

Systematics mitigation for catalogue-based angular power spectra

Thomas Cornish,^{1,2*} David Alonso,¹ Boris Leistedt² and Kevin Wolz¹

¹*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom*

²*Department of Physics, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, United Kingdom*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Recent work has developed a formalism for computing angular power spectra directly from catalogues containing field values at discrete positions on the sky, thereby circumventing the need to create pixelised maps of the fields, as well as avoiding aliasing and finite-resolution effects. We adapt this formalism to incorporate template deprojection for mitigating systematic biases in the measured angular power spectra. We also introduce an alternative method of mitigating the ‘deprojection bias’ – the loss of modes induced by deprojection – employing simple simulations to compute a transfer function. We find that this approach performs at least as well as existing methods, and is relatively insensitive to how well one can guess the true power spectrum of the observed field, except at the largest scales ($\ell \lesssim 3$). Additionally, we develop exact expressions for the bias introduced by deprojection in the shot-noise component, which further improves the accuracy of this approach. We test our formalism on simulated datasets, demonstrating its applicability both to discretely sampled fields, and to the special case of galaxy clustering, with the survey selection function defined in terms of a random catalogue or as a continuous sky map. After removing the bias in the shot noise and correcting for the remaining mode loss using a transfer function, our formalism produces unbiased measurements of the angular power spectrum in all scenarios tested here. Finally, we apply our formalism to real data and show it produces results consistent with the standard map-based pseudo- C_ℓ formalism. We implement our method in the public code **NaMaster**.

Key words: methods: data analysis – methods: numerical

1 INTRODUCTION

The accurate measurement of projected fields on the sky has become a cornerstone of modern cosmology, with perhaps the most famous example being the cosmic microwave background (CMB), for which measurements of anisotropies in its temperature and polarisation enable constraints on models of inflation and the primordial density perturbations, as well as on the natures of neutrinos and dark energy (e.g. Planck Collaboration et al. 2020; Louis et al. 2025; Camphuis et al. 2025). Our knowledge of the large-scale structure (LSS) in the Universe is also heavily dependent on our ability to measure projected fields on the sky: tomographic measurements of galaxy clustering exploit the fact that galaxies are biased tracers of the underlying dark matter distribution while leveraging photometric redshifts as a means of estimating their radial positions, thereby enabling indirect measurements of the matter density field in ‘slices’ across cosmic time without the need for spectroscopic observations, which are time-consuming and costly across large areas (e.g. Nicola et al. 2020; Rodríguez-Monroy et al. 2022; Abbott et al. 2022; Vakili et al. 2023; Sugiyama et al. 2023; Yan et al. 2025). Furthermore, the effects of weak gravitational lensing can be expressed in terms of projected fields, such as the lensing shear and convergence, both of which are sensitive probes of the intervening dark matter distribution (Heymans et al. 2021; Abbott et al. 2022; Miyatake et al. 2023; Stölzner et al. 2025; Faga et al. 2025). These are only two examples

of a myriad other projected probes of the LSS, which include other CMB secondary anisotropies, the Cosmic Infrared Background, etc. (e.g. Stompor & Efstathiou 1999; Adam et al. 2017; Bocquet et al. 2019; Zubeldia & Challinor 2019; Maniyar et al. 2019; Carron et al. 2022; Bleem et al. 2022; Jegou et al. 2023; Fang et al. 2024; Shaikh et al. 2024; Qu et al. 2024; Coulton et al. 2024; Ge et al. 2025).

It is no surprise, then, that the angular power spectrum (denoted C_ℓ), which quantifies the covariance between angular fluctuations in two fields as a function of angular multipole ℓ (and encodes all information in the case of Gaussian fields), has become an essential summary statistic in cosmology. Unbiased estimators of the angular power spectrum have been extensively explored, and optimal quadratic minimum variance (QMV) estimators have been developed which are capable of recovering angular power spectra with minimal loss of information (Tegmark 1997). However, these estimators require the inversion of a pixel-space covariance matrix in order to optimally weight the data, an operation which can be computationally expensive when working with high-resolution maps. The prohibitive computational demand of these estimators often necessitates an alternative, an example of which can be found in the pseudo- C_ℓ (PCL) algorithm (Hivon et al. 2002; Alonso et al. 2019). The PCL approach simplifies things by assuming a diagonal data covariance matrix – an assumption which holds if the data in question are spatially uncorrelated and/or dominated by white noise (Leistedt et al. 2013). This simplification significantly improves the speed of computation, making the PCL algorithm an often attractive approach to C_ℓ estimation.

The characterisation and mitigation of systematic effects is becom-

* E-mail: thomas.cornish@physics.ox.ac.uk

ing a problem of increasing relevance in cosmological surveys: as the hoards of data being collected continually increase in size, the statistical uncertainties have the potential to become comparable to those arising from systematics (e.g. Ho et al. 2012a; Pullen & Hirata 2013; Huterer et al. 2013; Agarwal et al. 2014; Shafer & Huterer 2015; Ho et al. 2015; Weaverdyck et al. 2018). A plethora of techniques have been developed to tackle this issue by attempting to quantify the dependence of the desired field on maps of the survey properties or other contaminants thought to be contributing to the observed signal. These range from techniques interpretable through a linear regression framework (Weaverdyck & Huterer 2021) such as mode deprojection (also called template deprojection; Rybicki & Press 1992; Slosar et al. 2004; Elsner et al. 2016, 2017), template subtraction (Ross et al. 2011; Ho et al. 2012b), and the iterative systematics decontamination (ISD) technique employed by the Dark Energy Survey (DES; Elvin-Poole et al. 2018; Rodríguez-Monroy et al. 2022), to more sophisticated methods employing machine learning tools such as artificial neural networks (Rezaie et al. 2020) or self-organising maps (Johnston et al. 2021).

The approaches discussed so far typically involve data in the form of pixelised maps that quantify the value of a projected field at a given position on the sky. However, it is often the case in cosmology that the fields whose angular power spectra we would like to measure are sampled at discrete positions on the sky, corresponding to the positions of sources in a catalogue. The typical course of action to obtain power spectrum measurements in this case is to bin these discrete points into pixels and produce a map of the mean field value in each pixel. However, recent work by Baleato Lizancos & White (2024); Wolz et al. (2025) allows one to circumvent this step, through an update to the PCL formalism which facilitates the computation of the PCL directly from catalogue data. This approach naturally avoids any bias introduced by the process of pixelising discretely sampled observables, which would otherwise influence measurements of the power spectrum at scales comparable to or smaller than the pixel scale. This extension to the PCL formalism was implemented in the publicly available code NaMaster¹ (Alonso et al. 2019), and its validity subsequently demonstrated on various discretely-sampled astrophysical fields, including cosmic shear, proper motions, dispersion measure of fast radio bursts (FRBs), and galaxy positions. However, these validation tests were performed using idealised mock data with no contamination due to systematics; indeed, the catalogue-based formalism as it is described in Wolz et al. (2025) includes no means of mitigating systematic effects.

The aim of this paper is thus to extend the catalogue-based PCL formalism of Wolz et al. (2025) in order to incorporate systematics mitigation, and to implement this in NaMaster. In particular, we adapt the template deprojection approach for applicability to catalogue data, as this is already implemented in NaMaster for map-based fields.

The structure of the paper is as follows. In Section 2 we outline the existing PCL formalism and its extension to catalogue-based data, before introducing our integration of template deprojection into the latter. In Section 3 we validate our approach using simulated data with systematic effects factored in. Section 4 contains our conclusions.

2 METHODS

2.1 Overview of pseudo- C_ℓ methods

This section contains a brief review of the pseudo- C_ℓ power spectrum estimator, including the concept of template deprojection, and the techniques used to extract power spectra directly from catalogues. This provides the background needed to present the extension of deprojection techniques to catalogue-based spectra described in Section 2.2. Readers are referred to Hivon et al. (2002); Alonso et al. (2019) for a more detailed account of these techniques.

We will focus our discussion on the calculation of auto-correlations of a given field, since many of the new results presented in this paper are most relevant in this scenario, and all others can be trivially generalised to the case of cross-correlations.

2.1.1 Standard pseudo- C_ℓ estimator

The pseudo- C_ℓ (PCL) or MASTER algorithm for power spectrum estimation Hivon et al. (2002); Alonso et al. (2019) is a simplified version of the optimal QMV estimator. The optimal inverse-variance weighting of the data map, prescribed by the QMV estimator, is a computationally expensive operation, effectively requiring the inversion of a large pixel-space covariance matrix. In the PCL approach, this matrix is assumed to be diagonal, turning the inverse variance weighting step into a simple multiplication of two maps: the data map containing the field measurements, and the “mask”, which plays the role of the local inverse variance of these measurements. Not only is this a significantly simpler and faster operation, but the statistical coupling between different angular multipoles induced by this weighting can be estimated analytically using efficient methods. This makes the PCL estimator easy to deploy on a large variety of datasets to extract fast power spectrum measurements. This is particularly beneficial on small scales, where the QMV estimator becomes prohibitively expensive.

The steps involved in the PCL estimator can be summarised as follows (see Alonso et al. (2019) for a more complete description of the algorithm)

(i) Let $\mathbf{a}(\hat{\mathbf{n}})$ be a spin- s field defined on the sphere, and let $v(\hat{\mathbf{n}})$ be its associated mask. Construct the masked field, given by the product $\mathbf{a}^v(\hat{\mathbf{n}}) \equiv v(\hat{\mathbf{n}}) \mathbf{a}(\hat{\mathbf{n}})$. Note that we denote spin- s quantities as a vector $\mathbf{a}(\hat{\mathbf{n}}) \equiv (a_1(\hat{\mathbf{n}}), a_2(\hat{\mathbf{n}}))$, where a_1 and a_2 are the two components of the field (e.g. the Q and U Stokes parameters for CMB polarisation maps, or the γ_1 and γ_2 components of the cosmic shear field).

(ii) Compute the spherical harmonic transform of the masked field, $\mathbf{a}_{\ell m}^v$, ignoring the presence of a mask.

(iii) Calculate the pseudo- C_ℓ of $\mathbf{a}_{\ell m}^v$, $\text{PCL}_\ell(\mathbf{a}^v, \mathbf{a}^v)$, defined as the naïve estimator of the field’s power spectrum, averaging over all values of m for a given ℓ , ignoring the fact that $\mathbf{a}^v$ is in fact a masked field. In general, for two masked maps $\mathbf{a}^v$ and $\mathbf{b}^w$, the PCL² is defined as

$$\text{PCL}_\ell(\mathbf{a}^v, \mathbf{b}^w) \equiv \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \mathbf{a}_{\ell m}^v \mathbf{b}_{\ell m}^{w\dagger}. \quad (1)$$

Since the presence of a sky mask leads to the statistical coupling of different angular multipoles, the PCL is also often called the “coupled” C_ℓ .

¹ <https://github.com/LSSTDESC/NaMaster>

² Note that we use “PCL” as shorthand for both the “pseudo- C_ℓ ” formalism and for the cut-sky estimator of the power spectrum defined here.

