Primordial non-Gaussianity from the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey – I: Catalogue preparation and systematic mitigation

Mehdi Rezaie [1]★ Ashley J. Ross [2] Hee-Jong Seo [1] Eva-Maria Mueller,[3] Will J. Percival,[4,5,6]
Grant Merz,[1] Reza Katebi [1] Razvan C. Bunescu,[7] Julian Bautista [8] Joel R. Brownstein [9]
Etienne Burtin,[10] Kyle Dawson,[9] Héctor Gil-Marín [11,12] Jiamin Hou,[13] Eleanor B. Lyke,[14] Axel de la
Macorra,[15] Graziano Rossi,[16] Donald P. Schneider,[17,18] Pauline Zarrouk[10,19] and Gong-Bo Zhao[20,21]

Affiliations are listed at the end of the paper

ABSTRACT

We investigate the large-scale clustering of the final spectroscopic sample of quasars from the recently completed extended Baryon Oscillation Spectroscopic Survey (eBOSS). The sample contains 343 708 objects in the redshift range $0.8 < z < 2.2$ and 72 667 objects with redshifts $2.2 < z < 3.5$, covering an effective area of 4699 $\deg^2$. We develop a neural network-based approach to mitigate spurious fluctuations in the density field caused by spatial variations in the quality of the imaging data used to select targets for follow-up spectroscopy. Simulations are used with the same angular and radial distributions as the real data to estimate covariance matrices, perform error analyses, and assess residual systematic uncertainties. We measure the mean density contrast and cross-correlations of the eBOSS quasars against maps of potential sources of imaging systematics to address algorithm effectiveness, finding that the neural network-based approach outperforms standard linear regression. Stellar density is one of the most important sources of spurious fluctuations, and a new template constructed using data from the *Gaia* spacecraft provides the best match to the observed quasar clustering. The end-product from this work is a new value-added quasar catalogue with the improved weights to correct for non-linear imaging systematic effects, which will be made public. Our quasar catalogue is used to measure the local-type primordial non-Gaussianity in a companion paper.

Key words: inflation – large-scale structure of the Universe.

1 INTRODUCTION

The statistical properties of the initial conditions of the Universe are an open problem in modern cosmology. Single-field inflationary models predict that primordial fluctuations are almost Gaussian and any deviation towards non-Gaussianity is small (Guth & Pi 1982; Hawking 1982; Starobinsky 1982; Bardeen, Steinhardt & Turner 1983; Acquaviva et al. 2003; Maldacena 2003a; Creminelli & Zaldarriaga 2004; Scoccimarro, Sefusatti & Zaldarriaga 2004; Zaldarriaga 2004). However, some alternative inflationary models, which include more fields, generate non-Gaussian perturbations (Allen, Grinstein & Wise 1987; Kofman & Pogosyan 1988; Salopek, Bond & Bardeen 1989; Linde & Mukhanov 1997; Bernardeau & Uzan 2002; Lyth, Ungarelli & Wands 2003; Maldacena 2003b; Chen et al. 2007; see e.g. Desjacques & Seljak 2010 for a review). The local-type primordial non-Gaussianity is often parametrized as (Matarrese, Verde & Jimenez 2000; Verde et al. 2000; Komatsu &

Spergel 2001):

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

where Φ is the primordial gravitational field, ϕ is a Gaussian field, and f_{NL} is called the non-linear coupling constant. For a single field inflationary model, we have $f_{NL} = 5(1 - n_s)/12 \sim 0.01$, where n_s is the scalar spectral index of primordial power spectrum. Therefore, a non-zero detection of primordial non-Gaussianity, $f_{NL} \gtrsim 1$, will rule out single-field models of inflation and refine our understanding of the early Universe (Mukhanov & Chibisov (Mukhanov & Chibisov 1981; Guth & Pi 1982; Hawking 1982; Starobinsky 1982; Allen et al. 1987; Falk, Rangarajan & Srednicki 1992; Gangui et al. 1993; see e.g. Alvarez et al. 2014 for a review).

The current state-of-the-art constraint is $f_{NL} = -0.9 \pm 5.1$ at 68 per cent confidence level from measurements of the bispectrum of cosmic microwave background anisotropies as measured by the *Planck* satellite (Akrami et al. 2020). However, limited by cosmic variance, CMB measurements cannot reach the necessary precision to differentiate between various inflationary models (e.g. Baumann et al. 2009; Abazajian et al. 2016). As an alternative route to constrain f_{NL} is to measure the scale-dependent effect, it has on the large-scale

★ E-mail: mr095415@ohio.edu

clustering of biased tracers at lower redshift (Scoccimarro et al. 2004; Dalal et al. 2008; Grossi et al. 2008; Matarrese & Verde 2008; Slosar et al. 2008; Taruya, Koyama & Matsubara 2008),

$$\Delta b \propto f_{\mathrm{NL}}(b - p)\frac{1}{k^2 T(k)}, \qquad (2)$$

where $T(k)$ is the transfer function normalized to unity for small wavenumber k, b is the halo bias, and p is a correction factor accounting for the response of the tracer to the halo gravitational field, e.g. 1.6 for recent mergers (Slosar et al. 2008; Reid et al. 2010). Due to the k^{-2} dependence, the biasing effect is most apparent for fluctuations with large wavelengths (see e.g. Alvarez et al. 2014, for a review). Therefore, constraining primordial non-Gaussianity demands galaxy redshift experiments that survey a huge cosmic volume.

Quasars are bright quasi-stellar objects that are detectable from high redshifts thanks to the brightness of their active nuclei, and are consequently the ideal candidate for studying the distribution of matter on high redshifts, and constraining primordial non-Gaussianities with large-scale structure data (Agarwal, Ho & Shandera 2014; Giannantonio et al. 2014; Karagiannis, Shanks & Ross 2014; Leistedt, Peiris & Roth 2014). Using the scale-depend bias effect on galaxy power spectrum measurements, Slosar et al. (2008) obtained $f_{\mathrm{NL}} = 8^{+26}_{-37}$ at 68 per cent confidence level from the Sloan Digital Sky Survey (SDSS; Blanton et al. 2017) Data Release 5 (DR5) quasar sample (Adelman-McCarthy et al. 2007). Recently, Castorina et al. (2019) used 148 659 quasars with $0.8 < z < 2.2$ from the SDSS DR14 (Abolfathi et al. 2018) to obtain $-51 < f_{\mathrm{NL}} < 21$ at 95 per cent confidence level.

Quasar clustering measurements at large scales are sensitive to spurious density fluctuations caused by imaging properties (Ross et al. 2012; Leistedt et al. 2013; Pullen & Hirata 2013; Ross et al. 2013; Leistedt & Peiris 2014; Ho et al. 2015; Laurent et al. 2017; Kalus et al. 2019), and are the primary source of systematic error that impede a robust analysis of primordial non-Gaussianity. The reasons for data having systematic errors include, but are not limited to, Galactic foregrounds (such as Galactic extinction and stellar contaminations), seeing, and survey depth variations. These properties fluctuate across the survey footprint and introduce spurious variations in the observed density field of quasars. Because many of these effects have large-scale variations, they result in an excess clustering signal at large scales. For instance, Pullen & Hirata (2013) analysed the SDSS DR6 quasar sample (Richards et al. 2008) to find that systematic error in the sample does not allow a robust inference on f_{NL}. They found the 2 per cent rms fluctuations in the quasar density, while they argued that the fluctuations under 1 per cent (0.6 per cent) is required to measure f_{NL} less than 100 (10).

Conventional techniques for improving quasar clustering measurements rely on a regression analysis of observed quasar density against a set of mappable properties that describe imaging conditions during observing (see e.g. Ross et al. 2011, 2012, 2017; Delubac et al. 2017; Prakash et al. 2016; Laurent et al. 2017; Bautista et al. 2018). This approach has been further enhanced to account for intercorrelations among imaging variables, validated with simulations, and applied to other photometric surveys, such as the Dark Energy Survey (e.g. Elvin-Poole et al. 2018; Wagoner et al. 2021). The regression analysis fits for spurious fluctuations in target quasar density and then is employed as a selection function to assign an appropriate weight to each target to eliminate such variations. The method of mode-projection removes the modes that strongly correlate with imaging templates in covariance based estimators (e.g. Tegmark 1997; Leistedt et al. 2013; Leistedt & Peiris 2014; Kalus et al.

2019). Regression and mode-projection turn out to be mathematically equivalent (Kalus et al. 2016). An alternative technique uses cross-correlations of multiple tracers, or same tracer at different redshift bins, with the assumption that each sample responds differently to imaging systematics and thus cross-correlations are not affected (e.g. Rhodes et al. 2013). For a benchmark analysis of various cleaning methods, see e.g. Weaverdyck & Huterer (2021).

The extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) is part of the fourth phase of SDSS. The eBOSS programme is the final large-galaxy redshift survey proposed within SDSS aiming to measure the expansion history and energy contents of the Universe, using large-scale structure (LSS) (Ahumada et al. 2020). With a series of galaxy redshift surveys, SDSS has probed the distribution of matter traced by galaxies and quasars using the dedicated 2.5m Sloan Foundation Telescope (Gunn et al. 2006) at the Apache Point Observatory from 1998 to 2019. During four distinct phases, SDSS collected images and spectra of thousands to millions of astronomical objects and created the largest three-dimensional map of the cosmic web to date. The first phase of SDSS led to the detection of Baryon Acoustic Oscillations (BAO) in the large-scale clustering of galaxies (Eisenstein et al. 2005), at the same time as the 2-degree Field Galaxy Redshift Survey (Cole et al. 2005). Calibrated by CMB measurements, the BAO scale has been widely utilized as a standard ruler leading to robust distance-redshift measurements and constraints on the nature of Dark Energy with LSS surveys. The observations of the BAO signal in LSS have improved since the early detections, passing the 5σ detection threshold and now cover a wide range in redshift using only SDSS data (e.g. Percival et al. 2010; Anderson et al. 2014; Alam et al. 2017).

In the fourth phase of SDSS, eBOSS adopted the same 1000-fibre spectrograph (Smee et al. 2013) from its predecessor, BOSS (Dawson et al. 2012), and utilized improved pipelines for redshift estimation, background subtraction, and flux calibration (Bolton et al. 2012; Hutchinson et al. 2016; Jensen et al. 2016; Bautista et al. 2017). eBOSS introduced a new class of targets for actively star-forming galaxies with strong [O II] emission lines, known as Emission-Line Galaxies (ELGs; Raichoor et al. 2017). ELGs extend over the redshift range of $0.6 < z < 1.1$ (Raichoor et al. 2020) and fill the redshift gap between Luminous Red Galaxies (LRGs; Prakash et al. 2016) and Quasi-Stellar Objects (QSOs; Myers et al. 2015), referred to as quasars in this manuscript.

We present a careful assessment and treatment of imaging systematic effects in the final sample of quasars (Lyke et al. 2020; Ross et al. 2020) from the eBOSS Data Release 16 (Ahumada et al. 2020), the largest sample of quasars available to date. We improve upon a neural network-based cleaning approach, which was originally developed, validated, and applied to eBOSS-like emission-line galaxies in Rezaie et al. (2020). Compared to emission-line galaxies, quasars represent a class of sparser targets for galaxy surveys, particularly for spectroscopic surveys. This paper enhances the method to deal with the sparsity of the eBOSS DR16 quasars by modelling quasar counts per pixel with the Poisson distribution. We improve neural network training to be less prone to local minima by utilizing a cyclic learning rate. We perform a comprehensive benchmark of linear and non-linear treatments, and investigate each method effectiveness in reducing spurious fluctuations. Residual systematic errors are quantified using mean quasar density contrasts and cross-power spectra against imaging templates. For the significance of residual spurious fluctuations, we construct covariance matrices from realistic simulations. Our primary objective is to examine, quantify, and mitigate the potential sources of observational systematic error and enhance quasar power spectrum measurements for constraining

primordial non-Gaussianity. This work presents a new set of value-added catalogues with the enhanced systematic weights to account for imaging systematic effects more exquisitely compared to the standard linear regression which is used in the eBOSS pipeline. The new quasar catalogue is utilized in two accompanying papers for constraining the local-type primordial non-Gaussianity (Mueller et al. in preparation) and for exploring the impact of imaging systematic error on Baryon Acoustic Oscillations (Merz et al. in preparation).

This paper is structured as follows. Section 2 describes the eBOSS quasar sample and simulated data sets used in this work. Section 3 outlines the power spectrum estimator and our strategies for the treatment and characterization of imaging systematics. In Section 4, we present our statistical tests for quantifying residual systematic errors, assess the performance of different cleaning methods, and illustrate the impact of imaging properties on the measured power spectrum of the DR16 sample and that of the simulated catalogues. Finally, we conclude with a summary of our results and their significance for constraining primordial non-Gaussianity with quasar clustering in Section 5.

