MNRAS 000, 1–27 (2022) Preprint 24 May 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[3,4], Hui Kong[5], Anna Porredon[6], Lado Samushia[1], Edmond Chaussidon[7], Alex Krolewski[8,9,10], Arnaud de Mattia[7], Santiago Avila[5], Benedict Bahr-Kalus[11], Jose R. Bermejo-Climent[12], Florian Beutler[6], Klaus Honscheid[2,13], Eva-Maria Mueller[14,15], Adam Myers[16], Jeffrey A. Newman[17], Nathalie Palanque-Delabrouille[4,7], Will Percival[8,9,10], Christophe Yeche[7]

[1]*Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*
[2]*Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[3]*Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA*
[4]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[5] *Institut de Física d'Altes Energies, The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain*
[6]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[7] *IRFU, CEA, Universite Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[8]*Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[9]*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*
[10]*Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[11]*Korea Astronomy and Space Science Institute, Yuseong-gu, Daedeok-daero 776, Daejeon 34055, Republic of Korea*
[12] *Department of Physics and Astronomy, University of Rochester, 500 Joseph C. Wilson Boulevard, Rochester, NY 14627, USA*
[13]*Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA*
[14]*Institute of Cosmology and Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK*
[15]*Department of Physics & Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[16]*Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA*
[17]*Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We use 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 $f_{\rm NL}$. Our sample comprises over 12 million LRG targets, spanning approximately 14,000 square degrees of the sky, with redshifts ~~ranging from~~ in the range $0.2 < z < 1.35$. Galactic extinction, survey depth, and astronomical seeing are identified as the primary sources of systematic error using feature selection and cross-correlation techniques. Linear regression and artificial neural networks are applied to mitigate systematics and alleviate non-cosmological excess clustering signals on large scales. Our treatment methods are tested and calibrated rigorously against log-normal density simulations with and without $f_{\rm NL}$ and systematic effects. We find that the neural network treatment outperforms linear regression in reducing remaining systematics in the DESI LRG sample. Assuming the universality relation, we find $f_{\rm NL} = 47^{+14\,(+29)}_{-11\,(-22)}$ at 68% (95%) confidence for our fiducial method. Applying a more aggressive systematics treatment that includes regression against the full set of imaging maps we identified, our maximum likelihood value changes only slightly to $f_{\rm NL} \sim 50$, but the uncertainty on $f_{\rm NL}$ increases due to the aggressive treatment removing large-scale clustering information. We apply a series of robustness tests (e.g., cuts on imaging, declination, or scales) that show remarkable consistency in the obtained constraints. We measure $f_{\rm NL} > 0$ at 99.9 per cent even after various systematics mitigations, which could be either due to unforeseen systematics (e.g., from calibration errors in photometric zero-point determination or in Galactic extinction corrections) or a new cosmological signal (corresponding to some kind of scale-dependent $f_{\rm NL}$ model which induces a large non-gaussianity on scales around large-scale structure but leaves cosmic microwave background scales intact). ~~Our fiducial result can either be interpreted as a strong detection of non-zero $f_{\rm NL}$ with the probability $P(f_{\rm NL} > 0) = 99.9$ per cent, inconsistent with what is measured from Planck, or evidence for unknown sources of variation in the observed galaxy density (e.g., from calibration errors in photometric zero-point determination or in Galactic extinction corrections).~~ Our results motivate follow-up studies of $f_{\rm NL}$ with DESI spectroscopic samples, where the inclusion of 3D clustering modes should help separate imaging systematics.

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 (Guth 1981; Linde 1982; Albrecht & Steinhardt 1982). 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 compelling explanation, the characteristics of the field or fields that drove the inflationary expansion remain largely unknown in 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 (Komatsu et al. 2003; Tegmark et al. 2004; Guth & Kaiser 2005), some alternative classes of inflationary models predict different 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 the growth of structure and galaxies in the Universe (see, e.g., Verde 2010; Desjacques & Seljak 2010; Biagetti 2019).

In its simplest form, local primordial non-Gaussianity (PNG) is parameterized by the non-linear coupling constant $f_{\rm NL}$(Komatsu & Spergel 2001):

$$\Phi = \phi + f_{\rm NL}[\phi^2 - < \phi^2 >], \qquad (1)$$

where Φ is the primordial curvature perturbation and ϕ is assumed to be a Gaussian random field. Local-type PNG generates a primordial bispectrum, which peaks in the squeezed triangle configuration 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. The coupling between the short and long wavelengths induces a distinct bias in the galaxy distribution, which leads to a k^{-2}-dependent feature in the two-point clustering of galaxies and quasars (Dalal et al. 2008). Obtaining reliable, accurate, and robust constraints on $f_{\rm NL}$ is crucial in advancing our understanding of the dynamics of the early Universe. For instance, the standard single-field slow-roll inflationary model predicts a small value of $f_{\rm NL} \sim 0.01$ (see, e.g., Maldacena 2003). 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 $f_{\rm NL}$ that can reach up to 100 or higher (see, e.g., Chen 2010, for a review). With $\sigma(f_{\rm NL}) \sim 1$, we can rule out or confirm specific models of inflation and gain insight into the physics that drove the inflationary expansion (see, e.g., Alvarez et al. 2014; de Putter et al. 2017).

The current tightest bound on $f_{\rm NL}$ comes from Planck's bispectrum measurement of CMB anisotropies, $f_{\rm NL} = 0.9 \pm 5.1$ (Planck Collaboration et al. 2019). Limited by cosmic variance, CMB data cannot enhance the statistical precision of $f_{\rm NL}$ measurements enough to break the degeneracy amongst various inflationary paradigms (see, e.g., Abazajian et al. 2016; Simons Observatory et al. 2019). On the other hand, LSS surveys probe a 3D map of the Universe, and thus provide more modes to limit $f_{\rm NL}$. However, nonlinearities raised from structure formation pose a serious challenge for measuring $f_{\rm NL}$ with the three-point clustering of galaxies, and these nonlinear effects are non-trivial to model and disentangle from the primordial signal (Baldauf et al. 2011b,a). Currently, the most precise constraints on $f_{\rm NL}$ from LSS reach a level of $\sigma(f_{\rm NL}) \sim 20-30$, with the majority of the constraining power coming from the two-point clustering statistics that utilize the scale-dependent bias effect (Slosar et al. 2008; Ross et al. 2013; Castorina et al. 2019; Mueller et al. 2022; Cabass et al. 2022; D'Amico et al. 2022). Surveying large areas of the sky can unlock more modes and help improve these constraints.

The Dark Energy Spectroscopic Instrument (DESI) is ideally suited to enable excellent constraints on primordial non-Gaussianity from the galaxy distribution. ~~As an ongoing wide-area galaxy survey, the Dark Energy Spectroscopic Instrument (DESI)~~ DESI uses 5000 robotically-driven fibers to simultaneously collect spectra of extra-galactic objects (Levi et al. 2013; DESI Collaboration et al. 2016b; Silber et al. 2023). DESI is designed to deliver an unparalleled volume of spectroscopic data covering $\sim 14,000$ square degrees that promises to deepen our understanding of the energy contents of the Universe, neutrino masses, and the nature of gravity (DESI Collaboration et al. 2022). Moreover, DESI alone is expected to improve our constraints on local PNG down to $\sigma(f_{\rm NL}) = 5$, assuming systematic uncertainties are under control (DESI Collaboration et al. 2016a). With multi-tracer techniques (Seljak 2009), cosmic variance can be further reduced to allow surpassing CMB-like constraints (Alonso et al. 2015). For instance, the distortion of CMB photons around foreground masses, which is referred to as CMB lensing, provides an additional probe of LSS, but from a different vantage point. We can significantly reduce statistical uncertainties below $\sigma(f_{\rm NL}) \sim 1$ by cross-correlating LSS data with CMB-lensing, or other tracers of matter, such as 21 cm intensity mapping (see, e.g., Schmittfull & Seljak 2018; Heinrich & Doré 2022; Jolicoeur et al. 2023; Sullivan et al. 2023).

However, further work is needed to fully harness the potential of the scale-dependent bias effect in constraining $f_{\rm NL}$ with LSS. The amplitude of the $f_{\rm NL}$ signal in the galaxy distribution is proportional to ~~a~~ the bias parameter b_ϕ, such that $\Delta b \propto b_\phi f_{\rm NL} k^{-2}$. DH: here you really need to say what b_{phi} is in words. Assuming the universality relation, $b_\phi \sim (b-p)$, where b is the linear halo bias and $p = 1$ is a parameter that describes the response of galaxy formation to primordial potential perturbations in the presence of local PNG (see, e.g., Slosar et al. 2008). The ~~theoretical uncertainties on~~value of p is not very well constrained for other tracers of matter (Barreira et al. 2020; Barreira 2020), and Barreira (2022) showed that marginalizing over p even with wide priors leads to biased $f_{\rm NL}$ constraints because of parameter space projection effects. More simulation-based studies are necessary to investigate the halo-assembly bias and the relationship between b_ϕ and b for various galaxy samples. For instance, 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 $b_\phi f_{\rm NL}$ term (see, e.g., Barreira 2020). Nevertheless, the detection significance of local PNG remains unaffected by various assumptions regarding p. This means that a nonzero detection of $b_\phi f_{\rm NL}$ at a certain confidence level will still indicate a nonzero detection of $f_{\rm NL}$ at that same confidence level. In this work, we assume ~~the universality relation and do not marginalize over p~~ the relation that links b_ϕ to $b-p$ and, further, fix the value of p.

In addition to the theoretical uncertainties, measuring $f_{\rm NL}$

through the scale-dependent bias effect is a difficult task due to various imaging systematic effects that can modulate the galaxy power spectrum on large scales. The imaging systematic effects often induce wide-angle variations in the density field, and in general, any large-scale variations can translate into an excess signal in the power spectrum (see, e.g., Huterer et al. 2013), that can be misinterpreted as the signature of non-zero local PNG (see, e.g., Thomas et al. 2011). Such spurious variations can be caused by Galactic foregrounds, such as dust extinction and stellar density, or varying imaging conditions, such as astrophysical seeing and survey depth (see, e.g., Ross et al. 2011). ~~As an example for a false detection, the BICEP2 experiment initially claimed more than 5σ evidence for primordial gravitational waves through the detection of B-mode polarization in CMB~~ (BICEP2 Collaboration et al. 2014)~~, but it was later found that the signal was likely caused by dust, rather than by primordial gravitational waves~~ (Flauger et al. 2014; Mortonson & Seljak 2014). The imaging systematic issues have made it challenging to accurately measure $f_{\rm NL}$, as demonstrated in previous efforts to constrain it using the large-scale clustering of galaxies and quasars (see, e.g., Ross et al. 2013; Pullen & Hirata 2013; Ho et al. 2015), and it is anticipated that they 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 to accommodate fainter galaxies (see, e.g, Kitanidis et al. 2020).

The primary objective of this paper is to utilize the scale-dependent bias signature in the angular power spectrum of galaxies selected from DESI imaging data to constrain the value of $f_{\rm NL}$. With an emphasis on a careful treatment of imaging systematic effects, we aim to lay the groundwork for subsequent studies of local PNG with DESI spectroscopy. To prepare our sample for measuring such a delicate signal, we employ linear multivariate regression and artificial neural networks to mitigate spurious density fluctuations and ameliorate the excess clustering power caused by imaging systematics. We thoroughly investigate potential sources of systematic error, including survey depth, astronomical seeing, photometric calibration, Galactic extinction, and local stellar density. Our methods and results are validated against simulations, with and without imaging systematics.

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 modelling 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 haloes, 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 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 programmes observed the same footprint in the North Galactic Cap (NGC) while the DECaLS programme 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 programme integrates observations executed from the Blanco instrument under the Dark Energy Survey (DES Collaboration et al. 2016), which cover about 1130 deg^2 of the South Galactic Cap (SGC) footprint. The DESI imaging catalogues 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).

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 vary for each imaging survey, but they are designed to achieve a nearly consistent density of approximately 800 galaxies per square degree across a total area of roughly 14,000 square degrees. Table 2 summarizes the mean galaxy density and area for each region. This is accomplished despite variations in 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 (DESI Collaboration in prep), and is shown via the solid curve in Figure 1. Zhou et al. (2021) analyzed the DESI LRG targets and found that the redshift evolution of the linear bias for these targets is consistent with a constant clustering amplitude and varies via $1/D(z)$, where $D(z)$ is the growth factor (as illustrated by the dashed red line in Figure 1).

The LRG sample is masked rigorously for foreground bright stars, bright galaxies, and clusters of galaxies[1] to further reduce stellar contamination (Zhou et al. 2022). Then, the sample is binned into HEALPix (Gorski et al. 2005) pixels at NSIDE = 256, corresponding to pixels of about 0.25 degrees on a side, to construct 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. Moreover, the density of galaxies is matched to the randoms

[1] See `https://www.legacysurvey.org/dr9/bitmasks/` for maskbit definitions.

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, $z_{\rm fiber}$, 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

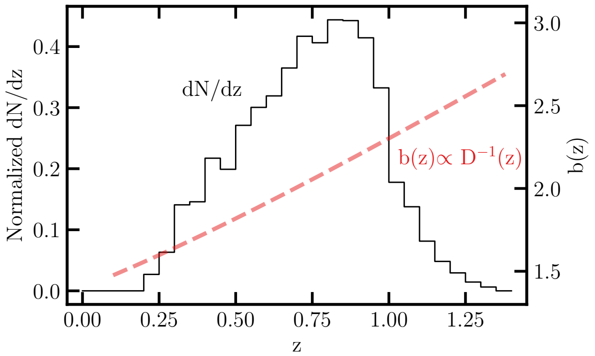

Figure 1. The redshift distribution (solid line and vertical scale on the left) and bias evolution (dashed line and vertical scale on the right) of the DESI LRG targets. The redshift distribution is determined from DESI spectroscopy (DESI Collaboration in prep). The redshift evolution of the linear bias is supported by HOD fits to the angular clustering of the DESI LRG targets (Zhou et al. 2021), where $D(z)$ represents the growth factor.

separately for each of the three data sections (BASS+MzLS, DECaLS North / South) so the mean density differences are mitigated (see Table 2). The DESI LRG targets are selected brighter than the imaging survey depth limits, e.g., $g = 24.4$, $r = 23.8$, and $z = 22.9$ for the median 5σ detection in AB mag in the DECaLS North region (Table 2); and thus the LRG density map does not exhibit severe spurious fluctuations.

2.1.1 *Imaging systematic maps*

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). Zhou et al. (2022) has previously identified nine astrophysical properties as potential sources of imaging systematic errors in the DESI LRG targets. These imaging properties are mapped into HEALPix of NSIDE= 256. As illustrated by the 3x3 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 (i.e., point spread function, or psfsize) in g, r, and z. The depth maps have been corrected for extinction using the coefficients adapted from Schlafly & Finkbeiner (2011). Table 2 summarizes the median values for the imaging properties in each region. 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 DESI Collaboration in (prep) to further test the robustness of our analysis against unknown systematics.

The fluctuations in each imaging map are unique and tend to be correlated with the LRG density map. For instance, large-scale LRG density fluctuations ~~can be associated with~~could be caused by stellar density, extinction, or survey depth; while small scale-fluctuations ~~can be related to~~could be caused by psfsize variations. 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 DEC < -10) or calibrated using different catalogues of standard stars (e.g., DEC < -30 in the SGC). The potential impact of not imposing these declination cuts on the LRG sample and our $f_{\rm NL}$ constraints is explored in Section 4.

We employ the Pearson correlation coefficient to characterize the correlation between the galaxy and imaging properties, which for two random variables x and y is given by,

$$\text{Pearson}\,(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \tag{2}$$

where $\bar{x}$ and $\bar{y}$ represent the mean estimates of the random variables. 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 for no correlation and are constructed by cross-correlating 100 synthetic lognormal density fields and the imaging systematic maps. Consistent among the different regions, 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.~~ Figure 3 (bottom panel) shows the correlation matrix among the imaging systematics maps for the entire DESI footprint. 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. Additionally, we compute the Spearman correlation coefficients between the LRG density and imaging systematic maps to assess whether or not the correlations are impacted by outliers in the imaging data, but find no substantial differences from Pearson.

