is a convolution of the inverse of a normalization matrix \mathbf{N}_{ij} and a weighted two-point function

$$\mathbf{p}_j \equiv \sum_{\alpha, \beta} F^*(\mathbf{k}_\alpha) \mathbf{E}_{\alpha\beta}(\mathbf{k}_j) F(\mathbf{k}_\beta). \quad (4)$$

The weight is given by the estimator matrix

$$\mathbf{E}(\mathbf{k}_j) = -\frac{\partial \mathbf{C}^{-1}}{\partial P(\mathbf{k}_j)}, \quad (5)$$

which describes how the inverse of the density field covariance matrix \mathbf{C} changes with respect to the prior of the power spectrum of the respective bin. If the QML normalization is proportional to the Fisher information, i.e.

$$\mathbf{N}_{ij} = \text{tr} \left\{ \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial P(k_i)} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial P(k_j)} \right\}, \quad (6)$$

the QML estimator is the optimal maximum likelihood estimator of the variance of a field that obeys a multivariate Gaussian distribution (Tegmark et al. 1998). Assuming a Gaussian density field $D(\mathbf{k})$, the QML estimator therefore provides an estimate of the power spectrum with minimal errors.

Under the assumption that all modes are independent, the covariance of the density field is given by the power spectrum (and the Kronecker delta $\delta_{\mu\nu}$)

$$\mathbf{C}_{\mu\nu} = \delta_{\mu\nu} P(k_\mu). \quad (7)$$

We assume for the derivative of \mathbf{C} with respect to $P(k_i)$ that it is unity if the modes \mathbf{k}_α and \mathbf{k}_β are equal and contained in the bin \mathbb{k}_i , and zero otherwise, which we write using the Heaviside function Θ as

$$\frac{\partial \mathbf{C}_{\alpha\beta}}{\partial P(k_i)} = \delta_{\alpha\beta} \Theta(\mathbf{k}_\alpha \in \mathbb{k}_i) \equiv \delta_{\alpha\beta} \Theta_{\alpha i}. \quad (8)$$

Given equation (7) and (8), we find

$$\mathbf{E}_{\alpha\beta}(\mathbf{k}_j) = \frac{\delta_{\alpha\beta}}{P^2(k_\alpha)} \Theta_{\alpha j} \quad (9)$$

and

$$\mathbf{N}_{ij} = \frac{N_{\mathbb{k}_i}}{P^2(k_i)} \delta_{ij}, \quad (10)$$

where $N_{\mathbb{k}_i}$ is the total number of modes in a given bin \mathbb{k}_i . Hence, the QML estimator of equation (3) reduces to the FKP estimator (Feldman et al. 1994) under the assumption that the covariance is constant within the \mathbf{k} -bin, where several modes (labelled with Greek indices) are combined into bins (denoted by \mathbb{k}_i and distinguished with lower case Latin indices) and the absolute values squared of the density field of each bin are summed:

$$\widehat{P}(k_i) = \frac{1}{N_{\mathbb{k}_i}} \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} |F(\mathbf{k}_\alpha)|^2. \quad (11)$$

The difference here is that the QML estimator uses a prior of the power spectrum $P(k_\alpha)$ to weight contributions from each mode optimally, which means that the covariance of the power spectrum is minimal. The FKP estimator is commonly applied even when the assumptions of equation (7)–(10) are not valid.

3 REMOVING CONTAMINANTS: MODE DEPROJECTION

We now describe how *mode deprojection* can be applied to estimate the 3D galaxy power spectrum. The method was first suggested in Rybicki & Press (1992) in the context of noisy, irregularly sampled data. Applications and extensions to angular power spectra can be found for *Wilkinson Microwave Anisotropy Probe* data in Slosar, Seljak & Makarov (2004), for SDSS-III data in Ho et al. (2012), for photometric quasars of the XDQSOz catalogue in Leistedt & Peiris (2014) and Leistedt, Peiris & Roth (2014) and for 2D galaxy clustering in general in Elsner et al. (2016). We use the notation of Elsner et al. (2016) for consistency.

Suppose we estimate the power spectrum using QML and that there is only a single contaminant. Then one can suppress contaminated modes in the covariance matrix updating the covariance matrix as (Elsner et al. 2016)

$$\mathbf{C}_{\alpha\beta} \rightarrow \tilde{\mathbf{C}}_{\alpha\beta} = \mathbf{C}_{\alpha\beta} + \lim_{\sigma \rightarrow \infty} \sigma f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta), \quad (12)$$

i.e. letting the covariances of contaminated modes tend to infinity. Making use of the Sherman–Morrison matrix inversion lemma (Sherman & Morrison 1950), one can see that (if $f(\mathbf{k}) \neq 0 \forall \mathbf{k}$) the inverse updated covariance matrix converges to

$$\tilde{\mathbf{C}}_{\alpha\beta}^{-1} = \mathbf{C}_{\alpha\beta}^{-1} - \frac{\sum_{\mu\nu} \mathbf{C}_{\alpha\mu}^{-1} f(\mathbf{k}_\mu) f^*(\mathbf{k}_\nu) \mathbf{C}_{\nu\beta}^{-1}}{\sum_{\mu\nu} f^*(\mathbf{k}_\mu) \mathbf{C}_{\mu\nu}^{-1} f(\mathbf{k}_\nu)}. \quad (13)$$

Now supposing that the modes are independent, i.e. equation (7) holds, we can insert it into equation (13) so that

$$\tilde{\mathbf{C}}_{\alpha\beta}^{-1} = \frac{\delta_{\alpha\beta}}{P(k_\alpha)} - \frac{1}{R_P} \frac{f(\mathbf{k}_\alpha) f^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\beta)}, \quad (14)$$

where we have defined

$$R_P \equiv \sum_{\mu} \frac{|f(\mathbf{k}_\mu)|^2}{P(k_\mu)}, \quad (15)$$

for simplicity. Taking the derivative of equation (14) with respect to $P(k_i)$, we obtain the updated estimator matrix³

$$\tilde{\mathbf{E}}_{\alpha\beta}(k_j) = \frac{\delta_{\alpha\beta}}{P_\alpha^2} \Theta_{\alpha j} - \frac{1}{R_P} \frac{f_\alpha f_\beta^*}{P_\alpha P_\beta} \left(\frac{\Theta_{\alpha j}}{P_\alpha} + \frac{\Theta_{\beta j}}{P_\beta} - \frac{t_j}{R_P} \right), \quad (16)$$

where

$$t_i \equiv \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \frac{|f(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)}. \quad (17)$$

After inserting equation (16) into equation (4), we obtain for the two-point function

$$\begin{aligned} P_i &= \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \left\{ \frac{|F(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)} - \frac{2}{R_P} \text{Re} \left[S_P \frac{F^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P^2(k_\alpha)} \right] \right. \\ &\quad \left. + \frac{|S_P|^2 |f(\mathbf{k}_\alpha)|^2}{R_P^2 P^2(k_\alpha)} \right\} \\ &= \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \frac{\left| F(\mathbf{k}_\alpha) - \frac{S_P}{R_P} f(\mathbf{k}_\alpha) \right|^2}{P^2(k_\alpha)}, \end{aligned} \quad (18)$$

where we have defined

$$S_P \equiv \sum_{\mathbf{k}_\alpha} \frac{F^*(\mathbf{k}_\alpha) f(\mathbf{k}_\alpha)}{P_\alpha}. \quad (19)$$

³ writing $f_\alpha \equiv f(\mathbf{k}_\alpha)$ and $P_\alpha \equiv P(k_\alpha)$ to save space

S_P is real, because $F(\mathbf{k})$ and $f(\mathbf{k})$ are Hermitian fields with real Fourier transforms.