(iv) Calculate the “mode-coupling matrix” (MCM), the matrix encoding the statistical coupling between different ℓ modes induced by the sky mask, using the analytical expressions presented in e.g. [Alonso et al. \(2019\)](#). The MCM depends solely on the PCL of the mask. If binning the power spectrum into bandpowers, calculate also the binned MCM. Invert the MCM and multiply the PCL calculated in the previous step by it.

(v) In most cases, the measured field receives both signal and noise contributions: $\mathbf{a} = \mathbf{s}_a + \mathbf{n}_a$, and more often than not we are interested in the power spectrum of the signal. Thus, an estimate of the noise power spectrum must be constructed and subtracted from the measured power spectrum. In certain cases (e.g. white noise) this may be done analytically. In other cases one must rely on noise simulations. Finally, noise bias may be avoided using only cross-correlations between different map “splits” (e.g. maps constructed from data taken during disjoint time periods), containing the same signal but different noise realisations ([Louis et al. 2025](#)). The specific case of uncorrelated noise from discretely-sampled sources is discussed in Section 2.1.3.

(vi) The power spectrum must be corrected for finite-resolution effects. These may be caused by a finite telescope beam, or by the finite size of the sky pixels over which the map has been measured (assuming that the pixel values correspond to an average of the field value over the pixel area). This can be done exactly if a sufficiently accurate model of the smoothing kernel can be constructed (e.g. [Datta et al. 2024](#)).

It is worth noting that, under the usual interpretation of the QMV estimator, the mask $v(\hat{\mathbf{n}})$ in step 1 above should in principle correspond to the local inverse variance of the measured field³. However, the estimator remains unbiased regardless of what $v(\hat{\mathbf{n}})$ is (as long as it is zero where the field is not measured), although its variance may become highly sub-optimal if $v(\hat{\mathbf{n}})$ deviates significantly from this prescription.

2.1.2 Template deprojection

Often the map under study is affected by contamination from non-cosmological sources. Although the level of contamination is usually not known a priori, it is often possible to construct a comprehensive library of templates representing the potential sources of contamination. If the level of contamination is relatively mild, we can then construct a linear model of the form

$$\mathbf{a}^{\text{obs}} = \mathbf{a}^v(\hat{\mathbf{n}}) + A_p \mathbf{f}^p(\hat{\mathbf{n}}), \quad (2)$$

where $\mathbf{a}^{\text{obs}}$ is the observed, contaminated field, $\mathbf{a}^v$ is the masked and uncontaminated field, $\mathbf{f}^p(\hat{\mathbf{n}})$ are the different contaminant template maps, and A_p are unknown amplitudes that control the level of contamination from each source. Note that we implicitly sum over the repeated index p (and we will continue to do so throughout the paper).

A self-consistent procedure to account for this contamination is then to calculate the power spectrum of $\mathbf{a}^{\text{obs}}$ after marginalising over all possible values of A_p ([Rybicki & Press 1992](#)). As shown in [Elsner et al. \(2017\)](#), in the context of the PCL framework, this can be achieved by calculating the best-fit value of A_p , and subtracting the resulting contamination from the original map. Mathematically, this is also equivalent to projecting the observed map onto the subspace

that is orthogonal to all the contaminant templates, effectively “de-projecting” these templates from the data. The resulting deprojected map is thus

$$\mathbf{a}^c(\hat{\mathbf{n}}) = \mathbf{a}^v(\hat{\mathbf{n}}) - \mathbf{f}^p(\hat{\mathbf{n}}) F_{pq} \langle \mathbf{f}^q, \mathbf{a}^v \rangle, \quad (3)$$

where the matrix F_{pq} is defined as the pseudo-inverse⁴ of

$$\left(\mathbf{F}^{-1} \right)_{pq} \equiv \langle \mathbf{f}^p, \mathbf{f}^q \rangle. \quad (4)$$

In these equations we have defined the map-level dot product as

$$\langle \mathbf{a}, \mathbf{b} \rangle \equiv \int d\hat{\mathbf{n}} \mathbf{a}^\dagger(\hat{\mathbf{n}}) \mathbf{b}(\hat{\mathbf{n}}). \quad (5)$$

Expanding the fields in terms of their harmonic coefficients, this dot product may also be written in harmonic space as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{\ell m} \mathbf{a}_{\ell m}^\dagger \mathbf{b}_{\ell m} = \sum_{\ell} (2\ell + 1) \text{Tr} [\text{PCL}_\ell(\mathbf{a}, \mathbf{b})]. \quad (6)$$

Since this will be important for our discussion in Section 2.2, note that the combination

$$\hat{A}_p \equiv F_{pq} \langle \mathbf{f}^q, \mathbf{a}^v \rangle \quad (7)$$

is the best-fit value of the amplitude coefficient A_p in Eq. 2 estimated from the data. Eq. 3 may thus be expressed directly in harmonic space:

$$\mathbf{a}_{\ell m}^c = \mathbf{a}_{\ell m}^v - \hat{A}_p \mathbf{f}_{\ell m}^p, \quad (8)$$

with $\hat{A}_p$ calculated in real or harmonic space. Finally, note that all the template maps $\mathbf{f}^p$ entering the equations above are implicitly masked using $v(\hat{\mathbf{n}})$ (the mask applied to the data).

Since contaminant deprojection effectively removes modes from the map (any mode that lies in the subspace of contaminant templates), it causes a bias in the estimated power spectrum that must be corrected for ([Elsner et al. 2017](#)). We will explore alternative approaches to correct for this bias in Section 2.3.

As first introduced in Section 2.4.2 of [Alonso et al. \(2019\)](#), in the case of uncorrelated noise, it is possible to calculate the correction to the noise power spectrum caused by contaminant deprojection exactly, even if the noise is inhomogeneous. The additional contribution to the noise PCL is:

$$\Delta N_\ell = -2 F_{pq} \text{PCL}_\ell(\mathbf{f}^p, v^2 \sigma_N^2 \mathbf{f}^q) + F_{pq} F_{rs} \left[\int d\hat{\mathbf{n}} v^2 \sigma_N^2 \mathbf{f}^{q\dagger} \mathbf{f}^r \right] \text{PCL}_\ell(\mathbf{f}^p, \mathbf{f}^s), \quad (9)$$

where $\sigma_N^2(\hat{\mathbf{n}})$ is a map of the local field variance in a unit solid angle.

2.1.3 Catalogue-based pseudo- C_ℓ s

We are often interested in fields that are not mapped over continuous areas, but instead are recorded at the positions of discrete sources. A prime example of this is cosmic shear, where the lensing shear field is effectively measured at the positions of background sources. The power spectrum of such fields can still be calculated using the standard PCL approach, generating continuous maps by binning the sources into pixels, and using the density of these sources as a proxy for the sky mask. Some subtleties about this approach can be found in [Nicola et al. \(2021\)](#).

Instead, a more accurate treatment of discretely-sampled fields

³ We assume scalar masks, although this can be generalised to the case of fully anisotropic weighting – see [Alonso \(2025\)](#).

⁴ The pseudo-inverse is used here to account for potential internal linear dependences in the library of contaminant templates.

has been introduced recently (Baleato Lizancos & White 2024; Wolz et al. 2025), enabled by new algorithms to estimate the spherical harmonic transforms (SHTs) of irregularly-sampled fields (Reinecke et al. 2023). In this case, the field and its mask can be expressed as a sum of delta functions centred at the source coordinates:

$$v(\hat{\mathbf{n}}) = \sum_{i=1}^{N_d} w_i \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i), \quad a^v(\hat{\mathbf{n}}) = \sum_{i=1}^{N_d} w_i a_i \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i), \quad (10)$$

where the index i runs over all sources in the sample, w_i is the weight of the i -th source, a_i is the value of field measured at it, $\hat{\mathbf{n}}_i$ is its position in the sky, and δ^D is the Dirac delta function on the sphere. For simplicity we consider only a scalar field here, and the generalisation to spin- s can be found in Wolz et al. (2025). The spherical harmonic coefficients of v and a^v are then simply

$$v_{\ell m} = \sum_{i=1}^{N_d} w_i Y_{\ell m, i}^*, \quad a_{\ell m}^v = \sum_{i=1}^{N_d} w_i a_i Y_{\ell m, i}^*, \quad (11)$$

which can be calculated using the irregular SHT methods implemented in e.g. Reinecke et al. (2023). Once computed, these harmonic coefficients can be used to calculate the PCL of the field and the associated MCM, as described in Section 2.1.1.

The most immediately obvious advantage of the catalogue-based approach is the absence of finite pixel resolution effects, allowing the resulting estimator to recover power spectra up to arbitrarily small scales (as long as the corresponding harmonic coefficients can be calculated and stored). In addition to this, Wolz et al. (2025) showed that the catalogue-based approach naturally allows for a more numerically stable treatment of uncorrelated noise bias. In particular, it was shown that all contributions from any uncorrelated noise sources can be removed exactly by subtracting the zero-lag contributions to the PCLs of the masked field and of the mask itself prior to estimating the mode-coupling matrix (i.e. the contributions to the PCLs coming from the product of v and a^v with themselves evaluated at the same source). These are given by

$$\tilde{N}^v \equiv \frac{1}{4\pi} \sum_{i=1}^{N_d} w_i^2, \quad \tilde{N}^a \equiv \frac{1}{4\pi} \sum_{i=1}^{N_d} w_i^2 a_i^2. \quad (12)$$

Doing so has the added advantage that, by removing the asymptotic value of the mask PCL at high ℓ (given by $\tilde{N}^v$), the associated MCM is significantly more numerically stable, and can be safely inverted.

Another application of the catalogue-based approach is the case of galaxy clustering, in which the density of sources is itself the field of interest. In the simplest scenario, a continuous map with the expected mean density of sources $\bar{n}(\hat{\mathbf{n}})$ can be constructed. The overdensity field is then given by

$$a(\hat{\mathbf{n}}) \equiv \delta_g(\hat{\mathbf{n}}) = \frac{1}{\bar{n}(\hat{\mathbf{n}})} \left[\sum_{i=1}^{N_d} w_i \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i) - \bar{n}(\hat{\mathbf{n}}) \right]. \quad (13)$$

Assuming the mean density is itself used as a mask for this field (since $\bar{n}(\hat{\mathbf{n}})$ is indeed proportional to the inverse variance of δ_g in the Poisson limit), the masked field is simply

$$a^v(\hat{\mathbf{n}}) = \sum_{i=1}^{N_d} w_i \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i) - \bar{n}(\hat{\mathbf{n}}). \quad (14)$$

Since in this case the mask is continuous, its shot noise component is simply $\tilde{N}^v = 0$, while the field noise is $\tilde{N}^a = \sum_{i=1}^{N_d} w_i^2 / (4\pi)$.

