

# Mitigating contamination in LSS surveys: a comparison of methods

Noah Weaverdyck\* and Dragan Huterer

*Department of Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040 and  
Leinweber Center for Theoretical Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040\**

(Dated: July 30, 2020)

Future large scale structure surveys will measure the locations and shapes of billions of galaxies. The precision of such catalogs will require meticulous treatment of systematic contamination of the observed fields. We compare several existing methods for removing such systematics from galaxy clustering measurements. We show how all the methods, including the popular pseudo- $C_\ell$  Mode Projection and Template Subtraction methods, can be interpreted under a common regression framework and use this to suggest improved estimators. We show how methods designed to mitigate systematics in the power spectrum can be used to produce clean maps, which are necessary for cosmological analyses beyond the power spectrum, and we extend current methods to treat the next-order multiplicative contamination in observed maps and power spectra. Two new mitigation methods are proposed, which incorporate desirable features of current state-of-the-art methods while being simpler to implement. Investigating the performance of all the methods on a common set of simulated measurements from Year 5 of the Dark Energy Survey, we test their robustness to various analysis cases. Our proposed methods produce improved maps and power spectra when compared to current methods, while requiring almost no user tuning. We end with recommendations for systematics mitigation in future surveys, and note that the methods presented are generally applicable beyond the galaxy distribution to any field with spatial systematics.

## I. INTRODUCTION

Over the past 40 years, cosmological surveys have produced increasingly detailed maps of the large-scale structure (LSS) in the Universe [1–12]. These observations have proven crucial for testing our understanding of gravity and cosmological structure formation, and helped to constrain cosmological parameters to the percent level [e.g. 12–14]. Recent observations from DES have for the first time imposed strong constraints on dark energy using an LSS survey alone, independently of the cosmic microwave background [15]. Upcoming ground-based missions like DESI [16], and the Rubin Observatory’s LSST [17], along with space-based missions like SPHEREx [18], Euclid [19], and RST (formerly WFIRST) [20] will truly herald the age of precision cosmology, mapping up to  $\sim 20$  billion galaxies across the sky and bringing unprecedented precision to measurements of the dark energy equation of state and modified gravity. Such statistical precision makes the control of systematic errors in these datasets of paramount importance to avoid biasing cosmological analyses.

Cosmological information is extracted from LSS observations in multiple ways. The most common approach is to calculate the two-point correlation function or its Fourier counterpart, the power spectrum, to characterize the spatial distribution of galaxies (galaxy clustering) or their shapes (weak lensing). To date, these have been used in cosmological analyses to great success [21–41]. The two-point function contains all available information when the field it characterizes is Gaussian, but nonlinear gravitational collapse induces non-Gaussianity

at late times and small scales. Therefore there is considerable cosmological information that is inaccessible to the two-point function. This has led to growing interest in using complementary statistical representations of LSS observations, such as higher order N-point functions [42–51], statistics of peaks [52–54] and voids [55–61], density-split statistics [62], marked power spectra [63–66], Minkowski functionals [67–72], wavelet transforms [73, 74] and more. Further information is also gleaned from the baryon acoustic oscillations and redshift-space distortions. These methods rely on the accurate mapping of the underlying cosmological fields from which they are derived, so there is increasing need for tools to mitigate systematic contamination at the levels of both the map and the two-point functions.

Here, we consider a very general class of systematics that describes an arbitrary spatial modulation of the observed field. Such generic sources of error are one of the most serious contaminants in our quest to probe cosmology with future surveys. For definiteness we focus on the case of galaxy clustering, where the systematic error corresponds to a modulation of the galaxy selection function in redshift or across the observing footprint. **However, the methods we test and develop in this paper are general enough to apply to any real or complex field for which there exist maps of potential contaminants** (e.g. shear or Sunyaev-Zeldovich-effect fields).

Spatially-varying systematics in LSS maps may be caused by a large variety of physical effects. These include observing conditions and dust extinction (both of which effectively create a position-dependent “screen,” obscuring background galaxies), bright objects and star-galaxy separation (which can eclipse, change the shape, or be confused for galaxies close to them on the sky),

---

\* Corresponding author: nweaverd@umich.edu

and variations in sensitivity of the detector (which include potentially time- and position-dependent variations in the focal plane), or imaging pipeline. In all of these cases, failure to fully account for variability in the selection function will result in residual artifacts — *calibration errors* — in the final data product and potentially bias results [75–77]. The presence of calibration errors is evidenced by a number of surveys [78–86] which have shown a significant excess of power at large scales where calibration errors are thought to be most prevalent. Recent observations (e.g. from the Dark Energy Survey [87]) demonstrate however that such contamination is by no means limited to large scales alone. In addition to adding power, calibration errors induce a multiplicative effect, coupling different scales and thus affecting all scales in the survey, including those smaller than the typical size of the calibration systematic itself [75, 76]. Much recent work [81, 82, 86, 88–103] has focused on mitigating these systematics in order to probe the underlying cosmology.

The simplest strategy to ameliorate the effects of calibration errors is to simply mask scales or data points suspected having large levels of contamination. More sophisticated strategies include using maps of suspected contaminants — so-called ‘templates’ — to correct the observations. An alternative and complementary approach (e.g. [104–111]) is to forward-model many possible realizations of the cosmic initial conditions. One then evolves these initial conditions in time (while adding realizations of nonlinearities, bias, and observational/instrument systematics), and performs joint inference of cosmology, the initial conditions and late-time “true” fields, given observations. Yet another forward-modeling approach involves the injection of false images into observations in order to sample the selection function [93, 102]. While such forward approaches are powerful and very general, they also require extensive computational resources and are complicated to implement. In contrast, using templates to clean contaminated observations and directly infer the underlying fields is straightforward to implement and can be readily incorporated into ongoing or completed analyses. They have been the dominant approach in the community thus far, and so these are the methods we focus on here.

In this paper we revisit and extend state-of-the-art LSS systematics-cleaning strategies. We interpret them through a regression framework to highlight commonalities and differences of the methods, as well as some tacit assumptions. In doing so, we show that the common pseudo- $C_\ell$  Mode Projection method is equivalent to linear regression. We use this framework to propose straightforward extensions that leverage the extensive body of literature and tools that have been developed for regression analyses. We rigorously test the performance of several existing methods, plus new ones that we propose, on a common set of simulated observations from current and future surveys.

We study performance using an ensemble of simulated galaxy overdensity maps, such that we can assess both

the accuracy and precision of each method. We provide a library of templates and a contaminated overdensity map as input to each cleaning method, which then produces an estimate of the true overdensity map and power spectrum that we assess for accuracy. We repeat the process over a large number of sky realizations and for various configurations of templates to asses the precision and robustness of each method. A schematic outline of this process is shown in Fig. 1.

The paper is organized as follows. In Sec. II we describe in detail our general model for contamination, which encompasses a wide range of systematics due to foregrounds or instrument calibration errors. In Sec. III we describe several existing methods for systematics mitigation; in Sec. IV we reinterpret the methods through a common framework to facilitate comparison, and in Sec. V we use this to map several aspects onto well-known techniques in statistics and propose two new mitigation methods. In Sec. VI we describe the fiducial synthetic surveys on which we test the efficacy of the methods that we study. Sec. VII shows the results of these performance comparisons, while Sec. IX has our conclusions. Several Appendices show important but more technical and detailed aspects of the investigation.

## II. CONTAMINATION MODEL

We first introduce the model for contamination of the observed LSS fields. It is very general, encompassing most known sources of real-world contamination.

Following Huterer *et al.* [75], we model the observed number density map as the product of the true galaxy number density map ( $N_{\text{true}}(\hat{\mathbf{n}})$ ) and a direction-dependent screen ( $1 + f_{\text{sys}}(\hat{\mathbf{n}})$ ),

$$N_{\text{obs}}(\hat{\mathbf{n}}) = (1 + f_{\text{sys}}(\hat{\mathbf{n}})) N_{\text{true}}(\hat{\mathbf{n}}). \quad (1)$$

Here,  $f_{\text{sys}}(\hat{\mathbf{n}})$  characterizes the systematic modulation of the true field, such that pixels with  $f_{\text{sys}}(\hat{\mathbf{n}}) = 0$  are free of contamination. Using  $N = \bar{N}(1 + \delta)$  and defining the ratio of true to observed mean number density as  $\gamma = \bar{N}_{\text{true}}/\bar{N}_{\text{obs}}$ , the observed overdensity can be written as

$$\delta_{\text{obs}}(\hat{\mathbf{n}}) = \gamma(\delta(\hat{\mathbf{n}}) + 1)(f_{\text{sys}}(\hat{\mathbf{n}}) + 1) - 1. \quad (2)$$

Here  $\gamma$  enforces the constraint that  $\langle \delta_{\text{obs}} \rangle_{\text{pix}} = 0$  across the survey footprint, even though this is not necessarily true for the *true* overdensity field  $\delta$ . This is due to the fact that we can only access the observed mean number density  $\bar{N}_{\text{obs}}$ , which differs from the true mean both because of systematic contamination and because of sample variance from a limited survey footprint (see Sec. IV B for details).

This model for contamination is similar to the one used in [75–77, 112] to assess the impacts of *residual* calibration errors that remain in the data after cleaning. Here we focus on the methods used to perform such cleaning,

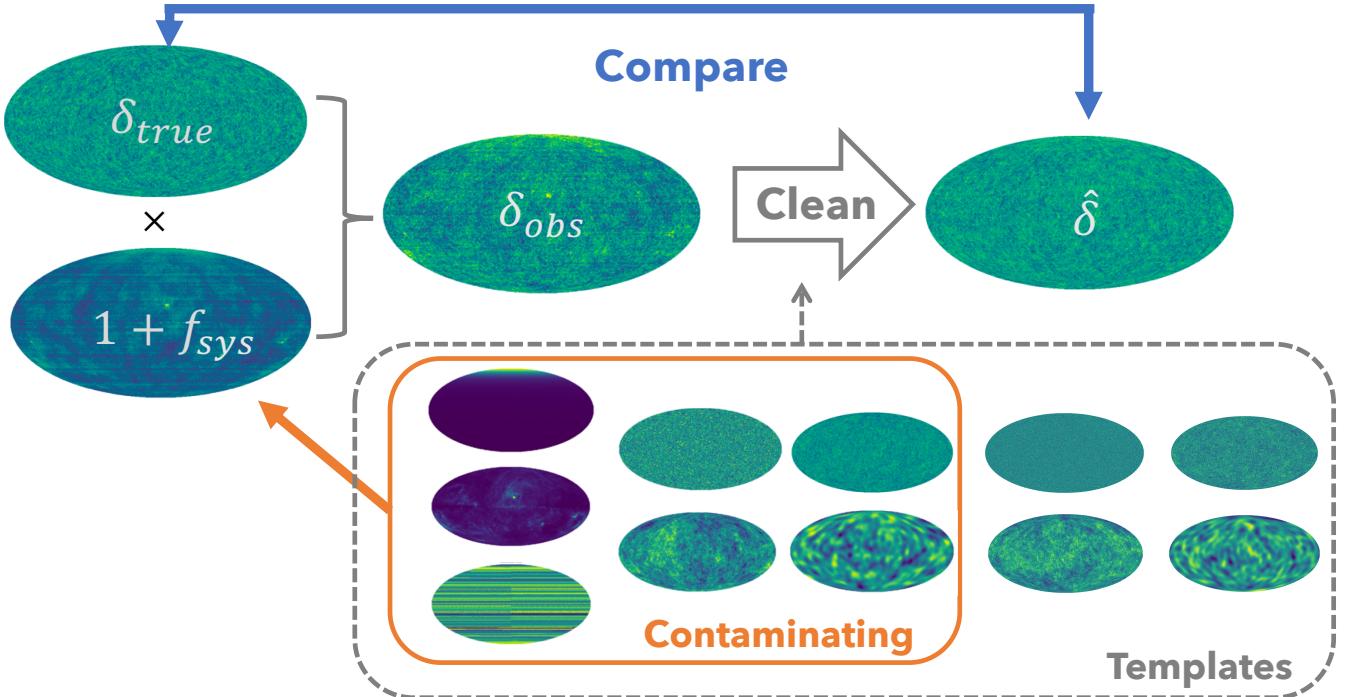


FIG. 1. Analysis procedure for a single map. A set of templates is generated (dashed box) along with a true overdensity map  $\delta_{true}$ . A subset of the templates (orange box) contaminate the true overdensity map to generate the observed overdensity field  $\delta_{obs}$ . We generate an estimated signal map  $\hat{\delta}$  using one of the cleaning methods, and compare it to the truth, either at a map-level or power-spectrum-level. This is repeated for many realizations of the signal map and the performance of each cleaning method is assessed.

and so use the screen model to describe contamination more generally.

We extend the screening formalism by considering that the total systematic modulation is comprised of  $N_{sys}$  individual systematics, each of which acts as its own screen. Thus we have

$$1 + f_{sys} = \prod_{i=1}^{N_{sys}} (1 + f_i) \quad (3)$$

$$\simeq 1 + \sum_{i=1}^{N_{sys}} f_i + \sum_{j \neq k}^{N_{sys}} f_j f_k + [\text{higher order terms}]$$

where we have suppressed  $\hat{\mathbf{n}}$  in the notation for convenience both here and in what follows. Note that even if a systematic individually contributes to  $f_{sys}$  linearly, there exist interaction terms with other systematics up to order  $N_{sys}$ . Here and in general,  $f_i \equiv f_i(\hat{\mathbf{n}})$  is a column vector with each element corresponding to a pixel, unless otherwise noted.

### III. BACKGROUND: EXISTING MITIGATION METHODS

The principal goal of this paper is to compare various proposed systematics mitigation methods. The methods that we test are all designed to use maps that trace potential contamination in order to mitigate the impact of systematics, i.e. they assume that the systematic  $f_i(\hat{\mathbf{n}})$  is a function of some tracer  $t_i(\hat{\mathbf{n}})$ . We refer to these tracer maps as *templates*, and examples include maps of stellar density, extinction, or summary statistics of observing conditions (e.g. mean g-band seeing) in each region of the sky throughout the duration of the survey (see Leistedt *et al.* [87] for a detailed description of the process for creating templates from multi-epoch observational data for the Dark Energy Survey). Sources of error for which we have no templates (e.g. shot noise) are implicitly subsumed into the overdensity field.

We will investigate how effectiveness depends on analysis choices and suggest improvements where possible. We start with three principal methods that have been applied in the literature: the Dark Energy Survey Year 1 method (henceforth DES-Y1), the Template Subtraction method (TS), and the Mode Projection method (MP). While at face value the algorithms associated with these methods seem quite different, we demonstrate that they

can be translated into a common mathematical framework of linear regression. Doing so allows us to distill commonalities and differences between the methods, as well to identify simplifications and extensions to them. We include three additional methods based on these insights.

For all the methods, we will work with maps that are divided into pixels in HEALPix<sup>1</sup>[113] format, which summarize the mean galaxy overdensity or template values within each pixel (see Sec. VI for details). Furthermore, while we work in the context of cleaning galaxy overdensity fields, the methods are applicable more generally to corrections of any field for which we have templates of potential contamination, and so we denote the true signal more generally as  $s$  and the observed field as  $d_{\text{obs}}$ . In our application, these correspond to the true and observed galaxy overdensity fields,  $\delta_{\text{true}}$  and  $\delta_{\text{obs}}$ . In the sections that follow, we use  $\hat{x}$  to denote an estimate of  $x$ , and  $\tilde{C}_{\ell}^{xx}$  to indicate a realization-specific measurement of the power spectrum, as compared to its theoretical mean  $C_{\ell}^{xx}$ .

### A. Dark Energy Survey Y1 Method

The method used to derive galaxy weights for the Year-1 DES release is one of the more sophisticated mitigation methods applied to date. It is described in detail in Elvin-Poole *et al.* [41], but we review its main features here. Hereafter referred to as ‘DES-Y1,’ it builds on the method first proposed as the ‘Weights’ method in Ross *et al.* [88], wherein 1-dimensional relationships between observed galaxy densities and systematic templates (there called ‘survey property maps’) are removed by iteratively applying multiplicative factors (‘weights’) to galaxies. Fig. 2 shows one example of how the observed overdensity varies with a template. Multiplicative weights are applied to galaxies to de-trend the data, shifting the blue line to lie atop the dashed line. This method is explicitly a regression method, with versions employing linear fits [114–116], splines [117] or higher-order polynomials [118] as fitting functions for the 1D relationships.

Here we describe the version we adopt, which closely follows the implementation in Elvin-Poole *et al.* [41] used on the DES-Y1 data. For each template  $t_i$ , we group pixels into 10 evenly-spaced bins based on their template values, and independent of location on the sky (e.g. all pixels with a mean  $i$ -band seeing value within 10% of the max would be grouped). We then find the mean galaxy overdensity over the pixels in each bin<sup>2</sup>. A  $10 \times 10$  covari-

ance matrix of these bin means is estimated by performing the same bin-averaging process on a set of 400 uncontaminated mock maps, generated with a fiducial power spectrum for the overdensity field (we assume the true overdensity power spectrum to generate these mocks).

Next, we use `scipy.optimize` and the estimated covariance to find the parameters  $\{m_i, b_i\}$  of the best-fit line<sup>3</sup> of the binned overdensity to each binned template  $i$ :

$$\frac{\langle N_{\text{obs}} \rangle_j}{N_{\text{obs}}} = m_i \langle t_i \rangle_j + b_i \quad (4)$$

where  $\langle \cdot \rangle_j$  indicates the average pixel value in bin  $j$  of the given template. See Fig. 2 (blue points and trend) for an illustration.

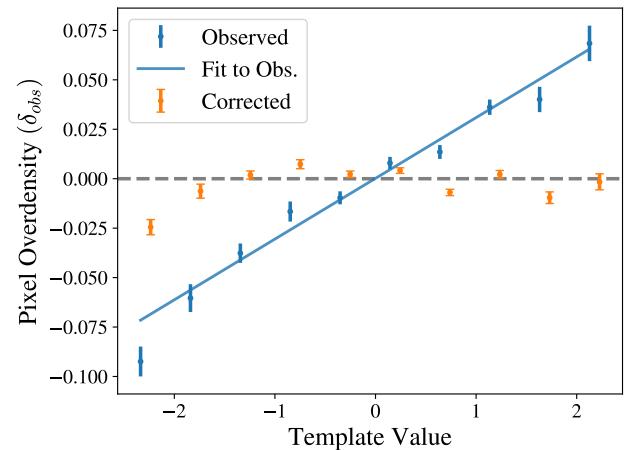


FIG. 2. Illustration of the DES-Y1 cleaning method, showing the total observed pixel overdensity ( $\delta_{\text{obs}}$ ) as a function of a template’s pixel overdensity, in ten evenly-spaced bins. Given the estimated covariance matrix (diagonals shown by blue error bars), the best-fit trend (blue line) can be calculated and used to reweight the observed map, producing a corrected map whose dependence on the template is removed (orange points, with corresponding standard errors on the pixel means). The process is then iterated for other templates until a satisfactory threshold is reached; see text for details.

The template with the most significant fit is used to reweight the number density in each pixel as  $N'_{\text{obs}}(\hat{\mathbf{n}}) = N_{\text{obs}}(\hat{\mathbf{n}})/(\hat{m}_i t(\hat{\mathbf{n}}) + \hat{b}_i)$ , where the significance metric is defined below. Having removed the effect of the dominant

<sup>1</sup> <http://healpix.sourceforge.net>

<sup>2</sup> In Elvin-Poole *et al.* [41], extreme regions are removed by eye: each template is inspected and bins that exhibit an average fluctuation in number density of  $> 20\%$  are masked, as are regions where visual inspection suggests a deviation from non-monotonic behavior (see their Fig. 3). We neglect this step, as it is difficult

to automate robustly and in our tests we found that it did not alter our results.

<sup>3</sup> Elvin-Poole *et al.* [41] also use linear fits for almost all templates, with only a couple exceptions. As noted in Sec. IV B, even if a template is thought to contaminate non-linearly, the relationship can usually be made linear through an appropriate transformation of the template.

systematic, the whole process is repeated: for each template, the pixels are assigned to bins and averaged, the new best-fit parameters are computed from Eq. (4), and the trend from the most significant template is removed from the data. The process stops when all templates are below a predefined significance threshold.

In general, the more contamination from template  $i$ , the stronger relationship the relationship with the observed galaxy density. However, some level of correlation is expected just by chance, and this depends on the spatial clustering of each template. The DES-Y1 method addresses this in two ways: (1) by using a different covariance matrix for the observed overdensity for each template as described above, and (2) by having a template-specific significance threshold, calibrated on mocks. Specifically, the significance statistic used is  $\Delta\chi_i^2/[\Delta\chi_i^2]_{68}$ , where  $\Delta\chi_i^2$  is the improvement in  $\chi^2$  for the binned fit on template  $i$ , compared to a null hypothesis of  $m_i = b_i = 0$ . It is normalized to the 68<sup>th</sup> percentile of the same quantity measured on uncontaminated signal mocks ( $[\Delta\chi_i^2]_{68}$ ). We use the stopping criterion  $\Delta\chi_i^2/[\Delta\chi_i^2]_{68} < \Delta\chi_{\text{threshold}}^2 = 2$ , but find that our results change little when changing this threshold between 1 and 4 (see App. E).

There are a number of required parameter choices in the DES-Y1-type method. These include the criterion for selecting the most significant template,<sup>4</sup> the significance threshold that determines when to stop weighting, the prior power spectrum for generating mocks, and choices associated with binning (e.g. number of bins, equally-spaced vs. equally-filled, etc). Here we use the fiducial choices from Elvin-Poole *et al.* [41], and investigate some of the effects of these choices in App. E.

## B. Template Subtraction

The Template Subtraction method uses the cross-power of templates with the observed sky to estimate contamination of each template at each angular scale. Contamination is subtracted directly from the two-point clustering statistics. The method was proposed in Ho *et al.* [82] and Ross *et al.* [88] where it was called the “cross-correlation” technique, and we review it here.

Template Subtraction assumes the observed overdensity  $d_{\text{obs}}$  is a linear combination of the true galaxy overdensity  $s$  and individual template overdensities  $t_i$ :

$$d_{\text{obs}} = s + \sum_{i=1}^{N_{\text{tpl}}} \alpha_i t_i. \quad (5)$$

Any systematics or noise not accounted for by templates are subsumed into the signal  $s$ . In Ho *et al.* [82],  $d_{\text{obs}}$

<sup>4</sup> E.g. one could consider an  $R^2$  statistic, the commonly-used  $F$ -statistic, Akaike or Bayesian information criteria, etc.

and  $t_i$  are taken to be in multipole space, such that  $\langle ss \rangle \rightarrow \langle s_{\ell m} s_{\ell m} \rangle = C_{\ell}^{ss}$  (where  $\langle \cdot \rangle$  is the ensemble average over many sky realizations), and  $\alpha \rightarrow \alpha_{\ell}$  is a function of  $\ell$ . The companion paper of Ross *et al.* [88] works in configuration space, so in their version of Template Subtraction, the data vectors are in pixel-space.

We will work in harmonic space and so follow Ho *et al.* [82], but we will keep the notation general until dealing with the two-point functions where we will explicitly work with power spectra. The treatment for configuration space is largely identical. To apply the method, one would simply substitute the correlation function for the power spectrum  $C_{\ell}^{ij} \rightarrow w^{ij}(\theta)$  and  $\alpha_{\ell} \rightarrow \alpha(\theta)$ . See e.g. Crocce *et al.* [119] for an application of template subtraction to the correlation function.

If we consider just a single contaminant for simplicity ( $N_{\text{tpl}} = 1$ ), and assume that it is uncorrelated with the underlying galaxy field, then from Eq. (5) the two point function of the observed field is

$$\langle d_{\text{obs}} d_{\text{obs}} \rangle = \langle ss \rangle + \alpha^2 \langle tt \rangle. \quad (6)$$

Then on average,

$$\langle td_{\text{obs}} \rangle / \langle tt \rangle = C_{\ell}^{td} / C_{\ell}^{tt} = \alpha_{\ell} \quad (7)$$

and the contamination at each multipole can be estimated as

$$\hat{\alpha}_{\ell} = \tilde{C}_{\ell}^{td} / C_{\ell}^{tt} \quad (8)$$

where the tilde in  $\tilde{C}_{\ell}$  indicates the power spectrum that is measured from the observed sky realization, and  $\tilde{C}_{\ell}^{tt} = C_{\ell}^{tt}$  since we take templates to be fixed.