2.1.2 *Treatment of imaging systematics*

There are several approaches for handling imaging systematic errors, broadly classified into data-driven and simulation-based modeling approaches (see e.g. Ross et al. 2011; Ross et al. 2012, 2017;

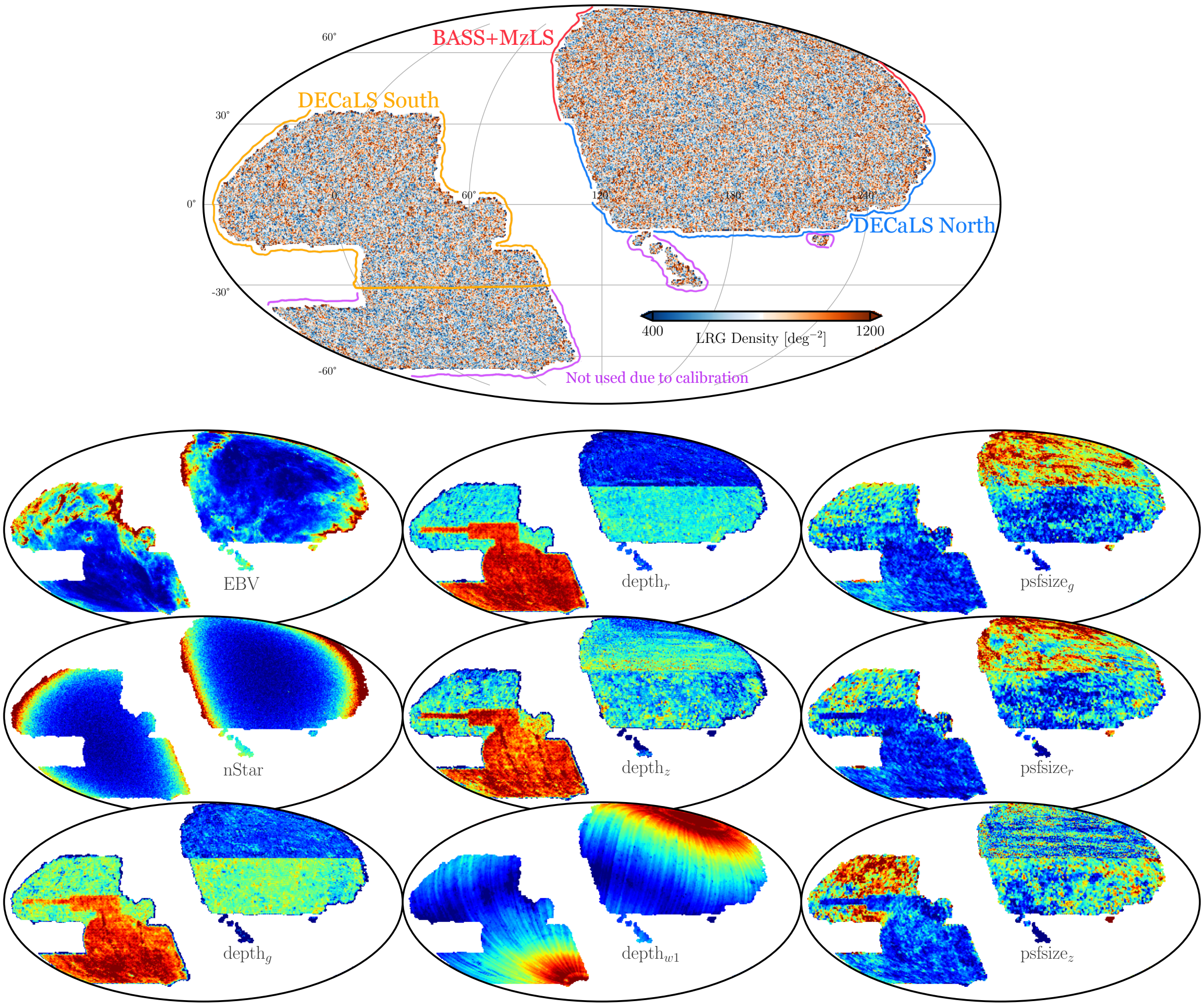

Figure 2. Top: The DESI LRG target density map before correcting for imaging systematic effects in Mollweide projection. The disconnected islands from the North footprint and parts of the South footprint with declination below -30 are removed from the sample for the analysis due to potential calibration 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.

Table 2. Statistics for DESI imaging data. Median depths are for galaxy/point sources detected at 5σ. Median psfsize values are computed with a depth-weighted average at each location on the sky.

	BASS+MzLS	DECaLS North	DECaLS South
Mean galaxy density [deg^{-2}]	804	808	796
Area [deg^2]	4525	5257	5188
Median extinction [mag]	0.02	0.03	0.05
Median stellar density [deg^{-2}]	667	629	629
Median g galaxy depth [mag]	24.0	24.4	24.5
Median r galaxy depth [mag]	23.4	23.8	23.9
Median z galaxy depth [mag]	23.0	22.9	23.1
Median $W1$ psf depth [mag]	21.6	21.4	21.4
Median g psfsize [arcsec]	1.9	1.5	1.5
Median r psfsize [arcsec]	1.7	1.4	1.3
Median z psfsize [arcsec]	1.2	1.3	1.3

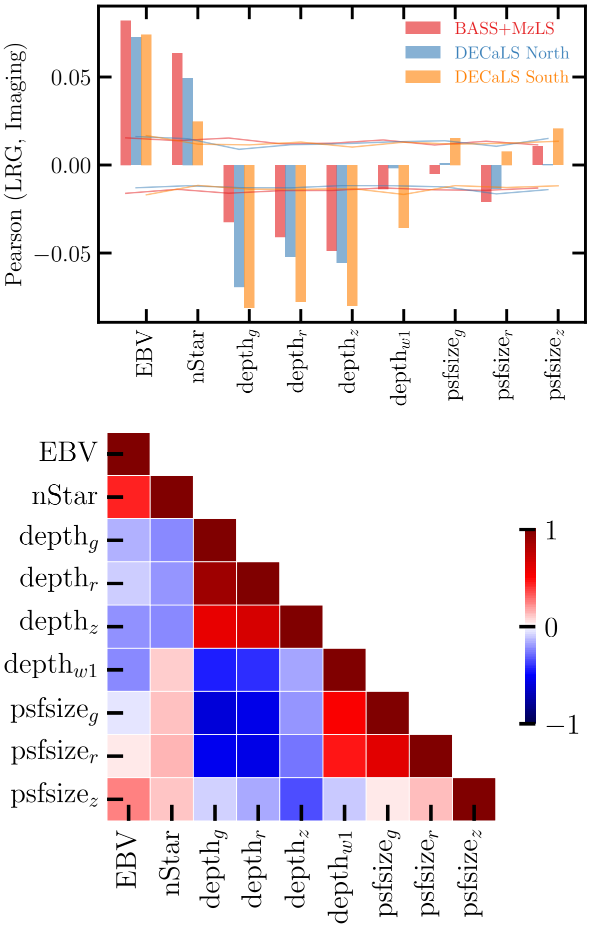

Figure 3. Top: The Pearson correlation coefficient between the DESI LRG target density and imaging properties in BASS+MzLS, DECaLS North, and DECaLS South. Solid horizontal curves represent the 95% confidence intervals estimated from simulations of lognormal density fields. Bottom: The Pearson correlation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

Ho et al. 2012; Suchyta et al. 2016; Delubac et al. 2016; Prakash et al. 2016; Raichoor et al. 2017; Laurent et al. 2017; Elvin-Poole et al. 2018; Bautista et al. 2018; Rezaie et al. 2020; Kong et al. 2020; Rezaie et al. 2021; Everett et al. 2022; Chaussidon et al. 2022; Eggert & Leistedt 2023). 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 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). 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.

Linear multivariate model: 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,

$$N_i = \log(1 + \exp[\mathbf{a}.\mathbf{x}_i + a_0]), \tag{3}$$

where a_0 is a global offset, and $\mathbf{a}.\mathbf{x}_i$ represents the inner product between the parameters, $\mathbf{a}$, and the values for imaging systematics in pixel i, $\mathbf{x}_i$. The functional form for N_i is adapted to enforce the predicted galaxy counts to be positive. Then, Markov Chain Monte Carlo (MCMC) ~~random~~ search is performed using the EMCEE package (Foreman-Mackey et al. 2013) to explore the parameter space by minimizing the negative Poisson log-likelihood between the actual and predicted number counts of galaxies.

~~No spatial coordinate is~~Spatial coordinates are not 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 for training the linear model. This aligns with the methodology used for training linear models in previous analyses (see, e.g., Zhou et al. 2022). The predicted galaxy counts are evaluated for each region using the marginalized mean estimates of the parameters, combined with those from other regions to cover the DESI footprint~~, and then normalized and inverted to represent the imaging weights into HEALPix of NSIDE = 256~~. The linear-based imaging weights are then defined as the inverse of the predicted target density, normalized to unity.

Neural network model: 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, $\mathbf{a}.\mathbf{x}_i$ in Equation 3 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 non-linear 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 model prediction 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 $\max(0, x)$ to introduce nonlinearities in the neural network (Nair & Hinton 2010). This simple form of nonlinearity is very effective in enabling deep neural networks to learn more complex, non-linear relationships between the input imaging maps and output galaxy counts.

Compared with linear regression, neural networks potentially are more prone to fitting noise, i.e., excellent performance on training data and poor performance on validation (or test) data. Therefore, our analysis uses a training-validation-testing split to avoid over-fitting and ensure that the neural network is well-optimized. 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 also test a geometrical split in which neighboring pixels belong to the same set of training, testing, or validation, but no significant performance difference is observed.

The neural networks are trained for up to 70 training epochs with the gradient descent ADAM optimizer (Loshchilov & Hutter 2017), which iteratively updates the neural network parameters 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 again, 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. 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). With the cross-validation technique, the model predictions from the different test sets are aggregated together to form the predicted target density map into the DESI footprint. 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.

2.2 Synthetic lognormal density fields

Density fluctuations of galaxies on large scales can be approximated with lognormal distributions (Coles & Jones 1991; Clerkin et al. 2017). Unlike N-body simulations, simulating lognormal density fields is not computationally intensive, and allows quick and robust validation of data analysis pipelines. Lognormal simulations are therefore considered efficient for our study since the signature of local PNG appears on large-scales and small-scale clustering is not used in our analysis. 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 $f_{\rm NL} = 0$ and 76.9 are considered. A set of 1000 realizations is produced for every $f_{\rm NL}$. The mocks are designed to match the clustering signal of the DESI LRG targets on scales insensitive to $f_{\rm NL}$. The analysis adapts the fiducial BOSS cosmology (BOSS Collaboration et al. 2017) which assumes a flat ΛCDM universe, including one massive neutrino with $m_\nu = 0.06$ eV, Hubble constant $h = 0.68$, 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 $8h^{-1}$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*

We employ a linear multivariate model to introduce synthetic spurious fluctuations in the lognormal density fields, and validate our imaging systematic mitigation methods. 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 less severe imaging systematic effects. The imaging systematic maps considered for the contamination model are extinction, depth in z, and psfsize in r. As shown in the Pearson correlation and will be discussed later in §3.4, the DESI LRG targets correlate strongly with these three maps. We fit for the parameters of the linear models with the MCMC process, executed separately on each imaging survey (BASS+MzLS, DECaLS North, and DECaLS South). Then, the imaging selection function for contaminating each simulation is uniquely determined by randomly drawing from the parameter space probed by MCMC, and then the results from each imaging survey are combined to form the DESI footprint. The clean density is then multiplied by the contamination model to induce systematics. The same contamination model is used for both the $f_{\rm NL} = 0$ and 76.9 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 $f_{\rm NL}$ biases due to over-correction. We briefly summarize two statistical tests based on the mean galaxy density contrast and the cross power spectrum between the galaxy density and the imaging systematic maps to assess the quality of the data and the significance of the remaining systematic effects (see, also, Rezaie et al. 2021). We calculate these statistics and compare the values to those measured from the clean mocks before looking at the auto power spectrum of the DESI LRG targets.

3 ANALYSIS TECHNIQUES

We address imaging systematics in DESI data by performing a separate treatment for each imaging region (e.g., DECaLS North) within the DESI footprint to reduce the impact of systematic effects specific to that region. Once the imaging regions have been mitigated for systematics, we combine the data from all regions to compute the power spectrum for the entire DESI footprint to increase the overall statistical power and enable more robust measurements of $f_{\rm NL}$. We then conduct robustness tests on the combined data to assess the significance of any remaining systematic effects.

3.1 Power spectrum estimator

We first construct the density contrast field from the LRG density, ρ,

$$\delta_g = \frac{\rho - \overline{\rho}}{\overline{\rho}}, \tag{4}$$

where the mean galaxy density $\overline{\rho}$ is estimated from the entire LRG sample. As a robustness test, we also analyze the power spectrum from each imaging region individually, in which $\overline{\rho}$ is calculated separately for each region. Then, we use the pseudo angular power spectrum estimator (Hivon et al. 2002),

$$\tilde{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2, \tag{5}$$

where the coefficients $a_{\ell m}$ are obtained by decomposing δ_g into spherical harmonics, $Y_{\ell m}$,

$$a_{\ell m} = \int d\Omega\, \delta_g W Y^*_{\ell m}, \tag{6}$$

where W represents the survey window that is described by the number of randoms normalized to the expected value.

We use the implementation of `anafast` from the HEALPix package (Gorski et al. 2005) to do fast harmonic transforms (Equation 6) and estimate the pseudo angular power spectrum of the LRG targets and the cross power spectrum between the LRG targets and the imaging systematic maps.

3.2 Modelling

The estimator in Equation 5 yields a biased power spectrum when the survey sky coverage is incomplete. Specifically, the survey mask causes correlations between different harmonic modes (Beutler et al. 2014; Wilson et al. 2017), and the measured clustering power is smoothed on scales near the survey size. An additional potential cause of systematic error arises from the fact that the mean galaxy density used to construct the density contrast field (Equation 4) 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). Since the clustering power on these scales are highly sensitive to $f_{\rm NL}$, it is crucial to account for these systematic effects in the model galaxy power spectrum to obtain unbiased $f_{\rm NL}$ constraints (see, also, Riquelme et al. 2022), which we describe below.

The other theoretical systematic issues are however subdominant in the angular power spectrum. For instance, relativistic effects generate PNG-like scale-dependent signatures on large scales, which interfere with measuring $f_{\rm NL}$ with the scale-dependent bias effect using higher order multipoles of the 3D power spectrum (Wang et al. 2020). Similarly, matter density fluctuations with wavelengths larger than survey size, known as super-sample modes, modulate the galaxy 3D 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).