Often, however, the mean source density is not provided as a continuous map, but as a catalogue of “random” objects (or simply “randoms”). Catalogues of randoms have long been a critical component of configuration-space galaxy clustering analyses (e.g. Totsuji

& Kihara 1969; Landy & Szalay 1993). These are point processes distributed according to the survey selection function and without any intrinsic clustering. They are typically generated by uniformly sampling across the survey footprint to create a catalogue of positions with a density significantly higher than that of the data, such that correlations between the randoms and observed galaxies can be used to mitigate systematics caused by masking or edge effects. It is worth noting that recent efforts have also been made to construct catalogues of ‘organised randoms’ which additionally account for spatial correlations induced by more complex survey properties such as inhomogeneous depth or seeing (Johnston et al. 2021; Yan et al. 2025). However, since the aim of this work is to employ template deprojection to mitigate these more complex effects, we adopt the former approach to generating randoms when applying our formalism to simulations (Section 3.2) and observations (Section 3.5).

If randoms are used in place of a continuous map, the mean density map (and hence the mask) entering the equations above is

$$\bar{n}(\hat{\mathbf{n}}) = \alpha \sum_{i=1}^{N_r} w_i^r \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i), \quad (15)$$

where w_i^r is the weight of the i -th random object, and $\alpha \equiv \sum_{i=1}^{N_d} w_i / \sum_{i=1}^{N_r} w_i^r$ corrects for the different number of sources in the data and random catalogues. To simplify the notation, in what follows, we will omit the factor α , and absorb it into the random weights⁵. In this case the mask has a shot noise component, and the randoms also contribute to the field noise:

$$\tilde{N}^w = \frac{1}{4\pi} \sum_{i=1}^{N_r} (w_i^r)^2, \quad \tilde{N}^a = \frac{1}{4\pi} \sum_{i=1}^{N_d} w_i^2 + \tilde{N}^w. \quad (16)$$

In summary, the treatment of catalogue-based fields is straightforward to extend to the case of galaxy clustering, by simply replacing all field values with 1, and including the contribution from random objects in the calculation of the field spherical harmonic coefficients and in the shot-noise components of field and mask (if randoms are used).

2.2 Template deprojection for catalogue-based C_ℓ s

We now present the extension of the template deprojection technique, outlined in Section 2.1.2 for continuous fields, to the case of fields sampled at the positions of discrete sources. Specifically, our aim is to extend Eqs. 8 and 9, describing the harmonic coefficients of the contaminant-deprojected field and the modification to the white noise level brought about by deprojection, respectively. We will do so for both sampled fields and for galaxy clustering.

2.2.1 Sampled fields

Consider first a discretely-sampled field $a^v(\hat{\mathbf{n}})$ as given in Eq. 10. The masked contaminant templates in this case are

$$\tilde{f}^p(\hat{\mathbf{n}}) = \sum_i w_i f_i^p \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i), \quad (17)$$

where f_i^p is the value of the template at the position of the i -th source in the catalogue. Substituting this in Eq. 3, we find that the deprojected field is simply

⁵ In other words, we will implicitly assume that the random weights are normalised such that $\sum_{i=1}^{N_r} w_i^r = \sum_{i=1}^{N_d} w_i$.

$$a^c(\hat{\mathbf{n}}) = \sum_i w_i a_i^c \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i), \quad (18)$$

where we have defined the deprojected sampled field values a_i^c are

$$a_i^c \equiv a_i - \hat{A}_p f_i^p, \quad (19)$$

and the estimated linear amplitude is

$$\hat{A}_p \equiv \tilde{F}_{pq} \sum_j w_j^2 f_j^q a_j, \quad (\tilde{F}^{-1})_{pq} \equiv \sum_j w_j^2 f_j^p f_j^q. \quad (20)$$

Thus, we see that deprojection simply affects the values of the field sampled at the source positions, and it only depends on the values of the contaminant templates sampled at the same positions. The dot product introduced in Eq. 5 is replaced by a sum over all sources of the product of two weighted fields. Deprojection can therefore be achieved easily, without requiring any contaminant maps explicitly, and simply modifying the sampled field values as above. Likewise, the white noise amplitude, accounting for deprojection, can be calculated simply using Eq. 12, replacing a_i with a_i^c as given above.

The white-noise power spectrum receives a further scale-dependent correction from deprojection due to mode loss. In the case of continuous fields, this is given by Eq. 9. To calculate this correction in the case of a discretely-sampled field, we start by writing the PCL of a^c :

$$\text{PCL}_\ell(a^c, a^c) = \sum_{ij} (w_i w_j a_i^c a_j^c) \sum_{m=-\ell}^{\ell} \frac{|Y_{\ell m, i}^* Y_{\ell m, j}|}{2\ell + 1}. \quad (21)$$

We then expand a_i^c in terms of a_i and the templates as in Eq. 19, and then take the ensemble average of the result, assuming a_i to consist solely of uncorrelated noise: $\langle a_i a_j \rangle = \sigma_i^2 \delta_{ij}$, where σ_i^2 is the noise variance of the field at the i -th discrete source. Writing the result as

$$\langle \text{PCL}_\ell(a^c, a^c) \rangle = \tilde{N}^a + \Delta N_\ell^a, \quad (22)$$

the correction is

$$\begin{aligned} \Delta N_\ell^a = & -2\tilde{F}_{pq} \text{PCL}_\ell(\tilde{f}^p, \tilde{g}^q) \\ & + \tilde{F}_{pq} \tilde{F}_{rs} \left[\sum_j w_j^2 \sigma_j^2 (w_j f_j^q) (w_j f_j^s) \right] \text{PCL}_\ell(\tilde{f}^p, \tilde{f}^r), \end{aligned} \quad (23)$$

where

$$\tilde{f}_{\ell m}^p \equiv \sum_i w_i f_i^p Y_{\ell m, i}^*, \quad (24)$$

and where we have defined the noise-weighted field

$$\tilde{g}_{\ell m}^p \equiv \sum_i w_i^2 \sigma_i^2 (w_i f_i^p) Y_{\ell m, i}^*. \quad (25)$$

As before, this contribution to the noise bias due to deprojection depends only on the values of the contaminant templates evaluated at the source positions, as well as an estimate of the noise variance at each source σ_i^2 . For noise-dominated fields, this may be estimated directly as $\sigma_i^2 \sim a_i^2$, but care must be exercised when this is not the case. For modest numbers of contaminant templates, this white noise correction is relatively small, and may be neglected. However, as we will see in Section 3.1, when many ($O(100)$) templates are deprojected, the correction may become significant, particularly on large angular scales.

2.2.2 Galaxy clustering

We turn now to the case of galaxy clustering. We will start with the more complex scenario in which the mask is provided as a discrete set of random points, and then turn to the simpler case of a continuous mask at the end of this section.

Two characteristics of this type of data separate it from the standard sampled fields in a way that impacts how contaminant deprojection must be carried out: first, the source positions and weights do not constitute the field's mask, the random catalogue does instead. Secondly, this discretely-sampled mask features as an additive contribution to the masked field itself, quantifying the mean density subtracted from the data (the second term in Eq. 14). Note that although this latter choice is indeed the most common one when analysing galaxy data, one could also use two different random catalogues – one to construct the subtracted mean density, and a different one to define the mask – and the resulting estimator for the galaxy overdensity would still be unbiased⁶. Nevertheless, we will adopt the standard approach of using a single random catalogue for both purposes here.

This immediately leads to a problem if we simply apply Eq. 3 to the field defined in Eq. 14, with $\bar{n}$ given by Eq. 15 and, hence, with the masked templates given by⁷

$$\tilde{f}^p(\hat{\mathbf{n}}) = \sum_{i=1}^{N_r} w_i^r f_i^p \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i), \quad \tilde{f}_{\ell m}^p = \sum_{i=1}^{N_r} w_i^r f_i^p Y_{\ell m, i}^*.$$

Substituting this in the expression for the estimated linear coefficients $\hat{A}_p$ (Eq. 7) we find:

$$\begin{aligned} \hat{A}_p &= F_{pq} \int d\hat{\mathbf{n}} a^v(\hat{\mathbf{n}}) \tilde{f}^q(\hat{\mathbf{n}}) \\ &= F_{pq} \left(\sum_{i=1}^{N_d} \sum_{j=1}^{N_r} w_i w_j^r f_j^q \int d\hat{\mathbf{n}} \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i) \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_j) \right. \\ &\quad \left. - \sum_{i,j=1}^{N_r} w_i^r w_j^r f_j^q \int d\hat{\mathbf{n}} \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_i) \delta^D(\hat{\mathbf{n}}, \hat{\mathbf{n}}_j) \right) \\ &\propto -F_{pq} \sum_{i=1}^{N_r} (w_i^r)^2 f_i^p, \end{aligned}$$

since we assume that no two objects from either the source or random catalogues occupy the exact same position. Likewise, the elements of the F_{pq} matrix are

$$(F)_{pq}^{-1} \propto \sum_{i=1}^{N_r} (w_i^r)^2 f_i^p f_i^q. \quad (26)$$

This is problematic because, first, the estimated $\hat{A}_p$ does not depend on the data catalogue, only on the randoms, and hence it does not capture any real correlations between the galaxy overdensity and the contaminants. Secondly, had we chosen to use two different catalogues for the subtracted mean density and the mask, the quantities above would both be exactly zero.

The cause of this problem lies in the key assumption underlying the pseudo- C_ℓ power spectrum estimator: the covariance matrices of all fields involved are diagonal in real space, and the inverse diagonals are represented by the field masks (in this case, the random catalogue positions). We know that, in practice, this is not the case:

⁶ Assuming that both random catalogues represent the same underlying mean density field.

⁷ Recall that we assume the random weights to be normalised so that $\alpha = 1$.

LSS fields exhibit correlations over a range of real-space separations. Furthermore, we actually expect the observed galaxy overdensity to be affected by systematic contamination (and hence to correlate with the contaminant templates) on large scales. We thus propose to use these large-scale correlations to redefine the linear deprojection coefficients. In particular, we can use the harmonic-space expression for the dot product of two fields (Eq. 6), truncating it at a given maximum multipole $\ell_{\max}$, which characterises the scale over which we want to capture correlations between data and contaminants. The linear deprojection coefficients, in the case of galaxy clustering, are therefore calculated via

$$\hat{A}_p \equiv \tilde{F}_{pq} \sum_{\ell=0}^{\ell_{\max}} (2\ell+1) \left[\text{PCL}_{\ell}(a^v, \tilde{f}^q) + \sum_{i=1}^{N_r} \frac{(w_i^r)^2 f_i^q}{4\pi} \right], \quad (27)$$

$$(\tilde{F})_{pq}^{-1} = \sum_{\ell=0}^{\ell_{\max}} (2\ell+1) \left[\text{PCL}_{\ell}(\tilde{f}^p, \tilde{f}^q) - \sum_{i=1}^{N_r} \frac{(w_i^r)^2 f_i^p f_i^q}{4\pi} \right].$$

Note that we have subtracted the shot-noise contribution to the pseudo- C_{ℓ} s entering these equations, caused by coincident pairs of random points in the subtracted mean density and the mask. This contribution is completely accidental (e.g. it would be zero if we had decided to use two different random catalogues for both purposes), and must therefore be discarded.