An estimate of the power spectrum can then be found to be

$$\hat{C}_{\ell}^{ss} = \left( \tilde{C}_{\ell}^{dd} - \hat{\alpha}_{\ell}^2 C_{\ell}^{tt} \right) \left( 1 - \frac{1}{2\ell+1} \right)^{-1}. \quad (9)$$

Here  $[1 - 1/(2\ell+1)]^{-1} = [(2\ell+1)/(2\ell)]$  is a factor found by Elsner *et al.* [120] that is needed to debias the estimator.<sup>5</sup> The bias arises because the process is too aggressive — any chance correlation between template and the true signal is also removed, resulting in an underestimate of the true clustering power.

The Template Subtraction method is easily generalized to multiple templates by extending the dimensionality of terms as

$$\begin{aligned} \alpha_{\ell} \text{ (scalar)} &\rightarrow \boldsymbol{\alpha}_{\ell} \text{ (\textit{N}_tpl)} \\ C_{\ell}^{tt} \text{ (scalar)} &\rightarrow \mathbf{C}_{\ell}^{\text{TT}} \text{ (\textit{N}_tpl} \times \textit{N}_tpl) \end{aligned}$$

<sup>5</sup> In the case of the correlation function, the bias cannot be written in a signal-independent fashion, and so requires a prior signal power spectrum or simulations to estimate.

and Eqs. (8) and (9) become

$$\hat{\mathbf{a}}_\ell = [\tilde{\mathbf{C}}_\ell^{\text{TT}}]^{-1} [\tilde{\mathbf{C}}_\ell^{\text{TD}}], \quad (10)$$

$$\hat{C}_\ell^{ss} = \left( \tilde{C}_\ell^{dd} - \hat{\mathbf{a}}^\dagger \mathbf{C}_\ell^{\text{TT}} \hat{\mathbf{a}} \right) \left( \frac{2\ell+1}{2\ell+1-N_{\text{tpl}}} \right). \quad (11)$$

For the cut-sky equations, we refer the reader to Ref. [120].

While previous work on Template Subtraction has focused on the cleaned power spectrum, an estimate of cleaned overdensity field itself is also of interest for cosmological study, as it contains more information than just its power spectrum.

A map estimate from the Template Subtraction method can be produced as

$$\hat{s}_{\ell m}^{TS}(\hat{\mathbf{n}}) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} \hat{s}_{\ell m}^{TS} Y_{\ell m}(\hat{\mathbf{n}}), \quad (12)$$

where the harmonic coefficients of the map are given by

$$\hat{s}_{\ell m}^{TS} = (d_{\text{obs}})_{\ell m} - \sum_{i=1}^{N_{\text{tpl}}} (t_{\ell m})_i (\hat{\alpha}_\ell)_i \quad (13)$$

and the (biased) power spectrum of the cleaned map is equivalent to the first factor in Eq. (11).

### C. Mode Projection

Mode Projection (also often called Mode Deprojection [96, 98, 121–123]) assumes the same contamination model as Template Subtraction, given by Eq. (5). The original formulation [89, 124] cleans the map-level systematics by assigning infinite variance to contaminating templates. This procedure desensitizes the power spectrum estimate to the templates and is equivalent to marginalizing over the contamination amplitude of each template [89].

In particular, it updates the map-level covariance matrix  $\mathbf{C}$  as follows

$$\begin{aligned} \mathbf{C}' &= \left[ \mathbf{C} + \sum_k^{N_{\text{tpl}}} \lim_{\beta \rightarrow \infty} (\beta_k t_k t_k^\dagger) \right] \\ &= \lim_{\beta \rightarrow \infty} [\mathbf{C} + \beta T T^\dagger] \end{aligned} \quad (14)$$

where  $t_k$  are the individual template maps, which can represent either real spin-0 or complex spin-2 fields [122], and which can be assembled into a matrix  $T$ , with  $t_k$  as the  $k^{\text{th}}$  column. In previous works with Mode Projection, the maps have been represented in pixel space, but in principle the operations can also be performed in harmonic space, e.g. representing a spin-0 field by its complex harmonic coefficients. For clarity and continuity, we will assume the maps are  $N_{\text{pix}}$ -length vectors in

what follows, as opposed to their multipole transforms. There are some benefits to performing Mode Projection in harmonic space, however, which we explore in Sec. IV.

The main challenge with the original formulation of Mode Projection is that it requires the construction and inversion of a covariance matrix for the whole map, which is often intractable. To remedy this, Elsner *et al.* [125] extended Mode Projection to the popular (albeit sub-optimal) pseudo- $C_\ell$  estimator. In practice, this is achieved by computing the pseudo- $C_{\ell s}$  of the overdensity field after first applying a filter  $\mathbf{F}$ , where

$$\begin{aligned} \mathbf{F} &= \lim_{\beta \rightarrow \infty} (I + \beta T T^\dagger)^{-1} \\ &= I - T(T^\dagger T)^{-1} T^\dagger, \end{aligned} \quad (15)$$

where the second expression follows from the Sherman-Morrison-Woodbury formula. It is easy to see that  $T(T^\dagger T)^{-1} T^\dagger$  is a projection matrix, projecting an  $N_{\text{pix}}$ -dimensional map onto a  $N_{\text{tpl}}$ -dimensional subspace. The filter thus removes any components of the observed map within the subspace spanned by the templates (hence the alternate name of Mode Deprojection).

Taking the case of a single template map  $t$  for simplicity,  $\mathbf{F}$  then takes the form  $(I - (tt^\dagger)/(t^\dagger t))$ , resulting in a filtered overdensity map

$$\hat{s} = \mathbf{F} d_{\text{obs}} \quad (16)$$

$$= [I - t(t^\dagger t)^{-1} t^\dagger] d_{\text{obs}} \quad (17)$$

$$= d_{\text{obs}} - t \hat{\alpha}_{\text{mp}} \quad (18)$$

where

$$\hat{\alpha}_{\text{mp}} = (t^\dagger d_{\text{obs}})/(t^\dagger t) = \tilde{\sigma}_{td}^2/\sigma_{tt}^2, \quad (19)$$

and  $\tilde{\sigma}_{td}^2$  is a measure of the covariance of maps  $t$  and  $d$ . Note that this is very similar to the Template Subtraction estimate in Eq. (8), but here the covariances are taken over the whole footprint, rather than for a single mode  $\ell$ . We can make the connection even more explicit by noting that in the full-sky case,

$$\tilde{\sigma}_{td}^2 = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell+1) \tilde{C}_\ell^{td} \quad (20)$$

While Elsner *et al.* [125] introduce this filtered map only as a means to compute the power spectrum, it can be used on its own as an estimate for the cleaned overdensity field. However, as with Template Subtraction, the power spectrum of this cleaned map is a biased estimate of the true power spectrum, as some of the signal is removed in the cleaning process:

$$\langle C_\ell^{\hat{s}\hat{s}} \rangle = \langle (d_{\text{obs}} - t \hat{\alpha}_{\text{mp}})^\dagger (d_{\text{obs}} - t \hat{\alpha}_{\text{mp}}) \rangle \quad (21)$$

$$= C_\ell^{ss} - \frac{C_\ell^{tt}}{4\pi (\sigma_{tt}^2)^2} \left( 2C_\ell^{ss} \sigma_{tt}^2 - \frac{1}{4\pi} \sum_{\ell'} (2\ell'+1) C_{\ell'}^{ss} C_{\ell'}^{tt} \right). \quad (22)$$

In the full sky case, the power spectrum estimate can be debiased analytically [125]:

$$\hat{C}_\ell^{ss} = \sum_{\ell'} [(I + B)^{-1}]_{\ell\ell'} C_{\ell'}^{\hat{s}\hat{s}}, \quad (23)$$

where

$$B_{\ell\ell'} = \frac{C_\ell^{tt}}{4\pi(\sigma_{tt}^2)^2} \left( -2\sigma_{tt}^2 \delta_{\ell\ell'} + \frac{2\ell' + 1}{4\pi} C_{\ell'}^{tt} \right) \quad (24)$$

and  $\delta_{\ell\ell'}$  is the Kronecker delta. In the presence of a mask, one can debias via iteration or assuming a prior power spectrum [125]. As we work in the full-sky case, we debias analytically via Eq. (23), though we do not expect an iterative or prior-based debiasing to significantly alter our conclusions.

The procedure outlined above easily generalizes to multiple maps by extending the dimensionality of the terms:

$$\begin{aligned} \alpha \text{ (scalar)} &\rightarrow \alpha \text{ ($N_{\text{tpl}}$)}, \\ t \text{ ($N_{\text{pix}}$)} &\rightarrow \mathbf{T} \text{ ($N_{\text{pix}} \times N_{\text{tpl}}$)}, \\ \sigma_{td}^2 \text{ (scalar)} &\rightarrow \sigma_{\mathbf{Td}}^2 \text{ ($N_{\text{tpl}}$)}, \\ C_\ell^{tt} \text{ (scalar)} &\rightarrow \mathbf{C}_\ell^{\mathbf{TT}} \text{ ($N_{\text{tpl}} \times N_{\text{tpl}}$)}, \\ \sigma_{tt}^2 \text{ (scalar)} &\rightarrow \sigma_{\mathbf{TT}}^2 \text{ ($N_{\text{tpl}} \times N_{\text{tpl}}$)}. \end{aligned} \quad (25)$$

Hereafter, we will use ‘Mode Projection’ to refer to the pseudo- $C_\ell$  mode projection method described above, due to the popularity of the pseudo- $C_\ell$  power spectrum estimator and the adoption of this version into NaMaster<sup>6</sup>[122], in anticipation of LSST. We again refer the reader to Elsner *et al.* [125] for the modifications necessary to account for the mask, and specifically to their Eq. (21) for the multi-template version of the debiasing matrix, which we use to correct for all Mode Projection power spectrum estimates (see Ref. [122] for the equivalent formulae for spin-2 fields).

#### IV. PLACING INTO A COMMON MATHEMATICAL FRAMEWORK

To facilitate a comparison of the methods, it is useful to place them into a common mathematical framework. In this section, we show how all three methods presented so far can be interpreted through a regression analysis lens, and in doing so help identify different assumptions within each method and possible avenues for improvement. Moreover, we can leverage the powerful suite of tools that have already been developed and tested for regression to the task of systematics removal, facilitating and accelerating the process.

#### A. Connections to Regression

We have purposefully formulated the methods (e.g. Eqs. (19 and (8)) in a manner designed to make the connections between Mode Projection and Template Subtraction apparent. Template Subtraction is equivalent to running the Mode Projection algorithm, but with each original template ( $t_i(\hat{\mathbf{n}})$ ) decomposed into a set of independent templates ( $t_\ell^i(\hat{\mathbf{n}})$ ), where

$$t_\ell^i(\hat{\mathbf{n}}) = \sum_{m=-\ell}^{\ell} t_{\ell m}^i Y_{\ell m}(\hat{\mathbf{n}}). \quad (26)$$

Fig. 3 shows this schematically. In other words, (pseudo- $C_\ell$ ) Mode Projection can be considered a special case of Template Subtraction, where the contamination is assumed to be independent of scale *and* the full template map is used to estimate such contamination. It has been pointed out before in the context of 3D clustering estimates that Template Subtraction and Mode Projection can be related if they use equivalent templates [96].

Casting the two methods into this form allows us to make the connection to standard linear regression wherein a measured response  $\mathbf{y}$  is assumed to be a linear combination of predictors given by the  $\boldsymbol{\alpha}$  and a noise term  $\boldsymbol{\epsilon}$ :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}. \quad (27)$$

$\mathbf{X}$  is a  $n \times p$  matrix, where  $p$  is the number of predictors (potentially including a column of ones — the intercept term),  $\boldsymbol{\alpha}$  a vector of length  $p$ , and  $\mathbf{y}$  and  $\boldsymbol{\epsilon}$  vectors of length  $n$ .

Perhaps the most common regression method, Ordinary Least Squares (OLS), finds the vector  $\hat{\boldsymbol{\alpha}}$  that minimizes the squared residuals:

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\|^2 \quad (28)$$

$$= (\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{X}^\dagger \mathbf{y}, \quad (29)$$

where the second expression follows if  $X$  is full column-rank (i.e. the number of observations exceeds the degrees of freedom from the predictors). This is equivalent to the maximum likelihood solution if one assumes the noise of each element,  $\epsilon_i$ , is independent and identically Gaussian distributed,

$$P(\mathbf{y} | \mathbf{X}\boldsymbol{\alpha}) \sim \mathcal{N}(0, I\sigma_\epsilon^2). \quad (30)$$

such that the log-likelihood goes as  $\mathcal{L} \propto \|\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\|^2$ . Even if the assumption of Gaussianity is violated, by the Gauss-Markov theorem Eq. (28) still corresponds to the unbiased estimator with minimum variance if the errors  $\boldsymbol{\epsilon}$  are uncorrelated and have equal variance.

We can write Eq. (27) in terms of the OLS estimates as

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\epsilon}} = \mathbf{X}(\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{X}^\dagger \mathbf{y} + \hat{\boldsymbol{\epsilon}} \quad (31)$$

---

<sup>6</sup> <https://github.com/LSSTDESC/NaMaster>

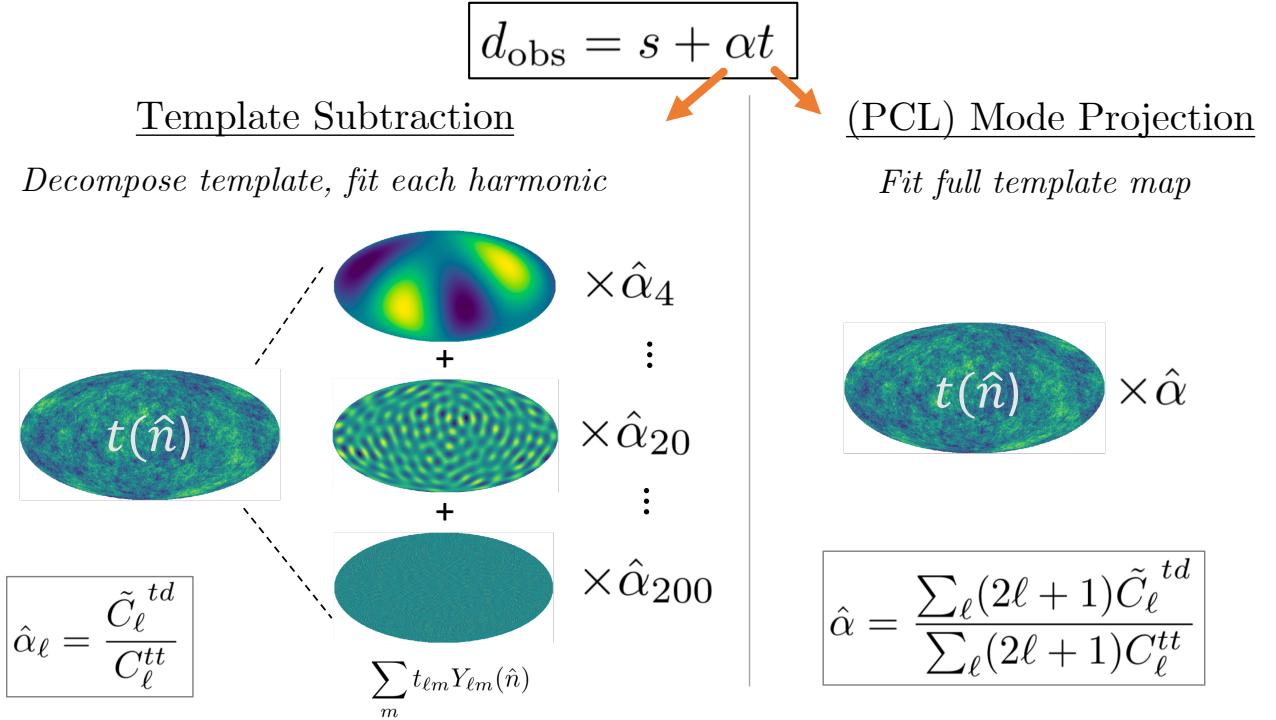


FIG. 3. Schematic illustration of the difference between the Template Subtraction and (pseudo- $C_\ell$ ) Mode Projection methods. Template Subtraction allows templates to have different levels of contamination at each scale. This is analogous to performing Mode Projection, but first decomposing each template map into a series of derived templates, each corresponding to a different harmonic  $\ell$ . See Sec. IV A for details.

where the residuals are defined as

$$\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\alpha}. \quad (32)$$

The quantities of interest in the typical regression problem are the coefficients  $\hat{\alpha}$  or the predicted response  $\hat{\mathbf{y}} = \mathbf{X}\hat{\alpha}$ , with the goal of understanding the influence of predictors or to predict future observations, and hence the residuals are largely used to assess whether the basic OLS assumptions hold. However comparing Eq. (32) to Eqs. (13) and (18), we see that **both Mode Projection and Template Subtraction can be interpreted as OLS regression methods where the observed overdensity signal is regressed onto the templates, and the reconstructed overdensity signal  $\hat{s}$  and power spectrum  $C_\ell^{\hat{s}\hat{s}}$  correspond to the map and power spectrum of the residuals  $\hat{\epsilon}$ .**

Mode Projection uses the full map footprint, with each pixel corresponding to a single observation, for a total of  $N_{\text{tpl}}$  fit coefficients. In contrast, Template Subtraction can be interpreted as performing multiple OLS regressions in parallel on smaller subspaces — one at each multipole in our case — for a total of  $N_\ell \times N_{\text{tpl}}$  fit coefficients (see Fig. 3).

We can write the Template Subtraction amplitudes computed by Eq. (10) in OLS form as

$$\hat{\alpha}_\ell = (\mathbf{T}_\ell^\dagger \mathbf{T}_\ell)^{-1} \mathbf{T}_\ell^\dagger d_\ell, \quad (33)$$

where  $\mathbf{T}_\ell$  is a  $(2\ell+1) \times N_{\text{tpl}}$  matrix, with each column corresponding to a template, consisting of all the harmonic coefficients for a fixed  $\ell$ :

$$\mathbf{T}_\ell = \begin{pmatrix} t_{\ell,-\ell}^1 & t_{\ell,-\ell}^2 & \cdots & t_{\ell,-\ell}^{N_{\text{tpl}}} \\ t_{\ell,-\ell+1}^1 & t_{\ell,-\ell+1}^2 & \cdots & t_{\ell,-\ell+1}^{N_{\text{tpl}}} \\ \vdots & \vdots & \ddots & \vdots \\ t_{\ell,\ell}^1 & t_{\ell,\ell}^2 & \cdots & t_{\ell,\ell}^{N_{\text{tpl}}} \end{pmatrix}. \quad (34)$$

In cases where the multipoles (or angular scales) are binned, the number of fit coefficients is reduced to  $N_{\text{bins}} \times N_{\text{tpl}}$ , which reduces the variance of the contamination estimate. Indeed, Mode Projection corresponds to a limiting case, where the modes of each template are averaged with equal weight before fitting. However in principle one could apply weights differently across scales, and as we will show, this can produce improved coefficient estimates. Alternatively, one could fit individual modes as in Template Subtraction but combine at the coefficient level — potentially useful if certain scales are of particular interest for a given analysis.<sup>7</sup>

<sup>7</sup> In their real-space analysis of SDSS galaxies, Ross *et al.* [88] seem

An immediate consequence of the OLS interpretation of these methods is in making explicit the assumptions that Mode Projection and Template Subtraction are making about the underlying density field — they are exactly the “OLS” assumptions for the error term  $\epsilon$  in the regression model: independent, Gaussian and of equal variance, in whatever basis the map is represented. These assumptions hold well for Template Subtraction, which performs a separate regression at each multipole  $\ell$ . In this case, the assumed OLS “noise” terms are the set of harmonic coefficients of the map ( $s_{\ell m}$ ) at that multipole, which have  $\text{Cov}[s_{\ell m_1}, s_{\ell m_2}] = C_{\ell}^{ss} \delta_{m_1 m_2}$ . For Mode Projection, these assumptions are violated, as the covariance matrix between overdensity pixels is not diagonal,  $\text{Cov}[s(\hat{\mathbf{n}}_i), s(\hat{\mathbf{n}}_j)] \neq \sigma_{\text{sig}}^2 \delta_{ij}$ .

Since the primary contribution to the “noise” of the OLS fit is the clustering signal itself, we can diagonalize it by performing Mode Projection in multipole space, with the maps  $d$ ,  $s$  and  $t_i$  becoming complex column vectors comprised of the map spherical harmonic coefficients. The noise of the observed overdensity  $d_{\ell m}$  is then  $\text{Cov}[s_{\ell_1 m_1}, s_{\ell_2 m_2}] = C_{\ell}^{ss} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}$ . While diagonal, this varies strongly with  $\ell$  and therefore violates the assumption of equal variance, a property known as ‘heteroskedasticity’ in the statistics literature.

However once the noise is diagonal, we can improve the Mode Projection estimate of  $\hat{\alpha}$  by weighting the observed data and template modes by (a prior-inferred)  $1/\sqrt{C_{\ell}^{ss}}$ :

$$\hat{\alpha} = \frac{\sum_{\ell=0}^{\infty} (2\ell+1) \tilde{C}_{\ell}^{td} / C_{\ell}^{ss}}{\sum_{\ell=0}^{\infty} (2\ell+1) \tilde{C}_{\ell}^{tt} / C_{\ell}^{ss}}. \quad (35)$$

This is equivalent to a weighted least-squares approach and recovers the maximum likelihood estimate of  $\hat{\alpha}$ , eschewing the erroneous assumption of a flat signal power spectrum. This of course only works in the ideal full-sky case, but in principle it should not be difficult to extend to a masked sky, e.g. using a predicted cut-sky  $C_{\ell}^{ss}$  computed using the standard coupling matrix from the mask (e.g. [125, 126]) along with the cut-sky harmonics of the templates and datavector, or appropriate binning of modes. This can be viewed as a form of ‘prewhitening’ the data, which accounts for the off-diagonal pixel covariance in the likelihood through an appropriate transform. We explore the potential improvement from such prewhitening in App. B, finding that it improves cleaning, but is subdominant to differences between cleaning methods and higher-order corrections we discuss below.

Finally we note that both the Template Subtraction bias from Elsner *et al.* [120], as well as the pseudo- $C_{\ell}$

Mode Projection bias from Elsner *et al.* [125] result trivially when interpreting them through the OLS lens, in which the variance of observed residuals is well-known to be biased low:

$$\langle \epsilon^{\dagger} \epsilon \rangle = \left( \frac{N_{\text{data}} - p}{N_{\text{data}}} \right) \epsilon^{\dagger} \epsilon. \quad (36)$$

For Template Subtraction, the regression at each harmonic has  $N_{\text{data}} = 2\ell + 1$  and number of predictors  $p = N_{\text{tpl}}$ , leading exactly to the debiasing terms for the signal power estimate in Eqs. (9) and (11). The debiasing terms for Mode Projection in Eq. (22) are more complicated and dependent on the signal and template clustering, but if we take both  $C_{\ell}^{ss}$  and  $C_{\ell}^{tt}$  to be independent of  $\ell$ , Eq. (22) reduces to

$$\langle \hat{s} \hat{s} \rangle = C_{\ell}^{ss} \left( 1 - \frac{1}{\sum_{\ell'=0}^{\ell_{\max}} (2\ell' + 1)} \right), \quad (37)$$

where  $p = N_{\text{tpl}} = 1$  and  $N_{\text{data}} = \sum_{\ell=0}^{\ell_{\max}} (2\ell + 1) = (\ell_{\max} + 1)^2$  is the total number of Fourier modes in the map. This is in keeping with the interpretation of Elsner *et al.* [125], wherein each template removes one degree of freedom from the number of observed Fourier modes. This interpretation can help to assess the risk of overfitting based on the size of the template library.

By making connections between current methods and linear regression explicit, we not only facilitate their interpretation, but can more easily identify the tacit assumptions within these methods, as well as readily improve upon them, drawing on the large body of research into the statistical properties of various regression approaches.

## B. Additive vs. Multiplicative Treatment

One fundamental way in which the methods described here differ is whether or not they assume the systematic contamination is purely additive. This amounts to neglecting a multiplicative term that, if unaddressed, can bias cosmological constraints in upcoming surveys [76]. Here we show how Mode Projection (or any regression method) and be readily adapted to account for the multiplicative contamination and so lead to improved map and power spectrum estimates.

