MNRAS 000, 1–19 (2022) Preprint 16 February 2023 Compiled using MNRAS LaTeX style file v3.0

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

Mehdi Rezaie[1], Ashley J. Ross[2], Hee-Jong Seo[3], Hui Kong[3], Edmond Chaussidon[4], Anna Porredon[3], Lado Samushia[1], Rongpu Zhou[5], Alex Krolewski[6,7,8], Arnaud de Mattia[4], Jose Bermejo[?], Florian Beutler[9], Christophe Yeche[4], Nathalie Palanque-Delabrouille[4,5], Klaus Honscheid[2,10], Santi Avila, Eva Mueller, Will Percival, Shadab Alam, Benedict Bahr-Kalus, Chia-Hsun Chuang, Noah Weaverdyck, Simone Ferraro, Tanveer Karim, and Adam Myers

[1]*Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*
[2]*Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA*
[3]*Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[4]*IRFU, CEA, Universite Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[5]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[6]*Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[7]*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*
[8]*Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[9]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[10]*Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

This paper uses the large-scale clustering of luminous red galaxies selected from the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys Data Release 9 to constrain the local primordial non-Gaussianity (PNG) parameter $f_{\rm NL}$. We thoroughly investigate the impact of various photometric systematic effects, such as those caused by Galactic extinction, local stellar density, varying survey depth, and astronomical seeing using spherical harmonics cross power spectrum and mean galaxy density statistics. Lognormal density fields are simulated with and without PNG to construct covariance matrices, evaluate the robustness of power spectrum modeling code, assess whether spurious fluctuations are properly mitigated, and calibrate imaging systematics cleaning methods. With harmonic modes from $\ell = 2$ to 300, we find $36.07(25.03) < f_{\rm NL} < 61.44(75.64)$ with a conservative cleaning approach and $13.09(-15.95) < f_{\rm NL} < 69.14(91.84)$ with an extreme treatment of imaging systematics, both at 68%(95%) confidence. We find significant remaining systematic error raised by calibration issues in the South Galactic Cap and local stellar density in the North Galactic Cap, which induce noticeable biases in $f_{\rm NL}$ constraints. While our constraints are consistent with zero PNG at 95% confidence for the extreme approach, we show that the characterization of stellar contamination and calibration issues are crucial to derive unbiased constraints on $f_{\rm NL}$ in the era of DESI and Rubin Observatory.

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

1 INTRODUCTION

Characteristics of the temperate anisotropy and polarization maps from the cosmic microwave background (CMB) and that of the galaxy density fluctuations probed by large-scale structure (LSS) surveys are explained to remarkable extents by a cosmological model for the universe that consists of dark energy, dark matter, and ordinary luminous matter, and has experienced a period of rapid expansion, known as *inflation*, in its early stages (see, e.g., Weinberg et al. 2013, for a review). The paradigm of inflation elegantly addresses fundamental issues, such as the isotropy, flatness, and homogeneity of the universe as well as the absence of magnetic monopoles (see, e.g., Weinberg 2008). At the end of inflation, the universe went through a reheating process, and primordial fluctuations were generated to seed the subsequent growth of structure (Kofman et al. 1994; Bassett et al. 2006; Lyth & Liddle 2009). Even though current observations imply that inflation has certainly happened, characteristics of the field or the fields driving the inflationary expansion remain vastly unknown, and statistical properties of primordial fluctuations pose an intriguing problem in modern observational cosmology (see, e.g., Biagetti 2019, for a review). Early studies of cosmological observables have suggested that initial conditions of the universe are consistent with Gaussian fluctuations (Tegmark et al. 2004; Guth & Kaiser 2005); however, alternative classes of inflationary models predict some levels of non-Gaussianities in the primordial gravitational field. In its simplest form, primordial non-Gaussianity (PNG) depends on the local value of the gravitational potential ϕ, and it is parameterized by a nonlinear coupling constant $f_{\rm NL}$ (Komatsu & Spergel 2001),

$$\Phi = \phi + f_{\rm NL}[\phi^2 - <\phi^2>]. \tag{1}$$

Standard slow-roll inflation predicts $f_{\rm NL}$ to be of order 10^{-2} (see, e.g., Alvarez et al. 2014, for a review), while multi-field inflationary scenarios anticipate considerably higher than unity $f_{\rm NL}$ values (see, e.g., de Putter et al. 2017). Therefore, getting robust constraints on $f_{\rm NL}$ is the first stepping stone toward better understanding the dynamics of the early universe. PNG alters the local number density of galaxies by coupling the long and small wavelength modes of the dark matter gravitational field, and consequently, it induces a scale-dependent feature in the two point clustering of biased tracers of the dark matter gravitational field (Dalal et al. 2008).

The current tightest bound on $f_{\rm NL}$ comes from the three-point clustering measurement of the CMB temperature anisotropies by the Planck satellite, $f_{\rm NL} = 0.9 \pm 5.0$ (Akrami et al. 2019). Upcoming generations of CMB experiments will improve this constraint, but since CMB is limited by cosmic variance, its data alone cannot enhance statistical precision of $f_{\rm NL}$ measurements enough to break the degeneracy amongst various inflationary paradigms (see, e.g., Ade et al. 2019). Combining CMB with LSS data could cancel cosmic variance, partially if not completely, and improve these results to a precision level needed to differentiate between alternative inflationary scenarios (see, e.g., Schmittfull & Seljak 2018). Constraining $f_{\rm NL}$ with the three-point clustering of LSS is likewise hindered by the late-time nonlinear systematic effects raised from structure growth, which is non-trivial to account for (Baldauf et al. 2011b,a). UV Luminosity Function is a novel approach for constraining $f_{\rm NL}$ by probing galaxy abundances and structure formation on small scales (e.g., $k \sim 2$ 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 catalogs (Bouwens et al. 2015) to find a 2σ bound of $-166 < f_{\rm NL} < 497$. Even though this 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 $f_{\rm NL}$ constraints from UV Luminosity Function. Given these limitations, the scale-dependent bias technique is the smoking gun for constraining local PNG with LSS. With this technique, Mueller et al. (2022) analyzed the large-scale clustering of quasars and obtained XXX at 65% confidence. cite recent measurements after eBOSS.

Measuring $f_{\rm NL}$ with the scale-dependent bias effect is nonetheless incredibly challenging due to various systematic effects that modulate clustering power on scales where there is a high sensitivity to $f_{\rm NL}$. These systematics are broadly classified into theoretical and observational. For instance, survey geometry entangles clustering power on different angular modes (Beutler et al. 2014; Wilson et al. 2017). Relativistic effects also generate scale-dependent signatures on large scales, identical to local PNG, which hinder measuring $f_{\rm NL}$ with the scale-dependent bias effect using higher order multipoles of power spectrum (Wang et al. 2020). Similarly, matter density fluctuations with wavelengths larger than survey volume, known as super-sample modes, modulate galaxy power spectrum (Castorina & Moradinezhad Dizgah 2020). Another source for systematic error is raised because the mean galaxy density for constructing the density contrast field is estimated from data directly rather than being known a priori. This integral constraint effect pushes clustering power on modes near the survey size to zero (Peacock & Nicholson 1991; De Mattia & Ruhlmann-Kleider 2019). Accounting for these effects in modeling power spectrum is crucial to derive unbiased $f_{\rm NL}$ constraints (see, e.g., Riquelme et al. 2022).

On the other hand, observational systematics are driven primarily by varying imaging properties across the sky (Ross et al. 2011) and photometric calibration issues that manifest as spurious fluctuations in the observed density field of galaxies (Huterer et al. 2013). This type of systematic error is much more complicated to model and mitigate, compared to integral constraint and survey geometry, and it has hampered previous studies of local PNG with the scale-dependent bias effect in the large-scale clustering of galaxies and quasars (see, e.g., Ho et al. 2015). For instance, Pullen & Hirata (2013) found that the contribution of stellar density and astronomical seeing is too high for a robust $f_{\rm NL}$ measurement using the quasar sample from the Sloan Digital Sky Survey Data Release 6. These imaging systematic issues are expected to be severe for wide-area galaxy surveys that observe the night sky closer to the Galactic plane and attempt to implement more relaxed selection criteria to include fainter galaxies (see, e.g, Kitanidis et al. 2020).

SEEMS OUT OF PLACE The Dark Energy Spectroscopic Instrument (DESI) uses robotically-driven fibers to collect 5000 spectra simultaneously and is designed to deliver an unparalleled volume of spectroscopic data that will enable future analyses to deepen our understanding of the energy contents of the universe, specifically, the dynamic of dark energy equation of state (Aghamousa et al. 2016). Even though the primary focus of DESI is dark energy, assuming imaging systematics are under control, DESI along with other upcoming surveys, such as Rubin Observatory and SphereX, are expected to yield unprecedented constraints on $f_{\rm NL}$ as well (see, e.g., Heinrich & Doré 2022). Luminous red galaxies (LRGs) constitute one of the primary targets for DESI spectroscopy. DESI is designed to collect spectra of millions of LRGs covering the redshift range of $0.4 < z < 1.0$ throughout its five-year mission. LRGs are massive galaxies that occupy massive halos, lack active star formation, and are one of the highly biased tracers of dark matter gravitational field. A distinct break around 4000 Å in the spectrum of LRGs 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 spectroscopy selects its LRG targets from photometry of three ground-based surveys that observed the sky in the optical g, r, and z bands between 2014 and 2019: the 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). The BASS and MzLS surveys observed the same footprint in the North Galactic Cap (NGC) while the DECaLS observed both caps around the galactic plane; the BASS+MzLS footprint is separated from the DECaLS NGC at 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 same instrument under the Dark Energy Survey (Abbott et al. 2016), which constitute about 1130 deg^2 of the South Galactic Cap (SGC) footprint. The DESI imaging catalogs also integrate the 3.4 (W1) and 4.6 μm (W2) infrared photometry from the Wide-Field Infrared Explorer (WISE; Wright et al. 2010; Meisner et al. 2018).

The characterization of potential sources for systematic error in the DESI imaging LRG sample is of paramount importance to DESI success since spectroscopic catalogs could inherit these issues from imaging catalogs, and thus negatively impact DESI science goals. The effects of observational systematics in DESI targets have been studied in great detail (see, e.g., Kitanidis et al. 2020; Zhou et al. 2021; Chaussidon et al. 2022). Improving techniques to characterize systematic error in these tracers is crucial for the science beyond dark energy, such as constraining $f_{\rm NL}$ and other features in the primordial power spectrum (Beutler et al. 2019). Some of the current methods seek to mitigate systematic effects by either cross-correlating target density and imaging maps (mode deprojection) or solving a least-square optimization to estimate the contribution from each imaging property to target density (template-based regression), ultimately to regress out the modes affected by imaging properties from target density. Another class of methods aims to cross-correlate different tracers of dark matter to enhance inferences by canceling cosmic variance and by reducing the effect of systematic error, 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). For instance, mode deprojection yields an unbiased clustering but can be employed for angular clustering only, and its involved matrix algebra could prove time-consuming for large survey sizes. Template-based regression is on the other hand computationally economic, but it returns biased clustering by removing some of clustering power, depending on the number of input templates and the flexibility of the regression model. Specifically related to the template-based regression method, there is little effort to calibrate and characterize the amount of clustering power removed during the cleaning process. For studies like BAO and RSD, these effects are demonstrated to be negligible (Merz et al. 2021); however, these effects introduce significant biases in $f_{\rm NL}$ constraints (Mueller et al. 2022) as they highly impact galaxy clustering on large scales (Rezaie et al. 2021).

