UC Berkeley
UC Berkeley Previously Published Works

Title

Learning to concentrate: multi-tracer forecasts on local primordial non-Gaussianity with machine-learned bias

Authors

Sullivan, James M
Prijon, Tijan
Seljak, Uroš

Peer reviewed

Learning to Concentrate: Multi-tracer Forecasts on Local Primordial Non-Gaussianity with Machine-Learned Bias

James M. Sullivan, ^{a,b,1} **Tijan Prijon**^{b,c} **Uroš Seljak**^{a,b,d,e}

^{a}Department of Astronomy, University of California, Berkeley, CA 94720, USA

^{b}Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA

^{c}Faculty of Mathematics and Physics, University of Ljubljana, Jadranska ulica 19, 1000 Ljubljana, Slovenia

^{d}Department of Physics, University of California, Berkeley, CA 94720, USA

^{e}Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

E-mail: jmsullivan@berkeley.edu, tp7058@student.uni-lj.si, useljak@berkeley.edu

Abstract. Local primordial non-Gaussianity (LPNG) is predicted by many non-minimal models of inflation, and creates a scale-dependent contribution to the power spectrum of large-scale structure (LSS) tracers, whose amplitude is characterized by b_ϕ. Knowledge of b_ϕ for the observed tracer population is therefore crucial for learning about inflation from LSS. Recently, it has been shown that the relationship between linear bias b_1 and b_ϕ for simulated halos exhibits significant secondary dependence on halo concentration. We leverage this fact to forecast multi-tracer constraints on $f_{NL}^{\rm loc}$. We train a machine learning model on observable properties of simulated Illustris-TNG galaxies to predict b_ϕ for samples constructed to approximate DESI emission line galaxies (ELGs) and luminous red galaxies (LRGs). We find $\sigma(f_{NL}^{\rm loc}) = 2.3$, and $\sigma(f_{NL}^{\rm loc}) = 3.7$, respectively. These forecasted errors are roughly factors of 3, and 35% improvements over the single-tracer case for each sample, respectively. When considering both ELGs and LRGs in their overlap region, we forecast $\sigma(f_{NL}^{\rm loc}) = 1.5$ is attainable with our learned model, more than a factor of 3 improvement over the single-tracer case, while the ideal split by b_ϕ could reach $\sigma(f_{NL}^{\rm loc}) < 1$. We also perform multi-tracer forecasts for upcoming spectroscopic surveys targeting LPNG (MegaMapper, SPHEREx) and show that splitting tracer samples by b_ϕ can lead to an order-of-magnitude reduction in projected $\sigma(f_{NL}^{\rm loc})$ for these surveys.

---

[1]Corresponding author.

Contents

1 Introduction

Cosmological observations of the Cosmic Microwave Background (CMB) and Large-scale Structure (LSS) are currently in complete concordance with Gaussian initial conditions [1]. Yet, the tantalizing possibility that there are deviations from Gaussianity lurking beneath observational uncertainties remains. These deviations - or Primordial non-Gaussianity (PNG) - come in several well-known forms, one of which is the *local* type (LPNG) parameterized by the amplitude $f_{NL}^{\rm loc}$:

$$\phi = \phi_G + f_{NL}^{\rm loc} \left[\phi_G^2 - \langle \phi_G^2 \rangle \right], \tag{1.1}$$

where ϕ is the Bardeen potential [2] and ϕ_G a Gaussian random field. A detection of $f_{NL}^{\rm loc} \sim 1$ would definitively indicate the presence of multiple fields during inflation, and provides a natural theoretical sensitivity target [3, 4].

    The Planck satellite has placed the most stringent constraints on $f_{NL}^{\rm loc}$ to date, finding $f_{NL}^{\rm loc} = -0.9 \pm 5.1$ [5], consistent with no local PNG. While future CMB missions should improve on this uncertainty, they are not forecasted to reach $\sigma(f_{NL}^{\rm loc}) \lesssim 1$ [6, 7]. However, LSS surveys promise to supersede this sensitivity in the near-term [8–14]. In particular, the multi-tracer technique [13–22] provides a powerful tool for using large-scale modes most affected by survey sample variance.

The presence of local PNG modulates the halo density field in a scale-dependent manner [23, 24]. This "scale-dependent bias" effect (LPNG bias) produces a signal in the halo power spectrum that scales like k^{-2}, becoming more important on larger scales. This behavior is fundamentally due to LPNG inducing a bispectrum that peaks in the squeezed triangle configuration, which couples a long wavelength mode with two short wavelength modes, effectively linking very large scales with the small-scale process of halo formation. The amplitude of this effect is controlled by the LPNG bias b_ϕ, and while initially it was believed that it only depends on the halo mass [23], it was soon realized secondary halo properties such as merger history also affect its value [24]. This effect has been exploited over the last 15 years in several LSS analyses to make significant progress toward a precision measurement of $f_{NL}^{\rm loc}$ [24–29], though those constraints remain looser than those from Planck - LSS has yet to take the lead in constraining $f_{NL}^{\rm loc}$. The above analyses do not make use of the multi-tracer technique, and therefore can in principle be pushed further in sensitivity. However, all of these analyses (save Ref. [28]) make a simplifying assumption that the relationship between the LPNG bias b_ϕ and the linear tracer bias b_1 is perfectly known.

A slew of recent papers has called this knowledge, which was always approximate, into significant question for simulated galaxies and other tracers [30–33]. Building on the pioneering work of Ref. [34], Ref. [35] investigated to what extent several commonly considered assembly bias parameters affect the relationship between the linear halo bias b_1 and b_ϕ in dark matter only N-body simulations. For halo spin and sphericity, the authors found only a modest effect, but, especially for lower halo masses and redshifts, they found a significant secondary dependence of b_ϕ on halo concentration. There, the authors of Ref. [35] argue that the strong sensitivity of b_ϕ to changes in concentration indicates that existing constraints assuming a fixed form for $b_\phi(b_1)$ are unreliable, but also acknowledge that this sensitivity is an opportunity to better constrain $f_{NL}^{\rm loc}$ with suitably defined galaxy populations. The goal of this work is to identify such populations with the aid of a machine learning algorithm trained on observable galaxy properties and to use its predictions to forecast the improvement on $\sigma(f_{NL}^{\rm loc})$ that such a definition furnishes.

By using simulated galaxy samples that approximate the DESI selections for emission line galaxies (ELGs) and luminous red galaxies (LRGs), we can get a realistic estimate for the current and near-term prospects of constraining $f_{NL}^{\rm loc}$ by using observable quantities beyond halo mass to which b_ϕ responds. We will show that, depending on the information available and the galaxy sample properties, we are able to improve the forecasted error on $f_{NL}^{\rm loc}$ by factors of several for DESI-like galaxies using machine-learned b_ϕ predictions. We also comment on the prospects for using this strategy in future spectroscopic surveys, where naively one could expect an order of magnitude improvement on $\sigma(f_{NL}^{\rm loc})$ if halo concentration were perfectly recoverable.

This paper is structured as follows: We briefly review the LPNG parameter b_ϕ in Section 2 before moving on to the construction of the DESI-like simulated galaxy samples and our machine learning methodology in Section 3. We then provide forecasts for $\sigma(f_{NL}^{\rm loc})$ in Section 4, and provide some concluding remarks in Section 5.

2 Local PNG bias b_ϕ

We briefly review the salient facts about the scale-dependent bias signal induced by local primordial non-Gaussianity and its amplitude b_ϕ.

In the peak-background split (PBS) formalism [36], the parameter b_ϕ is defined as the response of the tracer mean density $\bar{n}_t$ to the presence of a long-wavelength perturbation of the Bardeen potential ϕ

$$b_\phi = \frac{\partial \log \bar{n}_t}{\partial (f_{NL}^{\text{loc}} \phi)} = 2 \frac{\partial \log \bar{n}_t}{\partial \log \sigma_8}. \tag{2.1}$$

The second equality follows from the argument that the effect of LPNG on the tracer abundance is equivalent to the effect of a rescaling of the amplitude of linear fluctuations [36]. We will accept this argument for the purposes of this work.

This bias enters the tracer power spectrum at linear order in δ and f_{NL}^{loc}:

$$P_{tt}(k) = \langle \delta_t(\mathbf{k}) \delta_t^*(\mathbf{k}') \rangle' \tag{2.2}$$

$$\stackrel{L.O.}{=} b^2 P_L(k) + 2 b_\phi b \mathcal{M}^{-1}(k) P_L(k), \tag{2.3}$$

where the $'$ indicates that we drop the missing momentum-conserving Dirac delta and associated $(2\pi)^3$ prefactor, b is the linear bias, and $\mathcal{M}^{-1}(k) = \frac{3 \Omega_m H_0^2}{2 c^2 D_{md} k^2 \tilde{T}(k)}$, where $\tilde{T}(k)$ is the transfer function normalized to 1 on large scales, and D_{md} is the linear total matter growth factor normalized to 1 in matter domination.

