

# Local primordial non-Gaussianity from the large-scale clustering of photometric DESI luminous red galaxies

Mehdi Rezaie<sup>1</sup>, Ashley J. Ross<sup>2</sup>, Hee-Jong Seo<sup>3,5</sup>, Hui Kong<sup>2</sup>, Anna Porredon<sup>2</sup>, Lado Samushia<sup>1</sup>, Edmond Chaussidon<sup>4</sup>, Alex Krolewski<sup>6,7,8</sup>, Arnaud de Mattia<sup>4</sup>, Santiago Avila<sup>8,5</sup>, Benedict Bahr-Kalus<sup>9</sup>, Jose Bermejo-Climent<sup>7</sup>, Florian Beutler<sup>10</sup>, Klaus Honscheid<sup>2,11</sup>, Eva-Maria Mueller<sup>14,15</sup>, Adam Myers<sup>13</sup>, Nathalie Palanque-Delabrouille<sup>4,5</sup>, Will Percival<sup>6,7,8,?</sup>, Christophe Yèche<sup>4</sup>

<sup>1</sup>Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA

<sup>2</sup>Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA

<sup>3</sup>Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA

<sup>4</sup>IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>5</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>6</sup>Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>7</sup>Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada

<sup>8</sup>Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

<sup>8,5</sup>Institut de Física d'Altes Energies, The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain

<sup>9</sup>Korea Astronomy and Space Science Institute, Yuseong-gu, Daedeok-daero 776, Daejeon 34055, Republic of Korea

<sup>10</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

<sup>11</sup>Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA

<sup>13</sup>Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA

<sup>14</sup>Institute of Cosmology and Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK

<sup>15</sup>Department of Physics & Astronomy, University of Sussex, Brighton BN1 9QH, UK

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

This paper uses the angular power spectrum of luminous red galaxies (LRGs) selected from the Dark Energy Spectroscopic Instrument (DESI) imaging surveys to constrain the local primordial non-Gaussianity parameter  $f_{\text{NL}}$ . Our dataset comprises over 14 million LRG targets, spanning approximately 18,000 square degrees of the sky, with redshifts ranging from  $0.2 < z < 1.35$ . The FFTLog algorithm is employed to enable modeling the power spectrum on large scales. Galactic extinction, survey depth, and astronomical seeing are identified as the primary sources of systematic error using feature selection and cross-correlation techniques. Artificial neural networks are applied to alleviate excess clustering signals on large scales. Our treatment methods are tested and calibrated rigorously against lognormal density simulations with  $f_{\text{NL}} = 0$  and 76.9, and systematic effects included. Assuming halo bias only depends on mass, we find  $36.08 (25.03) < f_{\text{NL}} < 61.44 (75.64)$  at 68% (95%) confidence. Our constraints weaken to  $13.10 (-15.96) < f_{\text{NL}} < 69.14 (91.84)$  if the spurious fluctuations in the galaxy density field are regressed out against all available imaging systematic maps. We also test the robustness of the  $f_{\text{NL}}$  constraints against different assumptions and variations in the data analysis pipeline, only finding significant photometric calibration issues in the South Galactic Cap below declination of  $-30$ . Assuming Planck's constraint on the value of  $f_{\text{NL}}$  is accurate a priori, our best fit estimates of  $f_{\text{NL}} \sim 47 - 50$  provide evidence for some unknown calibration issues which cannot be addressed with available imaging maps. Our results show that constraining  $f_{\text{NL}}$  requires a careful analysis of imaging systematic effects and a thorough understanding of astrophysical factors that impact the measured clustering signal on large scales. **While more theoretical work and simulations are required to better understand halo assembly bias, our results motivate follow-up studies of  $f_{\text{NL}}$  with DESI spectroscopic samples, which are less prone to contamination issues and provide a more uniform selection of galaxies.**

**Key words:** cosmology: inflation - large-scale structure of the Universe

## 1 INTRODUCTION

Inflation is a widely accepted paradigm in modern cosmology that explains many important characteristics of our Universe. It predicts that the early Universe underwent a period of accelerated expansion, resulting in the observed homogeneity and isotropy of the Universe on large scales (see, e.g., Weinberg et al. 2013, for a review). After the period of inflation, the Universe entered a phase of reheating in which primordial perturbations were generated, setting the initial seeds for structure formation (Kofman et al. 1994; Bassett et al. 2006; Lyth & Liddle 2009). Although inflation is widely accepted as a convincing explanation, the characteristics of the field or fields that drive the inflationary expansion remain largely unknown in contemporary observational cosmology. While early studies of the cosmic microwave background (CMB) and large-scale structure (LSS) suggested that primordial fluctuations are both Gaussian and scale-invariant (Tegmark et al. 2004; Guth & Kaiser 2005), some alternative classes of inflationary models predict deviations from these predictions, including some levels of non-Gaussianities in the primordial gravitational field. Non-Gaussianities are a measure of the degree to which the distribution of matter in the Universe deviates from a Gaussian distribution, which would have important implications for our understanding of the early Universe and the formation of structure (see, e.g., Desjacques & Seljak 2010; Biagetti 2019).

In its simplest form, primordial non-Gaussianity (PNG) depends on the local value of the primordial potential  $\phi$  (Komatsu & Spergel 2001),

$$\Phi = \phi + f_{\text{NL}}[\phi^2 - \langle \phi^2 \rangle], \quad (1)$$

where  $f_{\text{NL}}$  is a nonlinear coupling constant and  $\phi$  is assumed to be a Gaussian random field. Local-type PNG generates a primordial bispectrum, which peaks in the squeezed limit where one of the three wave vectors is much smaller than the other two. This means that one of the modes is on a much larger scale than the other two, and this mode couples with the other two modes to generate a non-Gaussian signal, which then affects the local number density of galaxies. This coupling induces a distinct scale-dependent bias in the galaxy distribution, which can be detected and measured through two-point clustering statistics (Dalal et al. 2008).

Different inflationary models involve different assumptions about the physics that drove the early expansion, and as a result, they predict different values for  $f_{\text{NL}}$ . For instance, the standard single-field slow-roll inflationary model predicts a small value of  $f_{\text{NL}} \sim 0.01$  (see, also, Maldacena 2003, for a discussion). On the other hand, some alternative inflationary scenarios involve multiple scalar fields that can interact with each other during inflation, leading to the generation of larger levels of non-Gaussianities. These models predict considerably larger values of  $f_{\text{NL}}$  that can reach up to 100 or higher (see, e.g., Alvarez et al. 2014; de Putter et al. 2017). Obtaining accurate and robust constraints on  $f_{\text{NL}}$  is therefore crucial in advancing our understanding of the dynamics of the early Universe. With  $\sigma(f_{\text{NL}}) \sim 1$ , we can rule out or confirm specific models of inflation and gain insight into the physics that drove the inflationary expansion.

The current tightest bound on  $f_{\text{NL}}$  comes from Planck's bispectrum measurement of CMB anisotropies,  $f_{\text{NL}} = 0.9 \pm 5.0$  (Planck Collaboration et al. 2019). Limited by cosmic variance, CMB data alone cannot enhance statistical precision of  $f_{\text{NL}}$  measurements enough to break the degeneracy amongst various inflationary paradigms (see, e.g., Simons Observatory et al. 2019). However, by combining LSS with CMB data, we can partially or

even completely cancel out the effects of cosmic variance, and thus significantly reduce statistical uncertainties down to  $\sigma(f_{\text{NL}}) \sim 1$  (see, e.g., Schmittfull & Seljak 2018). This is because LSS provides a measurement of the same underlying signal, but from a different vantage point and at a later time in Universe history. Also, constraining  $f_{\text{NL}}$  with the three-point clustering of LSS is hindered by nonlinearities raised from structure formation, which is non-trivial to model and disentangle from the primordial signal (Baldauf et al. 2011b,a). As a novel approach, UV Luminosity Function probes galaxy abundances and structure formation on small scales (e.g.,  $k \sim 2 \text{ Mpc}^{-1}$ ), which are otherwise impossible to explore with the scale-dependent bias. Sabti et al. (2021) used UV Luminosity Function from the Hubble Space Telescope catalogues (Bouwens et al. 2015) to find  $-166 < f_{\text{NL}} < 497$  at 95% confidence. Even though this approach is still not competitive with the current bounds from CMB and LSS, upcoming surveys such as the James Webb Space Telescope and the Nancy Grace Roman Space Telescope are forecast to yield up to four times improvements on  $f_{\text{NL}}$  constraints from the UV Luminosity Function.

Currently, the most precise constraints on  $f_{\text{NL}}$  from LSS reach a level of  $\sigma(f_{\text{NL}}) \sim 20 - 30$ , with the majority of the constraining power coming from the two-point statistics that utilize the scale-dependent bias (Castorina et al. 2019; Mueller et al. 2022; Cabass et al. 2022; D'Amico et al. 2022). However, further work is needed to fully harness the potential of the scale-dependent bias in constraining  $f_{\text{NL}}$ . Specifically, more simulation-based studies of halo-assembly bias are necessary to properly model the galaxy power spectrum in the presence of local PNG (Barreira et al. 2020; Barreira 2020, 2022; Lazeyras et al. 2023). In addition to the theoretical challenges, measuring  $f_{\text{NL}}$  with the scale-dependent bias effect is nonetheless incredibly challenging due to various observational systematic effects that modulate the galaxy power spectrum on large scales, where the primordial signal is sensitive. For instance, survey geometry mixes clustering power on different angular scales (Beutler et al. 2014; Wilson et al. 2017). Relativistic effects also generate PNG-like scale-dependent signatures on large scales, which interfere with measuring  $f_{\text{NL}}$  with the scale-dependent bias effect (Wang et al. 2020). Similarly, matter density fluctuations with wavelengths larger than survey size, known as super-sample modes, modulate the galaxy power spectrum (Castorina & Moradinezhad Dizgah 2020). In a similar way, the peculiar motion of the observer can mimic a PNG-like scale-dependent signature through aberration, magnification and the Kaiser-Rocket effect, i.e., a systematic dipolar apparent blue-shifting in the direction of the observer's peculiar motion (Bahr-Kalus et al. 2021). However, acting radially, this effect is subdominant in angular clustering measurements. An additional potential cause of systematic error arises from the fact that the mean galaxy density used to construct the density contrast field is estimated from the available data, rather than being known a priori. This introduces what is known as an integral constraint effect, which can cause the power spectrum on modes near the size of the survey to be artificially suppressed, effectively pushing it towards zero (Peacock & Nicholson 1991; De Mattia & Ruhlmann-Kleider 2019). Accounting for these theoretical systematic effects in the galaxy power spectrum is crucial to obtain unbiased inference about  $f_{\text{NL}}$  (see, e.g., Riquelme et al. 2022).

Other observational systematics are driven by varying imaging conditions across the sky (Ross et al. 2011) and photometric calibration issues in data reduction (Huterer et al. 2013) that appear as spurious fluctuations in the observed density field of galaxies. Compared to the integral constraint and survey geometry effects, this particular type of systematic error is much more complex to

deal with, and has proven to be a significant obstacle in past efforts to constrain local primordial non-Gaussianity using the scale-dependent bias (see, e.g., Pullen & Hirata 2013; Ho et al. 2015). It is anticipated that imaging systematic errors will be particularly problematic for wide-area galaxy surveys that observe regions of the night sky closer to the Galactic plane and that seek to incorporate more lenient selection criteria in order to include fainter galaxies (see, e.g., Kitandis et al. 2020). As an ongoing wide-area galaxy survey, the Dark Energy Spectroscopic Instrument (DESI) uses 5000 robotically-driven fibers to simultaneously collect spectra of extra-galactic objects. DESI is designed to deliver an unparalleled volume of spectroscopic data over its five year mission that promises to deepen our understanding of the equation of state for dark energy (DESI Collaboration et al. 2016). If systematic errors can be effectively controlled, DESI has the potential to offer competitive constraints on local PNG, with an anticipated precision of  $\sigma(f_{\text{NL}}) = 5$  (DESI Collaboration et al. 2016). By combining DESI data with those of other forthcoming surveys, such as the Rubin Observatory and SphereX, it may be possible to achieve an unparalleled level of precision, potentially reaching  $\sigma(f_{\text{NL}}) \sim 1$ . (see, e.g., Heinrich & Doré 2022).

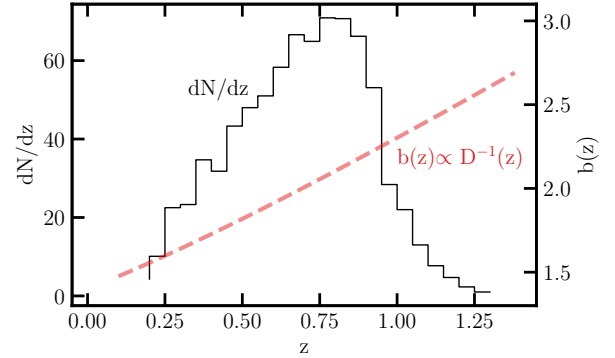
The primary objective of this paper is to utilize the scale-dependent bias signature in the angular power spectrum of luminous red galaxies selected from DESI imaging data to constrain the value of  $f_{\text{NL}}$ . By prioritizing the treatment of imaging systematic errors, we aim to lay the groundwork for subsequent studies of local PNG with DESI spectroscopy. To prepare our sample for such a delicate signal, we employ linear multivariate regression and neural networks to clean and refine the data from spurious density fluctuations caused by varying imaging conditions. Our findings are validated using suites of lognormal simulations featuring values of  $f_{\text{NL}} = 0$  and 76.9. With cross correlation techniques, we thoroughly characterize and examine potential sources of systematic error, including survey depth, astronomical seeing, photometric calibration, Galactic extinction, and local stellar density.

This paper is structured as follows. Section 2 describes the galaxy sample from DESI imaging and lognormal simulations with, or without, PNG and synthetic systematic effects. Section 3 outlines the theoretical framework for modeling the angular power spectrum, strategies for handling various observational and theoretical systematic effects, and statistical techniques for measuring the significance of remaining systematics in our sample after mitigation. Our results are presented in Section 4, and Section 5 summarizes our conclusions and directions for future work.

## 2 DATA

Luminous red galaxies (LRGs) are massive galaxies that populate massive halos, lack active star formation, and are highly biased tracers of the dark matter gravitational field. A distinct break around 4000 Å in the LRG spectrum is often utilized to determine their redshifts accurately. LRGs are widely targeted in previous galaxy redshift surveys (see, e.g., Eisenstein et al. 2001; Prakash et al. 2016), and their clustering and redshift properties are well studied (see, e.g., Ross et al. 2020; Gil-Marín et al. 2020; Bautista et al. 2021; Chapman et al. 2022).

DESI is designed to collect spectra of millions of LRGs covering the redshift range  $0.2 < z < 1.35$ . DESI selects its targets for spectroscopy from the DESI Legacy Imaging Surveys, which consist of three ground-based surveys that provide photometry of the sky in the optical  $g$ ,  $r$ , and  $z$  bands. These surveys include the



**Figure 1.** The redshift distribution (solid line) and bias evolution (dashed line) of the DESI LRG targets. The redshift distribution is determined by DESI spectroscopy. The redshift evolution of the linear bias is determined by the growth factor  $D(z)$ , assuming a constant clustering amplitude.

Mayall  $z$ -band Legacy Survey using the Mayall telescope at Kitt Peak (MzLS; Dey et al. 2018), the Beijing–Arizona Sky Survey using the Bok telescope at Kitt Peak (BASS; Zou et al. 2017), and the Dark Energy Camera Legacy Survey on the Blanco 4m telescope (DECaLS; Flaugher et al. 2015). As shown in Figure 2, the BASS and MzLS surveys observed the same footprint in the North Galactic Cap (NGC) while the DECaLS program observed both caps around the galactic plane; the BASS+MzLS footprint is separated from the DECaLS NGC at  $\text{DEC} > 32.375$  degrees, although there is an overlap between the two regions for calibration purposes (Dey et al. 2018). Additionally, the DECaLS program integrates observations executed from the Blanco instrument under the Dark Energy Survey (DES Collaboration et al. 2016), which cover about 1130 deg<sup>2</sup> of the South Galactic Cap (SGC) footprint. The DESI imaging catalogues also integrate the 3.4 (W1) and 4.6  $\mu\text{m}$  (W2) infrared photometry from the Wide-Field Infrared Explorer (WISE; Wright et al. 2010; Meisner et al. 2018).