2 DATA

This section describes the final sample of quasars from the completed eBOSS DR16 data set (Ross et al. 2020; Lyke et al. 2020) and the simulated EZmock catalogues (Zhao et al. 2021) used in our analysis.

2.1 eBOSS DR16 quasars

The DR16 quasar sample is the final release of large-scale structure data from SDSS-IV eBOSS and provides twice the number of quasars and sky coverage over the previous Data Release DR14. The target selection of quasars for eBOSS is presented in Myers et al. (2015), and only the main details are briefly summarized here. It uses optical and infrared imaging, respectively, from SDSS and *Wide-Field Infrared Survey Explorer* (*WISE*; Wright et al. 2010). The photometric data are taken in five bands (*u, g, r, i, z*; Fukugita et al. 1996) and calibrated to account for the Galactic dust effect using correction factors presented in Schlafly et al. (2012). The ability to obtain a quasar with redshift $z > 0.9$ is improved by 20 per cent using the XDQSOZ algorithm (Bovy et al. 2012). The magnitude selection applies the extinction-corrected flux cuts in the *g* and *r* bands, namely $g < 22$ and $r < 22$, to choose the CORE eBOSS quasars. The targeting strategy enables the observation of the Ly-α high-z quasars by relaxing the maximum redshift cut and reduces stellar contaminations in the sample by incorporating a mid-infrared cut. The redshifts are estimated using the REDVSBLUE[1] principal component analysis algorithm described in Lyke et al. (2020), with 95 per cent completeness and 2 per cent false-positive rate. The main quasar sample spans the redshift range of $0.8 < z < 2.2$ and is used for various cosmological analyses of the BAO and RSD features (du Mas des Bourboux et al. 2020; Neveux et al. 2020; Smith et al. 2020; Hou et al. 2021). Additionally, the high-z quasars with $2.2 < z < 3.5$ are employed for measuring the BAO signal in the Ly α forest (Blomqvist et al. 2019; Chabanier et al. 2019; de Sainte Agathe et al. 2019).

The preparation of the large-scale clustering catalogue of quasars is outlined in Ross et al. (2020). As well as associating attributes and *weights* per object, the process includes generating a set of unclustered synthetic objects (often referred to as the *random catalogue*

or *randoms*), matching the expected weighted density of quasars and accounting for the radial and angular survey geometry. Standard procedures account for the veto masking, completeness cuts, fibre collision correction by nearest neighbour upweighting, redshift failure correction through weighting, and imaging systematics by linear regression against templates. The products are the tabulated coordinates of the quasars and randoms along with appropriate columns for per-object weights, separately for the North Galactic Cap (NGC) and South Galactic Cap (SGC) regions. Each weight column is intended to address a particular systematic effect in the observed data, which is discussed briefly in the following.

The physical size of a fibre limits the ability to observe a pair of quasars within 62 arcsec of separation. This *fibre collision* effect is not random and affects quasar clustering. Up-weighting the nearest neighbour corrects for much of the large-scale effect, although small-scales remain affected and require a more complicated procedure to achieve unbiased removal (Bianchi & Percival 2017; Hahn et al. 2017; Mohammad et al. 2020). In our analysis, we use the simple nearest neighbour upweighting scheme as we are only interested in large scales (as used, for example, by Neveux et al. 2020).

The completeness of redshift estimation varies among the fibres across the spectrographs, with fibres near the edge having a higher rate of redshift measurement failures. The weight w_{noz} is determined to mitigate this systematic error by weighting by the reciprocal of the probability that each fibre accurately measures a redshift. The impact of redshift completeness on quasar clustering is investigated in Hou et al. (2021).

The remaining source of systematic uncertainty is associated with the properties of the imaging data from which targets were selected including, but not limited to, stellar contamination, Galactic extinction, and inaccurate photometric calibrations. The standard method uses a multivariate linear model to find the intercorrelation between target density and imaging templates, and provides a per-object weight w_{systot} to reduce this systematic effect (Ross et al. 2012; Bautista et al. 2018). We explain the standard treatment and how the systematic weight w_{systot} is obtained in Section 3.1. Collectively, to account for all of these observational effects, each quasar and random object must be weighted by,

$$w = w_{systot} \times w_{noz} \times w_{FKP} \times w_{cp}, \qquad (3)$$

where $w_{FKP} = [1 + n(z)P_0]^{-1}$ is the FKP weight (Feldman, Kaiser & Peacock 1994) with $P_0 = 6000\,Mpc^3h^{-3}$ based on the expected power[2] on scales around $0.01-0.3\,h\,Mpc^{-1}$, and $n(z)$ is the number density at redshift z. We require completeness >0.5 (for both parameters COMB_BOSS and SECTOR_SSR[3]) as well as $0.8 < z < 3.5$ to avoid low-quality samples. The quasar catalogue contains 343708 objects in the redshift range $0.8 < z < 2.2$ and 72 667 in the redshift interval $2.2 < z < 3.5$, covering an effective area of 4699 square degrees. Hereafter, we refer to the sample with $0.8 < z < 2.2$ as *main*, and the one with $2.2 < z < 3.5$ is referred to as *high-z*.

Fig. 1 shows the mean number density of quasars from the DR16 catalogue as a function of redshift for the NGC (solid blue) and SGC (dashed orange) regions. The main and high-z samples are separated by vertical dotted lines. Targeting efficiency varies slightly between

---

[1] https://github.com/londumas/redvsblue

[2] This is the default value for the BAO and RSD studies with eBOSS (see e.g. Neveux et al. 2020; Hou et al. 2021). The impact of the FKP weight on f_{NL} constraints is investigated in our companion paper (Mueller et al. in preparation).

[3] These parameters describes the survey completeness and the probability of detecting an object with a good redshift in spite of other instrumental factors.
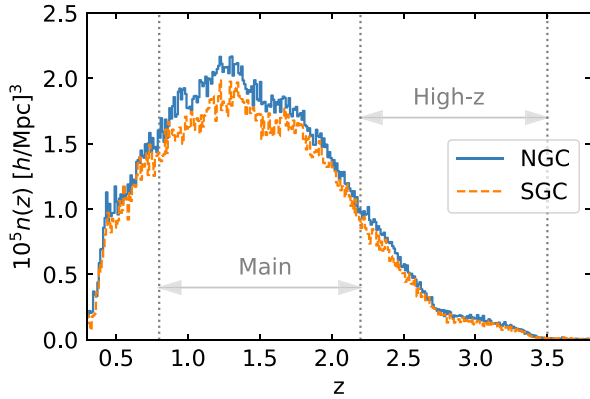
Figure 1. Mean density of the DR16 quasars as a function of redshift for the NGC (solid blue) and SGC (dashed orange) regions. The vertical dotted lines represent the redshift cuts for selecting the main and high-z samples.
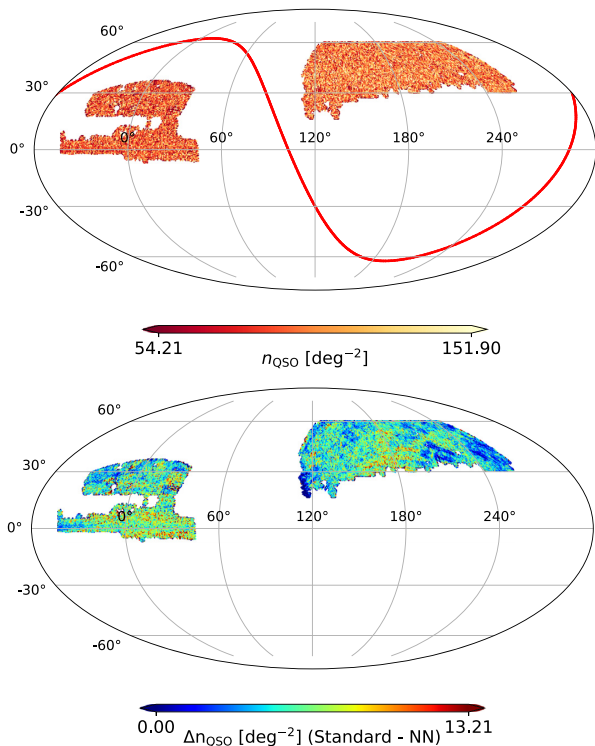

Figure 2. *Top*: Mollweide projection of the DR16 sample in the redshift interval of $0.8 < z < 3.5$ and the effective area of 4700 deg^2 with the standard treatment for imaging data systematics. The solid red curve represents the Galactic plane. *Bottom*: The difference in corrected quasar density between applying the standard or neural network treatments for imaging data systematics.

the two caps that causes small differences in $n(z)$. Fig. 2 shows the sky coverage of the DR16 quasars in equatorial coordinates. The top panel is the Mollweide projection[4] of the quasar density field in deg^{-2} after the standard treatment. The solid red curve represents the Milky Way plane. The effective area of the NGC and SGC regions are 2860 and 1839 deg^2, respectively (see table 3 in Ross et al. 2020).

----

[4]The density map is in HEALPIX (Gorski et al. 2005) with NSIDE = 128, which corresponds to a pixel area of 0.21 deg^2.

The quasar density in each pixel is corrected for pixel completeness, using the number of random objects. The bottom panel describes the residual quasar density between the standard and new neural network treatments. These methods are thoroughly described in Section 3.1. Interestingly, the residual map indicates that the size of the correction depends on angular direction, with the non-linear approach leading to more correction in particular regions.

2.2 Synthetic catalogues

We use synthetic catalogues (Zhao et al. 2021), generated using the extended Zeldovich approximation (Zel'Dovich 1970; Chuang et al. 2015), to construct covariance matrices of our statistics and perform robustness tests, characterizing the significance of residual systematic uncertainties. Throughout this manuscript, we refer to these mocks as the *EZmocks*. The EZmocks are tuned to reproduce accurate two-point clustering statistics, e.g. within 1 per cent of an N-body simulation on $k < 0.55 \, h \, \mathrm{Mpc}^{-1}$ and $r > 10 \, \mathrm{Mpc} \, h^{-1}$ (Chuang et al. 2015), and thus are suitable for studying the large-scale clustering of galaxies and quasars. These simulations are sufficient for this work since the impact of imaging systematics on quasar power spectrum measurements appears primarily on long-wavelength modes, e.g. $k < 0.01 \, h \, \mathrm{Mpc}^{-1}$. A detailed description for the creation of the eBOSS EZmocks is presented in Zhao et al. (2021).

An EZmock realization is created by adopting the Zeldovich approximation to generate a density field. Then, stochastic and parametric techniques are applied to simulate the scale-dependent non-linear biasing effects. A flat ΛCDM universe defined by cosmological parameters $h = 0.678$, $\Omega_M = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_b = 0.0482$, $\sigma_8 = 0.822$, and $n_s = 0.961$ is considered as the input cosmology for the EZmocks. The mock realizations do not contain primordial non-Gaussianity, i.e. $f_{NL} = 0$. Each mock catalogue is constructed by combining seven periodic boxes with the comoving side length $L = 5 \, \mathrm{Gpc}/h$ to mimic a light-cone geometry that accounts for the redshift evolution of quasars. Then, the mock catalogue is sub-sampled along the line of sight to simulate the redshift distributions of the DR16 quasar sample.

We use two sets of simulations in this paper. Each set contains roughly 1000 independent realizations per Galactic cap. We have the mocks only for the redshift interval $0.8 < z < 2.2$. One set of the mocks is manipulated to reflect the effects of observational systematics, including fibre collision, stellar contamination, redshift failures, and angular imaging systematics. While the other set lacks any observational systematics, except the survey geometry effect. Throughout this manuscript, we refer to these two sets as the *contaminated* or *null* mocks, indicating whether or not observational systematic effects are added to the simulations. In particular importance to our work, angular imaging systematics are simulated based on a linear multivariate fit to the DR16 quasar density against two imaging templates[5] for the SDSS stellar density and depth-g with HEALPIX resolution of NSIDE=512 (see 2.3). Similar to the DR16 sample, eBOSS pipelines are applied to the EZmock realizations to calculate and assign the appropriate attributes, which are intended to mitigate the simulated systematic effects and recover the ground truth clustering.

----

[5]This is a conservative approach for simulating imaging systematics in the mock catalogues since the treatment of the real sample has shown that more than two templates are required to achieve a desirable cleaning.