3.2.1 *Angular power spectrum*

Santi: would nonlinear power spectrum matter? MR: Somewhere here we should mention that the nonlinear P(k) matter does not matter, as those scales are not used. Maybe provide a reference. I can also re-run using a nonlinear P(k). The relationship between the linear matter power spectrum $P(k)$ and the projected angular power spectrum of galaxies is expressed by the following equation:

$$C_\ell = \frac{2}{\pi}\int_0^\infty \frac{dk}{k} k^3 P(k)|\Delta_\ell(k)|^2 + N_{\rm shot}, \tag{7}$$

where $N_{\rm shot}$ is a scale-independent shot noise term. The projection kernel $\Delta_\ell(k) = \Delta^{\rm g}_\ell(k) + \Delta^{\rm RSD}_\ell(k)$ includes redshift space distortions and determines the contribution of each wavenumber k to the galaxy power spectrum on mode ℓ. For more details, refer to Padmanabhan et al. (2007). The FFTLog[2] 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 $l^{\rm th}$ order spherical Bessel functions, $j_\ell(kr)$, and its second derivatives,

$$\Delta^{\rm g}_\ell(k) = \int \frac{dr}{r} r(b+\Delta b)D(r)\frac{dN}{dr} j_\ell(kr), \tag{8}$$

$$\Delta^{\rm RSD}_\ell(k) = -\int \frac{dr}{r} r f(r) D(r)\frac{dN}{dr} j''_\ell(kr), \tag{9}$$

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 normalized to unity and described in terms of comoving distance[3] (solid curve in Figure 1). The PNG-induced scale-dependent shift is given by (Slosar et al. 2008)

$$\Delta b = b_\phi(z) f_{\rm NL}\frac{3\Omega_m H_0^2}{2k^2 T(k) D(z) c^2}\frac{g(\infty)}{g(0)}, \tag{10}$$

where Ω_m is the matter density, H_0 is the Hubble constant[4], $T(k)$ 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 Λ because of our normalization of D (see, e.g., Reid et al. 2010; Mueller et al. 2019). We assume the universality relation which directly relates b_ϕ to b via $b_\phi = 2\delta_c(b-p)$ with $\delta_c = 1.686$ representing the critical density for spherical collapse (Fillmore & Goldreich 1984). We fix $p=1$ in our analysis and marginalize over b (see, also, Slosar et al. 2008; Reid et al. 2010; Ross et al. 2013). ~~Assuming the universality relation, $b_\phi = 2\delta_c(b-p)$, where $p=1$ and $\delta_c = 1.686$ is the critical density for spherical collapse~~ (Fillmore & Goldreich 1984).~~Provided that the uncertainty on b_ϕ only impacts the error on $f_{\rm NL}$, and not the maximum likelihood estimate, we fix $p=1$ in our analysis for our sample of DESI LRG targets~~ (see, also, Slosar et al. 2008; Reid et al. 2010; Ross et al. 2013). ~~We do not marginalize over p to avoid parameter-space 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

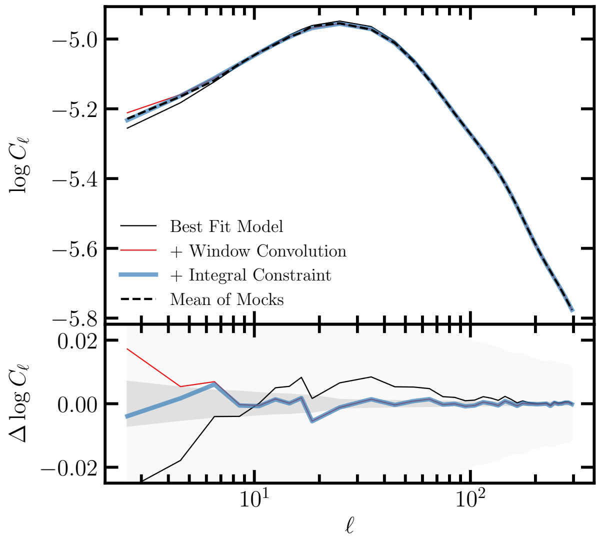

Figure 4. The mean power spectrum from the $f_{\rm NL}=0$ mocks (no contamination) and best-fitting theoretical prediction after accounting for the survey geometry and integral constraint effects. The dark and light shades represent the 68% error on the mean and one realization, respectively. Bottom panel shows the residual power spectrum relative to the mean power spectrum. No imaging systematic cleaning is applied to these mocks.

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

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

[4] $H_0 = 100\ ({\rm km\ s^{-1}})/(h^{-1}{\rm Mpc})$ and k is in unit of $h{\rm Mpc}^{-1}$

via a mode-mode coupling matrix, $\mathrm{M}_{\ell\ell'}$,

$$< \tilde{C}_\ell >= \sum_{\ell'} \mathrm{M}_{\ell\ell'} < C_{\ell'} > . \tag{11}$$

We convert this convolution in the spherical harmonic space into a multiplication in the correlation function space. Specifically, we first transform the theory power spectrum (Equation 7) to the correlation function, $\hat{\omega}^{\rm model}$. Then, we estimate the survey mask correlation function, $\hat{\omega}^{\rm window}$, and obtain the pseudo-power spectrum,

$$\tilde{C}_\ell^{\rm model} = 2\pi \int \hat{\omega}^{\rm model} \hat{\omega}^{\rm window} \, P_\ell(\cos\theta) d\cos\theta. \tag{12}$$

The integral constraint is another systematic effect which is induced since the mean galaxy density is estimated from the observed galaxy density, and therefore is biased by the limited sky coverage (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 as,

$$\tilde{C}_\ell^{\rm model,IC} = \tilde{C}_\ell^{\rm model} - \tilde{C}_{\ell=0}^{\rm model} \left(\frac{\tilde{C}_\ell^{\rm window}}{\tilde{C}_{\ell=0}^{\rm window}} \right), \tag{13}$$

where $\tilde{C}^{\rm window}$ is the survey mask power spectrum, i.e., the spherical harmonic transform of $\hat{\omega}^{\rm window}$.

The lognormal simulations are used to validate our survey window and integral constraint correction. Figure 4 shows the mean power spectrum of the $f_{\rm NL} = 0$ simulations (dashed) and the best-fitting 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, which covers around 40% of the sky, is applied to the simulations. We find that the survey window effect modulates the clustering power on $\ell < 200$ and the integral constraint alters the clustering power on $\ell < 6$.

3.3 Parameter estimation

Our parameter inference uses standard MCMC sampling. A constant clustering amplitude is assumed to determine the redshift evolution of the linear bias of our DESI LRG targets, $b(z) = b/D(z)$, which is supported by the HOD fits to the angular power spectrum (Zhou et al. 2021). In MCMC, we allow $f_{\rm NL}$, $N_{\rm shot}$, 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.

The expected large-scale power is highly sensitive to the value of $f_{\rm NL}$ such that the the amplitude of the covariance for C_ℓ is influenced by the true value of $f_{\rm NL}$, see also Ross et al. (2013) for a discussion. As illustrated in the top row of Figure 5, we find that the distribution of the power spectrum at the lowest bin, $2 \leq \ell < 4$, is highly asymmetric and its standard deviation varies significantly from the simulations with $f_{\rm NL} = 0$ to 76.9. We can make the covariance matrix less sensitive to $f_{\rm NL}$ by taking the log transformation of the power spectrum, $\log C_\ell$. As shown in the bottom panels in Figure 5, the log transformation reduces the asymmetry and the difference in the standard deviations between the $f_{\rm NL} = 0$ and 76.9 simulations. Therefore, we minimize the negative log likelihood defined as,

$$-2\log\mathcal{L} = (\log\tilde{C}(\Theta) - \log\tilde{C})^\dagger \mathbb{C}^{-1} (\log\tilde{C}(\Theta) - \log\tilde{C}), \tag{14}$$

where Θ represents a container for the parameters $f_{\rm NL}$, b, and $N_{\rm shot}$; $\tilde{C}(\Theta)$ is the (binned) expected pseudo-power spectrum; $\tilde{C}$ is the (binned) measured pseudo-power spectrum; and $\mathbb{C}$ is the covariance on $\log\tilde{C}$ constructed from the $f_{\rm NL} = 0$ log-normal simulations. Log-normal simulations have been commonly used and validated to estimate the covariance matrices for galaxy density fields, and non-linear effects are subdominant on the scales of interest to $f_{\rm NL}$ (see, e.g., Clerkin et al. 2017; Friedrich et al. 2021). We also test for the robustness of our results against an alternative covariance constructed from the $f_{\rm NL} = 76.9$ mocks. Flat priors are implemented for all parameters: $f_{\rm NL} \in [-1000, 1000]$, $N_{\rm shot} \in [-0.001, 0.001]$, and $b \in [0, 5]$.

3.4 Characterization of remaining systematics

One potential problem that can arise in the data-driven mitigation approach is *over-correction*, which occurs when the corrections applied to the data 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 its increased degrees of freedom. As illustrated in the bottom panel of Figure 3, the significant correlations among the imaging systematic maps may pose additional challenges for modeling the spurious fluctuations in the galaxy density field. Specifically, using highly correlated imaging systematic maps increases the statistical noise in the imaging weights, which elevates the potential for over subtracting the clustering power. These over-correction effects are estimated to have a negligible impact on baryon acoustic oscillations (Merz et al. 2021); however, they can significantly modulate the galaxy power spectrum on large scales, and thus lead to biased $f_{\rm NL}$ constraints (Rezaie et al. 2021; Mueller et al. 2022). Although not explored thoroughly, the over-correction 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 in the hope of ensuring that the constraints are as accurate and reliable as possible; one such approach is to reduce the dimensionality of the problem.

Our goal is to reduce the correlations between the DESI LRG target density and the imaging systematic maps while controlling the over-correction effect. Below, we describe how we achieve this objective, by employing a series of simulations along with the residual systematics that we construct based on the cross power spectrum between the LRG density and imaging maps, and the mean LRG density as a function of imaging. We test different sets of the imaging systematic maps to identify the optimal set of the feature maps:

(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 photometric 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.

It is imperative to note that these sets are selected prior to examining the auto power spectrum of the LRG sample and unblinding the $f_{\rm NL}$ constraints, and that the auto power spectrum and $f_{\rm NL}$ measurements are unblinded only after our mitigation methods passed our rigorous tests for residual systematics.

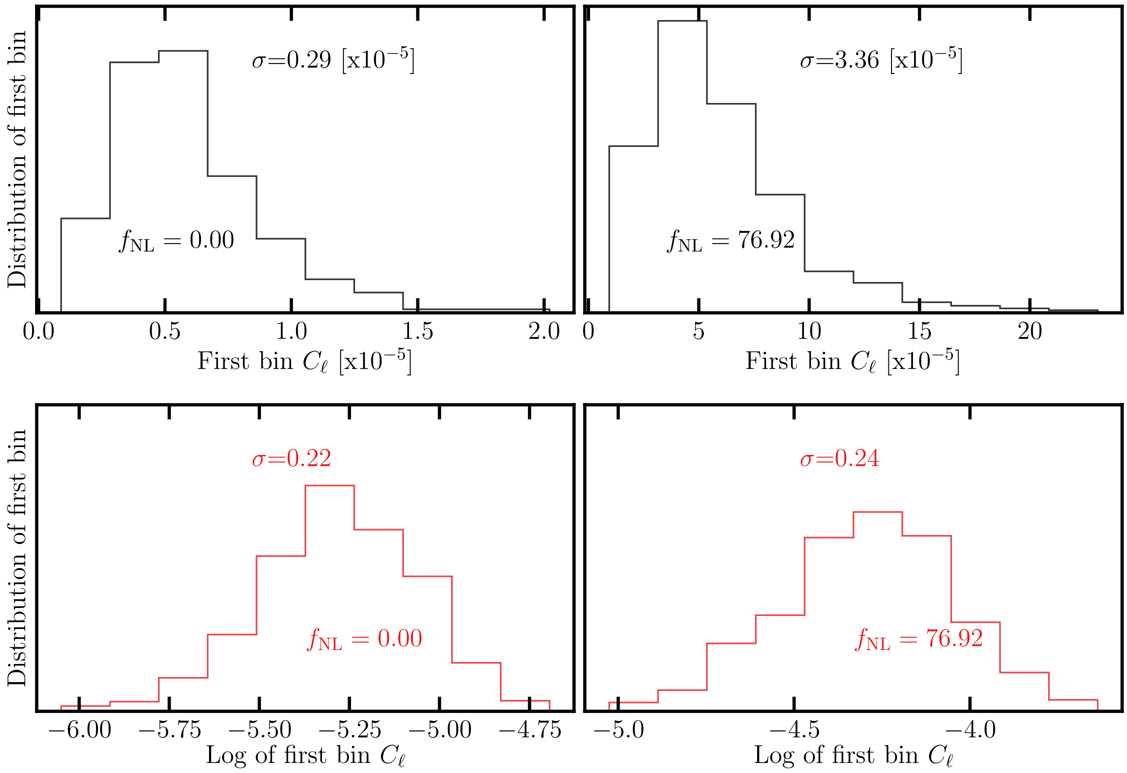

Figure 5. The distribution of the first bin power spectra and its log transformation from the simulations with $f_{\rm NL} = 0$ (left) and 76.9 (right). The log transformation ~~alleviates~~largely removes the asymmetry in the distributions.

3.4.1 *Cross power spectrum*

We characterize the cross correlations between the galaxy density and imaging systematic maps by

$$\tilde{C}_{X,\ell} = [\tilde{C}_{x_1,\ell}, \tilde{C}_{x_2,\ell}, \tilde{C}_{x_3,\ell}, ..., \tilde{C}_{x_9,\ell}], \tag{15}$$

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

$$\tilde{C}_{x_i,\ell} = \frac{(\tilde{C}_{gx_i,\ell})^2}{\tilde{C}_{x_i x_i,\ell}}. \tag{16}$$

With this normalization, $\tilde{C}_{x_i,\ell}$ estimates the contribution of systematics at every multipole up to the ~~first~~linear order to the galaxy power spectrum. Then, the χ^2 value for the cross power spectra is calculated via,

$$\chi^2 = \tilde{C}^T_{X,\ell}\mathbb{C}^{-1}_X\tilde{C}_{X,\ell}, \tag{17}$$

where the covariance matrix $\mathbb{C}_X =< \tilde{C}_{X,\ell}\tilde{C}_{X,\ell'} >$ is constructed from the lognormal mocks. These χ^2 values are measured for every clean mock realization with the *leave-one-out* technique and compared to the values observed in the LRG sample with various imaging systematic corrections. Specifically, we use 999 realizations to estimate a covariance matrix and then apply the covariance matrix from the 999 realizations to measure the χ^2 for the one remaining realization. This process is repeated for all 1000 realizations to construct a histogram for χ^2. We only include the bandpower bins from $\ell = 2$ to 20 with $\Delta\ell = 2$, and test for the robustness with higher ℓ modes in Appendix A1.