### 2.1 DESI imaging LRGs

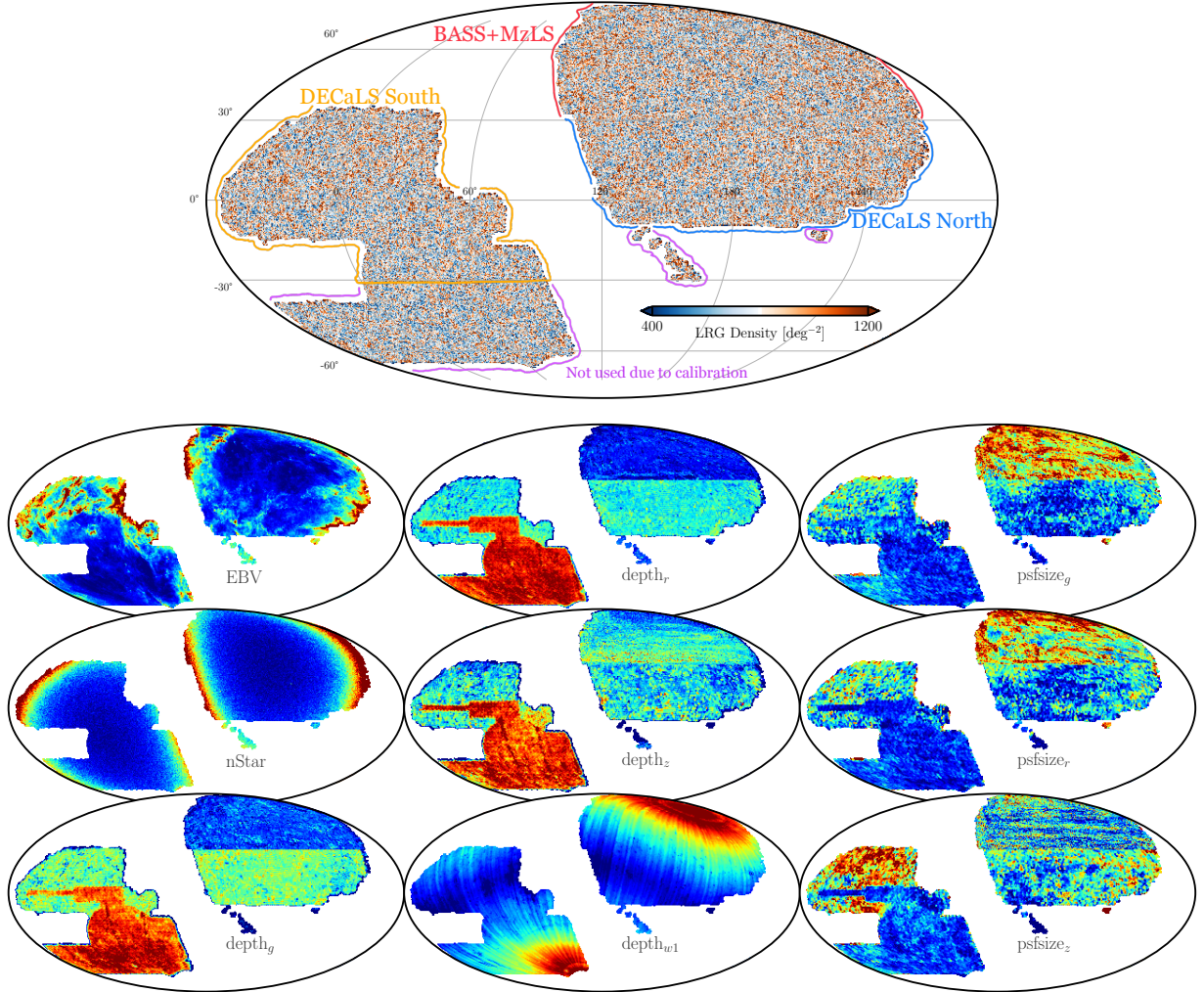
Our sample of LRGs is drawn from the DESI Legacy Imaging Surveys Data Release 9 (DR9; Dey et al. 2018) using the color-magnitude selection criteria designed for the DESI 1% survey (DESI Collaboration in prep), described as the Survey Validation 3 (SV3) selection in more detail in Zhou et al. (2022). The color-magnitude selection cuts are defined in the  $g$ ,  $r$ ,  $z$  bands in the optical and W1 band in the infrared, as summarized in Table 1. The selection cuts are developed differently for each imaging survey to reach an almost uniform target surface density, with an average density of 800 galaxies per square degree covering around 18000 square degrees, despite different survey efficiency and photometric calibration between DECaLS and BASS+MzLS. The implementation of these selection cuts in the DESI data processing pipeline is explained in Myers et al. (2022). The redshift distribution of our galaxy sample are inferred respectively from DESI spectroscopy during the Survey Validation phase. The linear bias is assumed to vary with redshift through the growth factor  $D(z)$ , i.e.,  $b(z) = 1.43/D(z)$ , such that the clustering amplitude remains constant over redshift (see, e.g., Zhou et al. 2021). Figure 1 shows the redshift distribution (solid black line) and the evolution of the linear bias (dashed red line) over redshift for our sample of LRGs.

The LRG sample is masked rigorously for foreground bright



**Table 1.** Color-magnitude selection criteria for the DESI LRG targets (Zhou et al. 2022). Magnitudes are corrected for Galactic extinction. The z-band fiber magnitude,  $z_{\text{fiber}}$ , corresponds to the expected flux within a DESI fiber.

Footprint	Criterion	Description
DECaLS	$z_{\text{fiber}} < 21.7$	Faint limit
	$z - W1 > 0.8 \times (r - z) - 0.6$	Stellar rejection
	$[(g - r > 1.3) \text{ AND } ((g - r) > -1.55 \times (r - W1) + 3.13)] \text{ OR } (r - W1 > 1.8)$	Remove low-z galaxies
	$[(r - W1 > (W1 - 17.26) \times 1.8) \text{ AND } (r - W1 > W1 - 16.36)] \text{ OR } (r - W1 > 3.29)$	Luminosity cut
BASS+MzLS	$z_{\text{fiber}} < 21.71$	Faint limit
	$z - W1 > 0.8 \times (r - z) - 0.6$	Stellar rejection
	$[(g - r > 1.34) \text{ AND } ((g - r) > -1.55 \times (r - W1) + 3.23)] \text{ OR } (r - W1 > 1.8)$	Remove low-z galaxies
	$[(r - W1 > (W1 - 17.24) \times 1.83) \text{ AND } (r - W1 > W1 - 16.33)] \text{ OR } (r - W1 > 3.39)$	Luminosity cut

**Figure 2.** Top: The DESI LRG target density map before correcting for imaging systematic effects in Mollweide projection. The disconnected islands from the North footprint and parts of the South footprint with declination below  $-30^\circ$  are removed from the sample for the analysis due to potential calibration issues (see text). Bottom: Mollweide projections of the imaging systematic maps (survey depth, astronomical seeing/psfsize, Galactic extinction, and local stellar density) in celestial coordinates. Not shown here are two external maps for the neutral hydrogen column density and photometric calibration, which are only employed for the robustness tests. The imaging systematic maps are colour-coded to show increasing values from blue to red.

stars, bright galaxies, and clusters of galaxies<sup>1</sup> to further reduce stellar contamination (Zhou et al. 2022). Then, the sample is binned into HEALPIX (Gorski et al. 2005) pixels at  $N_{\text{SIDE}} = 256$  to con-

struct the 2D density map (as shown in the top panel of Figure 2). The LRG density is corrected for the pixel incompleteness and lost areas using a catalogue of random points, hereafter referred to as randoms, uniformly scattered over the footprint with the same cuts and masks applied. The DESI LRG targets are selected brighter than the imaging survey depth limits, and thus the LRG density map does not exhibit severe spurious density fluctuations.

<sup>1</sup> See <https://www.legacysurvey.org/dr9/bitmasks/> for maskbit definitions.

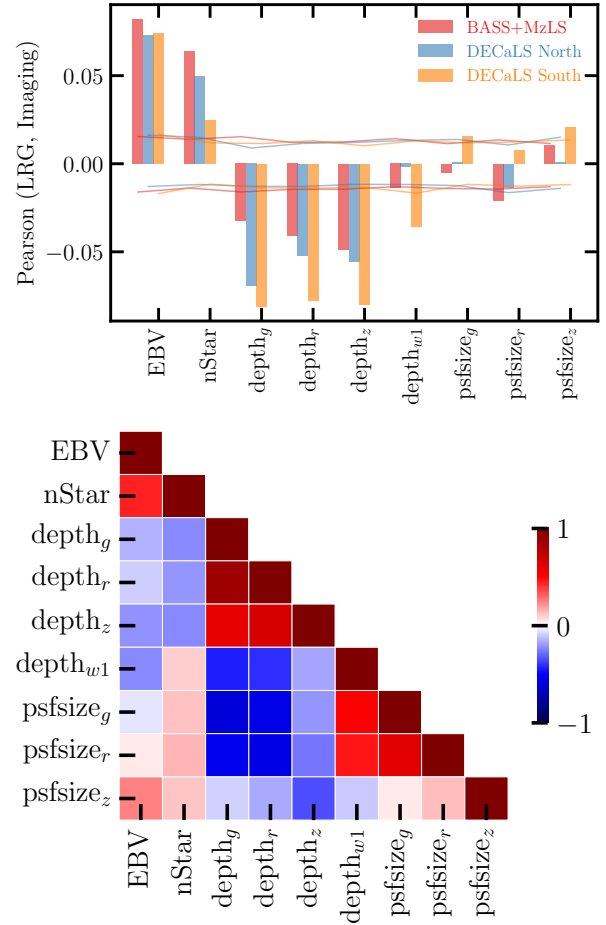
### 2.1.1 Imaging systematic maps

In this paper, nine imaging properties, which can be considered as potential sources of systematic error, are mapped into HEALPix of  $N_{\text{SIDE}} = 256$  (as illustrated by a  $3 \times 3$  grid in the bottom panel of Figure 2). The maps include local stellar density constructed from point-like sources with a G-band magnitude in the range  $12 \leq G < 17$  from the *Gaia* DR2 (see, [Gaia Collaboration et al. 2018](#); [Myers et al. 2022](#)); Galactic extinction  $E[B-V]$  from [Schlegel et al. \(1998\)](#); survey depth (galaxy depth in  $g$ ,  $r$ , and  $z$  and PSF depth in W1) and astronomical seeing (psfsize) in  $g$ ,  $r$ , and  $z$ . The depth maps have been corrected for extinction using the coefficients adapted from [Schlafly & Finkbeiner \(2011\)](#). [Zhou et al. \(2022\)](#) has previously identified these nine imaging systematic maps as potential sources of imaging systematic errors in the DESI LRG targets. In addition to these nine maps, we consider two external maps for the neutral hydrogen column density (HI) from [HI4PI Collaboration et al. \(2016\)](#) and photometric calibration in the  $z$ -band (CALIBZ) from [CITE](#) to further test the robustness of our analysis against unknown systematics.

Each imaging map carries its characteristic fluctuations, which correlate with the LRG density map. For instance, large-scale fluctuations can be associated with stellar density, extinction, or survey depth; while small scale-fluctuations can be related to psfsize variations. Upon visual inspection, the under-dense regions of LRGs in the SGC can be associated with varying survey depth, while the over-dense regions in the NGC can be connected to galactic dust or extinction. Some regions of the DR9 footprint are removed from our analysis to avoid potential photometric calibration issues. These regions are either disconnected from the main footprint (e.g., the islands in the NGC with  $\text{DEC} < -10$ ) or calibrated using different catalogues of standard stars (e.g.,  $\text{DEC} < -30$  in the SGC). The potential impact of not imposing these declination cuts on the LRG sample and our  $f_{\text{NL}}$  constraints is explored in Section 4.

Figure 3 shows the Pearson correlation coefficient between the DESI LRG target density map and the imaging systematics maps for the three imaging regions (DECaLS North, DECaLS South, and BASS+MzLS) in the top panel. The horizontal curves represent the 95% confidence regions and are constructed by cross-correlating 100 synthetic lognormal density fields and the imaging systematic maps. Figure 3 (bottom panel) shows the correlation matrix among the imaging systematics maps for the DESI footprint. There are statistically significant correlations between the LRG density and depth, extinction, and stellar density. There are less significant correlations between the LRG density and the W1-band depth and psfsize in DECaLS South and DECaLS North. Significant inner correlations exist among the imaging systematic maps themselves, especially between local stellar density and Galactic extinction; also, the  $r$ -band and  $g$ -band survey properties are more correlated with each other than with the  $z$ -band counterpart. **To ensure the robustness of the analysis, we computed the Spearman correlation coefficient as well to evaluate the monotonic relationship between the DESI LRG target density and imaging systematic maps. However, we found no substantial differences in the correlation coefficients.**

The effects of observational systematics in the DESI targets have been studied in great detail (see, e.g., [Kitanidis et al. 2020](#); [Zhou et al. 2021](#); [Chaussidon et al. 2022](#)). There are several approaches for handling imaging systematic errors, broadly classified into data-driven and simulation-based modeling approaches. The general idea behind these approaches is to use the available data or simulations to learn or forward model the relationship between the observed target density and the imaging systematic maps, and



**Figure 3.** Top: The Pearson correlation coefficient between the DESI LRG target density and imaging properties in BASS+MzLS, DECaLS North, and DECaLS South. Solid horizontal curves represent the 95% confidence intervals estimated from simulations of lognormal density fields. Bottom: The Pearson correlation matrix of imaging properties for the DESI footprint.

to use this relationship, which is often described by a set of *imaging weights*, to mitigate spurious fluctuations in the observed target density. Another techniques for reducing the effect of imaging systematics rely on cross-correlating different tracers of dark matter to ameliorate excess clustering signals, as each tracer might respond differently to a source of systematic error (see, e.g., [Giannantonio et al. 2014](#)). These methods have their limitations and strengths (see, e.g., [Weaverdyck & Huterer 2021](#), for a review).

### 2.1.2 Treatment of imaging systematics

**MR: make this section clear.** In this paper, data-driven approaches, including linear multivariate regression and artificial neural networks, are applied to the data to correct for imaging systematic effects. Both linear and neural network models are trained on each imaging region separately since the Pearson correlation coefficient analysis indicated the level of imaging systematic effects in each region is different, as illustrated in the top panel of Figure 3.

The linear multivariate model only uses the imaging systematic maps up to the linear power to predict the number counts of the DESI LRG targets in pixel  $i$ ,

$$\rho_i = \log(1 + \exp[\mathbf{a} \cdot \mathbf{x}_i]), \quad (2)$$

where  $\mathbf{a} \cdot \mathbf{x}_i$  represents the inner product between the parameters,  $\mathbf{a}$ , and imaging systematic values for pixel  $i$ ,  $\mathbf{x}_i$ . The implementation of Markov Chain Monte Carlo (MCMC) random search from EMCEE (Foreman-Mackey et al. 2013) is used to explore the parameter space by minimizing the negative Poisson log-likelihood between the actual and predicted number counts of targets. No spatial coordinate is included in  $\mathbf{x}_i$  to help avoid over correction, and as a result, the predicted number counts solely reflect the spurious density fluctuations that arise from varying imaging conditions. The number of pixels is substantially larger than the number of parameters for the linear model, and thus no training-validation-testing split is applied to the data. This aligns with the methodology used for training linear models in previous analyses (see, e.g., Zhou et al. 2022). The marginalized mean of the parameters from MCMC are then utilized to compute the imaging weights.

Our neural network-based mitigation approach uses the implementation of fully connected feedforward neural networks from Rezaie et al. (2021). With the neural network approach,  $\mathbf{a} \cdot \mathbf{x}_i$  in Equation 2 is replaced with  $NN(\mathbf{x}_i|\mathbf{a})$ , where  $NN$  represents the fully connected neural network and  $\mathbf{a}$  denotes its parameters. The implementation, training, validation, and application of neural networks on galaxy survey data are presented in Rezaie et al. (2021). We briefly summarize the methodology here.

A fully connected feedforward neural network (also called a *multi-layer perceptron*) is a type of artificial neural network where the neurons are arranged in layers, and each neuron in one layer is connected to every neuron in the next layer. The imaging systematic information flows only in one direction, from input to output. Each neuron applies a nonlinear activation function (i.e., transformation) to the weighted sum of its inputs, which are the outputs of the neurons in the previous layer. The output of the last layer is the prediction of the model for the number counts of galaxies. Our architecture consists of three hidden layers with 20 rectifier activation functions on each layer, and a single neuron in the output layer. The rectifier is defined as  $\max(0, x)$  (Nair & Hinton 2010). This simple form of nonlinearity is very effective in enabling deep neural networks to learn more complex, nonlinear relationships between the input imaging maps and output galaxy counts.

Unlike linear regression, neural networks are prone to fitting noise, i.e., excellent performance on training data and poor performance on unseen data. Therefore, our analysis uses a training-validation-testing split to ensure that the network is well-optimized and generalizes well to unseen data. Specifically, 60% of the LRG data is used for training, 20% is used for validation, and 20% is used for testing. The split is performed randomly aside from the locations of the pixels. We test a geometrical split in which neighboring pixels beyond to the same set of training, testing, or validation, but no significant performance difference is observed. The neural network models are tested on the entirety of the LRG sample with the technique of permuting the choice of the training, validation, or testing sets (Arlot & Celisse 2010). The neural networks are trained for up to 70 training epochs with the gradient descent ADAM optimizer (Loshchilov & Hutter 2017). The neural network parameters are adjusted iteratively following the gradient of the negative Poisson log-likelihood. The step size of the parameter updates is controlled via the learning rate hyper-parameter, which is initialized with a grid search and is designed to dynamically vary between two boundary values of 0.001 and 0.1 to avoid local minima (see, also, Loshchilov & Hutter 2016). At each training epoch, the neural network model is applied to the validation set, and ultimately the model with the best performance on validation is identified and applied to the test set. With the cross-validation technique, the model predictions from

the different test sets are aggregated together to form the predicted target density map into HEALPIX of  $N_{\text{SIDE}} = 256$ . To reduce the error in the predicted number counts, we train an ensemble of 20 neural network models and average over the predictions. The imaging weights are then defined as the inverse of the predicted target density, normalized to unity.

One potential problem that can arise in the data-driven mitigation approach is *over-correction*, which occurs when the corrections applied to the data are too strong that remove the clustering signal and induce additional biases in the inferred parameter. The neural network approach is more prone to this issue compared to the linear approach, due to the increased flexibility. As illustrated in the bottom panel of Figure 3, the significant correlations among the imaging systematic maps may pose additional challenges for modeling these spurious density fluctuations. Specifically, using highly correlated imaging systematic maps increases the input noise to data-driven regression, which elevates the potential for over subtracting the clustering power. These over-correction effects are identified to have a negligible impact on Baryon Acoustic Oscillations and Redshift Space Distortions (Merz et al. 2021); however, they can significantly modulate the galaxy power spectrum on large scales, and thus lead to biased  $f_{\text{NL}}$  constraints (Rezaie et al. 2021; Mueller et al. 2022). Although not explored thoroughly, these issues could limit the detectability of primordial features in the galaxy power spectrum and that of parity violations in higher order clustering statistics (Beutler et al. 2019; Cahn et al. 2021; Philcox 2022). Therefore, it is crucial to develop, implement, and apply techniques to minimize and control over-correction, if possible, by reducing the dimensionality of the problem, in the hope of ensuring that the constraints are as accurate and reliable as possible. Our goal is to reduce the correlations between the DESI LRG target density and the imaging systematic maps, while minimizing the chance of over correction. To achieve this objective, we employ a series of simulations along with statistical methods that involve cross power spectrum and mean galaxy density to identify different sets of the imaging systematic maps: Santi: Elaborate how these maps are selected

- (i) **Two maps:** Extinction, depth in  $z$ .
- (ii) **Three maps:** Extinction, depth in  $z$ , psfsize in  $r$ .
- (iii) **Four maps:** Extinction, depth in  $z$ , psfsize in  $r$ , stellar density.
- (iv) **Five maps:** Extinction, depth in  $z$ , psfsize in  $r$ , neutral hydrogen density, and calibration in  $z$ .
- (v) **Eight maps:** Extinction, depth in  $grzW1$ , psfsize in  $grz$ .
- (vi) **Nine maps:** Extinction, depth in  $grzW1$ , psfsize in  $grz$ , stellar density.