This paper presents robust constraints on $f_{\rm NL}$ using the large-scale clustering of photometric galaxies from DESI imaging with exquisite emphasis on the treatment of imaging systematic error and mitigation biases. We measure the significance of residual systematic error in our data using angular cross-power spectrum (between galaxy density and imaging properties) and mean density contrast of galaxies. Specifically, the robustness of our results is validated against various sources of systematic error, including but not limited to photometric calibration and Milky Way foregrounds. We cross-correlate the density map of galaxies with the template maps of imaging properties to evaluate the effectiveness of different treatment methods and to characterize the significance level of remaining systematic error. Various linear and nonlinear data cleaning approaches are applied with different combinations of imaging templates to quantify the sensitivity of $f_{\rm NL}$ constraints to alternative foreground removal methods. The redshift distribution of galaxies for modeling power spectrum is determined by early spectroscopy from the DESI Survey Validation. This paper is structured as follows. Section 2 describes the sample of LRGs from DESI imaging

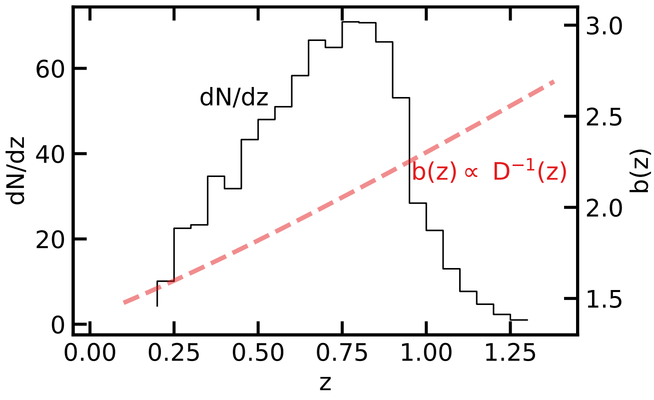

Figure 1. The redshift distribution (solid) and bias evolution (dashed) of the DR9 LRG sample. The redshift distribution is determined by DESI spectroscopy, and the model for bias assumes a constant clustering amplitude (see, e.g., Zhou et al. 2021, 2022).

and lognormal simulations of galaxy density field with and without PNG and imaging systematic effects. Section 3 outlines the theoretical framework for modeling angular power spectrum and analysis techniques to account for various observational and theoretical systematic error. Finally, we present $f_{\rm NL}$ constraints in Section 4, and conclude with a comparison to previous $f_{\rm NL}$ studies in Section 5.

2 DATA

2.1 DESI imaging LRG sample

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 (CITE), described as the 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 Tab. 1. The selection cuts are developed differently for each imaging survey to reach an almost uniform target surface density 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). Fig. 1 shows the redshift distribution of the DR9 LRGs (solid black), inferred from the spectroscopic DESI Survey Validation data (CITE), and the evolution of halo bias (red dashed), adopted from Zhou et al. (2021).

DESI-like LRGs are selected brighter than the imaging survey depth limits; therefore, the DR9 LRG density field is nearly homogenous, unlike the other DESI tracers. To further reduce stellar contamination, the LRG sample is masked rigorously for foreground bright stars, galaxies, and clusters of galaxies[1]. Then, the sample is binned into HEALPix (Gorski et al. 2005) pixels at NSIDE = 256 to construct the 2D density map with an average surface density of 800 deg^{-2} with sky coverage around 14000 square degrees. We correct for the pixel incompleteness and lost areas in the LRG density field using a catalog of random points, hereafter referred to as randoms, uniformly scattered over the footprint with the same cuts and masks

[1] See the maskbits at https://www.legacysurvey.org/dr9/bitmasks/

Table 1. Selection criteria for the DESI-like LRG targets (Zhou et al. 2022). Magnitudes are corrected for MW extinction. $z_{\rm fiber}$ represents the z-band fiber magnitude which corresponds to the expected flux within a DESI fiber.

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

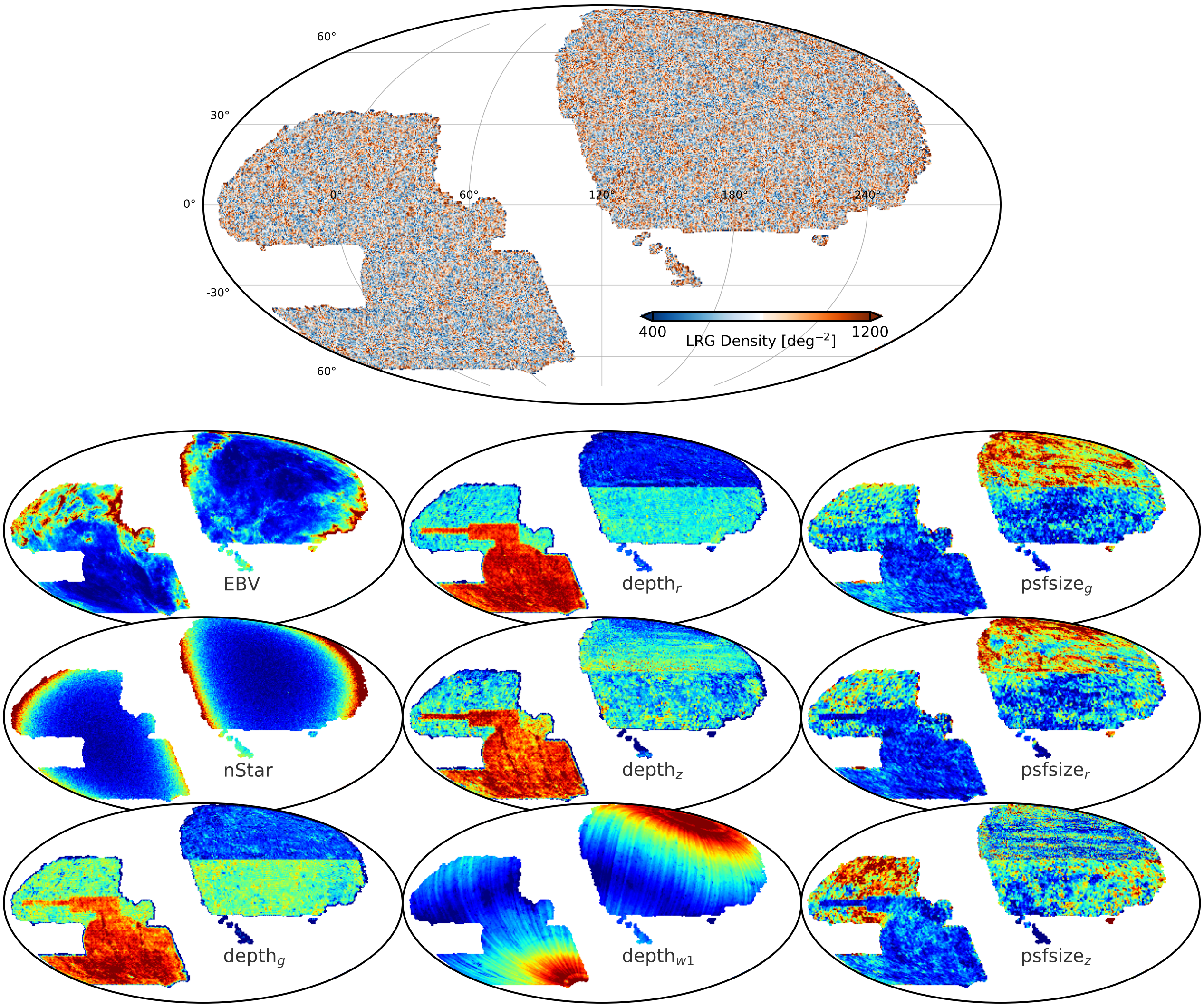

Figure 2. Top: The DESI imaging LRG density field in Mollweide projection. Spurious disconnected islands from the North footprint and declination below −30 from the South footprint are removed for the analysis due to potential calibration issues (see text). Bottom: Mollweide projections of the DR9 catalog imaging properties (survey depth and astronomical seeing/psfsize) and MW foregrounds (extinction and local stellar density) in celestial coordinates.

applied. Fig. 2 (top) shows the observed density field of the DR9 LRGs in deg^{-2} before applying any correction weights to account for imaging systematic effects. The DR9 LRG density exhibits large-scale spurious fluctuations, which are unlikely to be of cosmological origin. Specifically, the SGC footprint exhibits some systematic under-density while there is some systematic over-density near the survey boundaries in the NGC.

2.1.1 *Correlation coefficients*

First, we study the linear correlation between the LRG density map and various imaging properties which are considered as potential sources of systematic error. The imaging properties are mapped into HEALPix pixels at the same NSIDE. Following Zhou et al. (2022), the imaging properties investigated in this work are 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); and survey-related imaging properties include survey depth (galaxy depth in the g, r, and z bands and PSF depth in W1) and astronomical seeing (psfsize) in the g, r, and z bands. Templates for the survey-related imaging properties are produced by binning the randoms into HEALPix, and averaging over the imaging attributes of the randoms in each pixel.

Fig. 2 (bottom) illustrates the imaging templates investigated as potential sources of systematic error. Each map shows its own characteristic large-scale spurious fluctuations. For instance, the under-dense part of the DR9 LRG sample in the SGC can be associated with survey depth, while the over-density in the NGC can be linked to the extinction map. We reject some parts of the DR9 sample to minimize the potential for photometric calibration systematics. There are some disconnected islands, hereafter referred to as *spurious islands*, in the DECaLS North region at Dec < -11. Additionally, some parts of the DECaLS South footprint with Dec < -30 are removed from the sample, because a different catalog of standard stars is employed to calibrate images below that region. We discuss how these quality cuts influence $f_{\rm NL}$ constraints from the DR9 LRG sample in Section 4.