2.3 Imaging templates

In total, we have 17 templates of photometric variables, which are involved in the SDSS imaging and targeting, as potential sources of systematic uncertainties. These templates comprise three maps tracking the structure of the Milky Way: *Galactic extinction* (Schlegel, Finkbeiner & Davis 1998), *neutral hydrogen column density* (HI4PI Collaboration et al. 2016), and *stellar density* from either the SDSS program (as used in Bautista et al. 2018) or the *Gaia* spacecraft (Gaia Collaboration et al. 2018). The survey-specific templates are *seeing*, *sky brightness*, and *survey depth* in four bands (*griz*). The remaining two maps are *run* and *airmass*.

We produce these imaging templates in HEALPIX by using a catalogue of uniformly distributed random objects painted with imaging features. For each pixel, the average of each imaging property is computed over the random objects in the pixel. The maps are created in the ring ordering format and two resolutions, NSIDE = 256 and 512, respectively corresponding to 0.23 and 0.11 deg. Although the standard treatment uses templates with NSIDE = 512 for cleaning the eBOSS DR16 quasars, the preparation of templates in a different resolution, e.g. 256, will enable us to investigate how changing pixel size might influence the mitigation process and quasar clustering measurements.

A caveat of all template-based cleaning methods is that the available templates are not perfect, and this limitation might affect the performance of such treatments; for instance, the default stellar density map is an incomplete sample of stars, which is constructed as the number of type PSF objects with $i < 19.9$ from SDSS. Therefore, our proxy of stellar artefacts may be an incomplete representation of existing stellar contaminations in the DR16 sample. Indeed, as we discuss later in Section 4, we find a noticeable reduction in residual systematic error by swapping the SDSS map for a different stellar map from the *Gaia* DR2 with $12 < g < 17$ (Gaia Collaboration et al. 2018). The other caveat is that using all imaging maps for training might add the input noise to the neural network, increase chance cross-correlation between cosmological signal and imaging fluctuations, and reduce the effective number of modes. Therefore, we decide to follow a conservative approach to use a minimum number of templates in regression analysis. Later in Section 4, our results show that spurious fluctuations against imaging properties are alleviated by using only fives maps including *sky-i*, *seeing-i*, *extinction*, *depth-g*, and *Gaia stellar density*. Based on the χ^2 of the mean density residuals, the first four maps are identified as the primary sources of systematic error in the standard cleaning procedure (Ross et al. 2020), and hereafter we refer to this set as the *known* templates.

3 ANALYSIS TECHNIQUES

This section presents the techniques and statistics employed in this work for the characterization and mitigation of residual spurious fluctuations, and outlines the estimator for measuring the large-scale clustering of the DR16 sample and that of the EZmock realizations.

3.1 Template-based mitigation

3.1.1 Linear multivariate regression

The default systematic weights in the DR16 catalogue (Ross et al. 2020) are obtained based on a linear multivariate regression using the four known maps, i.e. *sky-i*, *seeing-i*, *extinction*, and *depth-g*, as the primary sources of systematic fluctuations. Linear regression is the most commonly used approach to reduce the effects of imaging

systematics in previous SDSS catalogues (e.g. Ross et al. 2011; Ross et al. 2012; Myers et al. 2015; Prakash et al. 2016; Ross et al. 2017). Despite some minor differences among the various implementations, a common assumption is that the observed number density of quasars in pixel i is a linear combination of imaging quantities in the pixel (Bautista et al. 2018, equation 6),

$$y_i = \bar{n} + \sum_j p_j x_{j,i}, \tag{4}$$

where $\bar{n}$ is the mean density of quasars, $x_{j,i}$ is the *j*'th imaging variable in pixel i, and p_j is the corresponding linear coefficient. These methods mostly differ either on their cost function[6] or on the number of templates used in the model. For instance, the least-squares error is evaluated over either the binned or pixelated mean density of quasars. Sometimes, the regression routine fits against all templates simultaneously or one template at a time in each iteration.

Specifically, the method presented in Bautista et al. (2018) implements a simultaneous fit using the four known maps as the independent variables in the model and chooses the least-squares error on the binned quasar density as the cost function. A subtle aspect of the cost function is that the residual fluctuations are summed over not only the four known maps but also the SDSS stellar density and airmass. This design will train the model parameters to explain the trends against the six templates while providing only the four known maps as input. Given the four known maps as the input variable x, the systematic weight of quasars in pixel i is defined as, $w_{\mathrm{systot},i} = 1/(c + \sum_j p_j x_{j,i})$, where c is the intercept. The optimal coefficients and intercept c are then derived by minimizing the least-squares error[7] on the binned quasar density after applying the systematic weight w_{systot} to each quasar object. The binned density is accounted for other observational and pixel completeness effects by weighting the quasars and randoms, respectively, by $w_{\mathrm{tot,g}}$ and $w_{\mathrm{tot,r}}$ (see equation 9).

3.1.2 Non-linear approximation using neural networks

The DR16 quasar sample is very sparse, with a surface density of $73\,\mathrm{deg}^{-2}$ in the redshift range $0.8 < z < 2.2$ (main) and $15\,\mathrm{deg}^{-2}$ in $2.2 < z < 3.5$ (high-*z*). The high sparsity poses a major challenge for regression-based treatments by making it difficult to disentangle the effect of systematics and the cosmological signal. This paper expands the neural network-based methodology described in Rezaie et al. (2020) to obtain non-linear systematic weights. In what follows, we first describe the specifics of the neural network architecture, and then elaborate on how we further improve the methodology specifically with the implementation of the Poisson statistics in the cost function to properly deal with the high sparsity eBOSS QSO data and a cyclic learning rate to enhance training against local minima.

Feed forwards neural networks: Neural networks are universal approximators that can model a wide variety of non-linear mappings from a set of independent variables to a target variable. Neural networks can account for non-linear spurious fluctuations caused by imaging and the cross-correlation among imaging properties. We use y_i and x_i to, respectively, denote the observed quasar count and

---

[6]Cost function is used to find the best parameters of a model.

[7]It is worth noting that unlike the neural network method, this approach uses all data for training and does not apply any form of training, validation, and testing. We believe linear regression is not in a limit prone to over-fitting, i.e. the degree of freedom is much smaller than the number of data points.

imaging features[8] in pixel i. We implement a fully connected feed-forward neural network to model the observed quasar count y_i from imaging templates as input features $\boldsymbol{x}_i$. At its standard configuration, a fully connected feed-forward neural network is a system of neurons organized in a series of interconnected layers, where each neuron is connected to all neurons in the previous layer and the following layer.[9] For instance, the output value of neuron μ in layer $l + 1$, a_μ^{l+1}, is obtained from the linear combination of the output values from the previous layer neurons, a_ν^l, after applying a non-linear activation function f,

$$a_\mu^{l+1} = \psi \left(b_\mu^l + \sum_\nu w_{\mu\nu}^l a_\nu^l \right), \tag{5}$$

where $w_{\mu\nu}^l$ and b_μ^l are, respectively, the associated weights and bias connecting to neuron μ. The non-linear function ψ determines how much of the signal travels from layer l to $l + 1$. In this notation, the first layer input values are imaging properties x_i and the last layer output is then compared to y_i. The network architecture comprises three fully connected hidden layers[10] with twenty neurons with the rectified linear activation function, $\psi(u) = \max(0, u)$, on each hidden layer and a single neuron with the softplus activation function, $\text{Softplus}(u) = \log[1 + \exp(u)]$, on the output layer. The softplus activation in the last layer will ensure the network output is always positive. We also add a batch normalization layer after each layer, except the output layer. Batch normalization stabilizes and expedites training by scaling the inputs to each layer to have a mean of zero and a standard deviation of one (e.g. Ioffe & Szegedy 2015).

Poisson statistics in the cost function: For the DR16 quasars, we decide to use the negative Poisson log-likelihood as the default cost function and the Mean Squared Error for benchmarking. Assuming the quasar counts per pixel are independent and identically distributed variables, we can write the joint probability for N pixels as

$$L = f(y_1, ..., y_N | \theta) = \prod_{i=1}^N f(y_i | \theta, \boldsymbol{x}_i), \tag{6}$$

which is a reasonable assumption provided that the input features $\boldsymbol{x}_i$ do not include any spatial information, and we do not intend to provide any feature that contains cosmological clustering in our modelling. Otherwise, the mitigation procedure will learn the clustering signal and remove it as well. The objective is then to find the best set of parameters θ that maximizes the likelihood L, or minimizes its negative logarithm $-\log L$. Assuming y_i follows a Poisson distribution, we have

$$f(y_i | \theta, x_i) = \frac{\lambda(\theta, \boldsymbol{x}_i)^{y_i} \exp^{-\lambda(\theta, \mathbf{x}_i)}}{y_i!}, \tag{7}$$

where λ is the expected number of quasars, $\lambda > 0$ by definition, and a function of imaging features. We use $\lambda_i \equiv \lambda(\theta, \boldsymbol{x}_i)$ for brevity and

obtain the negative log-likelihood[11] as the cost function J,

$$J \equiv -\log(L) = \sum_{i=1}^N [\lambda_i - y_i \log(\lambda_i)]. \tag{8}$$

In this notation, the quasar number count y_i is assumed to be an integer value. This assumption breaks down in practice since each object in the DR16 catalogue needs to be weighted for other observational systematics and pixel completeness beforehand, and that turns y_i into a non-integer value. To circumvent this issue and still account for the pixel completeness and other observational effects, we decide to use the raw number count of quasars for y_i, and instead weight the model prediction in pixel i, i.e. λ_i, by the ratio of the weighted number of randoms to that of quasars in that pixel. To this end, the quasar and random objects are, respectively, weighted by $w_{\text{tot, g}}$ and $w_{\text{tot, r}}$,

$$w_{\text{tot,g}} = w_{\text{noz}} \times w_{\text{cp}} \times w_{\text{FKP}},$$
$$w_{\text{tot,r}} = w_{\text{FKP}} \times \text{comp}_{\text{BOSS}}, \tag{9}$$

where $\text{comp}_{\text{BOSS}}$ is the survey completeness that describes the likelihood of obtaining a good redshift, regardless of any other instrumental deficiencies.

Cyclic learning rate: The weights $w_{\mu\nu}^l$ and biases b_μ^l of the neural network are trained using the Decoupled Weight Decay Regularization optimizer (AdamW; Loshchilov & Hutter 2017), which is an iterative gradient descent approach for optimization. Specifically, the parameters are updated in the opposite direction of the gradient of the cost function with respect to the parameters,

$$m_{t+1} = \beta_1 m_t + (1 - \beta_1) \nabla J(\theta_t), \tag{10}$$

$$v_{t+1} = \beta_2 v_t + (1 - \beta_2) [\nabla J(\theta_t)]^2, \tag{11}$$

$$\theta_{t+1} = \theta_t - \eta_t \frac{m_{t+1}}{\sqrt{v_{t+1}} + \epsilon}, \tag{12}$$

where $\epsilon = 10^{-8}$ and the learning rate η_t controls the magnitude of each parameter update per iteration t. The first and second moments of the gradients, i.e. m_t and v_t, are initialized as zero. The parameters β_1 and β_2 determine the average history of the first and second moments of the gradients, and are fixed at 0.9 and 0.999, respectively (see e.g. Ruder 2016, for a review of gradient descent methods).

We incorporate a cyclic learning rate to prevent the optimizer from being trapped in a local minimum. Specifically, the learning rate η at epoch t is scheduled to vary following the method presented in (Loshchilov & Hutter 2016),

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left[\cos \left(\frac{T_{\text{cur}}}{T_i} \pi \right) + 1 \right], \tag{13}$$

where $\eta_{\max}$ is the initial (maximum) learning rate, $\eta_{\min}$ is the minimum learning rate, T_{cur} is the number of epochs since the last restart, and T_i is the number of epochs between two subsequent restarts. The neural network training begins with $T_{i=0} = 10$, but then we increase T_i by a factor of 2 after each restart (i.e. T_i changes like 10, 20, 40, ...). We follow the procedure presented in Smith (2015) to search an exponential grid[12] for the optimal values of $\eta_{\min}$ and $\eta_{\max}$. The optimal values are chosen such that the variation of loss per training epoch is maximized. We separately perform the learning rate finding procedure for each Galactic cap, redshift

---

[8]We use the z-score normalization scheme to standardize our imaging templates, i.e. $z = (x - \mu)/\sigma$ where μ and σ are the mean and standard deviation of imaging feature x determined from the training set.

[9]Bias neurons are an exception and connect only to the subsequent layer.

[10]We follow a grid search approach to experiment with a various number of hidden layers ranging from two up to five hidden layers; however, we do not observe any significant change in validation loss. Therefore, we fix the architecture at three hidden layers with 20 units on each hidden layer.