Edmond: You can comment a bit on the different combinations; why / how do you decide which combination we want ? (maybe just say that it is explained below, your sentence on visual inspection)  
 Alex: Throughout—I think it would be clearer if you always referred to every method with 2 labels, “[linear/nonlinear] [map name]”. We emphasise that these sets are established prior to examining the auto power spectrum of the LRG sample. The auto power spectrum measurements are *unblinded* only after our mitigation methods passed our rigorous tests for residual systematics. To start, we will use the linear model and two maps (extinction and depth in  $z$ ), which is the most conservative method in terms of both the model flexibility and the number of parameters. We clean the sample using the imaging weights obtained from *linear two maps*. We find that the linear two maps approach mitigates most of the spurious fluctuations in the LRG density and reduces the cross-correlations between the LRG density and the imaging systematic maps, except for the trends against psfsize in the  $r$  and  $z$  bands. Adding the  $r$ -band ps-



fszsize improves the linear model performance such that the cross correlations are similar to those obtained from linear eight maps, which indicates no further information can be extracted from eight maps. Therefore, we identify extinction, depth in  $z$ , and psfsize in  $r$  (*three maps*) as the primary sources of systematic effects. Then, we adapt *neural network three maps* to model nonlinear systematic effects, and find that the neural network-based weights significantly reduce the cross correlations and spurious density fluctuations. As explained in Section 3, we consider neural network with four, five, and nine maps to further test for the robustness of our cleaning methods.

## 2.2 Synthetic lognormal density fields

Density fluctuations of galaxies on large scales can be approximated with lognormal distributions (Coles & Jones 1991). Unlike N-body simulations, simulating lognormal density fields is not computationally intensive, and allows quick and robust validation of data analysis pipelines. These mocks are therefore considered useful for our study since the signature of local PNG appears on large-scales and small scale clustering is not used. The package FLASK (Full-sky Lognormal Astro-fields Simulation Kit; Xavier et al. 2016) is employed to generate ensembles of synthetic lognormal density maps that mimic the bias, redshift, and angular distributions of the DESI LRG targets, as illustrated in Figure 1 and 2. Two universes with  $f_{\text{NL}} = 0$  and 76.92 are considered. A set of 1000 realizations are produced for every  $f_{\text{NL}}$ . The analysis adapts a fiducial cosmology from a flat  $\Lambda$ CDM universe, including one massive neutrino with  $m_\nu = 0.06$  eV, Hubble constant  $h = 0.67$ , matter density  $\Omega_M = 0.31$ , baryon density  $\Omega_b = 0.05$ , and spectral index  $n_s = 0.967$ . The amplitude of the matter density fluctuations on a scale of  $8h^{-1}\text{Mpc}$  is set as  $\sigma_8 = 0.8225$ . The same fiducial cosmology is used throughout this paper unless specified otherwise. **Santi: how sensitive is the fnl signal to the choice of fiducial cosmology.**

### 2.2.1 Contaminated mocks

The linear model is employed to introduce synthetic spurious fluctuations in the lognormal density fields. The motivation for choosing a linear contamination model is to assess how much of the clustering signal can be removed by applying more flexible and rigorous models, based on neural networks, for correcting imaging systematic effects. The parameter space of the linear model is explored using MCMC, from the relationship between the LRG sample and the imaging systematic maps. The MCMC process is executed separately on each imaging survey using only *three maps* (extinction, depth in  $z$ , and psfsize in  $r$ ) as  $\mathbf{x}_i$ . The imaging selection function for contaminating each simulation is uniquely and randomly drawn from the parameter space probed by MCMC, and then the results from each imaging survey are combined to form the DESI footprint. The same contamination model is used for both the  $f_{\text{NL}} = 0$  and 76.92 simulations.

Similar to the imaging systematic treatment analysis for the DESI LRG targets, the neural network methods with various combinations of the imaging systematic maps are applied to each simulation, with and without PNG, and with and without systematics, to derive the imaging weights. Section 3 presents how the simulation results are incorporated to calibrate  $f_{\text{NL}}$  biases due to over-correction.

## 3 ANALYSIS TECHNIQUES

### 3.1 Power spectrum estimator

The pseudo angular power spectrum (Hivon et al. 2002) is utilized to extract information from the galaxy density contrast field,  $\delta_g$ ,

$$\delta_g = \frac{\rho - \bar{\rho}}{\bar{\rho}}, \quad (3)$$

by decomposing it into spherical harmonics,  $Y_{\ell m}$ ,

$$a_{\ell m} = \int d\Omega \delta_g W Y_{\ell m}^*. \quad (4)$$

The mean galaxy density  $\bar{\rho}$  is estimated from the entire LRG sample<sup>2</sup> and the survey window  $W$  is determined by comparing the number of randoms to the expected number. Then, the angular power spectrum is estimated by

$$\tilde{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2. \quad (5)$$

We use the implementation of *anafast* from the HEALPIX package (Gorski et al. 2005) to do fast harmonic transforms and estimate the pseudo angular power spectrum and cross power spectrum. When the sky coverage of a survey is incomplete, this estimator yields a biased power spectrum. Specifically, the survey mask causes correlations between different harmonic modes and results in the measured power on scales near the survey size being pushed towards zero. Since these scales are highly sensitive to local primordial non-Gaussianity, it is crucial to account for these systematic effects in the model galaxy power spectrum to obtain unbiased  $f_{\text{NL}}$  constraints.

### 3.2 Modeling

#### 3.2.1 Angular power spectrum

**Would nonlinear P(k) matter?** The relationship between the linear matter power spectrum  $P(k)$  and the projected angular power spectrum of galaxies is expressed by the following equation:

$$C_\ell = \frac{2}{\pi} \int_0^\infty \frac{dk}{k} k^3 P(k) |\Delta_\ell(k)|^2 + N_{\text{shot}}, \quad (6)$$

where  $N_{\text{shot}}$  is a scale-independent shot noise term. The projection kernel  $\Delta_\ell(k) = \Delta_\ell^g(k) + \Delta_\ell^{\text{RSD}}(k)$  includes redshift space distortions and determines the contribution of each wavenumber  $k$  to the galaxy power spectrum on mode  $\ell$ . For more details, refer to Padmanabhan et al. (2007). The FFTLog<sup>3</sup> algorithm and its extension as implemented in Fang et al. (2020) are employed to calculate the integrals for the projection kernel  $\Delta_\ell(k)$ , which includes the  $l^{\text{th}}$  order spherical Bessel functions,  $j_\ell(kr)$ , and its second derivatives,

$$\Delta_\ell^g(k) = \int \frac{dr}{r} r(b + \Delta b) D(r) \frac{dN}{dr} j_\ell(kr), \quad (7)$$

$$\Delta_\ell^{\text{RSD}}(k) = - \int \frac{dr}{r} r f(r) D(r) \frac{dN}{dr} j_\ell''(kr), \quad (8)$$

where  $b$  is the linear bias (dashed curve in Figure 1),  $D$  represents the linear growth factor normalized as  $D(z=0) = 1$ ,  $f(r)$  is the growth rate, and  $dN/dr$  is the redshift distribution of galaxies

<sup>2</sup> The mean galaxy density is calculated separately for each region when we fit the power spectrum from each region individually.

<sup>3</sup> [github.com/xfangcosmo/FFTLLog-and-beyond](https://github.com/xfangcosmo/FFTLLog-and-beyond)

normalized to unity and described in terms of comoving distance<sup>4</sup> (solid curve in Figure 1). The PNG-induced scale-dependent shift is given by (see, also, Slosar et al. 2008)

$$\Delta b = b_\phi(z) f_{\text{NL}} \frac{3\Omega_m H_0^2}{2k^2 T(k) D(z) c^2} \frac{g(\infty)}{g(0)}, \quad (9)$$

where  $\Omega_m$  is the matter density,  $H_0$  is the Hubble constant<sup>5</sup>,  $T(k)$  is the transfer function, and  $g(\infty)/g(0) \sim 1.3$  with  $g(z) \equiv (1+z)D(z)$  is the growth suppression due to non-zero  $\Lambda$  because of our normalization of  $D$  (see, e.g., Reid et al. 2010; Mueller et al. 2019). The bias parameter  $b_\phi$  describes the response of galaxy formation to primordial potential perturbations in the presence of local PNG. Assuming that only mass determines the halo occupation function,  $b_\phi = 2\delta_c(b-p)$ , where  $p = 1$  and  $\delta_c = 1.686$  is the critical density for spherical collapse (Fillmore & Goldreich 1984). Santi:  $p=1$  for galaxies assumes this on top of assuming that halos follow a universal halo mass function (only depending on peak height) The theoretical uncertainty on  $p$  is not very well constrained, and Barreira (2022) showed that marginalizing over this parameter even with wide priors leads to biased  $f_{\text{NL}}$  constraints because of parameter space projection effects. Lazeyras et al. (2023) used N-body simulations to investigate secondary halo properties, such as concentration, spin and sphericity of haloes, and found that halo spin and sphericity preserve the universality of the halo occupation function while halo concentration significantly alters the halo function. Without better-informed priors on  $p$ , it is argued that the scale-dependent bias effect can only be used to constrain the  $b_\phi f_{\text{NL}}$  term (see, e.g., Barreira et al. 2020; Barreira 2020). However, the detection significance of local PNG remains unaffected by various assumptions regarding  $p$ . This means that a nonzero detection of  $b_\phi f_{\text{NL}}$  at a certain confidence level will still indicate a nonzero detection of  $f_{\text{NL}}$  at that same confidence level. This paper is focused on how a careful assessment of imaging systematic effects, or lack thereof, can bias our PNG constraints. Therefore, we choose  $p = 1$  for our sample of DESI LRG targets (see, also, Slosar et al. 2008; Reid et al. 2010; Ross et al. 2013), and do not marginalize over  $p$  to avoid projection effects (Barreira 2022).

### 3.2.2 Survey geometry and integral constraint

We employ a technique similar to the one proposed by Chon et al. (2004) to account for the impact of the survey geometry on the theoretical power spectrum. The ensemble average for the partial sky power spectrum is related to that of the full sky power spectrum via a mode-mode coupling matrix,  $M_{\ell\ell'}$ ,

$$\langle \tilde{C}_\ell \rangle = \sum_{\ell'} M_{\ell\ell'} \langle C_{\ell'} \rangle. \quad (10)$$

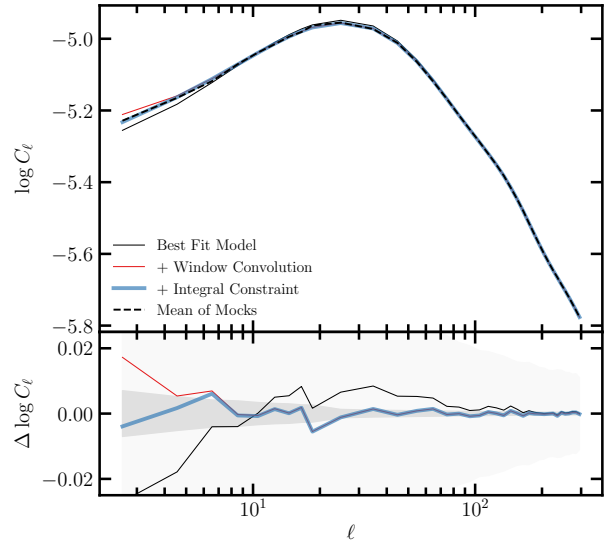
We convert this convolution in the spherical harmonics space into a multiplication in the correlation function space. Specifically, we first express the theory power spectrum (Equation 6) in terms of the correlation function,  $\hat{\omega}^{\text{model}}$ , using the spherical harmonics transform. Then, we estimate the survey mask correlation function,  $\hat{\omega}^{\text{window}}$ , and obtain the pseudo-power spectrum,

$$\tilde{C}_\ell^{\text{model}} = 2\pi \int \hat{\omega}^{\text{model}} \hat{\omega}^{\text{window}} P_\ell(\cos\theta) d\theta. \quad (11)$$

The integral constraint is another systematic effect which is

<sup>4</sup>  $dN/dr = (dN/dz) * (dz/dr) \propto (dN/dz) * H(z)$

<sup>5</sup>  $H_0 = 100 \text{ (km s}^{-1}\text{)/(h}^{-1}\text{Mpc)}$  and  $k$  is in unit of  $h\text{Mpc}^{-1}$



**Figure 4.** The mean power spectrum from the  $f_{\text{NL}} = 0$  mocks (no contamination) and best fit theoretical prediction after accounting for the survey geometry and integral constraint effects. The dark and light shades represent  $1\sigma$  error on the mean and one realization, respectively. Bottom panel shows the residual power spectrum relative to the mean power spectrum. No imaging systematic cleaning is applied to these mocks.

induced since the mean galaxy density is estimated from the observed galaxy density. The estimate of the mean density is biased by the limited sky coverage. This issue was first raised in Peacock & Nicholson (1991). To account for the integral constraint, the survey mask power spectrum is used to introduce a scale-dependent correction factor that needs to be subtracted from the power spectrum. Finally, the pseudo power spectrum with the integral constraint correction is obtained as

$$\tilde{C}_\ell^{\text{model,IC}} = \tilde{C}_\ell^{\text{model}} - \tilde{C}_{\ell=0}^{\text{model}} \left( \frac{\tilde{C}_\ell^{\text{window}}}{\tilde{C}_{\ell=0}^{\text{window}}} \right), \quad (12)$$

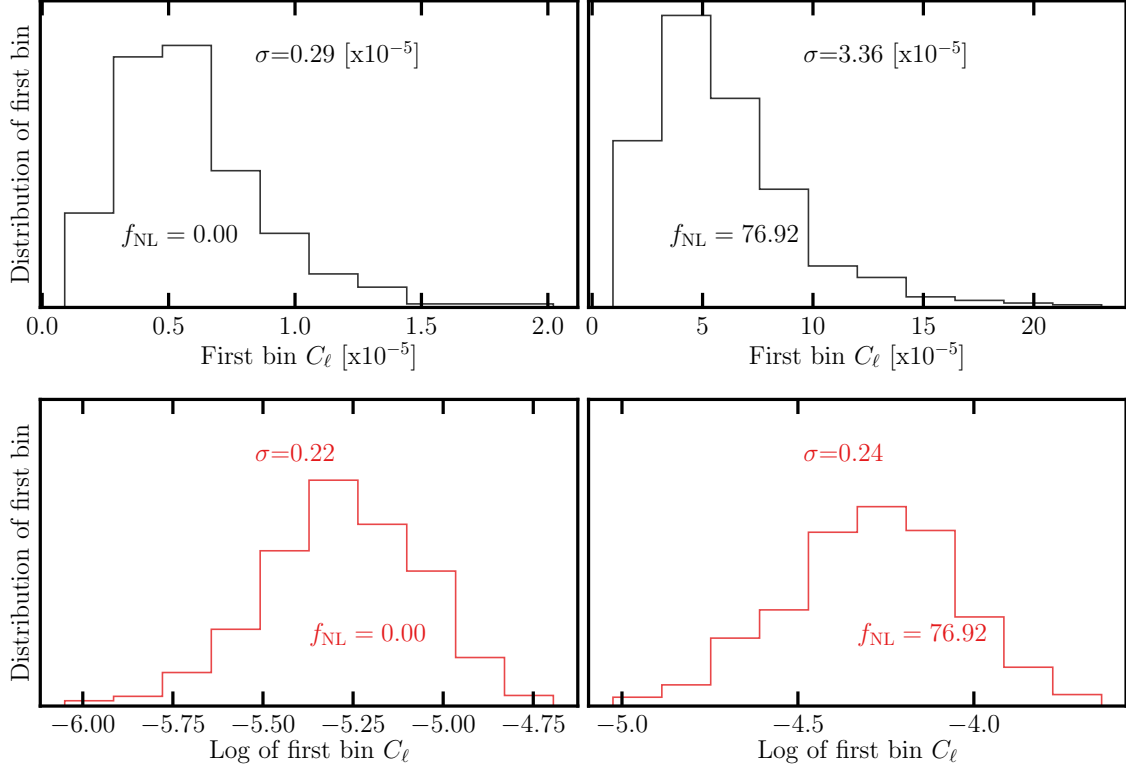
where  $\tilde{C}^{\text{window}}$  is the spherical harmonic transform of  $\hat{\omega}^{\text{window}}$ .

The lognormal simulations are used to validate our survey window and integral constraint correction. Figure 4 shows the mean power spectrum of the  $f_{\text{NL}} = 0$  simulations (dashed) and the best fit theory prediction before and after accounting for the survey mask and integral constraint. The simulations are neither contaminated nor mitigated. The light and dark shades represent the 68% estimated error on the mean and one single realization, respectively. The DESI mask is applied to the simulations, which covers around 40% of the sky. We find that the survey window effect affects the clustering power on  $\ell < 200$  and the integral constraint modulates the clustering power on  $\ell < 6$ .

### 3.3 Parameter estimation

Our parameter inference uses standard MCMC sampling. A constant clustering amplitude is assumed for the linear bias of our DESI LRG targets,  $b(z) = b/D(z)$ . In MCMC, we allow  $f_{\text{NL}}$ ,  $N_{\text{shot}}$ , and  $b$  to vary, while all other cosmological parameters are fixed at the fiducial values (see §2.2). The galaxy power spectrum is divided into a discrete set of bandpower bins with  $\Delta\ell = 2$  between  $\ell = 2$  and 20 and  $\Delta\ell = 10$  from  $\ell = 20$  to 300. Each clustering mode is weighted by  $2\ell + 1$  when averaging over the modes in each bin.





**Figure 5.** The distribution of the first bin power spectra and its log transformation from the simulations with  $f_{\text{NL}} = 0$  (left) and 76.92 (right). The log transformation alleviates the asymmetry in the distributions.