Reducing to the case of halos ($\bar{n}_t = \bar{n}_h$) with a universal mass function that only allows halo abundance to depend on halo mass, gives the universality relation (or UMF [24]):

$$b_\phi(b_1, p) = 2 \delta_c (b_1 - p),$$

where the critical overdensity is $\delta_c = 1.686$. For the case where we select all of the halos of a given mass the universality relation gives $p = 1$. For dark matter halos there are some deviations of this $b_\phi(b_1)$ relationship observed [33, 37]. It has also been argued that other values of p are more appropriate for certain sub-populations of halos, such as for halos that have undergone a recent merger, which may host quasars [24]. However, the picture becomes even more complicated for simulated galaxies, for which it was recently shown that when selecting by stellar mass, a value of $p = 0.55$ is preferred when using the IllustrisTNG galaxy formation model [30]. Presumably, if the stellar mass selection can be replaced with halo mass selection, this would restore $p = 1$, which suggests that by choosing additional observables one may be able to split galaxies by b_ϕ.

One such additional observable that has been found to be very sensitive to b_ϕ is halo concentration c. With a fixed galaxy formation model, a value of the LPNG bias can be estimated from Separate Universe (SU) simulations [e.g., 38–45] in which two N-body simulations are run with two values of σ_8 and the bias is then estimated via finite difference with respect to the mean number density $\bar{n}_h(M, c)$. We use the reported results of Refs. [30, 35] at $z = 1$ to relate linear halo bias b_1 (estimated from power spectra) and SU LPNG bias b_ϕ to halo mass M and concentration c.

3 b_ϕ with machine learning

In this Section, we describe our machine learning method for obtaining predictions for b_ϕ. Before diving into the details, we provide a brief overview of this process here.

First, we obtain $b_1(M, c), b_\phi(M, c)$ from the gravity-only Separate Universe (SU) results of [35], and assign $b_1(M, c), b_\phi(M, c)$ to each TNG galaxy's host halo of mass M and concentration c. We then train the ML model described below on galaxies where the values of b_ϕ

are known to learn the relationship between input features such as galaxy magnitude and b_ϕ. After making a prediction for b_ϕ on test data, we rank order the b_ϕ predictions to determine a split of the sample into tertiles. We then obtain predictions for a mean value of b_ϕ in each tertile by using the trained ML model evaluated on the observable input features of galaxies in that tertile (magnitude, position, etc.) and averaging. This final set of b_ϕ values in the tertiles are used as input to the Fisher forecasts of Section 4.

3.1 Simulated galaxy sample
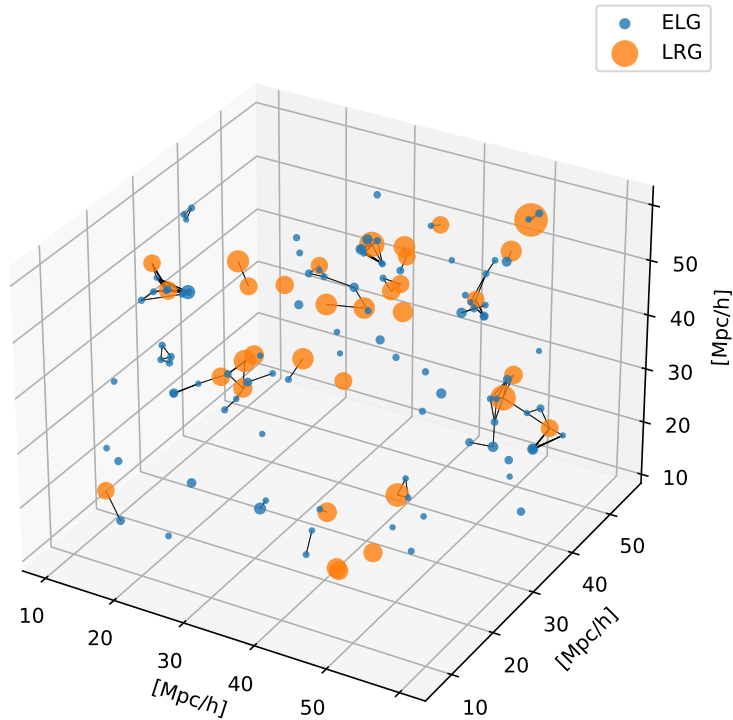
3.1.1 Simulation

Figure 1. ELG and LRG-like galaxies in the part of the volume of the IllustrisTNG simulations at redshift $z = 1$. Galaxies are presented in real space, and those closer than $5\ h^{-1}$ Mpc are linked for ease of presentation. The size of each galaxy marker is directly proportional to its logarithmic stellar mass ($\log M^*$).

We use simulated galaxy data from the IllustrisTNG-300 [46–51] simulation, a cosmological magnetohydrodynamical simulation that models the formation and evolution of galaxies and covers the volume of $V = 205^3\ [h^{-1}\ \mathrm{Mpc}]^3$. The snapshot that we use is at redshift $z = 1$, which is a compromise between being able to mimic Dark Energy Spectroscopic Instrument (DESI) survey [52] observations of both emission-line galaxies (ELGs) and luminous red galaxies (LRGs). A sub-volume of the simulation containing ELG and LRG- like galaxies, as we define in Section 3.1.3, is shown in Fig. 1.

To ensure that we correctly model the local environment surrounding the halos, we drop the galaxies near the edges of the simulation (near 1 h^{-1} Mpc)[1]. We do this to eliminate boundary bias and edge effects and choose the value of 1 h^{-1} Mpc based on the tradeoff between the loss of data and the potential for edge effects. We observe no noticeable variation in results upon increasing this distance further.

3.1.2 Observables

The halos in the IllustrisTNG simulation are determined with the friend-of-friends (FoF) algorithm [53], with linking length $b = 0.2$.

We use the results of Ref. [35] to link b_ϕ and b_1 to halo mass and concentration. We take halo concentration from the IllustrisTNG supplementary data catalog [54]. It is defined as $c = R_{200c}/R_s$, where R_s is determined by fitting a Navarro-Frenk-White (NFW) [55] profile to the dark matter density profile. We use the R_{200c} halo mass definition. We note that Ref. [35] used several halo mass definitions when using different simulations, though since concentration was logarithmized, the effect of these differences should be reduced.

To avoid using halo properties that are unobservable by spectroscopic surveys to infer halo LPNG bias, we only use stellar masses M^*, the r, g, and z magnitudes, and redshift-space positions of the galaxies. We use the word "observable" somewhat loosely, as galaxy stellar mass must be estimated from spectra, and dust models that affect the observed r, g, z magnitudes are uncertain - we do not account for these aspects that would be relevant for a real data analysis here. We choose the IllustrisTNG z-coordinate as the line-of-sight direction to obtain positions in redshift space. The real-to-redshift space transformation is

$$\mathbf{s} = \mathbf{x} + \frac{v(\mathbf{x})\mu}{\mathcal{H}}\hat{z} \tag{3.1}$$

where $\mathbf{x}$ is the real-space galaxy position, $v(x)$ is the galaxy velocity magnitude, $\hat{z}$ is the line-of-sight direction, $\mu = \hat{\mathbf{v}} \cdot \hat{z}$, $\mathcal{H}$ is the conformal Hubble parameter, and $\mathbf{s}$ is the redshift-space position. To simulate the effect of redshift-space distortions on galaxy observations, we transform the galaxies to redshift space before computing input features and training the machine learning algorithm to predict b_ϕ.

While each galaxy lies in exactly one halo, a halo can contain multiple galaxies inside. When we report results in later sections, we predict b_ϕ of the halo for each galaxy separately. Thus, it can happen that in the results, we have several (either the same or different) predictions for the same halo if several galaxies lie in it. This approach is justified because we, in reality, observe galaxies and cannot a priori know whether two galaxies belong to the same halo. We emphasize that in the training we do not use any information about the halos (for example, halo position), except the knowledge of to which parent halo a galaxy belongs.

3.1.3 ELG and LRG sample

Our data sample mimics the emission-line galaxies (ELGs), and luminous red galaxies (LRGs), as observed by the DESI survey at redshift $z = 1$. We can construct these mock galaxy samples either by color-space cuts or by galaxy selection based on specific star-formation rate (sSFR, i.e., the SFR per stellar mass) and stellar mass (M^*) cuts [56–58]. We adopt the latter strategy and use sSFR and stellar mass cuts.