We identify extinction and depth in the z band as two primary ~~maps~~potential contaminants, and run the linear model with these two maps to derive the systematic weights. Linear two maps is the most conservative systematic treatment method in terms of both the model flexibility and the number of input maps. We clean the sample using the imaging weights obtained from *linear two maps*, and find that the linear two maps approach mitigates most of the spurious density fluctuations 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 psfsize 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 in the DESI LRG targets. Then, we adapt *neural network three maps* to model non-linear systematic effects. Compared with the linear three maps method, we find that the neural network-based weights significantly reduce the cross correlations and spurious density fluctuations of the DESI LRG sample and imaging systematic maps. Additionally, we consider neural network with four, five, and nine maps to further examine the robustness of our cleaning methods. Figure 6 shows $\tilde{C}_X$ from the DESI LRG targets 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. Without imaging weights, the LRG sample has the highest cross-correlations against extinction, stellar density, and depth in z (solid black curve). There are 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

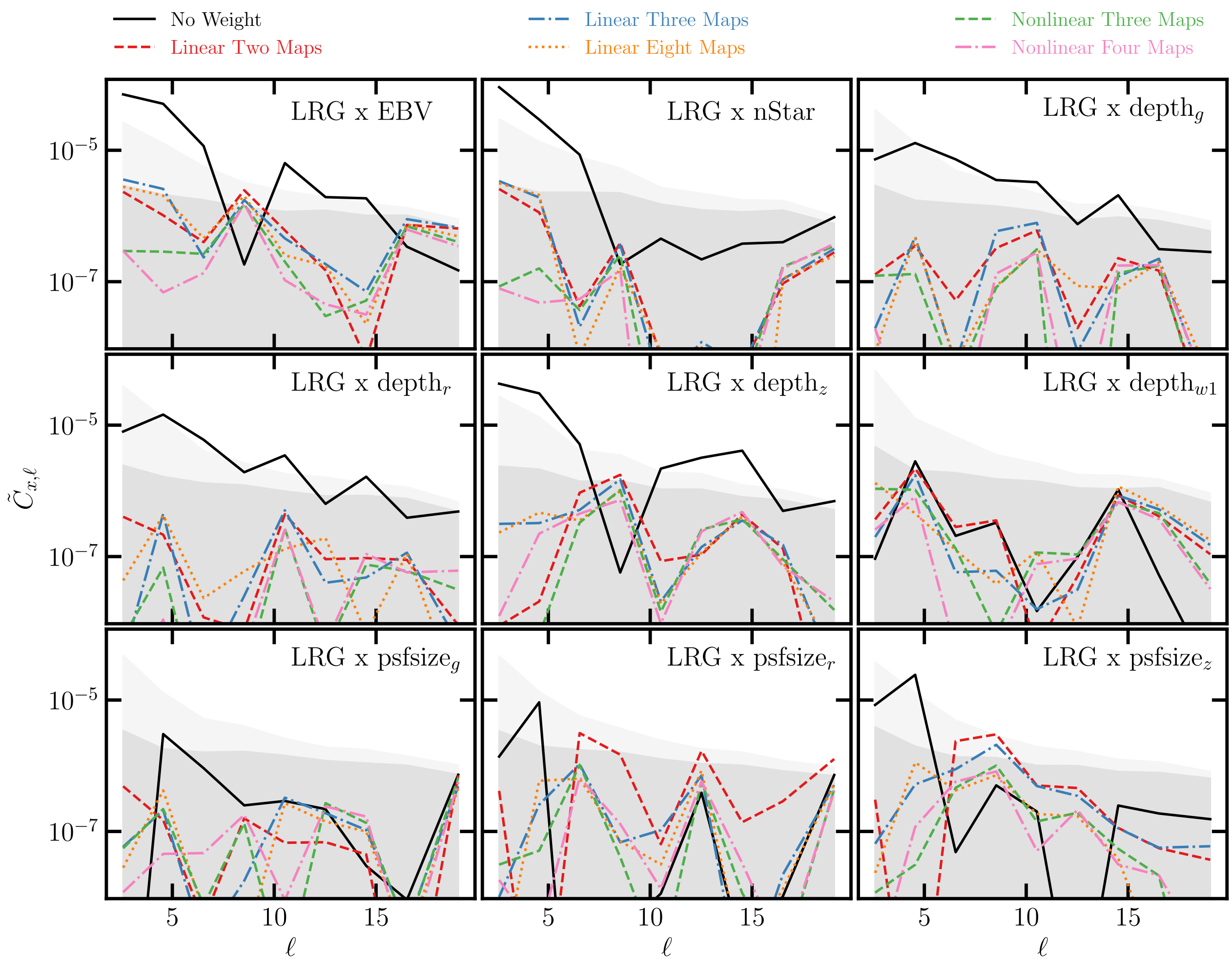

Figure 6. The square of the cross power spectra between the DESI LRG targets and imaging systematic maps normalized by the auto power spectrum of the imaging systematic maps; see equation 16. The systematic maps considered are Galactic extinction (EBV), stellar density (nStar), depth in *grzw1* (depth$_{grzw1}$), 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 $f_{\rm NL}$ = 0 and 76.9 lognormal density fields, respectively.

linear model using two maps (extinction and depth in z) as identified from the Pearson correlation. With linear two maps (red dashed 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. Regression against extinction and depth in the z-band helps mitigate large-scale cross correlations ($\ell < 6$)~~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 (blue dot-dashed curve) approach alleviates the cross power spectrum against psfsize in r. ~~On the other hand~~Additionally, nonlinear three maps (green dashed curve) can reduce the cross correlations against both the r and z-band psfsize maps, which ~~indicates~~shows 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 (orange dotted curve) and nonlinear four maps (pink dot-dashed curve).

3.4.2 *Mean density contrast*

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

$$\delta_{x_j} = (\overline{\rho})^{-1} \frac{\sum_i \rho_i W_i}{\sum_i W_i}, \tag{18}$$

where $\overline{\rho}$ is the global mean galaxy density, W_i is the survey window in pixel i, and the summations over i are evaluated from the pixels in every bin of x_j. We compute the histograms against all nine imaging properties (see Figure 2). We use a set of eight equal-width bins for every imaging map, which results in a total of 72 bins. Then, we construct the total mean density contract as,

$$\delta_X = [\delta_{x_1}, \delta_{x_2}, \delta_{x_3}, ..., \delta_{x_9}], \tag{19}$$

and the total residual error as,

$$\chi^2 = \delta_X^T \mathbb{C}_\delta^{-1} \delta_X, \tag{20}$$

where the covariance matrix $\mathbb{C}_\delta =< \delta_X \delta_X >$ is constructed from the lognormal mocks. Figure 7 shows the mean density contrast

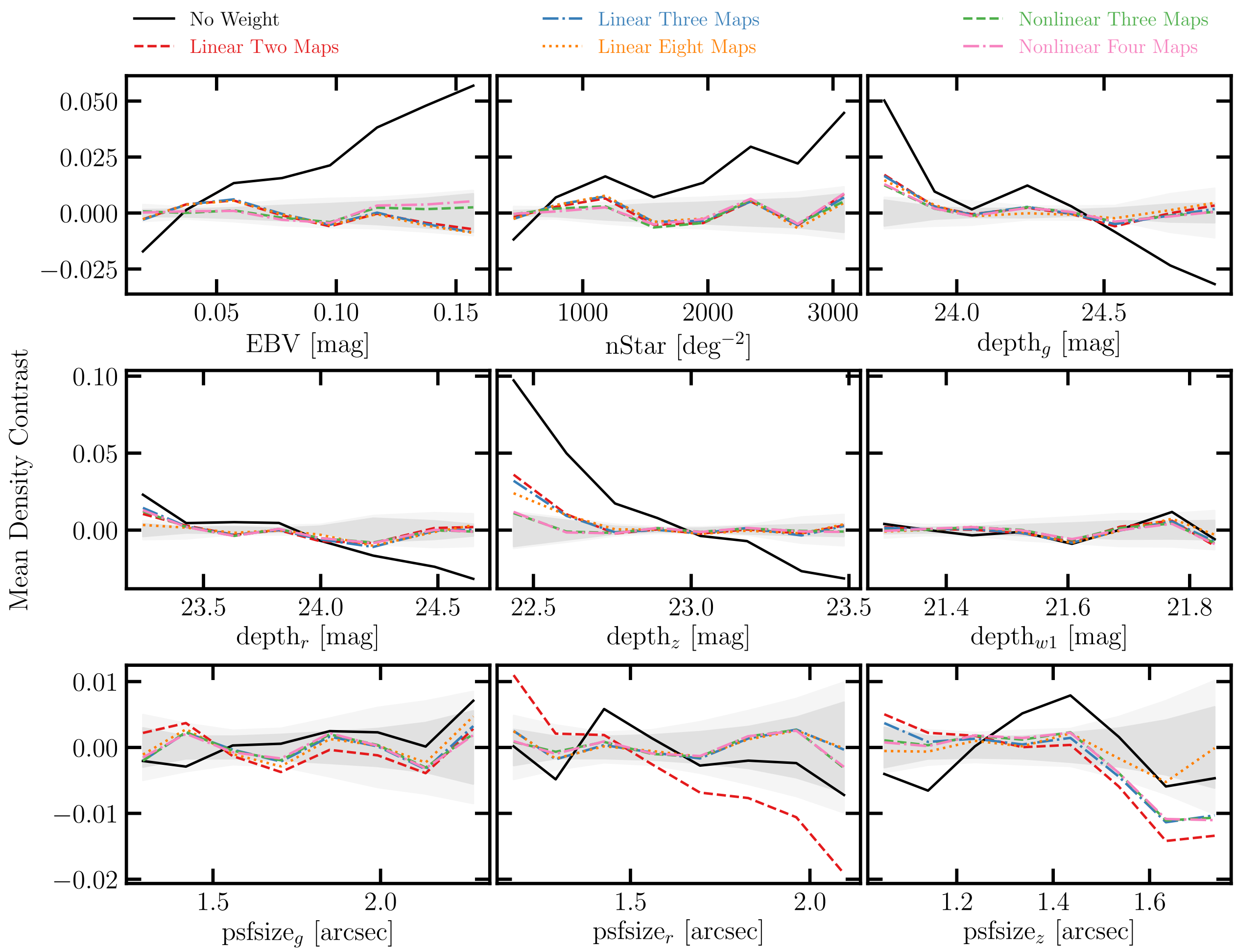

Figure 7. The mean density contrast of the DESI LRG targets as a function of the imaging systematic maps: Galactic extinction (EBV), stellar density (nStar), depth in *grzw1* (depth$_{grzw1}$), and seeing in *grz* (psfsize$_{grz}$). The black curves display the results before imaging systematic correction. The red, blue and orange curves represent the relationships after applying the imaging weights from the linear models trained with *two maps*, *three maps*, and *eight maps*, respectively. 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 $f_{\rm NL}$ = 0 and 76.9, respectively.

against the imaging properties for the DESI LRG targets. The dark and light shades represent the 1σ level fluctuations observed in 1000 lognormal density fields respectively with $f_{\rm NL}$ = 0 and 76.9. The DESI LRG targets 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 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 (red dashed 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 depth$_z$ and around 1% in the high end of psfsize$_z$~~, in the z band~~, which implies the presence of 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~~are capable of reducing the fluctuations below 2%. Even with the nonlinear three maps, we have about 1% remaining systematic fluctuations against the z-band psfsize. We use the χ^2 statistics to assess how significant these fluctuations are in comparison to the clean mocks.

DH: nothing incorrect here (apart from the comment just above), but the text just seems tedious and maybe repetitive, with lots of numbers thrown around. It would be good to give it a more streamlined rewrite. Lines 798-808 are particularly in need of a re-form. Figure 8 shows the χ^2 histograms from the normalized cross spectrum (top) and mean density contrast (bottom) statistics which are obtained from the $f_{\rm NL}$ = 0 and 76.9 lognormal mocks, respectively with solid and dashed lines. The χ^2 tests ~~prove~~turn out to be insensitive to the true $f_{\rm NL}$ of the mocks, as we obtain consistent distributions for both statistics, regardless of the true $f_{\rm NL}$ used for the mock generation. This provides a robust metric to separate the effect of the systematics from the effect of cosmology. No mitigation is applied to these mocks, and thus the χ^2 values are expected to be unbiased. The χ^2 values observed in the DESI LRG targets are shown via the vertical lines for comparison, and summarized in Table 3. The corresponding p-values are inferred from the comparison to the clean $f_{\rm NL}$ = 0 mocks. Before cleaning, our LRG sample has a cross power spectrum's χ^2 error of 20014.8. After correction with the linear two maps approach, the cross power spectrum's χ^2 is reduced

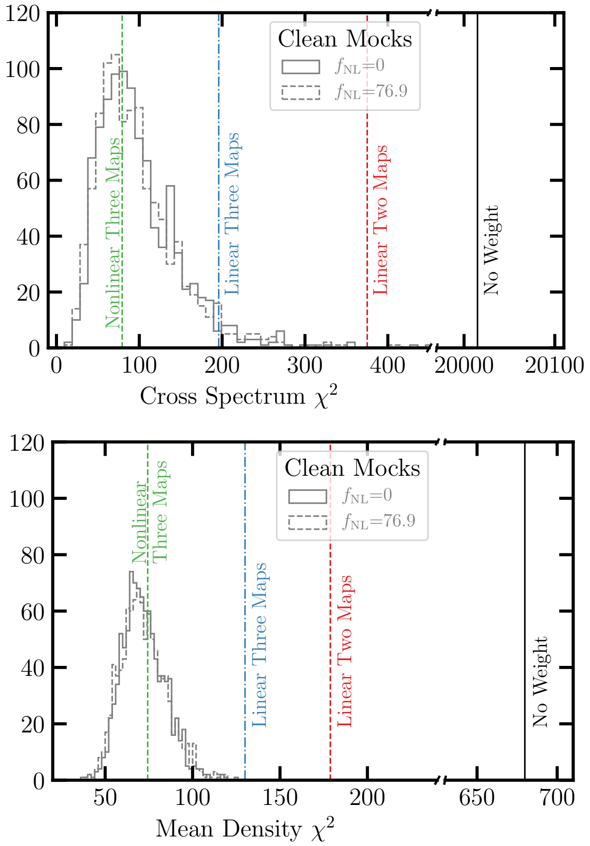

Figure 8. The remaining systematic error χ^2 from the galaxy-imaging cross power spectrum (top) and the mean galaxy density contrast (bottom). The values observed in the DESI LRG targets before and after linear and non-linear treatments are represented via vertical lines, and the histograms are constructed from 1000 realizations of clean lognormal mocks with $f_{\rm NL}$ = 0 and 76.9.

to 375.1 with p-value < 0.001. Adding the r-band psfsize, the linear three maps approach reduces the χ^2 down to 195.9 with p-value = 0.04; however, we can reject the null hypothesis that our sample with the linear three maps is properly cleaned at 95% confidence. Even though cleaning with linear all maps gives the lowest cross power spectrum's χ^2 of 129.2 (and p-value = 0.24), it potentially makes the analysis more prone to regressing out the true clustering signal, given the inner correlations 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's χ^2 is reduced to 79.3 with p-value = 0.59. Adding the stellar density map reduces the cross power spectrum's χ^2 error to 70.9 (p-value = 0.69). 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 A1).

Figure 8 (bottom) shows the mean galaxy density's χ^2 observed in the mocks with or without $f_{\rm NL}$. Similar ~~to~~as in the cross power spectrum test, we find consistent results for the mean density diagnostic regardless of the underlying true $f_{\rm NL}$, which supports the finding that our diagnostic is not sensitive to the fiducial cosmology. The values measured in the DESI LRG targets before and after applying imaging weights are shown via the vertical lines for comparison, and summarized in Table 3. The *linear two maps* weights reduce the χ^2 value from 679.8 (before correction) to 178.8. The p-value < 0.001 indicates severe remaining systematic effects. Adding the r-band psfsize reduces the error to χ^2 = 130 but the remaining systematics remain statistically significant with p-value < 0.001. Training the linear model with all imaging systematic maps returns a more reasonable χ^2 = 90 and p-value of 0.08. However, as noted earlier, regression with all imaging systematic maps as input can lead to the removal of the true clustering signal. On the other hand, we obtain a χ^2 value of 74.3 with p-value = 0.39 with the imaging weights from the *nonlinear three maps* approach. Re-training the nonlinear approach while adding the stellar density map (*nonlinear four maps*) yields minor improvement: χ^2 = 73.2 and p-value = 0.42. The minor impact on χ^2 indicates that the stellar density trend in the mean LRG density can be explained via extinction because of the strong correlation between these two properties, such that in regions with high stellar density, there is likely to be a higher concentration of dust, which can cause greater extinction of light.

Table 3. Mean density and cross power spectrum χ^2 and p-values that are inferred from the comparison to the $f_{\rm NL}$ = 0 clean mocks.

Method	Mean Density		Cross Spectrum	
	χ^2	p-value	χ^2	p-value
No Weight	679.8	< 0.001	20014.8	< 0.001
Linear Two Maps	178.8	< 0.001	375.1	< 0.001
Linear Three Maps	130	< 0.001	195.9	0.04
Linear Eight Maps	90	0.08	129.2	0.24
Nonlinear Three Maps	74.3	0.39	79.3	0.59
Nonlinear Four Maps	73.2	0.42	70.9	0.69

The mean density contrast and cross power spectrum tests presented here show the effectiveness of the different cleaning approaches for the LRG sample before measuring the galaxy power spectrum and using it to constrain $f_{\rm NL}$. In other words, none of these tests require unblinding the measured power spectrum or $f_{\rm NL}$ constraints. As summarized in Table 3, the results of these tests reveal that cleaning the LRG sample with nonlinear three maps produces statistically consistent χ^2 values with the clean mocks, regardless of the true $f_{\rm NL}$ used in the generation of the mocks. HJ: there are no clear indications for the need to add the fourth map. Moreover, we should not select the model with four maps, which includes stellar density, as our best model to avoid the confirmation bias, since we do not have other indicators showing this model is more reasonable than the one with three maps. On the other hand, considering p = 0.05 as our threshold for clean maps, the linear three maps approach is not able to thoroughly mitigate systematics, as characterised by p-value < 0.001 and = 0.04, respectively, for the mean density and cross power spectrum diagnostics. Based on these findings, we identify nonlinear three maps as the optimal cleaning approach for the LRG sample and the rest of this manuscript, because including any additional imaging systematics maps could potentially further complicate the over-correction issue.