Equation (18) is in a considerably simpler form than equation (4) and does not require calculating many matrix elements of the estimator matrix \mathbf{E} . We show in the next section that we can consider this equation as a best fit of the contaminants in the data.

We can normalize the updated mode deprojected QML estimator by replacing \mathbf{C} by $\tilde{\mathbf{C}}$ in equation (6). As the term that suppresses contaminated modes from the covariance matrix in equation (12) does not depend on the power $P(k)$, we have $\frac{\partial \tilde{\mathbf{C}}_{\alpha\beta}}{\partial P(k_i)} = \frac{\partial \mathbf{C}_{\alpha\beta}}{\partial P(k_i)}$ and hence the normalization is

$$\begin{aligned} \tilde{\mathbf{N}}_{ij} &= \sum_{\alpha\mu\nu\rho} \tilde{\mathbf{C}}_{\alpha\mu}^{-1} \delta_{\mu\nu} \Theta_{\mu i} \tilde{\mathbf{C}}_{\nu\rho}^{-1} \delta_{\rho\alpha} \Theta_{\alpha j} \\ &= \sum_{\alpha\mu} \left| \tilde{\mathbf{C}}_{\alpha\mu}^{-1} \right|^2 \Theta_{\mu i} \Theta_{\alpha j} \\ &= \sum_{\alpha\mu} \Theta_{\alpha i} \Theta_{\mu j} \left[\frac{\delta_{\alpha\mu}}{P^2(k_\alpha)} \left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)} \right) \right. \\ &\quad \left. + \frac{1}{R_P^2} \frac{|f(\mathbf{k}_\alpha) f(\mathbf{k}_\mu)|^2}{P^2(k_\alpha) P^2(k_\mu)} \right], \end{aligned} \quad (20)$$

where we have used the Hermitian property of $\tilde{\mathbf{C}}^{-1}$ in the third equality. In the first term in the square brackets, \mathbf{k}_α has to be in both \mathbb{k}_i and \mathbb{k}_j ; hence, we can replace one Θ with δ_{ij} , such that equation (20) can be written as a diagonal matrix with diagonal elements \tilde{n} and the outer product of a vector with itself:

$$\begin{aligned} \tilde{\mathbf{N}}_{ij} &= \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} \frac{\delta_{ij}}{P^2(k_\alpha)} \left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)} \right) + \frac{t_i t_j}{R_P^2} \\ &\equiv \tilde{n}_i \delta_{ij} + \frac{t_i t_j}{R_P^2}. \end{aligned} \quad (21)$$

This means that we can apply the Sherman–Morrison matrix inversion lemma (Sherman & Morrison 1950):

$$\tilde{\mathbf{N}}_{ij}^{-1} = \frac{\delta_{ij}}{\tilde{n}_i} - \frac{1}{R_P^2 + \sum_{\ell} \frac{t_\ell^2}{\tilde{n}_\ell}} \frac{t_i t_j}{\tilde{n}_i \tilde{n}_j}. \quad (22)$$

As $\tilde{\mathbf{N}}^{-1}$ is not diagonal, it does not reduce to a simple FKP-style estimator, i.e. if we have N_{bin} bins, we have to calculate for each bin the $N_{\text{mode}} \times N_{\text{mode}}$ estimator matrix \mathbf{E} and we have to invert the $N_{\text{bin}} \times N_{\text{bin}}$ normalization matrix. This is not feasible for 3D clustering, because of the large number of modes to be considered, especially if we want to choose narrow bins. Including several contaminants makes it even more costly.

One way around this is a new framework introduced by Leistedt & Peiris (2014) which they call *extended mode projection* and that selectively removes modes based on cross-correlations with the data. However, this procedure reintroduces a small bias (Elsner et al. 2016).

Another possibility is using the methodology of the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS) collaboration, which is similar to that described in the next section, but applied at the power spectrum level. However, this method is also biased (Elsner et al. 2016). Although Ross et al. (2016) show that, for the completed SDSS-III BOSS DR12, the bias is much smaller than the statistical uncertainty, it was shown in the appendix of Ross et al. (2012) that the bias is significant when one attempts to correct for many systematics. Furthermore, we expect smaller statistical uncertainties with future surveys, so in the next two sections we

consider a computationally cheaper way of removing this small bias.

4 REMOVING CONTAMINANTS: MODE SUBTRACTION

Here we will consider *mode subtraction* and its link to *mode deprojection*. In order to remove contaminants, we start by treating the true, but unknown, amplitude of the contamination $\varepsilon_{\text{true}}$ in equation (1) as a free parameter ε , so that an estimate of the true density field $D(\mathbf{k})$ reads

$$\widehat{D}(\mathbf{k}) = F(\mathbf{k}) - \varepsilon f(\mathbf{k}). \quad (23)$$

Note that this is different from the *template subtraction* method introduced by Ho et al. (2012), which is used by the BOSS collaboration and works entirely at the level of power spectra, whereas equation (23) works at the map level. We can write a simplified model of the Gaussian likelihood whose maximum is given by the QML (cf. equation (3) and Tegmark et al. 1998) in the approximation of a diagonal covariance matrix, with a small contaminant that does not affect the covariance. This is given by

$$-2 \ln \mathcal{L} = \ln \left(\prod_{\mathbf{k}} P(k) \right) + \sum_{\mathbf{k}} \frac{|F(\mathbf{k}) - \varepsilon f(\mathbf{k})|^2}{P(k)}. \quad (24)$$

We can therefore find ε by minimizing equation (24), which is equivalent to simultaneously fitting ε and the model parameters entering the model power spectrum. The derivative of $\ln \mathcal{L}$ with respect to ε reads

$$\frac{\partial \ln \mathcal{L}}{\partial \varepsilon} = \sum_{\mathbf{k}} \frac{\text{Re} [F_f(\mathbf{k})F^*(\mathbf{k})] - \varepsilon |F_f(\mathbf{k})|^2}{P(k)}. \quad (25)$$

This expression is equal to zero and the likelihood maximized if

$$\varepsilon^{(\text{BF})} = \frac{S_p}{R_p}. \quad (26)$$

The uncontaminated estimate of the density field is hence given by

$$\widehat{D}(\mathbf{k}) = F(\mathbf{k}) - \frac{S_p}{R_p} f(\mathbf{k}), \quad (27)$$

and we can estimate the power as

$$\widehat{P}(k_i) = \frac{1}{N_{k_i}} \sum_{\mathbf{k}_\alpha} \left| F(\mathbf{k}_\alpha) - \frac{S_p}{R_p} f(\mathbf{k}_\alpha) \right|^2. \quad (28)$$

This is similar to the *mode deprojection* result of equation (18) with a bias, missing the inverse noise matrix convolution of equation (3). The bias of this estimate comes about because S_p is correlated with the true density field $D(\mathbf{k})$. This correlation is similar to that created by the internal linear combination method (e.g. Bennett et al. 2003) for the analysis of cosmic microwave background data. Based on this knowledge, we build an unbiased FKP-style estimator in the next section.

5 AN UNBIASED FKP-STYLE ESTIMATOR

We present in this section a simple, although sub-optimal, way to remove the bias on the power spectrum estimate resulting from imperfectly removing systematics using either equation (18) or (28). A straightforward way to remove the bias consists of calculating the expectation value of the power from each mode analytically,

Table 1. Expectation values of quantities entering equation (30).