With the lognormal simulations, we find that the distribution of the power spectrum at the lowest bin,  $2 \leq \ell < 4$ , is asymmetric and its standard deviation varies significantly from the simulations with  $f_{\text{NL}} = 0$  to those with 76.9 (Figure 5). Therefore, we attempt to fit the log transformed power spectrum,  $\log C_\ell$ , to make our  $f_{\text{NL}}$  constraints less sensitive to the choice of covariance matrix. The parameter  $f_{\text{NL}}$  is constrained by minimizing a negative likelihood defined as,

$$-2 \ln \mathcal{L} = (\log \tilde{C}(\Theta) - \log \tilde{C})^\dagger \mathbb{C}^{-1} (\log \tilde{C}(\Theta) - \log \tilde{C}), \quad (13)$$

where  $\Theta$  represents a container for the parameters  $f_{\text{NL}}$ ,  $b$ , and  $N_{\text{shot}}$ ;  $\tilde{C}(\Theta)$  is the (binned) expected pseudo-power spectrum;  $\tilde{C}$  is the (binned) measured pseudo-power spectrum; and  $\mathbb{C}$  is the covariance matrix constructed from the lognormal simulations. Flat priors are implemented for all parameters.

### 3.4 Characterization of remaining systematics

**Santi:** I feel like this section would make more sense in "2.Data" In the absence of systematic effects, a) the mean galaxy density should be uniform across the footprint within the statistical fluctuations regardless of imaging conditions and b) the cross power spectrum between the galaxy density and the imaging systematic maps should be consistent with zero within the statistical fluctuations. In the following, two statistical tests are implemented and applied to quantify remaining systematic effects in our sample (see, also, Rezaie et al. 2021).

#### 3.4.1 Cross power spectrum

We characterize the cross correlations between the galaxy density and imaging systematic maps by [Could we call them something else here (maybe contaminant maps)? I first got confused and thought you were cross-correlating redshift slices in your imaging survey, but maybe our readers will read more carefully than me]

$$\tilde{C}_{X,\ell} = [\tilde{C}_{x_1,\ell}, \tilde{C}_{x_2,\ell}, \tilde{C}_{x_3,\ell}, \dots, \tilde{C}_{x_9,\ell}], \quad (14)$$

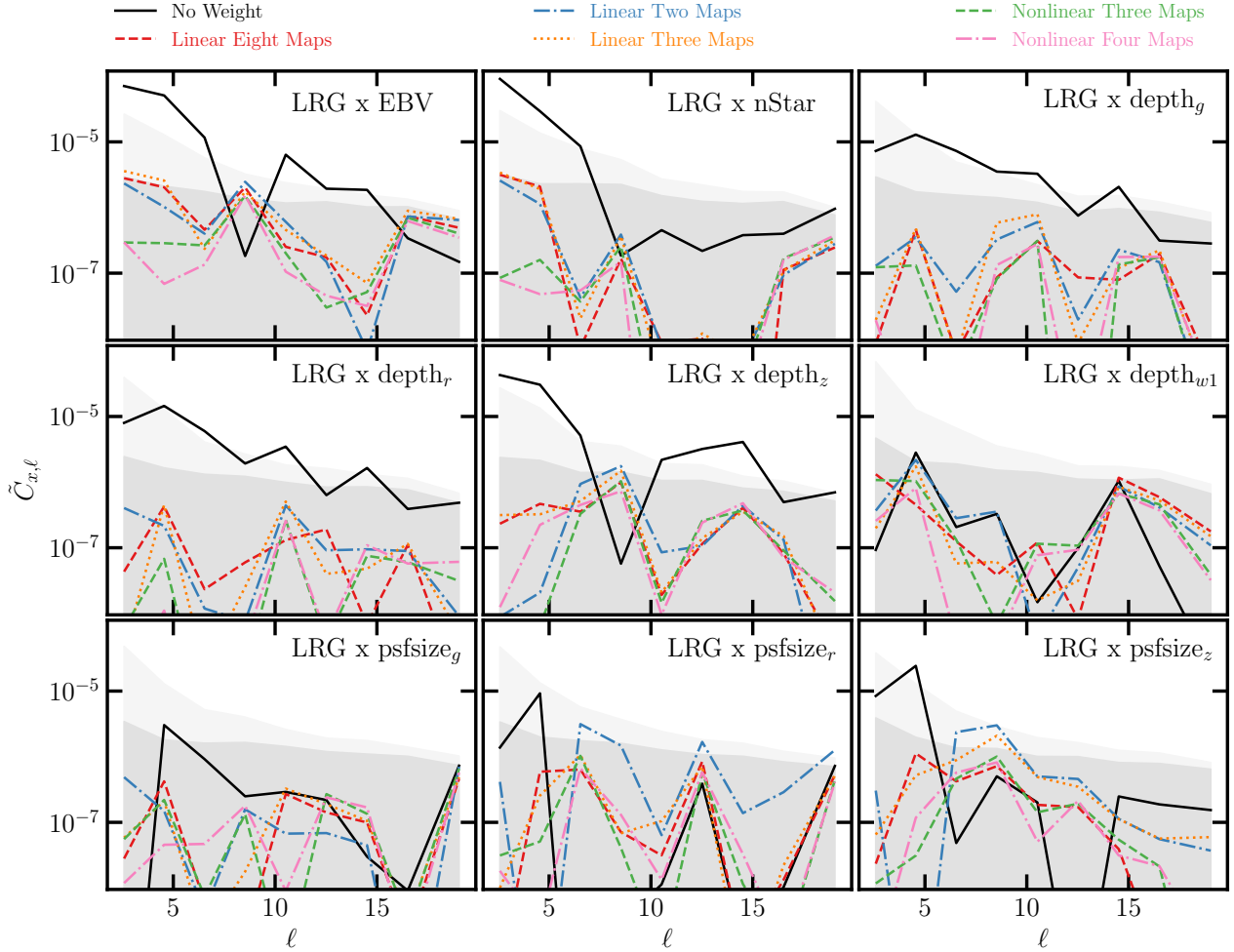
where  $\tilde{C}_{x_i,\ell}$  represents the normalized cross power spectrum, determined from the square of the cross power spectrum between the galaxy density and  $i^{\text{th}}$  imaging map,  $x_i$ , divided by the auto power spectrum of  $x_i$ :

$$\tilde{C}_{x_i,\ell} = \frac{(\tilde{C}_{gx_i,\ell})^2}{\tilde{C}_{x_i x_i,\ell}}. \quad (15)$$

Then, the  $\chi^2$  value for the cross power spectra is calculated via,

$$\chi^2 = \tilde{C}_{X,\ell}^T \mathbb{C}_X^{-1} \tilde{C}_{X,\ell}, \quad (16)$$

where the covariance matrix  $\mathbb{C}_X = \langle \tilde{C}_{X,\ell} \tilde{C}_{X,\ell'} \rangle$  is constructed from the lognormal mocks. These  $\chi^2$  values are measured for every mock realization with the *leave-one-out* technique and compared to the values observed in the DR9 sample with various imaging systematic corrections. [Can you elaborate a more what you mean by that?] Specifically, we use 999 realizations to estimate a covariance matrix and then apply the covariance matrix from the 999 realizations to measure the  $\chi^2$  for the one remaining realization. This process is repeated for all 1000 realizations to construct a histogram for  $\chi^2$ . We only include the bandpower bins from  $\ell = 2$  to 20 with  $\Delta\ell = 2$ , and test for the robustness with higher  $\ell$  modes in Appendix A.

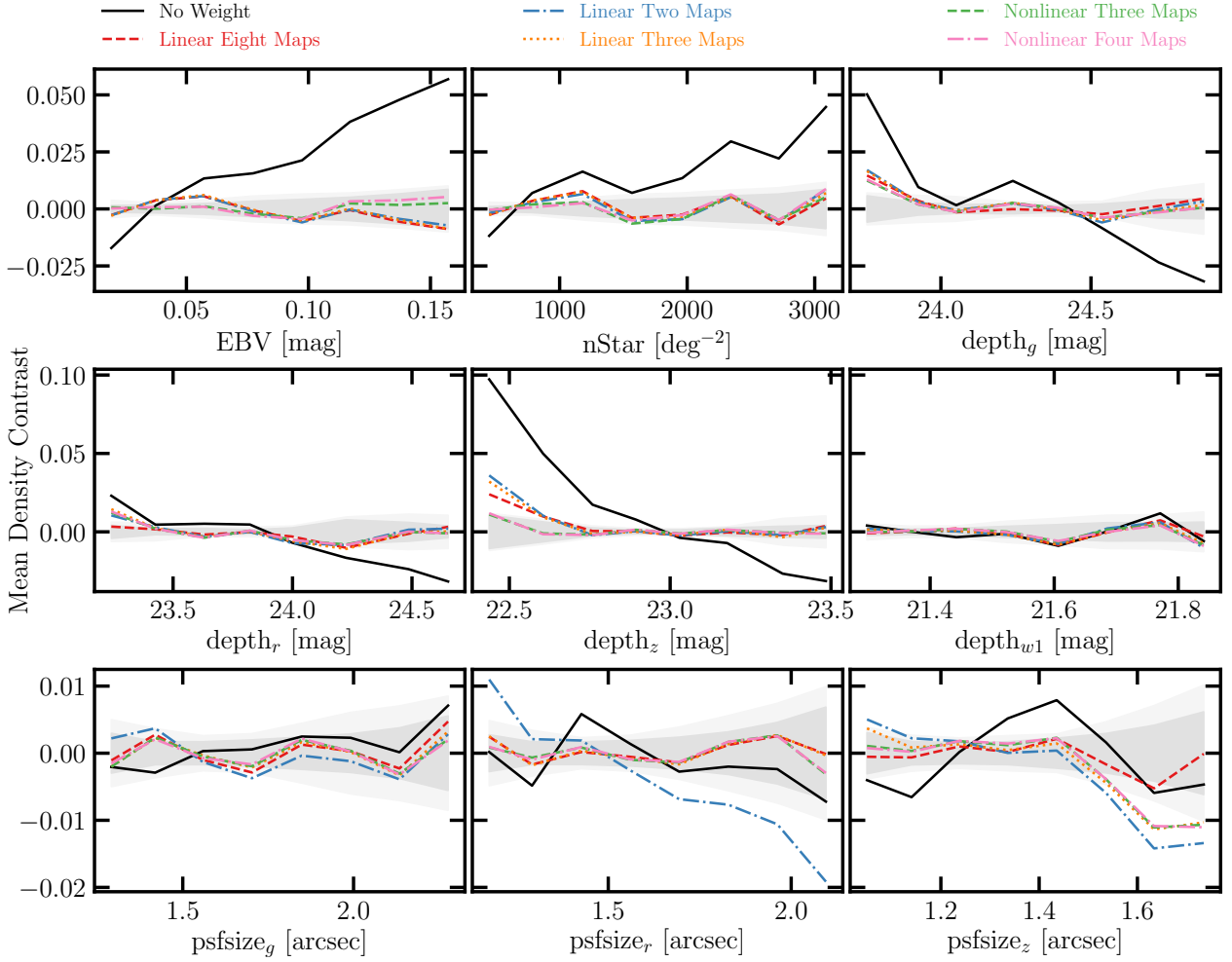


**Figure 6.** The normalized cross power spectra between the DR9 LRG sample and imaging systematic maps: Galactic extinction (EBV), stellar density (nStar), depth in  $grzwl$  (depth $_{grzwl}$ ), and seeing in  $grz$  (psfsize $_{grz}$ ). The black curves display the cross spectra before imaging systematic correction. The red, blue and orange curves represent the results after applying the imaging weights from the linear models trained with *eight maps*, *two maps*, and *three maps*. The green and pink curves display the results after applying the imaging weights from the nonlinear models trained with *three maps* and *four maps*. The dark and light shades represent the 97.5 percentile from cross correlating the imaging systematic maps and the  $f_{NL} = 0$  and 76.9 lognormal density fields, respectively.

Figure 6 shows  $\tilde{C}_X$  from the DR9 LRG sample before and after applying various corrections for imaging systematics. The dark and light shades show the 97.5<sup>th</sup> percentile from the  $f_{NL} = 0$  and 76.9 mocks, respectively. **Without imaging weights, the LRG sample has the highest cross-correlations against extinction, stellar density, and depth in z (solid black curve). There is less significant correlations against depth in the g and r bands, and psfsize in the z band, which could be driven because of the inner correlations between the imaging systematic maps.** First, we consider cleaning the sample with the linear model using two maps (extinction and depth in r) as identified from the Pearson correlation. With linear two maps (dot-dashed blue curve), most of the cross power signals are reduced below statistical uncertainties, especially against extinction, stellar density, and depth. However, the cross power spectra against psfsize in r and z increases slightly on  $6 < \ell < 20$  and  $6 < \ell < 14$ , respectively. Very likely, large-scale cross correlations ( $\ell < 6$ ) are reduced using extinction and depth in the z-band, but there are some residual cross correlations on smaller scales ( $\ell > 6$ ) which cannot be mitigated with our set of two maps. The linear three maps (dotted orange curve) approach alleviates the cross power spectrum against

psfsize in r, but not psfsize in z. On the other hand, nonlinear three maps can reduce the cross correlations against both the r and z-band psfsize maps, which indicates the benefit of using a nonlinear approach. For benchmark, we also show the normalized cross spectra after cleaning the LRG sample with linear eight maps and nonlinear four maps.

Figure 8 (top) [Figure 7 hasn't been discussed yet] shows the histogram of the cross spectrum  $\chi^2$  from 1000 mocks with and without  $f_{NL}$ . The  $\chi^2$  values observed in the DR9 LRG sample are quoted for comparison. Before cleaning, our LRG sample has a cross power spectrum  $\chi^2$  error of 20014.8. After correction with the linear two maps approach, the cross spectrum  $\chi^2$  is reduced to 375.1 with p-value = 0.002. Adding the r-band psfsize, the linear model reduces the  $\chi^2$  down to 195.9 with p-value = 0.044; we can reject the null hypothesis that the DR9 sample with the linear three maps is properly cleaned at 95% confidence. Even though training the linear model with all imaging systematic maps as input gives the lowest cross spectrum  $\chi^2$  of 129.2 (and p-value = 0.239), it potentially makes the analysis more prone to over-fitting and regressing out the true clustering signal, given the inner correlations



**Figure 7.** The mean density contrast of the DR9 LRG sample as a function of the imaging systematic maps: Galactic extinction (EBV), stellar density (nStar), depth in *grzwl* (depth<sub>grzwl</sub>), and seeing in *grz* (psfsize<sub>grz</sub>). The black curves display the results before imaging systematic correction. The red, blue and orange curves represent the relationships after applying the imaging weights from the linear models trained with *eight maps*, *two maps*, and *three maps*. The green and pink curves display the results after applying the imaging weights from the nonlinear models trained with *three maps* and *four maps*. The dark and light shades represent the 68% dispersion of 1000 lognormal mocks with  $f_{\text{NL}} = 0$  and 76.92, respectively.

among the imaging properties (Figure 3). As an alternative, we apply the imaging weights from the nonlinear method with the extinction, z-band depth, and r-band psfsize maps (*nonlinear three maps*). The cross power spectrum  $\chi^2$  is reduced to 79.3 with p-value = 0.594. Adding the stellar density map reduces the cross power spectrum  $\chi^2$  error to 70.9 (p-value = 0.687). Our cross power spectrum diagnostic supports the idea that a nonlinear cleaning approach is needed to properly regress out the remaining spurious fluctuations. We investigate the test with the cross power spectrum up to higher multipoles but find no evidence of remaining systematic errors (see Appendix A).

### 3.4.2 Mean density contrast

We calculate the histogram of the mean density contrast relative to the  $j^{\text{th}}$  imaging property:

$$\delta_{x_j} = (\hat{\rho})^{-1} \frac{\sum_i \rho_i f_{\text{pix},i}}{\sum_i f_{\text{pix},i}}, \quad (17)$$

where the summations are over HEALPix pixels in the bin with similar imaging values for  $x_j$ . For instance, in the absence of systematic error, the mean density contrast from one part of the sky with high extinction, the mean density contrast from one part of the sky with low extinction. We compute the histograms against all other imaging properties (see Figure 2), and construct the total mean density contrast as,

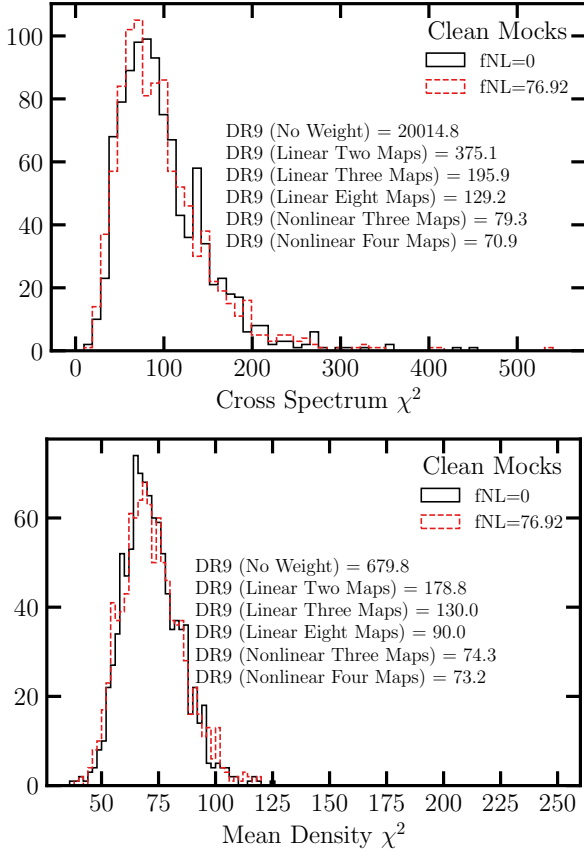
$$\delta_X = [\delta_{x_1}, \delta_{x_2}, \delta_{x_3}, \dots, \delta_{x_9}], \quad (18)$$

and the total residual error as,

$$\chi^2 = \delta_X^T \mathbb{C}_\delta^{-1} \delta_X, \quad (19)$$

where the covariance matrix  $\mathbb{C}_\delta = \langle \delta_X \delta_X \rangle$  is constructed from the lognormal mocks. Figure 7 shows the mean density contrast against the imaging properties for the DR9 LRG sample. The dark and light shades represent the  $1\sigma$  level fluctuations observed in 1000 lognormal density fields respectively with  $f_{\text{NL}} = 0$  and 76.92. The DR9 LRG sample before treatment (solid curve) exhibits a strong trend around 10% against the z-band depth which is consistent with the cross power spectrum. Additionally, there are significant

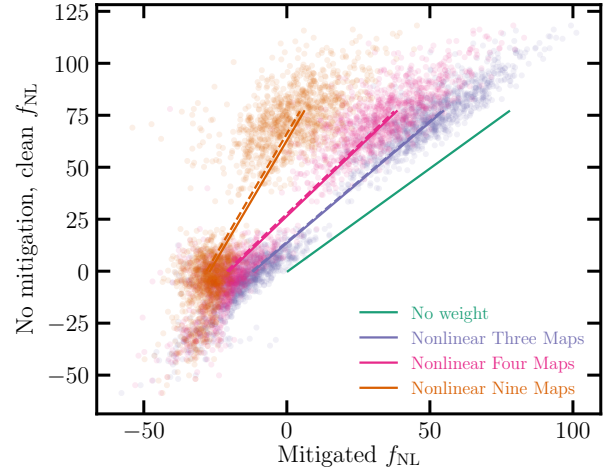




**Figure 8.** The remaining systematic error  $\chi^2$  from the galaxy-imaging cross power spectrum (top) and the mean galaxy density contrast (bottom). The values observed in the DR9 LRG sample before and after linear and non-linear treatments are quoted, and the histograms are constructed from 1000 realizations of clean lognormal mocks with  $f_{\text{NL}} = 0$  and 76.92. **Alex: can you show the DR9 chi2s as vertical lines rather than as text?**

spurious trends against extinction and stellar density at about 5 – 6%. The linear approach is able to mitigate most of the systematic fluctuations with only extinction and depth in the z-band as input; however, a new trend appears against the r-band psfsize map with the *linear two maps* approach (dot-dashed blue curve), which is indicative of the psfsize-related systematics in our sample. This finding is in agreement with the cross power spectrum. We re-train the linear model with three maps, but we still observe around 2% residual spurious fluctuations in the low end of the z-band depth, which implies nonlinear systematic effects exist. We find that the imaging weights from the nonlinear model trained with the three identified maps (or four maps including the stellar density) is capable of reducing the fluctuations below 2%. We experiment with different binning schemes but find consistent results.