Fig. 3 shows the Pearson correlation coefficient between the DR9 LRG density and DESI imaging properties for the three imaging surveys (DECaLS North, DECaLS South, and BASS+MzLS) in the top panel. The horizontal curves are constructed from lognormal simulations (see, subsection 2.2) to quantify the significance of correlations. Fig. 3 (bottom) shows the correlation matrix among imaging properties for the DESI footprint. There is a strong correlation between the LRG density and depth maps, and next correlated properties seem to be Galactic foregrounds. There is a small correlation between the LRG density and the $W1$-band depth and psfsize properties. We observe a significant inner correlation among the imaging properties themselves, especially between the local stellar density and Milky Way extinction; also, the r-band and g-band survey properties are more correlated with each other than with the z-band. We also investigate the correlations using the Spearman-r correlation, but find no significant differences.

2.1.2 Imaging weights

We follow a regression approach to model the spatial dependence of spurious fluctuations in the DR9 density map to the information encoded by the imaging templates. Both linear multivariate and nonlinear regression are applied to assess the level of nonlinear systematic effects.

Regression analyses are performed separately for each imaging survey as our preliminary analysis indicated that each region responds differently to each template (see, e.g., Fig. 3 top panel). The optimal parameters associated with the linear and nonlinear models are found by optimizing the negative Poisson log-likelihood, $\lambda - \rho \log(\lambda)$, between the observed galaxy density ρ and the output predicted density λ from the models given imaging properties $\mathbf{x}$ as input. We do not provide spatial information as input to avoid subtracting clustering signal. With $\lambda(\mathbf{x}) = \log(1+e^{f(\mathbf{x})})$, we investigate the use of linear multivariate and nonlinear models to approximate f. For the linear model, we perform a Monte Carlo Markov Chain (MCMC) search using the EMCEE package (Foreman-Mackey et al. 2013) and all of the DR9 data to calculate the Poisson log-likelihood.

For the nonlinear approximation, we employ the implementation of artificial neural networks from Rezaie et al. (2021); specifically, the nonlinear model is composed of an ensemble of 20 neural network models. Each neural network consists of imaging maps on the first layer, three hidden layers with 20 rectifier units on each layer, and single unit with the identity function in the output layer. The rectifier is the identity function for positive input and zero for negative, and it introduces nonlinearities in the neural network architecture. Unlike the linear regression, we use 60% of the LRG data for training, 20% for validation, and 20% for testing in order to minimize the chance of over-subtraction, i.e.g, regressing out some of clustering signal by the nonlinear model. However, we are able to test the nonlinear model on the entire LRG data with the technique of permuting the choice of the training, validation, or testing sets. We train the neural networks for up to 70 training epochs with the gradient descent ADAM optimizer (Loshchilov & Hutter 2017), i.e., the parameters are updated iteratively following the gradient of the negative Poisson log-likelihood. We initialize the learning rate by minimizing the loss on the validation set and adjust it to dynamically vary between two boundary values of 0.001 and 0.1

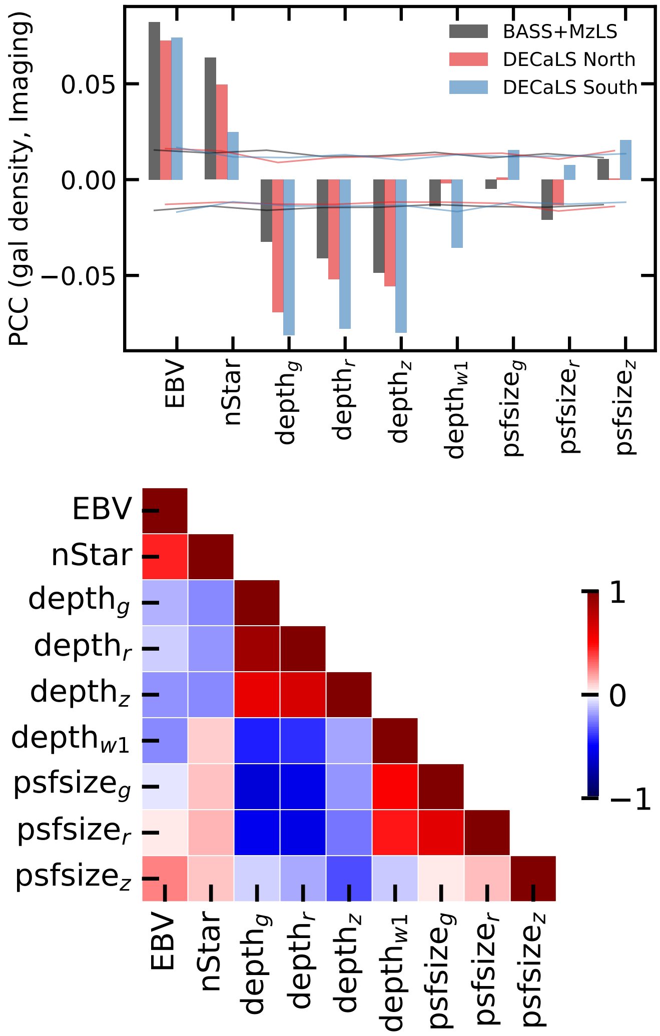

Figure 3. Top: Pearson correlation coefficients between the DR9 LRG density and imaging properties in the three imaging regions. Solid curves represent the 95% spread of correlation coefficients observed in 100 randomly selected lognormal mock density realizations. Bottom: Pearson correlation matrix from imaging properties for the full DESI footprint.

to avoid local minima during gradient descent. The best neural network model is then selected from the lowest prediction error when applied to the validation set. Finally, we run the ensemble of 20 best-fit models through the test set and average over the predictions to construct the predicted galaxy density map in HEALPix. The predicted galaxy density is normalized and applied as imaging weights to down-weight the observed density map of LRGs to reduce spurious fluctuations in the LRG data.

Because of the inner correlation amongst the maps (see, e.g., Fig. 3 bottom panel), a few subsets of imaging maps are considered as input for more conservative treatment of systematic errors. These subsets are selected to minimize the cross-correlations between the cleaned LRG density field and imaging properties while avoiding highly correlated templates:

- **Conservative I**: Extinction and z-band depth
- **Conservative II**: Extinction, z-band depth, and r-band psfsize
- **All Maps**: Extinction, depth in $grzW1$, and psfsize in grz

Upon inspecting the predicted density maps, we find that while most of the large-scale spurious fluctuations are explained by just the extinction map and depth in the z band, adding the r-band psfsize results in a finer structure in the predicted density map, and as we show later it reduces the cross-correlation between the LRG density and the z band psfsize. We observe that using all imaging maps as input features for regression does not add more information, as expected due to the strong correlations between different bands. Comparing the weights from the linear model to that of the nonlinear approach for the same input maps, we find that the nonlinear approach yields finer structures due to higher flexibility. Overall consistent with the LRG data, both models predict higher galaxy density near the boundaries where the DESI imaging surveys observed the high extinction regions of the Milky Way. These overdense regions are likely contaminated artifacts entering the LRG selection, e.g., stellar contaminants or other artifacts because of obscured photometry by MW extinction. The assessment of these imaging weights is further analyzed in 3.4. We also investigate the use of all imaging maps as a case which is highly prone to over-subtracting the true clustering signal. Additionally, we assess the robustness of our results against remaining systematic errors by adding external templates for the neutral hydrogen column density (HI4PI Collaboration et al. 2016) and the photometric calibration (e.g., in the z band; *CALIBZ*). The effects on $f_{\rm NL}$ constraints are discussed in Section 4.

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. We use FLASK (Full-sky Lognormal Astro-fields Simulation Kit; Xavier et al. 2016) to generate ensembles of synthetic galaxy density fields that mimic the redshift and angular distributions of the DR9 LRG sample (see, Fig. 1). We create 1000 realizations with $f_{\rm NL}=0$ and 76.92 using a redshift dependent bias $b(z)=1.43/D(z)$, consistent with Zhou et al. (2021). We adapt the fiducial cosmology from a flat ΛCDM universe, including one massive neutrino with $m_\nu=0.06$ eV, and the rest of cosmological parameters is deducted from Planck 2018 (Aghanim et al. 2020),

$$h=0.67, \Omega_M=0.31, \sigma_8=0.8, \text{and } n_s=0.97.$$

The same cosmology is employed throughout the rest of the manuscript. We demonstrate later in Section 4 that our $f_{\rm NL}$ constraints are quite robust against the choice of our fiducial cosmology.

2.2.1 *Contaminated mocks*

The linear multivariate model with the templates for the extinction, depth$_z$, psfsize$_r$ (conservative II) is used to induce observational imaging spurious fluctuations in the lognormal density fields. We choose the linear model for the contamination to assess how much of the clustering signal is removed by applying more flexible models (based on neural networks) for cleaning. The parameters of the linear model are fit on the DR9 LRG sample, separately on each imaging survey. The contamination model is drawn from the posteriors sampled by MCMC, and thus is distinct for each realization of lognormal density field. The same contamination model is applied to both the $f_{\rm NL}=0$ and 76.92 mocks.

Similar to the DR9 sample, the linear and nonlinear mitigation methods are applied to the simulations, with and without PNG or with and without the imaging-dependent systematics, to derive the imaging weights. Section 4 presents how we use these mock results to calibrate the amount of the clustering power that is removed by the cleaning methods. We use no prior knowledge regarding the underlying contamination model or the templates used.

3 ANALYSIS TECHNIQUES

3.1 Power spectrum estimator

We follow the procedure described in Hivon et al. (2002) to estimate the angular power spectrum. We begin with constructing the density contrast field δ_i from the observed density of galaxies ρ_i,

$$\hat{\delta}_i = \frac{\rho_i}{\hat{\overline{\rho}}} - 1, \tag{2}$$

where $\hat{\overline{\rho}}$ is the mean galaxy density estimated directly from the observed density field,

$$\hat{\overline{\rho}} = \frac{\sum_i \rho_i f_{\rm pix,i}}{\sum_i f_{\rm pix,i}}, \tag{3}$$

a with the pixel fractional completeness $f_{{\rm pix},i}$ represented by randoms. Next, the density contrast field δ_i is expanded into spherical harmonics $Y_{\ell m}$,

$$\hat{a}_{\ell m} = \int d\Omega\, \delta f_{\rm pix} Y^*_{\ell m}. \tag{4}$$

Then, we estimate the angular power spectrum via,

$$\hat{C}_\ell = \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell} |\hat{a}_{\ell m}|^2. \tag{5}$$

We also follow a similar approach to construct density contrast fields for imaging properties and cross correlate with the galaxy density map. 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. This power spectrum estimator is biased for a partial sky coverage, and thus often referred to as the pseudo power spectrum. Because of the survey mask, different harmonic modes are no longer independent and the measured power on scales near the survey size is pushed to zero. These effects impact the galaxy power spectrum on large scales, where the sensitivity to PNG is high. Therefore, correcting for these survey geometry-related effects is crucial for our analysis, and later we will describe how our model power spectrum accounts for these effects.