3.5 Calibration of over-correction ~~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 used for the regression. We calibrate the over-correction effect using the mocks presented in §2. We apply the neural network model to both the $f_{\rm NL}$ = 0 and 76.9 simulations, with and without imaging systematics, using various sets of imaging systematic maps. Specifically, we consider *nonlin-*

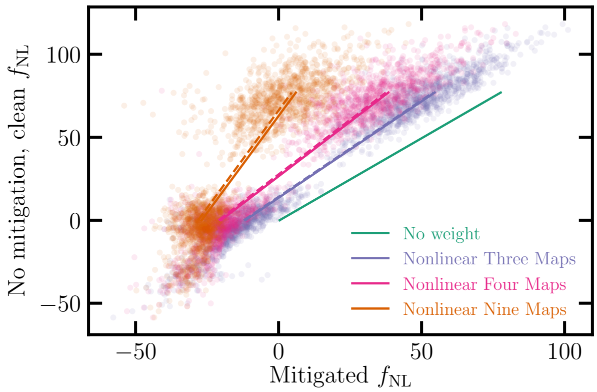

Figure 9. The *No mitigated, clean* vs *mitigated* $f_{\rm NL}$ values from the $f_{\rm NL} = 0$ and 76.9 mocks. The solid (dashed) lines represent the best-fitting estimates from fitting the mean power spectrum of the clean (contaminated) mocks. The scatter points show the best-fitting estimates from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

Table 4. Linear parameters employed to de-bias the $f_{\rm NL}$ constraints to account for the over-correction issue.Ashley: You could also add the m1/m2 to the table for the clean mock case?

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

ear three maps, *nonlinear four maps*, and *nonlinear nine maps*. Then, we measure the power spectra from the mocks. We fit both the mean power spectrum and each individual power spectrum from the mocks. Figure 9 displays a comparison between the estimates of $f_{\rm NL}$ before and after mitigation for the clean mocks. The best-fitting estimates are represented by the solid curves, and the individual spectra results are displayed as the scatter points. The results from fitting the mean power spectrum of the contaminated mocks are also shown via the dashed curves. We find nearly identical results for the biases caused by mitigation, whether or not the mocks have any contamination, which can be seen by observing the solid and dashed curves displayed on Figure 9 (see, also, Figure B4, for a comparison of the mean power spectrum). For clarity, the best-fitting estimates for the individual contaminated data are not shown in the figure.

To calibrate our methods, we fit a linear curve to the $f_{\rm NL}$ estimates from the mean power spectrum of the mocks, $f_{\rm NL,no\ mitigation,clean} = m_1 f_{\rm NL,mitigated} + m_2$. The m_1 and m_2 coefficients for nonlinear three, four, and nine maps are summarized in Table 4. These coefficients represent the impact of the cleaning methods on the likelihood. ~~We find that~~The uncertainty in $f_{\rm NL}$ after mitigation increases by $m_1 - 1$. Figure 9 also shows that the choice of our cleaning method can have significant implications for the accuracy of the measured $f_{\rm NL}$, and careful consideration should be given to the selection of the primary imaging systematic maps and the calibration of the cleaning algorithms in order to minimize systematic uncertainties.

4 RESULTS

~~This section presents~~We now present our $f_{\rm NL}$ constraints obtained from the power spectrum of the DESI LRG targets. The treatment of the imaging systematic effects is performed on each imaging region (BASS+MzLS, DECaLS North/South) separately. After cleaning, the regions are combined for the measurement of the power spectrum. We unblind the galaxy power spectrum and the $f_{\rm NL}$ values after our cleaning methods are validated and vetted by the cross power spectrum and mean galaxy density diagnostics. We also conduct additional tests to check the robustness of our constraints against various assumptions, such as analyzing each region separately, applying cuts on imaging conditions, and changing the smallest mode used in fitting for $f_{\rm NL}$.

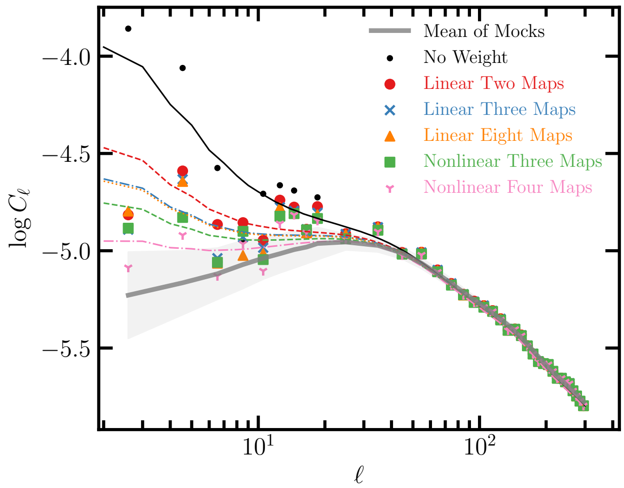

Figure 10. DH: what are the curves? Are they some fits through the points? Also the whole figure seems fuzzy - seems some kind of png file type issue, would be good to make it sharper. The angular power spectrum of the DESI LRG targets before (*No weight*) and after correcting for imaging systematics using the linear and non-linear methods. The curves represent their corresponding best-fitting theory predictions. The solid curve and grey shade respectively represent the mean power spectrum and 68% error from the $f_{\rm NL} = 0$ mocks.

4.1 DESI imaging LRG sample

We find that the excess clustering signal in the power spectrum of theDESI LRG targets is mitigated after correcting for the imaging systematic effects. Figure 10 shows the measured power spectrum of the DESI LRG targets before and after applying imaging weights and the best-fitting theory curves. The solid line and the grey shade represent respectively the mean power spectrum and 1σ error, estimated from the $f_{\rm NL} = 0$ 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 *non-linear three maps* to *linear three maps*, we find that the measured spectra on $4 \leq \ell < 6$ are very different, probably indicating some non-linear spurious fluctuations with large scale characteristics due to extinction. Adding stellar density in the non-linear approach (*non-linear* *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 the over-subtraction due to stellar density is reversed after accounting for over-correction.

4.1.1 *Calibrated constraints*

All $f_{\rm NL}$ constraints presented here are calibrated for the effect of over-correction using the lognormal simulations. Table 5 describes the best-fitting and marginalized mean estimates of $f_{\rm NL}$ from fitting the power spectrum of the DESI LRG targets before and after cleaning with the non-linear approach given various combinations for the imaging systematic maps. Figure 11 shows the marginalized probability distribution for $f_{\rm NL}$ in the top panel, and the 68% and 95% probability contours for the linear bias parameter and $f_{\rm NL}$ 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 $36(25) < f_{\rm NL} < 61(76)$ at 68%(95%) confidence with $\chi^2 = 34.6$ for *non-linear three maps* ~~over~~with 34 degrees of freedom. Accounted for over-correction, we obtain $37(25) < f_{\rm NL} < 63(78)$ with $\chi^2 = 35.2$ with the additional stellar density map in the *non-linear four maps*. With or without ~~*nStar*~~stellar density, the confidence intervals are consistent with each other and significantly off from zero PNG; specifically, the probability that $f_{\rm NL}$ is greater than zero, $P(f_{\rm NL} > 0) = 99.9$ per cent. We also apply a more aggressive systematics treatment that includes regression using the nonlinear approach against the full set of imaging maps we identified, nonlinear nine maps, and find that our maximum likelihood value changes only slightly to $f_{\rm NL} \sim 50$, but the uncertainty on $f_{\rm NL}$ increases due to the aggressive treatment removing large-scale clustering information. For comparison, we obtain $98(84) < f_{\rm NL} < 133(152)$ at 68%(95%) confidence with $\chi^2 = 44.4$ for the *no weight* case.

4.1.2 *Uncalibrated constraints: robustness tests*

Figure 12 shows the probability distributions of $f_{\rm NL}$ 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 aggressively and return biased $f_{\rm NL}$ constraints. As expected, non-linear nine maps yields the smallest maximum likelihood estimate of $f_{\rm NL} = -6$. Our non-linear three maps returns a best-fitting estimate of $f_{\rm NL} = 29$ with the 68%(95%) confidence of $19(9) < f_{\rm NL} < 41(53)$ and $\chi^2 = 34.6$. With the stellar density map included, non-linear four maps yields a smaller best-fitting estimates of $f_{\rm NL} = 17$ with the error of $7(-2) < f_{\rm NL} < 27(38)$. The non-linear nine maps gives an asymetric posterior with the marginalized mean $f_{\rm NL} = -9$, best estimate $f_{\rm NL} = -6$ with the error of $-21(-34) < f_{\rm NL} < 2(12)$.

Now we proceed to perform some robustness tests and assess how sensitive the $f_{\rm NL}$ 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 are only interested in the relative shift in the $f_{\rm NL}$ constraints after changing the assumptions. Therefore, the absolute scaling of the $f_{\rm NL}$ constraints presented here are biased because of the over-correction effect. Table 6 summarizes the uncalibrated $f_{\rm NL}$ constraints from the DESI LRG targets. Our tests are as follows:

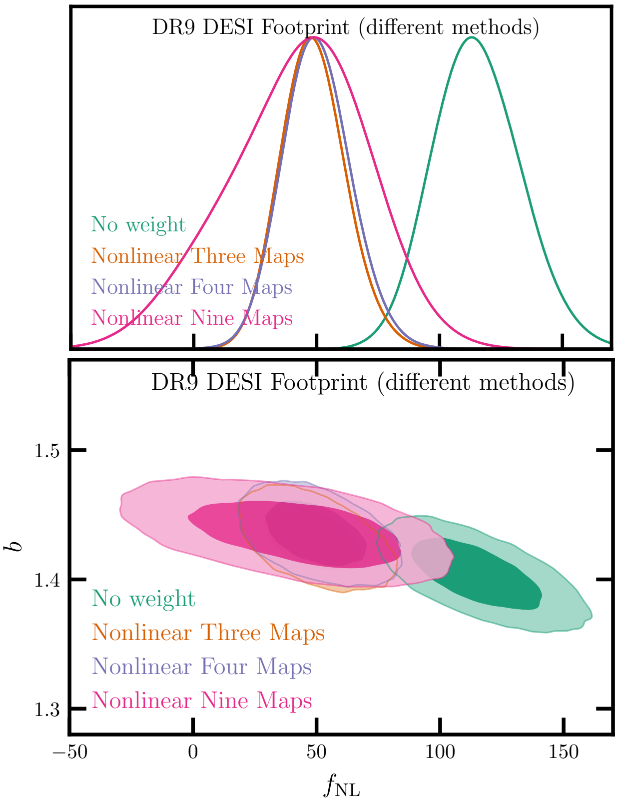

Figure 11. The calibrated constrains from the DESI LRG targets. *Top*: probability distribution for $f_{\rm NL}$ marginalized over the shotnoise and bias. *Bottom*: 68% and 95% probability distribution contours for the bias and $f_{\rm NL}$ from the DESI LRG targets before and after applying non-linear cleaning methods. The lognormal mocks are used to calibrate these distributions for over-correction.

- **Linear methods**: Even though the linear three maps approach shows significant remaining systematics (Figure 8), ~~but~~ we obtain identical constraints from *linear eight maps* and *linear three maps*, respectively, $26(16) < f_{\rm NL} < 49(62)$ and $26(16) < f_{\rm NL} < 50(63)$ at 68%(95%) confidence.
- **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 $f_{\rm NL}$ and b from each individual region, compared with that from DESI. The cleaning method here is *non-linear three maps*, and the covariance matrices are estimated from the $f_{\rm NL} = 0$ mocks. The bias in DECaLS North is lower than the ones from DECaLS South and BASS+MzLS, which might indicate some remaining systematic effects that could not be mitigated with the available imaging systematic maps. This is because given the negative correlation between b and $f_{\rm NL}$, a larger value of $f_{\rm NL}$ due to excess clustering power needs to be compensated by a smaller value of b. Overall, we find that the constraints from analyzing each imaging survey separately

Table 5. The calibrated best-fitting, marginalized mean, and marginalized 68% (95%) confidence estimates for $f_{\rm NL}$ from fitting the power spectrum of the DESI LRG targets before and after correcting for imaging systematic effects. The lowest mode is $\ell_{\rm min} = 2$.

		$f_{\rm NL}$				
Footprint	Method	Best fit	Mean	68% CL	95% CL	χ^2/dof
DESI	No Weight	113	115	$98 < f_{\rm NL} < 133$	$84 < f_{\rm NL} < 152$	44.4/34
DESI	Nonlinear Three Maps	47	49	$36 < f_{\rm NL} < 61$	$25 < f_{\rm NL} < 76$	34.6/34
DESI	Nonlinear Four Maps	49	50	$37 < f_{\rm NL} < 63$	$25 < f_{\rm NL} < 78$	35.2/34
DESI	Nonlinear Nine Maps	50	42	$13 < f_{\rm NL} < 69$	$-16 < f_{\rm NL} < 92$	39.5/34

are consistent with each other and DESI within 68% confidence. Ignoring the over-correction effect, we find that the results from the DECaLS North region to be the only one that finds PNG nonzero at greater than 95%, which motivates follow-up studies with the spectroscopic sample of LRGs in DECaLS North.

• **Stellar density template (*nStar*)**: When not accounting for ~~mitigation bias~~over-correction, adding the stellar density map appears to result in significant changes in the $f_{\rm NL}$ constraints, e.g., compare non-linear three maps with non-linear four maps in Table 6. But these changes disappear when we account for the mitigation bias and we find all methods recover the same maximum likelihood estimate for $f_{\rm NL} \sim 50$ within 69% confidence, see Table 5, which implies that these changes can be associated with 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 $f_{\rm NL}$. This pixel completeness cut removes 0.6% of the survey area, and no changes in the $f_{\rm NL}$ 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-fitting $f_{\rm NL}$ from fitting the DESI power spectrum. However, when each region is fit individually, the BASS+MzLS constraint shift toward higher values of $f_{\rm NL}$ by approximately $\Delta f_{\rm NL} \sim 10$, 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 *non-linear three maps* correction, but use the covariance matrix constructed from the $f_{\rm NL} = 76.9$ mocks. With the alternative covariance, a 12% increase in the $\sigma(f_{\rm NL})$ is observed. We also find that the best-fitting and marginalized mean estimates of $f_{\rm NL}$ 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-fitting $f_{\rm NL}$ 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-fitting and marginalized $f_{\rm NL}$ mean estimates from DECaLS South shift significantly to higher values of $f_{\rm NL}$ by $\Delta f_{\rm NL} \sim 10$, which supports ~~the issue of~~the case that there are remaining 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 do not induce significant contaminations.

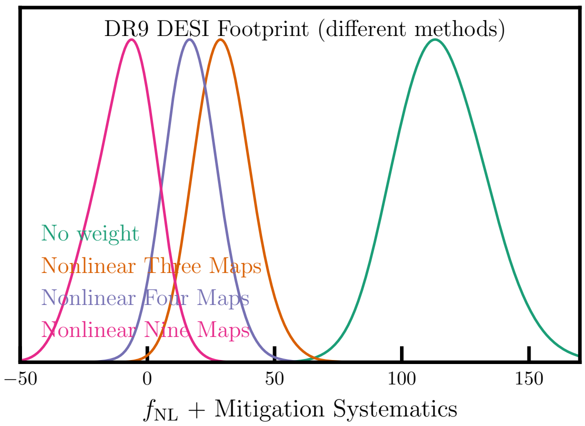

Figure 12. Same as Figure 11 but without accouting for over-correction.