$\langle F(\mathbf{k}_\alpha)F^*(\mathbf{k}_\beta) \rangle$	$\delta_{\alpha\beta} P(k_\alpha) + \varepsilon_{\text{true}}^2 f(\mathbf{k}_\alpha)f^*(\mathbf{k}_\beta)$
$\langle \varepsilon_{\text{BF}} \rangle$	$\frac{\langle S_p \rangle}{R_p} = \varepsilon_{\text{true}}$
$\langle \varepsilon_{\text{BF}}^2 \rangle$	$\frac{1}{R_p} + \varepsilon_{\text{true}}^2$
$\langle S_p F(\mathbf{k}_\alpha) \rangle$	$f(\mathbf{k}_\alpha) + R_p f(\mathbf{k}_\alpha)\varepsilon_{\text{true}}^2$

assuming equation (1), and divide out the bias. We start with calculating some useful expectations which we need for the final result, summarized in Table 1.

With these equations at hand, we can calculate the expectation of equations (18) and (28), i.e. the two-point function of equation (1):

$$\begin{aligned} & \left\langle |F(\mathbf{k}_\alpha) - \frac{S_p}{R_p} f(\mathbf{k}_j)|^2 \right\rangle \\ &= \langle |F(\mathbf{k}_\alpha)|^2 \rangle - \frac{2}{R_p} \langle S_p F(\mathbf{k}_\alpha) \rangle f^*(\mathbf{k}_\alpha) + \langle \varepsilon_{\text{BF}}^2 \rangle |f(\mathbf{k}_\alpha)|^2 \\ &= P(k_\alpha) - \frac{|f(\mathbf{k}_\alpha)|^2}{R_p}, \end{aligned} \quad (29)$$

hence, we can build an unbiased estimator of the power by dividing each mode in equations (18) and (28) by

$$1 - \frac{1}{R_p} \frac{|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}. \quad (30)$$

If we want to debias the two-point function using this factor, we have to assume a prior power spectrum. Note that the QML approach also requires the prior knowledge of the power spectrum. We will see in Section 7 that the impact of adopting a slightly wrong prior is indeed small. Our final estimator of the power spectrum is then

$$\widehat{P}(k_i) = \frac{1}{N_{k_i}} \sum_{\mathbf{k}_\alpha} \frac{\left| F(\mathbf{k}_\alpha) - \frac{S_p}{R_p} f(\mathbf{k}_\alpha) \right|^2}{1 - \frac{1}{R_p} \frac{|f(\mathbf{k}_\alpha)|^2}{P(k_\alpha)}}. \quad (31)$$

Equation (31) is one of the key results of this paper: this is an extension of the FKP estimator that removes potential contaminants from the data in an unbiased way, without the need for large matrices. Moreover, as it is in the same form as the well-established FKP estimator, this can easily be folded into estimators for redshift-space clustering such as those by Bianchi et al. (2015) and Scoccimarro (2015).

The same debiasing factor can also be derived from the QML Fisher information matrix \mathbf{N} , which in the QML approach performs both the debiasing and optimization effects. Without binning, equation (21) simplifies to

$$\tilde{\mathbf{N}}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P^2(k_\alpha)} \left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_p P(k_\alpha)} \right) + \frac{1}{R_p^2} \frac{|f(\mathbf{k}_\alpha)|^2}{P^2(k_\alpha)} \frac{|f(\mathbf{k}_\beta)|^2}{P^2(k_\beta)}. \quad (32)$$

The difference between the two approaches is that QML provides an unbiased optimal power estimate, whereas equation (31) has been constructed such that it is only unbiased, i.e. the powers in the denominators of equation (32) act as optimal weights to each mode. If we allow for some information loss within bins, by assuming that the expected power is constant within each bin, we can replace $P^2(k_\beta)$ by $P(k_\alpha)P(k_\beta)$, such that

$$\tilde{\mathbf{N}}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{P^2(k_\alpha)} \left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_p P(k_\alpha)} \right) + \frac{1}{R_p^2} \frac{|f(\mathbf{k}_\alpha)|^2}{P^3(k_\alpha)} \frac{|f(\mathbf{k}_\beta)|^2}{P(k_\beta)}. \quad (33)$$

This normalization is proportional to the Fisher information matrix (Tegmark et al. 1998), from which we marginalize out contributions

from other modes by summing over all modes \mathbf{k}_β :

$$\begin{aligned} \sum_{\beta} \tilde{\mathbf{N}}_{\alpha\beta} &= \frac{1}{P^2(k_\alpha)} \left(1 - \frac{2|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)} \right) + \frac{1}{R_P} \frac{|f(\mathbf{k}_\alpha)|^2}{P^3(k_\alpha)} \\ &= \frac{1}{P^2(k_\alpha)} \left(1 - \frac{|f(\mathbf{k}_\alpha)|^2}{R_P P(k_\alpha)} \right). \end{aligned} \quad (34)$$

This is exactly equation (30) with a factor of $\frac{1}{P^2(k_\alpha)}$ that cancels out the difference between equations (18) and (28). We have therefore shown that equation (31) is a non-optimal, but unbiased, approximation to using the QML normalization with *mode deprojection*. In the limit of narrow bins, when the power spectrum does not change significantly within the bin, equation (31) is mathematically identical to the QML result. We shall study the impact of this sub-optimality in examples later in Section 7. In fact, we will argue later that this is actually a weaker effect than many common approximations applied when using the FKP estimator, such as ignoring large-scale window effects in the QML approach, when averaging large-scale modes.

Note that, in the absence of systematics, we have assumed a diagonal covariance matrix in the derivation of both the *mode subtraction* and the debiasing step. In practice, the covariance matrix has off-diagonal terms due to the effect of the survey window. However, this is usually not included when calculating the data power spectrum but, instead, it is included as a convolution in the model power spectrum. We show in Appendix C1 that equation (31) still holds in the general case of having a non-diagonal covariance matrix, as long as R_P is generalized as in equation (C8). This generalized R_P requires the inversion of the full N_{mode}^2 covariance matrix. However, we show in Appendix C2 that the effect of assuming a diagonal covariance matrix is either small or can be corrected for using the covariance matrix, without inversion.

6 REMOVING MULTIPLE CONTAMINANTS

We have shown the equivalence between *mode deprojection* and *debaised mode subtraction* for one contaminant. A realistic survey has several sources of potential contaminants, so we show here this equivalence holds for an arbitrary number of templates. For *mode deprojection*, we have to update the covariance matrix with a sum over all templates, and thus we have to replace equation (12) with

$$\tilde{\mathbf{C}}_{\alpha\beta} = \mathbf{C}_{\alpha\beta} + \lim_{\sigma \rightarrow \infty} \sigma \sum_{A=1}^{N_{\text{sys}}} f_A(\mathbf{k}_\alpha) f_A^*(\mathbf{k}_\beta). \quad (35)$$

Starting from equation (35), we derive in Appendix A the unbinned *mode deprojection* power spectrum

$$\hat{P}(k_\alpha) = \left| F(k_\alpha) - \sum_{AB} \mathbf{S}_A \mathbf{R}_{AB}^{-1} f_B(k_\alpha) \right|^2, \quad (36)$$

where $\mathbf{R}_{AB} \equiv \sum_{\mu} \frac{f_A^*(k_\mu) f_B(k_\mu)}{P(k_\mu)}$ and $\mathbf{S}_A \equiv \sum_{\alpha} \frac{f_A(k_\alpha) F^*(k_\alpha)}{P(k_\alpha)}$ are matrix and vector equivalents of R_P and S_P , respectively, in contaminant space.