3.2 Modeling

3.2.1 *Power spectrum*

The projected angular power spectrum of galaxies is related to the 3D power spectrum $P(k)$ (see, e.g., Padmanabhan et al. 2007) and shotnoise $N_{\rm shot}$ by,

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

where the shotnoise is assumed to be scale-independent, and with redshift space distortions included, $\Delta_\ell(k) = \Delta_\ell^{\rm g}(k) + \Delta_\ell^{\rm RSD}(k)$ is the projection kernel that determines how much each wavenumber k contributes to mode ℓ by integrating over the $l^{\rm th}$ order spherical Bessel functions, $j_\ell(kr)$,

$$\Delta_\ell^{\rm g}(k) = \int \frac{dr}{r} r b(k,z) D(r) \frac{dN}{dr} j_\ell(kr), \tag{7}$$

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

The linear growth factor, $D(z)$, is scaled such that $D(0) = 1$, $f(r)$ is the growth rate, and dN/dr is the normalized redshift distribution of galaxies[2] (see, Fig. 1). The galaxy bias, $b(k,z)$, can be written as the redshift dependent linear bias (Fig. 1) and the scale-dependent shift due to PNG (Slosar et al. 2008),

$$b(k,z) = b(z) + 3(b(z)-p) f_{\rm NL} \frac{\delta_c \Omega_m H_0^2}{k^2 T(k) D(z) c^2} \frac{g(\infty)}{g(0)}, \tag{9}$$

where Ω_m is the matter density, H_0 is the Hubble constant[3], $T(k)$ is the transfer function, $\delta_c = 1.686$ represents the critical density for spherical collapse (Fillmore & Goldreich 1984), and $g(\infty)/g(0) \sim 1.3$ with $g(z) \equiv (1+z)D(z)$ (see, e.g., Mueller et al 2018). The parameter p is the correction due to galaxy selection beyond a Poisson sampling of the haloes of a given mass; if only mass determines how galaxies populate a halo, $p = 1$, which is often referred to as the universality of the halo occupation distribution. However, numerical simulations indicate that the halo occupation distribution for other tracers, e.g., quasars, which are from recent mergers, could depend on other properties besides mass, and thus p might take different values (see, e.g., Slosar et al. 2008). The theoretical uncertainty on p is not very well understood, and Barreira (2022) showed that marginalizing over this parameter even with wide priors leads to biased constraints because of parameter space projection effects. Using N-body simulations, Lazeyras et al. (2023) investigated secondary halo properties such as concentration, spin and sphericity of haloes, and found that halo spin and sphericity preserve the universality of halo occupation function while halo concentration significantly alters the halo function. Without better priors on p, it is argued that the scale-dependent bias effect can only be used to constrain the $p f_{\rm NL}$ term (see, e.g., Barreira et al. 2020; Barreira 2020). However, the significance of detection of nonzero PNG is not affected by different assumptions on p, i.e., a nonzero detection of $p f_{\rm NL}$ will still imply a nonzero detection of $f_{\rm NL}$. This paper is focused on how a thorough treatment of imaging systematic effects or lack thereof can impact the PNG constraints. Therefore, we choose $p = 1$ for our sample of LRGs. We use the FFTLog[4] algorithm and its extension as implemented in Fang et al. (2020) to handle the Bessel function integrals.

[2] $dN/dr = (dN/dz) * (dz/dr) \propto (dN/dz) * H(z)$

[3] Because k is in unit of h/Mpc, $H_0 = 100$ (km s^{-1})/(h^{-1}Mpc)

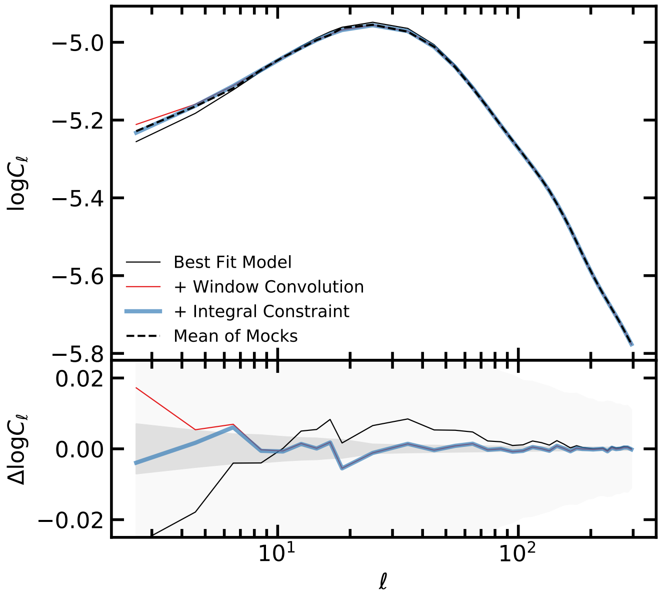

Figure 4. Mean power spectrum of the lognormal density fields with $f_{\rm NL} = 0$ and best fit theoretical prediction after accounting for the survey geometry and integral constraint effects. Dark and light shades represent 1σ error on the mean and one realization, respectively. Bottom panel shows the residual power spectrum relative to the mean power spectrum of the mocks. No imaging systematic effects are added to these mocks.

3.2.2 *Survey geometry and integral constraint*

The ensemble average for the partial sky power spectrum is related to that of the full sky power spectrum via the mode-mode coupling matrix $M_{\ell\ell'}$,

$$<\hat{C}_\ell> = \sum_{\ell'} M_{\ell\ell'} < C_{\ell'} > . \tag{10}$$

In general, the mode-mode coupling matrix is singular for a large sky cut, but a common approach is to use a discrete set of ℓ bins and assume the angular power spectrum is constant in each bin. This is not a bad assumption, but it might not be ideal for large scale modes and small bin widths. Therefore, we following a similar approach to Chon et al. (2004), and convolve the theoretical power spectrum C_ℓ with the survey mask window power to obtain the theoretical pseudo-power spectrum, rather than trying to de-convolve the observed power spectrum. The convolution in the spherical harmonics space becomes a multiplication in the correlation function space. To this end, we first estimate the paircount of the survey mask. The window paircount is normalized such that it is equal to one at the zero degree separation. Next, we multiply the model correlation function by the survey mask paircount. Then, the window-convolved power

[4] github.com/xfangcosmo/FFTLog-and-beyond

spectrum is obtained from,

$$\hat{C}_{\ell}^{\rm model} = 2\pi \int \hat{\omega}^{\rm model}\hat{\omega}^{\rm mask}\, P_{\ell}(\cos\theta) d\theta. \tag{11}$$

Finally, we account for the effect of integral constraint by subtracting the power spectrum of the survey mask,

$$\hat{C}_{\ell}^{\rm model, IC} = \hat{C}_{\ell}^{\rm model} - \hat{C}_{\ell=0}^{\rm model}(\frac{\hat{C}_{\ell}^{\rm window}}{\hat{C}_{\ell=0}^{\rm window}}). \tag{12}$$

We validate the pipeline for modeling the angular power spectrum against the lognormal simulations. Fig. 4 shows the log-transformed mean of 1000 spectra from the simulations (dashed) and the best fit theory prediction before and after accounting for survey geometry and integral constraint. light and dark shades represent the 68% error on the mean and one single realization, respectively. DESI footprint mask is applied to the mocks, and even though DESI covers around 40% of the sky, but the window effect is affecting modes down to $\ell = 200$. On the other hand, integral constraint only alters the power in the first two bins.

3.3 Parameter estimation

The signature of local PNG in the two-point function is unique and cannot be reproduced with other standard cosmological parameters. For parameter inference, we use standard Monte-Carlo Markov Chain sampling while allowing $f_{\rm NL}$, shotnoise, and bias to vary. For the bias, we assume a constant clustering amplitude for our sample of LRGs, $b(z) = b/D(z)$, and fit for b. Throughout this manuscript, we use a discrete set of bandpower bins with $\Delta\ell = 2$ between $\ell = 2$ and 20 and $\Delta\ell = 10$ from $\ell = 20$ to 300, while weighting each mode by $2\ell + 1$. We also find that the distribution of power spectrum at the lowest bin, $2 \leq \ell < 12$, is not Gaussian and its standard deviation varies significantly from mocks with $f_{\rm NL} = 0$ to those with 76.9 (see, Fig. 5). Therefore, we attempt to fit $\log C_{\ell}$ to make our constraints less sensitive to the choice of covariance matrix. The parameter $f_{\rm NL}$ is constrained by maximizing a posterior defined as,

$$-2\ln\mathcal{L} = (\log C(\Theta) - \log\hat{C})^{\dagger}\mathbb{C}^{-1}(\log C(\Theta) - \log\hat{C}) + \chi^2_{\rm priors}, \tag{13}$$

where Θ represents the parameters, $f_{\rm NL}$, bias coefficient, and shotnoise, all of which are associated with a flat prior, $\chi^2_{\rm priors}$; $C(\Theta)$ is the (binned) theoretical power spectrum including the effects for survey geometry and integral constraint; $\hat{C}$ is the (binned) measured power spectrum; and $\mathbb{C}$ is the covariance matrix constructed from simulations and corrected for the effect with Hartlap CITE.

3.4 Remaining systematic errors

In the absence of systematic effects, a) the mean density of galaxies should be uniform within the statistical fluctuations regardless of the imaging condition and b) the cross power spectrum between the galaxy density and imaging properties should be consistent with zero with statistical fluctuations. With these priors, we apply two statistical tests to look for remaining systematic effects in our sample of LRGs.