[11]We omit the term $\log(y_i!)$ from our objective function since it does not depend on θ.

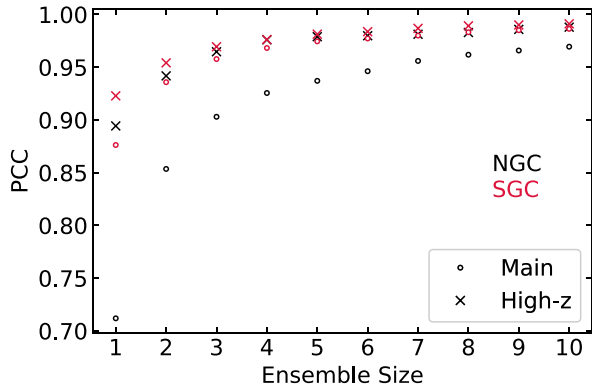[12]github.com/davidtvs/pytorch-lr-finder; github.com/fastai/fastai

Figure 3. The Pearson correlation coefficient between the averaged predicted quasar counts from neural networks for various ensemble size in the NGC (black) and SGC (red) regions, and the main and high-z samples. The correlation increases beyond 90 per cent after averaging over three networks.

selection, template resolution since the balance between systematics and noise changes.

Cross-validation: We apply fivefold cross-validation to perform training, validation, and testing the network on the entire footprint, while ensuring no overlap among any of the training, validation, or testing sets. Specifically, we randomly split the entire footprint into 60 per cent training, 20 per cent validation, and 20 per cent testing sets. The training set is used to compute the gradients and update the parameters. In each epoch, the model is applied to the validation set to assess the prediction error when the model is applied on unseen data. We let the networks to train for 150 epochs, and finally we apply the model with the lowest validation error on the test set. By changing the arrangement of the training and validation sets, the neural network will be tested on the entire footprint. Similar to Ross et al. (2020), we perform training, validation, and testing for the main and high-z samples in the NGC and SGC regions separately and finally aggregate the results. This approach assumes that the effects of imaging systematics along the line-of-sight do not vary much. We relax this assumption later by further splitting the main sample into $0.8 < z < 1.5$ and $1.5 < z < 2.2$, and analysing each subsample separately. As shown later in Section 4, we find no significant difference. Therefore, we leave an extensive 3D regression of target density and imaging properties for future work.

Unlike linear regression, the objective function of a neural network is non-convex, and thus there is no unique set of best-fitting parameters. This results in an intrinsic scatter on the predicted quasar density from a single neural network. We create an ensemble of 20 networks, through random initializations of the parameters, to mitigate the issue of model uncertainty, reduce the dispersion of the predicted quasar counts, and increase the stability of the selection function. We then take the mean predicted number of quasars across all networks in the ensemble as the final selection function. Fig. 3 illustrates the Pearson correlation coefficient (PCC[13]) between the predicted quasar counts of two independent ensemble subsets as a function of the ensemble size for the main and high-z samples in the NGC (black) and SGC (red) regions. The Pearson coefficient

illustrates that the mean predicted densities between two ensembles, each contains only four or more networks, correlate more than 90 per cent. Then, the systematic weight for quasars in pixel i is defined by $w_{\mathrm{systot}} \propto 1/\lambda(\theta, \boldsymbol{x}_i)$, where w_{systot} is normalized such that the total weighted number of quasars stays the same as before treatment.[14] The figure shows that with our default ensemble size of 20, we are introducing very little intrinsic scatter in the process of the neural network.

3.2 Residual systematics tests

3.2.1 Mean density contrast

This test is sensitive to the uniformity of the DR16 sample by clustering pixels with similar imaging properties. We compute the mean density contrast of quasars against each imaging variable to quantify the deviations and fluctuations in the sample before and after cleaning. For a given imaging quantity x_j, the mean density contrast δ is,

$$\delta(x_j) = \frac{\sum_i n_{g,i}}{\alpha \sum_i n_{r,i}} - 1, \tag{14}$$

where $n_{g,i}$ and $n_{r,i}$ are the weighted number of quasars and randoms in pixel i; α is the factor to normalize the number of randoms to that of quasars; and the summation $\sum_i$ is computed over pixels with $x_{j,i} \in [x_j, x_j + \Delta x_j]$. We adopt an equal frequency binning scheme in which pixels are initially sorted from the minimum to maximum imaging value, and then the width Δx_j is varied such that each bin contains the same area, i.e. the same number of randoms. By design, this statistic is not sensitive to the underlying true power spectrum, since the mean of density contrast is computed over a large number of pixels in the imaging phase space. Therefore, the cosmological component in δ cancels out.

The quasar and random objects are weighted differently before and after mitigation; initially, the quasars are weighted by $w_{\mathrm{tot,g}}$ and the randoms are weighted by $w_{\mathrm{tot,r}}$, equation (9). After mitigation, we use equation (3) to weight both the quasars and randoms. Finally, we compute the mean density contrast against all 17 templates, $\delta = [\delta(x_1), \delta(x_2), ..., \delta(x_{17})]$, and quantify the total mean density residuals with χ^2 statistics,

$$\chi^2_{\mathrm{tot}} = \delta^\dagger C^{-1} \delta, \tag{15}$$

where the covariance matrix C is estimated from the null EZmock mean density contrasts, $C = <\delta^\dagger \delta>$, where the angle brackets represent ensemble average over the EZmock realizations. The estimated covariance matrix is then unbiased by the Hartlap factor (Hartlap, Simon & Schneider 2007),

$$C = \frac{N_{\mathrm{mocks}} - 1}{N_{\mathrm{mocks}} - N_{\mathrm{bins}} - 2} C_{\mathrm{biased}}, \tag{16}$$

where N_{mocks} is the number of mock realizations, e.g. 1000 for the NGC and 999 for the SGC, and N_{bins} is the number of imaging bins, i.e. 136. Fig. 4 shows the estimated covariance matrix of the mean density contrast from the 1000 null EZmock realizations for the NGC. This matrix shows the complex correlation among the attributes. This

---

[13]$\mathrm{PCC}(X, Y) = \mathrm{cov}(X, Y)/\sqrt{\mathrm{cov}(X, X)\mathrm{cov}(Y, Y)}$, where $\mathrm{cov}(X, Y)$ is the covariance between variable X and variable Y.

[14]We also limit the systematic weights to $0.5 < w_{\mathrm{systot}} < 2.0$ to avoid extreme corrections. For pixels where we do not have imaging information, we use the mean value over nearest neighbours to estimate the systematic weight. We do not observe a significant impact on quasar clustering.
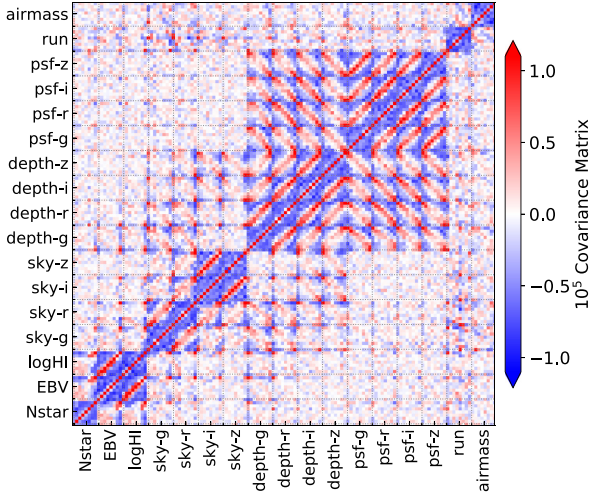
Figure 4. Covariance matrix of the mean density contrast for the NGC region estimated from the null EZmock realizations.

covariance matrix will be used to assign the error-bars presented in Fig. 8 and obtain the χ^2 values presented in Fig. 9

3.2.2 Angular cross-correlation

The anisotropies in target quasar density of a pure cosmological signal should not correlate with Galactic foregrounds and survey-related properties. To test this assumption, we measure the cross-power spectrum between quasar density and imaging templates. Due to imaging systematics, the cross-power spectrum measurements are not zero, and we use this tool to quantify spurious fluctuations. The cross-spectrum is squared and then normalized by the autopower spectrum of the imaging map,

$$\tilde{C}_\ell^{q,x} = (C_\ell^{q,x})^2 / C_\ell^{x,x}, \tag{17}$$

to convert it to the estimated contribution of systematics up to the first order to the autopower spectrum of the quasar density. The autopower or cross-power spectrum is computed over the coefficients $a_{\ell m}$ from spherical harmonic decomposition (e.g. Hobson 1931),

$$C_\ell^{p,q} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m}^p a_{\ell m}^q, \tag{18}$$

with $p = q$ for autocorrelation and $p \neq q$ for cross-correlation. We bin the cross-correlation and autocorrelation, and use only the four lowest ℓ bins centered at $\ell = 1.6, 3.7, 8.3$, and 18.8, as we are only interested in large-scale modes[15] around $k < 0.01\,h\,\mathrm{Mpc}^{-1}$. Similar to the mean density contrast diagnostic, the covariance matrix C is constructed from the null EZmock cross-power spectra, and then unbiased by the Hartlap factor. Finally, we quantify the significance of the residual cross-power against the 17 imaging maps by

$$\chi_{\mathrm{tot}}^2 = \tilde{C}_\ell^\dagger C^{-1} \tilde{C}_\ell, \tag{19}$$

where

$$\tilde{C}_\ell = \left[\tilde{C}_\ell^{q,x_1}, \tilde{C}_\ell^{q,x_2}, ..., \tilde{C}_\ell^{q,x_{17}} \right]. \tag{20}$$

[15]We use the approximation $\ell + 1/2 \sim kD_A$ with D_A being the angular size distance.

3.3 Power-spectrum multipoles

We use the estimator presented in Hand et al. (2017) and implemented in NBODYKIT[16] (Hand et al. 2018) to measure the power spectrum of the DR16 sample for the NGC and SGC regions, separately. We provide a summary of the algorithm below.

We assume a flat ΛCDM cosmology with $\Omega_M = 0.31$, $h = 0.676$, $\Omega_b h^2 = 0.022$, $\sigma_8 = 0.8$, and $n_s = 0.97$ (Alam et al. 2017) to convert each redshift to distance. After transforming the coordinates, we first begin with digitizing the quasar and random catalogues over a 3D cubic box with a side length of 6600 Mpc h^{-1} and 512^3 voxels. Then, we construct the Feldman–Kaiser–Peacock field (FKP; Feldman et al. 1994) and interpolate it using the Triangular Shaped Cloud (TSC) scheme,

$$F(\boldsymbol{r}) = n_g(\boldsymbol{r}) - \alpha n_r(\boldsymbol{r}), \tag{21}$$

where $\alpha = \sum n_g(\boldsymbol{r}) / \sum n_r(\boldsymbol{r})$. The terms n_g and n_r represent the observed density of the quasar sample and random objects weighted by equation (3), respectively. The power spectrum multipoles are then computed from the FKP field as (Yamamoto et al. 2006)

$$P_\ell(k) = \frac{2\ell+1}{4\pi A} \iiint d\Omega d\boldsymbol{r}_1 d\boldsymbol{r}_2 F(\boldsymbol{r}_1) F(\boldsymbol{r}_2) e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}_2-\boldsymbol{r}_1)} \mathcal{L}_\ell(\boldsymbol{k}\cdot\hat{\boldsymbol{r}}_h), \tag{22}$$

where $\boldsymbol{r}_h = (\boldsymbol{r}_1 + \boldsymbol{r}_2)/2$ is the line-of-sight distance to the middle point of a given pair, $\mathcal{L}$ is the Legendre polynomial of order ℓ, and $A = \int d\boldsymbol{r} \bar{n}^2(\boldsymbol{r})$ is the normalization for the field F, which can be approximated by a discrete sum over the synthetic objects weighted by w_r (see equation 3)

$$A = \alpha \sum_i w_r(\boldsymbol{r}_i) n_r(\boldsymbol{r}_i). \tag{23}$$

Equation (22) can be simplified even further. Under the small angle approximation, $\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}_h \sim \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}_2$, the two integrals on r_1 and r_2 are separable. Using the decomposition of the Legendre function into spherical harmonics,

$$\mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{r}}) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\hat{\boldsymbol{k}}) Y_{\ell m}^*(\hat{\boldsymbol{r}}), \tag{24}$$

we obtain,

$$P_\ell(k) = \frac{2\ell+1}{A} \int \frac{d\Omega}{4\pi} F_0(\boldsymbol{k}) F_\ell(-\boldsymbol{k}), \tag{25}$$

where

$$F_\ell(\boldsymbol{k}) = \int d\boldsymbol{r} F(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \mathcal{L}_\ell(\hat{\boldsymbol{k}}.\hat{\boldsymbol{r}}) \tag{26}$$

$$= \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\hat{\boldsymbol{k}}) \int d\boldsymbol{r} F(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} Y_{\ell m}^*(\hat{\boldsymbol{r}}). \tag{27}$$

We measure the power spectrum using equation (25) in k bands of width $\Delta k = 0.0019\,h\mathrm{Mpc}^{-1}$. This method requires less number of Fourier's transforms compared to other estimators in Bianchi et al. (2015) and Scoccimarro (2015). When computing the transforms, the NBDOYKIT software mitigates aliasing with the interlace grid technique presented in Sefusatti et al. (2016) and accounts for any effects caused by the TSC gridding using the factor presented in Jing (2005). Finally, we estimate the shotnoise by summing over the

[16]https://nbodykit.readthedocs.io

quasar and random catalogue objects,

$$P_{\text{shotnoise}} = \frac{1}{A}\left[\sum_i w_g^2(\boldsymbol{r}_i) + \alpha^2 \sum_j w_r^2(\boldsymbol{r}_j)\right], \qquad (28)$$

and subtract it only from the monopole power spectrum ($\ell = 0$).