To apply multiple mode subtraction, we extend the likelihood given in equation (24) to

$$-2 \ln \mathcal{L} = \sum_{\alpha} \frac{|F(k_\alpha) - \sum_A \varepsilon_A f_A(k_\alpha)|^2}{P(k_\alpha)}. \quad (37)$$

Writing ε as a vector, the joint maximum likelihood solution fitting all contaminants is given by (cf. Appendix B)

$$\varepsilon^{(\text{BF})} = \mathbf{R}^{-1} \mathbf{S}. \quad (38)$$

Note that this would require fitting the amplitude of all contaminants simultaneously. The absolute value squared of the best-fitting signal is hence equal to equation (36). Hence, we also do not need large $N_{\text{mode}} \times N_{\text{mode}}$ matrices when we have to remove several potential contaminants.

We can calculate the debiasing factor

$$\sum_j \tilde{\mathbf{N}}_{ij} P^2(k_i) = 1 - \sum_{AB} \frac{f_A(\mathbf{k}_i) \mathbf{R}_{AB}^{-1} f_B^*(\mathbf{k}_i)}{P(k_i)} \quad (39)$$

analogously to Section 4 from the *mode deprojection* normalization matrix without binning.

7 TESTING CONTAMINANT REMOVAL

In this section, we show how simple contaminants can be removed in power spectrum measurements from simulated density fields, using the hitherto described methodologies.

7.1 Gaussian spike contaminant

As a first test, we generate three-dimensional Gaussian random fields according to an input power spectrum that we calculate using CAMB (Lewis, Challinor & Lasenby 2000). Each of these fields consists of a $16 \times 16 \times 16$ grid, in a box of length $3136 h^{-1}$ Mpc. An example of such a field is shown in the left-hand panel of Fig. 1. We contaminate these Gaussian random fields by adding a real Gaussian spike in k -space with width $\sigma^2 = 10^{-5} h \text{Mpc}$, centred around $k = 0.01 h \text{Mpc}^{-1}$, such that its maximum lies within a bin with sufficiently good statistics. The Fourier transform of this contaminant field is again a Gaussian spike in the centre of the box with some long-wavelength fluctuations around it. The amplitude of the real part overdensity in k -space is 100, thus having the same order of magnitude as the ‘true’ density field. An example of this setup can be seen in the central panel of Fig. 1. We calculate four different power spectra.

- (i) We do not account for the contaminants and just average the absolute values squared of the density field in each bin (cf. equation 11).
- (ii) We perform a naïve *mode subtraction*, i.e. we subtract off the template, but do not debias the two-point function (cf. equation 28).
- (iii) We debias the previous power spectrum by applying equation (30).
- (iv) We use the full QML estimator with *mode deprojection*.

In the cases (ii) to (iv), we have to assume a prior power spectrum, which we take as equal to the input power. We shall test the effect of this assumption with the next example. As each bin contains modes with a range of different k -values, we have to clarify what we mean by the prior power spectrum $P(k_i)$ for a specific bin. We find that the power spectrum measurements are closest to the input values, when we assume that the input power spectrum $P(k_i)$ is given by the average of the prior power spectrum values for each mode in the respective bin, i.e.

$$P(k_i) \equiv \frac{1}{N_{\mathbb{k}_i}} \sum_{\mathbf{k}_\alpha \in \mathbb{k}_i} P(k_\alpha). \quad (40)$$

In Fig. 2, we can clearly see an increase of power in the bins around $k = 0.01$ in case (i). Subtracting off the template in the naïve way (method (ii)) is biased in the bins affected by the spike. However, this bias is only a one part in a thousand effect. Methods (iii) and (iv) both reproduce the input power spectrum well, removing

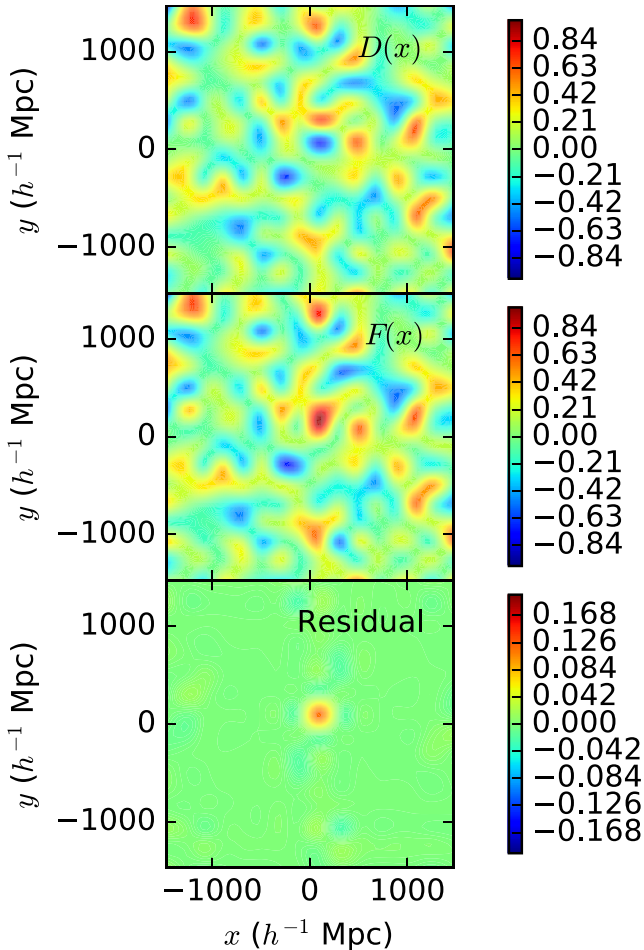


Figure 1. A slice through a realization of a Gaussian random field contaminated with a Gaussian spike used in Section 7.1. The top panel shows the ‘clean’ Gaussian random field (corresponding to $D(\mathbf{k})$ through Fourier transform) in configuration space. In the central panel, we have plotted the contaminated field (Fourier pair of $F(\mathbf{k})$) with an obvious Gaussian overdensity in the centre. The bottom panel shows the residual, i.e. the difference of the field after *mode subtraction* [i.e. the Fourier transform of $F(\mathbf{k}) - \varepsilon^{(\text{BF})} f(\mathbf{k})$, cf. equation (28)] and the input field. The best-fitting $\varepsilon^{(\text{BF})}$ for this particular realization amounts to 1.078. Although differences between the top and bottom panels are hard to spot by eye, in Fourier space the differences correspond to the bias in the *mode subtraction* estimator.

the bias. A significant difference between their error bars cannot be observed. It is therefore sufficient in this case to use the FKP-style estimator we introduced in Section 5.

7.2 Single contaminated mode

As a second example we use equation (30) to construct a contaminant that would lead to a strong bias in the recovered $P(k)$ without the debiasing step. Equation (30) only contains positive quantities and is normalized such that the bias is a value between 0 and 1. 1 corresponds to an unbiased estimate; hence, 0 is the maximal bias. This extreme case would be fulfilled if f is large for one mode and 0 otherwise. Therefore, we construct a contaminant that is a large number at the modes corresponding to $\mathbf{k} = \pm(0.003, 0.003, 0.003) h \text{ Mpc}^{-1}$. An example of this setup can be found in Fig. 3. The top panel again shows an uncontaminated Gaussian random field; the central

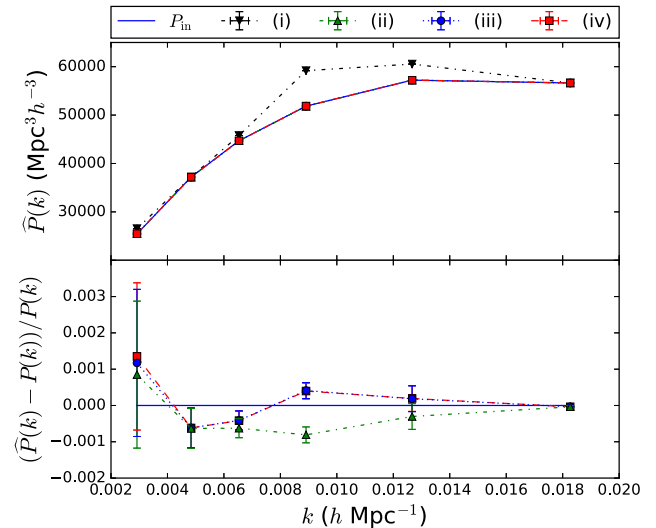


Figure 2. Means and standard deviations of the power spectra of 70 000 realizations of Gaussian random fields contaminated with a real Gaussian spike. The top panel shows the input power spectrum as a solid blue line, as well as the power spectra obtained with methods (i)–(iv) as described in Section 7.1. In the lower panel, we plot fractional errors for methods (i)–(iv).