To obtain both ELG and LRG samples, we first make sSFR cuts and then take the heaviest galaxies based on M^* to get the desired target number densities. We made two

---

[1]But take them into account when calculating neighborhood information of other halos/galaxies
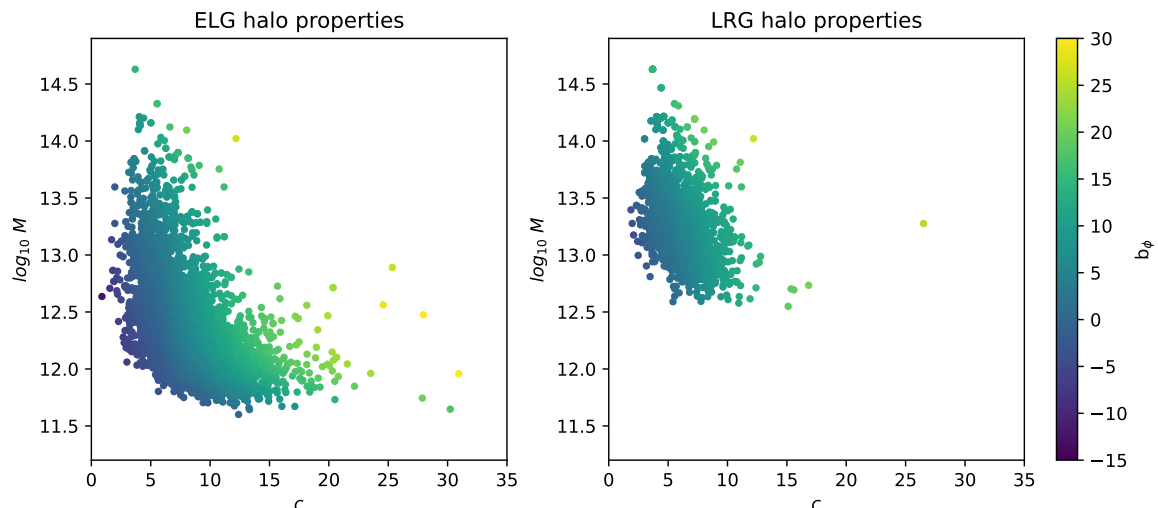
Figure 2. Parent halo properties of ELGs and LRGs. As expected we see that LRGs live in more massive halos with lower concentration. We also show the b_ϕ value associated each halo. We can see that while it is a function of both halo mass and concentration, the latter has a much greater impact. The halo mass is reported in the units of 10^{10} M$_\odot/h$.

different sample selections using sSFR cut values of $\log_{10}[sSFR] = -9.09$ from [58], and $\log_{10}[sSFR] = -9.23$ from [56]. The target number densities are taken from [59, 60]. The selection of ELG and LRG samples is summarized in Table 1. Here we find that the mean b_1 values of our mock samples are similar to those of Ref. [52] for ELGs (1.4) and LRGs (2.6) at $z = 1$. With this selection, we also ensure that no galaxy is classified as an ELG and an LRG simultaneously. In the next sections we report the results for all eight galaxy samples presented here. We optimize hyperparameter separately for each sample (see Section 3.2.3). However, because the samples are relatively similar, most hyperparameters remain the same.

Type	$\log_{10} sSFR$	n [h^{-1} Mpc]$^{-3}$	min $\log_{10} M^*$	$\bar{b}_1$
LRG	< -9.09	2×10^{-4}	1.06	2.40
LRG	< -9.23	2×10^{-4}	1.06	2.40
ELG	> -9.09	5×10^{-4}	0.02	1.44
ELG	> -9.09	7×10^{-4}	-0.09	1.42
ELG	> -9.09	1×10^{-3}	-0.23	1.39
ELG	> -9.23	5×10^{-4}	0.23	1.46
ELG	> -9.23	7×10^{-4}	0.17	1.44
ELG	> -9.23	1×10^{-3}	0.07	1.42

Table 1. Selection criteria with number densities, minimum obtained stellar mass M^* (units of 10^{10} M$_\odot h$), and average b_1 for each sample. We first apply the sSFR cut (units of $h/$yr), as specified in the second column, and then select galaxies with the highest M^* to achieve the target number density. The number of galaxies ("size of dataset") in each sample can be calculated from number densities and the volume of the simulation ($V = 205^3$ [h^{-1} Mpc]^3), and lie between 1500 and 7000.

In Fig. 2, we compare the halo properties in which LRG and ELG galaxies lie, and show their expected b_ϕ in the mass and concentration plane. We see an apparent trend that LRGs lie in heavier halos and that their halos have lower concentrations. In Fig. 3, we compare the

stellar mass of ELGs and LRGs and its difference when applying different sSFR cuts. While the difference in sSFR affects only the M^* distribution of ELGs, it is evident that LRGs are significantly heavier, in accordance with Refs. [52, 61].

All the results in this Section are reported for the samples selected with the sSFR cut of $\log_{10} sSFR = -9.09$, and with number densities of $n_{ELG} = 5 \times 10^{-4}$ $[h^{-1}$ Mpc$]^{-3}$, and $n_{LRG} = 2 \times 10^{-4}$ $[h^{-1}$ Mpc$]^{-3}$, as these are the DESI number densities at $z = 1$ in Refs. [59, 60].
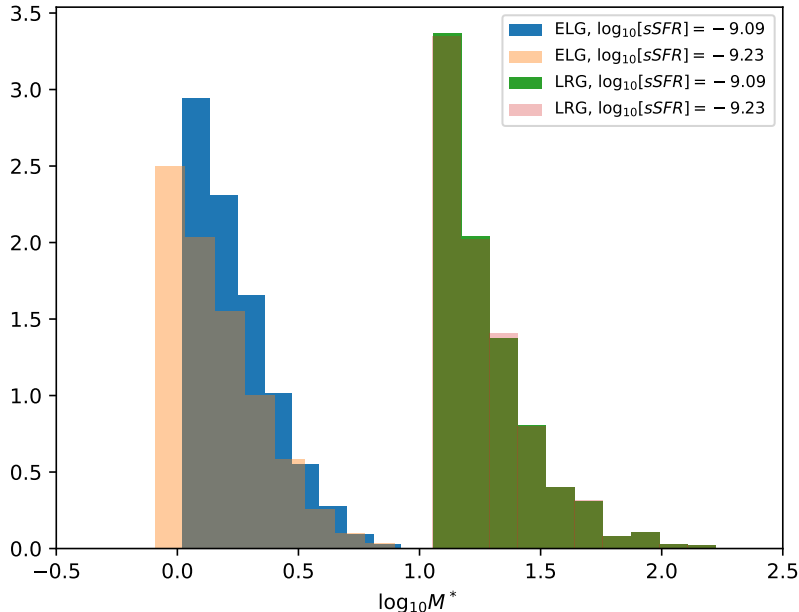

Figure 3. The stellar mass of ELG and LRG-like galaxies selected by two different sSFR cuts. LRGs are significantly heavier in both cuts, and it is apparent that the ELG and LRG selected samples do not overlap. While the different sSFR cuts significantly affect the M^* of the ELGs, they have almost no influence on the M^* of LRGs. The histograms are normalized for easier visualization, as ELG and LRG samples do not contain the same number of galaxies.

3.2 Machine learning

Our objective is to predict the value of the continuous variable b_ϕ, based on M continuous features (inputs) $\mathbf{x_i} = (x_{i,1}, x_{i,2}, ..., x_{i,M})$ on N training examples, which are in our case galaxies. We want to find a function $f : \mathbb{R}^M \to \mathbb{R}$ that maps the inputs $\mathbf{x_i}$ to the scalar value of b_ϕ. When constructing such a function, we want to minimize the difference between its output and the actual value of b_ϕ. The metric of the difference we are minimizing is our loss function, which we take to be the most common root-mean-square error function (RMSE), defined as $L = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$, where y_i and $\hat{y}_i$ are the actual and predicted values of b_ϕ for the ith training example.

The described problem is a single-target regression problem. However, rather than first interpolating b_ϕ from halo mass and concentration (as described in Section 3.1.2) and directly predicting it, we could make predictions for mass and concentration, and interpolate b_ϕ from predicted values instead. In this case, we would be dealing with a multi-target regression problem. We tried this approach and verified that it leads to worse results according to

the RMSE metric. The difference between the two approaches can be imagined only as a difference in the loss function that is optimized. However, because we are only interested in b_ϕ it is both simpler and more effective to focus on a single target regression.

The problem of inferring the halo properties from observable data has already been studied in the literature using various machine learning (ML) techniques. One approach is to use structured data with a combination of graph neural networks (GNN) or convolutional neural networks (CNN), while another approach uses unstructured data by calculating statistics of the halo neighborhood [e.g., 62–68]. We focus on the latter and used several different ML algorithms in combination with unstructured input features. The main reason for this is that GNN and CNN would generally need a larger data sample size, with higher number densities. We further justify this selection when we discuss the interpretability of our model in Section 3.3.

3.2.1 Input features

Each galaxy belongs to exactly one halo whose b_ϕ we want to infer. However, it is not necessarily true that the halo's center coincides with this galaxy's position. We define a central galaxy as the galaxy with the largest M^* within $R = 1 \ h^{-1}$ Mpc of the galaxy whose halo properties we want to infer. R is a hyperparameter of our model, which we optimize as discussed in Section 3.2.3. We then use the central galaxy surroundings to describe the halo environment more accurately.

We use 13 features that we found useful for predicting b_ϕ:

- r, g, z – AB magnitudes of r, g, and z bands of the central galaxy

- $\log M^*$ of the central galaxy

- N_R; number of neighbor galaxies within R, for $R \in \{0.5, 1, 2, 3, 4, 5\} \ h^{-1}$ Mpc

- Metric of anisotropy, defined as $A = \left\| \sum_i \frac{\mathbf{R_i}}{\|\mathbf{R_i}\|} \cdot H(R_H - R_i) \right\|_2$

- Sum of stellar masses of the galaxies in the neighborhood; $\log \left(\sum_i M_i^* \cdot H(R_H - R_i) \right)$

- Sum of stellar mass to distances ratios; $\log \left(\sum_i \frac{M_i^*}{R_i} \cdot H(R_H - R_i) \right)$.