3.4.1 Cross power spectrum

We calculate the cross power spectra between galaxy density contrast field and imaging maps,

$$\hat{C}_{X,\ell} = [\hat{C}_{x_1,\ell}, \hat{C}_{x_2,\ell}, \hat{C}_{x_3,\ell}, ..., \hat{C}_{x_9,\ell}], \tag{14}$$

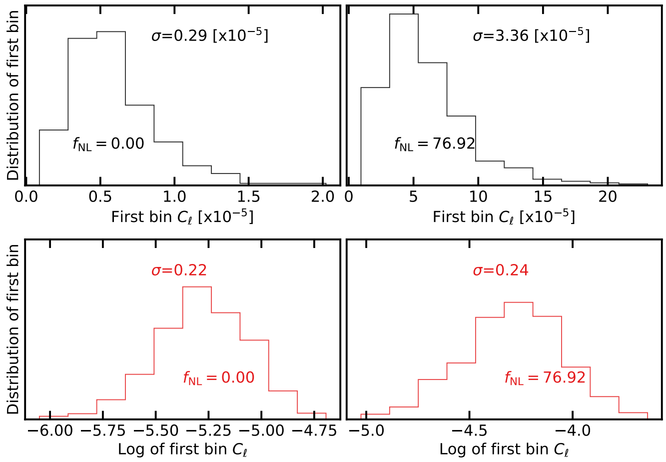

Figure 5. Distribution of the first bin power spectrum and its log transformation from the mocks with $f_{\rm NL} = 0$ (left) and 76.92 (right). Differences in the standard deviations become less significant, and power spectrum measurements follow a more symmetric distribution after the log transformation.

where $C_{x_i,\ell}$ represents the cross power spectrum between the galaxy density and $i^{\rm th}$ imaging map, normalized by the auto power spectrum of x_i:

$$\hat{C}_{x_i,\ell} = \frac{(\hat{C}_{gx_i,\ell})^2}{\hat{C}_{x_ix_i,\ell}}. \tag{15}$$

Then, the χ^2 value for the cross power spectra is calculated via,

$$\chi^2 = \hat{C}^T_{X,\ell}\mathbb{C}_X^{-1}\hat{C}_{X,\ell}, \tag{16}$$

where the covariance matrix $\mathbb{C}_X = < \hat{C}_{X,\ell}\hat{C}_{X,\ell'} >$ is constructed from the lognormal mocks. This χ^2 statistics is measured for every mock realization with the *leave-one-out* technique[5], and then is compared to the χ^2 value observed from the DR9. We only use the bandpower bins from $\ell = 2$ to 20 with $\Delta\ell = 2$, although higher modes are included for a robustness test later.

Fig. 6 shows the measured C_X from the DR9 sample before and after applying various corrections for imaging systematics. The dark and light shades show the 97.5$^{\rm th}$ percentile from the $f_{\rm NL} = 0$ and 76.9 mocks, respectively. The LRG sample has the highest cross correlation against the extinction, stellar density, and the z-band depth (solid black curve). As the most conservative cleaning approach, a linear model is trained with the extinction and z-band depth (*linear conservative I*). With the correction applied, we find that the cross power spectrum against the r-band psftsize increases (dot-dashed blue curve), which indicates that only the extinction and the z-band depth are not sufficient to regress out all of the residual cross correlations between the LRG density and imaging. With the linear model re-trained using the three maps (*linear conservative II*), we are able to reduce the cross power spectra (dotted orange curve). For comparison, we also show the cross spectra for a case in which the DR9 is cleaned using the linear model trained with all imaging maps (dashed red curve). There are minor differences between the *linear conservative II* and *linear all maps* cross spectra, further supporting the idea that only three imaging maps are sufficient to regress out spurious correlations.

[5] 999 realizations are used to estimate the covariance matrix and applied to measure the χ^2 for the left-out realization.

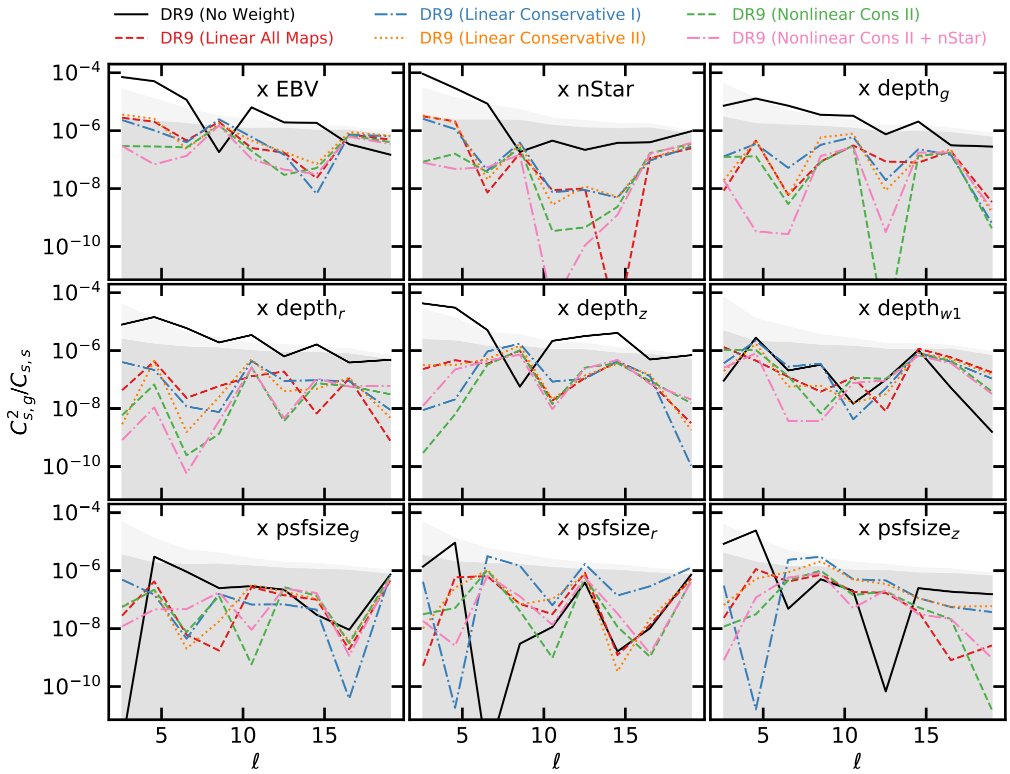

Figure 6. Cross power spectra between the DR9 LRG sample and imaging maps. Dark and light shades represent the 97.5 percentile of 1000 lognormal mocks without and with PNG, respectively.

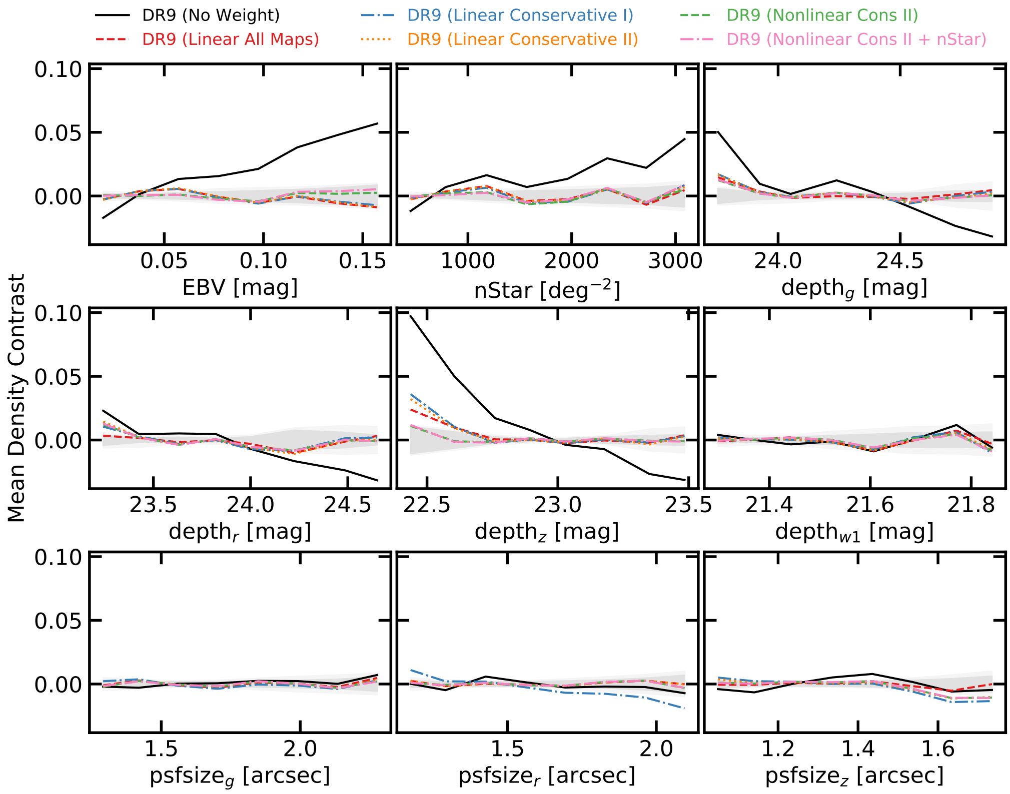

Figure 7. Mean density contrast of the DR9 LRG sample as a function of imaging maps. Dark and light shades represent the 1σ dispersion of 1000 lognormal mocks without and with PNG, respectively.

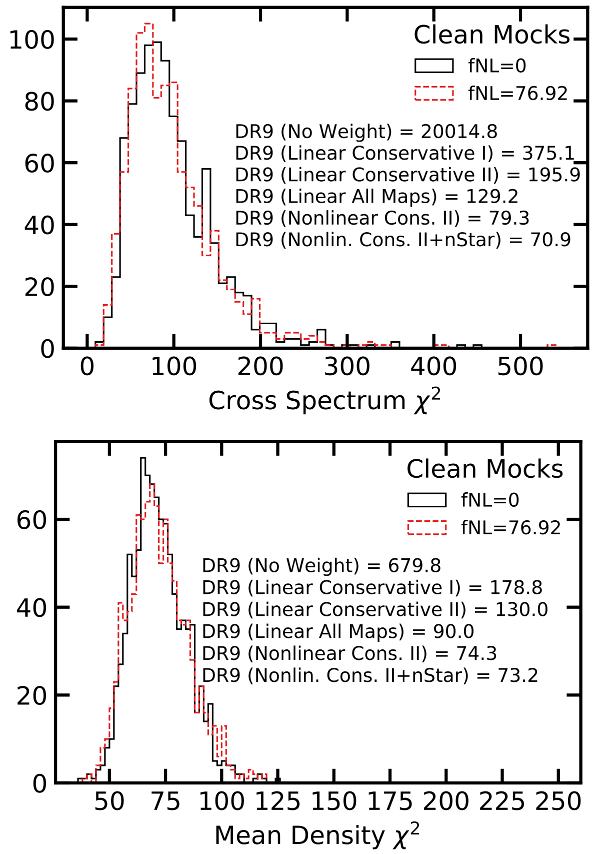

Figure 8. Remaining systematic error χ^2 from the galaxy-imaging cross power spectrum (top) and the mean galaxy density contrast (bottom). The values observed in the DR9 sample before and after linear and nonlinear treatments are quoted, and the histograms are constructed from 1000 realizations of clean lognormal mocks with $f_{\rm NL}$ = 0 and 76.92.