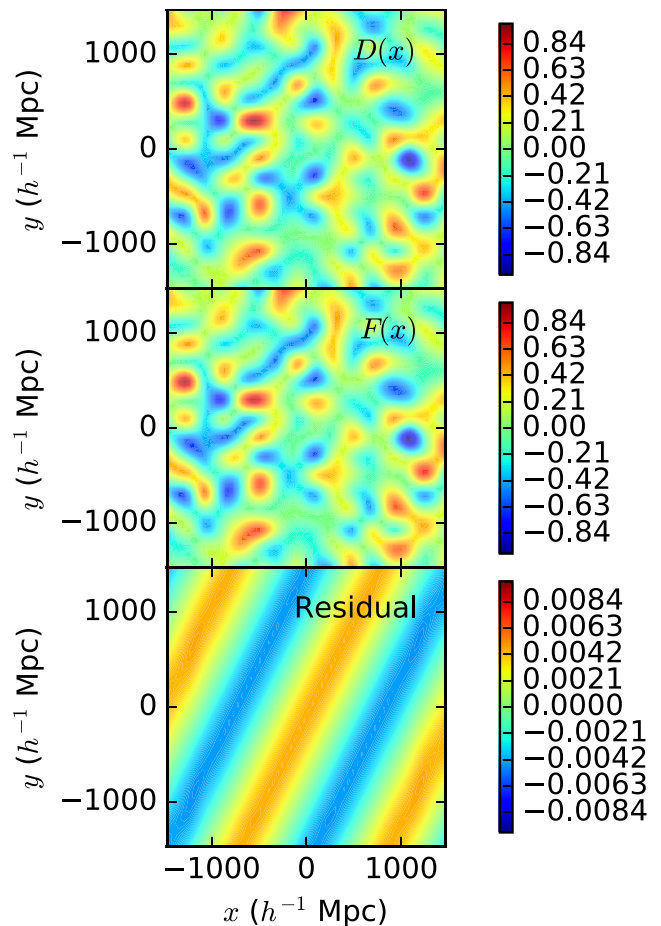


Figure 3. This plot is similar to Fig. 1, but shows a slice through a field with a single contaminated mode as described in Section 7.2. The best-fitting $\varepsilon^{(\text{BF})}$ for this particular realization amounts to 1.005. All panels appear very similar; this is quantified in Fig. 4.

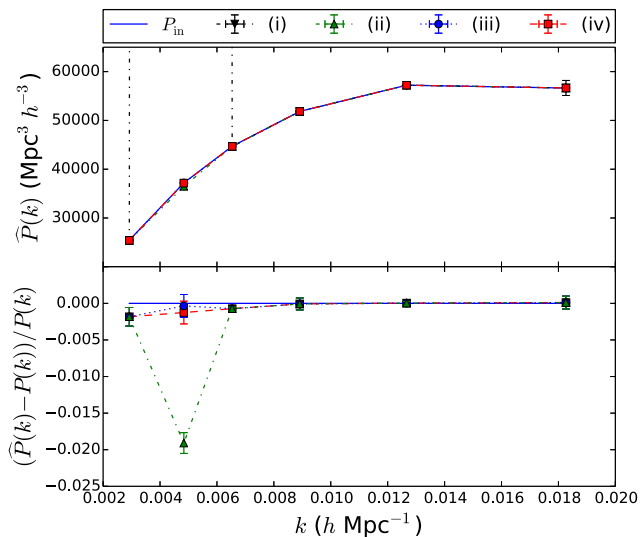


Figure 4. Means and standard deviations of the power spectra of 1000 realizations of Gaussian random fields contaminated with Hermitian Gaussian spikes. The red dots represent measurements, where the contamination has not been taken into account. For the blue dots, *mode deprojection* has been used to remove the spikes. For the green dots, we used *debaised mode subtraction*. The solid blue line shows the input power spectrum.

panel shows the same field with the contaminant added. The contaminant itself is not as prominent as the one in Fig. 1, because this single contaminated mode just adds a long-wavelength contribution in real space. The bottom panel shows the field after subtracting the template.

We measure the same cases (i)–(iv) as in the previous subsection, which we plot in Fig. 4. The prior power is again the input power. If we were to apply this to a real survey, we would not know the true power, so we perform a few runs, where we first assume a flat prior power spectrum $P(k) = 1 \forall k$, and then iteratively compute the power with the power from the previous run as the prior power spectrum. The effect of the prior power spectrum is negligible, because the result in the first step provides the same result as assuming the input power as prior.

The data points for all cases (i)–(iv) are close to the input power in all bins but the second. In the second bin, the power spectrum for case (i) extends beyond the plotted range, chosen to highlight differences between the other approaches. In case (ii), the power is significantly underestimated. The bias amounts to about 2 per cent, i.e. it highly affects measurements where small- k modes are crucial, such as f_{NL} measurements. The difference between the cases (iii) and (iv) is much smaller, even in this extreme example.

8 CONCLUSIONS

We have considered methods to remove contaminants when measuring the 3D galaxy power spectrum from a given density field, focusing on *mode deprojection* and *mode subtraction*. In order to understand how these are related, we have decomposed the problem into separate steps. In particular, we have separated *mode deprojection* from power spectrum estimation – they are often considered together – arguing that this split makes sense given the mathematical equivalence of *mode deprojection* and *mode subtraction*. We argue that the QML estimation is not practical for modern surveys with

large numbers of observed modes, but that we can apply *mode deprojection* to the FKP estimator, using the mathematical equivalence of *mode deprojection* and *mode subtraction*, thus avoiding having to create large estimator and covariance matrices for all modes. The resulting estimate is biased, but can easily be made unbiased with a simple correction, again that can be implemented without the inversion of large matrices. This correction is easily extended to the case of multiple contaminants and is not affected if the modes are correlated even without the effects of contaminants. The final result of our short paper is the suggestion that 3D galaxy power spectrum should be estimated using equation (31),

$$\hat{P}(k_i) = \frac{1}{N_{k_i}} \sum_{k_\alpha} \frac{\left| F(k_\alpha) - \frac{S_p}{R_p} f(k_\alpha) \right|^2}{1 - \frac{1}{R_p} \frac{|f(k_\alpha)|^2}{P(k_\alpha)}}. \quad (41)$$

While theoretically it is sub-optimal, in practice the degradation of signal is expected to be less than ignoring window effects in the optimization of mode averaging when using the standard FKP estimator.

ACKNOWLEDGEMENTS

We would like to thank Franz Elsner, Héctor Gil-Marín and Ashley Ross for valuable comments.