Here $\mathbf{R_i}$ is the separation vector between the i-th neighbor galaxy and central galaxy in redshift space, and $H(R_H - R_i)$ is the Heaviside step function that ensures that only galaxies within the radius R_H are included in the sum. R_H is a hyperparameter of our model, for which we use $R_H = 5 \ h^{-1}$ Mpc. The results are mostly insensitive to the choice of R_h, however when considering a neighborhood much more distant than $5 \ h^{-1}$ Mpc, the results do not improve, and rather start to deteriorate. The machine learning models can in theory learn to ignore less important input features, however, we are limited by the size of our dataset. For this reason, using fewer, only dominantly important features turns out to be better. All the described features use the distances calculated in redshift space. The predictive power of the model would increase if the features are calculated in real space, however we do not focus on them, since spectroscopic surveys observe galaxies in redshift space.

The correlations between the most representative input features are presented in Fig. 4. The intrinsic features of the galaxies are strongly correlated, with the correlation between luminosity bands being almost exactly 1. For this reason, we tried performing principal

component analysis (PCA), which would reduce the dimensionality of the dataset. However, the results did not improve, and we do not use PCA to produce final predictions.

The stellar mass of the central galaxy, sum of neighborhood stellar masses, and sum of stellar-mass-to-distance ratios are all logarithmized before being used for machine learning. They are approximately exponentially distributed, therefore very large (small) numerical values could lead to potential numerical instabilities during the ML learning process. Furthermore, we standardize all features, $\hat{x}_i = \frac{x_i - \mu_i}{\sigma_i}$. Normalizing instead of standardizing features lead to slightly worse results, probably due to the outliers present.

We experimented with utilizing alternative input features. We already commented on the rationale behind restricting the input information to neighboring galaxies within a radius of R_H. However, it would be possible to instead select the N nearest galaxies with the objective of optimizing N. This approach appears to be less intuitive from a physical standpoint, as it may not accurately capture the influence of clustering effects. We have indeed checked and observed that it leads to slightly worse predictions for b_ϕ. Furthermore, we have checked and observed that incorporating information on the luminosities and stellar masses of the neighboring galaxies does not lead to any discernible improvements in the obtained predictions.
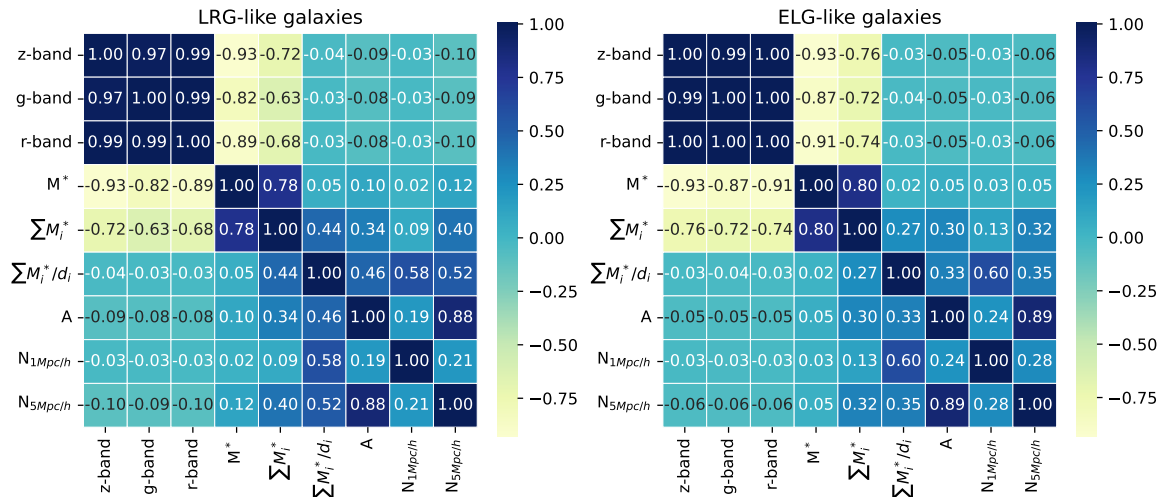

Figure 4. Pearson correlation coefficients of the most representative input features. The coefficients are rounded and are not exactly 1 between different luminosity bands. While the intrinsic features of galaxies (luminosity bands and stellar mass) are quite strongly correlated, they are mostly uncorrelated with the distributions of the galaxies in the neighborhood.

In Table 1, we have 8 different datasets for which we want to make separate b_ϕ predictions. When we make predictions for ELG galaxies, we also include information about nearby LRG galaxies (in redshift-space) and vice versa. In other words, we use both ELGs and LRGs to calculate the number of neighbor galaxies or local anisotropy. However, we train the model and make predictions only on one type of galaxy at a time - this procedure could be performed on real data.

3.2.2 Models

We have tried using several different machine learning algorithms to predict b_ϕ from the input features as described in Section 3.2.1. The top three best-performing models are Artificial

Neural Networks (ANN), Support Vector Machines (SVMs), and gradient-boosted tree models (XGBoost [69]). We discuss the process of hyperparameter optimization in Section 3.2.3.

We employ a technique known as model stacking to enhance the predictive performance of the three best-performing models; ANN, SVM, and XGBoost. Model stacking integrates individual predictions of multiple models and produces the final output from them. By stacking models, we can reduce both prediction bias and variance. The prediction errors arising from the individual models stem from distinct sources and thus can be partially offset.

The effectiveness of model stacking relies heavily on the diversity and quality of the base learners (the models whose predictions we are combining). A decrease in variance and new sources of prediction bias may be introduced if the base models are not good enough. ANN, XGBoost, and SVM all perform well independently and have distinct modeling techniques and assumptions (XGBoost is an example of a tree-based ensemble method, ANN is a type of neural network, and SVM is a discriminative model). We tried including additional models in the ensemble, which did not lead to an increase in predictive power.

After obtaining the predictions of the base models, we want to train a new meta-model on them and map "base predictions" to b_ϕ. We want to find a function $f : \mathbb{R}^3 \to \mathbb{R}$, which takes three continuous inputs (predictions of base learners) and outputs b_ϕ. We use a simple ridge regression for this task to avoid overfitting.

3.2.3 Training

To optimize the hyperparameters of the algorithms described in Section 3.2.2, and features described in Section 3.2.1, we perform a train-validation-test split of the dataset (of each galaxy sample). The test set is held aside and is only used for the final evaluation. We use 3-fold cross-validation for training the meta-model.

Regarding the size of the train-test-validation split, we use a ratio of 70-20-10 %. Because we are dealing with spatially distributed galaxies, having a random split would lead to data leakage. The galaxies in the test (or validation) set may lie nearby the galaxies from the training set and would thus be in an identical environment. It would allow the model to perform well on the test set by simply memorizing the patterns from the training set, but would not be able to generalize well. We adopt the solution of splitting the data into spatially distinct regions based on x and y spatial coordinates, which are not the line-of-sight directions.

Using the training and validation sets, we employ grid search to tune the hyperparameters. We also evaluate the performance of other combinations of input features, different options of feature transformations, different loss functions, different models, and the performance of adopting a multi-target regression problem and predicting halo mass and concentration over b_ϕ.

The ANN architecture that we use has four hidden layers with 20 neurons per layer. In addition, we use dropout and the LeakyReLU activation function. The kernel of the SVM model used is the radial basis function. All other values of the hyperparameters and the values in the grid search grid are available on GitHub[2].

3.3 Results

The final predictions made by the stacked model on the test set are presented in Fig. 5.. We also show the separate predictions for halo concentration and mass. We do not use mass and concentration predictions elsewhere, but it is useful to see them since b_ϕ is a function of these

---

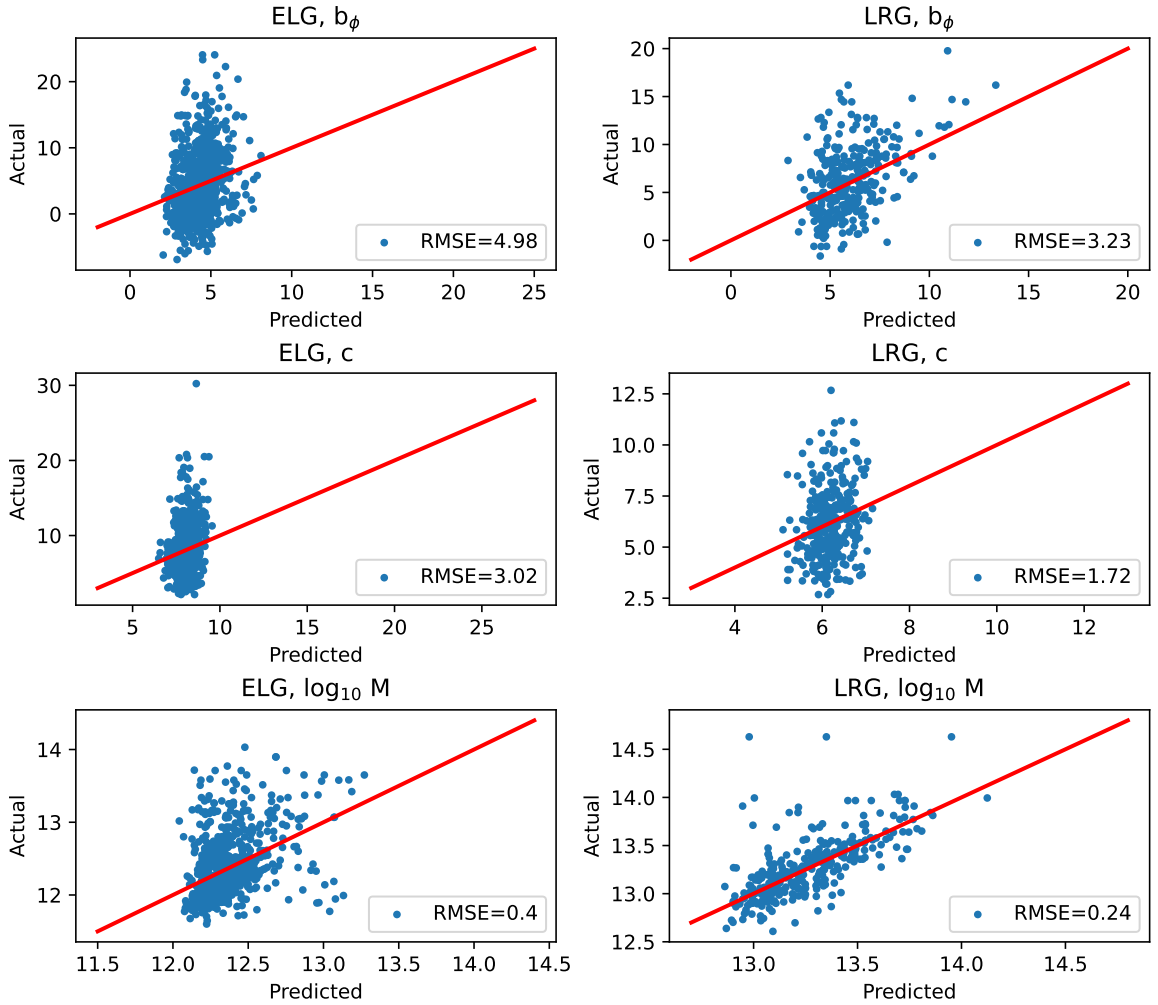[2]https://github.com/jmsull/ml_fnl_forecast