Using the same notation we have used for the additive methods, the general expression for the observed overdensity is

$$d_{\text{obs}} = \gamma(1+s)(1+f_{\text{sys}}) - 1 \quad (38)$$

where again we have suppressed the pixel index. The factor  $\gamma = \bar{N}_{\text{true}}/\langle N_{\text{obs}} \rangle_{\text{pix}}$  accounts for the so-called integral constraint, wherein the mean *observed* number density is used to compute the overdensity field, rather than

---

to implement a version of this. They use Template Subtraction to produce fit coefficients for a large number of scales and templates, but ultimately select one coefficient to apply to all scales for each template. However it is unclear how they compute the single summary coefficient.

the true full-sky mean density (see App. D for a more in depth look at the impact of this monopole term).

To compare with the additive methods, it is convenient to define a zero-centered systematic as:

$$f'_i \equiv \frac{f_i - \bar{f}_i}{1 + \bar{f}_i}, \quad (39)$$

and write Eq. (38) in an equivalent but zero-centered form:

$$d_{\text{obs}} = \gamma'(1+s)(1+f'_{\text{sys}}) - 1, \quad (40)$$

with the new prefactor

$$\gamma' \equiv \gamma(1+\bar{f}_{\text{sys}}) = (1 + \langle s' f'_{\text{sys}} \rangle_{\text{pix}} + s_0)^{-1} \quad (41)$$

ensuring that the monopole in  $d_{\text{obs}}$  is zero, and having the property that  $\langle \gamma' \rangle \approx 1$ .<sup>8</sup> Here  $s_0 \equiv \langle s \rangle_{\text{pix}}$  characterizes the global overdensity in which the footprint resides, and  $s' \equiv s - s_0$  is the deviation from that local overdensity.

Thus the observed overdensity field contaminated with a generic systematic  $f_i$  can be equivalently written as contamination from a *zero-centered* systematic with a rescaled amplitude,  $f'_i$ .

Expanding Eq. (40), we have

$$d_{\text{obs}} = s + \gamma' f'_{\text{sys}} + \gamma' s f'_{\text{sys}} + (\gamma' - 1)(s + 1), \quad (42)$$

Comparing to additive models like Mode Projection and Template Subtraction, which take

$$d_{\text{obs}}^{\text{add}} = s + f'_{\text{sys}} = s + \sum_{i=1}^{N_{\text{tpl}}} f'_i, \quad (43)$$

we see that they not only assume  $\gamma' = 1$  (a vanishing correlation between signal and systematics over the full footprint, as well as no mean local overdensity), but also that  $s f'_{\text{sys}} = 0$  for every pixel, a much stronger assumption. However, despite these assumptions, additive estimates of the total contamination are unbiased, provided the templates  $T$  span the space of the true contamination:

$$\langle \hat{f}_{\text{sys}} \rangle = \langle T \hat{\alpha} \rangle = \langle T(T^\dagger T)T^\dagger d_{\text{obs}} \rangle \quad (44)$$

$$\approx T(T^\dagger T)T^\dagger f'_{\text{sys}} = f'_{\text{sys}}. \quad (45)$$

Intuitively this makes sense, since in the ensemble average the multiplicative term  $s f'$  will vanish.

From Eq. (40), we can then make an improved estimate of the signal map as

$$\hat{s} = \frac{1 + d_{\text{obs}}}{1 + \hat{f}_{\text{sys}}} - 1, \quad (46)$$

$$= \frac{d_{\text{obs}} - \hat{f}_{\text{sys}}}{1 + \hat{f}_{\text{sys}}} \quad (47)$$

where the second form makes clear that this is a simple rescaling of the additive signal estimate,  $d_{\text{obs}} - \hat{f}_{\text{sys}}$ . Therefore, *in a model with multiplicative contamination, signal estimates from additive methods can be improved by weighting the estimated signal map by  $1/(1 + \hat{f}_{\text{sys}})$* .

To explicitly close the loop on the aforementioned methods, the DES-Y1 method performs a series of 1-D regressions and iteratively weights the observed overdensity in a manner equivalent to Eq. (46) for each template, whereas Mode Projection estimates contamination via a single  $N_{\text{tpl}}$ -dimensional regression, with a signal estimate that can be improved via Eq. (47).<sup>9</sup> Applying the multiplicative correction makes Mode Projection equivalent to the Weights model where the coefficients are derived from a simultaneous multiple regression on all the templates (such as in Refs. [97, 103]), but with an additional correction to debias the inferred two-point function. Thus a pixelized weights map for Mode Projection can be produced<sup>10</sup> as

$$w(\hat{\mathbf{n}}) = (\hat{s}(\hat{\mathbf{n}}) + 1)/(d_{\text{obs}}(\hat{\mathbf{n}}) + 1), \quad (49)$$

Fig. 4 illustrates the effect of the multiplicative terms — as well as the impact of neglecting them — on the residuals of a map with a single linear, multiplicative contaminant. The diagonal, dotted line shows the expected relation that would be precisely followed by a purely additive contaminant. A multiplicative contamination adds

<sup>9</sup> As noted in Sec. II even linear contaminants will have interaction terms up to order  $N_{\text{tpl}}$ , such that in principle, for Eq. (44) to fully capture  $f_{\text{sys}}$ , additional templates up to  $t_i t_j t_k \dots t_{N_{\text{tpl}}}$  would need to be included in the template library. A more precise and efficient approach would be to not add any interaction templates, but instead combine the base systematic estimates as

$$\hat{f}_{\text{sys,alt}} = \prod_{i=1}^{N_{\text{tpl}}} (1 + \hat{f}_i) = \prod_{i=1}^{N_{\text{tpl}}} (1 + \hat{\alpha}_i t_i), \quad (48)$$

where recall  $t_i$  corresponds to the  $i^{\text{th}}$  template and  $i^{\text{th}}$  column of  $T$ , and  $\hat{\alpha}_i$  the  $i^{\text{th}}$  element of  $\hat{\alpha}$ . This is closer to the treatment of the DES-Y1 method, wherein weightings for each  $\hat{f}_i$  are applied in series and thus cumulatively.

In practice we find that Eq. (44) is a very good approximation since  $\sigma_{\text{sys}}^2 \lesssim \mathcal{O}(10^{-2})$ , so the nonlinear interaction contributions to  $f_{\text{sys}}$  due to each systematic acting as its own multiplicative screen are fairly negligible, i.e.  $f'_{\text{sys}} \approx \left( \sum_{i=1}^{N_{\text{tpl}}} f'_i \right)$ , as long as the templates sufficiently capture the form of contamination:  $f'_i = \alpha_i t_i$ . Of course this latter condition is a basic requirement of all of the methods we describe here, one that can and should be verified through standard residual plots and other regression diagnostic techniques to ensure an appropriate contamination model for each template (though using methods that incorporate template selection criteria, such as the proposed Elastic Net, can help to satisfy this by allowing a large number of templates to be included in order to address potential higher order terms with little penalty).

<sup>10</sup> This can be released on its own or, as with the DES-Y1 data release, as an additional column at the catalogue level. c.f. <https://des.ncsa.illinois.edu/releases/y1a1/key-catalogs/key-redmagic>.

<sup>8</sup> Here the approximation stems from making the assumption  $\langle x^{-1} \rangle \approx \langle x \rangle^{-1}$ , which holds very well for the cases we are studying where the mean is taken over a footprint with  $N_{\text{pix}} \gtrsim 10^5$  and shot noise is subdominant.

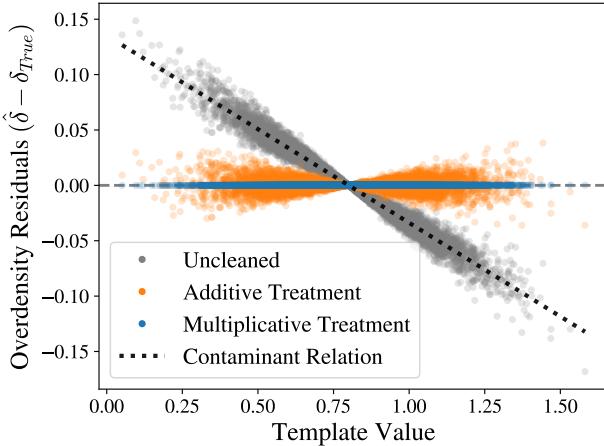


FIG. 4. The error in estimates of the overdensity  $\delta$  in a toy Gaussian map when contaminated with a single template. Gray points indicate the pixel-based difference between the observed, uncleaned overdensity and the true overdensity when the contamination is multiplicative (additive contamination would lie directly along the dotted line). Orange points are the result when erroneously assuming the contamination is only additive. Blue points are the result when correctly treating the multiplicative component.

significant scatter around this relation, shown as the gray points. This scatter remains when the contamination is cleaned with an additive method (orange), but is effectively removed when the multiplicative component is taken into account (blue). Fig. 5 shows how errors on the estimated overdensity field are drastically reduced when applying the multiplicative correction of Eq. (47) to a realistic use case with multiple contaminating systematics (see Sec. VI for details of implementation).

### C. Multiplicative Effect on Likelihood

While the multiplicative term vanishes in the ensemble average, resulting in the same ensemble pixel mean as the additive-only prediction ( $\langle d_{\text{obs}} \rangle = f'_{\text{sys}}$ ), the pixel variance is modulated:

$$\begin{aligned} \text{Var}[d_{\text{obs}_i}] &\approx \langle [s\gamma'(1 + f'_{\text{sys}})]^2 \rangle \\ &\approx \langle s^2 \rangle (1 + f'_{\text{sys}})^2 \\ &= \sigma_{\text{sig}}^2 (1 + f'_{\text{sys}})^2, \end{aligned} \quad (50)$$

where for large  $N_{\text{pix}} \gtrsim 10^5$ ,  $\langle \gamma' \rangle \approx \langle \gamma'^2 \rangle \approx 1$ . The corresponding covariance between pixels is

$$\text{Cov}(d_{\text{obs}}, d_{\text{obs}_j}) \approx \left\langle (\gamma')^2 [s_i(1 + f'_{\text{sys}_i})] [s_j(1 + f'_{\text{sys}_j})] \right\rangle. \quad (51)$$

This is the source of the systematic-dependent scatter in Fig. 4, which will result in biased two-point statistics

from additive methods. Because the contamination estimate is unbiased, the correction of Eq. (47) almost fully suppresses this variance, but the multiplicative terms also impact the likelihood when performing the regression. In pixel-space a simple fix would be to iterate: use an initial estimate of  $\langle \hat{f}_{\text{sys}} \rangle$  with Eq. (50) to apply inverse variance weights to the maps before making a second estimate of  $\langle \hat{f}_{\text{sys}} \rangle$ . In practice these are “errors on the errors” and so the impacts will be subdominant to the multiplicative correction to the datavector itself.

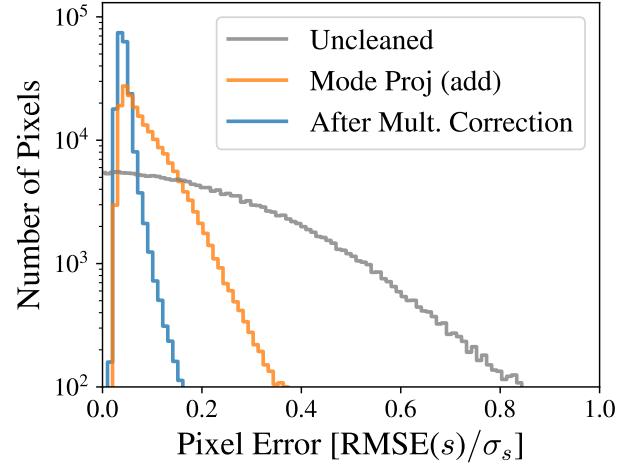


FIG. 5. Distribution of pixel errors before cleaning (gray), after cleaning with Mode Projection but *before* multiplicative correction (orange), and *after* multiplicative correction (blue). The errors have been calculated as the RMSE of each pixel across 100 cleaned mocks in our fiducial configuration of a DES-like survey as described in Sec. VI, and have been normalized to the expected dispersion from the true overdensity field.

## V. APPLICATIONS

We can use the insights of the previous sections to propose two additional methods, as well as to estimate the errors on the cleaned map. We now describe these in turn.

### A. Iterative Forward Selection

We include an iterative Forward Selection method that incorporates some of the main features of the DES-Y1 method, but adopts some of the simplifying assumptions of Mode Projection. The result is greatly simplified and easier to implement than the full DES-Y1 method.

We keep the core of the template selection algorithm, but modify the fit procedure and significance criterion to eliminate the need to generate mocks. We do this by

adopting the same implicit assumptions of Mode Projection: that pixels are uncorrelated and have equal variance. This allows for an analytical solution for the best-fit parameters  $\theta = \{m_i, b_i\}$  and their covariance  $\text{Cov}_\theta$  for each template, which we obtain using `numpy.polyfit`<sup>11</sup>. We then adopt a simplified significance criterion of  $\Delta\chi^2_{\text{FS}} = \theta^T[\text{Cov}_\theta]^{-1}\theta$ , and use the same stopping threshold as the DES-Y1 method.<sup>12</sup>

This Iterative Forward Selection method is a fast and simple method that incorporates some of the key aspects of the DES-Y1 method, the iterative weighting and template selection, while avoiding the most computationally expensive parts, the generation of mocks. We expect some loss of precision by not including a covariance matrix in the fitting step, but on the other hand to gain some precision by not having to bin pixels, so this method can help to benchmark the importance of including the covariance matrix in a DES-Y1-like method.

## B. Elastic Net

We also propose a method that closely mimics Mode Projection but incorporates template selection, thereby reducing the impact of overfitting when the template library is large.<sup>13</sup> Having shown that Mode Projection is equivalent to linear regression, we adopt a regression method specifically designed to automatically select predictors based on the data.

This selection is accomplished by modifying the *Loss* function that is optimized when fitting, which is equivalent to applying a prior to the template coefficients and finding their maximum a-posteriori (MAP) estimate. Specifically, instead of finding  $\hat{\alpha}$  that minimizes the square of the residuals ( $\|d_{\text{obs}} - T\alpha\|^2$ ), we instead minimize

$$\text{Loss} = \frac{1}{2N_{\text{pix}}} \|d_{\text{obs}} - T\alpha\|_2^2 + \lambda_1 \|\alpha\|_1 + \frac{\lambda_2}{2} \|\alpha\|_2^2, \quad (52)$$

where

$$\|\alpha\|_1 = \sum_i^{N_{\text{tpl}}} |\alpha_i| \quad (53)$$

<sup>11</sup> To estimate  $\text{Cov}_\theta$ , `numpy.polyfit` assumes a diagonal Gaussian covariance of the pixels, scaled so that the best-fit model has reduced  $\chi^2$  of  $\chi^2_{\text{red}} = \chi^2/(N_{\text{pix}} - 2) = 1$

<sup>12</sup> As with the DES-Y1 method, this method can suffer from a lack of convergence when the threshold is low, where chance correlations between the signal realization and templates result in a loop of the same series of templates being repeatedly reweighted. We adopted a limit of  $10 \times N_{\text{tpl}}$  reweightings for each signal realization before breaking the loop and using the resulting signal estimate as is. This occurred occasionally and at very low thresholds, with no discernible effect on the estimated maps or power spectra.

<sup>13</sup> See Leistedt and Peiris [90] for an alternative approach that pre-selects templates for projection using a  $\chi^2$  threshold.

is the L1-norm of  $\alpha$ , and

$$\|\alpha\|_2 = \left( \sum_i^{N_{\text{tpl}}} |\alpha_i^\dagger \alpha_i| \right)^{\frac{1}{2}} \quad (54)$$

is the usual vector L2-norm of  $\alpha$ . Here  $\lambda_1$  and  $\lambda_2$  are hyperparameters that are tuned from the data, which we now discuss in turn:

1. The L1-norm term incentivizes sparsity in  $\alpha$  by penalizing non-zero coefficients of templates, thus naturally performing template *selection*. This is useful because the number of templates in modern surveys can be enormous — e.g. Leistedt and Peiris [90] produce  $\sim 3700$  templates for their analysis of SDSS quasars — and so it is common to pre-select only a handful to use, for fear of removing true signal. Since we don't know *a priori* which templates are contaminating, the incorporation of an automated selection scheme enables a more agnostic, data-centric approach to cleaning a large library of templates, while mitigating the risk of overfitting. The use of this penalty term in isolation (i.e. setting  $\lambda_2 = 0$ ) is often called the Least Absolute Shrinkage and Selection Operator (LASSO)[127], and has a Bayesian interpretation of applying a zero-centered Laplace prior on the elements of  $\alpha$ , with a width  $\propto 1/\lambda_1$  (see e.g. [128] for a discussion). L1 priors to induce sparsity have been used in a variety of astrophysical problems, such as for source separation in cosmic microwave background analyses [129–131] or in reconstructing mass maps from weak lensing data [132–134].
2. The L2-norm term helps address collinearity (i.e. correlation) between template maps which, when present, can cause the matrix  $T^\dagger T$  to be ill-conditioned and the variance of contamination estimates to be large. When it is the only additional penalty term (i.e.  $\lambda_1 = 0$ ), this is often called Ridge Regression, or Tikhonov Regularization. It is straightforward to show that, from a Bayesian perspective, this method is equivalent to placing a zero-centered Gaussian prior on the elements of  $\alpha$ , with a width  $\propto 1/\lambda_2$ .

Since each penalty term addresses a different issue with standard regression, it is not uncommon to combine them, as proposed by Zou and Hastie [135], in a method known as the “Elastic Net”. We use the `scikit-learn`[136] implementation, `ElasticNetCV`, with a hyperparameter space of  $\lambda_1/(\lambda_1 + \lambda_2) \in \{0.1, 0.5, 0.9\}$  and 100 values of  $(\lambda_1 + \lambda_2)$  spanning three orders of magnitude, which are automatically determined from the input data (the default setting). We use 5-fold cross-validation to determine the best  $\lambda_1$  and  $\lambda_2$ , trained on a random selection of 30% of the input map pixels.

In this 5-fold cross-validation scheme, the training sample (30% of the map) is itself partitioned into five

equal subsamples. For each combination of hyperparameters, one subsample is withheld for validation, while the other four are used to train the model by minimizing Eq. (52). The mean squared error (MSE) of the validation sample is then computed and stored (i.e. the first term in Eq. (52)). One of the four training subsamples is then withheld as the new validation set, and the process is repeated until each of the five subsamples has been used exactly once for validation, with their average MSE used to characterize the goodness-of-fit for the given set of hyperparameters  $\lambda_1$  and  $\lambda_2$ .

Setting  $\lambda_1 = \lambda_2 = 0$  reduces to OLS regression and hence to the pseudo- $C_\ell$  Mode Projection method, while sampling extreme values for the relative weight of the L1 vs. L2 penalty allows for the effective use of only one of the penalty terms, if preferred by the data. The use of cross-validation on a subset of the map allows the data to dictate which model is most appropriate, with minimal risk of overfitting. We illustrate the utility of this in Fig. 6, which shows how the cross-validation scheme naturally increases the L1 penalty when fitting for more (*uncontaminating*) templates. We found that the L2 penalty became increasingly important when the correlation between templates increased beyond  $\rho_{\text{tpl}} \gtrsim 0.9$ .

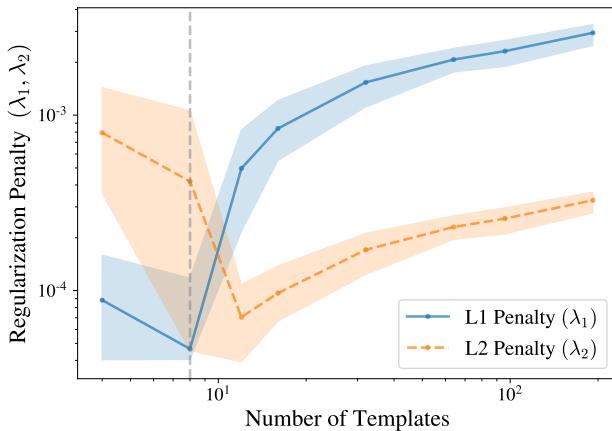


FIG. 6. Best-fit L1 and L2 penalty coefficients in the regularization technique described in Sec. V B, as a function of the number of templates used for cleaning,  $N_{\text{tpl}}$  (new signal and template maps are generated at each value of  $N_{\text{tpl}}$ ). In all cases, 12 templates are contaminating the observed data (vertical dashed line). The importance of the L1 penalty, facilitating template selection, becomes increasingly important as more templates are included for cleaning. Lines and shaded region indicate the median and central 68% probability mass of 50 mocks at each  $N_{\text{tpl}}$  for the central bin of our fiducial DES-like survey. Here,  $\rho_{\text{tpl}} = 0.2$  within template groups, though plots for other  $\rho_{\text{tpl}}$  look similar. See Sec. VI for details of implementation.

### C. Map errors

We can use the regression framework to gain insight into how errors in the estimated overdensity map are distributed across pixels. This aids the propagation of map errors in cross-correlation studies and summary statistics beyond the two-point functions, as well as helps to identify regions that may benefit from masking.

For simplicity, we assume additive contamination and correction and ignore higher-order terms:

$$d_{\text{add}} = s + f_{\text{sys}} = s + T\alpha \quad (55)$$

The estimated contamination amplitude is then

$$\hat{\alpha}_{\text{mp}} = (T^\dagger T)^{-1} T^\dagger d_{\text{add}} \quad (56)$$

$$= \alpha + (T^\dagger T)^{-1} T^\dagger s \quad (57)$$

such that our signal estimate is

$$\hat{s}_{\text{mp}} = d_{\text{add}} - T\hat{\alpha}_{\text{mp}} \quad (58)$$

$$= s - T(T^\dagger T)^{-1} T^\dagger s \quad (59)$$

$$\equiv (I - H)s \quad (60)$$

where the matrix  $H \equiv T(T^\dagger T)^{-1} T^\dagger$  is often called the ‘Hat’ or ‘Projection’ matrix in the statistics literature. Then

$$\text{Var}[(\hat{s}_{\text{mp}} - s)_i] = \text{Var}[(Hs)_i] = [H\text{Var}[s]H^\dagger]_{ii} \quad (61)$$

If we make the assumption that the signal covariance is diagonal, then  $\text{Var}[s] \approx \sigma_{\text{sig}}^2 I$  and

$$\text{Var}[(\hat{s}_{\text{mp}} - s)_i] \approx \sigma_{\text{sig}}^2 (HH^\dagger)_{ii} = \sigma_{\text{sig}}^2 H_{ii} \quad (62)$$

where we have used the fact that  $H$  is both Hermitian and idempotent so that  $HH^\dagger = HH = H$ .

Despite a number of simplifying assumptions and the fact that some of the methods only fit for some of the templates, we find that with the exception of Template Subtraction,  $H_{ii}$  is a remarkably good predictor<sup>14</sup> of how the errors in the overdensity estimates are distributed for all the methods. The errors arise from removing real signal during the cleaning process, with  $H_{ii}$  as a measure of how susceptible pixel  $i$  is to such overcorrection. This also indicates that while to first order all correlation with templates is removed from the estimated overdensity field, the templates remain imprinted on the map through their absence; there is missing signal in precisely their spatial configuration.

Intuitively,  $H_{ii}$  as a distance measure of pixel  $i$  from the center of mass of other pixels in the  $N_{\text{tpl}}$ -dimensional

<sup>14</sup> Note that  $H_{ii}$  only requires the diagonal elements of  $H$ , which are far more tractable to calculate than the full  $N_{\text{pix}} \times N_{\text{pix}}$  matrix.

space spanned by the templates. This is sometimes referred to as ‘leverage’, as pixels with higher  $H_{ii}$  have larger impact when performing a regression.<sup>15</sup> This can be seen by observing that the estimated systematic field can be written as

$$\hat{f}_{\text{sys}} = H d_{\text{obs}} \quad (63)$$

such that the leverage

$$H_{ii} = \frac{\partial \hat{f}_{\text{sys}}^{(i)}}{\partial d_{\text{obs}}^{(i)}} \quad (64)$$

encodes the sensitivity of the contamination estimate to an observed over- or underdensity at pixel  $i$ . Because pixels with high leverage can have an outsized effect on the estimated contamination, we expect leverage to be a useful tool for identifying potentially problematic pixels that should be masked before cleaning, in addition to providing error estimates for those pixels that remain.