Fig. 8 (top) shows the histogram of the cross spectrum χ^2 from 1000 mocks with and without $f_{\rm NL}$. The χ^2 values observed in the DR9 sample are quoted for comparison. Before cleaning, the DR9 sample has a cross power spectrum χ^2 error of 20014.8. After correction with the linear conservative I approach, the cross spectrum χ^2 is reduced to 375.1 with p-value = 0.002. Adding the r-band psfsize, the linear model reduces the χ^2 down to 195.9 with p-value = 0.044; we can reject the null hypothesis that the DR9 sample with the linear conservative II is properly cleaned at 95% confidence. Although training the linear model with all imaging maps as input gives the lowest cross spectrum χ^2 of 129.2 (and p-value = 0.239), but it potentially makes the analysis more prone to over-fitting and regressing out the true clustering signal, given the inner correlations among the imaging properties (see, Fig. 3). As an alternative, we train the nonlinear method with the extinction, z-band depth, and r-band psfsize maps (*nonlinear conservative II*) and clean the DR9 sample. The cross power spectrum χ^2 is reduced to 79.3 with p-value = 0.594. Adding the stellar density map reduces the cross power spectrum χ^2 error to 70.9 (p-value = 0.687). This cross power spectrum diagnostic supports the idea that a nonlinear cleaning approach is desired to null out the remaining spurious fluctuations. We investigate cross spectrum to higher multipoles but find no evidence of remaining systematic error (see Appendix A).

3.4.2 Mean density contrast

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

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

where the summations are over HEALPix pixels in the bin with similar imaging values. For instance, In the absence of systematic error, the mean density contrast from one part of the sky with high extinction should be consistent with that from another patch with low extinction. We compute the histograms against all other imaging properties (see Fig. 2), and construct the total mean density contract as,

$$\delta_X = [\delta_{x_1}, \delta_{x_2}, \delta_{x_3}, ..., \delta_{x_9}], \tag{18}$$

and the total residual error as,

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

where the covariance matrix $\mathbb{C}_\delta =< \delta_X \delta_X >$ is constructed from the lognormal mocks. Fig. 7 shows the mean density contrast against the imaging properties for the DR9 LRG sample. The dark and light shades represent the 1σ level fluctuations observed in 1000 lognormal density fields respectively with $f_{\rm NL}$ = 0 and 76.92. The DR9 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 spurious trends against the extinction and stellar density at about 5−6%. The linear approach is able to mitigate most of the systematic fluctuations with only the extinction and z-band depth as input maps; however, a new trend appears against the r-band psfsize with the linear conservative I approach (dot-dashed blue curve), which is indicative of the psfsize-related systematics in the LRG sample. This finding is in agreement with the cross power spectrum. We retrain the linear model with three maps identified as *conservative II*, 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 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.

Fig. 8 (bottom) shows the mean density χ^2 observed in the mocks with or without $f_{\rm NL}$. We find consistent results regardless of the underlying $f_{\rm NL}$, which supports that our diagnostic is not sensitive to the fiducial cosmology. The values measured in the DR9 sample before and after applying imaging weights are quoted for comparison. The *linear conservative I* weights reduce the χ^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 χ^2 = 130. Training the linear model with all imaging maps returns a more reasonable χ^2 = 90 and p-value of 0.084; however, regression with all imaging maps as input can lead to the removal of the true clustering signal. We try the *nonlinear conservative II* approach. We obtain a χ^2 value of 74.3 with p-value = 0.392. Retraining the nonlinear approach while adding the stellar map yields minor improvement: χ^2 = 73.2 and p-value = 0.422. This indicates that the stellar density trend in the mean density of LRGs can be explained via the extinction map.

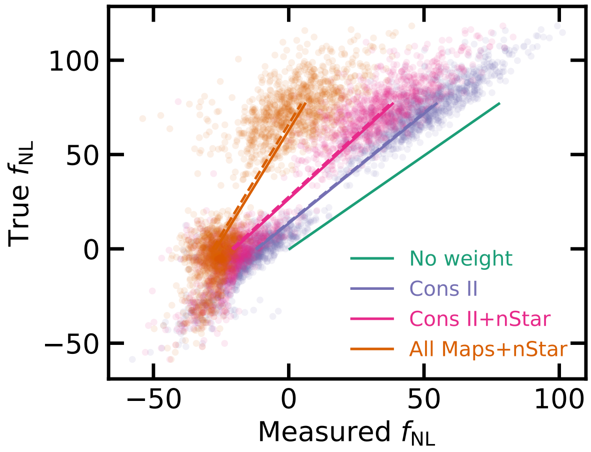

Figure 9. True vs measured f_{NL} values from the $f_{NL} = 0$ and 76.9 mocks. The solid (dashed) lines represent the best fit estimates from fitting the mean power spectrum of the clean (contaminated) mocks. The scatter points show the best fit estimates from fitting the individual spectra for the clean mocks.

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

Cleaning Method	m_1	m_2
Nonlinear Conservative II	1.17	13.95
Nonlinear Conservative II+nStar	1.32	26.97
Nonlinear All Maps+nStar	2.35	63.5

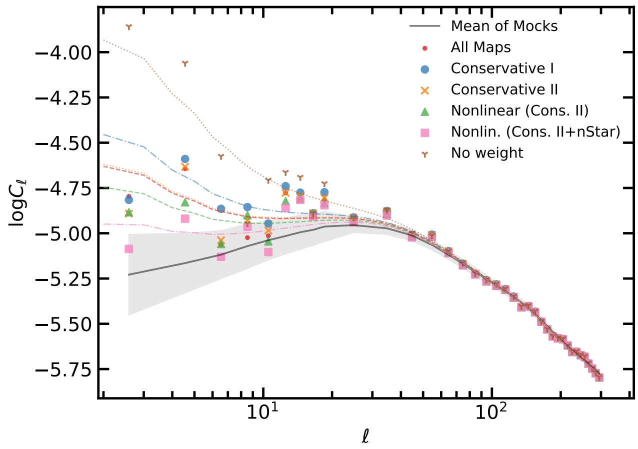

Figure 10. The angular power spectrum of the DR9 LRG sample before (*No weight*) and after correcting for imaging systematics using various methods with their corresponding best fit theory curves. The shade represents 1σ error constructed from the $f_{NL} = 0$ mocks.

3.5 Calibration of mitigation bias

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. Below we describe an approach to calibrate our cleaning methods and de-bias our f_{NL} constraints.

We utilize our series of lognormal density fields with and without PNG, with and without systematic effects. The contamination models in the simulations are based on a linear multivariate model with the extinction, z-band depth, and r-band psfsize and are drawn from the sampled posterior constrained from fitting the DR9 sample. The idea is to simulate systematic effects that reflect realistic spurious fluctuations in the real data. For correction, the nonlinear model is trained and applied to the simulations with various sets of imaging maps as input, namely, we consider *nonlinear conservative II*, *nonlinear conservative II + nStar*, and *nonlinear All Maps + nStar*. We fit 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_{NL} values and the estimates from the mocks with mitigation applied are considered as the measured values. Fig. 9 shows the true f_{NL} values and the measured f_{NL} values from fitting the mean power spectrum (solid lines) and individual spectra (points). The dashed lines show the best fit estimates from the contaminated realizations. The best fit estimates for the contaminated mocks before cleaning (*no weight*) or per realizations are not shown for visual clarity.

Then, a pair of linear coefficients are found to map the measured to the true values of f_{NL}, e.g., $f_{NL,true} = m_1 f_{NL,measured} + m_2$. These coefficients for the mean power spectrum are summarized in Tab. 2. The parameters for the *All Maps+nStar* approach are the highest, as more mitigation bias is expected when more imaging maps are used. The value for $m_1 - 1$ determines the added uncertainty in the f_{NL} inference due to mitigation, which is the highest for *All Maps+nStar* and smallest for *Conservative II*.

4 RESULTS

4.1 DESI imaging LRG sample

Fig. 10 shows the measured power spectrum of the DESI imaging DR9 LRG sample before and after applying imaging weights, the best fit theory curves, and the mean power spectrum and 1σ error estimated from the $f_{NL} = 0$ lognormal simulations. The power spectra are similar on small scales ($\ell > 20$), but the differences between various cleaning methods are significant on large scales. By comparing *linear conservative I* to *linear conservative II*, we find that the measure power spectra on modes with $6 \leq \ell < 10$ are noticeably different between the two methods. We associate the differences to the r-band psfsize template in *linear conservative II*. On other scales, the differences between the spectra after the linear-based cleaning are negligible, supporting the idea that our feature selection procedure has worked to identify the primary maps causing excess clustering signal. Comparing *nonlinear conservative II* to *linear conservative II*, we find that the measured spectra on $4 \leq \ell < 6$ are very different, probably pointing at nonlinear spurious fluctuations with large-scale characteristics due to the extinction. Adding stellar density to the nonlinear approach (*nonlinear conservative II + nStar*) results in less excess power relative to the mock power spectrym, with the modes on $2 \leq \ell < 4$ reflecting the biggest change. The flexibility of the nonlinear approach to correct for these effects ameliorates the clustering measurements on these scales.

4.1.1 Calibrated constraints

Tab. 3 describes the best fit and marginalized mean estimates of f_{NL} from fitting the power spectrum of the DR9 LRG sample before and after cleaning with the nonlinear approach given var-

Table 3. Calibrated best fit and marginalized mean estimates for $f_{\rm NL}$ from fitting power spectrum of the DESI DR9 LRG sample before and after correcting for systematics. Degree of freedom is 34 (37 data points - 3 parameters).

		$f_{\rm NL}$				
Footprint	Method	Best fit	Mean	68% CL	95% CL	χ^2
DESI	No Weight	113.18	115.49	$98.14 < f_{\rm NL} < 132.89$	$83.51 < f_{\rm NL} < 151.59$	44.4
DESI	Nonlinear (Cons. II)	47.38	48.81	$36.08 < f_{\rm NL} < 61.44$	$25.03 < f_{\rm NL} < 75.64$	34.6
DESI	Nonlin. (Cons. II+nStar)	48.92	50.10	$36.88 < f_{\rm NL} < 63.31$	$24.87 < f_{\rm NL} < 77.78$	35.2
DESI	Nonlin. (All Maps+nStar)	49.69	41.91	$13.10 < f_{\rm NL} < 69.14$	$-15.96 < f_{\rm NL} < 91.84$	39.5