Figure 5. Results for ELG (left) and LRG (right) predictions on the test set. We present predicted versus actual values for each galaxy. The perfect predictions should lie on the red line, which has a slope of 1. Only the b_ϕ predictions (top row) are relevant for the analysis. However, we also show halo concentration and mass predictions (bottom two rows).

variables. We can see that for both ELGs and LRGs, halo mass predictions are significantly better than both b_ϕ and concentration predictions. While the b_ϕ and concentration RSME are of the same magnitude, the RMSE of mass prediction is roughly 10 times smaller.

It is hard to make direct comparison between ELG and LRG results, since the distributions of the target values are different, as seen in Fig. 2. However, it seems that the results in terms of RMSE are generally better for LRGs. The reason for this could be that, as seen in Fig. 2, LRGs live in more massive halos, which are correlated with larger environments and higher central fractions. Moreover, LRGs are older than ELGs and are thus less affected by hydrodynamical effects, which are hard to capture with the properties we are looking at.

In Fig. 6 we can see the average b_ϕ values of the tertiles, split based on predicted b_ϕ from Fig. 5. The "Ideal" dots in Fig. 6 correspond to the values of b_ϕ for splits made by actual b_ϕ values of the galaxy host halos. After these tertilles are defined we compute the mean b_1 of all halos in each tertile.
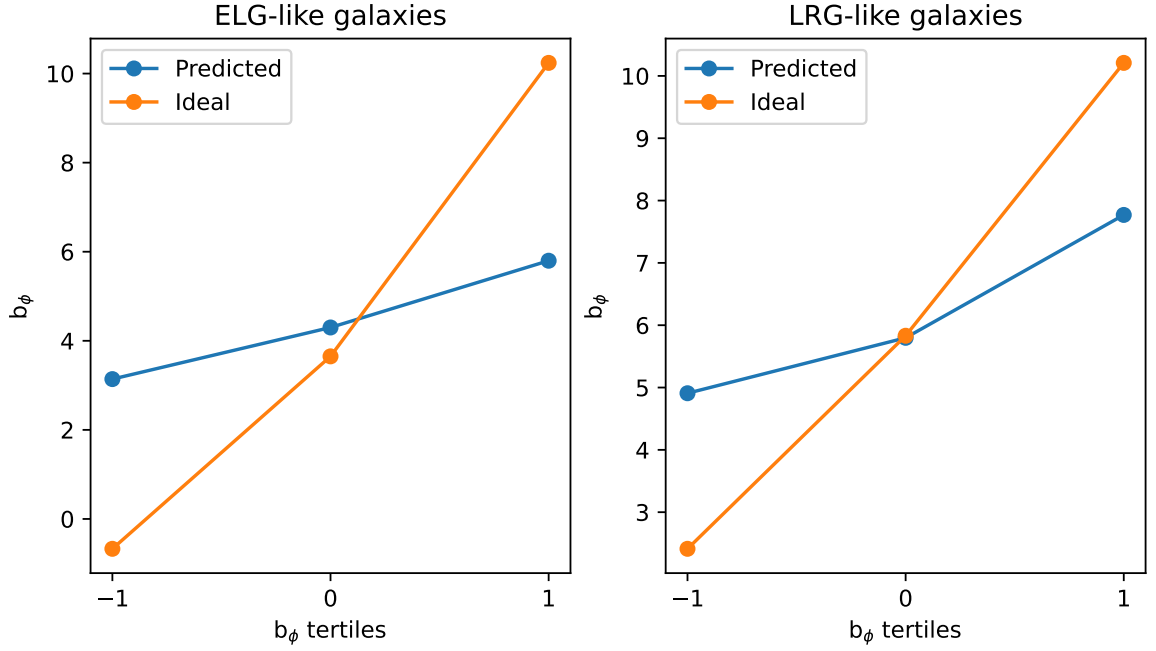
Figure 6. $\bar{b}_\phi$, for ELG and LRG samples. We split the sample into tertiles for the "Ideal" case based on the actual b_ϕ value and report the average b_ϕ for each tertile. For the "Predicted" case, we make the split based on predicted b_ϕ, as shown in Fig. 5. The horizontal axis on the plots is arbitrary and just denotes different tertiles.

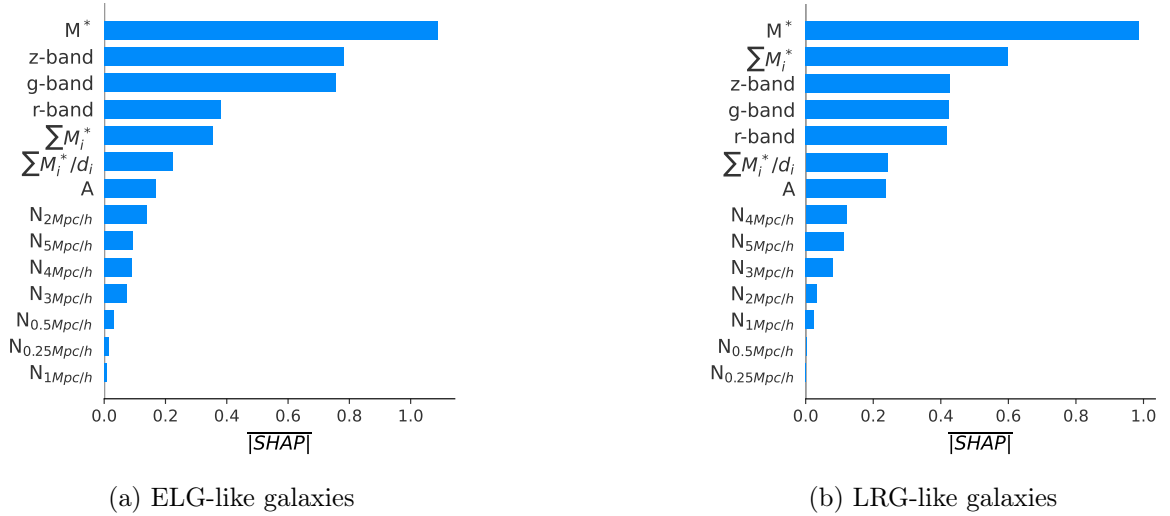(a) ELG-like galaxies

(b) LRG-like galaxies

Figure 7. SHAP values of feature importance for predictions in redshift space. We can see that for both ELG and LRG- like galaxies, intrinsic features of galaxies are of much greater importance than the properties of the neighborhood. We can also observe that the feature importance is similar for both ELGs and LRGs.

We want to look at the interpretability of our model and see which information is most important for the obtained results. For this purpose, we use Shapley Additive explanations (SHAP) values, a technique for determining feature importance [70]. The SHAP feature

importance is calculated by permutating all features and summing the average contribution of each feature over all permutations. It provides a fair and accurate estimation of the contribution of each feature, even if features are correlated, and can account for complex interactions between features[3].

The average SHAP values for each feature are presented in Fig. 7. The feature importance is presented for only two galaxy samples, however it remains practically the same for others. We can notice that, for both ELG and LRG, the intrinsic properties of the galaxies are significantly more important than environmental properties. Another interesting thing to observe is that the feature importance is similar for ELG and LRG galaxies.

We can observe that the importance of N_R does not converge to 0. This makes sense since there is a high correlation between the number of neighbors within different R. Moreover, we have explicitly checked that adding information about a larger environment does not improve results. It just takes some importance from other N_R features, while their sum would remain the same. The reasons are that the information too far from the halo does not affect the results, the dataset size is inadequate, and the correlation between halo properties and those features is noisy.