• **Scale dependence (*varying* $\ell_{\rm min}$)**: We raise the value of the lowest harmonic mode $\ell_{\rm min}$ used for the likelihood evaluation during MCMC. This is equivalent to ~~decreasing the highest scale of measurement in the power spectrum~~utilizing smaller spatial scales in the measurements of the power spectrum. By doing so, we anticipate a reduction in the impact of imaging systematics on $f_{\rm NL}$ inference as lower ℓ modes are more likely to be contaminated. Figure 14 illustrates the power spectra before and after the correction with *non-linear three maps* in the top panel. The bottom panel shows the marginalized mean and 68% error on $f_{\rm NL}$ with *non-linear three maps* for the DESI, BASS+MzLS, DECaLS North, and DECaLS South regions. We discover that a slight upward shift in the mean estimates of $f_{\rm NL}$ on scales ranging from 12 to 18 for DECaLS North and BASS+MzLS when we utilized a higher $\ell_{\rm min}$. This outcome might imply that the imaging systematic maps do not contain enough information to help the cleaning method null out the contaminating signal in the NGC. We also find that the bump is resilient against an alternative correction, in which we apply the neural networks trained on the DECaLS South to the DECaLS North region (see A3). Overall, this result is contrary to what one might predict if a significant systematic-induced spike existed at the very low ℓ. As a result, it suggests that the underlying issue is more subtle than originally anticipated.

Table 6. The uncalibrated best-fitting and marginalized mean estimates for $f_{\rm NL}$ from fitting the power spectrum of the DESI LRG targets 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 $f_{\rm NL} = 0$ clean mocks (no mitigation) except for the case with '+cov' in which the covariance matrix is from the $f_{\rm NL} = 76.9$ clean mocks (no mitigation).

		$f_{\rm NL}$ + Mitigation Systematics				
Footprint	Method	Best fit	Mean	68% CL	95% CL	χ^2
DESI	**No Weight**	**113**	**115**	$\mathbf{98 < f_{\rm NL} < 133}$	$\mathbf{84 < f_{\rm NL} < 152}$	**44.4**
DESI	Linear Eight Maps	36	38	$26 < f_{\rm NL} < 49$	$16 < f_{\rm NL} < 62$	41.1
DESI	Linear Two Maps	50	51	$38 < f_{\rm NL} < 64$	$27 < f_{\rm NL} < 79$	38.8
DESI	**Linear Three Maps**	**37**	**38**	$\mathbf{26 < f_{\rm NL} < 50}$	$\mathbf{16 < f_{\rm NL} < 63}$	**39.6**
DESI	**Nonlinear Three Maps**	**29**	**30**	$\mathbf{19 < f_{\rm NL} < 41}$	$\mathbf{9 < f_{\rm NL} < 53}$	**34.6**
DESI (imag. cut)	Nonlinear Three Maps	29	31	$19 < f_{\rm NL} < 42$	$9 < f_{\rm NL} < 55$	35.8
DESI (comp. cut)	Nonlinear Three Maps	28	29	$18 < f_{\rm NL} < 40$	$9 < f_{\rm NL} < 53$	34.5
DESI	Nonlinear Four Maps	17	18	$8 < f_{\rm NL} < 27$	$-2 < f_{\rm NL} < 38$	35.2
DESI	Nonlinear Nine Maps	-6	-9	$-21 < f_{\rm NL} < 2$	$-34 < f_{\rm NL} < 12$	39.5
DESI	Nonlinear Three Maps+$f_{\rm NL} = 76.9$ Cov	32	33	$21 < f_{\rm NL} < 45$	$11 < f_{\rm NL} < 59$	33.5
BASS+MzLS	**Nonlinear Three Maps**	**15**	**19**	$\mathbf{-1 < f_{\rm NL} < 39}$	$\mathbf{-19 < f_{\rm NL} < 64}$	**35.6**
BASS+MzLS	Nonlinear Four Maps	13	15	$-5 < f_{\rm NL} < 36$	$-25 < f_{\rm NL} < 59$	34.7
BASS+MzLS	Nonlinear Nine Maps	-4	-6	$-27 < f_{\rm NL} < 14$	$-47 < f_{\rm NL} < 34$	36.8
BASS+MzLS (imag. cut)	Nonlinear Three Maps	25	29	$6 < f_{\rm NL} < 52$	$-14 < f_{\rm NL} < 81$	36.2
BASS+MzLS (comp. cut)	Nonlinear Three Maps	17	21	$0 < f_{\rm NL} < 42$	$-18 < f_{\rm NL} < 67$	35.8
DECaLS North	**Nonlinear Three Maps**	**41**	**45**	$\mathbf{23 < f_{\rm NL} < 67}$	$\mathbf{5 < f_{\rm NL} < 93}$	**41.1**
DECaLS North	Nonlinear Four Maps	31	35	$14 < f_{\rm NL} < 56$	$-6 < f_{\rm NL} < 81$	41.2
DECaLS North	Nonlinear Five Maps	55	60	$37 < f_{\rm NL} < 84$	$18 < f_{\rm NL} < 113$	38.4
DECaLS North	Nonlinear Nine Maps	1	-6	$-30 < f_{\rm NL} < 17$	$-53 < f_{\rm NL} < 36$	45.1
DECaLS North (no DEC cut)	Nonlinear Three Maps	41	45	$24 < f_{\rm NL} < 66$	$6 < f_{\rm NL} < 91$	40.7
DECaLS North (imag. cut)	Nonlinear Three Maps	43	48	$25 < f_{\rm NL} < 72$	$5 < f_{\rm NL} < 101$	35.1
DECaLS North (comp. cut)	Nonlinear Three Maps	41	45	$22 < f_{\rm NL} < 67$	$4 < f_{\rm NL} < 94$	41.4
DECaLS South	**Nonlinear Three Maps**	**31**	**33**	$\mathbf{15 < f_{\rm NL} < 52}$	$\mathbf{-5 < f_{\rm NL} < 74}$	**30.2**
DECaLS South	Nonlinear Four Maps	14	6	$-21 < f_{\rm NL} < 30$	$-54 < f_{\rm NL} < 50$	31.9
DECaLS South	Nonlinear Five Maps	34	37	$18 < f_{\rm NL} < 57$	$0 < f_{\rm NL} < 81$	30.8
DECaLS South	Nonlinear Nine Maps	-37	-32	$-49 < f_{\rm NL} < -14$	$-65 < f_{\rm NL} < 8$	31.5
DECaLS South (no DEC cut)	Nonlinear Three Maps	44	47	$30 < f_{\rm NL} < 63$	$16 < f_{\rm NL} < 83$	23.8
DECaLS South (imag. cut)	Nonlinear Three Maps	26	23	$3 < f_{\rm NL} < 48$	$-58 < f_{\rm NL} < 71$	30.0
DECaLS South (comp. cut)	Nonlinear Three Maps	30	32	$13 < f_{\rm NL} < 52$	$-9.78 < f_{\rm NL} < 74$	29.7

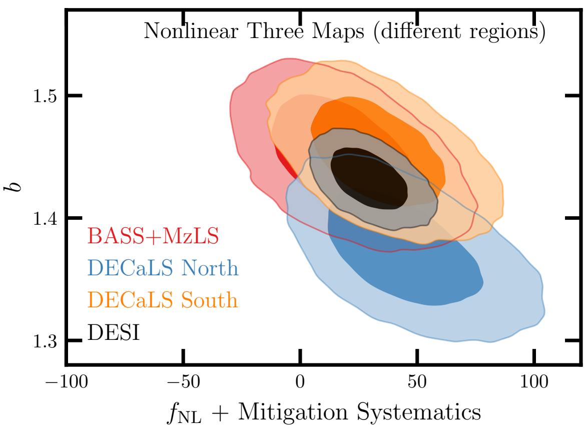

Figure 13. The uncalibrated 2D constraints from the DESI LRG targets for each imaging survey compared with that for the whole DESI footprint. The dark and light shades represent the 68% and 95% confidence intervals, respectively.

5 DISCUSSION AND CONCLUSION

We have measured the local PNG parameter $f_{\rm NL}$ using the scale-dependent bias in the angular clustering of LRGs selected from the DESI Legacy Imaging Survey DR9. Our sample includes more than 12 million LRG targets covering around 14,000 square degrees in the redshift range of $0.2 < z < 1.35$. We leverage early spectroscopy during DESI Survey Validation (DESI Collaboration in prep) to infer the redshift distribution of our sample (Figure 1). In our fiducial analysis, we have obtained a maximum likelihood value of $f_{\rm NL} = 47$ with a significant probability that $f_{\rm NL}$ is greater than zero, $P(f_{\rm NL} > 0) = 99.9$ (Figure 11 and Table 5).

The signature of local PNG is very sensitive to excess clustering signals caused by imaging systematic effects. We have applied a series of robustness tests on the impact of how we estimate the selection function of our LRG sample, including: the methods (linear and nonlinear), the set of imaging systematic maps used (Galactic extinction, stellar density, depth in $grzW1$, psfsize in grz, and neutral hydrogen column density), and data quality cuts on the accepted regions. We find no change in the analysis that shifts the maximum likelihood value of $f_{\rm NL}$ to a significantly lower value (Figure 13, Figure 14, and Table 6). The only manner in which the significance of nonzero PNG decreases is due to the uncertainty on the measurement increasing when we employ more imaging systematic maps to the selection function estimation and by doing so remove large-scale clustering information (the effect of which on $f_{\rm NL}$ recovery we have calibrated with mocks, as shown in Figure 9).

When comparing our fiducial results to recent CMB and QSO measurements, as shown in Figure 15, we find a significant tension with CMB but consistent constraints with LSS within 95% confidence. Either we have measured an $f_{\rm NL}$ signal that is inconsistent with CMB measurements or there is a hidden source of systematic contamination in our data which cannot be mitigated with available

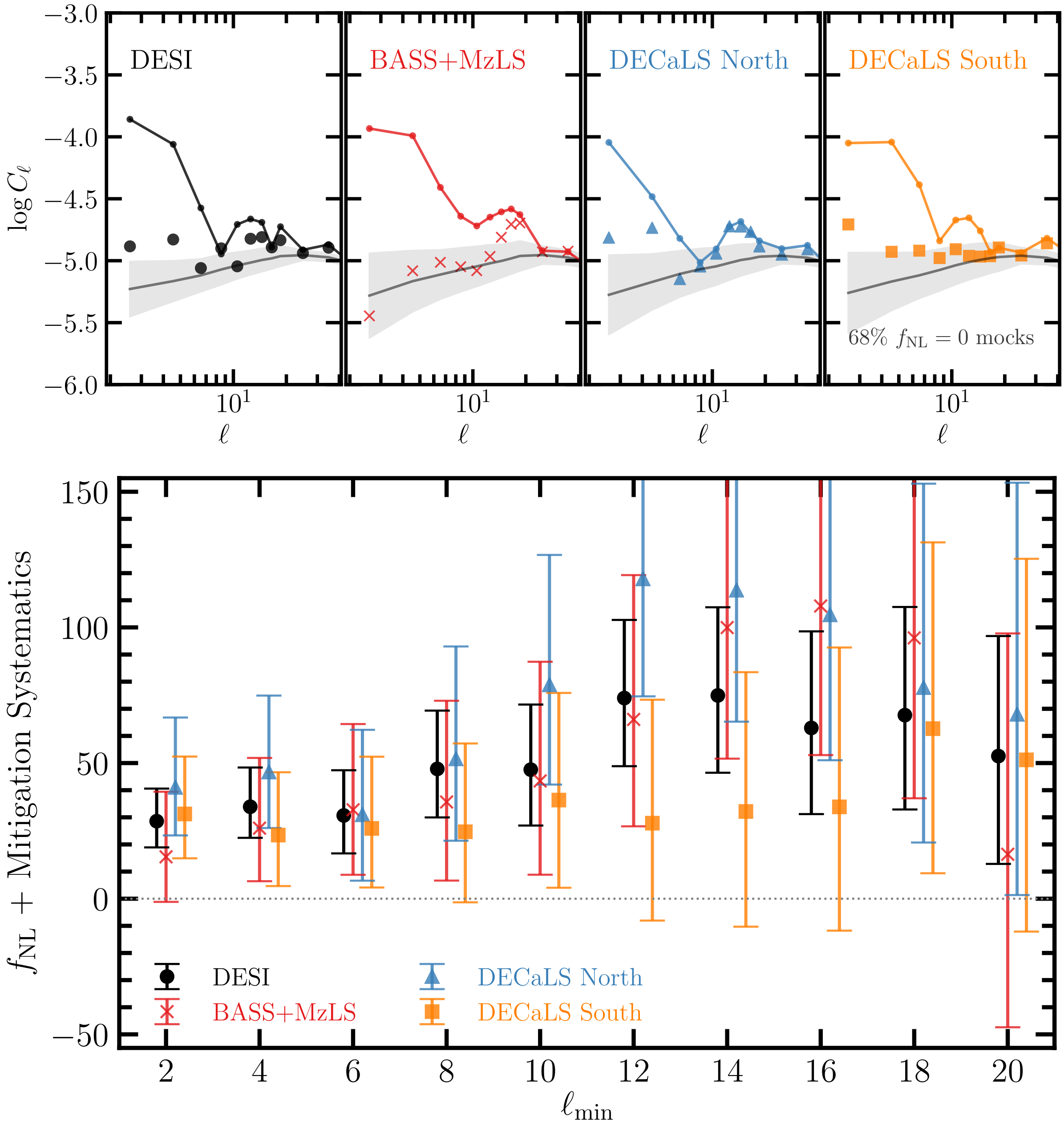

Figure 14. Top: The measured power spectrum of the DESI LRG targets before (solid curves) and after *non-linear three maps* (scatter points) for the DESI, BASS+MzLS, DECaLS North, and DECaLS South regions. Bottom: The uncalibrated $f_{\rm NL}$ constraints vs the lowest ℓ mode used for fitting $f_{\rm NL}$. The points represent marginalized mean estimates of $f_{\rm NL}$ and error bars represent 68% confidence estimated from the $f_{\rm NL} = 0$ mocks. The scaling of $f_{\rm NL}$ is not calibrated to account for over-correction caused by mitigation.

imaging systematic maps. To be consistent with Planck, it would need to correspond to some kind of scale-dependent $f_{\rm NL}$ model which has a larger non-gaussianity on LSS scales but negligible one at CMB scales (possible references: Sefusatti, Liguori, Yadav, Jackson and Pajer (2009), Becker, Huterer and Kadota (2011)). Our analysis can be considered as the first attempt to identify major systematics in DESI, so we can be ready for constraining $f_{\rm NL}$ with DESI spectroscopy. Internal DESI tests of the photometric calibration ~~are~~were unable to uncover DESI-specific issues, e.g., when comparing to Gaia data. The most significant trends that we find are with the E(B-V) map. The source of such a trend would be a mis-calibration of the E(B-V) map itself or the coefficients applied to obtain Galactic extinction corrected photometry. ~~Such issues are generically likely to be proportional in amplitude to the estimated E(B-V), but may not strictly follow the estimated E(B-V)~~Such a mis-calibration would plausibly be proportional in amplitude to the estimated E(B-V) map, though it may not have E(B-V)'s spatial distribution. In order to explain the $f_{\rm NL}$ signal we measure, such an effect would need to be approximately twice that of the trend we find with E(B-V). There are ongoing efforts within DESI to obtain improved Galactic extinction information, which will help ~~discover~~establish if this is indeed the cause.