We used matplotlib (Hunter 2007) to generate plots. The CAMB package (Lewis et al. 2000) has been used to generate model and prior power spectra. We made use of the facilities and staff of the UK Sciama High Performance Computing cluster supported by the ICG, SEPNet and the University of Portsmouth.

BK thanks the Faculty of Technology of the University of Portsmouth for support during his PhD studies. WJP and DB acknowledge support from UK STFC through the consolidated grant ST/K0090X/1. WJP also acknowledges support from the European Research Council through the Darksurvey grant and the UK Space Agency through grant ST/N00180X/1. LS is grateful for support from SNSF grant SCOPES IZ73Z0-152581, GNSF grant FR/339/6-350/14 and DOE grant DEFG 03-99EP41093.

REFERENCES

- Bennett C. et al., 2003, ApJS, 148, 97
- Bianchi D., Gil-Marín H., Ruggeri R., Percival W. J., 2015, MNRAS, 453, L11
- Elsner F., Leistedt B., Peiris H. V., 2016, MNRAS, 456, 2095
- Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
- Ho S. et al., 2012, ApJ, 761, 14
- Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
- Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
- Leistedt B., Peiris H. V., 2014, MNRAS, 444, 2
- Leistedt B., Peiris H. V., Roth N., 2014, Phys. Rev. Lett., 113, 221301
- Levi M. et al., 2013, preprint (arXiv:1308.0847)
- Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
- Ross A. J. et al., 2012, MNRAS, 424, 564
- Ross A. J. et al., 2016, MNRAS, in press
- Rybicki G. B., Press W. H., 1992, ApJ, 398, 169
- Schlegel D. et al., 2011, preprint (arXiv:1106.1706)
- Scoccimarro R., 2015, Phys. Rev. D, 92, 083532
- Sherman J., Morrison W. J., 1950, Ann. Math. Stat., 21, 124
- Slosar A., Seljak U., Makarov A., 2004, Phys. Rev. D, 69, 123003
- Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, ApJ, 499, 555

APPENDIX A: DERIVATION OF MODE DEPROJECTION WITH MULTIPLE TEMPLATES

In this appendix, we want to derive equation (36) from equation (35). We start by rewriting equation (35) in matrix notation

$$\tilde{\mathbf{C}} = \mathbf{C} + \lim_{\sigma \rightarrow \infty} \sigma \mathbf{f} \mathbf{I}_{N_{\text{sys}}} \mathbf{f}^\dagger, \quad (\text{A1})$$

defining an $N_{\text{mode}} \times N_{\text{sys}}$ matrix $\mathbf{f}_{\alpha A} \equiv f_A(\mathbf{k}_\alpha)$, such that we can invert $\tilde{\mathbf{C}}$ using the Woodbury matrix identity

$$\begin{aligned} \tilde{\mathbf{C}}^{-1} &= \mathbf{C}^{-1} - \mathbf{C}^{-1} \lim_{\sigma \rightarrow \infty} \sigma \mathbf{f} \left(\mathbf{I}_{N_{\text{sys}}}^{-1} + \mathbf{f}^\dagger \mathbf{C}^{-1} \sigma \mathbf{f} \right)^{-1} \mathbf{f}^\dagger \mathbf{C}^{-1} \\ &= \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{f} \left(\mathbf{f}^\dagger \mathbf{C}^{-1} \mathbf{f} \right)^{-1} \mathbf{f}^\dagger \mathbf{C}^{-1} \\ &\equiv \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{f} \mathbf{R}^{-1} \mathbf{f}^\dagger \mathbf{C}^{-1}. \end{aligned} \quad (\text{A2})$$

If we assume $\mathbf{C}_{\alpha\beta} = \delta_{\alpha\beta} P(k_\alpha)$, $\mathbf{R} \equiv \mathbf{f}^\dagger \mathbf{C}^{-1} \mathbf{f}$ becomes a matrix equivalent to the factor R_p in the previous sections:

$$\mathbf{R}_{AB} = \sum_{\mu\nu} f_A^*(\mathbf{k}_\mu) \frac{\delta_{\mu\nu}}{P(k_\mu)} f_B(k_\nu) = \sum_{\mu} \frac{f_A^*(\mathbf{k}_\mu) f_B(k_\mu)}{P(k_\mu)}. \quad (\text{A3})$$

The inverse updated covariance matrix then reads

$$\tilde{\mathbf{C}}_{\alpha\beta}^{-1} = \frac{\delta_{\alpha\beta}}{P(k_\alpha)} - \sum_{AB} \frac{f_A(\mathbf{k}_\alpha) \mathbf{R}_{AB}^{-1} f_B^*(\mathbf{k}_\beta)}{P(k_\alpha) P(k_\beta)}. \quad (\text{A4})$$

If we do not bin, but apply mode deprojection to each mode separately, the matrix $\tilde{\mathbf{E}}$ simplifies to

$$\tilde{\mathbf{E}}_{\alpha\beta}(k_j) = \sum_{\mu\nu} \tilde{\mathbf{C}}_{\alpha\mu}^{-1} \delta_{\mu j} \delta_{\nu\beta} \tilde{\mathbf{C}}_{\nu\beta}^{-1} = \tilde{\mathbf{C}}_{\alpha j}^{-1} \tilde{\mathbf{C}}_{j\beta}^{-1}. \quad (\text{A5})$$

After inserting equation (A4) into equation (A5), we obtain

$$\begin{aligned} &P^2(k_j) \sum_{\alpha\beta} F^*(\mathbf{k}_\alpha) \tilde{\mathbf{E}}_{\alpha\beta}(k_j) F(\mathbf{k}_\beta) \\ &= |F(\mathbf{k}_j)|^2 \\ &\quad - \sum_{AB\alpha} \frac{F^*(\mathbf{k}_\alpha) f_A(\mathbf{k}_\alpha)}{P(k_\alpha)} \mathbf{R}_{AB}^{-1} f_B^*(\mathbf{k}_j) F(\mathbf{k}_j) \\ &\quad - \sum_{AB\beta} F^*(\mathbf{k}_j) f_A(\mathbf{k}_j) \mathbf{R}_{AB}^{-1} \frac{f_B^*(\mathbf{k}_\beta) F(\mathbf{k}_\beta)}{P(k_\beta)} \\ &\quad + \sum_{ABCD\alpha\beta} \frac{F^*(\mathbf{k}_\alpha) f_A(\mathbf{k}_\alpha)}{P(k_\alpha)} \mathbf{R}_{AB}^{-1} f_B^*(\mathbf{k}_j) f_C(\mathbf{k}_j) \mathbf{R}_{CD}^{-1} \frac{f_D^*(\mathbf{k}_\beta) F(\mathbf{k}_\beta)}{P(k_\beta)} \\ &= |F(\mathbf{k}_j)|^2 \\ &\quad - 2 \text{Re} \left[\sum_{AB} \mathbf{S}_A \mathbf{R}_{AB}^{-1} f_B^*(\mathbf{k}_j) F(\mathbf{k}_j) \right] \\ &\quad + \left| \sum_{AB} \mathbf{S}_A \mathbf{R}_{AB}^{-1} f_B(\mathbf{k}_j) \right|^2 \\ &= \left| F(\mathbf{k}_j) - \sum_{AB} \mathbf{S}_A \mathbf{R}_{AB}^{-1} f_B(\mathbf{k}_j) \right|^2, \end{aligned} \quad (\text{A6})$$

where we defined $\mathbf{S}_A \equiv \sum_{\alpha} \frac{f_A(\mathbf{k}_\alpha) F^*(\mathbf{k}_\alpha)}{P(k_\alpha)}$ analogous to S_p .