In Fig. 7, we can see that the information from the local neighborhood is much less influential than other features are when making predictions. Furthermore, we observed that adding information about luminosities and stellar masses of nearby galaxies does not improve the results. We can conclude that it thus makes sense to approach the problem with the unstructured data instead of using GNN or CNN. Those could work much better in real space, with a larger simulation (larger data size) and higher number densities.

4   Multi-tracer $f_{NL}^{\rm loc}$ forecasts

In this Section we provide Fisher forecast results for $\sigma(f_{NL}^{\rm loc})$ using a slightly extended version of the multi-tracer approach of Ref. [15]. In particular, we first provide forecasts for DESI LRGs and ELGs using the $b_\phi(b_1)$ relations found in Section 3 for mock versions of these samples. We then provide more speculative forecasts for future planned spectroscopic surveys (MegaMapper [71], SPHEREx [72]) that will target LPNG.

4.1   Forecast setup

Following [15], we perform forecasts of idealized constraints on $f_{NL}^{\rm loc}$, slightly generalizing the expressions there to a 3-tracer analysis. To be consistent with IllustrisTNG, we will use the Planck15 fiducial flat ΛCDM cosmology with $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $\Omega_\Lambda = 0.6911$, $h = 0.6774$, $\sigma_8 = 0.8159$, and $n_s = 0.9667$ in these forecasts [73, 74], and do not marginalize over ΛCDM parameters. While this choice might be considered optimistic, the multi-tracer method is relatively insensitive to marginalizing over ΛCDM parameters (see Appendix C of Ref. [75]). Furthermore, even without the multi-tracer method, we find that our single-tracer forecasts produce $\sigma(f_{NL}^{\rm loc})$ that is 10% lower than the (marginalized) values quoted in Ref. [9] or less (see Fig. 14 there) except for DESI ELGs, where the difference is 25%, though in this case, we do not use the same redshift distribution or redshift range as in Ref. [9]. We will work entirely in linear theory at the power spectrum level (including linear redshift space distortions $b \to b + f\mu^2$), and use a simple linear bias model only. We do not account for the Alcock-Paczynski effect [76]. We use 101 linear μ bins between 0 and 1, $N_z = 65$ linear z

---

[3]The more popular technique used is "permutation importance", which has the main drawback of being unable to deal with correlated features.

bins between $z_{\min}$ and $z_{\max}$, and 100 k values between $k_{\min}$ (which is survey dependent) and 0.075 h/Mpc. We verified that doubling the number of z or μ bins or k points beyond these values changes the final forecasted error for the single-tracer forecasts by less than 1%. To be consistent with Ref. [9], we perform all single-tracer forecasts with the UMF prediction with $p = 1$. For a more detailed discussion of a careful treatment of the choice of p in the context of real data, including the choice of priors, we refer the reader to Ref. [33]. We neglect wide-angle effects and large-scale relativistic effects on the galaxy power spectrum. We neglect any additional systematic effects that could contaminate a large-scale measurement of the power spectrum (e.g. stellar contamination), though it is of course extremely important to characterize these well in a real data analysis.

Here we describe the galaxy samples used in our forecasts. The primary samples we consider are the Dark Energy Spectroscopic Instrument (DESI) Luminous Red Galaxies (LRGs) and Emission Line Galaxies (ELGs). For the DESI LRG-like sample we use the redshift distribution of [77], and for the DESI ELG-like sample we use the redshift distribution of [60]. For DESI $f_{\rm sky} = 0.34$ [52], $k_{\min} = 0.0023$ h/Mpc, and $z_{\min} = 0.0$, $z_{\max} = 2.0$. For the linear galaxy bias of the two DESI samples in the single-tracer analysis, we make the empirically calibrated choice [52, 78] of $b_{\rm LRG} = \frac{1.7}{D(z)}$, $b_{\rm ELG} = \frac{0.84}{D(z)}$, where $D(z)$ is the linear growth factor (normalized to 1 today) computed in our fiducial cosmology. For the multi-tracer forecasts using the results of Section. 3, we assume the results evolve similarly with $D(z)$ (i.e. $b(z) = \frac{D(z=1)}{D(z)}b$, and similarly for b_ϕ).

For the MegaMapper forecasts, we follow the prescriptions of Refs. [9, 71, 79, 80] for the galaxy density $n(z)$ and linear bias $b(z)$, we assume $f_{\rm sky} = 0.34$, $k_{\min} = 0.0017$ h/Mpc, and $z_{\min} = 2.0$, $z_{\max} = 5.0$. For SPHEREx, we use $f_{\rm sky} = 0.65$, $k_{\min} = 0.001$ h/Mpc, and $z_{\min} = 0.1$, $z_{\max} = 3.0$. For the SPHEREx forecasts, we assign linear bias as a function of redshift based on the procedure outlined in Ref. [72], but use the fitting functions of Ref. [81] as implemented in COLOSSUS [82] (and have verified that this choice changes the halo bias of individual redshift error samples by less than 1% with respect to the fitting functions of Refs. [83, 84]). We consider the redshift-error samples (labeled by $\frac{\sigma_z}{1+z}$) on their own and in combination. We incorporate the SPHEREx redshift errors as described in Ref. [72] for the various redshift error samples, multiplying the shot noise term by $\exp(-\left[k\mu\sigma_z\frac{d\chi}{dz}\right]^2)$. All non-SPHEREx forecasts in this paper assume zero redshift error. When combining several redshift error samples, we always use the largest redshift error. For the fully-combined SPHEREx sample, this corresponds to an assumed redshift error of $\frac{\sigma_z}{1+z} = 0.2$, so we deviate slightly from the treatment of Ref. [9], who used $\frac{\sigma_z}{1+z} = 0.05$ for this full sample, and therefore we are more pessimistic in our forecasts. We compute an effective bias for combined samples by weighting the bias by the sample number density.

We work in linear theory and at a fiducial cosmology where $f_{NL}^{\rm loc} = 0$, and so use the Fisher matrix for $\delta_X(k, \mu, z)$ which, in the Gaussian approximation is [85]:

$$F_{\alpha\beta} = \sum_{i_z=1}^{N_z-1} \int_0^1 d\mu \int_{k_{\min}}^{k_{\max}} N_{k,i_z} {\rm Tr}\left[\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\beta}\mathbf{C}^{-1}\right] \tag{4.1}$$

where $\mathbf{C}$ can be either the two-tracer version $\mathbf{C}^{(2)}$ or the three-tracer version $\mathbf{C}^{(3)}$, with

$$\mathbf{C}^{(2)} = \begin{pmatrix} C_{XX} & C_{XY} \\ & C_{YY} \end{pmatrix}, \quad \mathbf{C}^{(3)} = \begin{pmatrix} C_{XX} & C_{XY} & C_{XZ} \\ & C_{YY} & C_{YZ} \\ & & C_{ZZ} \end{pmatrix}, \tag{4.2}$$

where the field covariance and number of modes N_{k,i_z} in a volume shell V_{i_z} are

$$C_{XY} = \left(b_1^X + \mu^2 f + b_\phi^X f_{NL}^{\mathrm{loc}} \mathcal{M}^{-1}(k)\right) \left(b_1^Y + \mu^2 f + b_\phi^Y f_{NL}^{\mathrm{loc}} \mathcal{M}^{-1}(k)\right) P_L(k) + \frac{\delta_{XY}^D}{\bar{n}_X}, \quad (4.3)$$

$$N_{k,i_z} = \frac{dk k^2 V_{i_z}}{2\pi^2}, \quad (4.4)$$

respectively, and we neglect both tracer non-Poisson stochasticity and cross-stochasticity between different tracer fields [75, 86, 87]. Here i_z gives the index of the $z-$bin midpoint used to compute Fisher matrix element, and V_{i_z} is a spherical shell computed using the adjacent redshift bin edges and is scaled by f_{sky}.

Here, in the notation of Ref. [15], $\alpha = \frac{b_1}{b_2}$ and $P_2 = b_2^2 P$ (at our fiducial value of $f_{NL}^{\mathrm{loc}} = 0$), where b_1 and b_2 are the *linear* biases of the first and second samples under consideration. We define $\beta \equiv \frac{b_3}{b_2}$ analogously for the three-tracer case. Explicit expressions for the Fisher matrix element $F_{f_{NL}^{\mathrm{loc}} f_{NL}^{\mathrm{loc}}}$ are provided in Appendix A and in the public code accompanying this paper (linked above).

4.2 Results for $\sigma(f_{NL}^{\mathrm{loc}})$

4.2.1 DESI

Sample	$\sigma_{\mathrm{ST}}(f_{NL}^{\mathrm{loc}})$	$\sigma_{2,\mathrm{MT}}(f_{NL}^{\mathrm{loc}})$	$\sigma_{2,\mathrm{MT}}^{(P)}(f_{NL}^{\mathrm{loc}})$	$\sigma_{3,\mathrm{MT}}(f_{NL}^{\mathrm{loc}})$	$\sigma_{3,\mathrm{MT}}^{(P)}(f_{NL}^{\mathrm{loc}})$
ELG, ideal	7.1	1.3	1.4	1.7	1.5
ELG, pred.	.	10	2.5	12	2.3
LRG, ideal	5.0	3.0	2.6	3.5	3.4
LRG, pred.	.	16	3.7	17	4.7

Table 2. Local primordial non-Gaussianity forecasts for single- and multi-tracer forecasts for ELG and LRG-like simulated galaxies. For each column $\sigma_{i,XT}^{(P)}$, i denotes the number of tracers used in each forecast, (P) indicates whether the forecast includes P_2 as a parameter, and $X = \{S, M\}$ is S for single-tracer and M for multi-tracer. For the 2-tracer forecasts, we use the highest and lowest b_ϕ tertiles.