This generalisation of the standard contaminant deprojection approach is perfectly legitimate from a mathematical point of view: it is equivalent to a standard deprojection operation in harmonic space in which some data points (multipoles with $\ell, |m| > \ell_{\max}$) are assigned zero weight (or, equivalently, a very large variance).

Having defined the deprojection procedure, we must now derive the correction to the shot noise power spectrum caused by deprojection. We can do so following the same steps outlined in the previous section for sampled fields, with two caveats:

- The noise variance in this case is sourced by Poisson noise in the data and random catalogues. Furthermore, if different random catalogues are used to quantify the subtracted mean density and to define the mask, only the randoms in the former (i.e. those entering $a_{\ell m}^v$) contribute to this variance, and not those in the latter (i.e. those entering $\tilde{f}_{\ell m}$).
- We must take care to consistently apply the truncation scale $\ell_{\max}$ used to find the deprojection coefficients.

The resulting expression for the shot noise deprojection bias is

$$\Delta N_{\ell} = -2\tilde{F}_{pq} \text{PCL}_{\ell}(\tilde{f}^p, \tilde{g}^q) + \tilde{F}_{pq} \tilde{F}_{rs} \left[\sum_{i=1}^{N_d} w_i^2 \tilde{f}_{Fi}^q \tilde{f}_{Fi}^s + \sum_{i=1}^{N_r} (w_i^r)^2 \tilde{f}_{Fi}^q \tilde{f}_{Fi}^s \right] \text{PCL}_{\ell}(\tilde{f}^p, \tilde{f}^q), \quad (28)$$

where we have defined the filtered-weighted templates $\tilde{f}_{Fi}^p$ as a truncated inverse spherical harmonic transform of the masked templates

$$\tilde{f}_{Fi}^p \equiv \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \tilde{f}_{\ell m}^p Y_{\ell m, i}, \quad (29)$$

and where

$$\tilde{g}_{\ell m}^p \equiv \sum_{i=1}^{N_d} w_i^2 \tilde{f}_{Fi}^p Y_{\ell m, i}^* + \sum_{i=1}^{N_r} (w_i^r)^2 \tilde{f}_{Fi}^p Y_{\ell m, i}^*. \quad (30)$$

Note that, as in the case of sampled fields, we only require the

values of the contaminant templates at the discrete positions of the points defining the field's mask (i.e. the random catalogue), and not explicitly in the form of maps.

Now, in the simpler case in which the mask (and hence the mean number density) is provided as a continuous map, deprojection can be carried out following the steps outlined above, with the following simplifications:

- The templates are also provided as continuous maps, and the masked templates are simply $\tilde{f}^p(\hat{\mathbf{n}}) = \tilde{n}(\hat{\mathbf{n}}) f^p(\hat{\mathbf{n}})$, where $\tilde{n}(\hat{\mathbf{n}})$ is the mean number density map, which also serves as our mask.
- No shot-noise contributions must be subtracted when estimating the deprojection coefficients (i.e. the rightmost terms in both expressions in Eq. 27).
- Only the data weights contribute to the shot noise variance in Eqs. 28 and 30.

2.3 Correcting for mode loss

The process of template deprojection indiscriminately removes all power from modes proportional to any of the deprojected templates, introducing a bias in the final power spectrum estimate. As discussed in [Elsner et al. \(2017\)](#), this *deprojection bias* can be estimated analytically and subsequently removed: for a scalar field a with corresponding mask v and contaminants f^p , it is given by

$$\Delta C_{\ell} = -2F_{pq} \text{PCL}_{\ell}(\tilde{f}_{\ell m}^q, f_{\ell m}^p) + F_{pq} F_{rs} \left[\int d\hat{\mathbf{n}} f^{q\dagger}(\hat{\mathbf{n}}) \tilde{f}^r(\hat{\mathbf{n}}) \right] \text{PCL}_{\ell}(f^p, f^s), \quad (31)$$

where F_{pq} is as defined via Eq. 4, and

$$\tilde{f}_{\ell m}^p \equiv \mathcal{S}[v(\hat{\mathbf{n}}) \mathcal{S}^{-1}[C_{\ell'}^{\text{true}} \mathcal{S}[v f^p]_{\ell' m'}^{s_b}]_{\hat{\mathbf{n}}}^{s_a}]_{\ell m}. \quad (32)$$

with $\mathcal{S}$ denoting a spherical harmonic transform and $\mathcal{S}^{-1}$ its inverse⁸. Note that we are again only considering autocorrelations of a single scalar field here; a generalisation to cross-correlations between two fields with arbitrary spin can be found in [Alonso et al. \(2019\)](#).

This method of correcting for the loss of modes induced by deprojection has two drawbacks. Firstly, it requires some knowledge of the true angular power spectrum C_{ℓ}^{true} , for which a reliable estimate is not always achievable. Secondly, while the above prescription works for map-based fields, an equivalent procedure for catalogue-based data is non-trivial to implement, owing to the inverse spherical harmonic transforms involved.

We therefore explore in this work an alternative approach, in which we use simulations to estimate a *transfer function* by which one can divide the measured power spectra to correct for mode loss following template deprojection. The process involves randomly generating many Gaussian realisations of some angular power spectrum C_{ℓ}^{input} , and applying to each realisation the same mask as is being used for the data in question. For each realisation, the angular power spectrum is then estimated both before and after deprojecting the set of contaminants believed to be influencing the data; we will denote these power spectra as C_{ℓ}^{before} and C_{ℓ}^{after} , respectively. The transfer function T_{ℓ} is then defined as

$$T_{\ell} = \frac{\langle C_{\ell}^{\text{after}} \rangle_{\text{sim}}}{\langle C_{\ell}^{\text{before}} \rangle_{\text{sim}}}, \quad (33)$$

⁸ Defined as $\mathcal{S}[a(\hat{\mathbf{n}})]_{\ell m} \equiv \int d\hat{\mathbf{n}} Y_{\ell m}^*(\hat{\mathbf{n}}) a(\hat{\mathbf{n}})$ and $\mathcal{S}^{-1}[a_{\ell m}]_{\hat{\mathbf{n}}} \equiv \sum_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}) a_{\ell m}$ for a scalar field $a(\hat{\mathbf{n}})$.

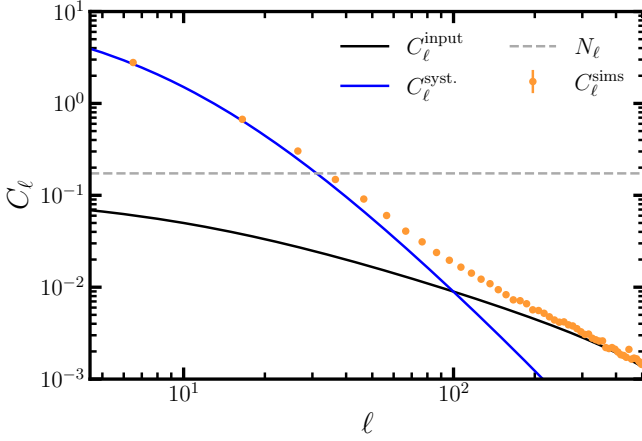


Figure 1. Input power spectrum used to validate the approach to contaminant deprojection in the case of sampled fields (black line). The blue line shows the power spectrum of the contaminant maps used, while the horizontal dashed line shows the simulated white noise component. The orange points show the mean power spectrum measured from all contaminated simulations.

where $\langle \cdot \rangle_{\text{sim}}$ denotes the average across Gaussian realisations. Dividing the power spectrum measured from the data by T_ℓ should then correct it for mode loss (see e.g. [Elsner et al. 2016](#), in which this approach is used to validate analytical expressions for the deprojection bias).

In Section 3.3 we test this approach and verify it as a viable alternative to calculating the deprojection bias, which can in turn be applied to power spectra measured from catalogue-based data.

3 RESULTS

We now validate the catalogue-based adaptation of template deprojection described above using both simulated and real data. Sections 3.1 and 3.2 do this for the cases of sampled fields and galaxy clustering, respectively. Next, in Section 3.3 we examine the relative performances of the mode loss correction methods described in Section 2.3. Finally in Section 3.5 we apply this approach to real data from the *Gaia-unWISE* quasar catalogue (*Quaia*; [Storey-Fisher et al. 2024](#)).

3.1 Contaminant deprojection in sampled fields

We start by validating the formalism described in the previous section in the case of sampled fields. To do so, we generate a number of simulated catalogue-based observations including the presence of sky contamination, and show that we are able to recover unbiased power spectra by applying deprojection techniques and accounting for the impact of deprojection bias in both the signal and uncorrelated noise components.

Specifically, in our simulations we use the positions of sources in the *Quaia* catalogue with spectro-photometric redshifts in the range $z_p < 1.47$, corresponding to a sample of 644786 quasar candidates. This ensures that the simulated data have a realistic source distribution, including large-scale sky coverage and small-scale clustering. Using a HEALPix⁹ ([Górski et al. 2005](#); [Zonca et al. 2019](#)) pixelisation

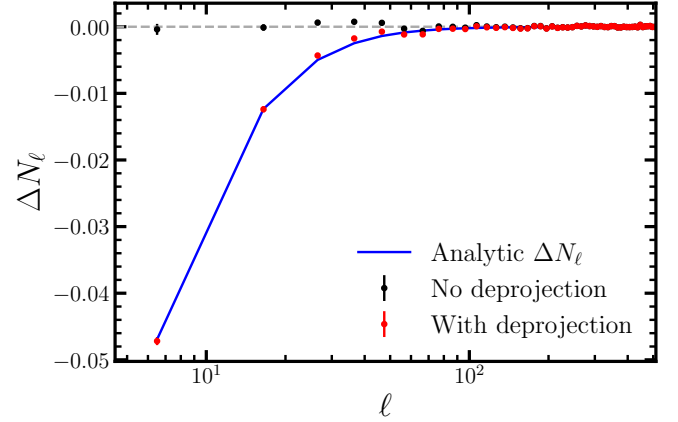


Figure 2. Power spectrum of noise-only simulations with and without contaminant deprojection (black and red points, respectively). The loss of power due to deprojection leads to a negative power spectrum (note that the catalogue-based estimator automatically removes the white noise component), which is accurately predicted by Eq. 23 (blue line).

scheme with a resolution parameter of $N_{\text{side}} = 256$ (corresponding to a pixel area of $\sim 0.05 \text{ deg}^2$), we then generate signal maps as Gaussian realisations with an angular power spectrum of the form

$$C_\ell^{\text{input}} = \frac{1}{\ell + 10}, \quad (34)$$

and sample them at the positions of the *Quaia* sources. We generate 1000 such realisations. Likewise, we generate 100 Gaussian maps from a steeper power spectrum of the form

$$C_\ell^{\text{syst}} = \frac{A_{\text{syst}}}{(\ell + 10)^3}, \quad (35)$$

which we sample at the same source positions and use as contaminant templates to deproject. We then construct a specific linear combination of these contaminant templates to contaminate the signal assigned to each source. For simplicity, we use the direct sum of all templates, scaling the proportionality constant A_{syst} to ensure that the total variance due to systematic contamination in each source is 30% of the signal. Finally, we generate uncorrelated Gaussian numbers for each source to serve as a noise component, with a standard deviation equal to 10 times that of the signal. Figure 1 shows the input and contaminant power spectra, as well as the expected noise level, together with the mean power spectrum of all simulations. We note that the amplitudes of the contaminant and noise components have been chosen to exacerbate the impact of systematic deprojection on the simulations: the measurements are systematics-dominated on scales $\ell \simeq 100$. Furthermore, the data are noise dominated on a source-by-source basis, as well as on all scales. This, together with the large number of contaminant templates being deprojected, increases the amplitude of the correction to the noise bias due to deprojection (ΔN_ℓ^a in Eq. 23).