4 RESULTS

This section presents the analyses for the characterization of systematic error and the treatment of spurious fluctuations with various cleaning approaches. We compare the measured power spectra before and after accounting for imaging systematic effects with various linear and non-linear cleaning methods. In the end, we present the clustering analysis of the EZmock realizations in order to assess the impact of systematic treatments on the measured power spectrum and how one can account for such effect in theoretical modelling.

4.1 Method benchmarks

4.1.1 Linear versus non-linear treatment

Our series of tests begin with applying various linear and nonlinear cleaning methods to the main sample of quasars in the NGC region. To conduct a fair comparison, the same set of the four known maps are employed as the input templates. We use the mean-squared error and the Poisson negative log-likelihood as two alternatives for the cost function. When using PNLL as the cost function, *Softplus* is applied to the output to satisfy the boundary condition $\lambda > 0$. We also experiment with a non-linear model without the learning rate annealing to test the sensitivity of training to local minima.

Fig. 5 shows the mean density contrast as a function of Galactic extinction in the top panel and the measured power spectrum in the bottom panel, after applying different cleaning methods. The 1σ uncertainty is shown via the grey shades and is constructed from the null EZmock realizations. Linear-MSE here is equivalent to the standard treatment except that the cost function of the former is based on the pixelated density, while the latter is based on the binned density. This figure illustrates that our implementations of linear-MSE and linear-PNLL yield similar residual fluctuations to that of the standard treatment. This result implies that with linear regression, adopting the Poisson statistics does not help. Interestingly, the residual fluctuations are reduced substantially after accounting for non-linearities by the hidden layers in the *NN-MSE* and *NN-PNLL* architectures. We also find that turning off the learning rate cycling does not hugely impact the neural network performance (*NN-PNLL-lr*) with marginal effect on the power spectrum, which proves the robustness of the pipeline against local minima and saddle points. This test shows that the most substantial improvement in the measured power spectrum is enabled by accounting for nonlinear systematic effects using the neural network-based methods and then by adopting PNLL in the cost function. Interestingly, our linear cleaning method yields a lower clustering power than the standard linear approach. This test demonstrates that cost function needs to be defined accurately for different levels of model complexity; the choice of PNLL over MSE makes more substantial difference for the NN approach compared with the linear model.

4.1.2 Stellar density from Gaia DR2

The quality of the selection function derived from a template-based cleaning method relies on the available input templates. We may
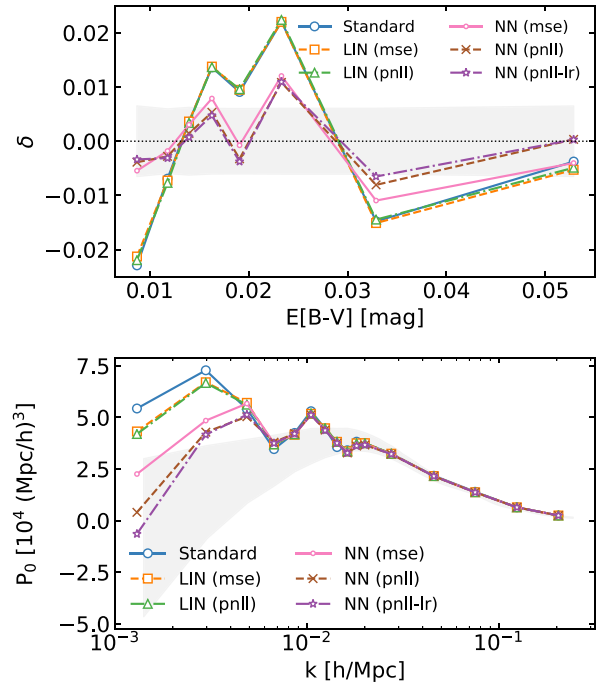

Figure 5. *Top*: Mean density contrast of the NGC main sample against Galactic extinction after applying various mitigation techniques, including linear with MSE, linear with PNLL, NN with MSE, NN with PNLL, and NN with PNLL but without learning rate annealing (NN-PNLL-lr). All methods use the *known* set of imaging maps. *Bottom*: Monopole of the NGC main sample after applying the same techniques. The shades represent 1σ statistical uncertainty constructed from the null EZmock realizations.

fail to properly eliminate spurious fluctuations if a primary imaging map is missed or our available imaging maps are not completely representing the underlying systematic effects. In this test, we experiment with the available SDSS and non-SDSS imaging maps to find the optimal number of crucial imaging maps, which one would need to explain the residual trends in the mean density contrast. As mentioned before, the standard cleaning approach uses only four maps, i.e. Galactic extinction, depth in *g*-band, sky brightness in *i* band, and seeing in *i* band.

Fig. 6 top panel shows the mean density contrast against the *Gaia* stellar density from *Gaia* Collaboration et al. (2018) after training a neural network with various combinations of the imaging maps using PNLL as the cost function. We also plot the measured density contrast after the standard treatment (*standard*) as our reference for comparison. The measured monopole power spectrum is shown in the bottom panel. This exercise illustrates that the four known maps are not sufficient to eliminate the systematic trend in the mean density against the *Gaia* stellar map. We also show that incorporating the SDSS stellar map (*NN known + sdss*) and even all of the SDSS maps (*NN all sdss*) are not adequate to obtain satisfactory cleaning. On the other hand, we note that *NN all sdss* performs as well as *Lin Gaia*, implying that the nonlinear nature of NN can mitigate the effect of the missing input to some extent. Interestingly, adding the SDSS stellar map to the four known maps impacts the result adversely, especially on the low density end of the *Gaia* DR2 stars. This result might mean that the SDSS stellar density map is not a proper proxy of the stellar contamination effect in the regions with a lower stellar density. This test indicates that the *Gaia* DR2 stellar map is proved pivotal to perform a robust cleaning of data, that is consistent with the statistical tests of the mocks. For comparison, we apply the linear
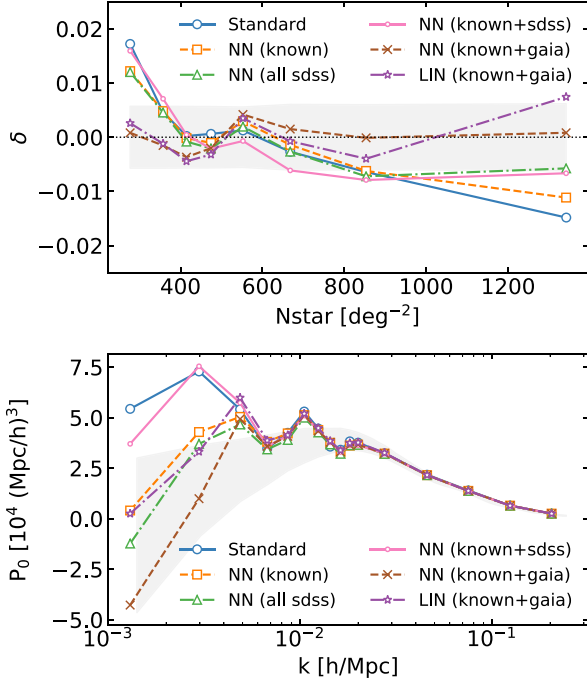
Figure 6. *Top*: Mean density contrast of the NGC main sample against the *Gaia* stellar density after accounting for systematics, using different combinations of imaging templates. *Bottom*: Monopole power spectrum of the NGC main sample for the same techniques. The shades represent 1σ statistical uncertainty constructed from the null EZmock realizations.

model with MSE and the additional *Gaia* map (*LIN known + gaia*), and obtain a total chi-squared value of $\chi^2 = 196.9$ which is still significant (see Table 1). This test demonstrates that capabilities of linear treatment is limited even after including the Gaia map, and the nonlinear treatment is crucial, not only when we know the proper set of the input templates a priori but also when we do not know.

4.1.3 Redshift slicing and pixel resolution

The effects of imaging systematics on target density might vary along the line of sight, and alter the slope of the redshift distribution of quasars as well as its overall magnitude. This effect can be ideally investigated by slicing the sample into smaller redshift bins to construct a 3D selection function of imaging systematics. However, further splitting the sample increases the sparsity and makes it
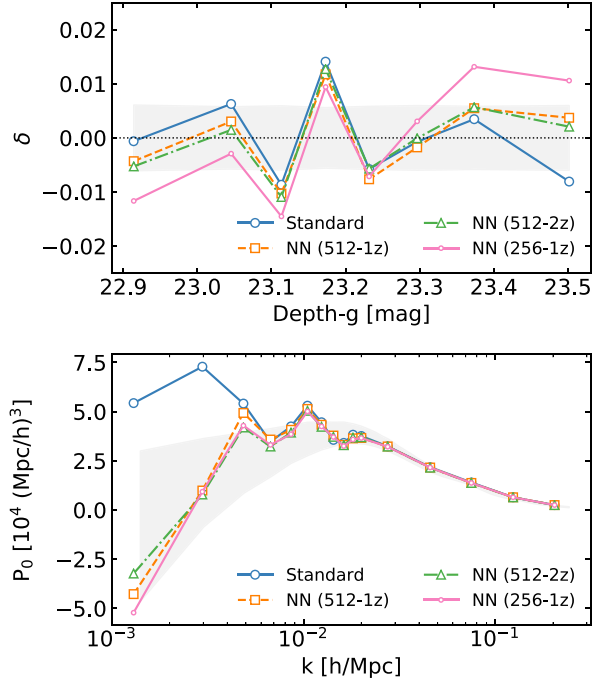
Figure 7. *Top*: Mean density contrast of the NGC main sample against depth in the *g*-band after training the neural network methods with two redshift slices with NSIDE = 512 (*NN-512-2z*), one redshift slice with NSIDE = 256 (*NN-256-1z*), and one redshift slice with NSIDE = 512 (*NN-512-1z*). *Bottom*: Monopole power spectrum after the same techniques. The shades represent 1σ statistical uncertainty constructed from the null EZmock realizations.

difficult to separate the effects of imaging systematics and noise. Thus, we begin with splitting the main sample in the NGC with NSIDE = 512 into $0.8 < z < 1.5$ and $1.5 < z < 2.2$ and then we perform regression on each slice separately. We use PNLL as the cost function with the four known maps and the *Gaia* stellar map as input. Fig. 7 compares the mean density contrast as a function of depth in g band (*top*) and the measured monopole power spectrum (*bottom*) after this treatment (*NN 512-2z*) with that of the standard method and the neural network, trained on $0.8 < z < 2.2$ with NSIDE = 512 (*NN 512-1z*). This plot demonstrates that there is no evidence for redshift-dependent systematic effects due to imaging.

We also train the neural network with coarser imaging templates in NSIDE = 256 on $0.8 < z < 2.2$ (*NN 256-1z*). The mean density contrast and measured power spectrum after this treatment are shown in Fig. 7

Table 1. Total χ^2 of the mean density residuals of the main sample in the NGC after various mitigation configurations. The chi-squared value before accounting for imaging systematics is 1344.9, which is much larger than the 95th quantile observed in the null EZmocks, i.e. 178.