Figure 8 (bottom) shows the mean density  $\chi^2$  observed in the mocks with or without  $f_{\text{NL}}$ . We find consistent results regardless of the underlying  $f_{\text{NL}}$ , which supports that our diagnostic is not sensitive to the fiducial cosmology. The values measured in the DR9 LRG sample before and after applying imaging weights are quoted for comparison. The *linear two maps* weights reduce the  $\chi^2$  value from 679.8 (before correction) to 178.8. The p-value = 0 indicates severe remaining systematic effects. Adding the r-band psfsize does not reduce the p-value enough (e.g., greater than 0.05) even though the cleaning method yields a lower  $\chi^2 = 130$ . Training



**Figure 9.** The No mitigated, clean vs mitigated  $f_{\text{NL}}$  values from the  $f_{\text{NL}} = 0$  and 76.9 mocks. The solid (dashed) lines represent the best fit estimates from fitting the mean power spectrum of the clean (contaminated) mocks. The scatter points show the best fit estimates from fitting the individual spectra for the clean mocks.

**Table 2.** Linear parameters employed to de-bias the  $f_{\text{NL}}$  constraints to account for the over-correction issue.

Cleaning Method	$m_1$	$m_2$
Nonlinear Three Maps	1.17	13.95
Nonlinear Four Maps	1.32	26.97
Nonlinear Nine Maps	2.35	63.5

the linear model with all imaging systematic maps returns a more reasonable  $\chi^2 = 90$  and p-value of 0.084. However, regression with all imaging systematic maps as input can lead to the removal of the true clustering signal. With the imaging weights from the *nonlinear three maps* approach, we obtain a  $\chi^2$  value of 74.3 with p-value = 0.392. Re-training the nonlinear approach while adding the stellar density map (*nonlinear four maps*) yields minor improvement:  $\chi^2 = 73.2$  and p-value = 0.422. This indicates that the stellar density trend in the mean LRG density can be explained via the extinction map.

### 3.5 Calibration of mitigation bias

**Santi: How all of this was computed did not seem very clear to me. It seems that this part is what later determines that  $f_{\text{NL}}=0$  from the data. So I think it would be worth explaining it more.** The template-based mitigation of imaging systematics removes some of the true clustering signal, and the amount of the removed signal increases as more maps are fed to the regression. Supported by our remaining systematic test, *nonlinear three maps* is therefore chosen as our default approach and is applied to the mocks (with and without contaminations) to calibrate the  $f_{\text{NL}}$  biases induced by over-correction. Below we describe an approach for the calibration and de-biasing of our  $f_{\text{NL}}$  constraints.

To calibrate for over-correction, we utilize our series of log-normal density fields with and without PNG, with and without systematic effects. The contamination model is based on the linear multivariate approach with the extinction, z-band depth, and r-band psfsize maps as input and parameters drawn from the likelihood constrained by the DR9 LRG sample. The idea is to simulate systematic effects that reflect spurious fluctuations as realistic as

the DR9 LRG sample. For correction, the neural network model is trained and applied to the simulations with various sets of imaging systematic maps as input. Particularly, we consider *nonlinear three maps*, *nonlinear four maps*, and *nonlinear nine maps*. We fit both the mean power spectrum and each individual power spectrum of 1000 realizations. The best fit estimates from the mocks without systematics (and no mitigation applied) are considered as the true  $f_{\text{NL}}$  values and the estimates from the mocks (with the mitigation procedure applied) are considered as the measured  $f_{\text{NL}}$  values. Figure 9 shows the best fit estimates of  $f_{\text{NL}}$  before mitigation vs the best fit  $f_{\text{NL}}$  estimates after mitigation from the mean power spectrum (solid lines) and individual spectra (points) for the mocks without systematics. The dashed lines represent the best fit estimates from the contaminated realizations. The best fit estimates for the individual contaminated mocks are not shown for visual clarity.

Then, a pair of linear parameters are found to transform the  $f_{\text{NL}}$  values after mitigation to those before mitigation,  $f_{\text{NL, no mitigation, clean}} = m_1 f_{\text{NL, mitigated}} + m_2$ . These  $m_1$  and  $m_2$  coefficients for nonlinear three, four, and nine maps are summarized in Table 2. The uncertainty in  $f_{\text{NL}}$  increases by  $m_1 - 1$ . We find that  $m_1 - 1$  determines the added uncertainty in the  $f_{\text{NL}}$  constraints, once the correction coefficients are applied. We expect this effect to be maximum for *nonlinear nine maps* and minimum for *nonlinear three maps*, as the effect of over-correction is more significant when the number of imaging systematics maps as input to the neural network increases.

Emphasize that the mitigation bias or over-correction effect is the same regardless of the mocks being contaminated or not.

## 4 RESULTS

This section presents our  $f_{\text{NL}}$  constraints from the DESI LRG targets. The analysis is not carried out blindly. However, the cleaning methods are decided only based on the cross power spectrum and mean density contrast statistics. Alex: Might be nice to have a table listing all cleaning methods used—since there are more than just the 3 map combinations in 2.1.2 Edmond: Fig12: DESI is : you compute the CI on all the footprint? it is not a combination of the independent measurement of the three regions? same remark for the other figures/tabs. Ashley: explain how the data is combined.

### 4.1 DESI imaging LRG sample

Figure 10 shows the measured power spectrum of the DR9 LRG sample before and after applying imaging weights and the best fit theory curves. The solid line and the grey shade represent respectively the mean power spectrum and  $1\sigma$  error, estimated from the  $f_{\text{NL}} = 0$  lognormal simulations. The differences between various cleaning methods are significant on large scales ( $\ell > 20$ ), but the small scale clustering measurements are consistent. By comparing *linear two maps* to *linear three maps*, we find that the measured clustering power on modes with  $6 \leq \ell < 10$  are noticeably different between the two methods. We associate the differences to the additional map for psfsize in the r-band, which is included in *linear three maps*. On other scales, the differences between *linear three maps* and *linear eight maps* are negligible, supporting the idea that our feature selection procedure has been effective in identifying the primary maps which cause the large-scale excess clustering signal. Comparing *nonlinear three maps* to *linear three maps*, we find that the measured spectra on  $4 \leq \ell < 6$  are very different, probably indicating some nonlinear spurious fluctuations with large

scale characteristics due to extinction. Including stellar density in the nonlinear approach (*nonlinear four maps*) further reduces the excess power relative to the mock power spectrum, in particular on modes between  $2 \leq \ell < 4$ . However, when calibrated on the lognormal simulations, we find that these differences are reversed after accounting for over-correction. Therefore, we associate the subtraction after *nStar* to over fitting.

#### 4.1.1 Calibrated constraints

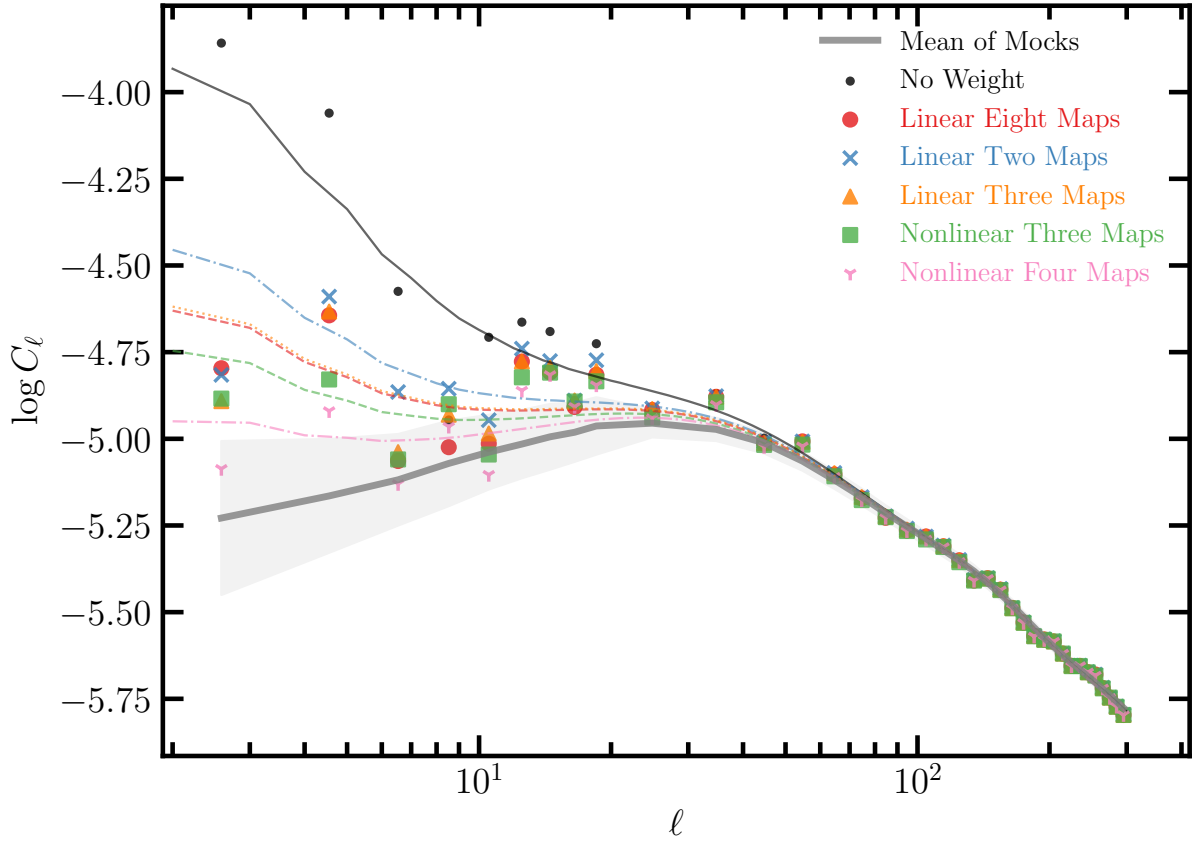
All  $f_{\text{NL}}$  constraints presented here are calibrated for the effect of over correction using the lognormal simulations. Table 3 describes the best fit and marginalized mean estimates of  $f_{\text{NL}}$  from fitting the power spectrum of the DR9 LRG sample before and after cleaning with the nonlinear approach given various combinations for the imaging systematic maps. Figure 11 shows the marginalized probability distribution for  $f_{\text{NL}}$  in the top panel, and the 68% and 95% probability contours for the linear bias parameter and  $f_{\text{NL}}$  in the bottom panel, from our sample before and after applying various corrections for imaging systematics. Overall, we find the maximum likelihood estimates to be consistent among the various cleaning methods. We obtain  $36.08(25.03) < f_{\text{NL}} < 61.44(75.64)$  with  $\chi^2 = 34.6$  for *nonlinear three maps* over 34 degrees of freedom. Accounted for over-correction, we obtain  $36.88(24.87) < f_{\text{NL}} < 63.31(77.78)$  with  $\chi^2 = 35.2$  with the additional stellar density map in the *nonlinear four maps*. With or without *nStar*, the confidence intervals are consistent with each other and more than  $3\sigma$  off from zero PNG. Cleaning the sample with *nonlinear nine maps* weakens our constraints to  $13.10(-15.96) < f_{\text{NL}} < 69.14(91.84)$  with  $\chi^2 = 39.5$ . For comparison, we obtain  $98.14(83.51) < f_{\text{NL}} < 132.89(151.59)$  at 68%(95%) confidence with  $\chi^2 = 44.4$  for the *no weight* approach.

#### 4.1.2 Uncalibrated constraints: robustness tests

Alex: I may have missed this—but do you ever quote the marginalized fnl constraints without mitigation bias (corresponding to Fig. 12)? I think they should probably be in a table and commented on in the text.

Figure 12 shows the probability distributions of  $f_{\text{NL}}$  for various treatments before accounting for the over-correction effect. The method with the largest flexibility and more number of imaging systematic maps is more likely to regress out the clustering signal and return biased  $f_{\text{NL}}$  constraints. As expected, nonlinear nine maps yields a smaller maximum likelihood estimate of  $f_{\text{NL}}$ . Our nonlinear three maps returns a best fit estimate of  $f_{\text{NL}} = 26$  with the 68%(95%) confidence of  $19(9) < f_{\text{NL}} < 41(53)$  and  $\chi^2 = 34.6$ . With the stellar density map included, nonlinear four maps yields a smaller best fit estimates of  $f_{\text{NL}} = 17$  with the error of  $7(-2) < f_{\text{NL}} < 27(38)$ . The nonlinear nine maps gives an asymmetric posterior with the marginalized mean  $f_{\text{NL}} = -9$ , best estimate  $f_{\text{NL}} = -6$  with the error of  $-21(-34) < f_{\text{NL}} < 2.4(12)$ .

Now we proceed to perform some robustness tests and assess how sensitive the  $f_{\text{NL}}$  constraints are to the assumptions made in the analysis or the quality cuts applied to the data. For each case, we re-train the cleaning methods and derive new sets of imaging weights. Accordingly, for the cases where a new survey mask is applied to the data, we re-calculate the covariance matrices using the new survey mask to account for the changes in the survey window and integral constraint effects. Calibrating the mitigation biases for all of these experiments is beyond the scope of this work and redundant, as we



**Figure 10.** The angular power spectrum of the DESI LRG targets before (*No weight*) and after correcting for imaging systematics using the linear and nonlinear methods with their corresponding best fit theory curves. The solid curve and grey shade respectively represent the mean power spectrum and 68% error from the  $f_{\text{NL}} = 0$  mocks.

**Table 3.** The calibrated best fit, marginalized mean, and marginalized 68% (95%) confidence estimates for  $f_{\text{NL}}$  from fitting the DR9 LRG power spectrum before and after correcting for imaging systematic effects.

Footprint	Method	$f_{\text{NL}}$				$\chi^2/\text{dof}$
		Best fit	Mean	68% CL	95% CL	
DESI	No Weight	113.18	115.49	$98.14 < f_{\text{NL}} < 132.89$	$83.51 < f_{\text{NL}} < 151.59$	44.4/34
DESI	Nonlinear Three Maps	47.38	48.81	$36.08 < f_{\text{NL}} < 61.44$	$25.03 < f_{\text{NL}} < 75.64$	34.6/34
DESI	Nonlinear Four Maps	48.92	50.10	$36.88 < f_{\text{NL}} < 63.31$	$24.87 < f_{\text{NL}} < 77.78$	35.2/34
DESI	Nonlinear Nine Maps	49.69	41.91	$13.10 < f_{\text{NL}} < 69.14$	$-15.96 < f_{\text{NL}} < 91.84$	39.5/34

are only interested in the relative shift in the  $f_{\text{NL}}$  constraints after changing the assumptions. Therefore, the absolute scaling of the  $f_{\text{NL}}$  constraints presented here are biased because of the over correction effect. Table 4 summarizes the uncalibrated  $f_{\text{NL}}$  constraints from the DR9 LRG sample. Our tests are as follows:

• **Linear methods:** We find consistent constraints from *linear eight maps* and *linear three maps*,  $26(16) < f_{\text{NL}} < 49(62)$  vs  $26(16) < f_{\text{NL}} < 50(63)$  at 68%(95%) confidence, which suggests that not all of the eight imaging systematic maps are needed to completely mitigate systematic effects. We find  $\sigma(f_{\text{NL}}) \sim 25$  for the linear methods. On the other hand, the nonlinear method with the same set of imaging systematic maps gives a smaller constraint and better fit. For instance, nonlinear three maps gives a smaller best fit value of  $f_{\text{NL}}$  by  $\Delta f_{\text{NL}} = -8$  and  $\chi^2 = 34.6$  compared with 39.6 from linear three maps for the same degrees of freedom and covariance matrix.