To first validate our analytical expression used to calculate ΔN_ℓ^a , we study the power spectrum of noise-only simulations, in which we only include the contribution from uncorrelated noise to the field value for each source. Fig. 2 shows the mean power spectrum of all such simulations before and after deprojecting all 100 contaminant templates (black and red points, respectively). We remind the reader that the white noise component is avoided by the catalogue-based estimator by construction. The loss of modes due to deprojection

⁹ <http://healpix.sf.net/>

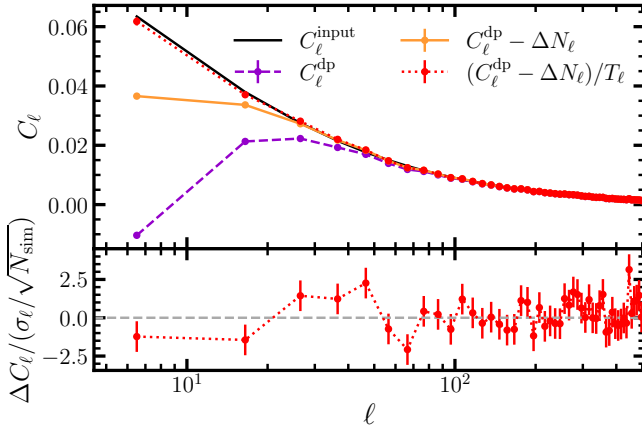


Figure 3. Mean and standard deviation of the power spectra from contaminated simulations after contaminant deprojection (purple points). The orange points show the result of accounting for the impact of deprojection in the white noise component (via Eq. 23), while the red points show the power spectra after accounting for deprojection bias in both the noise and signal components (the latter via the transfer function approach). The result is an unbiased estimate of the input power spectrum (black line). The bottom panel shows the residuals with respect to the input spectrum normalised by the error in the mean over all simulations.

leads to a negative power spectrum at low ℓ , which is accurately recovered by our analytical expression (blue line).

Finally, Fig. 3 shows the result of our validation of the formalism outlined here. The purple points show the average power spectrum for contaminated simulations after contaminant deprojection, before accounting for any sources of deprojection bias. On large scales, the power spectrum shows a very significant lack of power, even becoming negative at the lowest ℓ s. Much of this power is recovered by correcting for the white noise deprojection bias calculated via Eq. 23, as shown by the orange points. The remaining power loss can be recovered by correcting for the deprojection bias of the signal component. As described in Section 2.3, this may be done analytically if the true signal power spectrum is known, or through a simulation-based transfer function approach. The red points show the result of the latter method¹⁰, which is able to recover an unbiased power spectrum at all scales.

3.2 Contaminant deprojection in galaxy clustering

Here we validate the galaxy clustering formalism described in Section 2.2.2 by again using simulated catalogue-based observations, with contributions from sky contaminants imposed on the distribution of the sources. We assess the two possible treatments of the mask: using the positions and weights of randoms to define it, and providing it as a continuous map. We will hereafter refer to these as the randoms-based and mask-based approaches, respectively.

For each simulation, we begin by generating a Gaussian realisation of a map, $\delta_{g,i}^{\text{in}}$, from some input angular power spectrum, C_ℓ^{input} . For this test, we use the best-fit angular power spectrum for the *Quia*

‘low- z ’ ($0 \leq z < 1.47$) sample, described in Alonso et al. (2025); Fabbian et al. (2025), as the input for our simulations. A second map, $\delta_{g,i}^{\text{cont}}$, is also generated from C_ℓ^{input} and contaminated with templates f^p using known linear coefficients A_p such that the value in pixel i is given by

$$\delta_{g,i}^{\text{cont}} = \delta_{g,i}^{\text{in}} + \sum_{p=1}^{N_{\text{syst}}} A_p f_i^p. \quad (36)$$

We use the same 10 templates as used in Alonso et al. (2025); Fabbian et al. (2025), which the authors deproject from *Quia* overdensity maps using the standard formalism as outlined in Section 2.2. The use of these templates here is motivated by the desire to test our catalogue-based formalism in a realistic scenario for galaxy clustering. Details of these templates can be found in Appendix A, including Mollweide projections of each (Fig. A1).

We then use a mask defining the *Quia* survey geometry to convert each of these maps into a probability map, P , whose value in a pixel i is given by:

$$P_i^X = (1 + \delta_{g,i}^X) w^i, \quad (37)$$

where w is a map describing the survey mask, and $X \in \{\text{in}, \text{cont}\}$. As in Section 3.1, all maps are created using a HEALPix pixelisation scheme with a resolution parameter of $N_{\text{side}} = 256$. These maps are then normalised such that $\max(P^X) = 1$, and subsequently Poisson-sampled to create realisations of the catalogue with and without contamination from systematics. We note that it is possible, when generating these probability maps, for a pixel to end up with a value less than zero, and cannot therefore be Poisson-sampled. To test whether this is likely to be an issue, we use the prescription outlined above to generate two sets of 10^5 probability maps, one with contamination and one without, and count how many pixels in each have values below zero. We find that there are no such pixels in any of the maps, and thus conclude that they are unlikely to contribute significantly to our analysis.

As highlighted in Section 2.2.2, it is necessary in the case of galaxy clustering systematics to truncate at some maximum multipole ℓ_{max} the harmonic-space expression for the dot product of two fields in order to avoid losing information about real correlations between the galaxy overdensity and the contaminants (see Eq. 27). The choice of ℓ_{max} depends on the contaminants in question. To determine a suitable value for the templates used here, we rerun the suite of simulations using several different values of ℓ_{max} ranging from $\ell_{\text{max}} = 10$ to $\ell_{\text{max}} = 3N_{\text{side}} - 1 = 767$ ¹¹.

The results of this test are shown in Fig. 4, wherein each panel shows the means and standard deviations in the deprojection coefficient $\hat{A}_p$ for a given contaminant as measured using the randoms- and mask-based approaches (teal circles and red squares, respectively), along with the input value used when generating the simulated data. Both the randoms- and mask-based deprojection methods produce biased estimates of the deprojection coefficients for several of the contaminants evaluated here if one uses an ℓ_{max} that is either too low or too high. We find that the most stable and least biased estimates tend to occur at $\ell_{\text{max}} \sim 100$, and as such adopt this value going forward with our simulations.

The reason for this bias in the estimated $\hat{A}_p$ is related to the pixel window function, and is a by-product of the fact that our contaminant templates take the form of maps. While in the map-based formalism

¹⁰ It is worth noting that the true signal power spectrum is used here to generate the simulations required for deriving the transfer function (see Section 2.3). However, as we will show in Section 3.3, this approach is relatively insensitive to this choice of power spectrum.

¹¹ HEALPix maps generated with a resolution parameter of N_{side} are approximately band-limited to multipoles with $\ell \lesssim 3N_{\text{side}} - 1$.

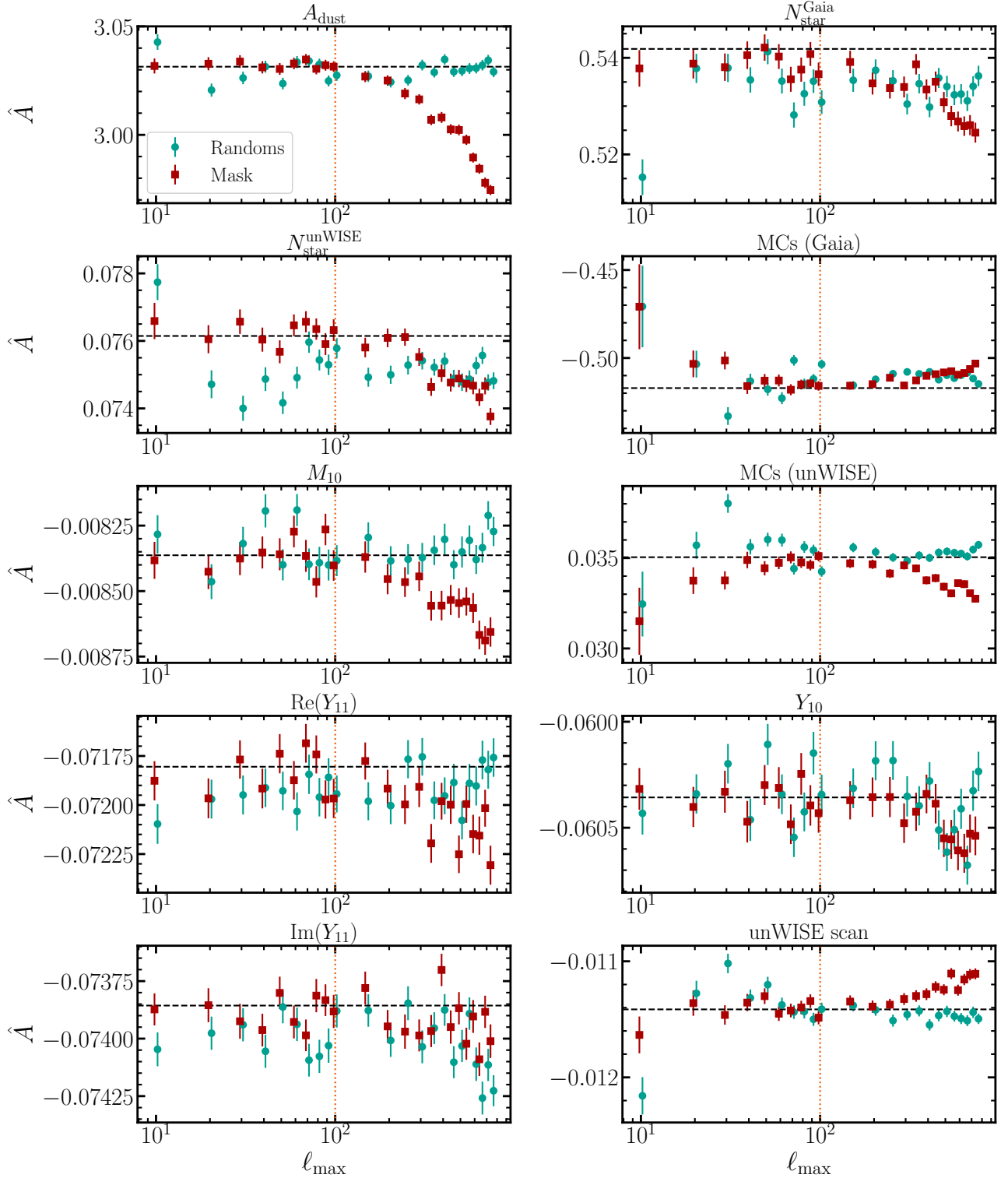


Figure 4. Linear deprojection coefficients measured as a function of $\ell_{\max}$ from simulations using the randoms-based (teal circles) and mask-based (red squares) approaches. The procedure is outlined in Section 3.2, and details of the templates (indicated at the top of each panel) can be found in Appendix A. The dashed black line in each panel shows the input deprojection coefficient used to contaminate the simulated data. Small horizontal offsets have been applied to the data for clarity. Both methods show significant variation in the deprojection coefficients for several contaminants, particularly as one moves to the lowest and highest multipoles. The dotted orange line indicates $\ell_{\max} = 100$, which we deem a suitable value for minimising the bias in our measured deprojection coefficients for the majority of templates.