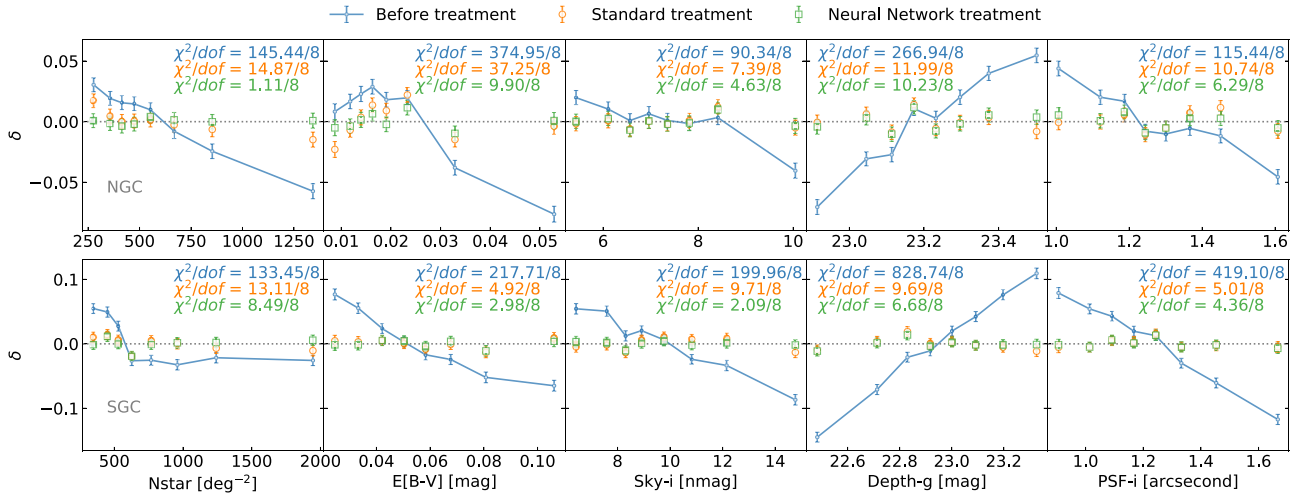|			Templates		
NSIDE-*Split*		*known*	*Known + SDSS*	*All SDSS*	*Known + Gaia*
512-1*z*	standard	218.1	–	–	–
	linear-mse	213.5	–	–	196.9
	linear-pnll	210.2	–	–	–
	nn-mse	194.6	–	–	–
	nn-pnll-lr	**168.99**	–	–	–
	nn-pnll	**163.97**	184.6	153.9	**151.7**
512-2*z*	nn-pnll	–	–	–	**165.5**
256-1*z*	nn-pnll	–	–	–	217.6

Figure 8. Density contrast of the main sample as a function of the primary imaging properties in the NGC (top) and SGC (bottom) before and after accounting for systematic effects using the standard method or neural network. The error-bars are estimated from the null EZmocks and used to calculate the χ^2 of the mean density residuals against each imaging property.

in the top and bottom panels, respectively. The mean density contrast histogram is computed against the depth template in NSIDE = 512 and indicates noticeable residual variations from zero in the regions at low and high *depth-g*. This test demonstrates that the systematic weights obtained in the lower resolution cannot properly reduce the trends in the higher resolution. As a caveat, if the top panel was drawn against the depth template in NSIDE = 256, we would have gotten a reasonable performance of NN-256. The bottom figure shows the resulting power spectrum accounting for the effects in all 17 imaging maps. We find a reasonable stability against the varied pixel resolutions and choose NN-512 as our default.

The total χ^2 of the mean density residuals for various systematic treatment methods are summarized in Table 1. The χ^2 value is 1344.9 before correcting for systematic effects. After using the linear cleaning techniques, the χ^2 value drops below 220. The non-linear approach lowers the error below 200, and changing the cost function from MSE to PNLL improves the performance even further by returning a value around 170, which is less than what is observed in the null EZmock realizations. As a comparison, purely cosmological signal without systematics, estimated from the null EZmocks, returns the χ^2 value of 178.

4.2 Significance of residual fluctuations

For testing the significance of residual systematics, and as our default NN approach, we focus on the neural network method trained with PNLL and cyclic learning rate on the DR16 sample covering $0.8 < z < 2.2$ with the five imaging maps as input features. From the previous analyses, see Table 1, we identify this configuration as the optimal approach.

Fig. 8 illustrates the observed density contrast of the DR16 quasars as a function of imaging properties for the NGC (top) and SGC (bottom). Respectively from left to right, the imaging quantities are the *Gaia* DR2 stellar density, Galactic extinction, sky brightness in *i* band, depth in *g* band, and seeing in *i* band. Each panel is annotated with the residual squared errors that are calculated against a zero model as the ground truth.[17] The error-bars are obtained from the

EZmock catalogues. We observe the biggest variation is against the extinction for about 8 per cent with $\chi^2/\text{dof} = 374.95/8$ in the NGC and 15 per cent against depth-g with $\chi^2/\text{dof} = 828.74/8$ in the SGC. Interestingly, the neural network treatment is capable of modelling and removing the non-linear effects in the sample. On the other hand, the standard treatment leaves a significant chi-squared value against the extinction with $\chi^2/\text{dof} = 37.25/8$ in the NGC.

We compute the total χ^2 of the mean density residuals against all of the 17 imaging maps to determine the significance of spurious fluctuations in the observed mean density of quasars. Fig. 9 shows the distributions of χ^2_{tot}, which are constructed from the null (Null) and contaminated EZmocks (Cont), before and after applying imaging systematics mitigation for the NGC (left) and SGC (right). The values observed in the DR16 sample before and after cleaning are represented with vertical lines. We use the distribution of the null mocks (Null Truth) to compute *p*-value. In the NGC, the standard treatment yields $\chi^2 = 218.1$ with *p*-value = 0.2 per cent, while the neural network treatment cleans the sample substantially and returns a smaller χ^2 and higher *p*-value, respectively, 151.7 and 27.1 per cent. In the SGC, we observe that the standard method returns 132.5 with *p*-value = 65.0 per cent. This residual is somewhat expected since the trends against imaging maps in the SGC are mostly linear, and thus a linear model is sufficient for cleaning (see Fig. 8). In the SGC, both methods return statistics that are in agreement with the χ^2 distribution of the null mocks. No remaining systematic error is observed within the statistical uncertainty of the mocks. The total χ^2 of the mean density residuals for the main and high-z samples are reported in Table 2. The 95th percentile for the main sample is estimated from the mocks and reported in the last column. Note that for the NGC region, the standard approach yields a χ^2 value that is larger than the 95th quantile of the mocks.

Similarly, we cross-correlate the quasar density map with all of the 17 imaging templates. The cross-correlations are then binned to decrease statistical fluctuations, and normalized by the auto power spectrum of the imaging maps. The first four bins are then used to compute the residual squared error against zero. Fig. 10 demonstrates the distribution of χ^2_{tot} constructed from the null and contaminated mocks before and after applying imaging systematics mitigation for the NGC (left) and SGC (right). The values observed in the DR16 sample are represented with vertical lines. In the NGC, the standard treatment returns $\chi^2 = 338.5$ with *p*-value = 1.1 per cent,

---

[17]We assume that the density contrast must be zero when averaged over many pixels in the absence of imaging systematics.
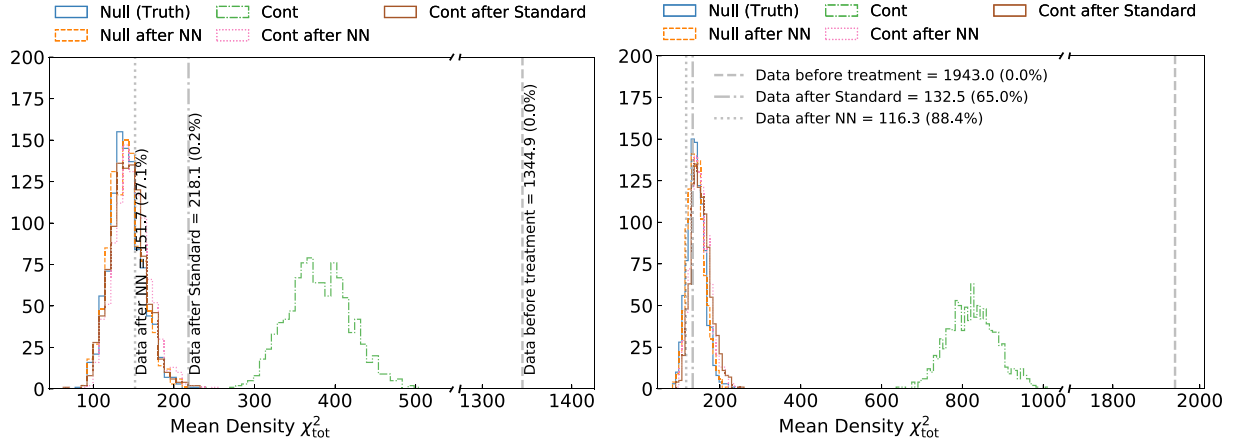
Figure 9. Total χ^2 of the mean density residuals for the EZMocks with systematics (Cont) and without systematics (Null) for the NGC (left) and SGC (right). The same statistics observed in the eBOSS DR16 quasar sample (before and after treatment) are shown via vertical lines with the associated p-values, which are derived by comparing with the mocks without systematics, Null (Truth). There is substantial remaining systematics in the NGC with the standard linear treatment. Also, the contaminated simulations do not reflect the same level of systematic effects as the DR16 sample.

Table 2. Total χ^2 of the mean density residuals for the main and high-z samples, before and after mitigation, and the 95th percentile of null EZmocks. The null EZmock covariance matrix is used to calculate these statistics.

		Noweight	Standard	NN	95th percentile
NGC	$0.8 < z < 2.2$	1344.9	218.1	151.7	178.0
	$2.2 < z < 3.5$	1752.0	121.6	104.2	–
SGC	$0.8 < z < 2.2$	1943.0	132.5	116.3	179.2
	$2.2 < z < 3.5$	2553.7	146.0	130.1	–

while the neural network treatment provides a cleaner sample with $\chi2 = 49.8$ and p-value $= 57.9$, respectively. In the SGC, we observe that the standard method is incapable of removing the systematics by returning $\chi^2 = 404.3$ with p-value $= 0.7$ per cent. On the other hand, the neural network approach enables rigorous cleaning with $\chi^2 = 53.2$ and p-value $= 52.4$ per cent. This test motivates further investigations of linear systematic treatment methods in future galaxy

surveys since the 1D diagnostic based on the mean density contrast is not sufficiently sensitive to unveil these issues with the standard treatment in the SGC (see Fig. 9). Interestingly, these histograms show that the magnitude of simulated systematic effects for the contaminated mocks are stronger in the SGC (cf. the left-hand and right-hand panels of Figs 9 and 10). Similarly, the DR16 sample shows a stronger spurious fluctuation around 15 per cent against depth in the SGC region, compared with 8 per cent in the NGC.

In summary, we find that the nonlinear aspect of our cleaning approach is the primary reason for efficiently reducing spurious fluctuations and systematic error. The neural network-based approach can model the non-linear feedback of observed quasar density to imaging templates, which results in a cleaner sample with a significantly lower χ^2 value. Although the standard treatment passes the null test based on mean density contrasts, however, the test based on cross-power shows that the catalogue with the standard systematic weights is not properly cleaned. We find no improvement in the mean density residual after including all SDSS maps for training; however, our analysis shows that the *Gaia* stellar density is required
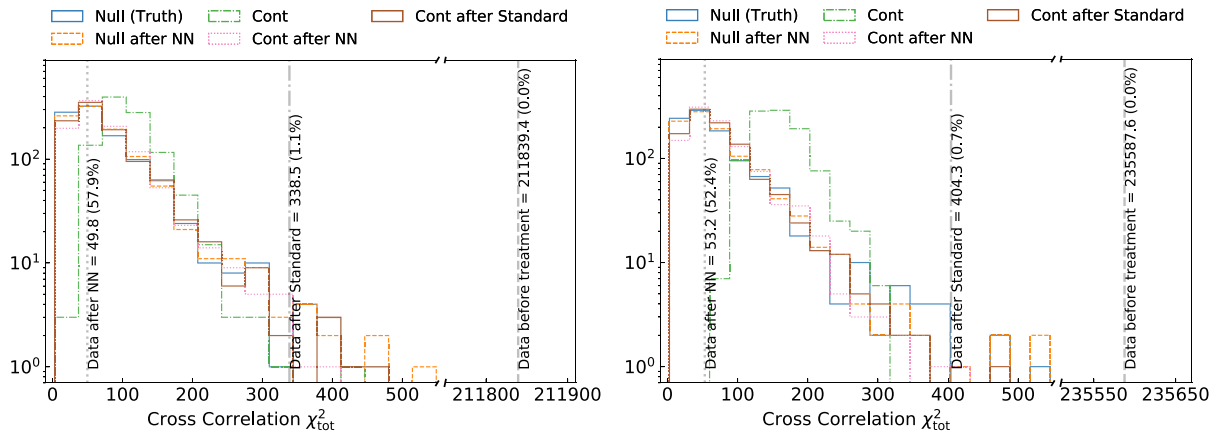