It is straightforward to derive the mean leverage value as

$$\bar{H}_{ii} \leq N_{\text{tpl}} / N_{\text{pix}}, \quad (65)$$

with the equality holding if  $T$  is full rank, since

$$\sum_i^{N_{\text{pix}}} H_{ii} = \text{Tr}(H) \leq N_{\text{tpl}}, \quad (66)$$

providing a basis on which to determine extreme leverage values.

The main panel of Fig. 7 shows the RMS error (RMSE =  $\sqrt{\langle (\hat{s} - s)^2 \rangle}$ ) of each pixel computed over 100 cleaned DES-like mock maps plotted against leverage  $H_{ii}$  from 27 cleaning templates, 11 of which are contaminating. Pixels are grouped into 1000 bins of 197 pixels, according to their leverage value, and we show the mean and standard error of the RMSE for each bin. We see that pixels with low leverage value have much smaller error in the estimated overdensity map, and that the error goes roughly as  $\propto H_{ii}^{1/2}$  (diagonal dashed line), as predicted by Eq. (62). The Template Subtraction method is an exception to this trend likely because the regression happens in a different space, at each harmonic separately, and so does not relate cleanly to the pixel leverage.<sup>16</sup>

The top panel of Fig. 7 shows the fraction of map pixels below a given leverage (note the log scale), with the vertical dotted line indicating  $3 \times \bar{H}_{ii}$ , which is one of

<sup>15</sup> This phenomenon is very familiar from the simple case of fitting a 1D line to a scatter of 2D points  $\{x, y\}$ , where the best-fit line is ‘pulled’ preferentially to points that lie farther from  $\bar{x}$ .

<sup>16</sup> In principle, one could construct the analogous leverage quantity  $H_{\ell m} = t_{\ell m} [\mathbf{C}_\ell^T \mathbf{T}]^{-1} t_{\ell m}^\dagger$  in harmonic space for the analysis of errors in  $\hat{s}_{\ell m}$ , which may be useful for cross correlation analyses in harmonic space.

two common thresholds used in statistics to flag points that may bias a regression analysis ( $2 \times \bar{H}_{ii}$  being the other). Here, 0.5% of map pixels exceed  $3 \times \bar{H}_{ii}$ ; these pixels potentially merit further inspection or masking, as they are particularly prone to biasing the regression. The trend of the uncleansed data may be surprising, but as noted in Sec. IV B, because of the integral constraint,  $d_{\text{obs}}$  is insensitive to a monopole in  $f_{\text{sys}}$  and so as long as templates approximately trace the true contamination, overdensities near the mean of the templates (i.e. low  $H_{ii}$ ) will be most accurately measured, even if contamination is greater than at other points in the map (see App. D).

A complementary statistic is the ‘Cook’s distance’[137, 138] for each pixel, which uses  $H_{ii}$  and  $\hat{s}_i$  to provide a measure of the total change in the  $\hat{s}$  map if pixel  $i$  were to be masked (assuming additive contamination and correction). Along with the leverage, we expect this to be a useful tool when performing template-based mitigation of spatial systematics and for mask creation. We leave further investigation of these as diagnostic tools, as well as generalization to the multiplicative case, to a later work.

We next describe the fiducial survey on which we test the performance of foreground-cleaning methods.

## VI. EVALUATING PERFORMANCE

Our analysis is fully synthetic, with the procedure depicted in Fig. 1. We compare the cleaning methods described, including results for both the standard additive Mode Projection case (denoted ‘Mode Projection (add.)’) as well as one with the multiplicative correction from Eq. (47) (denoted simply ‘Mode Projection’). For the Elastic Net, we only show results that include the multiplicative correction.

We only consider full-sky maps in this paper. Extension to partial-sky surveys should be fairly straightforward, requiring the usual correction of cut-sky power spectra, but this applies equally across the full-sky spectra estimated with each method here and so we do not expect it to qualitatively change the main results.

### A. Templates

We first describe the fiducial set of templates that we use, for both contamination and cleaning purposes. We adopt several classes of templates in order to span a range of possible contaminants and their spectral behavior. In most cases, we use multiple templates of the same class by generating Gaussian realizations of maps from the same theoretical power spectrum. The classes of template we use are

- $C_\ell \propto (\ell + 1)^0$  (white noise)
- $C_\ell \propto (\ell + 1)^{-1}$

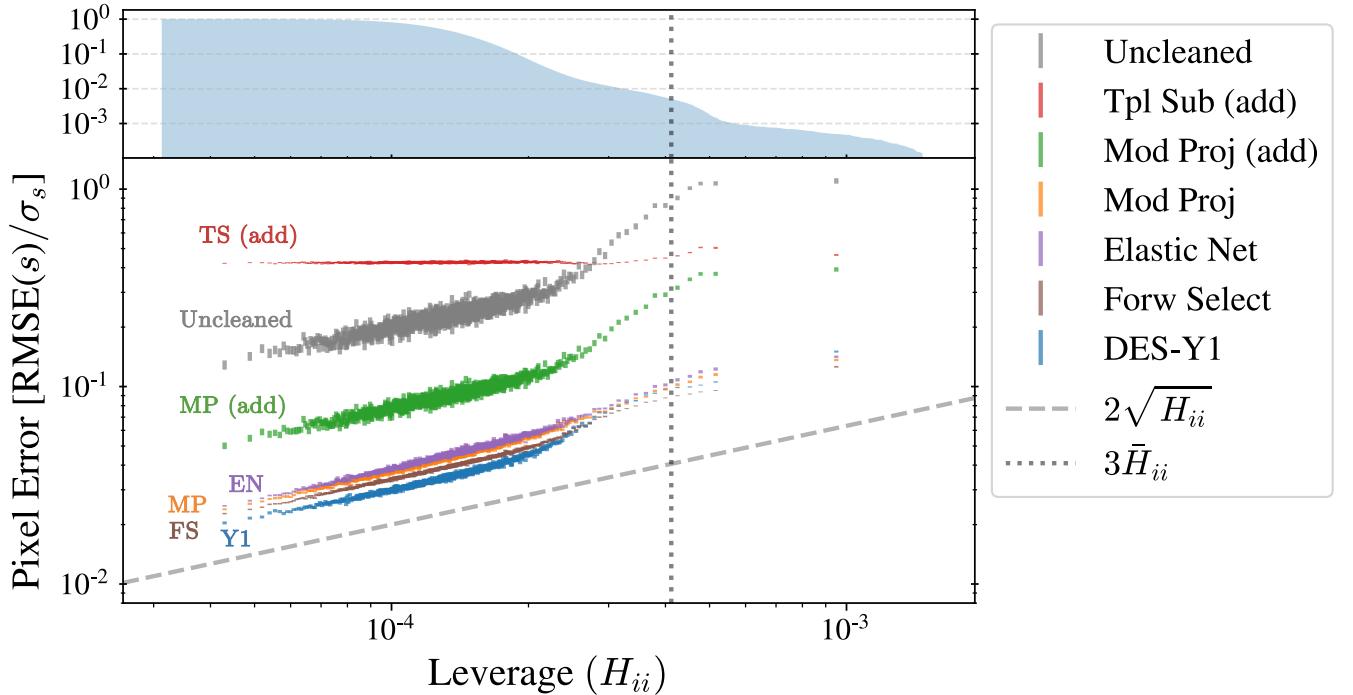


FIG. 7. Root-mean-square error of pixel overdensity estimates, normalized to expected dispersion from the true overdensity due to cosmic variance, vs. pixel leverage for 100 signal realizations. The vertical axis shows the standard error across pixels in 1000 equal-sized bins (in this case 197 pixels per bin at  $\text{Nside}=128$ ). The error in both observed and estimated overdensity scales as roughly  $\propto H_{ii}^{1/2}$  for all methods (dashed line, to guide the eye). The dotted vertical line indicates a commonly used threshold of  $3 \times$  the mean leverage across pixels to identify pixels that may have an undue impact on regression fit parameters. The histogram in the top panel indicates the number of pixels at a given leverage.  $N_{\text{tpl}} = 27$ ,  $N_{\text{sys}} = 11$ ,  $\sigma_{\text{sys}}^2 = 0.01$ .

- $C_\ell \propto (\ell + 1)^{-2}$
- $C_\ell \propto \exp[-(\ell/10)^2]$
- ‘a ‘Cat-scratch’ map, with 128 horizontal stripes to model a basic scanning pattern and/or differences in depth due to overlapping tiles
- a 2D Gaussian “spot” map
- a E(B-V) extinction map, with dependence on latitude removed.

The last three correspond to static maps which do not change throughout the analysis. We use the full-sky E(B-V) map<sup>17</sup> from Planck [139], but since this is dominated by emission near the galactic plane, which LSS surveys typically avoid, we reweight the map to remove its major latitudinal dependence.

We normalize the individual templates to the same overall variance, and construct a total systematic map

as a product of some or all of the individual template maps:

$$1 + f_{\text{sys}} = \prod_{i=1}^{N_{\text{sys}}} (1 + \alpha_i t_i) \quad (67)$$

Note that this model can generally encompass contamination to any polynomial order simply by including templates that are products of others (e.g.  $t_{\text{new}} \equiv t_i^2$ ), and incrementing  $N_{\text{sys}}$  accordingly. Similarly, nonlinear contamination can often be made linear through an appropriate transformation of the template map.<sup>18</sup> This total systematic map is then scaled to a desired overall map variance  $\sigma_{\text{sys}}^2$ , thus determining the overall contamination field  $f_{\text{sys}}$ . We use a fiducial level of contamination of  $\sigma_{\text{sys}}^2 = 0.01$ , as we found this to produce fluctuations similar to those seen in the DES-Y1 data [41]; this corresponds to an RMS error on  $\delta$  of  $\sim 10\%$ . Changing the level of contamination  $\sigma_{\text{sys}}^2$  did not significantly alter our results.

<sup>17</sup> [https://wiki.cosmos.esa.int/planckpla/index.php/CMB\\_and\\_astrophysical\\_component\\_maps#The\\_.5Bmath.5DE.28B-V.29.5B.2Fmath.5D\\_map\\_for\\_extra-galactic\\_studies](https://wiki.cosmos.esa.int/planckpla/index.php/CMB_and_astrophysical_component_maps#The_.5Bmath.5DE.28B-V.29.5B.2Fmath.5D_map_for_extra-galactic_studies)

<sup>18</sup> E.g. Elvin-Poole *et al.* [41] fit linear models to the square root of exposure time and sky brightness, based on how they contribute to the depth map.

We perform the contamination and cleaning procedure shown in Fig. 1 on each redshift bin and for each cleaning method over many sky realizations, and plot the mean and central 68% probability mass of the relevant quality statistic. We use the same set of templates and total systematic map for across redshift bins and sky realizations, but generate a new set for each unique combination of parameter choices (e.g. level of cross-correlation between templates, number of templates used, etc.) in order to minimize any effects from specific template realizations.

We use `CLASS` [140] to compute theoretical galaxy clustering power spectra for a mock LSS survey, including contributions from redshift-space and Doppler distortions and lensing. We found gravitational potential terms to contribute  $\lesssim 1\%$  to the resultant  $C_\ell$  for  $\ell > 7$  but increased computation time by an order of magnitude, so we neglect them. Since we find the cleaning procedures are not strongly sensitive to the signal power spectrum, this should not impact our results. We then use `Healpy`[141] to generate full-sky Gaussian realizations of large-scale structure overdensity ( $\delta \equiv \delta\rho/\rho$ ) maps for each redshift bin with `NSIDE` = 128. We compare the impact of using lognormal maps in App. A, finding it does not change our results.

## B. Cosmological model and simulated survey

We assume a standard  $\Lambda$ CDM cosmological model with one species of massive neutrino and parameter values from best-fit Planck 2018:  $\{\Omega_c, \Omega_b, h, n_s, \sigma_8, m_\nu/\text{eV}\} = \{0.26499, 0.04938, 0.6732, 0.96605, 0.8120, 0.06\}$ . Given the precise parameter constraints from current probes, the dependence of our results on cosmological parameters is expected to be very minimal. In contrast, the choice of the parameter set to be *determined* from the survey may be highly dependent on the residual systematics.

In general for comparing the methods, the exact form of the galaxy power spectra is not very consequential, so we use a fiducial survey comparable to the completed Y5 Dark Energy Survey, for which a realistic level of contamination can be estimated based on existing data. We assume the number density distribution of galaxies to be in the form

$$\frac{dn}{dz} \propto \left(\frac{z}{z_0}\right)^\alpha \exp[-(z/z_0)^\beta], \quad (68)$$

where  $z_0 = 0.55$ ,  $\alpha = 2.65$ , and  $\beta = 3.34$ . We assume five redshift bins centered at redshifts  $\{0.225, 0.375, 0.525, 0.675, 0.825\}$ , with galaxy bias of  $\{1.4, 1.6, 1.6, 1.95, 2\}$ , respectively, and containing galaxies with Gaussian redshift dispersion of  $\sigma_z = 0.05$ . These values were chosen to closely approximate the `REDMAGIC` redshift distribution given in Elvin-Poole *et al.* [41].

We choose to work primarily in harmonic space. Therefore, starting with some map with overdensity  $\delta \equiv \delta N/N$ , where  $N$  is the galaxy count over some patch, the

expansion in spherical harmonics gives

$$\delta(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}), \quad (69)$$

and the angular power spectrum is given by

$$C_\ell = \sum_{m=-\ell}^{\ell} \frac{|a_{\ell m}|^2}{2\ell + 1}. \quad (70)$$

Because we are working in the full-sky limit, all well-known estimators of power return the same result, so here we make use of the `anafast` and `alm2c1` functions in `Healpy`. To more accurately account for the cosmological impact of the cleaning methods on data from a DES Y5-like survey, we divide the assumed sample variance  $\sigma_{C_\ell}^2$  by a factor of  $f_{\text{sky}} = 0.116$ .

We add shot noise to the theoretical power spectrum as  $C_\ell \rightarrow C_\ell + \bar{n}^{-1}$ , with  $\bar{n} = 1.5 \times 10^8$ , but this is negligible at the large scales we work with ( $\ell \leq 350$ ). We are primarily interested in studying the systematic impacts of cleaning (or not) using spatial templates, so it is reasonable to focus on cases where the signal-to-noise is large (i.e. shot noise is negligible).<sup>19</sup>

## VII. SIMULATION RESULTS

To compare methods, we compare the fidelity of the cleaned data products to the truth, either at the map level or at the level of the power spectrum, rather than look for cosmological-parameter biases. We do this for a few reasons: (1) the map and power spectrum are more general, being independent of (but easily mapped to) any specific cosmological model one wants to test, or summary statistic one wants to use; (2) while we primarily study applications to galaxy clustering data here, the methods themselves are quite general and can easily be applied to other data sets for which one has tracers of potential contamination, such as shear or convergence maps; (3) galaxy clustering *alone* leads to relatively weak cosmological constraints and is rarely used on its own to constrain cosmology.

We therefore limit ourselves to investigating biases in data space and leave the investigation of impacts on cosmological constraints to a later work when weak lensing data can be incorporated in a more realistic fashion. At this stage, the test bed is sufficiently representative to compare foreground-cleaning methods in a manner to inform future LSS analyses.

---

<sup>19</sup> Shot noise may have the effect of (1) rendering the regression residuals more diagonal in pixel-space (or flattening them in harmonic space), which could actually improve the regression procedure, and/or (2) introduce significant skewness in the distribution. We would expect the impacts of these to be similar to those of prewhitening the data or using lognormal mocks, and so based on our results in Apps. A and B, we do not expect shot noise to significantly impact on our findings.

### A. Characterizing Performance

We first study the impacts of the different methods on the estimated maps and power spectra for a single configuration and compare the residual biases of each. For this fiducial comparison, we generate 50 mocks for each redshift bin and contaminate them with 11 systematics, two from each of the four Gaussian classes, plus the three static templates. We construct a template library that contains the contaminating templates, plus four additional realizations from each Gaussian class, for a total of 27 cleaning templates. Each method uses this library to produce estimates of the overdensity field and power spectra.

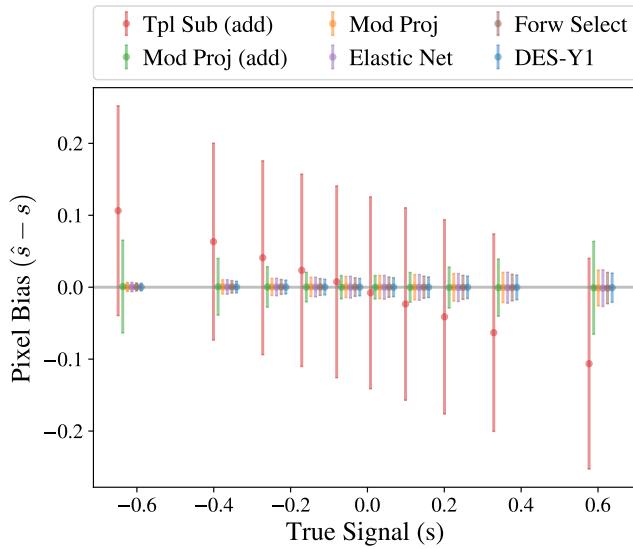


FIG. 8. Error in overdensity estimates for different cleaning methods, binned in deciles of the true overdensity and with points offset for clarity. Error bars show the standard deviation in each bin. Overcorrection at the map-level is only significant for Template Subtraction, which under-estimates the magnitude of both peaks and voids, while other methods are very close to unbiased. See text for details.

We show map residuals of each cleaning method for the lowest redshift bin in Fig. 8, where the residuals are binned into deciles of the true overdensity. Results for other redshift bins are similar. From left to right, in approximate order of performance, the figure shows the Template Subtraction method (red), Mode Projection without (green) and with (orange) multiplicative correction, the Elastic Net method (purple), Forward Selection (brown) and the DES-Y1 method (blue).

The overcorrection of Template Subtraction is evident, with density fluctuations consistently *under*-estimated (i.e. peaks and voids are both less extreme than they should be). The other methods are all very close to unbiased with respect to the true overdensity field, with bias of the mean  $\lesssim 0.001$  for each bin. The multiplicative methods show significantly reduced within-bin scat-

ter (i.e. smaller error bars) compared to the additive ones — the additive Template Subtraction and Mode Projection methods (leftmost, red and green) have typical errors in the overdensity of  $\sigma_s \sim 0.1$  and  $\sigma_s \sim 0.01 - 0.05$ , respectively, compared to the errors of  $\sigma_s \sim 0.005 - 0.02$  for the multiplicative methods. This suggests that applying the multiplicative correction results in significantly improved map estimates, making them excellent candidates for map-based analyses, such as as counts-in-cells or density-split statistics.

While the signal estimates are unbiased (with the exception of Template subtraction), the errors of the additive methods increase near extremes of the density field. This is similar to the result in Fig. 5, which showed larger errors at extreme template values, in part for the same reasons. Both Figs. 7 and 8 indicate a clear stratification of the methods, with the methods that fail to treat the multiplicative component of contamination showing significantly larger error.

We also compare the maps in harmonic space. The left panel of Fig. 9 shows the per-multipole performance of the cleaning algorithms as  $(1 - C_\ell^{\hat{s}\hat{s}} / C_\ell^{ss})$  vs. the multipole  $\ell$ , where

$$C_\ell^{s\hat{s}} = \langle s_{\ell m} \hat{s}_{\ell m}^* \rangle \quad (71)$$

This quantifies the fractional *missing* cross-power between the the true and estimated maps, such that a perfect reconstruction corresponds to 0, and pure noise corresponds to 1 (note the log scale). This conveys the approximate level of error expected when using cleaned maps for cross-correlation studies.

All of the methods that treat the multiplicative contamination perform significantly better than the additive methods. The corrected Mode Projection and Elastic Net, and the DES-Y1 method all have excellent performance at  $\ell \gtrsim 30$  or scales below about 0.2 degrees on the sky, showing  $\lesssim 0.1\%$  error. Maps cleaned with these methods should therefore be excellent candidates for cross-correlation studies. Even the additive mode projection method performs quite well with error of  $\lesssim 1\%$  in this case, and as such it may be adequate for many studies.

In the right panel of Fig. 9 we show the error in the power spectrum estimate as the difference between the estimated (after cleaning) and true angular power in bins of  $\Delta\ell = 10$  and normalized to sample variance  $(\hat{C}_\ell^{\text{est}} - \hat{C}_\ell^{ss})/\sigma_{C_\ell}$ , where for  $\sigma_{C_\ell}$  we use the standard Gaussian approximation for cosmic variance, scaled by  $1/f_{\text{sky}}$ . Unlike  $C_\ell^{s\hat{s}}$ , this quantity is insensitive to phase-differences between the true and reconstructed maps of the map.

To lowest order, all of the methods work well, and the residual biases are below cosmic variance for the large angles studied here (note that systematic shifts will become more significant with larger multipole bins). For mode projection, the performance is satisfactory only once it is corrected for the multiplicative bias via Eq. (46). We do not show Template Subtraction on the right for clarity —

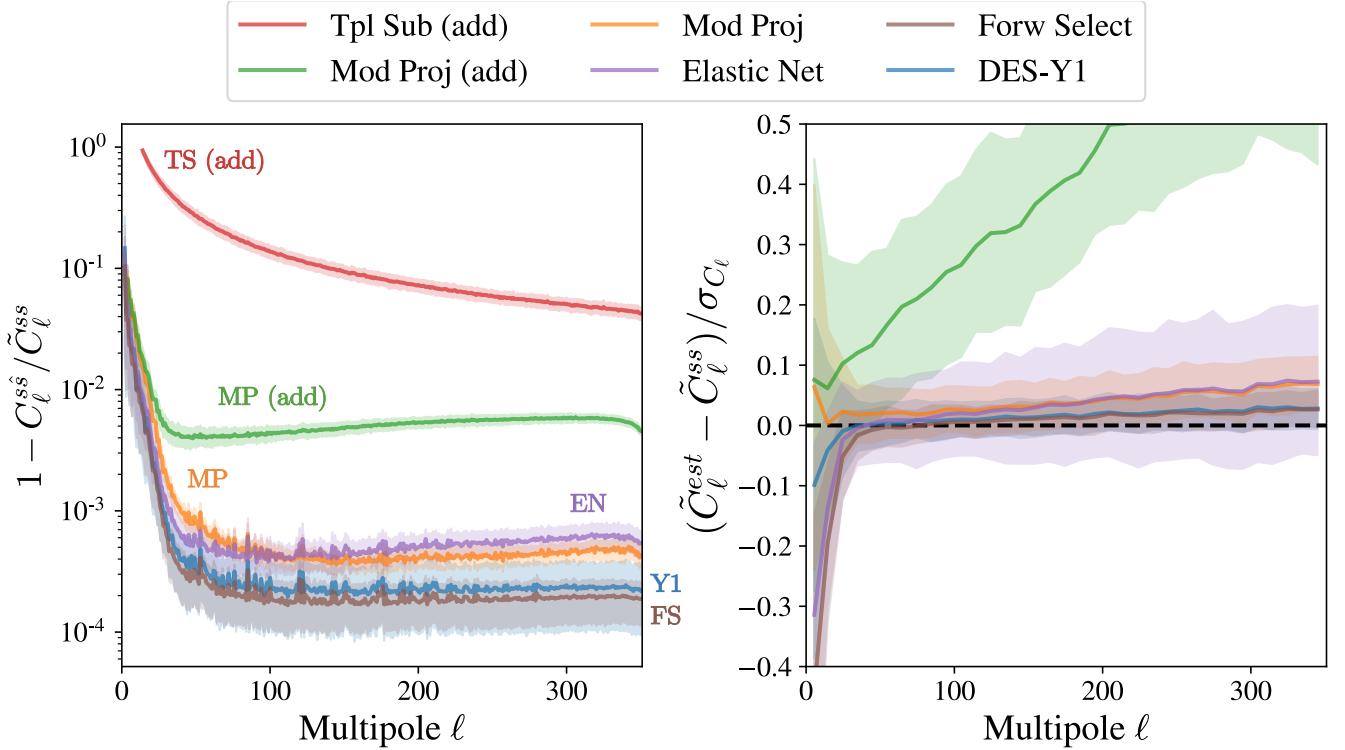


FIG. 9. *Left:* Error in map reconstruction for each method as a function of multipole  $\ell$  in a DES-like survey, shown as the deficit in correlation at each multipole between the true and cleaned maps ( $1 - C_\ell^{ss} / \tilde{C}_\ell^{ss}$ ). A perfect reconstruction corresponds to 0, whereas pure noise corresponds to 1. For all methods (except perhaps Template Subtraction), the cleaned map is a good approximation of the true map for cross-correlation purposes, especially at scales  $\ell \gtrsim 30$ . *Right:* Error in power spectrum estimation, shown as the residual angular power relative to sample variance  $(\tilde{C}_\ell^{est} - \tilde{C}_\ell^{ss}) / \sigma_{C_\ell}$  in bins of  $\Delta\ell = 10$ . Solid lines indicate means of cleaning performed on 50 signal realizations of each bin and shaded regions indicate the central 68% probability mass of the 250 total realizations. The multiplicative correction applied to Mode Projection removes most of the bias of the method (green to orange). Here we use 27 templates, of which 11 are contaminating the data.

its mean traces the mean for the additive Mode Projection method, but the dispersion is very large, exceeding the plot limits.