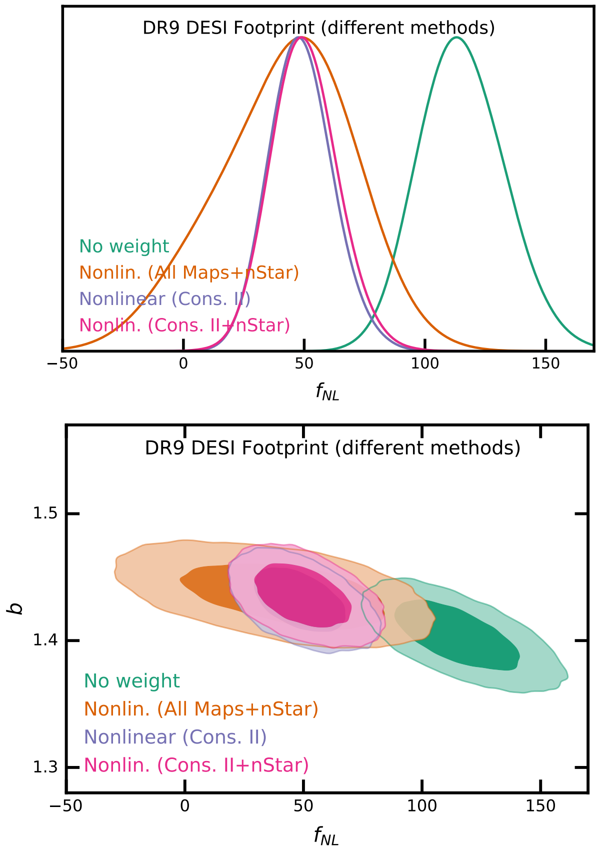

Figure 11. Calibrated constrains from the DR9 LRG sample. *Top*: probability distribution for $f_{\rm NL}$ marginalized over the shotnoise and bias. *Bottom*: 68% and 95% probability distribution contours for the bias and $f_{\rm NL}$ from the DR9 LRG sample before and after applying nonlinear cleaning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

ious combinations of imaging templates. All constraints are calibrated for the effect of mitigation bias using the lognormal simulations. With the corrections applied, we obtain $36.08(25.03) < f_{\rm NL} < 61.44(75.64)$ with $\chi^2 = 34.6$ for *nonlinear conservative II*, $36.88(24.87) < f_{\rm NL} < 63.31(77.78)$ with $\chi^2 = 35.2$ for *nonlinear conservative II + nStar*, and $13.10(-15.96) < f_{\rm NL} < 69.14(91.84)$ with $\chi^2 = 39.5$ for *nonlinear all maps + nStar* at 68%(95%) confidence over 34 degrees of freedom. here

Fig 11 shows the marginalized probability distribution for $f_{\rm NL}$ (top) and the 68% and 95% probability contours for the bias parameter and $f_{\rm NL}$ from the DR9 sample before (*no weight*) and after applying different cleaning schemes. No weight constraint at 68% is $98.14 < f_{\rm NL} < 132.89$ with a best fit of 113.18 and marginalized mean of 115.49, and is more than 2σ off from zero. Applying imaging weights shifts constraints to lower $f_{\rm NL}$ values, and b is slightly pulled upward since excess clustering due to systematics is removed. Using all maps with the linear model does not change the results, showing that three maps are sufficient at the linear level to mitigate systematics. As an alternative, using a nonlinear model with three maps shows around 1σ shift, with 68% confidence at $18.91 < f_{\rm NL} < 40.59$, inconsistent with zero for more than 2σ. Adding a template for the local stellar density shift constraints by 1σ, making it consistent with zero. As the most rigorous approach, using all maps and stellar density included results in more than 2σ shift. We emphasize that these shifts to lower $f_{\rm NL}$ are somewhat expected as more input maps results in regressing more modes from cosmological clustering signal. Therefore, we use lognormal mocks to calibrate the amount of signal that is removed in each case and attempt to undo the effect.

4.1.2 Robustness tests

These results are subject to mitigation bias. $f_{\rm NL}$ constraints from DR9 LRG sample is summarized in Table 4. First, we focus on the DESI footprint and then compare constraints obtained from each sub-survey. We also evaluate the robustness of constraints against various cuts and configurations. First, we compare how constraints from whole DESI footprint compares to those from each survey individually, namely BASS+MzLS, DECaLS Nouth, and DECaLS South. Fig. 12 shows 68% and 95% confidence on $f_{\rm NL}$ and b from each individual survey or all combined as DESI. Constraints from all surveys are consistent and agree with each other within 68%. Both BASS+MzLS and DECaLS South are consistent with zero PNG, but DECaLS North deviates from zero at more than 2σ. Adding the stellar density template does not change constraints from BASS+MzLS much, but it shifts DECaLS North and DECaLS South by 0.5σ and σ, respectively. This might indicate that there are some unresolved issues with stellar contamination in DECaLS North and DECaLS South. We note that differences are more significant when all maps and stellar density are used as input. This is expected as more maps mean the model has more freedom to take out clustering modes.

- **Pixel completeness** We remove pixels with low completeness from the DESI footprint by applying $f_{\rm pix} > 0.5$, and find that the impact is negligible. Specifically, the cut removes .6% survey area and causes best fit $f_{\rm NL}$ shifts only around 2%, from 28.58 to 28.07, see *comp cut* Table 4. When investigated this impact on each region separately, BASS+MzLS increases around 10%, DECaLS North decreases 1%, and DECaLS South decreases around 5%.
- **Imaging quality** We remove pixels with poor imaging from the DESI footprint by applying the following cuts on imaging properties; $E[B-V] < 0.1$, $nStar < 3000$, $depth_g > 23.2$, $depth_r > 22.6$, $depth_z > 22.5$, $psfsize_g < 2.5$, $psfsize_r < 2.5$, and $psfsize_z < 2$. Overall the constraints are consistent despite best fit and marginalized mean estimates shift. Quantitatively, we lose

Table 4. Uncalibrated best fit and marginalized mean estimates for $f_{\rm NL}$ from fitting power spectrum of the DR9 LRG sample before and after correcting for systematics. Degree of freedom is 34 (37 data points - 3 parameters).

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

about 8.2% survey area, and the best fit $f_{\rm NL}$ estimate changes about 2% from 28.58 to 29.16. See *imag cut* in Table 4. For BASS+MzLS only, the imaging cut increases the best fit by 62% from 15.43 to 25.03. For DECaLS North and DECaLS South, the best fit increases by 5% and 15% respectively.

- **Covariance** We now use the mocks with $f_{\rm NL} = 76.92$ to construct a covariance matrix, and with the new covariance we observe a 12% increase in the $f_{\rm NL}$ constraint uncertainties and 11% increase in the best fit estimate of $f_{\rm NL}$.
- **Lowest** ℓ We decrease the largest mode (or increase the lowest ℓ) used in estimating the best fit and 68% confidence intervals. Fig. 13 illustrates the results for the DESI footprint and how they compared to BASS+MzLS, DECaLS North, or DECaLS South only results. Points represent marginalized mean estimates of $f_{\rm NL}$ and errorbars represent 68% confidence from MCMC results.Overall we find that the constraints are robust against the largest mode.
- **External maps** We also derive imaging weights using additional external maps for the neutral hydrogen column density (HI) and magnitude calibration errors in the z band (CALIBZ). With the new weights, we find the best fit estimates increase from 41.02 to 55.46 for DECaLS North and from 31.24 to 33.79 for DECaLS South.
- **Declination cut** Our default analysis do not use the spurious islands in DECaLS North and DECaLS South below DEC = -30 to avoid potential calibration issues. PANASTARS are used for calibration below DEC of -30. Without these cuts, best fit $f_{\rm NL}$ estimates increase from 31.24 to 43.79 for DECaLS South and decrease from 41.02 to 41.05. This indicates that indeed there is an issue with DECaLS South below DEC of −30.

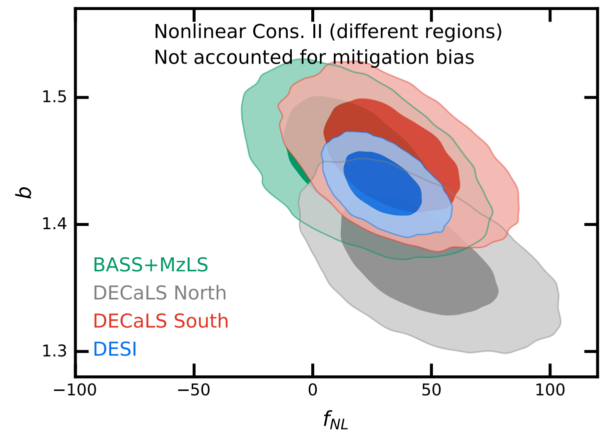

Figure 12. Uncalibrated 2D constraints from the DR9 LRG sample for each imaging survey compared with that for the whole DESI footprint. The dark and light shades represent the 68% and 95% confidence intervals, respectively.

Overall we find that the declination cut is necessary for DECaLS South, while adding external templates for HI and CALIBZ, using a different covariance, or applying imaging and completeness cuts do not alter the constraints significantly.

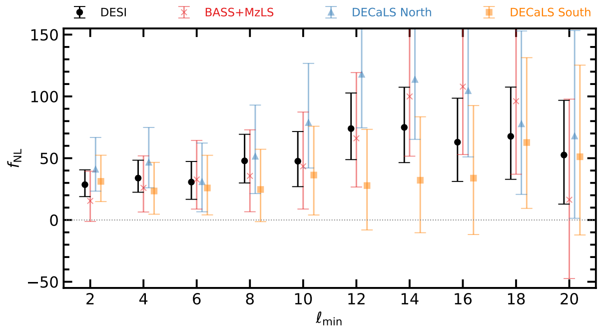

Figure 13. Robustness of the uncalibrated DR9 constraints w.r.t. the largest scale (lowest ℓ mode) used in MCMC regression. Points represent marginalized mean estimates of $f_{\rm NL}$ and errorbars represent 68% confidence.

5 CONCLUSIONS

We have presented constraints on $f_{\rm NL}$ using the scale-dependent bias effect in the large-scale clustering of DESI imaging DR9 LRGs. Methods from linear and nonlinear regressions are applied for data cleaning from foreground and imaging systematic effects. Same tools are tested on lognormal density fields to evaluate the sensitivity of signal to systematic error. As summarized in Table ??, we find that fitting $\log C_\ell$ rather than C_ℓ minimizes the dependence of constraints to the choice of covariance, and we are able to recover the truth $f_{\rm NL}$ at 95% confidence in both simulations with and without PNG. Table ??? summarizes the constraints from mocks undergone cleaning for systematics. We find that template-based regression removes clustering and thus biased constraints are obtained. We use the mock constraints to calibrate DR9 constraints. We obtain $36.07(25.03) < f_{\rm NL} < 61.44(75.64)$ when cleaning is performed with the nonlinear model using only three imaging maps. With more extreme cleaning using all maps and stellar density, we obtain $13.09(-15.95) < f_{\rm NL} < 69.14(91.84)$ at 68%(95%).

Various tests are performed to assess the robustness of constraints against analysis assumptions, and the results are summarized in Table ??. We find

- constraints from individual surveys are consistent with each other
- no significant shift observed in constraints after applying imaging cut, completeness cut
- no significant shift after including additional imaging templates for hydrogen column density or calibration in the z-band.
- region below dec of -30 indicates some issues probably due to unaccounted for calibration issues
- constraints are robust against the largest scales (lowest ℓ mode) used in fitting for $f_{\rm NL}$. Some signs of systematics on $10 < \ell < 18$.