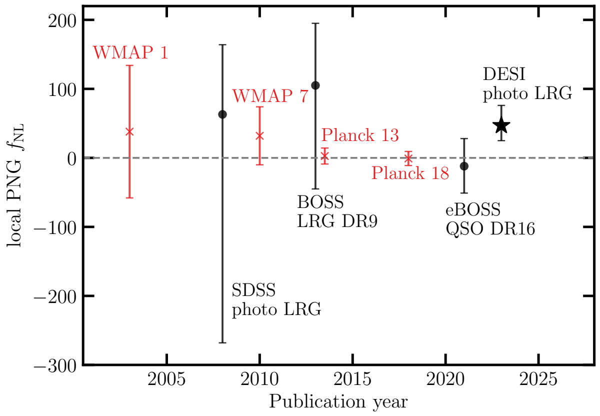

Figure 15. DH: what are the curves? Are they some fits through the points? Also the whole figure seems fuzzy - seems some kind of png file type issue, would be good to make it sharper. Add (this work) History of constraints on local PNG $f_{\rm NL}$ at 95% confidence from single-tracer LSS (Slosar et al. 2008; Ross et al. 2013; Mueller et al. 2022), including our analysis with $25 < f_{\rm NL} < 76$ (DESI photo LRG) and CMB surveys (Komatsu et al. 2003; Komatsu 2010; Planck Collaboration et al. 2014, 2019). The median $f_{\rm NL}$ value is used in case the maximum likelihood estimate was not reported in the reference.

ACKNOWLEDGEMENTS

MR is supported by the U.S. Department of Energy grants DE-SC0021165 and DE-SC0011840. H-JS acknowledges support from the U.S. Department of Energy, Office of Science, Office of High Energy Physics under grant No. DE-SC0019091 and No. DE-SC0023241. FB is a University Research Fellow, and has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 853291). BB-K is supported by the project 우주거대구조를 이용한 암흑우주 연구 ("Understanding Dark Universe Using Large Scale Structure of the Universe"), funded by the Ministry of Science of the Republic of Korea. We acknowledge the support and resources from the Ohio Supercomputer Center (OSC; Center 1987).

MR would like to thank CCAPP, in particular, John Beacom and Lisa Colarosa, for the hospitality and support. We would like to thank Tanveer Karim and Sukhdeep Singh for helpful discussions, and Rongpu Zhou for providing the galaxy density and imaging systematic maps. This research has made substantial 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.

The DESI 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: www.desi.lbl.gov/collaborating-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 complete acknowledgments can be found at www.legacysurvey.org/. 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. **All data points shown in the published graph will be also available in a machine-readable form on Github**.

REFERENCES

Abazajian K. N., et al., 2016, arXiv e-prints, p. arXiv:1610.02743

Albrecht A., Steinhardt P. J., 1982, Phys. Rev. Lett., 48, 1220

Alonso D., Bull P., Ferreira P. G., Maartens R., Santos M. G., 2015, ApJ, 814, 145

Alvarez M., et al., 2014, arXiv e-prints, p. arXiv:1412.4671

Arlot S., Celisse A., 2010, Statistics Surveys, 4, 40

BICEP2 Collaboration et al., 2014, Phys. Rev. Lett., 112, 241101

BOSS Collaboration et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617

Bahr-Kalus B., Bertacca D., Verde L., Heavens A., 2021, Journal of Cosmology and Astroparticle Physics, 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, Journal of Cosmology and Astroparticle Physics, 2020, 031

Barreira A., 2022, Journal of Cosmology and Astroparticle Physics, 2022, 013

Barreira A., Cabass G., Schmidt F., Pillepich A., Nelson D., 2020, Journal of Cosmology and Astroparticle Physics, 2020, 013

Bassett B. A., Tsujikawa S., Wands D., 2006, Reviews of Modern Physics, 78, 537

Bautista J. E., et al., 2018, ApJ, 863, 110

Bautista J. E., et al., 2021, Monthly Notices of the Royal Astronomical Society, 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

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, Journal of Cosmology and Astroparticle Physics, 2020, 007

Castorina E., et al., 2019, Journal of Cosmology and Astroparticle Physics, 2019, 010

Center O. S., 1987, Ohio Supercomputer Center, `http://osc.edu/ark:/19495/f5s1ph73`

Chapman M. J., et al., 2022, Monthly Notices of the Royal Astronomical Society, 516, 617

Chaussidon E., et al., 2022, Monthly Notices of the Royal Astronomical Society, 509, 3904

Chen X., 2010, Advances in Astronomy, 2010, 638979

Chon G., Challinor A., Prunet S., Hivon E., Szapudi I., 2004, Monthly Notices of the Royal Astronomical Society, 350, 914

Clerkin L., et al., 2017, MNRAS, 466, 1444

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., 2016a, arXiv preprint arXiv:1611.00036

DESI Collaboration et al., 2016b, arXiv e-prints, p. arXiv:1611.00037

DESI Collaboration et al., 2022, AJ, 164, 207

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

Delubac T., et al., 2016, Monthly Notices of the Royal Astronomical Society, p. stw2741

Desjacques V., Seljak U., 2010, Classical and Quantum Gravity, 27, 124011

Dey A., et al., 2018, arXiv preprint arXiv:1804.08657

Eggert E., Leistedt B., 2023, The Astrophysical Journal Supplement Series, 265, 30

Eisenstein D. J., et al., 2001, The Astronomical Journal, 122, 2267

Elvin-Poole J., et al., 2018, Phys. Rev. D, 98, 042006

Everett S., et al., 2022, The Astrophysical Journal Supplement Series, 258, 15

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

Flauger R., Hill J. C., Spergel D. N., 2014, Journal of Cosmology and Astroparticle Physics, 2014, 039

Flaugher B., et al., 2015, The Astronomical Journal, 150, 150

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Friedrich O., et al., 2021, MNRAS, 508, 3125

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, Monthly Notices of the Royal Astronomical Society, 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., 1981, Phys. Rev. D, 23, 347

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., 2012, APJ, 761, 14

Ho S., et al., 2015, Journal of Cosmology and Astroparticle Physics, 2015, 040

Huterer D., Cunha C. E., Fang W., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2945

Jolicoeur S., Maartens R., Dlamini S., 2023, arXiv e-prints, p. arXiv:2301.02406

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., 2010, Classical and Quantum Gravity, 27, 124010

Komatsu E., Spergel D. N., 2001, Physical Review D, 63, 063002

Komatsu E., et al., 2003, The Astrophysical Journal Supplement Series, 148, 119

Kong H., et al., 2020, Monthly Notices of the Royal Astronomical Society, 499, 3943

Laurent P., et al., 2017, Journal of Cosmology and Astroparticle Physics, 2017, 017

Lazeyras T., Barreira A., Schmidt F., Desjacques V., 2023, Journal of Cosmology and Astroparticle Physics, 2023, 023

Levi M., et al., 2013, arXiv e-prints, p. arXiv:1308.0847

Linde A., 1982, Physics Letters B, 108, 389

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: Cosmology, 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 American Astronomical Society, 2, 1

Merz G., et al., 2021, Monthly Notices of the Royal Astronomical Society, 506, 2503

Mortonson M. J., Seljak U., 2014, Journal of Cosmology and Astroparticle Physics, 2014, 035

Mueller E.-M., Percival W. J., Ruggeri R., 2019, Monthly Notices of the Royal Astronomical Society, 485, 4160

Mueller E.-M., et al., 2022, Monthly Notices of the Royal Astronomical Society, 514, 3396

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 conference on machine learning (ICML-10). pp 807–814

Padmanabhan N., et al., 2007, Monthly Notices of the Royal Astronomical Society, 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., 2014, A&A, 571, A24

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

Raichoor A., et al., 2017, MNRAS, 471, 3955

Reid B. A., Verde L., Dolag K., Matarrese S., Moscardini L., 2010, Journal of Cosmology and Astroparticle Physics, 2010, 013

Rezaie M., Seo H.-J., Ross A. J., Bunescu R. C., 2020, MNRAS, 495, 1613

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., 2012, MNRAS, 424, 564
Ross A. J., et al., 2013, Monthly Notices of the Royal Astronomical Society, 428, 1116
Ross A. J., et al., 2017, MNRAS, 464, 1168
Ross A. J., et al., 2020, Monthly Notices of the Royal Astronomical Society, 498, 2354
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
Seljak U. c. v., 2009, Phys. Rev. Lett., 102, 021302
Silber J. H., et al., 2023, AJ, 165, 9
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
Suchyta E., et al., 2016, Monthly Notices of the Royal Astronomical Society, 457, 786
Sullivan J. M., Prijon T., Seljak U., 2023, arXiv e-prints, p. arXiv:2303.08901
Tegmark M., et al., 2004, Phys. Rev. D, 69, 103501
Thomas S. A., Abdalla F. B., Lahav O., 2011, Phys. Rev. Lett., 106, 241301
Verde L., 2010, Advances in Astronomy, 2010, 768675
Wang M. S., Beutler F., Bacon D., 2020, Monthly Notices of the Royal Astronomical Society, 499, 2598
Weaverdyck N., Huterer D., 2021, Monthly Notices of the Royal Astronomical Society, 503, 5061
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, Phys. Rev. D, 95, 123507

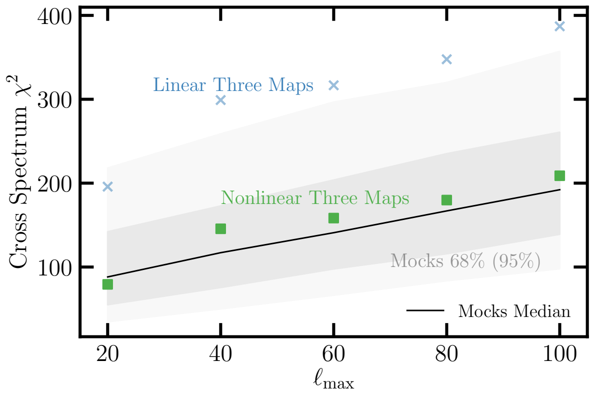

Figure A1. The cross power spectrum's χ^2 between the LRG density and imaging systematic maps as a function of the highest mode $\ell_{\rm max}$ when the sample is cleaned with the linear (triangles) and non-linear (squares) three maps. The lowest mode is fixed at $\ell_{\rm min} = 2$. The solid curve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

APPENDIX A: EXTRA ROBUSTNESS TESTS

A1 Scale dependent systematics

Our fiducial cross power spectrum diagnostic (equation 17) uses harmonic modes up to $\ell = 20$, which determines the smallest scale used for characterizing residual systematic errors. To investigate the statistical significance of the cross power spectrum's χ^2, we examine its dependence on the largest harmonic mode $\ell_{\rm max}$. We extend $\ell_{\rm max}$ from 20 to 100, where the latter scale corresponds to density fluctuations on scales smaller than 2 degrees. Figure A1 shows the median of the cross power spectrum's χ^2 from the clean $f_{\rm NL} = 0$ mocks as the highest mode $\ell_{\rm max}$ increases from 20 to 100 (represented by the solid line). The red circles and blue crosses show the χ^2 values for our sample cleaned respectively with the linear and neural network approaches, both with three maps. Overall, we find that for all scales up to $\ell = 100$, the nonlinear three maps approach yields consistent values with the clean mocks, whereas the linear three maps approach exhibits significant remaining systematics at more than 95% confidence.

A2 Redshift uncertainties

We use the Early Data Assembly Version 1 (EDA V1) to construct the redshift distribution for the DESI LRG targets. We find that the change in the maximum likelihood estimate of $f_{\rm NL}$ is negligible, $\Delta f_{\rm NL} = -1$. Figure A2 shows the measured power spectrum of the DESI targets and the corresponding best fit theory curves. The variations in dN/dz do not significicantly alter the conclusion of our paper.

A3 Spurious bump in NGC

We realize that the spurious feature at $\ell \sim 10 - 20$ is removed in the DECaLS South region after mitigation, but it remains in the BASS+MzLS and DECaLS North. We use the neural networks trained on the DECaLS South with three maps to mitigate the galaxy density in the DECaLS North region, and then measure the power spectrum. Figure A3 shows the power spectrum before treatment (No Weight) and after the nonlinear three maps and nine maps

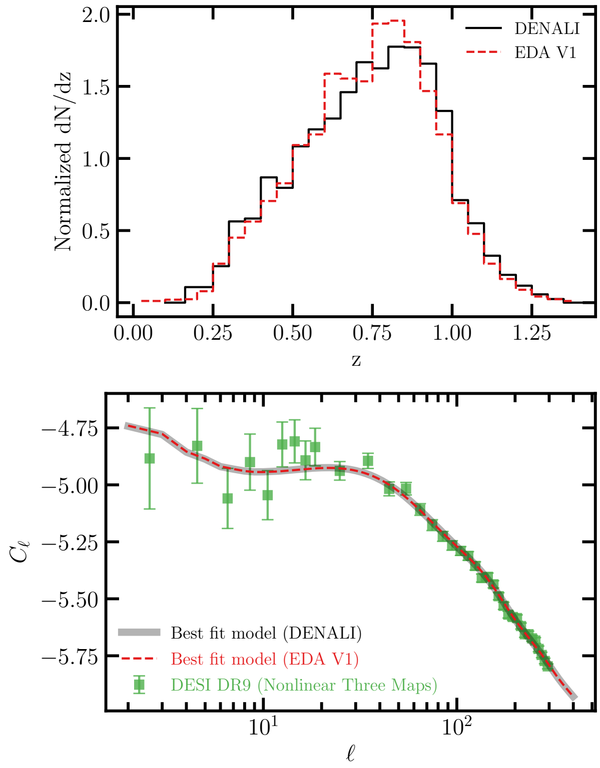

Figure A2. Top: Redshift distribution of the DESI LRG targets from the EDA V1 and Denali. Bottom: The measured power spectrum of the DESI DR9 LRG sample and the best fit theory models using different redshift distributions.

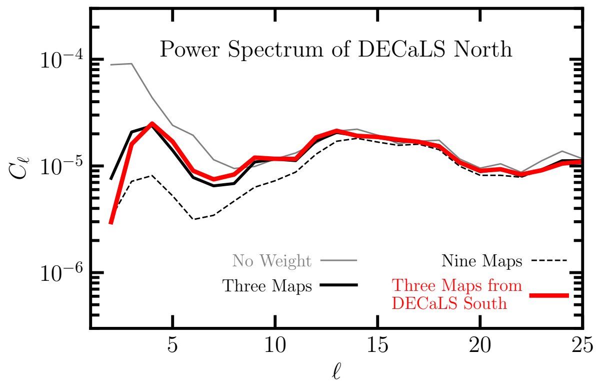

Figure A3. The unbinned measured power spectrum of the DESI LRG targets in the DECaLS North region before and after various mitigations.

methods for comparison. We find that whatever causing this bump is different between the DECaLS North and South.

APPENDIX B: LOGNORMAL MOCKS

We fit the mean power spectrum of the lognormal mocks to validate the modeling pipeline, and in particular the survey geometry and integral constraint treatments. We investigate the impact of covariance matrix on the inference of $f_{\rm NL}$. Finally, we show the impact of imaging systematic mitigation and the over-subtraction effect when the cleaning methods are applied to the mocks.

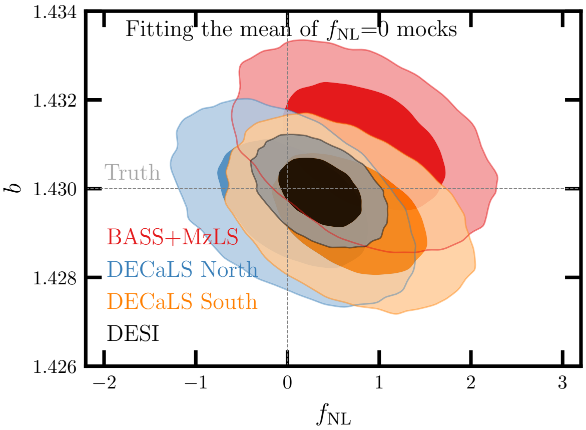

Figure B1. 68% and 95% confidence contours from the mean power spectrum of the $f_{\rm NL}=0$ mocks for the DESI footprint and sub-imaging surveys. The truth values are represented by vertical and horizontal lines.