it is straightforward to correct for the pixel window function, we can see from Eq. 27 that such a correction is not applicable here, owing to calculation of $\text{PCL}_\ell(a^v, \tilde{f}^q)$ required to obtain $\hat{A}_p$. We reiterate that a^v is the (masked) catalogue-based field being observed, and the $\tilde{f}^q$ is a given template (also masked), the nature of which depends on which approach is being used. In the mask-based approach, the template maps are used directly, resulting in a PCL cross-correlating a catalogue-based field with a map-based field, for which a correction for the pixel window function is not viable. One might think this is mitigated by using the randoms-based approach, in which $\tilde{f}^q$ is computed by sampling the template at the positions of the randoms, but since these values are still being sampled from a map of finite resolution, the pixel window function will still have an effect that is not easy to remove. While this effect is inherent to our simulations, it is also likely in a realistic scenario that one will be working with pixelised contaminant templates, in which case this will still be an issue. Since systematics tend to affect clustering measurements at larger scales, a relatively conservative ℓ_{max} of ~ 100 should be sufficient to capture the contribution from these systematics in most use cases, while significantly mitigating the effects of the pixel window function.

Figure 5 shows the result of applying the template deprojection procedure outlined in Section 2.2.2 to the set of contaminated catalogues from our simulations, using both the randoms-based (teal circles; top panel) and mask-based (red squares; middle panel) approaches. As was seen for sampled fields in Section 3.1, there is a significant loss of power at large scales, which is almost entirely mitigated by accounting for the noise deprojection bias via Eq. 28, since these data are noise-dominated in this regime. Consequently, attempts to apply the transfer function approach outlined in Section 2.3 appear to exacerbate the deviations from the true power spectrum, although these shifts are marginal and the resulting power spectra are still consistent within the errorbars. Overall, the final C_ℓ estimates are consistent with the input power spectrum to within $\sim 3\sigma$ at all scales. Since our simulated catalogues have been sampled from maps of finite resolution, aliasing effects are expected to arise beyond $\ell \approx 2N_{\text{side}}$; this can be seen for the mask-based case in the shaded grey region of the bottom panel in Fig. 5, while the randoms-based approach appears relatively unaffected. We therefore caution users against using the mask-based approach to probe scales of $\ell > 2N_{\text{side}}$. Importantly, we also note that the uncertainties in Fig. 5 correspond to the mean over 1000 simulated datasets, and are therefore $\sqrt{1000} \sim 30$ times smaller than the actual errors for a real-world dataset. The small deviations shown in the figure should therefore be completely negligible in a realistic scenario.

3.3 Comparison between debiasing methods

We now use simulations to validate the transfer function approach described in Section 2.3 (see Eq. 33) as a means of correcting for mode loss when deprojecting templates, and compare with the deprojection bias approach normally employed for this purpose (see Eq. 31; Rybicki & Press 1992; Elsner et al. 2017). In particular, we wish to test how dependent the transfer function approach is on assumptions about the true power spectrum.

As in Section 3.2, we begin with 1000 Gaussian realisations of the angular power spectrum obtained from the *Quia* survey, where we have contaminated each realisation with those same templates via Eq. 36 and applied the *Quia* mask w . We will refer to this as the ‘main’ suite of simulations henceforth. Using the standard map-based PCL approach (Sections 2.1.1 and 2.1.2), we then deproject the templates to obtain biased estimates of the input power spectrum, $\hat{C}_\ell$.

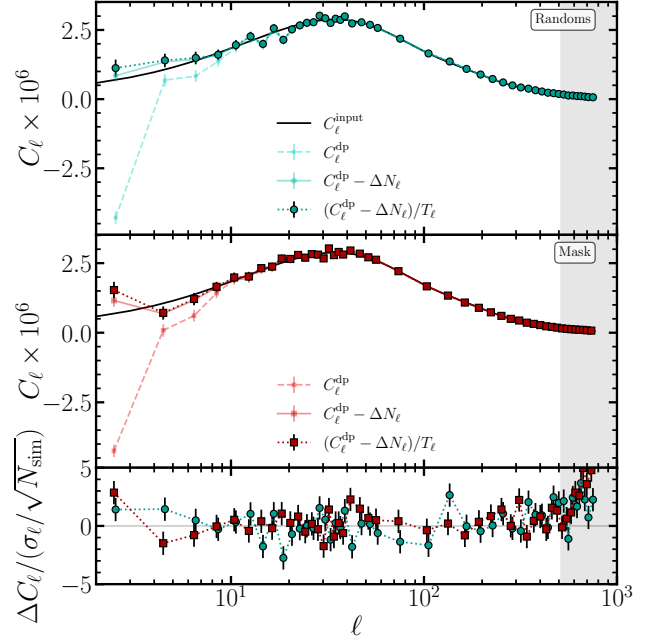


Figure 5. Means and standard deviations in angular power spectra from contaminated simulations measured at various stages of correction using the randoms-based (top panel) and mask-based (middle panel) approaches. Different marker sizes and line styles represent the different stages: after contaminant deprojection (small markers, dashed lines); after removing the noise deprojection bias (medium-sized markers, solid lines); after also correcting for deprojection bias in the signal component using the transfer function approach (large markers, dotted lines). The black solid line shows the angular power spectrum used to generate the (non-contaminated) data, and as such is indicative of the ‘true’ power spectrum. The shaded region begins at $\ell = 2N_{\text{side}}$, beyond which we expect aliasing effects to arise from the fact that we have sampled the catalogue from a map with finite resolution. The bottom panel shows the residuals of the fully corrected power spectra with respect to the input power spectrum, normalised by the error in the mean across the simulations. Both the randoms-based (teal circles) and mask-based (red squares) approaches are able to accurately reproduce the input power spectrum, with the randoms-based approach maintaining better accuracy than the mask-based at the higher multipoles. Exacerbated deviations from the input power spectrum after applying the transfer function (T_ℓ) at the largest scales are due to the data being noise-dominated in this regime (see text).

The transfer function or the deprojection bias estimates are then used to debias these angular power spectra. We use the debiased power spectra as a means of assessing the efficacy of the transfer function approach.

To calculate the various transfer functions we wish to test, we generate several more suites of simulated maps, each producing 1000 Gaussian realisations of a different input C_ℓ . For each realisation, we apply the same mask as for the main suite and use the standard PCL approach to measure the angular power spectrum of the masked field, C_ℓ^{before} . We then use the map-based approach to deproject the same templates used to contaminate the main suite and measure the angular power spectrum post-deprojection, C_ℓ^{after} . The transfer function is then computed using Eq. 33, averaging the relevant quantities across all simulations in the suite.

We use the following angular power spectra as inputs for each simulation suite:

- The *Quia* C_ℓ used to generate the the main suite, whose measured angular power spectra we aim to debias. This is the true power

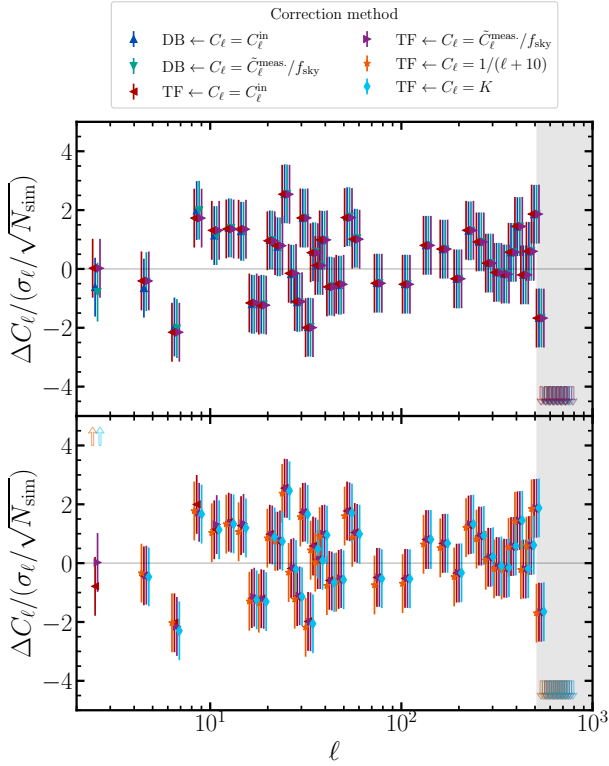


Figure 6. Residuals of the angular power spectra measured from simulations with respect to the power spectrum used to generate the simulated data, having mitigated the mode loss due to deprojection using different approaches. Namely, either the standard deprojection bias (DB) approach or the transfer function (TF) approach is used, with the chosen guess at the input power spectrum indicated in the legend. Small horizontal shifts have been applied to the data to aid visualisation, and the results have been partially split between two panels for ease of comparison between the most relevant methods. Arrows of a given colour denote scales at which the angular power spectrum deviates from the truth by more than 5σ after debiasing with the corresponding method. *Top panel:* The DB approach using the true angular power spectrum (blue upward-pointing triangles) and a data-driven guess at its form (teal downward-pointing triangles), along with the TF approach using the same angular power spectra (red left-pointing and purple right-pointing triangles, respectively). The DB and TF approaches produce similarly unbiased C_ℓ estimates in both cases. *Bottom panel:* The TF approach using the four different angular power spectra described in the text and indicated in the legend. All four perform similarly well overall, although the choice of guess C_ℓ matters more at low ℓ .

spectrum and therefore should, by definition, produce unbiased results after applying the transfer function.

- A guess at the true power spectrum, derived by measuring the PCL of one simulated *Quia*-like dataset and dividing by the covered sky fraction, estimated as $f_{\text{sky}} = \langle w^2 \rangle$.
- A power spectrum with an arbitrarily chosen functional form, $C_\ell = 1/(\ell + 10)$.
- A flat power spectrum with arbitrary amplitude, $C_\ell = K$.

In addition, we use the standard deprojection bias approach to debias the measured power spectra, using one of the first two C_ℓ s described above as the estimate of the true power spectrum (C_ℓ^{true} in Eq. 32).