Here we present $\sigma(f_{NL}^{\mathrm{loc}})$ forecast results - we first perform forecasts for LRGs and ELGs individually with several multi-tracer setup choices (Table 2), then present results illustrating the effect of selection choice on $\sigma(f_{NL}^{\mathrm{loc}})$ for ELGs (Table 3) before finally considering the case of ELGs and LRGs in their overlap region (Table 4).

Table 2 shows the improvement of the multi-tracer forecasts over the single-tracer forecasts for several choices of multi-tracer setup for our fiducial choice of DESI mock galaxy selection ($\bar{n}_{\mathrm{ELG}} = 5 \times 10^{-4}$ $[h^{-1} \mathrm{Mpc}]^{-3}, \log_{10}\left(\frac{sSFR}{h\ yr^{-1}}\right) = -9.09$). Here we report results for both the simulated LRG and ELG galaxy samples, for the ideal values of $b_\phi(b_1)$ in each tertile (i.e. concentration information is recovered perfectly) and for the predicted ("pred.") values using the machine learned relationship $b_\phi(b_1)$ from observable galaxy properties. While clearly the ML model cannot access all the f_{NL}^{loc} information contained in halo mass and concentration, it recovers enough information to significantly improve the forecasted LPNG amplitude error $\sigma(f_{NL}^{\mathrm{loc}})$. The improvement for the predicted relations is largest for ELGs, where we can obtain a forecasted error of $\sigma(f_{NL}^{\mathrm{loc}}) = 2.3$, a factor of 3 improvement over the single-tracer case.

We show both 2-tracer and 3-tracer forecasts. The 2-tracer forecasts use the upper and lower tertiles of the three-bin splits, while the 3-tracer forecasts use all three tertiles (both cases use number density $\bar{n}/3$ for each tracer) When making multi-tracer forecasts, we compute the error on f_{NL}^{loc} using two different procedures - first, when the only parameters considered in the Fisher matrix is (are) the relative amplitude(s) α, (β), and second, when both the relative amplitude(s) α, (β,) and P_2, the power spectrum corresponding to the second sample are used. Since in the first case we use only the relative amplitude for the two-tracer forecasts, we refer to the case of including P_2 as a parameter as "(P)", since we are including the power spectrum as a parameter. Including P_2 as a parameter generally reduces the forecasted error, however, in the sample-variance limit P_2 adds no information (see discussion around eqn. A.6). In multi-tracer forecasts used in the rest of this paper, we include P_2 as a parameter. Some brief further discussion of these forecasting aspects can be found in Appendix A.

$\log_{10}\left(\frac{sSFR}{h\ \text{yr}^{-1}}\right)$	-9.09		-9.23	
$n_{\text{ELG}}\ [h^{-1}\ \text{Mpc}]^{-3}$	ideal	pred.	ideal	pred.
5×10^{-4}	1.5	2.3	1.4	2.3
7×10^{-4}	1.6	2.3	1.5	2.2
1×10^{-3}	1.6	2.5	1.5	2.2

Table 3. Multi-tracer $\sigma(f_{NL}^{\text{loc}})$ Fisher forecasts for several choices of $\bar{n}_g$ and sSFR for DESI ELG-like simulated TNG galaxies. Results do not vary significantly with the choice of mock galaxy selection.

Table 3 characterizes the effect of the simulated galaxy sample selection strategy on the forecasts for $\sigma(f_{NL}^{\text{loc}})$. These selections were mentioned already in Section 3 (in the context of the learned $b_\phi(b_1)$ relation), but here we supplement this by checking the effect of the choice of specific star formation rate split and stellar mass threshold, which determines the number density for a fixed sSFR split, on the final forecasted f_{NL}^{loc} error. We find that the final results are mostly insensitive to these choices, as they vary only by a factor of 17% or less (within the ideal or predicted cases). It may appear counter-intuitive that $\sigma(f_{NL}^{\text{loc}})$ increases even in the ideal case for both specific star-formation rates as the number density increases, but this is due to the fact that the mean linear bias and, at first approximation, the LPNG bias both shift downward as the number density increases. In more detail, while the b_ϕ values drop uniformly across all tertiles with the mean, the b_1 value in the highest tertile hardly changes with increasing number density, but the lower tertile drops significantly, and this leads to a lower change in linear bias between the upper and lower tertiles at higher number density, which in turn leads to a smaller relative amplitude Jacobian factor in the Fisher matrix (see eqn. 4.5). In any event, the highest difference of 17% in $\sigma(f_{NL}^{\text{loc}})$ between the two choices of sSFR is for the highest number density sample (which is not used elsewhere).

Table 4 shows $\sigma(f_{NL}^{\text{loc}})$ Fisher forecasts for several multi-tracer sample configurations drawn from the ELG and LRG simulated galaxy samples. To represent the simplest possible multi-tracer forecast setup using the undivided LRG and ELG samples, we report $\sigma(f_{NL}^{\text{loc}}) = 4.0$ for the UMF bias prediction (eqn. 2) with the parameter p set to 1. This forecast does not make use of the learned $b_\phi(b_1)$ relations presented in this paper. We next consider a similar case that uses the mean of the $b_\phi(b_1)$ values for the simulated LRG sample and the simulated ELG sample (with the fiducial choice of sample parameters: $\bar{n}_{ELG} = 5 \times 10^{-4}$, $\bar{n}_{LRG} = 2 \times 10^{-4}$, and $\log_{10}(\text{sSFR}) = -9.09$).

ELG + LRG	$\sigma(f_{NL}^{\mathrm{loc}})$
$p = 1$	4.0
(2) ($\overline{\mathrm{LRG}}$, $\overline{\mathrm{ELG}}$), ideal	2.3
(2) ($\overline{\mathrm{LRG}}$, $\overline{\mathrm{ELG}}$), pred	2.3
(2) (LRG+, ELG+), ideal	1.4
(2) (LRG+, ELG+), pred	2.4
(3) (LRG-, LRG+, ELG-), ideal	0.8
(3) (LRG-, LRG+, ELG-), pred	2.0
(3) (LRG-, ELG+, ELG-), ideal	0.8
(3) (LRG-, ELG+, ELG-), pred	2.0
(3) (LRG-, ELG+, else), ideal	0.6
(3) (LRG-, ELG+, else), pred	1.5

Table 4. Local primordial non-Gaussianity amplitude f_{NL}^{loc} multi-tracer forecasts for ELG-LRG in their overlap region. The entries listed "($\overline{\mathrm{LRG}}$, $\overline{\mathrm{ELG}}$)" use the mean $b_\phi(b_1)$ prediction with the full number density of each sample Here "+" and "-" denote the top and bottom tertile values of $b_\phi(b_1)$ of the respective samples, respectively. The row entries following the ($\overline{\mathrm{LRG}}$, $\overline{\mathrm{ELG}}$) rows use (i)-tracer forecasts ($i = 2, 3$) with number densities of $\bar{n}/3$ for each sample. In the final two rows, the unused subsamples are combined with bias weighted by their number densities.

When we consider the multi-tracer forecasts that use two samples split by b_ϕ tertile, we find a lower $\sigma(f_{NL}^{\mathrm{loc}})$. When using the upper tertile samples for both ELGs and LRGs (LRG+,ELG+), we find a factor of 2.9 reduction in $\sigma(f_{NL}^{\mathrm{loc}})$ compared to the $p = 1$ UMF forecast in the ideal case, where concentration and halo mass information are perfectly recovered, and a factor of 1.7 reduction in $\sigma(f_{NL}^{\mathrm{loc}})$ when using the learned relation. We also consider three-tracer forecasts for the ELG and LRG tertile samples. Using the lower and upper tertiles for the LRGs and lower tertiles for the ELGs (LRG-,LRG+,ELG-), we find smaller forecasted errors than in the two-tracer case, and using the lower and upper tertiles for the ELGs along with the lower tertile for the LRGs (LRG-,ELG+,ELG-) provides similar results, reducing the forecasted error by a factor of 5 compared to the $p = 1$ UMF forecast in the ideal case and a factor of 2 in the predicted case.

Finally, based on the promising results of the (LRG-,ELG+,ELG-) choice of subsamples, we show in the last two rows of Table 4 forecasts for the case where ELG- in this triplet is replaced with all the other sub-samples combined. In particular, the third sample is the combination of ELG-, ELG0, LRG0, and LRG+, where "0" indicates the central tertile. Bias values for these samples are weighted by their number density (as for the SPHEREx redshift error $\frac{\sigma_z}{1+z}$ samples). In this case, we find the most constraining forecasts we report for LRGs and ELGs, with $\sigma(f_{NL}^{\mathrm{loc}}) = 0.6$ and $\sigma(f_{NL}^{\mathrm{loc}}) = 1.5$ for the ideal and predicted b_ϕ values, which are now factors of 6.7 and 2.7 reductions over the naive $p = 1$ multi-tracer forecast, respectively.