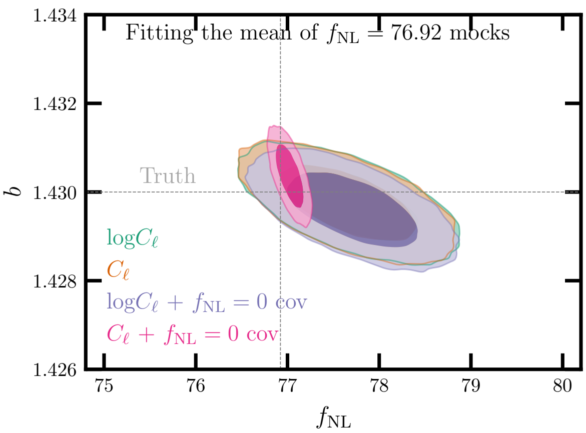

Figure B2. 68% and 95% confidence contours of fitting the mean power spectrum or its log transformation from the $f_{\rm NL}=76.9$ mocks for the DESI footprint. Using the log C_ℓ fitting yield constraints that are insensitive to the covariance used. The truth values are represented by vertical and horizontal lines.

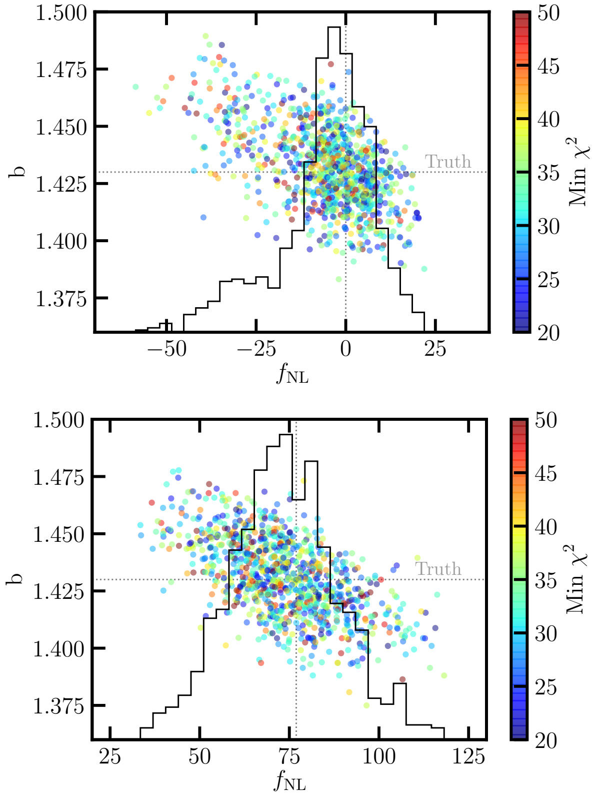

Figure B3. best-fitting estimates from fitting 1000 lognormal mocks with $f_{\rm NL}=0$ (top) and 76.9 (bottom) in the DESI footprint. The truth values are represented by vertical and horizontal lines. The truth values are represented by vertical and horizontal lines.

B1 Clean mocks

The 68% and 95% probability contours on the PNG parameter $f_{\rm NL}$ and bias coefficient b are shown in Figure B1 and B2, respectively, for the $f_{\rm NL}=0$ and 76.9 mocks. The best-fitting, marginalized mean estimates, as well as the 1σ and 2σ confidence intervals of $f_{\rm NL}$ are summarized in Table B1.

Measuring the power spectrum from the entire DESI footprint reduces the cosmic variance and thus improves the constraining power. Figure B1 compares the constraints from fitting the log of the mean power spectrum of the mocks when it is measured from the DESI footprint to those obtained from the sub imaging surveys. We find that the underlying true $f_{\rm NL}$ value is recovered within 95% confidence, and that the contours for the DESI region are smaller by a factor of two.

The power spectrum of the mocks at low ℓ is very sensitive to the cosmic variance and the true value of $f_{\rm NL}$. Consequently, a large value of $f_{\rm NL}$ can induce very large power on low ℓ, and thus significantly change the covariance matrix. We find that applying the log transformation on the power spectrum makes the result more robust against the choice of the covariance matrix. Figure B2 shows the confidence contours when we fit either the power spectrum or its log transform of the $f_{\rm NL}=76.9$ mocks, and use different covariance matrices. We consider the $f_{\rm NL}=0$ and 76.9 mocks to construct the covariance from one set and use it to fit the mean power spectrum of the other set. When the covariance matrix is constructed from the same set of mocks used for the mean power spectrum, we find that the difference in $f_{\rm NL}$ constraints between fitting the power spectrum and its log transformation is negligible at only 2%. If we use the $f_{\rm NL}=0$ mocks to estimate the covariance, and fit the log power spectrum of the $f_{\rm NL}=76.9$ mocks, we find that the error on $f_{\rm NL}$ increases only by 7%. However, when the 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 tighten by a factor of 5 due to a higher signal to noise ratio. Therefore, we argue that fitting the log power spectrum can help mitigate the need for having $f_{\rm NL}$-dependent covariance matrices and make the constraints less sensitive to covariance construction.

Figure B3 shows the best-fitting estimates for b vs $f_{\rm NL}$ for $f_{\rm NL}=0$ and $=76.9$ mocks in the top and bottom panels, 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-fitting $f_{\rm NL}$ estimates are plotted in the background. For the $f_{\rm NL}=0$ mocks, the best-fitting estimates are more symmetric. To understand this behaviour, we consider the first derivative of the likelihood (Equation 14), which is proportional to the first derivative of the log power spectrum. By simplifying the integrals involved in C_ℓ, we have $C_\ell = A_{0,\ell} + A_{1,\ell} f_{\rm NL} + A_{2,\ell} f_{\rm NL}^2$

Table B1. best-fitting 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 76.9	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 76.9	DESI	C_ℓ	77.67	77.65	$77.17 < f_{\rm NL} < 78.14$	$76.70 < f_{\rm NL} < 78.60$	39.0
Clean 76.9	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 76.9	DESI	$C_\ell + f_{\rm NL} = 0$ 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	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
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

where $A_{123,\ell}$ are ℓ-dependent terms. Then, the derivative of the likelihood is equivalent to

$$\frac{d}{df_{\rm NL}}\log(C_\ell) = \frac{A_{1,\ell} + 2A_{2,\ell}f_{\rm NL}}{A_{0,\ell} + A_{1,\ell}f_{\rm NL} + A_{2,\ell}f_{\rm NL}^2}. \tag{B1}$$

For infinitesimal values of $f_{\rm NL}$, the derivative becomes asymptotically independent from $f_{\rm NL}$ while for large values of $f_{\rm NL}$ is decreases as $2/f_{\rm NL}$. This implies that for the $f_{\rm NL} = 0$ mocks, the likelihood is more likely to be skewed toward negative values.

B2 Contaminated mocks

Our nonlinear neural network-based approach is applied to the $f_{\rm NL} = 0$ and 76.9 mocks. We only consider the methods that include running the neural network with three, four, and nine imaging systematic maps. The measured mean power spectrum of the mocks are shown in Figure B4 for $f_{\rm NL} = 0$ (left) and 76.9 (right). The solid and dashed curves show the measurements respectively from the clean and contaminated mocks.

We find that the imaging treatment has removed some of the true clustering signal, and the amount of the over-subtraction is almost the same regardless of whether the mocks have systematics. The over-subtraction induces biases in the $f_{\rm NL}$ constraints, as summarized in Table B2. The over-subtraction at low ℓ is so high that we get a poor fit after applying the mitigation with the nonlinear three maps approach, e.g., $\chi^2 = 86.8$ for the clean $f_{\rm NL} = 0$ mocks.

Using the calibration parameters presented in §3.5, we account for the shift in the $f_{\rm NL}$ constraints caused by the imaging systematic mitigation. We show the marginalized probability distributions on $f_{\rm NL}$ before and after accounting for the over-correction in the right and left panels of Figure B5.

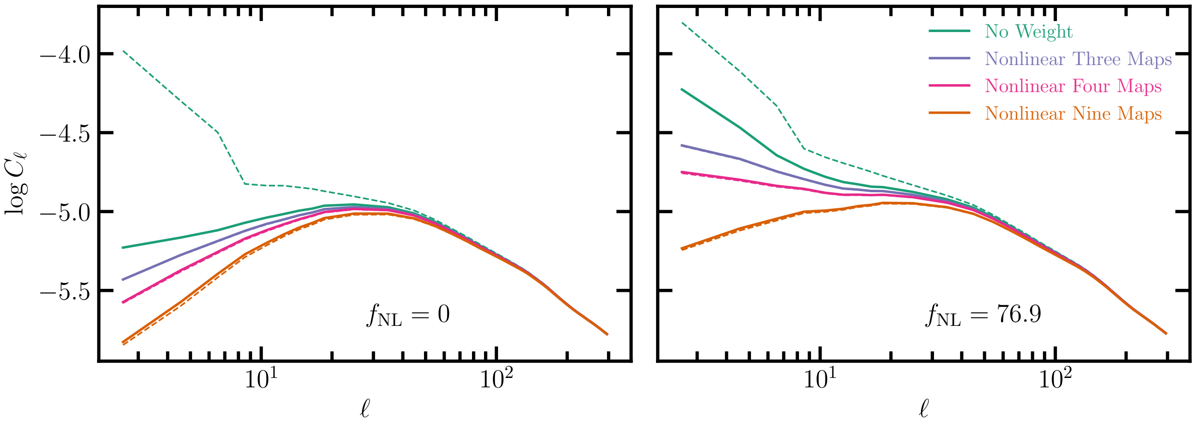

Figure B4. The mean power spectrum of the $f_{\rm NL} = 0$ and 76.9 mocks with (dashed) and without (solid) imaging systematics before ('No Weight') and after applying the non-linear cleaning method with three, four, and nine maps.

Table B2. best-fitting and marginalized estimates for $f_{\rm NL}$ from fitting the mean power spectrum of the mocks before and after corrections using the non-linear approach with various combinations of the imaging systematic maps. The estimates are not accounted for over-correction, and therefore are subject to mitigation systematics.

		$f_{\rm NL}$ + Mitigation Systematics				
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	Three Maps	−11.64	−11.65	$-12.00 < f_{\rm NL} < -11.30$	$-12.34 < f_{\rm NL} < -10.97$	86.8
Clean 0	Four Maps	−20.14	−20.13	$-20.44 < f_{\rm NL} < -19.82$	$-20.74 < f_{\rm NL} < -19.52$	472.8
Clean 0	Nine Maps	−26.91	−26.92	$-27.16 < f_{\rm NL} < -26.68$	$-27.39 < f_{\rm NL} < -26.46$	5481.0
Contaminated 0	Three Maps	−12.12	−12.13	$-12.48 < f_{\rm NL} < -11.78$	$-12.83 < f_{\rm NL} < -11.44$	94.0
Contaminated 0	Four Maps	−20.97	−20.98	$-21.28 < f_{\rm NL} < -20.67$	$-21.58 < f_{\rm NL} < -20.37$	556.3
Contaminated 0	Nine Maps	−28.13	−28.13	$-28.36 < f_{\rm NL} < -27.90$	$-28.59 < f_{\rm NL} < -27.67$	6760.5
Clean 76.9	No Weight	77.67	77.67	$77.17 < f_{\rm NL} < 78.16$	$76.71 < f_{\rm NL} < 78.64$	38.8
Clean 76.9	Three Maps	54.57	54.57	$54.14 < f_{\rm NL} < 55.01$	$53.72 < f_{\rm NL} < 55.45$	603.5
Clean 76.9	Four Maps	38.38	38.38	$37.99 < f_{\rm NL} < 38.78$	$37.60 < f_{\rm NL} < 39.16$	537.0
Clean 76.9	Nine Maps	6.04	6.04	$5.72 < f_{\rm NL} < 6.36$	$5.41 < f_{\rm NL} < 6.67$	694.0
Contaminated 76.9	Three Maps	54.01	54.00	$53.57 < f_{\rm NL} < 54.44$	$53.15 < f_{\rm NL} < 54.86$	588.0
Contaminated 76.9	Four Maps	37.48	37.49	$37.09 < f_{\rm NL} < 37.88$	$36.70 < f_{\rm NL} < 38.27$	510.7
Contaminated 76.9	Nine Maps	4.59	4.58	$4.26 < f_{\rm NL} < 4.90$	$3.95 < f_{\rm NL} < 5.22$	649.7

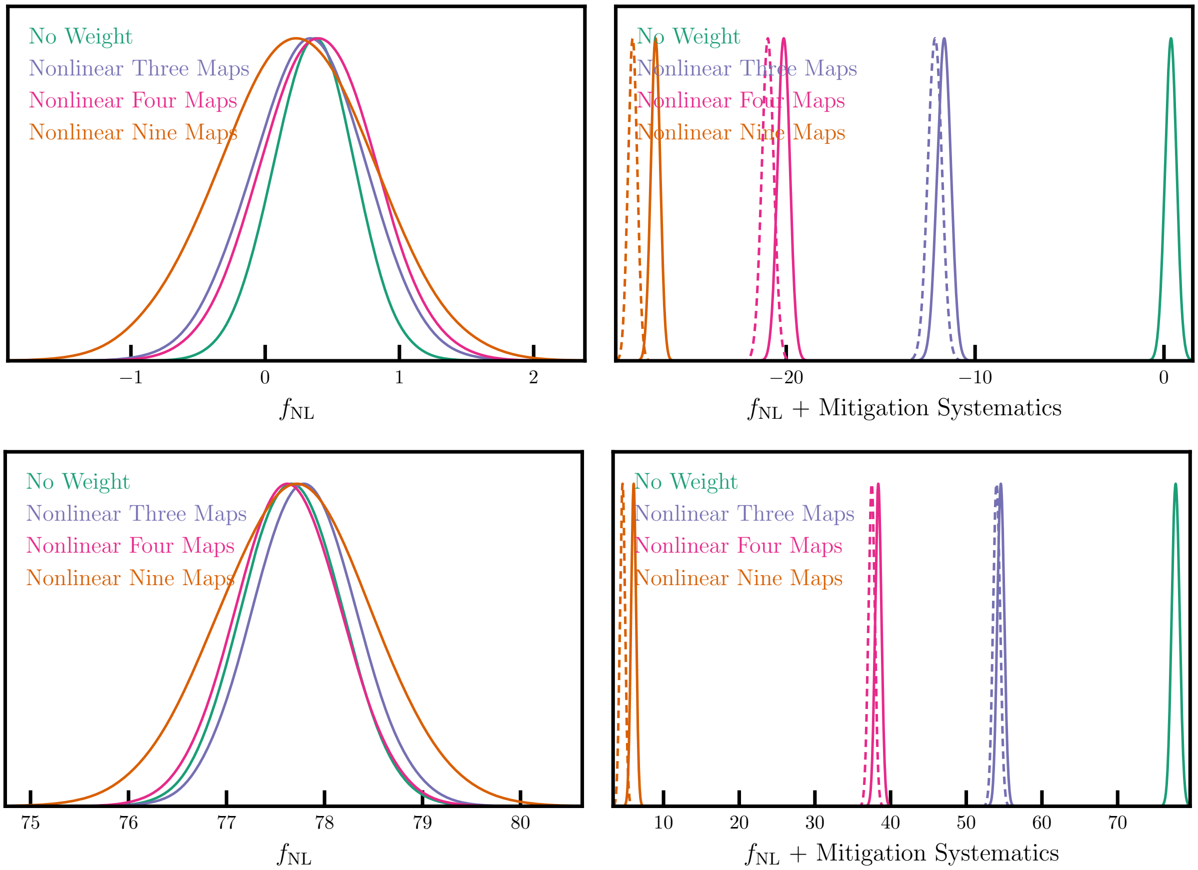

Figure B5. Probability distributions of $f_{\rm NL}$ from the mean power spectrum of the $f_{\rm NL}=0$ (top) and $f_{\rm NL}=76.9$ (bottom) mocks before and after mitigation with the non-linear methods using three, four, and nine maps. The dashed (solid) curves show the distributions for the contaminated (clean) mocks. 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 $f_{\rm NL}$ values is biased due to mitigation.

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