ACKNOWLEDGEMENTS

MR is supported by XXXXXX. We acknowledge the support and resources from the Ohio Supercomputer Center (OSC; Center 1987). MR is thankful for hospitality at the OSU CCAPP in particular facilitated by John Beacom and Lisa Colarosa. We would like to thank the open-source software and resources that were beneficial to this research: Pytorch, Nbodykit, HEALPix, Fitsio, Scikit-Learn, NumPy, SciPy, Pandas, IPython, Jupyter, arXiv, and GitHub.

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

REFERENCES

Abbott T., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1270
Ade P., et al., 2019, Journal of Cosmology and Astroparticle Physics, 2019, 056
Aghamousa A., et al., 2016, arXiv preprint arXiv:1611.00036
Aghanim N., et al., 2020, Astronomy & Astrophysics, 641, A6
Akrami Y., et al., 2019, arXiv preprint arXiv:1905.05697
Alvarez M., et al., 2014, arXiv e-prints, p. arXiv:1412.4671
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
Bassett 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
Castorina E., Moradinezhad Dizgah A., 2020, J. Cosmology Astropart. Phys., 2020, 007
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
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
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., Kilbinger 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 Society, 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., 2017, arXiv e-prints, p. arXiv:1711.05101
Lyth D. H., Liddle A. R., 2009, The primordial density perturbation: Cosmology, inflation and the origin of structure. Cambridge University Press
Meisner A. M., Lang D., Schlegel D. J., 2018, Research Notes of the American Astronomical Society, 2, 1
Merz G., et al., 2021, Monthly Notices of the Royal Astronomical Society, 506, 2503
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
Padmanabhan N., et al., 2007, MNRAS, 378, 852
Peacock J., Nicholson D., 1991, Monthly Notices of the Royal Astronomical Society, 253, 307
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
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., 2020, MNRAS, 498, 2354
Sabti N., Muñoz J. B., Blas D., 2021, J. Cosmology Astropart. Phys., 2021, 010
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, The Astrophysical Journal, 500, 525
Schmittfull M., Seljak U., 2018, Phys. Rev. D, 97, 123540
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 S., 2008, Cosmology. OUP Oxford
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo 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

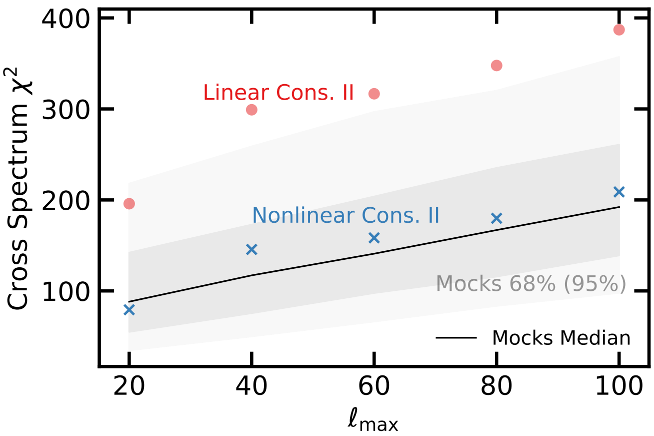

Figure A1. Cross power spectrum χ^2 as a function of the highest mode $\ell_{\rm max}$ for the DR9 LRG sample using the linear and nonlinear imaging weights with the conservative II maps. The lowest mode is fixed at $\ell_{\rm min} = 2$. Solid curve and dark (light) shade represent the median estimate and 68% (95%) confidence constructed from the $f_{\rm NL} = 0$ mocks.

APPENDIX A: SCALE DEPENDENT SYSTEMATICS

HERE We further test the stability of our results by extending the highest mode from $\ell = 20$ to 100, or fluctuations over scales as small as 1.8 degrees (see, Fig. A1). The solid line shows how the median of 1000 mocks changes as the highest ℓ increases from 20 to 100. The red circles show the chi2 for the linear approach with three maps and the blue crosses show the chi2 for the nonlinear approach with three maps. However, as we show later, our second diagnostic based on the mean density contrast reveals that there is a residual systematic error against the z-band depth with the linear cleaning approach even though the z-band depth was an input for training.

APPENDIX B: LOGNORMAL MOCKS

B1 Clean density fields

Corner plots of the PNG parameter $f_{\rm NL}$ and bias coefficient are shown in Fig. **??** for fitting the mean power spectrum of mocks, with and without $f_{\rm NL}$. Best fit estimates, marginalized mean, 1σ and 2σ confidence intervals are summarized in Tab. B1. Fig **??** (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 $f_{\rm NL}$ = 76.9 mocks or its log transformation using covariance matrices constructed from the same set of mocks or from the $f_{\rm NL} = 0$ 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 $f_{\rm NL} = 76.9$ mocks is fit using the covariance matrix estimated from the $f_{\rm NL} = 0$ 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 $f_{\rm NL}$-dependent covariance matrices and make the constraints less sensitive to covariance construction. Fig. **??** (left) shows the confidence contours for $f_{\rm NL} = 0$ mocks when fit is done to the log of mean spectra of $f_{\rm NL} = 0$ mocks for the different regions. We find that the underlying true $f_{\rm NL}$ value is recovered

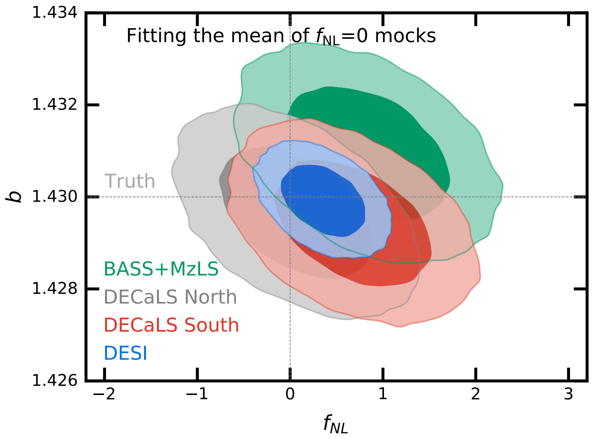

Figure B1. 68% and 95% confidence contours from the mean power spectrum of the $f_{\rm NL} = 0$ mocks for the DESI footprint and sub-imaging surveys. The truth values are represented by vertical and horizontal lines.

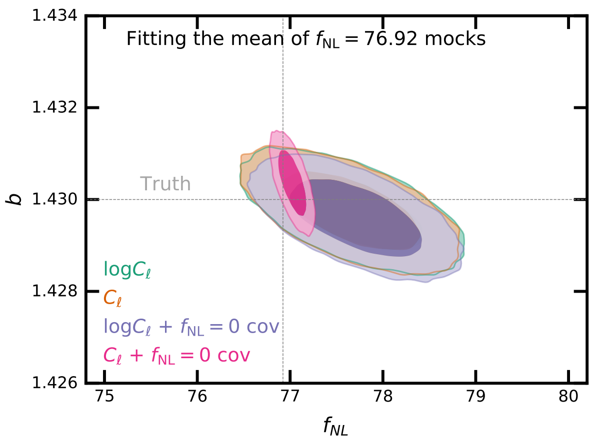

Figure B2. 68% and 95% confidence contours of fitting the mean power spectrum or its log transformation from the $f_{\rm NL} = 76.92$ mocks for the DESI footprint. Using the log C_ℓ fitting yield constraints that are insensitive to the covariance used. The truth values are represented by vertical and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

within 2σ confidence. Add a paragraph for the constraining power vs fsky.

Fig B3 shows the best fit estimates for b vs $f_{\rm NL}$ for $f_{\rm NL} = 0$ 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 χ^2 from fit for each realization. The histograms of best fit $f_{\rm NL}$ estimates are plotted in the background. We obtain $f_{\rm NL} = MU \pm STD$ and $= MU \pm STD$ for the left and right panels, respectively.

B1.1 Contaminated density fields

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

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

Table B2. Best fit and marginalized estimates for $f_{\rm NL}$ from fitting the mean power spectrum of the mocks before and after applying imaging weights.

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

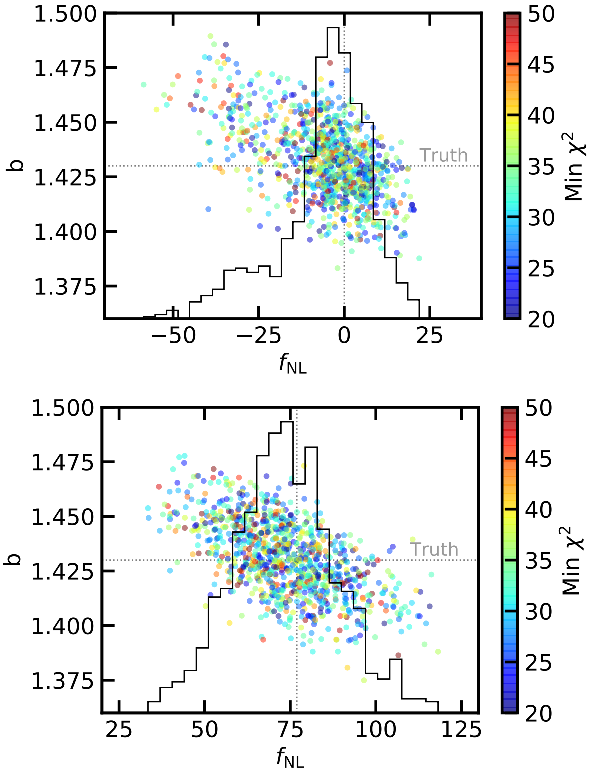

Figure B3. Top: 68% and 95% confidence contours for $f_{\rm NL} = 0$ (left) and 76.92 (right) mocks. Using the log C_ℓ fitting yield constraints that are insensitive to the covariance used. Bottom: best fit estimates from fitting 1000 lognormal mocks with $f_{\rm NL} = 0$ (left) and 76.92 (right) in the DESI footprint. The truth values are represented by vertical and horizontal lines.

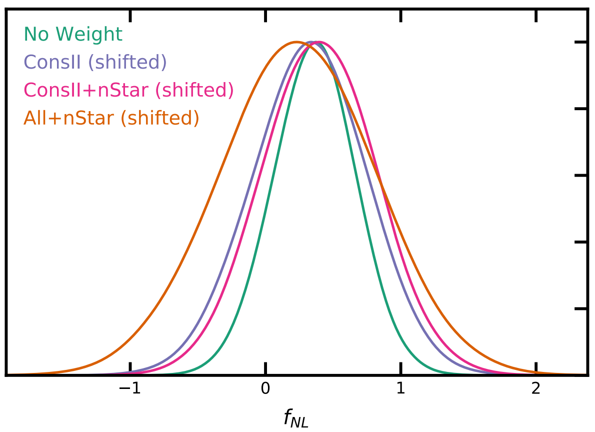

Figure B4. Shifted probability distributions of $f_{\rm NL}$ from the $f_{\rm NL} = 0$ mocks.

This paper has been typeset from a TEX/LATEX file prepared by the author.