APPENDIX B: DERIVATION OF MODE SUBTRACTION WITH MULTIPLE TEMPLATES

Here we derive the best-fitting $\varepsilon^{(\text{BF})}$ from the likelihood

$$-2 \ln \mathcal{L} = \sum_{\alpha} \frac{|F(\mathbf{k}_\alpha) - \sum_A \varepsilon_A f_A(\mathbf{k}_\alpha)|^2}{P(k_\alpha)} \quad (\text{B1})$$

to find the same result as in the previous appendix. Taking the derivative with respect to ε_B yields

$$\frac{\partial \chi^2}{\partial \varepsilon_B} = -2 \varepsilon_B \sum_{\alpha} \frac{f_B(\mathbf{k}_\alpha) F^*(\mathbf{k}_\alpha) - \sum_A \varepsilon_A f_B(\mathbf{k}_\alpha) f_A^*(\mathbf{k}_\alpha)}{P(k_\alpha)}. \quad (\text{B2})$$

This derivative is zero if

$$\sum_{\alpha} \frac{f_B(\mathbf{k}_\alpha) F^*(\mathbf{k}_\alpha)}{P(k_\alpha)} = \sum_{A\alpha} \frac{\varepsilon_A f_B(\mathbf{k}_\alpha) f_A^*(\mathbf{k}_\alpha)}{P(k_\alpha)}, \quad (\text{B3})$$

which reads

$$\mathbf{S} = \mathbf{R} \varepsilon \quad (\text{B4})$$

in matrix notation. The best-fitting ε value is therefore given by

$$\varepsilon^{(\text{BF})} = \mathbf{R}^{-1} \mathbf{S}. \quad (\text{B5})$$

The absolute value squared of the best-fitting signal is hence equal to equation (36):

$$\left| F(\mathbf{k}_\alpha) - \sum_A \varepsilon_A^{(\text{BF})} f_A(\mathbf{k}_\alpha) \right|^2 = \left| F(\mathbf{k}_\alpha) - \sum_{AB} \mathbf{R}_{AB}^{-1} \mathbf{S}_B f_A(\mathbf{k}_\alpha) \right|^2. \quad (\text{B6})$$

APPENDIX C: MODE SUBTRACTION AND THE DEBIASING STEP FOR A NON-DIAGONAL COVARIANCE MATRIX

C1 Including the covariance in the calculation of the best-fitting mode to subtract

Here we show a derivation similar to the one in Sections 4 and 5 for the more general case of a non-diagonal covariance matrix. We show that the debiasing works in the same way as in Section 5, just with a generalized definition of R_p .

Defining the covariance matrix of the true density

$$\mathbf{C}_{\alpha\beta} \equiv \langle D_\alpha D_\beta^* \rangle \quad (\text{C1})$$

and assuming that the true signal and the contaminant are uncorrelated, we can write

$$\langle F_\alpha F_\beta^* \rangle = \mathbf{C}_{\alpha\beta} + \varepsilon_{\text{true}}^2 f_\alpha f_\beta^*. \quad (\text{C2})$$

As we did in Section 4, we introduce a free parameter ε , such that

$$\hat{D}_\alpha \equiv F_\alpha - \varepsilon f_\alpha. \quad (\text{C3})$$

Assuming that the true density field is Gaussian, its log-likelihood reads

$$-2 \ln \mathcal{L} = \sum_{\alpha\beta} (F_\alpha - \varepsilon f_\alpha)^* \mathbf{C}_{\alpha\beta}^{-1} (F_\beta - \varepsilon f_\beta) + \text{const.} \quad (\text{C4})$$

To find the best-fitting $\varepsilon^{(\text{BF})}$, we take the derivative of the log-likelihood with respect to ε :

$$\begin{aligned}
 -2 \frac{\partial \ln \mathcal{L}}{\partial \varepsilon} &= - \sum_{\alpha\beta} f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} (F_{\beta} - \varepsilon f_{\beta}) \\
 &\quad - \sum_{\alpha\beta} (F_{\alpha} - \varepsilon f_{\alpha})^* \mathbf{C}_{\alpha\beta}^{-1} f_{\beta} \\
 &= 2\varepsilon \sum_{\alpha\beta} f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\beta} \\
 &\quad - \sum_{\alpha\beta} [f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} F_{\beta} + F_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\beta}]. \quad (C5)
 \end{aligned}$$

As \mathbf{C} is a covariance matrix of complex random variables, it is Hermitian positive-semidefinite, such that the second sum can be written as

$$\sum_{\alpha\beta} [f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} F_{\beta} + F_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\beta}] = 2 \operatorname{Re} \left[\sum_{\alpha\beta} f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} F_{\beta} \right]. \quad (C6)$$

For shortness and in analogy with Section 4, we call this sum

$$S_P \equiv \sum_{\alpha\beta} \operatorname{Re} [f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} F_{\beta}] \quad (C7)$$

and the first sum in equation (C5) we call

$$R_P \equiv \sum_{\alpha\beta} f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\beta}. \quad (C8)$$

We obtain the best-fitting, i.e. maximum likelihood, value

$$\varepsilon^{(\text{BF})} = \frac{S_P}{R_P} \quad (C9)$$

by equating equation (C5) to zero.

Now we want to calculate the expectation value

$$\begin{aligned}
 \left\langle \left| F_{\alpha} - \frac{S_P}{R_P} f_{\alpha} \right|^2 \right\rangle &= \langle |F_{\alpha}|^2 \rangle - \frac{2}{R_P} \operatorname{Re} [\langle S_P F_{\alpha}^* f_{\alpha} \rangle] \\
 &\quad + \left\langle \frac{S_P^2}{R_P^2} \right\rangle |f_{\alpha}|^2. \quad (C10)
 \end{aligned}$$

We calculate each term separately.

(i) The first term $\langle |F_{\alpha}|^2 \rangle = \mathbf{C}_{\alpha\alpha} + \varepsilon_{\text{true}}^2 |f_{\alpha}|^2$ is a special case of equation (C2).

(ii) To calculate the second term, we re-expand S_P and use the fact that $\operatorname{Re} [F_{\alpha} F_{\alpha}^*] = \operatorname{Re} [F_{\alpha}^* f_{\alpha}]$:

$$2 \operatorname{Re} [\langle S_P F_{\alpha}^* f_{\alpha} \rangle] = 2 \operatorname{Re} \left[\sum_{\gamma\beta} f_{\gamma}^* \mathbf{C}_{\gamma\beta}^{-1} \langle F_{\beta} F_{\alpha}^* \rangle f_{\alpha} \right]. \quad (C11)$$

After re-inserting equation (C2), we get

$$\begin{aligned}
 2 \operatorname{Re} [\langle S_P F_{\alpha}^* f_{\alpha} \rangle] &= 2 \operatorname{Re} \left[\sum_{\gamma\beta} f_{\gamma}^* \mathbf{C}_{\gamma\beta}^{-1} \mathbf{C}_{\beta\alpha} f_{\alpha} \right. \\
 &\quad \left. + \varepsilon_{\text{true}}^2 \sum_{\gamma\beta} f_{\gamma}^* \mathbf{C}_{\gamma\beta}^{-1} f_{\beta} f_{\alpha}^* f_{\alpha} \right]. \quad (C12)
 \end{aligned}$$

In the first term, we have $\sum_{\beta} \mathbf{C}_{\gamma\beta}^{-1} \mathbf{C}_{\beta\alpha} = \delta_{\gamma\alpha}$, and in the second term we find the definition of R_P . Thus, the second term of equation (C10) is

$$2 \operatorname{Re} [\langle S_P F_{\alpha}^* f_{\alpha} \rangle] = 2 |f_{\alpha}|^2 (1 + \varepsilon_{\text{true}}^2 R_P). \quad (C13)$$