• **Imaging regions:** We compare how our constraints from fitting the power spectrum of the whole DESI footprint compares to that from the power spectrum of each imaging region individually, namely BASS+MzLS, DECaLS North, and DECaLS South. Figure 13 shows the 68% and 95% probability contours on  $f_{\text{NL}}$  and  $b$  from each individual region, compared with that from DESI. The cleaning method here is *nonlinear three maps*, and the covariance matrices are estimated from the  $f_{\text{NL}} = 0$  mocks. Overall, we find that the constraints from all imaging surveys are consistent with each other and DESI within 68% confidence. Ignoring the over-correction effect, both BASS+MzLS and DECaLS South yield constraints consistent with  $f_{\text{NL}} = 0$  within 95%, but DECaLS North deviates from zero PNG at more than  $2\sigma$ . This motivates follow-up studies with the spectroscopic sample of LRGs in DECaLS North. **Edmond: maybe comment why the bias is lower in Decals-north? Have you a physical explanation?**

• **Stellar density template ( $n_{\text{Star}}$ ):** Adding the stellar density tem-



**Table 4.** The uncalibrated best fit and marginalized mean estimates for  $f_{\text{NL}}$  from fitting the power spectrum of the DR9 LRG sample before and after correcting for systematics. The estimates are not calibrated for over correction, and thus are subject to mitigation systematics. The number of degrees of freedom is 34 (37 data points - 3 parameters). The lowest mode is  $\ell = 2$  and the covariance matrix is from the  $f_{\text{NL}} = 0$  clean mocks (no mitigation) except for the case with ‘+cov’ in which the covariance matrix is from the  $f_{\text{NL}} = 76.9$  clean mocks (no mitigation).

Footprint	Method	$f_{\text{NL}} + \text{Mitigation Systematics}$				$\chi^2$
		Best fit	Mean	68% CL	95% CL	
DESI	No Weight	113.18	115.49	$98.14 < f_{\text{NL}} < 132.89$	$83.51 < f_{\text{NL}} < 151.59$	44.4
DESI	Linear Eight Maps	36.05	37.72	$26.13 < f_{\text{NL}} < 49.21$	$16.31 < f_{\text{NL}} < 62.31$	41.1
DESI	Linear Two Maps	49.58	51.30	$38.21 < f_{\text{NL}} < 64.33$	$27.41 < f_{\text{NL}} < 78.91$	38.8
DESI	Linear Three Maps	36.63	38.11	$26.32 < f_{\text{NL}} < 49.86$	$16.36 < f_{\text{NL}} < 63.12$	39.6
DESI	Nonlinear Three Maps	28.58	29.79	$18.91 < f_{\text{NL}} < 40.59$	$9.47 < f_{\text{NL}} < 52.73$	34.6
DESI (imag. cut)	Nonlinear Three Maps	29.16	30.57	$19.05 < f_{\text{NL}} < 42.18$	$9.01 < f_{\text{NL}} < 54.81$	35.8
DESI (comp. cut)	Nonlinear Three Maps	28.07	29.48	$18.38 < f_{\text{NL}} < 40.50$	$8.81 < f_{\text{NL}} < 53.10$	34.5
DESI	Nonlinear Four Maps	16.63	17.52	$7.51 < f_{\text{NL}} < 27.53$	$-1.59 < f_{\text{NL}} < 38.49$	35.2
DESI	Nonlinear Nine Maps	-5.87	-9.19	$-21.45 < f_{\text{NL}} < 2.40$	$-33.81 < f_{\text{NL}} < 12.06$	39.5
DESI	Nonlinear Three Maps+ $f_{\text{NL}} = 76.92$ Cov	31.62	33.11	$20.94 < f_{\text{NL}} < 45.24$	$10.56 < f_{\text{NL}} < 59.16$	33.5
BASS+MzLS	Nonlinear Three Maps	15.43	19.01	$-1.17 < f_{\text{NL}} < 39.43$	$-19.19 < f_{\text{NL}} < 63.56$	35.6
BASS+MzLS	Nonlinear Four Maps	13.12	15.39	$-4.59 < f_{\text{NL}} < 35.56$	$-24.88 < f_{\text{NL}} < 59.31$	34.7
BASS+MzLS	Nonlinear Nine Maps	-3.73	-6.34	$-27.11 < f_{\text{NL}} < 13.75$	$-47.44 < f_{\text{NL}} < 33.94$	36.8
BASS+MzLS (imag. cut)	Nonlinear Three Maps	25.03	29.12	$6.16 < f_{\text{NL}} < 52.44$	$-14.22 < f_{\text{NL}} < 80.54$	36.2
BASS+MzLS (comp. cut)	Nonlinear Three Maps	16.99	20.90	$0.26 < f_{\text{NL}} < 41.76$	$-18.30 < f_{\text{NL}} < 67.12$	35.8
DECaLS North	Nonlinear Three Maps	41.02	44.89	$23.33 < f_{\text{NL}} < 66.78$	$4.96 < f_{\text{NL}} < 93.02$	41.1
DECaLS North	Nonlinear Four Maps	31.45	34.78	$14.14 < f_{\text{NL}} < 55.79$	$-5.81 < f_{\text{NL}} < 80.80$	41.2
DECaLS North	Nonlinear Five Maps	55.46	60.44	$36.78 < f_{\text{NL}} < 84.05$	$17.86 < f_{\text{NL}} < 112.81$	38.4
DECaLS North	Nonlinear Nine Maps	0.81	-5.68	$-29.73 < f_{\text{NL}} < 16.71$	$-53.15 < f_{\text{NL}} < 36.19$	45.1
DECaLS North (no DEC cut)	Nonlinear Three Maps	41.05	44.82	$23.58 < f_{\text{NL}} < 66.08$	$6.40 < f_{\text{NL}} < 91.42$	40.7
DECaLS North (imag. cut)	Nonlinear Three Maps	43.27	48.39	$24.60 < f_{\text{NL}} < 72.50$	$4.71 < f_{\text{NL}} < 101.42$	35.1
DECaLS North (comp. cut)	Nonlinear Three Maps	40.55	44.63	$22.41 < f_{\text{NL}} < 67.11$	$3.95 < f_{\text{NL}} < 94.06$	41.4
DECaLS South	Nonlinear Three Maps	31.24	33.21	$14.89 < f_{\text{NL}} < 52.40$	$-5.11 < f_{\text{NL}} < 74.35$	30.2
DECaLS South	Nonlinear Four Maps	14.34	6.28	$-21.19 < f_{\text{NL}} < 30.01$	$-53.63 < f_{\text{NL}} < 49.51$	31.9
DECaLS South	Nonlinear Five Maps	33.79	37.50	$17.71 < f_{\text{NL}} < 57.42$	$-0.31 < f_{\text{NL}} < 80.94$	30.8
DECaLS South	Nonlinear Nine Maps	-36.76	-32.01	$-49.38 < f_{\text{NL}} < -13.61$	$-65.26 < f_{\text{NL}} < 7.52$	31.5
DECaLS South (no DEC cut)	Nonlinear Three Maps	43.79	46.79	$30.16 < f_{\text{NL}} < 63.41$	$16.38 < f_{\text{NL}} < 82.72$	23.8
DECaLS South (imag. cut)	Nonlinear Three Maps	26.47	23.36	$3.18 < f_{\text{NL}} < 47.84$	$-57.69 < f_{\text{NL}} < 71.39$	30.0
DECaLS South (comp. cut)	Nonlinear Three Maps	29.62	31.76	$13.00 < f_{\text{NL}} < 51.58$	$-9.78 < f_{\text{NL}} < 74.28$	29.7

plate (*nonlinear four maps*) does not change the constraints from BASS+MzLS much, but it shifts the  $f_{\text{NL}}$  distributions to lower values in DECaLS North and DECaLS South by  $0.5\sigma$  and  $\sigma$ , respectively, reconciling all constraints with  $f_{\text{NL}} = 0$ . We note that differences are more significant when all nine maps are used as input. This is somewhat expected as cleaning the data with more imaging systematic maps is more prone to the over-correction issue. We find that the shifts in  $f_{\text{NL}}$  from adding *nStar* are reversed after accounting for the over correction effect. Comparing the  $f_{\text{NL}}$  constraints from *nonlinear four maps* and *nonlinear three maps* in Table 3 to those in 4, we can argue that the  $f_{\text{NL}}$  shifts after adding *nStar* are probably caused by the over-correction issue from the chance correlations between the stellar density map and large-scale structure.

• **Pixel completeness (comp. cut):** We discard pixels with fractional completeness less than half to assess the effect of partially complete pixels on  $f_{\text{NL}}$ . This cut removes 0.6% of the survey area, and no changes in the  $f_{\text{NL}}$  constraints are observed.

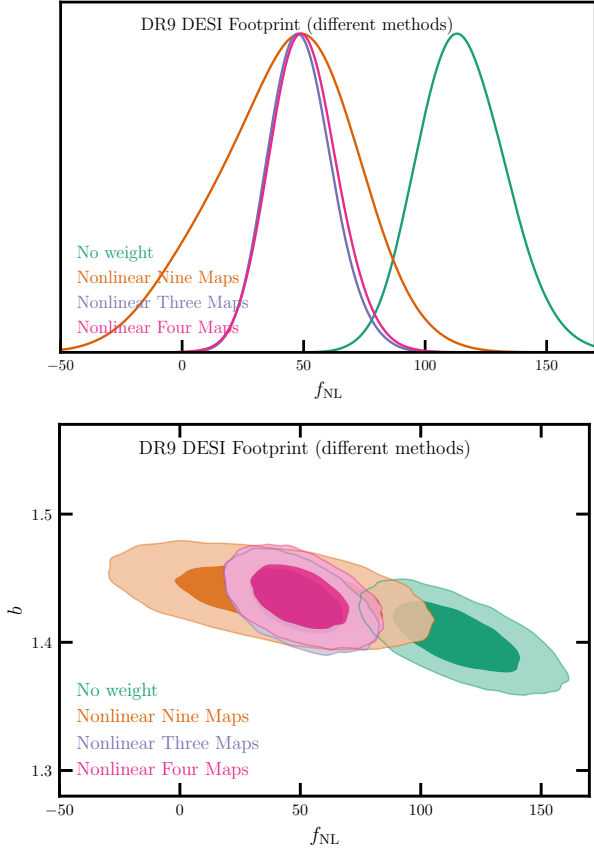
• **Imaging quality (imag. cut):** Pixels with poor photometry are removed from our sample by applying the following cuts on imaging;  $E[B-V] < 0.1$ ,  $nStar < 3000$ ,  $depth_g > 23.2$ ,  $depth_r > 22.6$ ,  $depth_z > 22.5$ ,  $psfsize_g < 2.5$ ,  $psfsize_r < 2.5$ , and  $psfsize_z < 2$ . Although these cuts remove 8% of the survey mask, there is a negligible impact on the best fit  $f_{\text{NL}}$  from fitting the DESI power spectrum. However, when each region is fit individually, the BASS+MzLS constraint shift toward higher values of  $f_{\text{NL}}$  by

approximately  $\Delta f_{\text{NL}} \sim 10$ , whereas the constraints from DECaLS North and DECaLS South do not change significantly.

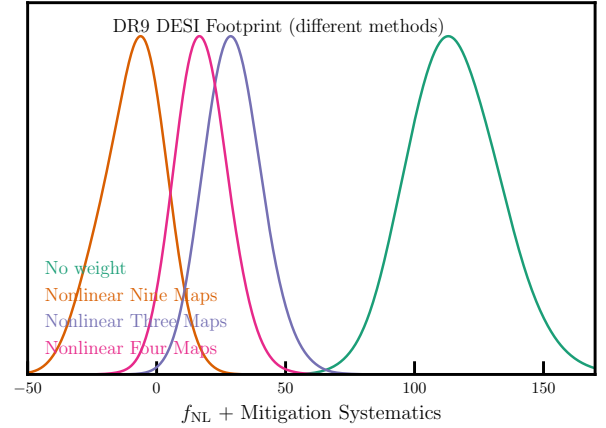
• **Covariance matrix (cov):** We fit the power spectrum of our sample cleaned with *nonlinear three maps* correction, but use the covariance matrix constructed from the  $f_{\text{NL}} = 76.92$  mocks. With the alternative covariance, a 12% increase in the  $\sigma_{f_{\text{NL}}}$  is observed. We also find that the best fit and marginalized mean estimates of  $f_{\text{NL}}$  increase by 10 – 11%. Overall, we find that the differences are not significant in comparison to the statistical precision.

• **External maps (CALIBZ+HI):** The neural network five maps correction includes the additional maps for HI and CALIBZ. With this correction, the best fit  $f_{\text{NL}}$  increases from 41.02 to 55.46 for DECaLS North and from 31.24 to 33.79 for DECaLS South, which might suggest that adding HI and CALIBZ increases the input noise, and thus negatively impacts the performance of the neural network model. This test is not performed on BASS+MzLS due to a lack of coverage from the CALIBZ map.

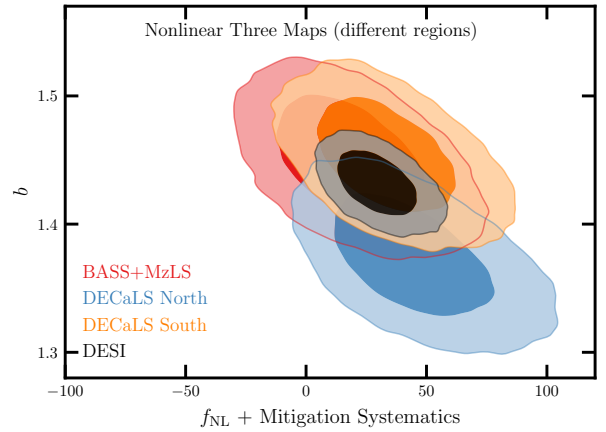
• **Declination mask (no DEC cut):** The fiducial mask removes the disconnected islands in DECaLS North and regions with  $DEC < -30$  in DECaLS South, where there is a high likelihood of calibration issues as different standard stars are used for photometric calibrations. We analyze our sample without these cuts, and find that the best fit and marginalized  $f_{\text{NL}}$  mean estimates from DECaLS South shift significantly to higher values of  $f_{\text{NL}}$  by  $\Delta f_{\text{NL}} \sim 10$ , which supports the issue of photometric systematics in the DECaLS South region below  $DEC = -30$ . On the other hand, the constraints from DECaLS North do not change significantly, indicating the islands



**Figure 11.** The calibrated constraints from the DR9 LRG sample. *Top*: probability distribution for  $f_{\text{NL}}$  marginalized over the shotnoise and bias. *Bottom*: 68% and 95% probability distribution contours for the bias and  $f_{\text{NL}}$  from the DR9 LRG sample before and after applying nonlinear cleaning methods. The lognormal mocks are used to calibrate these distributions for over correction.



**Figure 12.** Same as Figure 11 but without accounting for over correction.



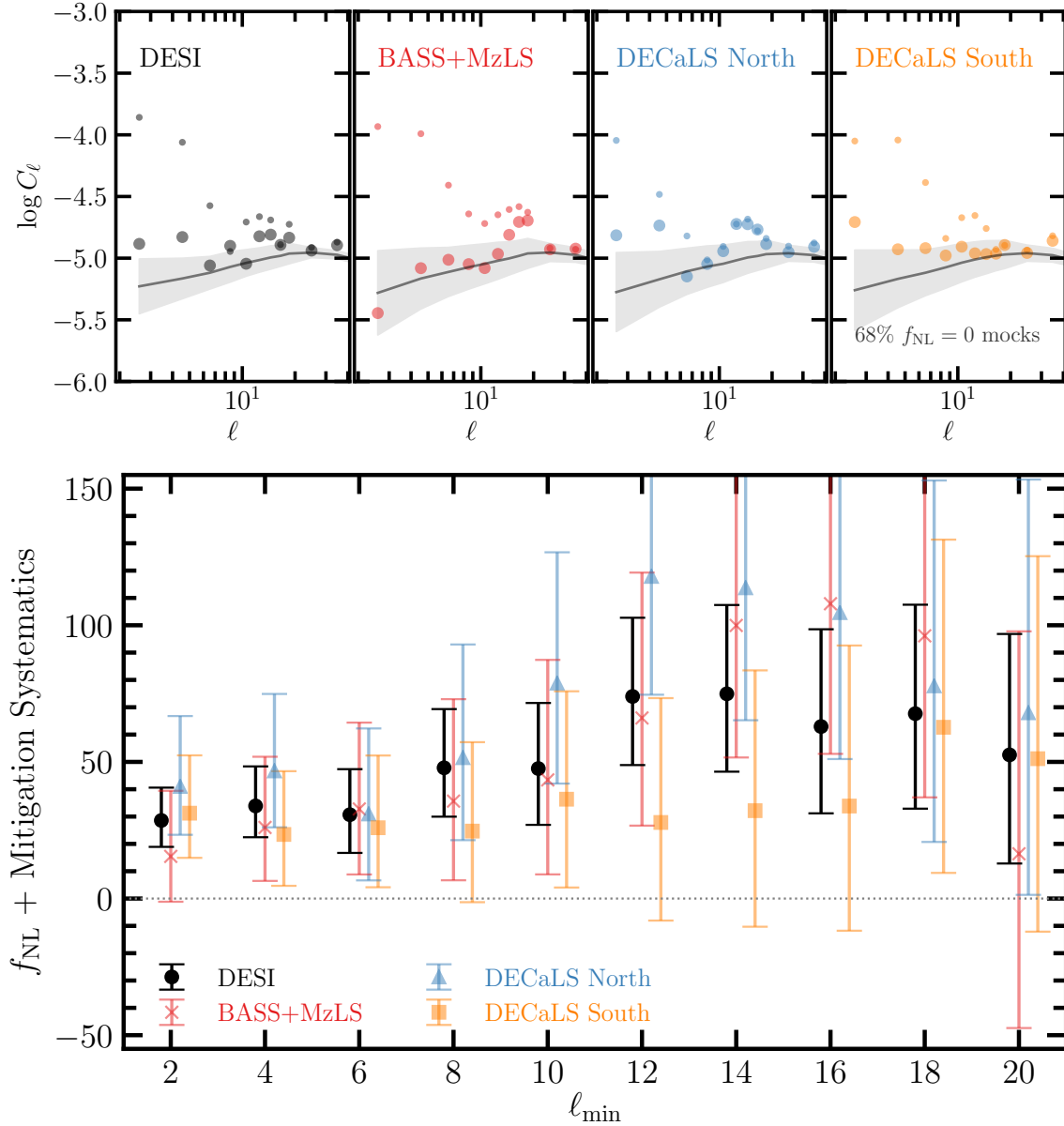
**Figure 13.** The uncalibrated 2D constraints from the DR9 LRG sample for each imaging survey compared with that for the whole DESI footprint. The dark and light shades represent the 68% and 95% confidence intervals, respectively. *Edmond: maybe comment why the bias is lower in Decals-north? Have you a physical explanation?*

do not induce significant contaminations. [Did you try cutting out more of DECaLS North, such as everything below the equator?]