Figure 6 compares the deviation (normalised by the standard error of the mean) of the measured angular power spectra from the input power spectrum, after correcting for mode loss due to deprojection using one of the methods outlined above. We find that all methods

perform similarly well at most scales, being consistent with each other within the errorbars. Ignoring scales at which $\ell \gtrsim 2N_{\text{side}}$ (demarcated by the grey shaded region in the figure) where aliasing effects are expected to arise from the pixelisation of our maps, it is only at the largest scales ($\ell \lesssim 3$) that we see any difference between the various methods. Namely, the transfer function approach fails to produce unbiased estimates at these scales if an unrealistic guess at the true angular power spectrum is used (see orange stars and cyan diamonds in Fig. 6). We thus conclude that the transfer function approach is a viable alternative to calculating the deprojection bias, so long as one uses a reasonable guess at the true power spectrum when generating the required simulations. For a realistic scenario in which the truth is not known, one can derive a sufficient guess by dividing the measured PCL by the observed sky fraction f_{sky} , as has been done here (teal downward-pointing and purple right-pointing triangles in Fig. 6, for the deprojection bias and transfer function approaches, respectively).

3.4 Catalogue-based transfer functions

While we find in Section 3.3 that the approach of using map-based simulations to debias catalogue-based data appears to work sufficiently well, one question that remains is whether one can instead use catalogue-based simulations such as the those used in Section 3.2 to estimate a transfer function that more fairly represents the data in question.

To test this, we again begin by generating 1000 Gaussian realisations of an input C_ℓ . Since we find that the transfer function derived using a data-driven guess for the true C_ℓ performs better than if some arbitrary C_ℓ is used (see Figure 6), we opt for this as our input C_ℓ here. This also represents a more realistic scenario one would encounter when working with real data, for which the true power spectrum is unknown. From each of these Gaussian realisations we then draw Poisson realisations of a catalogue with a mean of 644786 sources per catalogue (i.e. the same number of sources as in the *Quia* ‘low- z ’ sample), following the procedure outlined in Section 3.2. From these we compute the angular power spectra using the catalogue-based approach, both with and without deprojecting the contaminant templates. We do this for both the case where the mask is provided directly as a continuous map, and the case where randoms are used to define the mask, and compute the transfer function for each using Eq. 33.

In both cases, we find that the catalogue-based simulations produce significantly noisier transfer functions than map-based simulations using the same input C_ℓ , with the scatter across simulations being up to ~ 100 (1000) times greater for the randoms-based (mask-based) approach. This is due to the impact of shot noise on the measured spectra before and after deprojection. We therefore recommend that the map-based approach of Section 3.3 be used to derive transfer functions, particularly when working with shot noise-dominated datasets. Note that this also has the advantage of being less expensive computationally in most real-world applications.

3.5 Application to Quia

Having validated our method using simulations, we now test its application to real data. We again turn to the *Quia* catalogue (Storey-Fisher et al. 2024), focusing in particular on the ‘low- z ’ sample of 644768 quasars with spectro-photometric redshifts in the range $z_p < 1.47$. After removing any quasars with redshift estimates outside of this range, we deproject the same templates used by Alonso et al. (2025); Fabbian et al. (2025) (see Appendix A for details) to mitigate

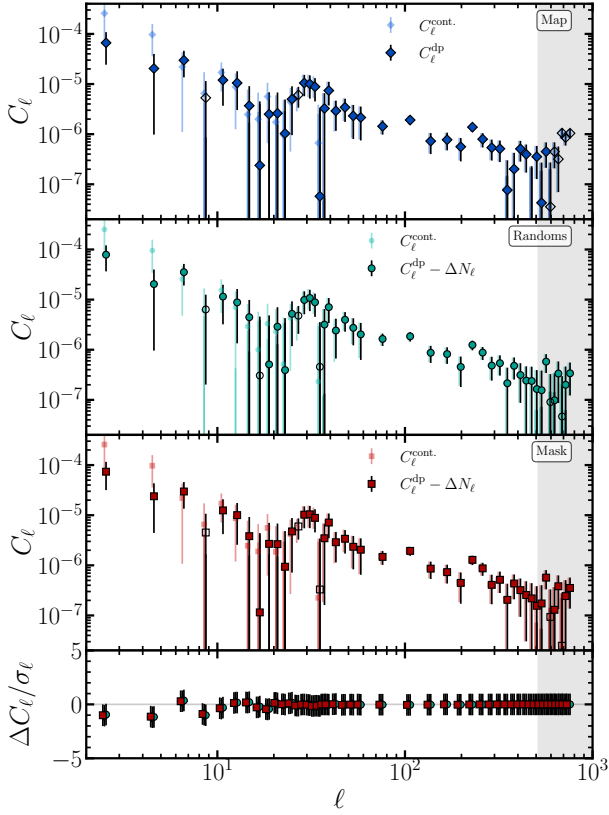


Figure 7. Application of three methods of deprojection to the low- z sample of quasars from *Quia*. The first three panels show (from top to bottom) the results of applying the traditional map-based approach to template deprojection, the catalogue-based approach in the case where randoms have been used to define the mask, and the catalogue-based approach where the mask has been provided as a continuous map. In each case, we show the angular power spectra pre- and post-deprojection (correcting for the noise deprojection bias in the catalogue-based approaches). The shot noise has already been subtracted in all cases. Open symbols denote negative power, for which we plot the absolute values. The bottom panel shows the difference between the pre- and post-deprojection angular power spectra, normalised by the corresponding uncertainties. All three methods are found to result in similar power deviations across all scales.

the contribution of systematics. We do this using three approaches for the sake of comparison: the standard map-based approach, the catalogue-based approach in the case of a continuous mask, and the catalogue-based approach using randoms to define the mask. In the latter two approaches, the shot noise is calculated with Eq. 12 (ignoring the contribution from randoms in the mask-based case) and subtracted from the measured power spectra. For the map-based approach, we assume that the shot noise is entirely Poissonian, in which case it can also be calculated analytically (Alonso et al. 2019; Nicola et al. 2020): the PCL of the noise is given by

$$\tilde{N}_\ell = \Omega_{\text{pix}} \frac{\langle w \rangle}{\langle n \rangle}, \quad (38)$$

where Ω_{pix} is the pixel area in steradians ($\approx 1.6 \times 10^{-5}$ sr for $N_{\text{side}} = 256$), $\langle w \rangle$ is the mean value of the *Quia* mask across the map, and $\langle n \rangle$ is the mean effective number of sources per pixel. The power spectrum of the shot noise is then obtained by multiplying it by the inverse mode-coupling matrix calculated from the *Quia* mask.

Figure 7 shows the noise-subtracted angular power spectra pre- and post-deprojection as measured with the three different approaches.

Uncertainties are estimated via the computation of Gaussian (or ‘disconnected’) power spectrum covariance matrices. Once the noise deprojection bias has been accounted for in the catalogue-based approaches, all three methods produce very similar measurements with respect to each other, both before and after template deprojection has been applied. The consistency of the deprojection methods themselves is exemplified in the bottom panel, in which we show the deviation in power induced by this step, normalised by the uncertainties: the deviations agree within errorbars for all three methods at all angular scales. This test thus validates our implementation of template deprojection in the catalogue-based PCL framework when applied to real data.

4 CONCLUSIONS

In this work, we present an extension to the catalogue-based PCL formalism of Wolz et al. (2025) that allows one to apply template deprojection in order to mitigate the contribution from systematic effects. In doing so, we combine the advantages of the existing catalogue-based formalism – i.e. avoiding biases stemming from the pixelisation of the observed fields – with systematics mitigation capabilities previously only available in the map-based formalism. We describe the application of this formalism to fields sampled at the positions of discrete sources, as well as to the particular case of galaxy clustering, in which the density of sources is the field of interest. Note that while the fields considered here are all scalar, our implementation is applicable to fields of arbitrary spin. A key part of this formalism lies in being able to account for mode loss in the noise component, induced by the deprojection procedure, which we demonstrate can be accurately computed analytically. We also introduce an alternative method to account for deprojection bias in the signal component, via the computation of a transfer function, and demonstrate its efficacy using simulations.

For the case of discretely-sampled fields, we validate our implementation by simulating 1000 Gaussian realisations of an arbitrary angular power spectrum, which are then contaminated with 100 synthetic templates and subsequently Poisson-sampled at the positions of galaxies from a real dataset (*Quia*) to create 1000 realisations of a catalogue with a realistic source distribution. We show that our implementation can be used to obtain unbiased estimates of the true power spectrum, after the mode loss induced by deprojection has been accounted for in both the noise and signal components.

We then simulate 1000 *Quia*-like catalogues and test the clustering-specific part of our formalism. We first demonstrate that care must be taken when choosing a maximum multipole, ℓ_{max} , for the deprojection procedure, as choosing a scale that is too small or too large can result in biased estimates of the deprojection parameters (which in turn will bias the final angular power spectrum). Having selected a suitable multipole ($\ell_{\text{max}} = 100$), we show that unbiased estimates of the true power spectrum can again be obtained after correcting for mode loss in the noise component. Since these simulated catalogues are shot noise-dominated, the effect of the transfer function is minimal in comparison, but we emphasise that for denser datasets this step will likely be necessary to achieve unbiased results. We find that all of this is true regardless of whether the mask is defined via a catalogue of randoms, or as a continuous map.

To assess the efficacy of the transfer function approach to mitigating mode loss caused by deprojection, we derive four transfer functions using different C_ℓ s as input for the simulations, and compare their performance with the standard deprojection bias approach by applying them each to a set of 1000 simulated *Quia*-like cata-

logues. All methods are found to produce similarly unbiased results, except at scales $\ell \lesssim 3$, where at least a reasonable approximation of the true angular power spectrum is required.

Finally, we apply our formalism to real data from *Quia* and compare it to the standard map-based approach. After correcting for the noise deprojection bias, both the randoms-based and mask-based approaches produce measurements consistent with the map-based approach, pre- and post-deprojection.

We have thus demonstrated that our formalism is a viable alternative to the traditional map-based approach when working with fields sampled at discrete positions. We anticipate that the method presented here will be useful in the analysis of large datasets incoming from Stage IV surveys such as the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), for which systematics mitigation will be a crucial step towards obtaining unbiased cosmological measurements, and the wealth of information contained at small scales will be better exploited by avoiding pixelisation. Our implementation is made available with the latest public release of the public PCL package *NaMaster*.

ACKNOWLEDGEMENTS

This work has been supported by STFC funding for UK participation in LSST, through grant ST/X00127X/1. DA and TC acknowledge support from the Beecroft Trust. BL is supported by the Royal Society through a University Research Fellowship. KW acknowledges support from the Science and Technology Facilities Council (STFC) under grant ST/X006344/1. Extensive use was made of computational resources at the University of Oxford Department of Physics, funded by the John Fell Oxford University Press Research Fund.

DATA AVAILABILITY

The quasar catalogue used in this work is publicly available at <https://zenodo.org/records/12805938>.