### B. Susceptibility to overfitting

Any template-fitting model faces a challenge to neither underfit nor overfit the data. In the case of underfitting, residual contamination will be left over in the map and inferred to be signal. In the case of overfitting, a portion of the signal will be inadvertently removed from the map, having been mistaken for systematics. Additionally, increasing the number of fitted templates increases the variance of the estimated power spectrum, which will increase the error of  $\tilde{C}_\ell^{est}$  in a mean-squared sense[125].

Mode Projection and Template Subtraction address the risk of overfitting by estimating how much signal power is lost from over-correction given the template library and scaling the power spectrum accordingly (Eqs. 11 and 23). In contrast, the DES-Y1 and Forward Selection methods use thresholds to limit the templates used for cleaning to only those that are most significant,

an approach that was also implemented in the *Extended* Mode Projection method of [90] for the QML power spectrum estimator (though as shown by [120], this comes at the cost of an unknown bias in the power spectrum). As described in Sec. V B, the Elastic Net reduces overfitting by adding a prior on the template coefficients to reduce the number of templates used.

While each of the methods addresses overfitting in its own way, the library of templates fed to them has in most cases already been narrowed from a much larger set of possible templates through decisions made by researchers. For example, almost all modern surveys observe any given patch of sky multiple times, resulting in multiple values for each observing condition for every pixel. To produce a scalar template map requires compressing these values into a summary statistic and, as it isn't known a priori which statistic will best capture systematic contamination of the data, multiple statistics may be computed, each corresponding to its own template (see e.g. Ref. [87]). If just one statistic (such as the mean) is chosen as representative as is often done, there is the very real risk of discarding potential templates that more accurately capture the contamination, resulting in

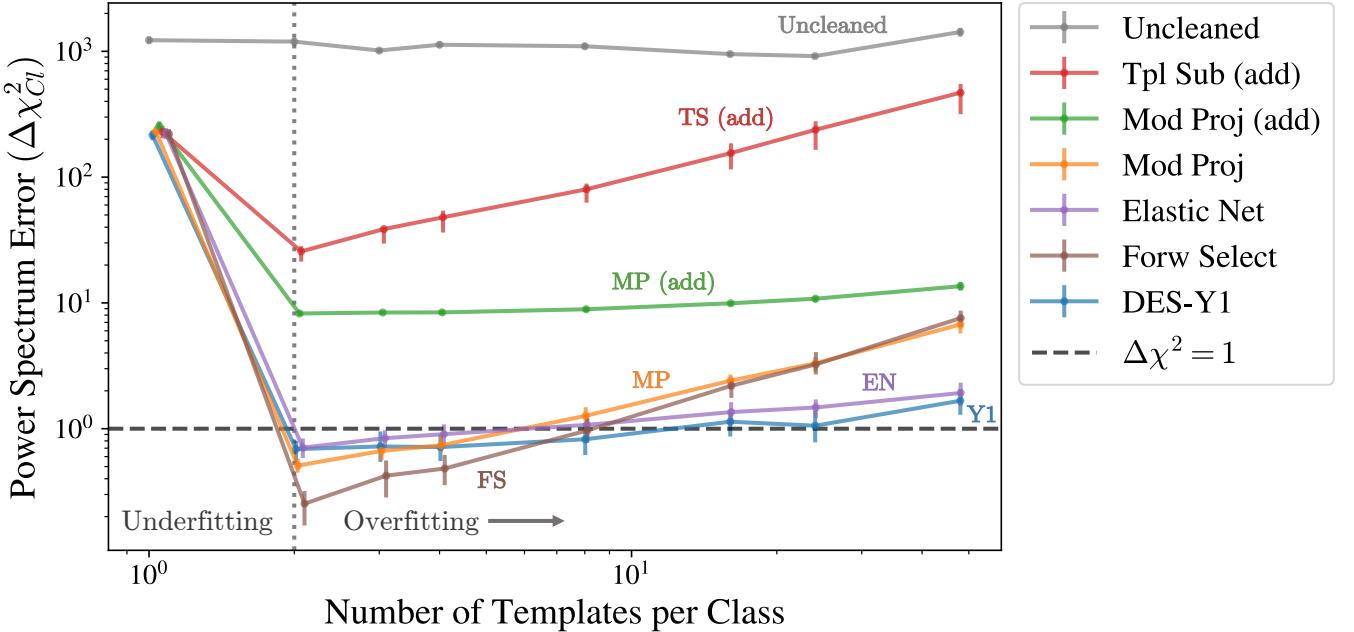


FIG. 10. Bias in the angular power spectrum,  $\Delta\chi_{C_\ell}^2$ , as a function of the number of templates fit to the map. We consider Gaussian templates which have a correlation of  $\rho_{tpl} = 0.2$  within each of the four template classes, as defined in Sec. VIA. We generate two template realizations per class with which to contaminate each signal map ( $N_{sys} = 2$ , denoted by the vertical dotted line). The templates used to perform the cleaning vary from one to 48 for each type of template spectrum, for a total of four to 196 templates. The Template Subtraction, Mode Projection, and Forward Selection methods are all mildly susceptible to overfitting — signaled by the increase in  $\Delta\chi_{C_\ell}^2$  for  $N_{tpl} > 2$  — though only Template Subtraction to a degree where it overcomes the penalty for neglecting a contaminating template ( $N_{tpl} = 1$ ). For the *additive* Mode Projection method,  $\Delta\chi_{C_\ell}^2$  is dominated by the bias from not addressing the multiplicative contribution to the power spectrum (see Fig. 9, right panel), while the other methods are dominated by increased variance from chance correlations. The bias from failing to correct for the multiplicative term dominates even when fitting for  $\sim 200$  templates. The DES-Y1 and Elastic Net display a lesser dependence on  $N_{tpl}$ , and so are more robust to overfitting. See Sec. VII B for details.

residual contamination, or *underfitting*.

*Therefore, one of the key performance metrics for these methods is their ability to handle increasing numbers of non-contaminating templates without degrading map or power spectrum estimates, and so simultaneously mitigate the risks of under- and over-fitting.*

To characterize the error in the reconstructed angular power spectrum, we use the sum of squared errors between the true and reconstructed power spectra, normalized by sample variance:

$$\Delta\chi_{C_\ell}^2 = \sum_{z\text{bins}} \sum_{\ell=\ell_{\min}}^{350} \frac{\left(\tilde{C}_\ell^{\text{est}}(z) - \tilde{C}_\ell^{\text{ss}}(z)\right)^2}{\sigma_{C_\ell^{\text{ss}}(z)}^2}, \quad (72)$$

where  $\ell_{\min} = 2$ , except for Template Subtraction where  $\ell_{\min} = \text{Ceil}[(N_{tpl} - 1)/2]$ , since with  $N_{tpl}$  templates, all signal is removed for  $\ell \leq (N_{tpl} - 1)/2$ .

In Fig. 10 we show  $\Delta\chi_{C_\ell}^2$  as a function of the number of templates used to clean the maps (Fig. 16 shows the same plot for map-level statistics, which demonstrate very similar behavior to  $\Delta\chi_{C_\ell}^2$ ). We generate two template realizations per class with which to contaminate each signal map, and vary the number of templates used

to perform the cleaning from one to 24 for *each template class*. The true contaminants are always ‘selected first’, such that  $N_{tpl} = N_{sys} = 2$  represents correctly fitting for the two contaminating templates from each class (vertical dotted line), whereas  $N_{tpl} > 2$  indicates the penalty for overfitting of non-contaminating templates. The error bars come from many signal realizations for the same template maps, and different template and signal map realizations are used for each value of  $N_{tpl}$ .

Fig. 10 demonstrates that all methods are susceptible to overfitting, as indicated by the fact that  $\Delta\chi_{C_\ell}^2$  increases for  $N_{tpl} > 2$ , but that some are more susceptible than others. Template Subtraction and additive Mode Projection are the worst-performing methods with  $\Delta\chi_{C_\ell}^2 \gtrsim 10$  for all cases, with Template Subtraction showing a strong dependence on  $N_{tpl}$ . Multiplicative Mode Projection and Forward Selection display approximately the same  $\Delta\chi_{C_\ell}^2 \propto N_{tpl}$  scaling as Template Subtraction, whereas The Elastic Net and DES-Y1 methods show a much weaker scaling, indicating that they are much more robust to a larger number of templates.

The trend for *additive* Mode Projection method indicates the importance of the multiplicative correction.

Here, the error in the power spectrum does not scale with  $N_{\text{tpl}}$  as strongly as that of Template Subtraction or the multiplicative Mode Projection method because it is dominated by the bias from not addressing the multiplicative contribution to the power spectrum (see Fig. 9, right panel), not the increased variance from a larger number of templates. The bias from failing to correct for the multiplicative term dominates the additive Mode Projection error even when overfitting by  $\sim 200$  templates. Were the plot to continue to the right, we would expect the error to begin to scale similarly to the other Mode Projection and Template Subtraction methods.<sup>20</sup>

Another key point is that for all cases except Template Subtraction, the penalty for overfitting is dwarfed by the penalty for neglecting contaminating templates ( $N = 1$  on the x-axis). This suggests that the researchers should err on the side of overfitting, rather than risk removing contaminating templates from the cleaning library. This is especially true if using a method that is more robust to overfitting, such as the Elastic Net or DES-Y1 method. In sum, the DES-Y1, Forward Selection, Mode Projection with multiplicative correction, and Elastic Net methods all perform very well relative to the uncleaned case, with the Elastic Net and DES-Y1 methods being most robust to overfitting and achieving the best performance with  $\Delta\chi^2_{C_\ell} \simeq 1$  even when  $N_{\text{tpl}} \gg N_{\text{sys}}$ .

### C. Impact of correlated templates

Real templates often have groups of templates that are highly similar to one another in their spectral behavior and/or in their correlation to one another, which we have modeled here as different template classes. The same tracer/property measured in different wavelength bands, or different summary statistics (e.g. the mean vs. median) for the same tracer in a multi-epoch survey are both common examples that can result in very similar templates. We wish to investigate the impact of selecting a non-optimal template for cleaning, which only partially describes the true systematic. This could be either through the choice of a non-optimal summary statistic, or through the apriori choice of a ‘representative’ template from a group of similar templates in order to mitigate the risk of overfitting, as is commonly done in current surveys.

We test this by cleaning with sets of templates that have varying levels of within-class correlation. For each template class (corresponding to one of the spectra listed in Sec. VI A) we use `Healpy.synfast` to generate template realizations with off-diagonal covariance terms be-

tween templates  $i$  and  $j$  of

$$C_\ell^{ij} = \begin{cases} \rho_{\text{tpl}} \sqrt{C_\ell^{ii} C_\ell^{jj}}, & \text{if } i \text{ and } j \text{ in same class} \\ 0, & \text{if } i \text{ and } j \text{ in different classes} \end{cases}$$

We only use the first four classes from Sec. VI A, which are defined by their spectrum and from which we can generate multiple Gaussian realizations with defined levels of cross-correlation.

Fig. 11 shows the performance of the methods when the within-class correlation between templates is varied. We again consider the case of two contaminating systematics from each of the four Gaussian template classes. The left panel shows the case where for each class we have chosen only one of the templates to clean with, deeming it “representative” of the template group. As within-class correlation between the systematics increases, the cleaning templates are more representative and can increasingly remove more of the unaccounted-for contamination. At  $\rho_{\text{tpl}} = 0.9$ , the multiplicative methods are able to reduce the error to  $\Delta\chi^2_{C_\ell} \sim 6$  compared to  $\Delta\chi^2_{C_\ell} \sim 300$  for the uncorrelated case.

Despite the additional freedom of the Template Subtraction method to fit multipoles independently, it does not do a better job than the other methods of correcting for the “unknown” systematics. The multiplicative methods have almost identical performance, with the dominant contributions to residual errors in the power spectrum resulting from the unaccounted-for systematics and, to a lesser extent, failing to treat the multiplicative term of the contamination.

The right panel in Fig. 11 illustrates the other approach of including many possible templates rather than preselecting a few: we use six cleaning templates from each class: the two true systematics, plus four more that are uncontaminating, for a total of 24. We find that with the exception of the DES-Y1 method, performance of the methods is largely independent of the correlation between templates.<sup>21</sup>

Comparing the panels, even if using a high threshold of similarity of  $\rho_{\text{tpl}} = 0.9$  to discard templates, significantly more error is introduced through neglecting a contaminating template than through overfitting, so it is better to not pre-select templates solely on the basis of similarity to others and instead err on the side of too many templates rather than too few. Template Subtraction is

---

<sup>20</sup> The multiplicative bias is not the dominant contribution for Template Subtraction because its *effective* number of templates is much larger, since it performs  $N_\ell$  regressions for each template.

<sup>21</sup> We found this to be true for both map-level and 2-pt reconstruction statistics, though we only show the latter here. It is not obvious from the outset that this would be the case — Forward Selection methods are often criticized for being less reliable when predictors are correlated, though this is in the context of the more typical regression scenario where it is the predictors themselves that are of interest, as opposed to the residuals which is our focus here. The source of the dependence of the DES-Y1 results on  $\rho_{\text{tpl}}$  is not entirely clear, but our investigations found it to be mildly impacted by both binning choices and the total monopole of systematic maps.

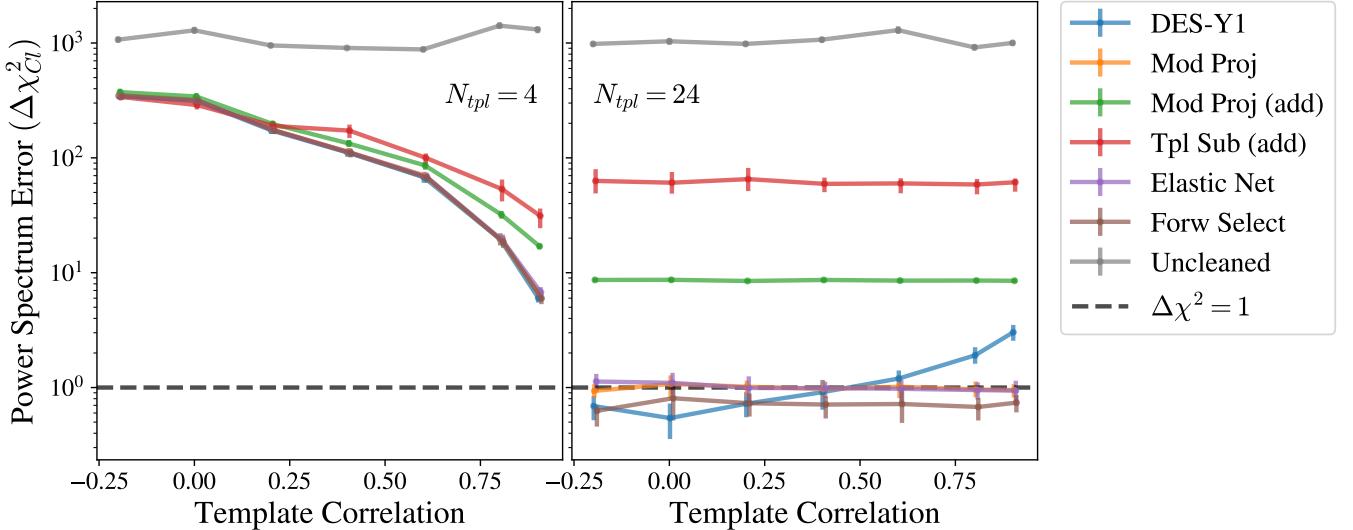


FIG. 11. Bias in the angular power spectrum,  $\Delta\chi_{Cl}^2$ , as a function of the level of cross-correlation imposed between the templates within the same class. We assumed contamination from two realizations from each of the four classes (i.e.  $N_{sys} = 8$ ). The left panel assumes cleaning with only one of the contaminating templates from each class, while in the right panel we clean for four templates from each class, including the contaminating ones. Note that in the case where template correlation  $\rho_{tpl} \rightarrow 1$ , the two templates are identical and it is equivalent to cleaning only for one contaminating template, an ideal scenario. In the right panel we see that while the DES-Y1 outperforms others when templates are completely orthogonal, it suffers as the level of correlation between templates increases. The Elastic Net method mitigates this problem.

the one exception to this, where each additional template results in  $N_\ell$  additional fits. While the additional freedom does not substantially protect against unknown systematics, it does result in a much steeper penalty for overfitting from higher  $N_{tpl}$ .

#### D. Extensions

By interpreting current LSS systematics cleaning methods in the context of regression, we have facilitated their comparison and interpretation, as well as motivated several possible extensions to them. We have explored some of these extensions in this work, such as the Elastic Net method in Sec. V B, and the use of leverage to predict overdensity errors, but with the extensive body of regression methods, there are many more that we must leave to future work. For example, one promising avenue for regression methods that use a threshold for template selection would be to motivate that threshold by controlling the ratio of Type I (false correction) to Type II (false omission) errors in the selection process via the False Discovery Rate[142], based on the relative impact of each type of error on the analysis.

We have noted individually multiple cases where the assumptions made by the methods do not hold and how they might be improved. A full treatment of these effects is beyond the scope of this paper and would include the full non-Gaussian likelihood of  $P(d_{obs}|\hat{f}_{sys})$ , including contributions from systematics, but as we show in

Apps. A and B, the corrections from these are minor compared to the methodological differences and the improvements we suggest. Generalized linear models may be a promising compromise for future mitigation routines, preserving off-the-shelf implementation and diagnostic tools, while providing greater specificity for the likelihood and relaxing some of the tacit assumptions of Mode Projection and OLS regression.

The methods presented here are general enough to be applicable in any situation where one has an external prediction (template) for systematic contamination of observational data, and is equally applicable to spin-2 fields. The insights gained can be used to further extend linear models like the ones in this work, or inform the formulation of nonlinear contamination models, non-parametric methods, or machine learning approaches such as that of Rezaie *et al.* [100].

## VIII. SUMMARY OF METHODS

Here we summarize our findings about the performance of systematic-cleaning methods.

- **DES-Y1 method:** The most complicated method of the ones we studied, the DES-Y1 method resulted in some of the lowest biases in the cleaned maps. It usefully includes prior information about the covariance between pixels in the fitting procedure, albeit in a coarse way. However it is also somewhat complicated to implement, as it

requires a large number of parameter choices on the part of the researcher (binning number and procedure, significance statistic and threshold, power spectrum prior) and the generation of realistic mocks. We observed some degradation of its performance as the correlation between templates increased. It is one of the two methods most robust to overfitting when using a large library of templates that are not actually contaminating the data (the other being Elastic Net).

- **Mode Projection:** The standard pseudo- $C_\ell$  Mode Projection method, as introduced in [125] and implemented in `NaMaster`[122]. We showed that it is equivalent to removing the result of an ordinary least squares regression of the observed data onto the template maps (thus providing a map estimate), with an additional step to debias the power spectrum. This removes most of the contamination present, but can be simply adapted to, and significantly improved by, treating the multiplicative component of contamination instead of just the additive term. We demonstrate how to do this in Sec. IV B. In all cases we studied, the error from not correcting the multiplicative term dominated over error induced from overfitting — as Fig. 10 illustrates, in the *ideal* case where our templates exactly matched the systematics, not treating the multiplicative term introduced as much error as using  $\sim 30\times$  more templates than systematics in the cleaning procedure.
- **Template Subtraction:** Equivalent to performing an individual OLS regression at each multipole, resulting in large variance and significant loss of signal from overfitting. As a result, it does not reconstruct maps well and generally performs most poorly in all of our tests. However, our implementation is a limiting case, where each harmonic from each template is allowed to contaminate independently, in contrast to Mode Projection where all modes contaminate identically. The work here should make it straightforward to construct a hybrid method where all modes contribute identically like in Mode Projection (as is physically motivated) and hence have small variance, but where certain modes are prioritized for cleaning, based on the analysis case.
- **Iterative Forward Selection:** This is a method we propose, which is a much simpler version of the DES-Y1 method that requires only a single tunable parameter (a significance threshold) and no mocks. We found that it produces excellent results and is robust to correlation between templates, but is not as robust to overfitting, displaying the same dependence of roughly  $\Delta\chi^2_{C_\ell} \propto N_{\text{tpl}}$  as the Mode Projection and Template Subtraction methods.

- **Cross-Validated Elastic Net:** A cleaning method we introduce, which we find has the best overall performance, being consistently low error and robust to overfitting. It is equivalent to Mode Projection, but with the amplitude of contamination for each template having a mixed Gaussian/Laplace prior applied to encourage sparsity and thus automatically select the important templates. The ‘priors’ are not strictly such in a Bayesian sense, as their strengths are determined by the data through cross-validation. It is easy to implement using out-of-the-box software and doesn’t require a user-defined prior for the power spectrum or debiasing step, providing the best balance of performance, ease of implementation, interpretability and robustness.

## IX. CONCLUSIONS

In this paper, we carried out a broad comparison of methods used to remove astrophysical, atmospheric, and instrumental systematic errors that affect galaxy-clustering measurements. We have generalized previous work by 1) showing how different methods can be interpreted under a common regression framework, 2) jointly assessing the robustness of methods on simulated data, 3) investigating the reconstruction fidelity of LSS *map(s)*, rather than just their clustering statistics, as the maps are useful points of departure for numerous other analyses (e.g. summary statistics beyond the power spectrum, cross-correlations, searches for signatures of dark matter or exotic new physics); and 4) proposing improvements to current methods, as well as new, hybrid and efficient methods for the systematics cleaning.

We employed a simple and general model for systematics, given in Eq. (1), which allows for spatially varying multiplicative and additive systematic errors with a range of clustering properties to any generic cosmological field. Equipped with that model, we defined a testing procedure that attempts to mimic real-world conditions for LSS surveys, where the true galaxy map is contaminated with an unknown set of systematics and a set of known templates is used to model and correct for the contamination. Given our methodology (pictorially described in Fig. 1) and a set of assumptions about the fiducial DES Y5-like survey used to generate the maps, we studied the performance of the systematics-cleaning methods under different conditions.

We showed that both Template Subtraction and Mode Projection, while developed independently, can be interpreted through a regression framework where the signal of interest corresponds to the noise term of a regression model. This allowed us to straightforwardly apply known statistical results and techniques to these methods. We used this to adapt additive methods to account for multiplicative errors (Fig. 5), and identify potentially highly contaminated map pixels as a function of their “lever-

age” (Fig. 7), while opening up avenues for further improvement. One such avenue we touched on was to optimize Mode Projection (or other regression methods) by prewhitening the maps in harmonic space. Recognizing that the noise of the regression is the clustering signal itself, we proposed that the maps could be efficiently and optimally inverse-variance weighted in harmonic space, where the clustering signal is diagonal. This is equivalent to accounting for the off-diagonal pixel covariance in the pixel-based regression methods, which is rarely done for tractability reasons (but see Wagoner *et al.* [101] for one approach). We found this to improve results (Fig. 14), but be subdominant to the multiplicative correction and differences between the cleaning methods.