Figure 10. Similar to Fig. 9 for the angular cross-power spectrum between the projected quasar density and imaging maps. Compared with the simulations without systematics (Null), there is a substantial remaining systematics both in the NGC and SGC with the standard linear treatment, while the 1D diagnostic (Fig. 9) shows no obvious issues with the SGC sample cleaned with the linear approach.
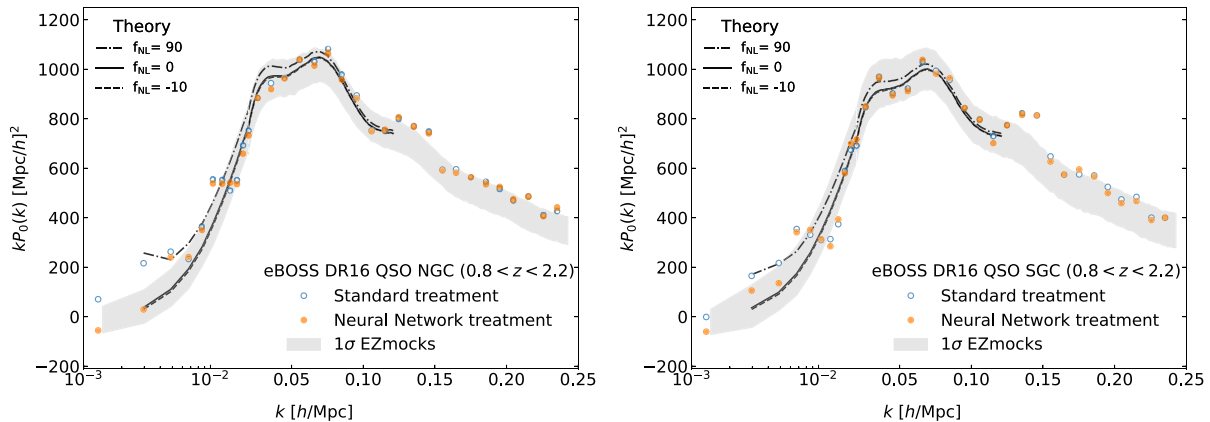
Figure 11. Monopole of the main sample in the NGC (left) and SGC (right) after treatment with the standard method and neural network. Various f_{NL} models are plotted to show the sensitivity of the signal on large scales. The shades represent 1σ statistical uncertainty estimated from the EZmocks. The x-axes are logarithmic for $k < 0.02\ h\mathrm{Mpc}^{-1}$ and linear otherwise.

to satisfy the null tests for residual systematic errors. We also find that computing the mean quasar density per pixel does not change our conclusion based on the mean quasar density per imaging bin, although the former quantity is subject to more fluctuations. Finally, we do not observe a significant change by splitting the main sample into redshift subsamples or using coarser imaging templates. In the following, our neural network-based treatment is applied on the entire $0.8 < z < 2.2$ and uses the Poisson cost function (nn-pnll), cyclic learning rate, and five imaging templates in NSIDE = 512 (Known + *Gaia*) as input features (see Table 1).

4.3 Power spectrum

4.3.1 Measurement

We now proceed to measure the power spectrum of the DR16 sample and EZmock simulations for each Galactic cap separately since each cap is subject to different targeting properties. Fig. 11 shows the measured monopole power spectrum P_0 of the main sample in the NGC (left) and SGC (right). We use the square root of the diagonal terms of the covariance matrices, constructed from the null EZmocks, as the error-bars on P_0. The open and filled circles represent the measured spectrum after cleaning the sample with the standard and neural network treatments, respectively. In both regions, the non-linear cleaning approach returns a lower power at small k. On the other hand, the effect on the small-scale clustering is very small. We also show various models with $f_{NL} = -10$, 0, or 90 to illustrate the sensitivity of the signal on the low-k measurements.

We apply the systematics treatment methods on both the null and contaminated EZmock catalogues to characterize the impact of the mitigation procedure on the measured clustering statistics. The measured power spectrum of the null mocks without any systematic treatment is considered as the ground truth clustering. Fig. 12 shows the mean and the standard deviation of the measured spectra from the EZmock realizations in the NGC (left) and SGC (right) regions. The top row illustrates the difference between the mean P_0 of the EZmocks after mitigation, including the null (*Truth after NN*) and contaminated catalogues using the neural network (*Cont after NN* and the standard approach *Cont after Standard*), and the truth clustering (*Truth*). The light and dark shades show the standard deviation of the null EZmock spectra and the 1σ uncertainties on the mean of the mock spectra, respectively. In the second row, we show the relative difference in the mean power spectrum. In the third row, we present
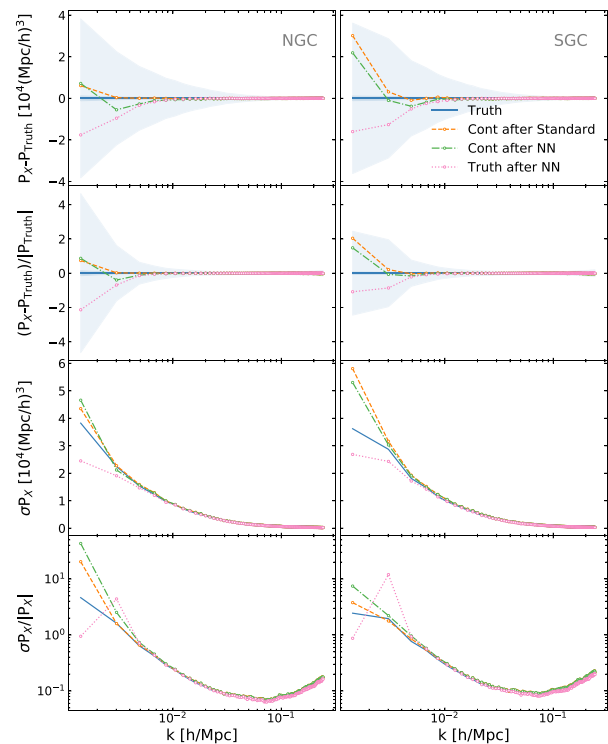

Figure 12. Measured power spectrum of the EZmock realizations before and after systematic treatment for the NGC (left) and SGC (right) regions. From top to bottom, we show the difference between the mean mitigated spectrum and the mean truth spectrum, the relative difference, the dispersion in the mock spectra, and the relative dispersion.

the standard deviation of the mock spectra. Finally, we show the relative dispersion in the bottom row. The measured spectra for the contaminated mocks before treatment is an order of magnitude larger than the truth clustering and thus is not visualized for clarity. We note that the magnitude of the excess power observed in the mocks is one order of magnitude smaller than what is observed in the real sample (cf. Fig. 11), primarily because a linear model was used to generate systematics. This implies that the actual systematics of the real sample is substantially more severe and complex than in these mocks.

Due to allowing the correction to account for more freedom, the neural-network treatment removes more of the modes, known as the
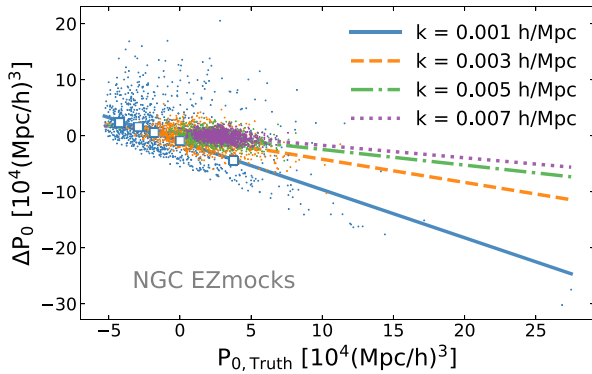
Figure 13. Difference between the measured spectra of the contaminated EZmock catalogues after cleaning and the spectra of the null catalogues as a function of the latter. The medians are used to obtain the best linear fit in each k bin, and are shown only for $k = 0.001$ h/Mpc with open squares.

over-fitting problem, when it is applied to the mocks. On the other hand, the standard treatment indicates less of this over-fitting issue. This is expected as the same linear model is used to produce the systematic effects in the mock realizations. The standard deviation of the mock spectra shows that the imaging treatments do not increase the fractional variance of the measured power spectrum down to $k = 0.003$ h/Mpc. Interestingly, we observe that the dispersion of the null mocks decreases after applying the NN treatment, which is due to the over-correction of the power itself .[18]

4.3.2 Mitigation Bias

Using the mocks, we attempt to estimate any residual or over-correction that might have been introduced in the measured eBOSS QSO power spectrum in the process of the imaging systematics treatment. This assessment is crucial for obtaining an unbiased clustering measurement, which will lead to an accurate inference of cosmological parameters.

We compare the spectra of the contaminated mocks after mitigation to that of the null mocks before treatment as the true power. Fig. 12 shows that the NN-based mitigation tends to introduce overcorrection for $k < 0.003$, especially if it is applied on the density field with no systematics (i.e. Truth with NN[19]). With this caveat, we focus on the overcorrection observed in the contaminated mocks and inspect its nature. The differences between the measured power of the 1000 contaminated mocks after cleaning and the true power for all mocks are shown in Fig. 13 as a function of the true power for the first few k bins. We find that the mitigation bias (or overcorrection) is almost linearly proportional to the true power. While this exercise shows that if the data has no systematics or is subject to simple, linear systematics, the neural network based method, we developed will potentially introduce a small degree of overcorrection and we can attempt to correct for such mitigation bias. However, from Figs 9 and 10, it is apparent that the DR16 sample is subject to much more severe and non-linear systematic effects compared to the mocks and the standard linear method is not sufficiently effective. Therefore, we believe that the overcorrection is less likely a problem for the real eBOSS data and an attempt to mitigate it might further bias the data. We present

---

[18]The error on the power spectrum is expected to be proportional to the power itself under a Gaussian limit.

[19]The standard method performs better by construction, as it assumes we know exactly the source of systematics.

the discussion of our mitigation bias on primordial non-Gaussianity constraints in a companion paper (Mueller et al. in preparation).

5 CONCLUSION

We have performed a thorough study of imaging systematic effects and various template-based mitigation techniques in the final sample of quasars (Lyke et al. 2020; Ross et al. 2020) from the eBOSS DR16 (Ahumada et al. 2020). We present a nonlinear cleaning approach, based on artificial neural networks, and compare the treatment effectiveness with the standard method, based on linear regression. The methods are applied to model the observed density of quasars given a set of templates for imaging properties, related to SDSS properties and Galactic foregrounds, which include stellar density, Galactic extinction, neutral hydrogen column density, depth, seeing, sky brightness, airmass, and run. As summarized in Table 1,

(i) We find that the neural network-based approach outperforms standard linear regression by allowing more freedom for correcting non-linear and complex variations in the quasar density caused by imaging properties, see Figs 5 and 8. The approach is also further improved by using the Poisson statistics to account for the sparsity of the DR16 sample.

(ii) Stellar density is one of the most important sources of spurious fluctuations, and a new template constructed using the *Gaia* DR2 (Gaia Collaboration et al. 2018) yields the best agreement to the observed chi-squared values in the simulations (see Fig. 6 and Table 1). We also show that linear treatment, with the *Gaia* map included, is still not able to properly remove systematics.

(iii) We find no evidence for redshift-dependent imaging systematics and no substantial difference after changing the pixel resolution of imaging templates, see Fig. 7. Therefore, we choose NN trained with PNLL, cyclic learning, and imaging templates in NSIDE = 512 as our default approach.

We utilize the EZmocks, both in the presence and absence of imaging systematics, to construct covariance matrices, quantify residual systematic error, and assess the quality of the DR16 sample for cosmological studies. We find

(iv) The mean density null test shows some remaining systematic error in the catalogue with the standard weights in the NGC region, specifically $\chi^2 = 218.1$ with p-value = 0.2 per cent. Although this test does not reveal any issues with the standard catalogue in the SGC region, $\chi^2 = 132.5$ with p-value = 65.0 per cent (see Table 2), our second null test based on cross-power spectra unveils a significant systematic error in the SGC, $\chi^2 = 404.3$ with p-value = 0.7 per cent.

(v) This work motivates further investigations of linear systematic treatment methods in future galaxy surveys since the 1D diagnostic based on the mean density contrast does not indicate any issues with the standard treatment in the SGC (see Fig. 9).

(vi) The catalogue with the neural network-based systematic weights passes both null tests by providing substantially lower χ^2 values (see Figs 9 and 10).