REFERENCES

- Abbott T. M. C., et al., 2022, *Phys. Rev. D*, **105**, 023520
 Adam R., et al., 2017, *A&A*, **598**, A115
 Agarwal N., et al., 2014, *J. Cosmology Astropart. Phys.*, **2014**, 007
 Alonso D., 2025, *The Open Journal of Astrophysics*, **8**, 53
 Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration 2019, *MNRAS*, **484**, 4127
 Alonso D., Hetmantsev O., Fabbian G., Slosar A., Storey-Fisher K., 2025, *The Open Journal of Astrophysics*, **8**, 42
 Baleato Lizancos A., White M., 2024, *J. Cosmology Astropart. Phys.*, **2024**, 010
 Bleem L. E., et al., 2022, *ApJS*, **258**, 36
 Bocquet S., et al., 2019, *ApJ*, **878**, 55
 Camphuis E., et al., 2025, *arXiv e-prints*, p. [arXiv:2506.20707](https://arxiv.org/abs/2506.20707)
 Carron J., Mirmelstein M., Lewis A., 2022, *J. Cosmology Astropart. Phys.*, **2022**, 039
 Chiang Y.-K., 2023, *ApJ*, **958**, 118
 Coulton W., et al., 2024, *Phys. Rev. D*, **109**, 063530
 Datta R., et al., 2024, *ApJS*, **273**, 26
 Elsner F., Leistedt B., Peiris H. V., 2016, *MNRAS*, **456**, 2095
 Elsner F., Leistedt B., Peiris H. V., 2017, *MNRAS*, **465**, 1847
 Elvin-Poole J., et al., 2018, *Phys. Rev. D*, **98**, 042006
 Fabbian G., Alonso D., Storey-Fisher K., Cornish T., 2025, *arXiv e-prints*, p. [arXiv:2504.20992](https://arxiv.org/abs/2504.20992)
 Faga L., et al., 2025, *MNRAS*, **536**, 1586

- Fang X., et al., 2024, *MNRAS*, **527**, 9581
 Ge F., et al., 2025, *Phys. Rev. D*, **111**, 083534
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, **622**, 759
 Heymans C., et al., 2021, *A&A*, **646**, A140
 Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, *ApJ*, **567**, 2
 Ho S., et al., 2012a, *ApJ*, **761**, 14
 Ho S., et al., 2012b, *ApJ*, **761**, 14
 Ho S., et al., 2015, *J. Cosmology Astropart. Phys.*, **2015**, 040
 Huterer D., Cunha C. E., Fang W., 2013, *MNRAS*, **432**, 2945
 Jogo B., Ruiz-Zapatero J., García-García C., Koukoufilippas N., Alonso D., 2023, *MNRAS*, **520**, 1895
 Johnston H., et al., 2021, *A&A*, **648**, A98
 Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64
 Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévy A., Pontzen A., 2013, *MNRAS*, **435**, 1857
 Louis T., et al., 2025, *arXiv e-prints*, p. [arXiv:2503.14452](https://arxiv.org/abs/2503.14452)
 Maniyar A., Lagache G., Béthermin M., Ilić S., 2019, *A&A*, **621**, A32
 Miyatake H., et al., 2023, *Phys. Rev. D*, **108**, 123517
 Nicola A., et al., 2020, *J. Cosmology Astropart. Phys.*, **2020**, 044
 Nicola A., García-García C., Alonso D., Dunkley J., Ferreira P. G., Slosar A., Spergel D. N., 2021, *J. Cosmology Astropart. Phys.*, **2021**, 067
 Planck Collaboration et al., 2020, *A&A*, **641**, A6
 Pullen A. R., Hirata C. M., 2013, *PASP*, **125**, 705
 Qu F. J., et al., 2024, *ApJ*, **962**, 112
 Reinecke M., Belkner S., Carron J., 2023, *A&A*, **678**, A165
 Rezaie M., Seo H.-J., Ross A. J., Bunescu R. C., 2020, *MNRAS*, **495**, 1613
 Rodríguez-Monroy M., et al., 2022, *MNRAS*, **511**, 2665
 Ross A. J., et al., 2011, *MNRAS*, **417**, 1350
 Rybicki G. B., Press W. H., 1992, *ApJ*, **398**, 169
 Shafer D. L., Huterer D., 2015, *MNRAS*, **447**, 2961
 Shaikh S., et al., 2024, *MNRAS*, **528**, 2112
 Slosar A., Seljak U., Makarov A., 2004, *Phys. Rev. D*, **69**, 123003
 Stölzner B., et al., 2025, *A&A*, **702**, A169
 Stomp R., Efstathiou G., 1999, *MNRAS*, **302**, 735
 Storey-Fisher K., Hogg D. W., Rix H.-W., Eilers A.-C., Fabbian G., Blanton M. R., Alonso D., 2024, *ApJ*, **964**, 69
 Sugiyama S., et al., 2023, *Phys. Rev. D*, **108**, 123521
 Tegmark M., 1997, *Phys. Rev. D*, **55**, 5895
 Totsuji H., Kihara T., 1969, *PASJ*, **21**, 221
 Vakili M., et al., 2023, *A&A*, **675**, A202
 Weaverdyck N., Huterer D., 2021, *MNRAS*, **503**, 5061
 Weaverdyck N., Muir J., Huterer D., 2018, *Phys. Rev. D*, **97**, 043515
 Wolz K., Alonso D., Nicola A., 2025, *J. Cosmology Astropart. Phys.*, **2025**, 028
 Yan Z., et al., 2025, *A&A*, **694**, A259
 Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *The Journal of Open Source Software*, **4**, 1298
 Zubeldia Í., Challinor A., 2019, *MNRAS*, **489**, 401

APPENDIX A: QUAIA TEMPLATES

We include here brief details of the *Quia* systematics templates used for the analysis carried out in Sections 3.2 and 3.3. These templates include: a Galactic dust extinction map (A_{dust}) from Chiang (2023); two maps of stellar contamination from *Gaia* and *unWISE* ($N_{\text{star}}^{\text{Gaia}}$ and $N_{\text{star}}^{\text{unWISE}}$, respectively); the scanning patterns of *Gaia* and *unWISE* (M_{10} and *unWISE scan*); stellar density maps in the regions around the Large and Small Magellanic Clouds measured with *Gaia* and *unWISE* (MCs (*Gaia*) and MCs (*unWISE*)); the three dipole moments associated with a multipole of $\ell = 1$ (Y_{10} , $\text{Re}(Y_{11})$, $\text{Im}(Y_{11})$).

We show Mollweide-projected maps of these templates in Fig. A1. In each case, we show the residual of a given pixel with respect to

the mean value of that systematic across the footprint ($S(\hat{\mathbf{n}}) - \bar{S}$), normalised by the rms deviation across the footprint (σ_S).

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

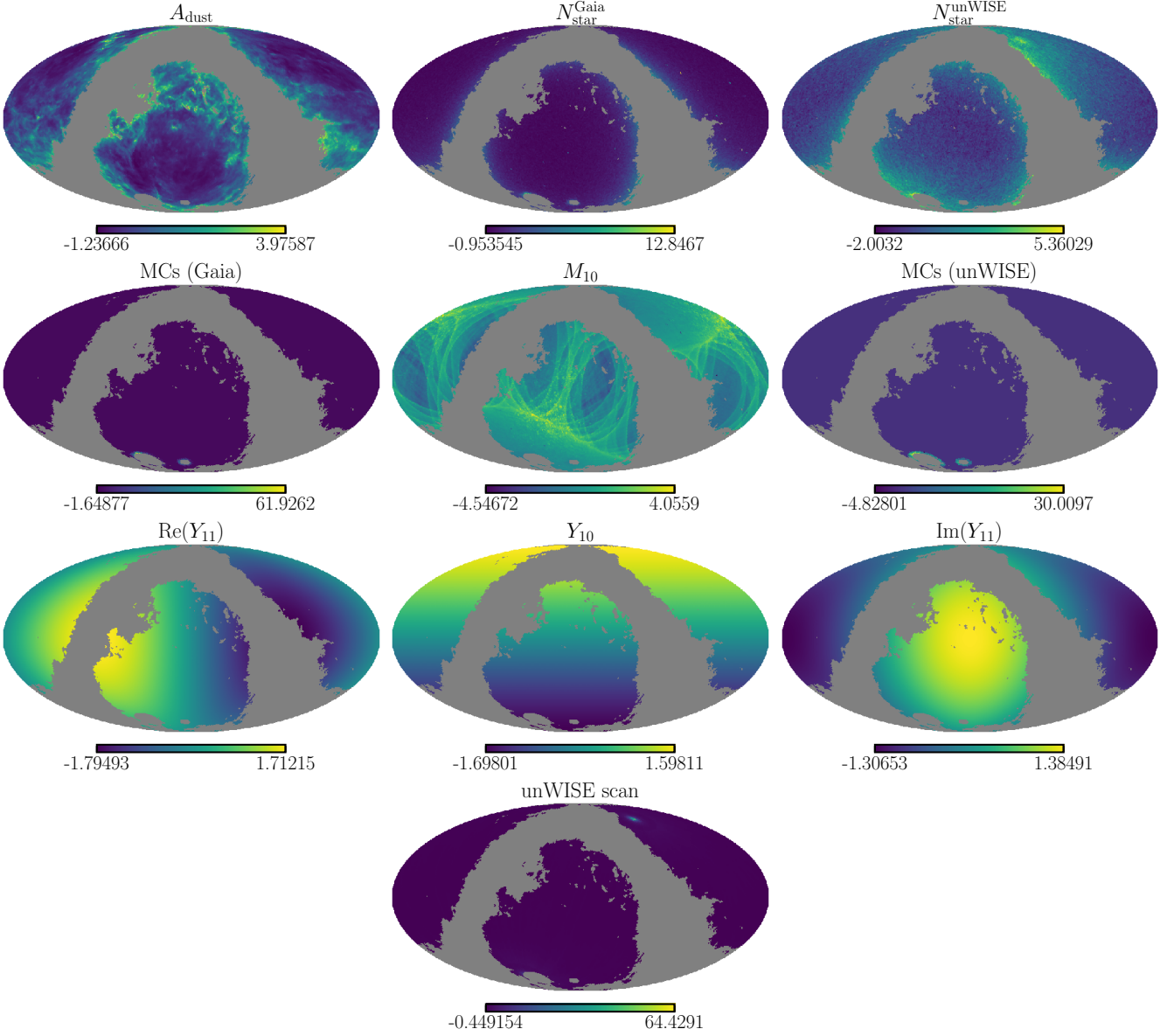


Figure A1. Systematic templates used when simulating clustering data in Sections 3.2 and 3.3, and when applying our catalogue-based deprojection formalism to *Quaia* data in Section 3.5. As in Fabbian et al. (2025), we show for each systematic template $S(\hat{\mathbf{n}})$ the quantity $(S(\hat{\mathbf{n}}) - \bar{S})/\sigma_S$, where $\bar{S}$ and σ_S are the mean and rms deviation within the footprint, respectively.