• **Scale dependence (varying  $\ell_{\text{min}}$ ):** We raise the value of the lowest harmonic mode  $\ell_{\text{min}}$  used for the likelihood evaluation during MCMC. This is equivalent to decreasing the highest scale of measurement in the power spectrum. By doing so, we anticipate a reduction in the impact of imaging systematics on  $f_{\text{NL}}$  inference as lower  $\ell$  modes are more likely to be contaminated. Figure 14 illustrates the power spectra before and after the correction with *nonlinear three maps* in the top panel. The bottom panel shows the marginalized mean and 68% error on  $f_{\text{NL}}$  with *nonlinear three maps* for the DESI, BASS+MzLS, DECaLS North, and DECaLS South regions. *ALEX: We find that the mean estimates of  $f_{\text{NL}}$  slightly shifts to higher values on scales  $12 < \ell < 18$  in DECaLS North and BASS+MzLS when higher  $\ell_{\text{min}}$  is used. This is the opposite behavior from what one would expect if there were just a giant systematics induced spike at low  $\ell$ . So it shows that the issue here is more subtle than what one would have initially suspected. Edmond: Fig10 / 13: do you have an explanation why there is this remaining bump at  $\ell=10-12$  ?? is a correction artefact ? you should mention it in the text and comment the effect since it increase the value of  $f_{\text{NL}}$  when you chose higher  $\ell_{\text{min}}$ .*

## 4.2 Summary

In summary, we find that the nonlinear methods outperform the linear methods in removing the excess clustering signal on large scales. Adding the stellar density map results in significant changes, however when accounted for the mitigation bias, all methods recover the same maximum likelihood estimate. With calibration on the lognormal mocks, the neural network three maps and four maps approaches show  $f_{\text{NL}}$  detection at more than  $2\sigma$  confidence. The most flexible neural network method with nine maps returns a bigger associated uncertainty which is consistent with  $f_{\text{NL}} = 0$ . We also run various tests with cuts on the DR9 sample or changing the configuration or details of the analysis. Overall, we find consistent results across sub imaging surveys within DESI. However, our results show that a declination cut at  $\text{DEC} = -30$  is necessary for DECaLS South to avoid potential calibration issues. Our analysis does not show a statistical demand for including external templates for HI and CALIBZ, using a different covariance matrix, or imposing additional cuts on the DR9 based on imaging and pixel completeness. We also obtain robust results regardless of the largest scale used for constraining  $f_{\text{NL}}$ .



**Figure 14.** Top: The measured power spectra before (small circles) and after *nonlinear three maps* (big circles). Bottom: The uncalibrated  $f_{NL}$  constraints vs the lowest  $\ell$  mode from the DR9 LRG sample cleaned with *nonlinear three maps* for the DESI, BASS+MzLS, DECaLS North, and DECaLS South footprints. The points represent marginalized mean estimates of  $f_{NL}$  and error bars represent 68% confidence. The  $f_{NL}$  values are subject to over correction.

## 5 CONCLUSIONS

We have presented constraints on the local primordial non-Gaussianity parameter  $f_{NL}$  from the angular power spectrum of LRGs from the DESI imaging DR9. We infer the redshift distribution of LRGs from early spectroscopy during DESI Survey Validation (Figure 1). Our LRG sample includes more than 14 million targets covering around 18,000 square degrees in the redshift range of  $0.2 < z < 1.35$ . Our analysis utilizes the scale-dependent bias effect that primarily comes from large scales; thus, it is very sensitive to systematic errors caused by photometric calibration issues, survey depth variations, and Milky Way foregrounds (Figure 2).

We use the FFTLog algorithm to model the angular clustering on large scales or multipoles as low as  $\ell = 2$  (Figure 4). We simulate lognormal density fields with DESI-like LRG angular and redshift

distributions to validate the pipeline, estimate covariance matrices, and characterize remaining systematic errors. Our mock test reveals that the distribution of power spectra on large scales is asymmetric (Figure 5). We demonstrate our likelihood inference benefits from fitting the log transformation of the power spectrum.

Multivariate linear and neural network-based regression models are applied to regress out spurious fluctuations in the LRG density field against various maps for the extinction, survey depth, astronomical seeing, neutral hydrogen column density, and stellar density. Feature selection uses the Pearson correlation and the Spearman correlation coefficients to reduce the likelihood of over-correction, i.e., removing the clustering signal (Figure 3). The LRG density map is cross-correlated against the imaging systematic maps (Figure 6), and the mean LRG density is calculated for different re-



gions with similar imaging to look for systematic trends in the mean density (Figure 7). Using the mean density and cross-power spectrum diagnostics, we quantify the remaining systematics against lognormal simulations (Figure 8). We identify the extinction, z-band depth, and r-band seeing as the primary sources of systematic error. Our simulation-based tests reveal that the DESI LRG targets cleaned with linear three maps suffers from significant remaining systematic error primarily due to depth variations. We observe that the nonlinear mitigation approach reduces the excess clustering signal more effectively.

We apply our cleaning methods to the lognormal mocks with and without PNG, with and without systematic effects, to calibrate the level of mitigation biases introduced in our constraints (Table 2 and Figure 9). With three maps, we obtain best fit estimates which are inconsistent with zero at more than 95% confidence. Adding local stellar density to the list of maps used for cleaning the LRG sample does not influence the constraints. However, using the combination of all imaging systematic maps and stellar density yields an asymmetric likelihood distribution with larger uncertainty and consistent with  $f_{\text{NL}} = 0$  at 95% confidence (Table 3). **This is interesting as the same covariance matrix is used for all; but we find that the nonlinear nine maps approach introduces a larger bias by having a larger multiplicative parameter,  $m_1$ , Table 2.** Overall, our nonlinear cleaning methods return consistent best fit estimates of  $f_{\text{NL}} \sim 47 - 50$  (Figure 11). We run multiple robustness tests but find no significant changes other than that there is spurious correlation against stellar density in the NGC and potential calibration issues in the SGC below  $\text{DEC} < -30$  (Table 4). Our constraints are consistent with each other when each imaging region is fit separately and/or the lowest mode used is increased (Figure 13 and Figure 14). Assuming Planck’s measurement of  $f_{\text{NL}}$  is accurate a priori, our results indicate some unknown calibration error. **Santi: Potentially, this could also indicate a more complicated type of PNG (scale dependent, beyond quadratic fnl, non-local...) I would also discuss possible implications from assuming a different cosmology.** Our results also suggest follow-up investigations of stellar contamination and depth-related variations in the spectroscopic sample of DESI LRGs. **Hui: cite LRG redshift success rate, stellar contamination rate. Noting possible improvements when we have all spectra data available.** On the other hand, calibrating simulations for the over correction effect might not be feasible for DESI spectroscopy. So, a simulation based forward model approach for estimating the imaging weights can become helpful to reduce the dimensionality of the imaging parameter space. We leave the idea of combining the forward-modeling and backward-modeling cleaning approaches to future work. Another source of theoretical uncertainty is that our analysis considers the halo bias depends on halo mass only, which is shown to be a naive assumption. Improving the methods for handling imaging systematics and calibrating the uncertainties around halo assembly bias, together, will enable us to derive reliable, accurate, and precise  $f_{\text{NL}}$  results.

## ACKNOWLEDGEMENTS

MR is supported by XXXXXX. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 853291). FB is a University Research Fellow. MR would like to thank CCAPP, in particular, John Beacom and Lisa Colarosa, for their hospitality and support. We would like to thank Tanveer Karim and Sukhdeep Singh for helpful discussions, and Rongpu Zhou for providing the raw galaxy density maps and imaging systematic maps. We acknowledge the support and resources from the Ohio Supercomputer Center (OSC; Center 1987). This research has made use of the arXiv preprint server, NASA's Astrophysics Data System, Github's online software development platform, and many open-source software, such as Pytorch, Nbodykit, HEALPix, Fitsio, Scikit-Learn, NumPy, SciPy, Pandas, IPython, and Jupyter.

This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions.

The DESI Legacy Imaging Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory. Legacy Surveys also uses data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), a project of the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Legacy Surveys was supported by: the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility; the U.S. National Science Foundation, Division of Astronomical Sciences; the National Astronomical Observatories of China, the Chinese Academy of Sciences and the Chinese National Natural Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy.

The authors are honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation."

## DATA AVAILABILITY

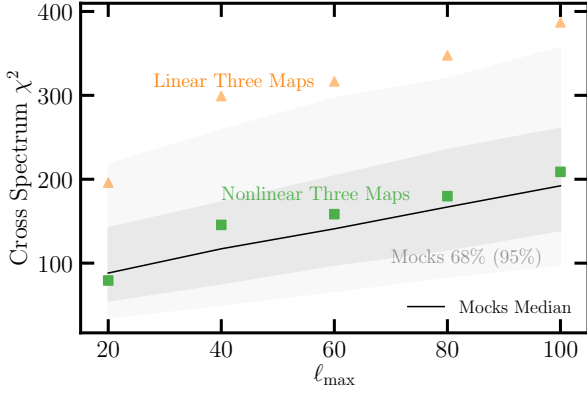
The DR9 catalogues from the DESI Legacy Imaging Surveys are publicly available at <https://www.legacysurvey.org/dr9/>. The software used for cleaning the imaging data is available at <https://github.com/mehdirezaie/sysnetdev>. The lognormal mock catalogs can be made available upon reasonable request.

## REFERENCES

- Alvarez M., et al., 2014, arXiv e-prints, p. arXiv:1412.4671  
Arlot S., Celisse A., 2010, *Statistics Surveys*, 4, 40  
Bahr-Kalus B., Bertacca D., Verde L., Heavens A., 2021, *J. Cosmology Astropart. Phys.*, 2021, 027  
Baldauf T., Seljak U., Senatore L., 2011a, *Journal of Cosmology and Astroparticle Physics*, 2011, 006  
Baldauf T., Seljak U., Senatore L., Zaldarriaga M., 2011b, *Journal of Cosmology and Astroparticle Physics*, 2011, 031  
Barreira A., 2020, *J. Cosmology Astropart. Phys.*, 2020, 031  
Barreira A., 2022, *J. Cosmology Astropart. Phys.*, 2022, 013  
Barreira A., Cabass G., Schmidt F., Pillepich A., Nelson D., 2020, *J. Cosmology Astropart. Phys.*, 2020, 013  
Basset B. A., Tsujikawa S., Wands D., 2006, *Reviews of Modern Physics*, 78, 537  
Bautista J. E., et al., 2021, *MNRAS*, 500, 736  
Beutler F., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 1065  
Beutler F., Biagetti M., Green D., Slosar A., Wallisch B., 2019, *Physical Review Research*, 1, 033209  
Biagetti M., 2019, *Galaxies*, 7, 71  
Bouwens R. J., et al., 2015, *ApJ*, 803, 34  
Cabass G., Ivanov M. M., Philcox O. H. E., Simonović M., Zaldarriaga M., 2022, *Phys. Rev. D*, 106, 043506  
Cahn R. N., Slepian Z., Hou J., 2021, arXiv preprint arXiv:2110.12004  
Castorina E., Moradinezhad Dizgah A., 2020, *J. Cosmology Astropart. Phys.*, 2020, 007  
Castorina E., et al., 2019, *J. Cosmology Astropart. Phys.*, 2019, 010  
Center O. S., 1987, Ohio Supercomputer Center, <http://osc.edu/ark:/19495/f5s1ph73>  
Chapman M. J., et al., 2022, *MNRAS*, 516, 617  
Chaussidon E., et al., 2022, *Monthly Notices of the Royal Astronomical Society*, 509, 3904  
Chon G., Challinor A., Prunet S., Hivon E., Szapudi I., 2004, *Monthly Notices of the Royal Astronomical Society*, 350, 914  
Coles P., Jones B., 1991, *Monthly Notices of the Royal Astronomical Society*, 248, 1  
D'Amico G., Lewandowski M., Senatore L., Zhang P., 2022, arXiv e-prints, p. arXiv:2201.11518  
DES Collaboration et al., 2016, *Monthly Notices of the Royal Astronomical Society*, 460, 1270  
DESI Collaboration et al., 2016, arXiv preprint arXiv:1611.00036  
DESI Collaboration in prep, XXXX  
Dalal N., Dore O., Huterer D., Shirokov A., 2008, *Physical Review D*, 77, 123514  
De Mattia A., Ruhlmann-Kleider V., 2019, *Journal of Cosmology and Astroparticle Physics*, 2019, 036  
Desjacques V., Seljak U., 2010, *Classical and Quantum Gravity*, 27, 124011  
Dey A., et al., 2018, arXiv preprint arXiv:1804.08657  
Eisenstein D. J., et al., 2001, *The Astronomical Journal*, 122, 2267  
Fang X., Krause E., Eifler T., MacCrann N., 2020, *Journal of Cosmology and Astroparticle Physics*, 2020, 010  
Fillmore J. A., Goldreich P., 1984, *Astrophysical Journal*, 281, 1  
Flaugher B., et al., 2015, *The Astronomical Journal*, 150, 150  
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
Gaia Collaboration et al., 2018, *A&A*, 616, A1

- Giannantonio T., Ross A. J., Percival W. J., Crittenden R., Bacher D., Kil-  
binger M., Nichol R., Weller J., 2014, *Physical Review D*, 89, 023511
- Gil-Marín H., et al., 2020, *MNRAS*, 498, 2492
- Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke  
M., Bartelmann M., 2005, *The Astrophysical Journal*, 622, 759
- Guth A. H., Kaiser D. I., 2005, *Science*, 307, 884
- HI4PI Collaboration et al., 2016, *A&A*, 594, A116
- Heinrich C., Doré O., 2022, in *American Astronomical Society Meeting  
Abstracts*, p. 202.03
- Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F.,  
2002, *The Astrophysical Journal*, 567, 2
- Ho S., et al., 2015, *J. Cosmology Astropart. Phys.*, 2015, 040
- Huterer D., Cunha C. E., Fang W., 2013, *Monthly Notices of the Royal  
Astronomical Society*, 432, 2945
- Kitanidis E., et al., 2020, *Monthly Notices of the Royal Astronomical Soci-  
ety*, 496, 2262
- Kofman L., Linde A., Starobinsky A. A., 1994, *Physical Review Letters*, 73,  
3195
- Komatsu E., Spergel D. N., 2001, *Physical Review D*, 63, 063002
- Lazeyras T., Barreira A., Schmidt F., Desjacques V., 2023, *J. Cosmology  
Astropart. Phys.*, 2023, 023
- Loshchilov I., Hutter F., 2016, *arXiv e-prints*, p. arXiv:1608.03983
- Loshchilov I., Hutter F., 2017, *arXiv e-prints*, p. arXiv:1711.05101
- Lyth D. H., Liddle A. R., 2009, *The primordial density perturbation: Cos-  
mology, inflation and the origin of structure*. Cambridge University  
Press
- Maldacena J., 2003, *Journal of High Energy Physics*, 2003, 013
- Meisner A. M., Lang D., Schlegel D. J., 2018, *Research Notes of the Amer-  
ican Astronomical Society*, 2, 1
- Merz G., et al., 2021, *Monthly Notices of the Royal Astronomical Society*,  
506, 2503
- Mueller E.-M., Percival W. J., Ruggeri R., 2019, *MNRAS*, 485, 4160
- Mueller E.-M., et al., 2022, *Monthly Notices of the Royal Astronomical  
Society*
- Myers A. D., et al., 2022, *arXiv e-prints*, p. arXiv:2208.08518
- Nair V., Hinton G. E., 2010, in *Proceedings of the 27th international con-  
ference on machine learning (ICML-10)*, pp 807–814
- Padmanabhan N., et al., 2007, *MNRAS*, 378, 852
- Peacock J., Nicholson D., 1991, *Monthly Notices of the Royal Astronomical  
Society*, 253, 307
- Philcox O. H., 2022, *Physical Review D*, 106, 063501
- Planck Collaboration et al., 2019, *arXiv preprint arXiv:1905.05697*
- Prakash A., et al., 2016, *The Astrophysical Journal Supplement Series*, 224,  
34
- Pullen A. R., Hirata C. M., 2013, *Publications of the Astronomical Society  
of the Pacific*, 125, 705
- Reid B. A., Verde L., Dolag K., Matarrese S., Moscardini L., 2010, *J. Cos-  
mology Astropart. Phys.*, 2010, 013
- Rezaie M., et al., 2021, *Monthly Notices of the Royal Astronomical Society*,  
506, 3439
- Riquelme W., et al., 2022, *arXiv preprint arXiv:2209.07187*
- Ross A. J., et al., 2011, *Monthly Notices of the Royal Astronomical Society*,  
417, 1350
- Ross A. J., et al., 2013, *MNRAS*, 428, 1116
- Ross A. J., et al., 2020, *MNRAS*, 498, 2354
- Sabti N., Muñoz J. B., Blas D., 2021, *J. Cosmology Astropart. Phys.*, 2021,  
010
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *The Astrophysical Journal*,  
500, 525
- Schmittfull M., Seljak U., 2018, *Phys. Rev. D*, 97, 123540
- Simons Observatory et al., 2019, *Journal of Cosmology and Astroparticle  
Physics*, 2019, 056
- Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, *Journal of  
Cosmology and Astroparticle Physics*, 2008, 031
- Tegmark M., et al., 2004, *Phys. Rev. D*, 69, 103501
- Wang M. S., Beutler F., Bacon D., 2020, *MNRAS*, 499, 2598
- Weaverdyck N., Huterer D., 2021, *MNRAS*, 503, 5061
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G.,  
Roza E., 2013, *Physics reports*, 530, 87
- Wilson M. J., Peacock J. A., Taylor A. N., de la Torre S., 2017, *Monthly  
Notices of the Royal Astronomical Society*, 464, 3121
- Wright E. L., et al., 2010, *AJ*, 140, 1868
- Xavier H. S., Abdalla F. B., Joachimi B., 2016, *Monthly Notices of the Royal  
Astronomical Society*, 459, 3693
- Zhou R., et al., 2021, *Monthly Notices of the Royal Astronomical Society*,  
501, 3309
- Zhou R., et al., 2022, *arXiv preprint arXiv:2208.08515*
- Zou H., et al., 2017, *Publications of the Astronomical Society of the Pacific*,  
129, 064101
- de Putter R., Gleyzes J., Doré O., 2017, *Physical Review D*, 95, 123507





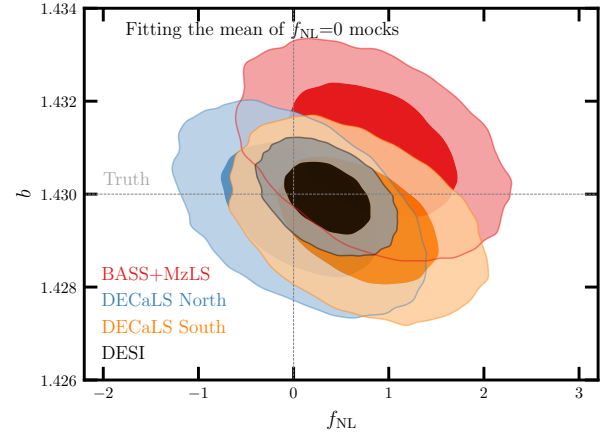
**Figure A1.** The cross power spectrum  $\chi^2$  between the LRG density and imaging systematic maps as a function of the highest mode  $\ell_{\max}$  when the sample is cleaned with the linear (triangles) and nonlinear (squares) three maps. The lowest mode is fixed at  $\ell_{\min} = 2$ . The solid curve and dark (light) shade represent the median value and 68% (95%) confidence regions, estimated from the  $f_{\text{NL}} = 0$  mocks.