We introduced two new methods for cleaning: (1) the ‘Forward Selection’ method, which is a greatly simplified version of the DES-Y1 method that achieves similar performance albeit being less robust to a large number of templates; and (2) the ‘Elastic Net’ method, a simple out-of-the-box method that implements Mode Projection, but which automatically selects important templates. We found that the Elastic Net method is very robust, with strong performance even when there is a large number of templates (Fig. 10) or templates are highly correlated (Fig. 11); both are cases where other methods display weaknesses. This method is very easy to implement, and we recommend it for future surveys.

On the whole, we found that all of the methods perform quite well, dramatically improving the chi-squared difference between the cleaned and true (uncontaminated) angular power spectrum. At the map level, Template Subtraction was the only method that did not significantly reduce the RMS overdensity error across pixels (Figs. 7 and 16), and so we do not recommend the version implemented here for map reconstruction. Once we adopted only the algorithms that take into account both additive and multiplicative errors, all of the methods improved  $\Delta\chi^2_{C_\ell}$  by three orders of magnitude relative to the uncleaned case. Moreover, overfitting did not lead to large degradation in the reconstructed power spectra (see Fig. 10), which is encouraging. Finally we found that the performance of the various systematics-cleaning methods is very weakly dependent on the level of cross-correlation between the template maps used for the cleaning, with the DES-Y1 method being mildly more susceptible.

We end with several recommendations based on this work:

- I. Current and future cleaning methods should account for multiplicative contamination. ‘Weights’ methods like the DES-Y1 method already do this and other methods like (Pseudo- $C_\ell$ ) Mode Projection can easily do so via Eqs. (46)-(47).
- II. Cleaning methods based on a single Ordinary Least Squares regression are equivalent to (Pseudo- $C_\ell$ ) Mode Projection and so should debias inferred two-point functions accordingly.

III. Analyses should err on the side of overfitting rather than underfitting for templates, as the error from the former tends to be small. Researchers should avoid arbitrarily removing templates from the library prior to cleaning based solely on their similarity to other templates. Larger template libraries result in increased variance of the map and power spectrum estimators, especially with the very large number of templates that will be available to future surveys. Therefore:

- IV. In scenarios where a very large template library is available, the data itself should be used to select a subset for cleaning. Among the methods that we studied, this is accomplished by either a DES-Y1 type method or the Elastic Net with cross validation. Both show good robustness, and the latter is simple to implement with common software. The theoretical connections we have made between methods should make alternative template selection routines such as those in Refs. [87, 100] simple to adapt and implement.
- V. The cleaning methods used thus far can — and should — be viewed in the context of regression, with the estimated overdensity field corresponding to the regression residuals. Researchers should make use of the powerful suite of existing tools and diagnostic measures to assess the validity of regression models when cleaning LSS data (e.g. leverage for outlier detection, Q-Q plots, partial regression/residual plots) and to aid mask creation. This is applicable to all methods studied in this paper.

## ACKNOWLEDGMENTS

We thank Boris Leistedt and Jessie Muir for helpful comments and encouragement. The authors have been supported by DOE under Contract No. DE-FG02-95ER40899. D. H. has also been supported by NASA under Contract No. 14-ATP14-0005, and N. W. has been supported through a Leinweber Graduate Fellowship.

## Appendix A: Lognormal vs. Gaussian Signal Maps

While the methods presented here are quite general for any case where systematic contamination can be traced using a template, we have specifically worked in the context of galaxy clustering. In this case, the signal map  $s$  that we are attempting to model is the galaxy overdensity  $\delta$ , which is subject to the constraint  $\delta > -1$  (as is the case for any overdensity statistic). Thus our assumption that  $s$  is Gaussian breaks down at low redshift and at small scales, when  $|\delta|$  can be large.

It is well known that galaxy and shear overdensities are better approximated by a lognormal distribution (see e.g. Refs. [143–146]), so we run the methods on a series

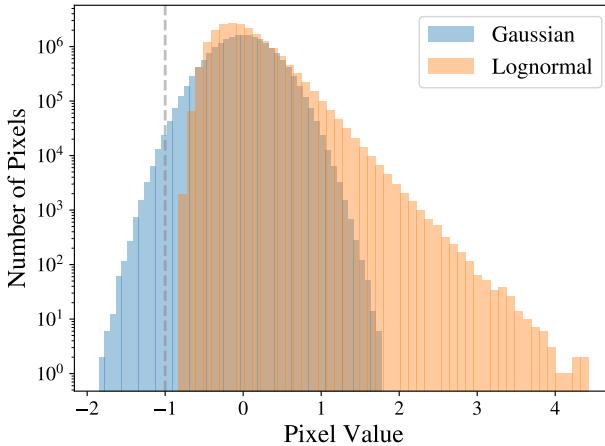


FIG. 12. Distribution of pixel overdensities across all 100 realizations of the lognormal (orange) and Gaussian (blue) maps of the galaxy overdensity in the lowest redshift bin of our fiducial DES-like survey. The Gaussian maps contain pixels with  $s < -1$ , which is nonphysical in cases like this where  $s$  corresponds to an overdensity.

of lognormal maps to see if the relative performance of the methods changes.

We generate 100 Gaussian signal realizations  $s_G(\hat{\mathbf{n}})$  of the lowest redshift bin of our fiducial DES survey, for which the cosmological signal will be most non-Gaussian. We generate lognormal versions of these maps by first computing the transformation that achieves zero-mean lognormal overdensity field *in the ensemble* [145], then centering and scaling so that each realization of the lognormal field has the same mean and variance as its Gaussian counterpart. The two steps correspond to the mathematical operations:

$$\begin{aligned} 1. \quad s'_{LN}(\hat{\mathbf{n}}) &= e^{s_G(\hat{\mathbf{n}})} - e^{\text{Var}[s_G(\hat{\mathbf{n}})]/2} \\ 2. \quad s_{LN}(\hat{\mathbf{n}}) &= \sqrt{\frac{\text{Var}[s_G(\hat{\mathbf{n}})]}{\text{Var}[s'_{LN}(\hat{\mathbf{n}})]}} (s'_{LN}(\hat{\mathbf{n}}) - \bar{s}'_{LN}(\hat{\mathbf{n}})). \end{aligned}$$

The resulting lognormal realizations are then of the form

$$s_{LN}(\hat{\mathbf{n}}) = \lambda_1 e^{s_G(\hat{\mathbf{n}})} - \lambda_0, \quad (\text{A1})$$

with scale and shift parameters of  $\lambda_1 = 0.9123 \pm 0.0017$  and  $\lambda_0 = 0.9697 \pm 0.0017$ , respectively for our lowest redshift bin, which is the most non-Gaussian.

Fig. 12 shows the distribution of pixel overdensities across all realizations of the lognormal and Gaussian signal maps. It is clear that the Gaussian maps contain many pixels with  $s < -1$ , which is nonphysical for our case, where  $s$  corresponds to an overdensity. The lognormal maps avoid this problem and are a better approximation of the true overdensity distribution. As we have shown, most of the cleaning methods can be viewed under

a regression framework wherein the signal distribution is assumed to be Gaussian, so we investigate whether our comparison of methods changes when using a more realistic lognormal distribution.

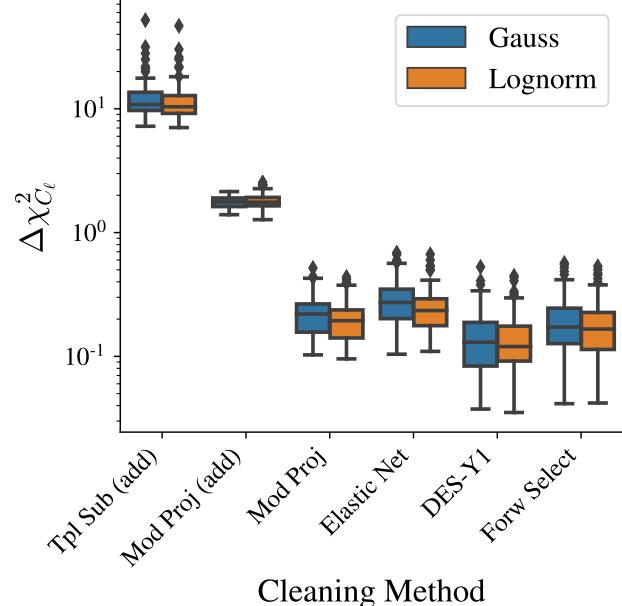


FIG. 13. Box plot showing the performance of each cleaning method when using Gaussian (blue, left) versus lognormal (orange, right) signal maps, as measured by  $\Delta\chi^2_{C_\ell}$  of the power spectrum. Filled boxes show the 25-50-75% quartiles, with whiskers encompassing the rest of the distribution out to  $1.5 \times$  the inter-quartile range. Points beyond this range are indicated by diamonds. Regardless of whether lognormal or Gaussian maps are used, the relative performance of the methods to one another is largely unchanged, and the Gaussian approximation is negligible compared to neglecting the multiplicative correction of Sec. IV B.

Fig. 13 shows the error in the power spectrum reconstruction, given by the  $\Delta\chi^2_{C_\ell}$  statistic, for the different methods. We find that while there is some overall shift, using the lognormal signal maps does not change the relative behavior of the methods; none of them display a unique susceptibility to the assumption of Gaussianity in the signal maps.

## Appendix B: Effect of Prewitening

In their derivation of the bias on the estimated power spectrum after (pseudo- $C_\ell$ ) Mode Projection, Elsner *et al.* [125] assume that the map  $d$  has been decorrelated (“prewhitened”) before projecting out the templates. This is quite difficult to do in practice, as it requires the inversion of an  $N_{\text{pix}} \times N_{\text{pix}}$  matrix, the same problem with QML estimators for the power spectrum. Indeed, one of the assumptions of pseudo- $C_\ell$  estimation is that pixels are uncorrelated (though individual pixels

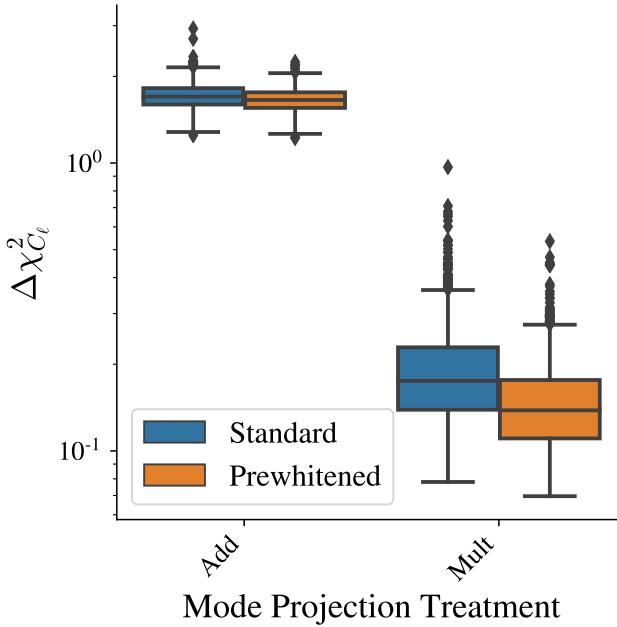


FIG. 14. Impact of prewhitening before cleaning with the multiplicative and additive versions of the Mode Projection method on 1000 realizations for our fiducial contamination model. The standard Mode Projection method assumes a flat power spectrum for the target signal, resulting in a sub-optimal estimate of contamination. This can be improved through ‘prewhitening’ the data vector and templates using a prior power spectrum, which can be shown to be equivalent to a standard weighted regression procedure in harmonic space. There is clear but modest improvement from the standard case (blue) to the nearly-optimal, prewhitened case (orange), with the most improvement seen for realizations that have large error. This can be seen by the preferential reduction of extreme points at the high end of the box plots in the prewhitened case (note the log scale).

are weighted by an estimate of their inverse noise variance and by the mask, see e.g. Ref. [122].)

As shown in Sec. IV A, however, the dominant ‘noise’ in our observations is actually our true clustering signal, so a true ‘prewhitening’ step should more appropriately inverse weight the data by the expected clustering variance. This can be done efficiently in harmonic space when there is no mask, as the clustering signal is diagonal, circumventing the need to invert a large covariance matrix.

We can define prewhitened data vectors for our observed overdensity field and templates as

$$(d_{\text{obs}})'_{\ell m} = (d_{\text{obs}})_{\ell m} / \sqrt{C_{\ell}^{ss}}. \quad (\text{B1})$$

$$(t_i)'_{\ell m} = (t_i)_{\ell m} / \sqrt{C_{\ell}^{ss}}, \quad (\text{B2})$$

which results in coefficient estimates of

$$\hat{\alpha} = (\mathbf{T}'^{\dagger} \mathbf{T}')^{-1} \mathbf{T}'^{\dagger} \mathbf{d}'_{\text{obs}}, \quad (\text{B3})$$

where  $\mathbf{T}'$  is a  $N_{\ell m} \times N_{\text{tpl}}$  matrix with complex entries defined in Eq. (B2). We can compute the amplitudes directly with

$$\hat{\alpha} = \frac{\sum_{\ell=0}^{\ell_{\text{max}}} (2\ell+1) \tilde{C}_{\ell}^{td} / C_{\ell}^{ss}}{\sum_{\ell=0}^{\ell_{\text{max}}} (2\ell+1) \tilde{C}_{\ell}^{tt} / C_{\ell}^{ss}}. \quad (\text{B4})$$

We found that prewhitening improved  $\Delta\chi_{C_{\ell}}^2$  by a mean of  $\sim 0.05$  with dispersion 0.08 across the mocks, with similar shifts regardless of whether the multiplicative correction was applied or not. Fig. 14 shows the improvement from the standard case (blue) to the prewhitened case (orange) for both additive and multiplicative mode projection. While we do not show it, we found that the benefit of prewhitening increased for realizations that had worse power spectrum estimates (higher  $\Delta\chi_{C_{\ell}}^2$ ), in effect catching and mitigating particularly bad realizations.

In practice, one would either assume a prior power spectrum for prewhitening or compute it iteratively, just as one does for the Mode Projection debiasing step, so this could easily be incorporated into existing Mode Projection routines such as **NaMaster**. As noted in Sec. IV A, since Mode Projection is equivalent to regression, this improvement also quantifies the expected level of improvement that would come from accounting for the covariance between pixels in pixel-based regression methods.

Analyses on real data will of course be complicated by the mask, which correlates different multipoles, but this can be addressed by suitable binning of the multipoles. Indeed, the standard pseudo- $C_{\ell}$  Mode Projection assumes a flat power spectrum and so can be thought of as the limiting case of using only a single bin across multipoles with equal weighting, such that even a rough estimate of the signal power spectrum should offer improvement.

The other methods tested here should benefit similarly from prewhitening, with the possible exception of the DES-Y1 method, which already incorporates an estimate of the covariance of  $s$  (which accounts for much of the methods’ complexity). The Forward Selection method we presented may be particularly impacted, since the estimated covariance of the fit parameters is underestimated when the pixel covariance is neglected, and this is used for the significance criterion for selecting a template. This could be one reason why the Forward Selection method sometimes failed to reduce all templates to below a significance of  $\Delta\chi^2 / \Delta\chi_0^2 = 2$  — such a threshold was artificially low compared to what would be expected from random variation.

As noted in Eq. 51, the prewhitening step in Eq. (B2) should optimally include contributions from the systematics as well. However as this represents minor perturbations to the major prewhitening correction above and is hence a small ‘error on the error’, the effects should be small. This is consistent with Elvin-Poole *et al.* [41], who found negligible impact on their method from neglecting the additional systematics contribution to their estimated covariance matrices.

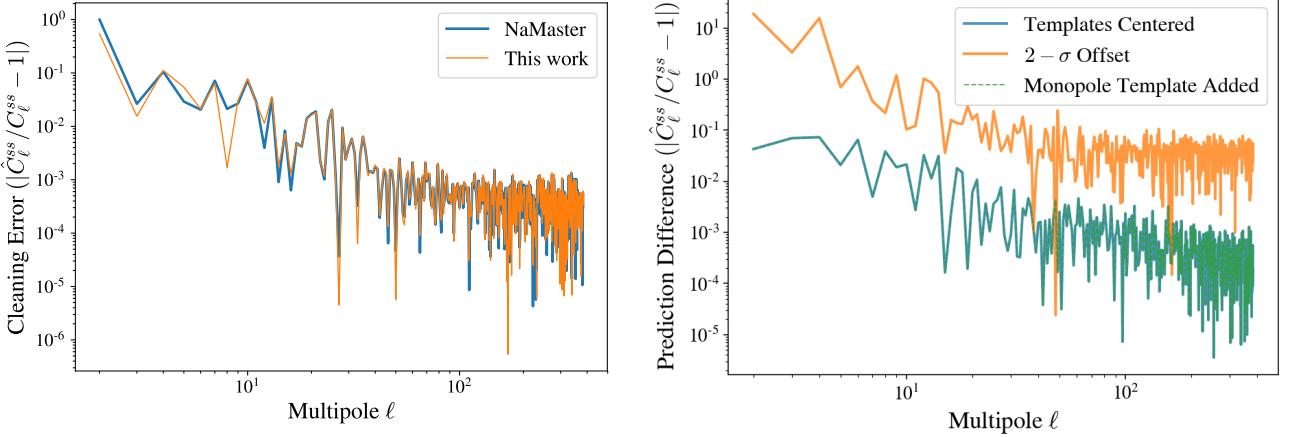


FIG. 15. Validation tests of the Mode Projection map-cleaning procedure. *Left panel:* Comparison of Mode Projection performance on an additive-only contaminated map (as assumed by the Mode Projection method), using **NaMaster** (blue) and our own implementation (orange). The agreement between the two is very good. *Right panel:* Impact of not pre-centering cleaning templates in **NaMaster**. The blue curve indicates the standard use case, where contamination is additive and completely described by the templates, which have been individually centered at zero. If templates are instead centered at another value (here we add a constant  $2\sigma_{tpl,i}$  offset to each template, where  $\sigma_{tpl,i}$  is the standard deviation of values in template map  $i$ ). Adding a monopole template completely mitigates the bias from non-centered templates.

### Appendix C: Comparison with NaMaster

We have used our own implementation of the Mode Projection method and have tested it against that of **NaMaster** [122], finding good agreement. **NaMaster** computes the power spectra given a set of templates and observations, but does not produce map estimates, so we compare the two implementations using the cleaned power spectrum only. The left panel of Fig. 15 shows the relative error of the estimated power spectrum when cleaned using **NaMaster** vs. our own implementation, using the exact same contaminated map and templates and we find good agreement (this held true for all realizations tested). There is very slight disagreement at larger scales (low  $\ell$ ), which may be numerical artifacts from the **Master** [126] algorithm implemented to account for mode coupling on a cut sky being applied to full-sky input maps. Regardless, the deviations between the two are small for  $\ell > 2$ .

### Appendix D: Accounting for the Monopole

It is worth saying a few words about the monopole term, both as in terms of prediction and as it relates to regression.

Firstly, the *overdensity* residuals do not correspond to the *number* density residuals. Even with a perfect reconstruction  $\hat{s} = s$ , the true number density will be unknown up to a factor of  $\gamma$ ,

$$N_{\text{true}} = \gamma \langle N_{\text{obs}} \rangle_{\text{pix}} (s + 1), \quad (\text{D1})$$

and as such the estimated number density could be quite

different from the truth. Fig. 4 shows a somewhat unintuitive consequence of this. The single systematic that contaminates the field has the form  $f_{\text{sys}} \propto -t$ , so that it only obscures galaxies from view ( $f_{\text{sys}} \leq 0$ ). At  $t = 0$ , there is no contamination and so  $N_{\text{obs}} = N_{\text{true}}$ , however as the figure shows the *over-density* residuals are quite large. This is because the *mean* number density is significantly underestimated, so pixels with no obscuration are preferentially (and wrongly) estimated to reside in overdense regions.<sup>22</sup>

Secondly, a net monopole in  $f_{\text{sys}}$  corresponds to the intercept in the regression methods (a column of ones in  $T$ ).

In OLS regression (or pseudo- $C_\ell$  Mode Projection), the fit is guaranteed to go through the center of mass of the points,  $(t, \bar{d}_{\text{obs}})$ , such that including a monopole is unnecessary with such methods if working with overdensities and zero-centered templates. In such cases, the ‘projection’ of the monopole has already been done by subtracting the mean from the density and template maps (consider Eq. (17) with a template of all 1s). We showed in Eqs. (38)-(44) how this also holds in the multiplicative case.

In realistic situations, there is high susceptibility to human error if a monopole term is not included — previously zero-centered maps can easily shift through template transformations, mask adjustments, and the application of a mask to mocks, resulting in wildly biased contamination estimates that may be difficult to detect. For example, it is easy to pass templates that are not zero-

<sup>22</sup> In other words,  $\gamma > 1$ , so from Eq. (38),  $\langle d_{\text{obs}} | f_{\text{sys}}=0 \rangle_{\text{pix}} > 0$ .

centered to current pseudo- $C_\ell$  Mode Projection methods such as implemented in **NaMaster** and receive highly biased spectra without warning (see right panel of Fig. 15).

The DES-Y1 and Forward Selection methods both already include an intercept term, in keeping with the original formulation of the DES-Y1 method, though in practice it should be very close to zero.

We therefore opt to include a monopole term in our Elastic Net method, as this ensures the method is robust and generalizes the process beyond overdensities to non-zero mean fields, and it will naturally be ignored as a template if it does not contribute information.

### Appendix E: Impact of $\Delta\chi^2/\Delta\chi_0^2$ on DES-Y1 Analysis

Here we investigate the effect of  $\Delta\chi^2/\Delta\chi_0^2$  and  $\sigma_{\text{sys}}^2$  on the efficacy of the DES-Y1 method, as described in Sec. III A. We describe the reconstruction quality with the residual chi squared between the cleaned and true model,  $\Delta\chi_{C_\ell}^2$ .

Fig. 17 shows how  $\Delta\chi^2/\Delta\chi_0^2$  affects the reconstruction quality for the DES-Y1 method, as a function of the level of contamination parameterized by the systematic-error variance  $\sigma_{\text{sys}}^2$ . We find little reduction in error by lowering the significance threshold below  $\Delta\chi_{\text{threshold}}^2 = 4$ .

At our fiducial level of contamination ( $\sigma_{\text{sys}}^2 = 10^{-2}$ ), almost all contaminating templates exceed the highest threshold displayed of  $\Delta\chi^2/\Delta\chi_0^2 = 32$  and so are corrected for. The larger the contamination, the more precisely its form can be determined, so as the level of contamination decreases, some contaminated templates are left uncorrected for. This results in the somewhat counter-intuitive turnover in the error for a given threshold level. We found that the lowest threshold of  $\Delta\chi^2/\Delta\chi_0^2 = 1$  consistently outperformed higher thresholds, despite the risk of overfitting, in agreement with our results in Sec. VII B, which showed that the extra power from residual contamination is likely more pernicious than the excess removal of power due to overfitting.