(iii) In the third term, we can again make use of equation (C2):

$$\begin{aligned}
 \langle S_P^2 \rangle &= \sum_{\alpha\beta\gamma\delta} \operatorname{Re} [f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\gamma} \mathbf{C}_{\gamma\delta}^{-1} \langle F_{\beta} F_{\delta}^* \rangle] \\
 &= \sum_{\alpha\beta\gamma\delta} \operatorname{Re} [f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\gamma} \mathbf{C}_{\gamma\delta}^{-1} \mathbf{C}_{\beta\delta} + \varepsilon_{\text{true}}^2 f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\gamma} \mathbf{C}_{\gamma\delta}^{-1} f_{\beta} f_{\delta}^*]. \quad (C14)
 \end{aligned}$$

In the first term, we have again $\sum_{\beta} \mathbf{C}_{\alpha\beta}^{-1} \mathbf{C}_{\beta\delta} = \delta_{\alpha\delta}$, and the second term is equal to R_P^2 , such that

$$\langle S_P^2 \rangle = \sum_{\alpha\gamma} \operatorname{Re} [f_{\alpha}^* f_{\gamma} \mathbf{C}_{\gamma\alpha}^{-1}] + \varepsilon_{\text{true}}^2 R_P^2 = R_P + \varepsilon_{\text{true}}^2 R_P^2. \quad (C15)$$

Recollecting (i)–(iii) and inserting into equation (C10) yields

$$\begin{aligned}
 \left\langle \left| F_{\alpha} - \frac{S_P}{R_P} f_{\alpha} \right|^2 \right\rangle &= \mathbf{C}_{\alpha\alpha} + \varepsilon_{\text{true}}^2 |f_{\alpha}|^2 \\
 &\quad - 2 |f_{\alpha}|^2 \left(\varepsilon_{\text{true}}^2 + \frac{1}{R_P} \right) + \left(\varepsilon_{\text{true}}^2 + \frac{1}{R_P} \right) |f_{\alpha}|^2 \\
 &= \mathbf{C}_{\alpha\alpha} - \frac{|f_{\alpha}|^2}{R_P}. \quad (C16)
 \end{aligned}$$

As the power spectrum $P(k_{\alpha}) = \mathbf{C}_{\alpha\alpha}$ is defined as the diagonal elements of the covariance matrix, the debiasing step is the same for a non-diagonal covariance matrix as for a diagonal one (cf. Section 5); we just have to use the generalized definition of R_P as in equation (C8).

C2 The independent mode approximation

We have seen in Appendix C1 that mode subtraction also works when the covariance matrix is non-diagonal. However, to compute the generalized R_P , one has to invert the full $N_{\text{mode}} \times N_{\text{mode}}$ covariance matrix, which makes this approach computationally almost as expensive as using the QML estimator. We will argue that, in most cases, equation (31) provides a good estimate of the power, even in the presence of covariant modes, and we will provide a further correction term that corrects for using equation (31) when off-diagonal covariances are important.

Suppose we apply equation (31) assuming a diagonal covariance matrix, even though there are covariances between different modes. Then, we find a best-fitting

$$\varepsilon'_{\text{BF}} = \frac{\sum_{\alpha} \frac{F_{\alpha}^* f_{\alpha}}{P_{\alpha}}}{\sum_{\mu} \frac{|f_{\mu}|^2}{P_{\mu}}} \quad (C17)$$

instead of the true

$$\varepsilon_{\text{BF}} = \frac{\sum_{\alpha\beta} f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} F_{\beta}}{\sum_{\alpha\beta} f_{\alpha}^* \mathbf{C}_{\alpha\beta}^{-1} f_{\beta}}. \quad (C18)$$

The expectations are the same $\langle \varepsilon'_{\text{BF}} \rangle = \langle \varepsilon_{\text{BF}} \rangle = \varepsilon_{\text{true}}$, but their variances are different. For the approximate estimate, we have

$$\begin{aligned}
 \langle \varepsilon_{\text{BF}}^2 \rangle &= \frac{\left\langle \sum_{\alpha\beta} \frac{F_{\alpha}^* f_{\alpha} F_{\beta} f_{\beta}^*}{P_{\alpha} P_{\beta}} \right\rangle}{R_P^2} \\
 &= \frac{\sum_{\alpha\beta} \frac{f_{\alpha} \mathbf{C}_{\alpha\beta} f_{\beta}^*}{P_{\alpha} P_{\beta}}}{R_P^2} + \frac{\sum_{\alpha\beta} \frac{\varepsilon_{\text{true}}^2 |f_{\alpha}|^2 |f_{\beta}|^2}{P_{\alpha} P_{\beta}}}{R_P^2} \\
 &= \frac{1}{R_P^2} \sum_{\alpha\beta} \frac{f_{\alpha} \mathbf{C}_{\alpha\beta} f_{\beta}^*}{P_{\alpha} P_{\beta}} + \varepsilon_{\text{true}}^2. \quad (C19)
 \end{aligned}$$

Unlike in the previous estimates, the covariance matrix does not cancel in the first term. Similarly,

$$\begin{aligned} \langle \varepsilon'_{\text{BF}} F_{\alpha}^* f_{\alpha} \rangle &= \frac{1}{R'_p} \left\langle \sum_{\beta} \frac{f_{\beta}^* F_{\beta} F_{\alpha}^* f_{\alpha}}{P_{\beta}} \right\rangle \\ &= \frac{1}{R'_p} \sum_{\beta} \frac{f_{\alpha} \mathbf{C}_{\alpha\beta} f_{\beta}^*}{P_{\beta}} + \varepsilon_{\text{true}}^2 |f_{\alpha}|^2. \end{aligned} \quad (\text{C20})$$

Combining the previous two equations, we obtain

$$\begin{aligned} \langle |F_{\alpha} - \varepsilon'_{\text{BF}} f_{\alpha}|^2 \rangle &= C_{\alpha\alpha} - \frac{2}{R'_p} \sum_{\beta} \frac{f_{\alpha} \mathbf{C}_{\alpha\beta} f_{\beta}^*}{P_{\beta}} \\ &\quad + \frac{|f_{\alpha}|^2}{R_p'^2} \sum_{\gamma\beta} \frac{f_{\gamma} \mathbf{C}_{\gamma\beta} f_{\beta}^*}{P_{\gamma} P_{\beta}}. \end{aligned} \quad (\text{C21})$$

Splitting the covariance matrix

$$C_{\alpha\beta} = P_{\beta} (\delta_{\alpha\beta} + \Delta_{\alpha\beta}) \quad (\text{C22})$$

into diagonal and off-diagonal elements yields

$$\begin{aligned} \langle |F_{\alpha} - \varepsilon'_{\text{BF}} f_{\alpha}|^2 \rangle &= P_{\alpha} - \frac{|f_{\alpha}|^2}{R_p'} \\ &\quad \times \left[1 + \sum_{\gamma\beta} f_{\gamma} \Delta_{\gamma\beta} f_{\beta}^* \left(\frac{2\delta_{\alpha\gamma}}{|f_{\alpha}|^2} - \frac{1}{R_p' P_{\gamma}} \right) \right]. \end{aligned} \quad (\text{C23})$$

Hence, one can perform mode subtraction assuming a diagonal covariance matrix and then apply another correction term which is linear in its off-diagonal elements. The advantage of this procedure is that it does not require any inversion of the N_{mode}^2 covariance matrix. If the off-diagonal elements are small, then the bias correction reverts back to the form of equation (30).

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