## APPENDIX A: SCALE DEPENDENT SYSTEMATICS

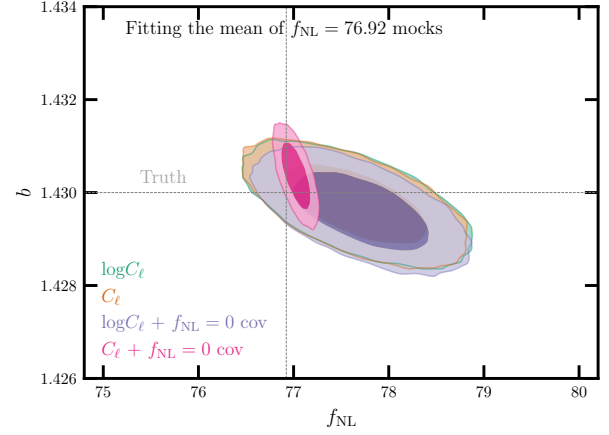
We investigate the dependence of statistical significance for the cross power spectrum  $\chi^2$  between the DR9 LRG targets and imaging maps. We extend the highest harmonic mode from  $\ell = 20$  to 100, which corresponds to density fluctuations on scales below 2 degrees. As shown in Figure A1, the solid line represents the median of the cross power spectrum  $\chi^2$  measured from the  $f_{\text{NL}} = 0$  mocks as a function of highest mode,  $\ell_{\max}$  increases from 20 to 100. The red circles and blue crosses show the chi2 values for our sample cleaned respectively with the linear and neural network approaches, both with *three maps*. This test supports that our cross spectrum  $\chi^2$  values are stable.

## APPENDIX B: LOGNORMAL MOCKS

Corner plots of the PNG parameter  $f_{\text{NL}}$  and bias coefficient are shown in Fig. B1 for fitting the mean power spectrum of mocks, with and without  $f_{\text{NL}}$ . Best fit estimates, marginalized mean,  $1\sigma$  and  $2\sigma$  confidence intervals are summarized in Tab. B1. Fig B1 (right) shows confidence contours for different combinations of target variable (e.g., either power spectrum or its log transform) and covariance matrix. First we attempt to understand the impact of covariance on confidence intervals. We fit the mean power spectrum of  $f_{\text{NL}} = 76.9$  mocks or its log transformation using covariance matrices constructed from the same set of mocks or from the  $f_{\text{NL}} = 0$  mocks. When covariance is consistent with mean, the difference between fitting power spectrum and log of it is only 2%. If a wrong covariance is used for the log power, the effect is only 7%. However, when mean power spectrum of the  $f_{\text{NL}} = 76.9$  mocks is fit using the covariance matrix estimated from the  $f_{\text{NL}} = 0$  mocks, the constraints improve by a factor of 5, simply due to a false higher signal to noise ratio. Therefore, we argue that fitting logarithm of power spectrum would remove the need for having  $f_{\text{NL}}$ -dependent covariance matrices and make the constraints less sensitive to covariance construction. Fig. B1 shows the confidence contours for  $f_{\text{NL}} = 0$  mocks when fit is done to the log of mean spectra of  $f_{\text{NL}} = 0$  mocks for the different regions. We find that the underlying true  $f_{\text{NL}}$  value is recovered within  $2\sigma$  confidence. Add a paragraph for the constraining power vs fsky.



**Figure B1.** 68% and 95% confidence contours from the mean power spectrum of the  $f_{\text{NL}} = 0$  mocks for the DESI footprint and sub-imaging surveys. The truth values are represented by vertical and horizontal lines.

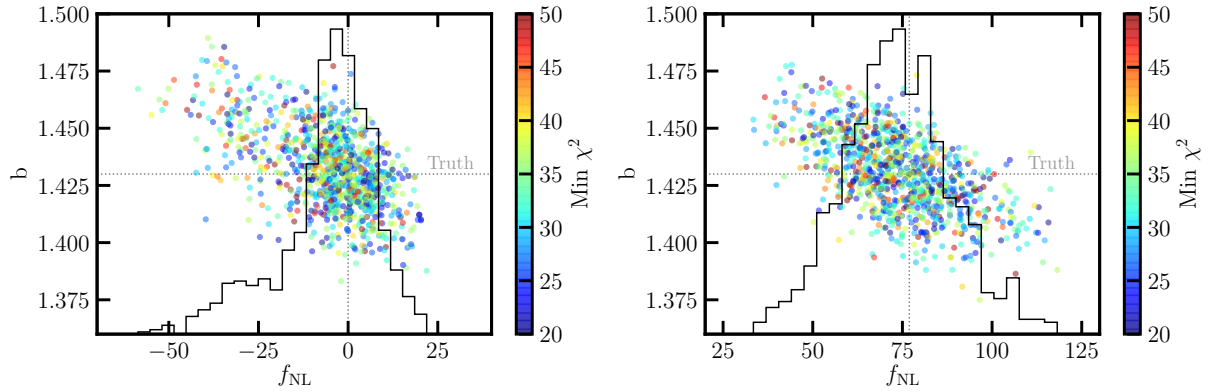


**Figure B2.** 68% and 95% confidence contours of fitting the mean power spectrum or its log transformation from the  $f_{\text{NL}} = 76.92$  mocks for the DESI footprint. Using the  $\log C_\ell$  fitting yield constraints that are insensitive to the covariance used. The truth values are represented by vertical and horizontal lines.

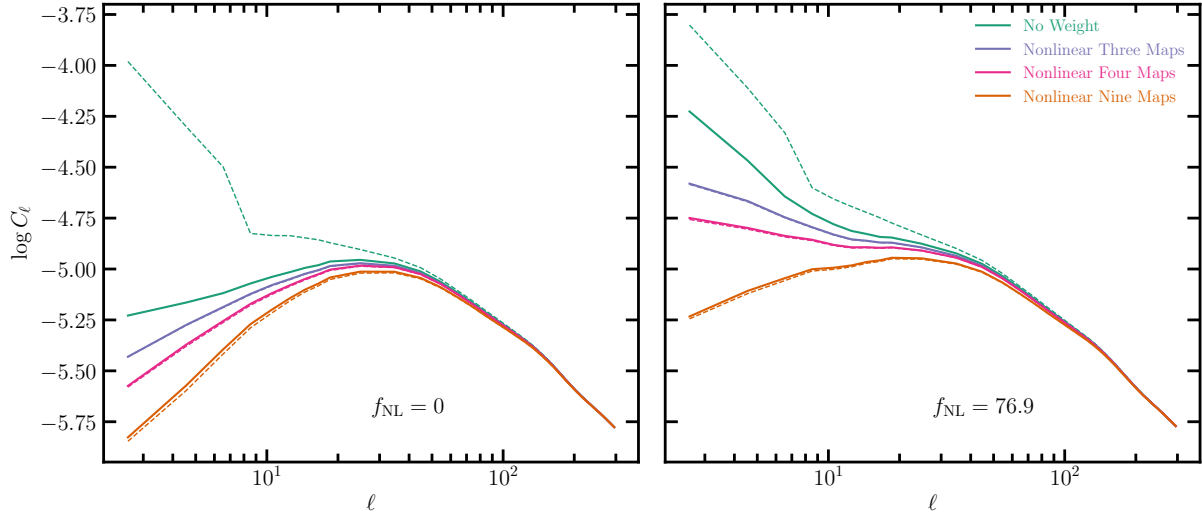
Fig B3 shows the best fit estimates for  $b$  vs  $f_{\text{NL}}$  for  $f_{\text{NL}} = 0$  and  $= 76.92$  mocks in the left and right, respectively. Truth values are represented via the dotted lines. The points are color-coded with the minimum  $\chi^2$  from fit for each realization. The histograms of best fit  $f_{\text{NL}}$  estimates are plotted in the background.

**Table B1.** Best fit and marginalized mean estimates for  $f_{\text{NL}}$  from fitting the mean power spectrum of the mocks. Degree of freedom is 34 (37 data points - 3 parameters).

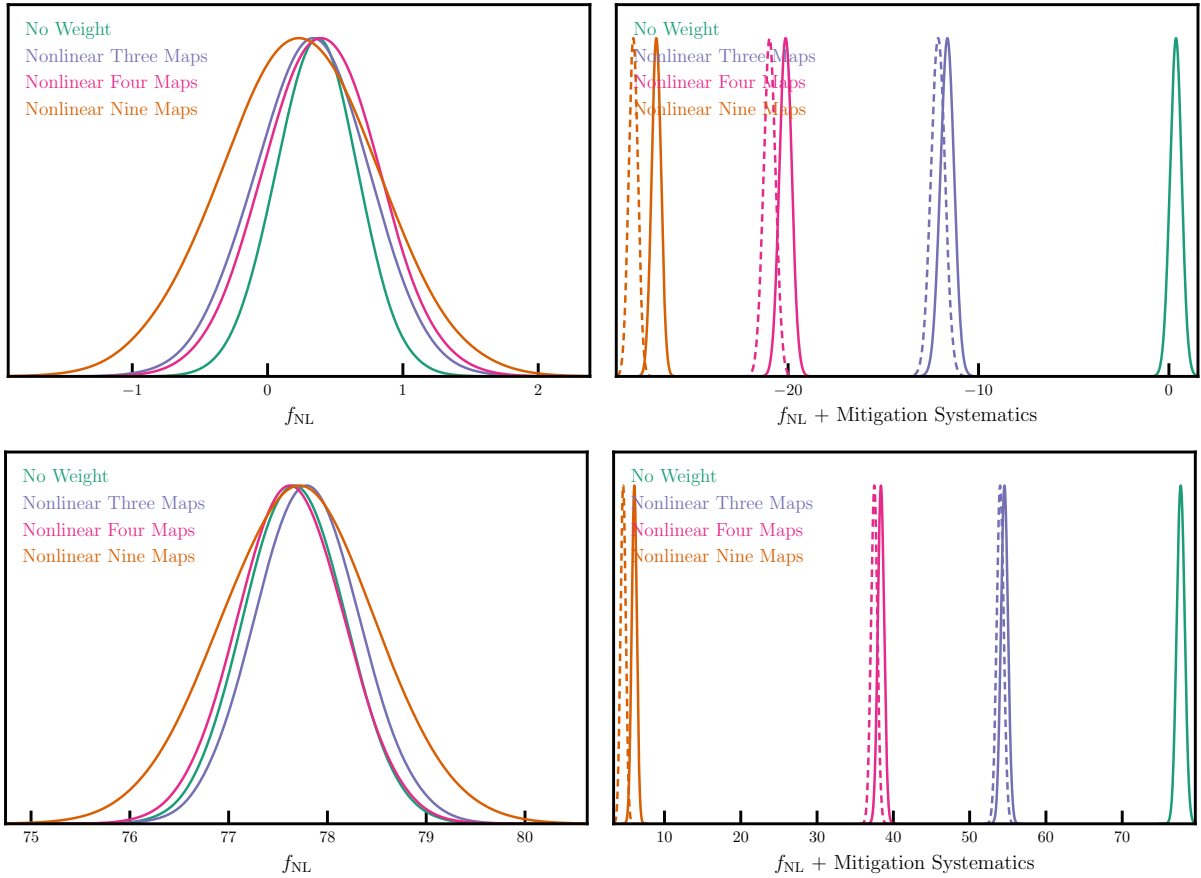
Mock / $f_{\text{NL}}$	Footprint	Observable	$f_{\text{NL}}$				$\chi^2$
			Best fit	Mean	68% CL	95% CL	
Clean 76.92	DESI	$\log C_\ell$	77.67	77.67	$77.17 < f_{\text{NL}} < 78.16$	$76.71 < f_{\text{NL}} < 78.64$	38.8
Clean 76.92	DESI	$C_\ell$	77.67	77.65	$77.17 < f_{\text{NL}} < 78.14$	$76.70 < f_{\text{NL}} < 78.60$	39.0
Clean 76.92	DESI	$\log C_\ell + f_{\text{NL}} = 0$ cov	77.70	77.71	$77.25 < f_{\text{NL}} < 78.17$	$76.81 < f_{\text{NL}} < 78.63$	39.9
Clean 76.92	DESI	$C_\ell + f_{\text{NL}} = 0$ cov	77.03	77.02	$76.93 < f_{\text{NL}} < 77.12$	$76.83 < f_{\text{NL}} < 77.22$	207.6
Clean 0	DESI	$\log C_\ell$	0.36	0.36	$0.06 < f_{\text{NL}} < 0.65$	$-0.23 < f_{\text{NL}} < 0.94$	35.7
Clean 0	BASS+MzLS	$\log C_\ell$	0.83	0.82	$0.25 < f_{\text{NL}} < 1.40$	$-0.31 < f_{\text{NL}} < 1.96$	39.4
Clean 0	DECaLS North	$\log C_\ell$	0.07	0.06	$-0.47 < f_{\text{NL}} < 0.60$	$-1.00 < f_{\text{NL}} < 1.12$	26.7
Clean 0	DECaLS South	$\log C_\ell$	0.67	0.67	$0.13 < f_{\text{NL}} < 1.22$	$-0.40 < f_{\text{NL}} < 1.75$	34.3

**Figure B3.** Best fit estimates from fitting 1000 lognormal mocks with  $f_{\text{NL}} = 0$  (left) and 76.92 (right) in the DESI footprint. The truth values are represented by vertical and horizontal lines. **up and down figures?****Table B2.** Best fit and marginalized estimates for  $f_{\text{NL}}$  from fitting the mean power spectrum of the mocks before and after corrections using the nonlinear approach with various combinations of the imaging systematic maps. The estimates are not accounted for over correction, and therefore are subject to mitigation systematics.

Mock / $f_{\text{NL}}$	Method	$f_{\text{NL}} + \text{Mitigation Systematics}$				$\chi^2$
		Best fit	Mean	68% CL	95% CL	
Clean 0	No Weight	0.36	0.36	$0.06 < f_{\text{NL}} < 0.65$	$-0.23 < f_{\text{NL}} < 0.94$	35.7
Clean 0	Three Maps	-11.64	-11.65	$-12.00 < f_{\text{NL}} < -11.30$	$-12.34 < f_{\text{NL}} < -10.97$	86.8
Clean 0	Four Maps	-20.14	-20.13	$-20.44 < f_{\text{NL}} < -19.82$	$-20.74 < f_{\text{NL}} < -19.52$	472.8
Clean 0	Nine Maps	-26.91	-26.92	$-27.16 < f_{\text{NL}} < -26.68$	$-27.39 < f_{\text{NL}} < -26.46$	5481.0
Contaminated 0	Three Maps	-12.12	-12.13	$-12.48 < f_{\text{NL}} < -11.78$	$-12.83 < f_{\text{NL}} < -11.44$	94.0
Contaminated 0	Four Maps	-20.97	-20.98	$-21.28 < f_{\text{NL}} < -20.67$	$-21.58 < f_{\text{NL}} < -20.37$	556.3
Contaminated 0	Nine Maps	-28.13	-28.13	$-28.36 < f_{\text{NL}} < -27.90$	$-28.59 < f_{\text{NL}} < -27.67$	6760.5
Clean 76.92	No Weight	77.67	77.67	$77.17 < f_{\text{NL}} < 78.16$	$76.71 < f_{\text{NL}} < 78.64$	38.8
Clean 76.92	Three Maps	54.57	54.57	$54.14 < f_{\text{NL}} < 55.01$	$53.72 < f_{\text{NL}} < 55.45$	603.5
Clean 76.92	Four Maps	38.38	38.38	$37.99 < f_{\text{NL}} < 38.78$	$37.60 < f_{\text{NL}} < 39.16$	537.0
Clean 76.92	Nine Maps	6.04	6.04	$5.72 < f_{\text{NL}} < 6.36$	$5.41 < f_{\text{NL}} < 6.67$	694.0
Contaminated 76.92	Three Maps	54.01	54.00	$53.57 < f_{\text{NL}} < 54.44$	$53.15 < f_{\text{NL}} < 54.86$	588.0
Contaminated 76.92	Four Maps	37.48	37.49	$37.09 < f_{\text{NL}} < 37.88$	$36.70 < f_{\text{NL}} < 38.27$	510.7
Contaminated 76.92	Nine Maps	4.59	4.58	$4.26 < f_{\text{NL}} < 4.90$	$3.95 < f_{\text{NL}} < 5.22$	649.7



**Figure B4.** The mean power spectrum of the  $f_{\text{NL}} = 0$  and 76.9 mocks with (dashed) and without (solid) imaging systematics before ('No Weight') and after applying the nonlinear cleaning method with three, four, and nine maps.



**Figure B5.** Probability distributions of  $f_{\text{NL}}$  from the mean power spectrum of the  $f_{\text{NL}} = 0$  (top) and  $f_{\text{NL}} = 76.9$  (bottom) mocks before and after mitigation with the nonlinear methods using three, four, and nine maps. Left: The posteriors are adjusted to account for the over correction effect. Right: The posteriors are subject to the over correction effect, and thus the scaling of  $f_{\text{NL}}$  values is biased due to mitigation.

1256 This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.