- 
- [1] C. Shane and C. Wirtanen, Pub. Lick. Obs **A1**, 22 (1967).
  - [2] V. de Lapparent, M. J. Geller, and J. P. Huchra, *Astrophys. J.* **302**, L1 (1986).
  - [3] S. J. Maddox, W. J. Sutherland, G. Efstathiou, and J. Loveday, *Mon. Not. R. Astron. Soc.* **243**, 692 (1990).
  - [4] S. L. Lumsden, R. C. Nichol, C. A. Collins, and L. Guzzo, *Mon. Not. R. Astron. Soc.* **258**, 1 (1992).
  - [5] M. Colless *et al.* (2DFGRS Collaboration), *Mon. Not. Roy. Astron. Soc.* **328**, 1039 (2001), arXiv:astro-ph/0106498 [astro-ph].
  - [6] D. H. Jones, M. A. Read, W. Saunders, M. Colless, T. Jarrett, Q. A. Parker, A. P. Fairall, T. Mauch, E. M. Sadler, F. G. Watson, D. Burton, L. A. Campbell, P. Cass, S. M. Croom, J. Dawe, K. Fiegert, L. Frankcombe, M. Hartley, J. Huchra, D. James, E. Kirby, O. Lahav, J. Lucey, G. A. Mamon, L. Moore, B. A. Peterson, S. Prior, D. Proust, K. Russell, V. Safouris, K.-I. Wakamatsu, E. Westra, and M. Williams, *Mon. Not. R. Astron. Soc.* **399**, 683 (2009), arXiv:0903.5451 [astro-ph.CO].
  - [7] M. J. Drinkwater, R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, *et al.*, *Mon. Not. Roy. Astron. Soc.* **401**, 1429 (2010), arXiv:0911.4246 [astro-ph.CO].
  - [8] D. G. York *et al.* (SDSS Collaboration), *Astron. J.* **120**, 1579 (2000), arXiv:astro-ph/0006396 [astro-ph].
  - [9] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouse, J. E. Bautista, A. r. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, A. S. Bolton, A. Borde, J. Bovy, W. N. Brandt, H. Brewington, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope, R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, T. Delubac, R. de Putter, S. Dhital, A. Ealet, G. L. Ebelke, D. J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, R. Génova-Santos, J. E. Gunn, H. Guo, D. Haggard, P. B. Hall, J.-C. Hamilton, B. Harris, D. W. Harris, S. Ho, D. W. Hogg, D. Holder, K. Honscheid, J. Huehnerhoff, B. Jordan, W. P. Jordan, G. Kauffmann, E. A. Kazin, D. Kirkby, M. A. Klaene, J.-P. Kneib, J.-M. Le Goff, K.-G. Lee, D. C. Long, C. P. Loomis, B. Lundgren, R. H. Lupton, M. A. G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, K. L. Masters, C. K. McBride, P. McDonald, I. D. McGreer, R. G. McMahon, O. Mena, J. Miralda-Escudé, A. D. Montero-Dorta, F. Montesano, D. Muna, A. D. Myers, T. Naugle, R. C. Nichol, P. Noterdaeme, S. E. Nuza, M. D. Olmstead, A. Oravetz, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, W. J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M. M. Pieri, F. Prada, A. M. Price-Whelan, M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, C. M. Rockosi, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martín, L. Samushia, A. G. Sánchez, C. Sayres, S. J. Schmidt, D. P. Schneider, C. G. Scóccola, H.-J. Seo, A. Sheldon, E. Sheldon, Y. Shen, Y. Shu, A. Slosar, S. A. Smee, S. A. Sneden, F. Stauffer, O. Steele, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, M. Tanaka, D. Thomas, J. L. Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D. A. Wake, M. Watson, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, W. M. Wood-Vasey, C. Yeh, I. Zehavi, G.-B. Zhao, and Z. Zheng, *Astro. Journal* **145**, 10 (2013), arXiv:1208.0022 [astro-ph.CO].

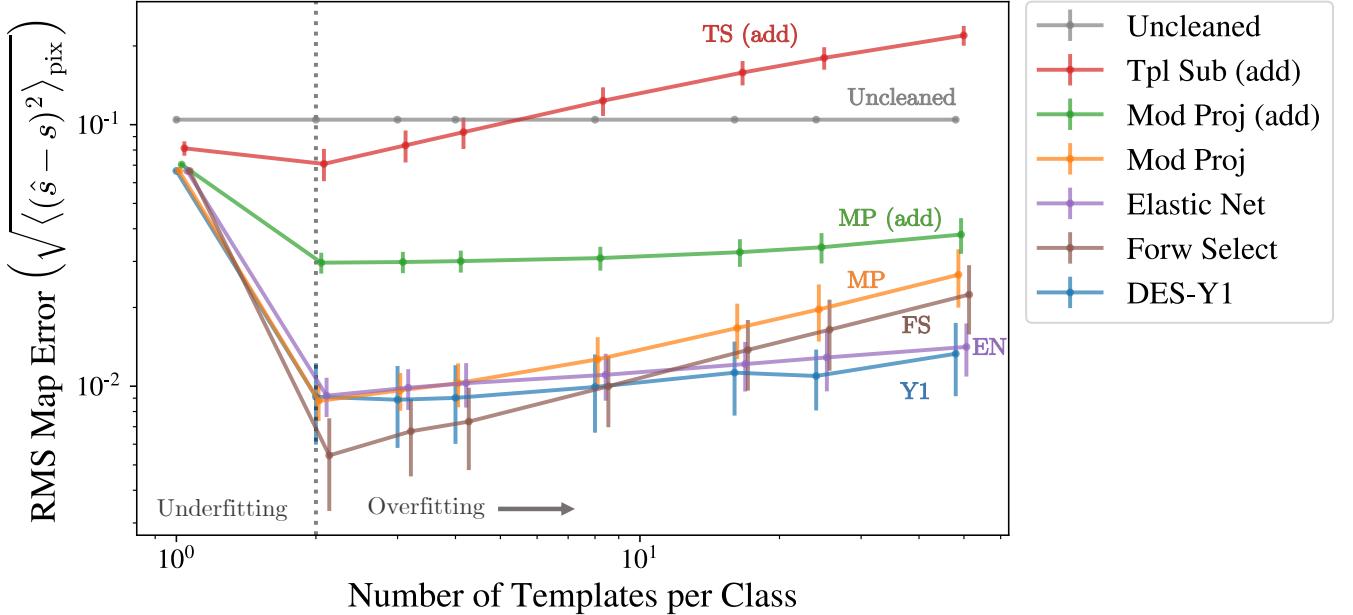


FIG. 16. Same as Fig. 10 but showing the RMSE in the estimated overdensity map for each method, rather than error in the power spectrum. Trends are very similar. See Sec. VII B for details.

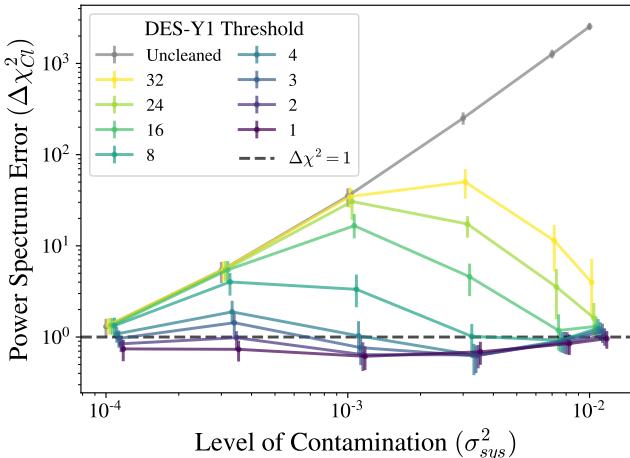


FIG. 17. Dependence of the power spectrum error ( $\Delta\chi_{Cl}^2$ ) on the level of contamination  $\sigma_{sys}^2$  ( $x$ -axis), and on the stopping criterion  $\Delta\chi^2/\Delta\chi_0^2$  used for the DES-Y1 method (colors). Points are offset for clarity. For comparison, the variance across pixels from the true overdensity in each bin ranges from  $\sigma_{sig}^2 \in [0.075, 0.122]$  for the 5 redshift bins of our fiducial survey, corresponding to factors of  $7.5 - 1220 \times$  larger than  $\sigma_{sys}^2$  for the points shown.

- [10] H. Aihara, N. Arimoto, R. Armstrong, S. Arnouts, N. A. Bahcall, S. Bickerton, J. Bosch, K. Bundy, P. L. Capak, J. H. H. Chan, M. Chiba, J. Coupon, E. Egami, M. Enoki, F. Finet, H. Fujimori, S. Fujimoto, H. Furusawa, J. Furusawa, T. Goto, A. Goulding, J. P. Greco, J. E. Greene, J. E. Gunn, T. Hamana, Y. Harikane,

Y. Hashimoto, T. Hattori, M. Hayashi, Y. Hayashi, K. G. Helminak, R. Higuchi, C. Hikage, P. T. P. Ho, B.-C. Hsieh, K. Huang, S. Huang, H. Ikeda, M. Imanishi, A. K. Inoue, K. Iwasawa, I. Iwata, A. T. Jaellani, H.-Y. Jian, Y. Kamata, H. Karoji, N. Kashikawa, N. Katayama, S. Kawanomoto, I. Kayo, J. Koda, M. Koike, T. Kojima, Y. Komiyama, A. Konno, S. Koshida, Y. Koyama, H. Kusakabe, A. Leauthaud, C.-H. Lee, L. Lin, Y.-T. Lin, R. H. Lupton, R. Mandelbaum, Y. Matsuo, E. Medezinski, S. Mineo, S. Miyama, H. Miyatake, S. Miyazaki, R. Momose, A. More, S. More, Y. Moritani, T. J. Moriya, T. Morokuma, S. Mukae, R. Murata, H. Murayama, T. Nagao, F. Nakata, M. Niida, H. Niikura, A. J. Nishizawa, Y. Obuchi, M. Oguri, Y. Oishi, N. Okabe, S. Okamoto, Y. Okura, Y. Ono, M. Onodera, M. Onoue, K. Osato, M. Ouchi, P. A. Price, T.-S. Pyo, M. Sako, M. Sawicki, T. Shibuya, K. Shimasaku, A. Shimono, M. Shirasaki, J. D. Silverman, M. Simet, J. Speagle, D. N. Spergel, M. A. Strauss, Y. Sugahara, N. Sugiyama, Y. Suto, S. H. Suyu, N. Suzuki, P. J. Tait, M. Takada, T. Takata, N. Tamura, M. M. Tanaka, M. Tanaka, M. Tanaka, Y. Tanaka, T. Terai, Y. Terashima, Y. Toba, N. Tominaga, J. Toshikawa, E. L. Turner, T. Uchida, H. Uchiyama, K. Umetsu, F. Uraguchi, Y. Urata, T. Usuda, Y. Utsumi, S.-Y. Wang, W.-H. Wang, K. C. Wong, K. Yabe, Y. Yamada, H. Yamanoi, N. Yasuda, S. Yeh, A. Yonehara, and S. Yuma, Publ. Astr. Soc. Japan **70**, S4 (2018), arXiv:1704.05858 [astro-ph.IM].

- [11] J. T. A. de Jong, G. A. Verdoes Kleijn, D. R. Boohorn, H. Buddelmeijer, M. Capaccioli, F. Getman, A. Grado, E. Helmich, Z. Huang, N. Irisarri, K. Kuijken, F. La Barbera, J. P. McFarland, N. R. Napolitano, M. Radovich, G. Sikkema, E. A. Valentijn, K. G. Bege-

- man, M. Brescia, S. Cavaoti, A. Choi, O.-M. Cordes, G. Covone, M. Dall’Ora, H. Hildebrandt, G. Longo, R. Nakajima, M. Paolillo, E. Puddu, A. Rifatto, C. Tortora, E. van Uitert, A. Buddendiek, J. Harnois-Déraps, T. Erben, M. B. Eriksson, C. Heymans, H. Hoekstra, B. Joachimi, T. D. Kitching, D. Klaes, L. V. E. Koopmans, F. Köhlinger, N. Roy, C. Sifón, P. Schneider, W. J. Sutherland, M. Viola, and W.-J. Vriend, *A&A* **582**, A62 (2015), arXiv:1507.00742 [astro-ph.CO].
- [12] T. M. C. Abbott *et al.* (DES), *Phys. Rev.* **D98**, 043526 (2018), arXiv:1708.01530 [astro-ph.CO].
- [13] L. Anderson *et al.* (BOSS), *Mon. Not. Roy. Astron. Soc.* **441**, 24 (2014), arXiv:1312.4877 [astro-ph.CO].
- [14] S. Alam *et al.* (BOSS), *Mon. Not. Roy. Astron. Soc.* **470**, 2617 (2017), arXiv:1607.03155 [astro-ph.CO].
- [15] T. Abbott *et al.* (DES), *Phys. Rev. Lett.* **122**, 171301 (2019), arXiv:1811.02375 [astro-ph.CO].
- [16] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland, O. Ballester, C. Baltay, L. Beaufore, C. Bebek, T. C. Beers, E. F. Bell, J. L. Bernal, R. Besuner, F. Beutler, C. Blake, H. Bleuler, M. Blomqvist, R. Blum, A. S. Bolton, C. Briceno, D. Brooks, J. R. Brownstein, E. Buckley-Geer, A. Burden, E. Burtin, N. G. Busca, R. N. Cahn, Y.-C. Cai, L. Cardiel-Sas, R. G. Carlberg, P.-H. Carton, R. Casas, F. J. Castander, J. L. Cervantes-Cota, T. M. Claybaugh, M. Close, C. T. Coker, S. Cole, J. Comparat, A. P. Cooper, M. C. Cousinou, M. Crocce, J.-G. Cuby, D. P. Cunningham, T. M. Davis, K. S. Dawson, A. de la Macorra, J. De Vicente, T. Delubac, M. Derwent, A. Dey, G. Dhungana, Z. Ding, P. Doel, Y. T. Duan, A. Ealet, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, S. Escoffier, M. Evatt, P. Fagrelius, X. Fan, K. Fanning, A. Farahi, J. Farihi, G. Favole, Y. Feng, E. Fernandez, J. R. Findlay, D. P. Finkbeiner, M. J. Fitzpatrick, B. Flaugher, S. Flender, A. Font-Ribera, J. E. Forero-Romero, P. Fosalba, C. S. Frenk, M. Fumagalli, B. T. Gaensicke, G. Gallo, J. Garcia-Bellido, E. Gaztanaga, N. Pietro Gentile Fusillo, T. Gerard, I. Gershkovich, T. Giannantonio, D. Gillet, G. Gonzalez-de-Rivera, V. Gonzalez-Perez, S. Gott, O. Graur, G. Gutierrez, J. Guy, S. Habib, H. Heetderks, I. Heetderks, K. Heitmann, W. A. Hellwing, D. A. Herrera, S. Ho, S. Holland, K. Honscheid, E. Huff, T. A. Hutchinson, D. Huterer, H. S. Hwang, J. M. Illa Laguna, Y. Ishikawa, D. Jacobs, N. Jeffrey, P. Jelinsky, E. Jennings, L. Jiang, J. Jimenez, J. Johnson, R. Joyce, E. Jullo, S. Juneau, S. Kama, A. Karcher, S. Karkar, R. Kehoe, N. Kennamer, S. Kent, M. Kilbinger, A. G. Kim, D. Kirkby, T. Kisner, E. Kitaniidis, J.-P. Kneib, S. Koposov, E. Kovacs, K. Koyama, A. Kremin, R. Kron, L. Kronig, A. Kueter-Young, C. G. Lacey, R. Lafever, O. Lahav, A. Lambert, M. Lampton, M. Landriau, D. Lang, T. R. Lauer, J.-M. Le Goff, L. Le Guillou, A. Le Van Suu, J. H. Lee, S.-J. Lee, D. Leitner, M. Lesser, M. E. Levi, B. L’Huillier, B. Li, M. Liang, H. Lin, E. Linder, S. R. Loebman, Z. Lukic, J. Ma, N. MacCrann, C. Magneville, L. Makarem, M. Manera, C. J. Manser, R. Marshall, P. Martini, R. Massey, T. Matheson, J. McCauley, P. McDonald, I. D. McGreer, A. Meisner, N. Metcalfe, T. N. Miller, R. Miquel, J. Moustakas, A. Myers, M. Naik, J. A. Newman, R. C. Nichol, A. Nicola, L. Nico-
- lati da Costa, J. Nie, G. Niz, P. Norberg, B. Nord, D. Norman, P. Nugent, T. O’Brien, M. Oh, K. A. G. Olsen, C. Padilla, H. Padmanabhan, N. Padmanabhan, N. Palanque-Delabrouille, A. Palmese, D. Papalardo, I. Pâris, C. Park, A. Patej, J. A. Peacock, H. V. Peiris, X. Peng, W. J. Percival, S. Perruchot, M. M. Pieri, R. Pogge, J. E. Pollack, C. Poppett, F. Prada, A. Prakash, R. G. Probst, D. Rabinoowitz, A. Raichoor, C. H. Ree, A. Refregier, X. Regal, B. Reid, K. Reil, M. Rezaie, C. M. Rockosi, N. Roe, S. Ronayette, A. Roodman, A. J. Ross, N. P. Ross, G. Rossi, E. Rozo, V. Ruhlmann-Kleider, E. S. Rykoff, C. Sabiu, L. Samushia, E. Sanchez, J. Sanchez, D. J. Schlegel, M. Schneider, M. Schubnell, A. Secroun, U. Seljak, H.-J. Seo, S. Serrano, A. Shafieloo, H. Shan, R. Sharples, M. J. Sholl, W. V. Shourt, J. H. Silber, D. R. Silva, M. M. Sirk, A. Slosar, A. Smith, G. F. Smoot, D. Som, Y.-S. Song, D. Sprayberry, R. Staten, A. Stefanik, G. Tarle, S. Sien Tie, J. L. Tinker, R. Tojeiro, F. Valdes, O. Valenzuela, M. Valluri, M. Vargas-Magana, L. Verde, A. R. Walker, J. Wang, Y. Wang, B. A. Weaver, C. Weaverdyck, R. H. Wechsler, D. H. Weinberg, M. White, Q. Yang, C. Yeche, T. Zhang, G.-B. Zhao, Y. Zheng, X. Zhou, Z. Zhou, Y. Zhu, H. Zou, and Y. Zu, arXiv e-prints , arXiv:1611.00036 (2016), arXiv:1611.00036 [astro-ph.IM].
- [17] L. D. E. S. Collaboration, “Large synoptic survey telescope: Dark energy science collaboration,” (2012), arXiv:1211.0310 [astro-ph.CO].
- [18] O. Doré, J. Bock, M. Ashby, P. Capak, A. Cooray, R. de Putter, T. Eifler, N. Flagey, Y. Gong, S. Habib, K. Heitmann, C. Hirata, W.-S. Jeong, R. Katti, P. Korngut, E. Krause, D.-H. Lee, D. Masters, P. Mauskopf, G. Melnick, B. Mennesson, H. Nguyen, K. Öberg, A. Pullen, A. Raccaelli, R. Smith, Y.-S. Song, V. Tolls, S. Unwin, T. Venumadhav, M. Viero, M. Werner, and M. Zemcov, arXiv e-prints , arXiv:1412.4872 (2014), arXiv:1412.4872 [astro-ph.CO].
- [19] L. Amendola, S. Appleby, A. Avgoustidis, D. Bacon, T. Baker, M. Baldi, N. Bartolo, A. Blanchard, C. Bonvin, and et al., *Living Reviews in Relativity* **21** (2018), 10.1007/s41114-017-0010-3.
- [20] D. Spergel, N. Gehrels, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, J. Centrella, W. Traub, C. Baltay, J. Colbert, D. Bennett, A. Kiessling, B. Macintosh, J. Merten, M. Mortonson, M. Penny, E. Rozo, D. Savransky, K. Stapelfeldt, Y. Zu, C. Baker, E. Cheng, D. Content, J. Dooley, M. Foote, R. Goullioud, K. Grady, C. Jackson, J. Krulk, M. Levine, M. Melton, C. Peddie, J. Ruffa, and S. Shaklan, “Wide-field infrared survey telescope-astrophysics focused telescope assets wfirst-afta final report,” (2013), arXiv:1305.5422 [astro-ph.IM].
- [21] M. G. Hauser and P. J. E. Peebles, *Astrophys. J.* **185**, 757 (1973).
- [22] P. J. E. Peebles and M. G. Hauser, *Astrophys. J. Suppl.* **28**, 19 (1974).
- [23] M. Davis and P. J. E. Peebles, *Astrophys. J.* **267**, 465 (1983).
- [24] W. Saunders, M. Rowan-Robinson, and A. Lawrence, *Mon. Not. R. Astron. Soc.* **258**, 134 (1992).
- [25] K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil, and

- J. P. Huchra, *Astrophys. J.* **402**, 42 (1993).
- [26] K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil, and J. P. Huchra, *Astrophys. J.* **402**, 42 (1993).
- [27] J. A. Peacock and D. Nicholson, *Mon. Not. R. Astron. Soc.* **253**, 307 (1991).
- [28] H. A. Feldman, N. Kaiser, and J. A. Peacock, *Astrophys. J.* **426**, 23 (1994), arXiv:astro-ph/9304022 [astro-ph].
- [29] C. M. Baugh and G. Efstathiou, *Mon. Not. R. Astron. Soc.* **265**, 145 (1993).
- [30] C. M. Baugh, *Mon. Not. R. Astron. Soc.* **280**, 267 (1996), arXiv:astro-ph/9512011 [astro-ph].
- [31] C. J. Miller and D. J. Batuski, *Astrophys. J.* **551**, 635 (2001), arXiv:astro-ph/0002295 [astro-ph].
- [32] S. Dodelson and E. Gaztañaga, *Mon. Not. R. Astron. Soc.* **312**, 774 (2000), arXiv:astro-ph/9906289 [astro-ph].
- [33] D. Huterer, L. Knox, and R. C. Nichol, *Astrophys. J.* **555**, 547 (2001), arXiv:astro-ph/0011069 [astro-ph].
- [34] D. J. Eisenstein and M. Zaldarriaga, *Astrophys. J.* **546**, 2 (2001), arXiv:astro-ph/9912149 [astro-ph].
- [35] J. A. Peacock, S. Cole, P. Norberg, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. D. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, W. J. Percival, B. A. Peterson, I. Price, W. Sutherland, and K. Taylor, *Nature (London)* **410**, 169 (2001), arXiv:astro-ph/0103143 [astro-ph].
- [36] S. Dodelson, V. K. Narayanan, M. Tegmark, R. Scranton, T. Budavári, A. Connolly, I. Csabai, D. Eisenstein, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, J. Loveday, R. C. Nichol, L. O'Connell, R. Scoccimarro, R. K. Sheth, A. Stebbins, M. A. Strauss, A. S. Szalay, I. Szapudi, M. S. Vogeley, I. Zehavi, J. Annis, N. A. Bahcall, J. Brinkman, M. Doi, M. Fukugita, G. Hennessy, Ž. Ivezić, G. R. Knapp, P. Kunszt, D. Q. Lamb, B. C. Lee, R. H. Lupton, J. A. Munn, J. Peoples, J. R. Pier, C. Rockosi, D. Schlegel, C. Stoughton, D. L. Tucker, B. Yanny, and D. G. York, *Astrophys. J.* **572**, 140 (2002), arXiv:astro-ph/0107421 [astro-ph].
- [37] A. J. Connolly, R. Scranton, D. Johnston, S. Dodelson, D. J. Eisenstein, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, S. Kent, J. Loveday, R. C. Nichol, L. O'Connell, M. Postman, R. Scoccimarro, R. K. Sheth, A. Stebbins, M. A. Strauss, A. S. Szalay, I. Szapudi, M. Tegmark, M. S. Vogeley, I. Zehavi, J. Annis, N. A. Bahcall, J. Brinkmann, I. Csabai, M. Doi, M. Fukugita, G. S. Hennessy, R. Hindsley, T. Ichikawa, Ž. Ivezić, R. S. J. Kim, G. R. Knapp, P. Kunszt, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, J. Munn, J. Peoples, J. Pier, C. Rockosi, D. Schlegel, C. Stoughton, D. L. Tucker, B. Yanny, and D. G. York, *Astrophys. J.* **579**, 42 (2002), arXiv:astro-ph/0107417 [astro-ph].
- [38] W. J. Percival, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, S. Moody, P. Norberg, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor, *Mon. Not. R. Astron. Soc.* **327**, 1297 (2001), arXiv:astro-ph/0105252 [astro-ph].
- [39] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, A. Berlind, T. Budavári, A. Connolly, D. J. Eisenstein, D. Finkbeiner, J. A. Frieman, J. E. Gunn, A. J. S. Hamilton, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R. C. Nichol, J. P. Ostriker, A. Pope, R. Scranton, U. Seljak, R. K. Sheth, A. Stebbins, A. S. Szalay, I. Szapudi, L. Verde, Y. Xu, J. Annis, N. A. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, I. Gott, J. Richard, G. Hennessy, D. W. Hogg, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, P. Kunszt, J. A. Munn, L. O'Connell, J. Peoples, J. R. Pier, M. Richmond, C. Rockosi, D. P. Schneider, C. Stoughton, D. L. Tucker, D. E. Vand en Berk, B. Yanny, D. G. York, and SDSS Collaboration, *Astrophys. J.* **606**, 702 (2004), arXiv:astro-ph/0310725 [astro-ph].
- [40] C. Blake, *Mon. Not. R. Astron. Soc.* **489**, 153 (2019), arXiv:1902.07439 [astro-ph.CO].
- [41] J. Elvin-Poole *et al.* (DES), *Phys. Rev. D* **98**, 042006 (2018), arXiv:1708.01536 [astro-ph.CO].
- [42] P. J. E. Peebles and E. J. Groth, *Astrophys. J.* **196**, 1 (1975).
- [43] H. A. Feldman, J. A. Frieman, J. N. Fry, and R. Scoccimarro, *Phys. Rev. Lett.* **86**, 1434 (2001), arXiv:astro-ph/0010205 [astro-ph].
- [44] R. Scoccimarro, H. A. Feldman, J. N. Fry, and J. A. Frieman, *Astrophys. J.* **546**, 652 (2001), arXiv:astro-ph/0004087 [astro-ph].
- [45] A. Cooray and W. Hu, *Astrophys. J.* **548**, 7 (2001), arXiv:astro-ph/0004151 [astro-ph].
- [46] E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro, *Phys. Rev. D* **74**, 023522 (2006), arXiv:astro-ph/0604505 [astro-ph].
- [47] L. Verde, A. F. Heavens, W. J. Percival, S. Matarrese, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, P. Norberg, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor, *Mon. Not. R. Astron. Soc.* **335**, 432 (2002), arXiv:astro-ph/0112161 [astro-ph].
- [48] F. A. Marín, C. Blake, G. B. Poole, C. K. McBride, S. Brough, M. Colless, C. Contreras, W. Couch, D. J. Croton, S. Croom, T. Davis, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I. h. Li, B. Madore, D. C. Martin, K. Pimbblet, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee, *Mon. Not. R. Astron. Soc.* **432**, 2654 (2013), arXiv:1303.6644 [astro-ph.CO].
- [49] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, *Mon. Not. R. Astron. Soc.* **451**, 539 (2015), arXiv:1407.5668 [astro-ph.CO].
- [50] Z. Slepian and D. J. Eisenstein, *Mon. Not. R. Astron. Soc.* **454**, 4142 (2015), arXiv:1506.02040 [astro-ph.CO].
- [51] Z. Slepian, D. J. Eisenstein, F. Beutler, A. J. Cuesta, J. Ge, H. Gil-Marín, S. Ho, F.-S. Kitaura, C. K. McBride, R. C. Nichol, W. J. Percival, S. Rodríguez-