Collectively, these tests demonstrate that the DR16 quasar catalogue with the standard systematic weights suffer from residual imaging systematics in both Galactic caps, and should not be used for measuring quasar clustering on large scales, i.e. $k < 0.01$ h Mpc^{-1}, as shown in Fig. 11. Nevertheless, it is expected that the impact of imaging systematics to be insignificant on the BAO measurements (e.g. analyses presented in Hou et al. 2021; Neveux et al. 2020), and a thorough investigation is conducted in a companion paper (Merz et al. in preparation).

We then apply our methods on the EZmocks to quantify the impact of systematics treatment on quasar clustering measurements

(see Fig. 12). The neural-network treatment removes some of the cosmological power due to allowing for more freedom in removing systematic effects. We find that the impacts of overfitting on the mean of the mock power spectra and its error are marginal for $k > 0.004\,h$/Mpc. We employ linear regression to model the impact of mitigation on recovering the ground truth clustering (see Fig. 13). We emphasize that the utility of the mitigation bias treatment is not clear, since the parameters are derived from the mocks without realistic imaging systematics (see Figs 9 and 10). However, we investigate the effect on primordial non-Gaussianity constraints in a companion paper (Mueller et al. in preparation).

The end-product from this work is a new value-added quasar catalogue with the enhanced weights to correct for nonlinear imaging systematic effects. The new weights are necessary to make a robust measurement of quasar clustering on large scales ($k < 0.01\,h\,\mathrm{Mpc}^{-1}$). This catalogue is used in a companion paper constraining the local-type primordial non-Gaussianity (Mueller et al. in preparation).

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The catalogue described in this article will be made public in Github at https://github.com/mehdirezaie/eBOSSDR16QSOE. The neural network pipeline utilized in this work is publicly available at https://github.com/mehdirezaie/sysnetdev.

REFERENCES

Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
Abolfathi B. et al., 2018, ApJS, 235, 42
Acquaviva V., Bartolo N., Matarrese S., Riotto A., 2003, Nucl. Phys. B, 667, 119
Adelman-McCarthy J. K. et al., 2007, ApJS, 172, 634
Agarwal N., Ho S., Shandera S., 2014, J. Cosmol. Astropart. P., 2014, 038
Ahumada R. et al., 2020, ApJS, 249, 3
Akrami Y. et al., 2020, A&A, 641, A9
Alam S. et al., 2017, MNRAS, 470, 2617
Allen T., Grinstein B., Wise M. B., 1987, Phys. Lett. B, 197, 66
Alvarez M. et al., 2014, preprint (arXiv:1412.4671)
Anderson L. et al., 2014, MNRAS, 441, 24
Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, Phys. Rev. D, 28, 679
Baumann D. et al., 2009, in CMB POLARIZATION WORKSHOP: THEORY AND FOREGROUNDS: CMBPol Mission Concept Study, AIP Conf. Proc., Vol. 1141. p. 10
Bautista J. E. et al., 2017, A&A, 603, A12
Bautista J. E. et al., 2018, ApJ, 863, 110
Bernardeau F., Uzan J.-P., 2002, Phys. Rev. D, 66, 103506
Bianchi D., Percival W. J., 2017, MNRAS, 472, 1106
Bianchi D., Gil-Marín H., Ruggeri R., Percival W. J., 2015, MNRAS, 453, L11
Blanton M. R. et al., 2017, AJ, 154, 28
Blomqvist M. et al., 2019, A&A, 629, A86
Bolton A. S. et al., 2012, AJ, 144, 144
Bovy J. et al., 2012, ApJ, 749, 41
Castorina E. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 010
Center O. S., 2016, Owens Supercomputer, Available at: http://osc.edu/ark:/19495/f5s1ph73
Chabanier S. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 017
Chen X., xin Huang M., Kachru S., Shiu G., 2007, J. Cosmol. Astropart. P., 2007, 002
Chuang C.-H., Kitaura F.-S., Prada F., Zhao C., Yepes G., 2015, MNRAS, 446, 2621
Cole S. et al., 2005, MNRAS, 362, 505
Creminelli P., Zaldarriaga M., 2004, J. Cosmol. Astropart. P., 2004, 006
Dalal N., Dore O., Huterer D., Shirokov A., 2008, Phys. Rev. D, 77, 123514
Dawson K. S. et al., 2012, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
de Sainte Agathe V. et al., 2019, A&A, 629, A85
Delubac T. et al., 2017, MNRAS, 465, 1831
Desjacques V., Seljak U., 2010, Class. Quant. Gravity, 27, 124011
du Mas des Bourboux H. et al., 2020, ApJ, 901, 153
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Elvin-Poole J. et al., 2018, Phys. Rev. D, 98, 042006
Falk T., Rangarajan R., Srednicki M., 1992, Phys. Rev. D, 46, 4232
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Fukugita M., Shimasaku K., Ichikawa T., Gunn J. et al., 1996, Technical Report, The Sloan Digital Sky Survey photometric system. SCAN-9601313
Gaia Collaboration et al., 2018, AAP, 616, A12
Gangui A., Lucchin F., Matarrese S., Mollerach S., 1993, ApJ, 430, 447

Giannantonio T., Ross A. J., Percival W. J., Crittenden R., Bacher D., Kilbinger M., Nichol R., Weller J., 2014, Phys. Rev. D, 89, 023511
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Grossi M., Branchini E., Dolag K., Matarrese S., Moscardini L., 2008, MNRAS, 390, 438
Gunn J. E. et al., 2006, AJ, 131, 2332
Guth A. H., Pi S.-Y., 1982, Phys. Rev. Lett., 49, 1110
Hahn C., Scoccimarro R., Blanton M. R., Tinker J. L., Rodríguez-Torres S. A., 2017, MNRAS, 467, 1940
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ 156 160
Hand N., Li Y., Slepian Z., Seljak U., 2017, J. Cosmol. Astropart. Phys., 2017, 002
Hartlap J., Simon P., Schneider P., 2007, AAP, 464, 399
Hawking S., 1982, Phys. Lett. B, 115, 295
HI4PI Collaboration et al., 2016, AAP, 594, A116
Ho S. et al., 2015, J. Cosmol. Astropart. P., 2015, 040
Hobson E. W., 1931, The Theory of Spherical and Ellipsoidal Harmonics. Cambridge University Press, London
Hou J. et al., 2021, MNRAS, 500, 1201
Hutchinson T. A. et al., 2016, AJ, 152, 205
Ioffe S., Szegedy C., 2015, preprint (arXiv:1502.03167)
Jensen T. W. et al., 2016, ApJ, 833, 199
Jing Y. P., 2005, ApJ, 620, 559
Kalus B., Percival W. J., Bacon D. J., Samushia L., 2016, MNRAS, 463, 467
Kalus B., Percival W., Bacon D., Mueller E., Samushia L., Verde L., Ross A., Bernal J., 2019, MNRAS, 482, 453
Karagiannis D., Shanks T., Ross N. P., 2014, MNRAS, 441, 486
Kofman L., Pogosyan D. Y., 1988, Phys. Lett. B, 214, 508
Komatsu E., Spergel D. N., 2001, Phys. Rev. D, 63, 063002
Laurent P. et al., 2017, J. Cosmol. Astropart. P., 2017, 017
Leistedt B., Peiris H. V., 2014, MNRAS, 444, 2
Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévy A., Pontzen A., 2013, MNRAS, 435, 1857
Leistedt B., Peiris H. V., Roth N., 2014, Phys. Rev. Lett., 113, 221301
Linde A., Mukhanov V., 1997, Phys. Rev. D, 56, R535
Loshchilov I., Hutter F., 2016, preprint (arXiv:1608.03983)
Loshchilov I., Hutter F., 2017, preprint (arXiv:1711.05101)
Lyke B. W. et al., 2020, ApJS, 250, 8
Lyth D. H., Ungarelli C., Wands D., 2003, Phys. Rev. D, 67, 023503
Maldacena J., 2003a, J. High Energy Phys., 2003, 013
Maldacena J., 2003b, J. High Energy Phys., 2003, 013
Matarrese S., Verde L., 2008, ApJ, 677, L77
Matarrese S., Verde L., Jimenez R., 2000, ApJ, 541, 10
Mohammad F. G. et al., 2020, MNRAS, 498, 128
Mukhanov V. F., Chibisov G. V., 1981, Sov. J. Exp. Theor. Phys. Lett., 33, 532
Myers A. D. et al., 2015, ApJS, 221, 27
Neveux R. et al., 2020, MNRAS, 499, 210
Percival W. J. et al., 2010, MNRAS, 401, 2148
Prakash A. et al., 2016, ApJS, 224, 34
Pullen A. R., Hirata C. M., 2013, Publ. Astron. Soc. Pac., 125, 705
Raichoor A. et al., 2017, MNRAS, 471, 3955
Raichoor A. et al., 2020, MNRAS, 500, 3254
Reid B. A., Verde L., Dolag K., Matarrese S., Moscardini L., 2010, J. Cosmol. Astropart. Phys., 2010, 013
Rezaie M., Seo H.-J., Ross A. J., Bunescu R. C., 2020, MNRAS, 495, 1613
Rhodes J. et al., 2013, preprint (arXiv:1309.5388)
Richards G. T. et al., 2008, ApJS, 180, 67
Ross A. J. et al., 2011, MNRAS, 417, 1350
Ross A. J. et al., 2012, MNRAS, 424, 564
Ross A. J. et al., 2013, MNRAS, 428, 1116
Ross A. J. et al., 2017, MNRAS, 464, 1168
Ross A. J. et al., 2020, MNRAS, 498, 2354
Ruder S., 2016, preprint (arXiv:1609.04747)
Salopek D., Bond J., Bardeen J. M., 1989, Phys. Rev. D, 40, 1753
Schlafly E. et al., 2012, ApJ, 756, 158
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Scoccimarro R., 2015, Phys. Rev. D, 92, 083532
Scoccimarro R., Sefusatti E., Zaldarriaga M., 2004, Phys. Rev. D, 69, 103513
Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, MNRAS, 460, 3624
Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, J. Cosmol. Astropart. P., 2008, 031
Smee S. A. et al., 2013, AJ, 146, 32
Smith A. et al., 2020, MNRAS, 499, 269
Smith L. N., 2015, preprint (arXiv:1506.01186)
Starobinsky A. A., 1982, Phys. Lett. B, 117, 175
Taruya A., Koyama K., Matsubara T., 2008, Phys. Rev. D, 78, 123534
Tegmark M., 1997, Phys. Rev. D, 55, 5895
Verde L., Wang L., Heavens A. F., Kamionkowski M., 2000, MNRAS, 313, 141
Wagoner E. L., Rozo E., Fang X., Crocce M., Elvin-Poole J., Weaverdyck N., 2021, MNRAS, 503, 4349
Weaverdyck N., Huterer D., 2021, MNRAS, 503, 5061
Wright E. L. et al., 2010, AJ, 140, 1868
Yamamoto K., Nakamichi M., Kamino A., Bassett B. A., Nishioka H., 2006, PASJ, 58, 93
Zaldarriaga M., 2004, Phys. Rev. D, 69, 043508
Zel'Dovich Y. B., 1970, A&A, 5, 84
Zhao C. et al., 2021, MNRAS, 503, 1149

[1]*Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA*
[2]*Center of Cosmology and AstroParticle Physics, The Ohio State University, Columbus, OH 43210, USA*
[3]*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[4]*Waterloo Centre for Astrophysics, University of Waterloo, Waterloo ON N2L 3G1, Canada*
[5]*Department of Physics and Astronomy, University of Waterloo, Waterloo ON N2L 3G1, Canada*
[6]*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo ON N2L 2Y5, Canada*
[7]*Department of Computer Science, College of Computing and Informatics, University of North Carolina, Charlotte, NC 28223, USA*
[8]*Institute of Cosmology & Gravitation, Dennis Sciama Building, University of Portsmouth, Portsmouth PO1 3FX, UK*
[9]*Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA*
[10]*IRFU,CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[11]*Institut de Ciències del Cosmos, Universitat de Barcelona, ICCUB, Martí i Franquès 1, E-08028 Barcelona, Spain*
[12]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[13]*Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstr., D-85748 Garching bei München, Germany*
[14]*Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA*
[15]*Instituto de Fisica, Universidad Nacional Autonoma de Mexico, Apdo. Postal 20-364, Mexico*
[16]*Department of Physics and Astronomy, Sejong University, Seoul 143-747, Korea*
[17]*Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA*
[18]*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*
[19]*Institute for Computational Cosmology, Dept. of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[20]*National Astronomy Observatories, Chinese Academy of Science, Beijing 100101, China*
[21]*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*

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