- Torres, A. J. Ross, R. Scoccimarro, H.-J. Seo, J. Tinker, R. Tojeiro, and M. Vargas-Magaña, arXiv e-prints , arXiv:1512.02231 (2015), arXiv:1512.02231 [astro-ph.CO].
- [52] B. Jain and L. Van Waerbeke, *Astrophys. J. Lett.* **530**, L1 (2000), arXiv:astro-ph/9910459 [astro-ph].
- [53] L. Marian, S. Hilbert, R. E. Smith, P. Schneider, and V. Desjacques, *Astrophys. J. Lett.* **728**, L13 (2011), arXiv:1010.5242 [astro-ph.CO].
- [54] J. Liu, A. Petri, Z. Haiman, L. Hui, J. M. Kratochvil, and M. May, *Phys. Rev. D* **91**, 063507 (2015), arXiv:1412.0757 [astro-ph.CO].
- [55] S. D. M. White, *Mon. Not. R. Astron. Soc.* **186**, 145 (1979).
- [56] J. N. Fry, *Astrophys. J.* **306**, 358 (1986).
- [57] R. Biswas, E. Alizadeh, and B. D. Wandelt, *Phys. Rev. D* **82**, 023002 (2010), arXiv:1002.0014 [astro-ph.CO].
- [58] E. G. P. Bos, R. van de Weygaert, K. Dolag, and V. Pettorino, *Mon. Not. R. Astron. Soc.* **426**, 440 (2012), arXiv:1205.4238 [astro-ph.CO].
- [59] A. Pisani, P. M. Sutter, N. Hamaus, E. Alizadeh, R. Biswas, B. D. Wandelt, and C. M. Hirata, *Phys. Rev. D* **92**, 083531 (2015), arXiv:1503.07690 [astro-ph.CO].
- [60] S. Nadathur and S. Hotchkiss, *Mon. Not. R. Astron. Soc.* **454**, 2228 (2015), arXiv:1504.06510 [astro-ph.CO].
- [61] F. Leclercq, J. Jasche, P. M. Sutter, N. Hamaus, and B. Wandelt, *JCAP* **2015**, 047 (2015), arXiv:1410.0355 [astro-ph.CO].
- [62] O. Friedrich, D. Gruen, J. DeRose, D. Kirk, E. Krause, T. McClintock, E. Rykoff, S. Seitz, R. Wechsler, G. Bernstein, and et al., *Physical Review D* **98** (2018), 10.1103/physrevd.98.023508.
- [63] R. K. Sheth, *Mon. Not. R. Astron. Soc.* **364**, 796 (2005), arXiv:astro-ph/0511772 [astro-ph].
- [64] M. White, *Journal of Cosmology and Astroparticle Physics* **2016**, 057057 (2016).
- [65] M. White and N. Padmanabhan, *Mon. Not. R. Astron. Soc.* **395**, 2381 (2009), arXiv:0812.4288 [astro-ph].
- [66] O. H. E. Philcox, E. Massara, and D. N. Spergel, arXiv e-prints , arXiv:2006.10055 (2020), arXiv:2006.10055 [astro-ph.CO].
- [67] I. Gott, J. Richard, D. H. Weinberg, and A. L. Melott, *Astrophys. J.* **319**, 1 (1987).
- [68] J. M. Kratochvil, E. A. Lim, S. Wang, Z. Haiman, M. May, and K. Huffenberger, *Phys. Rev. D* **85**, 103513 (2012), arXiv:1109.6334 [astro-ph.CO].
- [69] C. Hikage, E. Komatsu, and T. Matsubara, *Astrophys. J.* **653**, 11 (2006), arXiv:astro-ph/0607284 [astro-ph].
- [70] D. Munshi, L. van Waerbeke, J. Smidt, and P. Coles, *Mon. Not. R. Astron. Soc.* **419**, 536 (2012), arXiv:1103.1876 [astro-ph.CO].
- [71] A. Petri, J. Liu, Z. Haiman, M. May, L. Hui, and J. M. Kratochvil, *Phys. Rev. D* **91**, 103511 (2015), arXiv:1503.06214 [astro-ph.CO].
- [72] B. Mawdsley, D. Bacon, C. Chang, P. Melchior, E. Rozo, S. Seitz, N. Jeffrey, M. Gatti, E. Gaztanaga, D. Gruen, W. G. Hartley, B. Hoyle, S. Samuroff, E. Sheldon, M. A. Troxel, J. Zuntz, T. M. C. Abbott, J. Annis, E. Bertin, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, L. N. da Costa, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, A. E. Evrard, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, D. W. Gerdes, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. L. Hollowood, K. Honscheid, D. J. James, M. Jarvis, T. Jeltema, K. Kuehn, N. Kuropatkin, M. Lima, M. A. G. Maia, J. L. Marshall, R. Miquel, A. A. Plazas, A. Roodman, E. Sanchez, V. Scarpine, S. Serrano, I. Sevilla-Noarbe, M. Smith, R. C. Smith, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. L. Tucker, V. Vikram, A. R. Walker, and DES Collaboration, *Mon. Not. R. Astron. Soc.* **493**, 5662 (2020), arXiv:1905.12682 [astro-ph.CO].
- [73] E. Ally, T. Marchand, J. F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. Mallat, “New interpretable statistics for large scale structure analysis and generation,” (2020), arXiv:2006.06298 [astro-ph.CO].
- [74] S. Cheng, Y.-S. Ting, B. Mnard, and J. Bruna, “A new approach to observational cosmology using the scattering transform,” (2020), arXiv:2006.08561 [astro-ph.CO].
- [75] D. Huterer, C. E. Cunha, and W. Fang, *Mon.Not.Roy.Astron.Soc.* **432**, 2945 (2013), arXiv:1211.1015 [astro-ph.CO].
- [76] D. L. Shafer and D. Huterer, *Mon. Not. Roy. Astron. Soc.* **447**, 2961 (2015), arXiv:1410.0035 [astro-ph.CO].
- [77] N. Weaverdyck, J. Muir, and D. Huterer, *Phys. Rev. D* **97**, 043515 (2018), arXiv:1709.08661 [astro-ph.CO].
- [78] M. S. Vogeley, “Toward High-Precision Measures of Large-Scale Structure,” in *The Evolving Universe, Astrophysics and Space Science Library*, Vol. 231, edited by D. Hamilton (1998) p. 395.
- [79] R. Scranton, D. Johnston, S. Dodelson, J. A. Frieman, A. Connolly, D. J. Eisenstein, J. E. Gunn, L. Hui, B. Jain, S. Kent, J. Loveday, V. Narayanan, R. C. Nichol, L. O’Connell, R. Scoccimarro, R. K. Sheth, A. Stebbins, M. A. Strauss, A. S. Szalay, I. Szapudi, M. Tegmark, M. Vogeley, I. Zehavi, J. Annis, N. A. Bahcall, J. Brinkman, I. Csabai, R. Hindsley, Z. Ivezic, R. S. J. Kim, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. McKay, J. Munn, J. Peoples, J. Pier, G. T. Richards, C. Rockosi, D. Schlegel, D. P. Schneider, C. Stoughton, D. L. Tucker, B. Yanny, and D. G. York, *Astrophys. J.* **579**, 48 (2002), arXiv:astro-ph/0107416 [astro-ph].
- [80] T. Goto, I. Szapudi, and B. R. Granett, *Mon.Not.Roy.Astron.Soc.* **422**, L77 (2012), arXiv:1202.5306 [astro-ph.CO].
- [81] A. R. Pullen and C. M. Hirata, *Publications of the Astronomical Society of the Pacific*, Volume 125, Issue 928, pp. **705-718** (2013), 10.1086/671189, arXiv:1212.4500 [astro-ph.CO].
- [82] S. Ho, A. Cuesta, H.-J. Seo, R. de Putter, A. J. Ross, M. White, N. Padmanabhan, S. Saito, D. J. Schlegel, E. Schlafly, U. Seljak, C. Hernández-Monteagudo, A. G. Sánchez, W. J. Percival, M. Blanton, R. Skibba, D. Schneider, B. Reid, O. Mena, M. Viel, D. J. Eisenstein, F. Prada, B. A. Weaver, N. Bahcall, D. Bizyaev, H. Brewington, J. Brinkman, L. N. da Costa, J. R. Gott, E. Malanushenko, V. Malanushenko, B. Nichol, D. Oravetz, K. Pan, N. Palanque-Delabrouille, N. P. Ross, A. Simmons, F. de Simoni, S. Snedden, and C. Yeeche, *The Astrophysical Journal* **761**, 14 (2012).
- [83] S. Ho, N. Agarwal, A. D. Myers, R. Lyons, A. Disbrow, et al., 1311.2597 (2013), arXiv:1311.2597 [astro-ph.CO].
- [84] N. Agarwal, S. Ho, and S. Shandera, *JCAP* **1402**, 038 (2014), arXiv:1311.2606 [astro-ph.CO].
- [85] T. Giannantonio, A. J. Ross, W. J. Percival, R. Critten-

- den, D. Bacher, *et al.*, Phys.Rev. **D89**, 023511 (2014), arXiv:1303.1349 [astro-ph.CO].
- [86] N. Agarwal, S. Ho, A. D. Myers, H.-J. Seo, A. J. Ross, *et al.*, JCAP **1404**, 007 (2014), arXiv:1309.2954 [astro-ph.CO].
- [87] B. Leistedt *et al.* (DES), Astrophys. J. Suppl. **226**, 24 (2016), arXiv:1507.05647 [astro-ph.CO].
- [88] A. J. Ross *et al.*, Mon. Not. Roy. Astron. Soc. **417**, 1350 (2011), arXiv:1105.2320 [astro-ph.CO].
- [89] B. Leistedt, H. V. Peiris, D. J. Mortlock, A. Benoit-Lévy, and A. Pontzen, Mon.Not.Roy.Astron.Soc. **435**, 1857 (2013), arXiv:1306.0005 [astro-ph.CO].
- [90] B. Leistedt and H. V. Peiris, Mon.Not.Roy.Astron.Soc. **444**, 2 (2014), arXiv:1404.6530 [astro-ph.CO].
- [91] T. Delubac, A. Raichoor, J. Comparat, S. Jouvel, J.-P. Kneib, C. Yche, H. Zou, J. R. Brownstein, F. B. Abdalla, . Dawson, E. Jullo, A. D. Myers, J. A. Newman, W. J. Percival, F. Prada, A. J. Ross, D. P. Schneider, X. Zhou, Z. Zhou, and G. Zhu, Monthly Notices of the Royal Astronomical Society **465**, 1831 (2016), <https://academic.oup.com/mnras/article-pdf/465/2/1831/8364762/stw2741.pdf>.
- [92] E. S. Rykoff, E. Rozo, and R. Keisler, Submitted to: Astrophysical Journal (2015), arXiv:1509.00870 [astro-ph.IM].
- [93] E. Suchyta *et al.* (DES), Mon. Not. Roy. Astron. Soc. **457**, 786 (2016), arXiv:1507.08336 [astro-ph.GA].
- [94] A. Prakash, T. C. Licquia, J. A. Newman, A. J. Ross, A. D. Myers, K. S. Dawson, J.-P. Kneib, W. J. Percival, J. E. Bautista, J. Comparat, J. L. Tinker, D. J. Schlegel, R. Tojeiro, S. Ho, D. Lang, S. M. Rao, C. K. McBride, G. B. Zhu, J. R. Brownstein, S. Bailey, A. S. Bolton, T. Delubac, V. Mariappan, M. R. Blanton, B. Reid, D. P. Schneider, H.-J. Seo, A. C. Rosell, and F. Prada, The Astrophysical Journal Supplement Series **224**, 34 (2016).
- [95] H. Awan, E. Gawiser, P. Kurczynski, R. L. Jones, H. Zhan, N. D. Padilla, A. M. Muñoz Arancibia, A. Orsi, S. A. Cora, and P. Yoachim, Astrophys. J. **829**, 50 (2016), arXiv:1605.00555 [astro-ph.CO].
- [96] B. Kalus, W. J. Percival, D. J. Bacon, and L. Samushia, Monthly Notices of the Royal Astronomical Society **463**, 467 (2016), <http://oup.prod.sis.lan/mnras/article-pdf/463/1/467/18472203/stw2008.pdf>.
- [97] J. E. Bautista, M. Vargas-Magaa, K. S. Dawson, W. J. Percival, J. Brinkmann, J. Brownstein, B. Camacho, J. Comparat, H. Gil-Marn, E.-M. Mueller, and et al., The Astrophysical Journal **863**, 110 (2018).
- [98] B. Kalus, W. J. Percival, D. J. Bacon, E. M. Mueller, L. Samushia, L. Verde, A. J. Ross, and J. L. Bernal, Mon. Not. Roy. Astron. Soc. **482**, 453 (2019), arXiv:1806.02789 [astro-ph.CO].
- [99] H. Awan and E. Gawiser, “Angular correlation function estimators accounting for contamination from probabilistic distance measurements,” (2019), arXiv:1911.07832 [astro-ph.CO].
- [100] M. Rezaie, H.-J. Seo, A. J. Ross, and R. C. Bunescu, (2019), arXiv:1907.11355 [astro-ph.CO].
- [101] E. L. Wagoner, E. Rozo, and X. Fang, “Linear systematics mitigation in galaxy clustering in the dark energy survey year 1 data,” (2020), in Preparation.
- [102] H. Kong, K. J. Burleigh, A. Ross, J. Moustakas, C.-H. Chuang, J. Comparat, A. de Mattia, H. du Mas des Bourboux, K. Honscheid, S. Lin, A. Raichoor, G. Rossi, and C. Zhao, “Removing imaging systematics from galaxy clustering measurements with obiwan : Application to the sdss-iv extended baryon oscillation spectroscopic survey emission line galaxy sample,” (2020), arXiv:2007.08992 [astro-ph.CO].
- [103] A. J. Ross, J. Bautista, R. Tojeiro, S. Alam, S. Bailey, E. Burtin, J. Comparat, K. S. Dawson, A. de Mattia, H. du Mas des Bourboux, H. Gil-Marn, J. Hou, H. Kong, B. W. Lyke, F. G. Mohammad, J. Moustakas, E.-M. Mueller, A. D. Myers, W. J. Percival, A. Raichoor, M. Rezaie, H.-J. Seo, A. Smith, J. L. Tinker, P. Zarrouk, C. Zhao, G.-B. Zhao, D. Bizyaev, J. Brinkmann, J. R. Brownstein, A. C. Rosell, S. Chabanier, P. D. Choi, C.-H. Chuang, I. Cruz-Gonzalez, A. de la Macorra, S. de la Torre, S. Escouffier, S. Froementeau, A. Higley, E. Jullo, J.-P. Kneib, J. N. McLane, A. Muoz-Gutierrez, R. Neveux, J. A. Newman, C. Nitschelm, N. Palanque-Delabrouille, R. Paviot, A. R. Pullen, G. Rossi, V. Ruhlmann-Kleider, D. P. Schneider, M. V. Magaa, M. Vivek, and Y. Zhang, “The completed sdss-iv extended baryon oscillation spectroscopic survey: Large-scale structure catalogs for cosmological analysis,” (2020), arXiv:2007.09000 [astro-ph.CO].
- [104] J. Jasche and F. S. Kitaura, Mon. Not. Roy. Astron. Soc. **407**, 29 (2010), arXiv:0911.2496 [astro-ph.CO].
- [105] J. Jasche and B. D. Wandelt, Mon. Not. Roy. Astron. Soc. **432**, 894 (2013), arXiv:1203.3639 [astro-ph.CO].
- [106] F.-S. Kitaura, Mon. Not. Roy. Astron. Soc. **429**, 84 (2013), arXiv:1203.4184 [astro-ph.CO].
- [107] H. Wang, H. J. Mo, X. Yang, Y. P. Jing, and W. P. Lin, Astrophys. J. **794**, 94 (2014), arXiv:1407.3451 [astro-ph.CO].
- [108] J. Jasche, F. Leclercq, and B. D. Wandelt, JCAP **1501**, 036 (2015), arXiv:1409.6308 [astro-ph.CO].
- [109] H. Wang, H. J. Mo, X. Yang, Y. Zhang, J. Shi, Y. P. Jing, C. Liu, S. Li, X. Kang, and Y. Gao, Astrophys. J. **831**, 164 (2016), arXiv:1608.01763 [astro-ph.CO].
- [110] C. Modi, M. White, A. Slosar, and E. Castorina, JCAP **1911**, 023 (2019), arXiv:1907.02330 [astro-ph.CO].
- [111] N. Porqueres, D. Kodi Ramanah, J. Jasche, and G. Lavaux, Astronomy & Astrophysics **624**, A115 (2019).
- [112] J. Muir and D. Huterer, Phys. Rev. **D94**, 043503 (2016), arXiv:1603.06586 [astro-ph.CO].
- [113] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, Astrophys. J. **622**, 759 (2005), arXiv:astro-ph/0409513 [astro-ph].
- [114] A. J. Ross *et al.* (BOSS), Mon. Not. Roy. Astron. Soc. **464**, 1168 (2017), arXiv:1607.03145 [astro-ph.CO].
- [115] P. Laurent *et al.*, JCAP **1707**, 017 (2017), arXiv:1705.04718 [astro-ph.CO].
- [116] M. Ata *et al.*, Mon. Not. Roy. Astron. Soc. **473**, 4773 (2018), arXiv:1705.06373 [astro-ph.CO].
- [117] C. Hernández-Monteagudo *et al.*, Mon. Not. Roy. Astron. Soc. **438**, 1724 (2014), arXiv:1303.4302 [astro-ph.CO].
- [118] A. Nicola, A. Refregier, and A. Amara, Phys. Rev. **D94**, 083517 (2016), arXiv:1607.01014 [astro-ph.CO].
- [119] M. Crocce *et al.* (DES), Mon. Not. Roy. Astron. Soc. **455**, 4301 (2016), arXiv:1507.05360 [astro-ph.CO].
- [120] F. Elsner, B. Leistedt, and H. V. Peiris, Mon. Not. Roy. Astron. Soc. **456**, 2095 (2016), arXiv:1509.08933 [astro-ph.CO].

- [121] W. J. Percival (2018) arXiv:1810.04263 [astro-ph.CO].
- [122] D. Alonso, J. Sanchez, and A. Slosar (LSST Dark Energy Science), Mon. Not. Roy. Astron. Soc. **484**, 4127 (2019), arXiv:1809.09603 [astro-ph.CO].
- [123] A. Nicola, D. Alonso, J. Sanchez, A. Slosar, H. Awan, A. Broussard, J. Dunkley, Z. Gomes, E. Gawiser, R. Mandelbaum, H. Miyatake, J. A. Newman, I. Sevilla, S. Skinner, and E. Wagoner, “Tomographic galaxy clustering with the subaru hyper suprime-cam first year public data release,” (2019), arXiv:1912.08209 [astro-ph.CO].
- [124] G. B. Rybicki and W. H. Press, Astrophys. J. **398**, 169 (1992).
- [125] F. Elsner, B. Leistedt, and H. V. Peiris, Mon. Not. Roy. Astron. Soc. **465**, 1847 (2017), arXiv:1609.03577 [astro-ph.CO].
- [126] E. Hivon, K. M. Gorski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, Astrophys. J. **567**, 2 (2002), arXiv:astro-ph/0105302 [astro-ph].
- [127] D. P. Foster and E. I. George, Ann. Statist. **22**, 1947 (1994).
- [128] J.-L. Starck, D. L. Donoho, M. J. Fadili, and A. Rassat, Astronomy & Astrophysics **552**, A133 (2013).
- [129] J. Bobin, J. Starck, J. Fadili, and Y. Moudden, IEEE Transactions on Image Processing **16**, 2662 (2007).
- [130] J. Bobin, J.-L. Starck, F. Sureau, and S. Basak, Astronomy & Astrophysics **550**, A73 (2013).
- [131] S. Wagner-Carena, M. Hopkins, A. D. Rivero, and C. Dvorkin, “A novel cmb component separation method: Hierarchical generalized morphological component analysis,” (2019), arXiv:1910.08077 [astro-ph.CO].
- [132] A. Leonard, F. Lanusse, and J.-L. Starck, Monthly Notices of the Royal Astronomical Society **440**, 12811294 (2014).
- [133] Lanusse, F., Starck, J.-L., Leonard, A., and Pires, S., A&A **591**, A2 (2016).
- [134] N. Jeffrey, F. B. Abdalla, O. Lahav, F. Lanusse, J. L. Starck, A. Leonard, D. Kirk, C. Chang, E. Baxter, T. Kacprzak, S. Seitz, V. Vikram, L. White-way, T. M. C. Abbott, S. Allam, S. Avila, E. Bertin, D. Brooks, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D’Andrea, L. N. da Costa, C. Davis, J. De Viente, S. Desai, P. Doel, T. F. Eifler, A. E. Evrard, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, K. Honscheid, B. Hoyle, D. J. James, M. Jarvis, K. Kuehn, M. Lima, H. Lin, M. March, P. Melchior, F. Menanteau, R. Miquel, A. A. Plazas, K. Reil, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, A. R. Walker, and DES Collaboration, Mon. Not. R. Astron. Soc. **479**, 2871 (2018), arXiv:1801.08945 [astro-ph.CO].
- [135] H. Zou and T. Hastie, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**, 301 (2005).
- [136] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).
- [137] R. D. Cook, Technometrics **19**, 15 (1977).
- [138] R. D. Cook, Journal of the American Statistical Association **74**, 169 (1979).
- [139] A. Abergel *et al.* (Planck), Astron. Astrophys. **571**, A11 (2014), arXiv:1312.1300 [astro-ph.GA].
- [140] J. Lesgourges, (2011), arXiv:1104.2932 [astro-ph.IM].
- [141] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, Journal of Open Source Software **4**, 1298 (2019).
- [142] Y. Benjamini and Y. Hochberg, Journal of the Royal Statistical Society. Series B (Methodological) **57**, 289 (1995).
- [143] P. Coles and B. Jones, Mon. Not. R. Astron. Soc. **248**, 1 (1991).
- [144] A. Taruya, M. Takada, T. Hamana, I. Kayo, and T. Futamase, The Astrophysical Journal **571**, 638653 (2002).
- [145] S. Hilbert, J. Hartlap, and P. Schneider, Astronomy & Astrophysics **536**, A85 (2011).
- [146] H. S. Xavier, F. B. Abdalla, and B. Joachimi, Monthly Notices of the Royal Astronomical Society **459**, 36933710 (2016).