We do not list all possible permutations of the ELG and LRG tertile samples (and their combinations) for brevity, though some of the other combinations produce a lower $\sigma(f_{NL}^{\mathrm{loc}})$ in the predicted case than we show here. It would also be interesting to further consider optimally combining split subsamples in general [e.g. as for mass in 75, 88] - we took a first step in exploring this here by using the "else" samples in Table 4, which produces excellent results.

We now discuss the origin of the sizeable reduction in our forecasts for $\sigma(f_{NL}^{\mathrm{loc}})$ over the

single-tracer case. The Jacobian factor involved in converting from the Fisher information on the relative amplitude multi-tracer parameters (F_{ij} where $i, j \in \{\alpha, \beta\}$) to that of LPNG ($F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}$) involves the factor (reproduced from Appendix A):

$$\frac{\partial \alpha}{\partial f_{NL}^{\rm loc}} = \alpha \left(\frac{b_{\phi,1}}{b_1} - \frac{b_{\phi,2}}{b_2} \right) \mathcal{M}^{-1}(k), \qquad (4.5)$$

which is evaluated at $f_{NL}^{\rm loc} = 0$. For a fiducial choice of $f_{NL}^{\rm loc} = 0$, the Jacobian factor of eqn. 4.5 is in fact the only place that the LPNG bias enters the Fisher forecast calculation. A similar expression holds for $b_3, b_{\phi,3}$ and β in the 3-tracer case.

If we consider the limit where the relative amplitude $\alpha \to 1$, then clearly this factor is proportional to Δb_ϕ. This limit is exact in a 2-way split of a single halo mass bin if linear bias does not evolve across the bin or with the splitting parameter.

We now make the even further simplification, for illustrative purposes, that each b_ϕ of the two samples in question is described by a UMF relation for some choice of the parameter p, and so $b_{\phi,1} = \delta_c(b_1 - p_1)$, $b_{\phi,2} = \delta_c(b_2 - p_2)$, and for $\alpha \to 1$ we have $\Delta b_\phi \to \delta_c \Delta p$, and $\frac{\partial \alpha}{\partial f_{NL}^{\rm loc}} \to -\delta_c \Delta p \mathcal{M}^{-1}(k)$. From this demonstration, it would appear that we can drive our forecasted error on LPNG as low as we wish as we take $\Delta p \to \infty$. We further discuss the implications of this apparent conclusion in the next section.

4.2.2 Future LPNG surveys

To explore the potential of using samples with large differences in b_ϕ further, we apply our 2-way forecasting framework to two future surveys that will target LPNG, MegaMapper [71], and SPHEREx [72] using the setup described in Section 4.1. We use the UMF relations as input for $b_\phi(b_1)$ in these forecasts, as we do not have the information necessary to construct detailed simulated galaxies and corresponding learned $b_\phi(b_1)$ for the future MegaMapper or SPHEREx galaxy populations. Since we do not have a solid basis for selecting Δp for these forecasts, we instead consider forecasts at a range of Δp. For the following multi-tracer forecasts, we will set $b_{\phi,i} = \delta_c(b_i - [p \pm \Delta p])$ where $p = 1$ unless otherwise indicated[4]. The value $p = 1$ is the best choice to make contact with the literature, which mostly uses the UMF with $p = 1$. In principle, if we were to choose a lower p, we should only reduce $\sigma(f_{NL}^{\rm loc})$ (see Fig. 8) - this is closely related to the zero-bias effect discussed by Ref. [89].

Figure 8 shows the resulting dependence of the two-tracer forecasted error $\sigma(f_{NL}^{\rm loc})$ on Δp for MegaMapper. The solid blue line shows the two-tracer forecast for the scenario in which the number density of the MegaMapper sample is split into thirds, and the top and bottom tertiles of the sample are assigned the *same* linear bias b and an LPNG bias of either $b_\phi = b - (1 \pm \Delta p)$. When $\Delta p \to 0$, we find that we approach the single-tracer results at $p = 1$, (since for $\Delta p = 0$ the relative amplitude Jacobian factors vanish and we are left with only P_2 information). It is immediately apparent that there is no bound on how much the multi-tracer forecast can improve upon the single-tracer forecast at fixed p.

To make a fairer comparison to the single tracer case, we also plot in the dashed lines the single tracer forecasts with $p = 1 + \Delta p$ (green) and $p = 1 - \Delta p$ (orange). We can see that by drastically increasing or decreasing p in the single-tracer case we can obtain a large reduction in $\sigma(f_{NL}^{\rm loc})$ that is qualitatively similar to that of the multi-tracer case. In the limit of $\Delta p \to \infty$ we see that the multi-tracer case always outperforms the single-tracer case by a roughly constant factor, and both lines have similar slope. This is expected due to SV

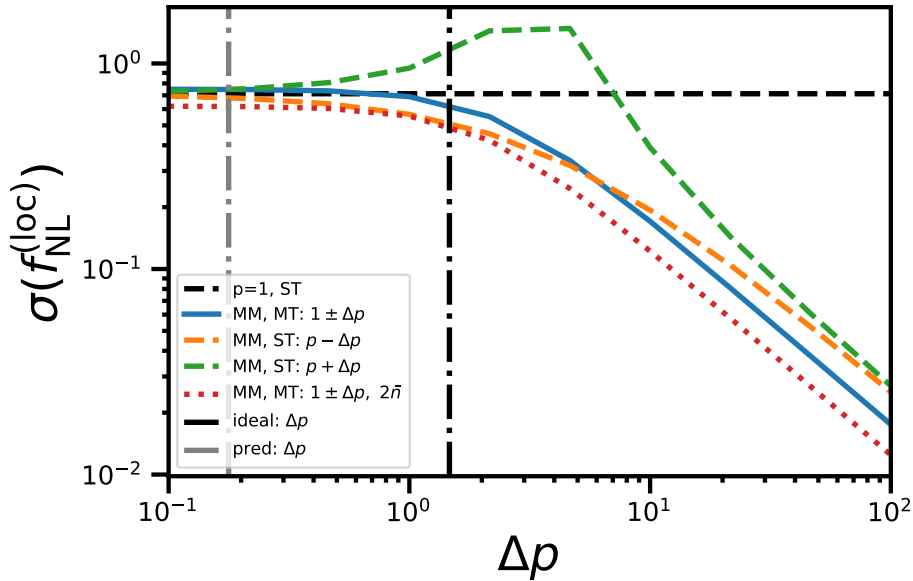[4]This is in a sense a somewhat pessimistic choice, see Appendix A for more discussion

Figure 8. (*MegaMapper*): The dependence of forecasted $\sigma(f_{NL}^{\rm loc})$ multi-tracer forecasts on Δp, the difference in the UMF parameter, for the planned MegaMapper survey. The black dashed line shows the single-tracer (ST) forecast with the original UMF scenario, for which $p = 1$. The solid blue line shows the multi-tracer (MT) forecast for the case in which the full MegaMapper sample is split into tertiles and the top and bottom tertiles are used with equal number density bins of $\frac{\bar{n}}{3}$, each of which is assigned $b_\phi = 1 \pm \Delta p$. For comparison, the dashed colored lines show the single tracer forecasts with $p = 1 + \Delta p$ (green dashed) and $p = 1 - \Delta p$ (orange dashed). The red dotted line is the same as the blue line (multi-tracer forecast), but with twice the number density. The horizontal dash-dotted lines in gray and black show the approximate values of Δp for the learned and ideal b_ϕ values for the mock DESI galaxy samples.

cancellation and the fact that (for fixed b) in the large-p limit, we have that $\sigma_{ST}(f_{NL}^{\rm loc}) \propto \frac{b}{p}$, and for large Δp, $\sigma_{MT}^{(2)}(f_{NL}^{\rm loc}) \propto \frac{b}{\Delta p}$. These single-tracer forecasts use the full number density of the sample, so to account for this difference when comparing to multi-tracer, we also artificially increase the number density by a factor of two in the multi-tracer case as an additional point of comparison (red dotted line). Here we can see that the slight improvement of single-tracer over multi-tracer around $\Delta p = 1$ is due to the reduced number density of the multi-tracer sample.

We also mark in vertical dash-dotted gray and dash-dotted black the values of Δp that are close to describing b_ϕ for the upper and lower tertiles of the ELG and LRG simulated galaxies in the predicted and ideal cases, respectively[5]. While this exercise illustrates in part where the improved constraints for DESI galaxies is coming from, the change in Δp is of course not the only thing driving a difference in the mulit-tracer and single-tracer forecasts. As in the usual halo-mass-based multi-tracer application, b_1 still plays a significant role in both the Jacobian factor of eqn. 4.5 and the Fisher matrix elements of the multi-tracer parameters (e.g. through factors of the form $b^2 P \bar{n}$). For anything but the idealized sample-variance limit, the increased number density from splitting the full sample can also strongly affect the final

---

[5]These values are simply calculated as $\Delta p = \frac{1}{2}(\delta^c)^{-1}(b_\phi^+ - b_\phi^-)$, so there is a symmetry here that isn't present in the ELG/LRG results. For more discussion see Appendix A
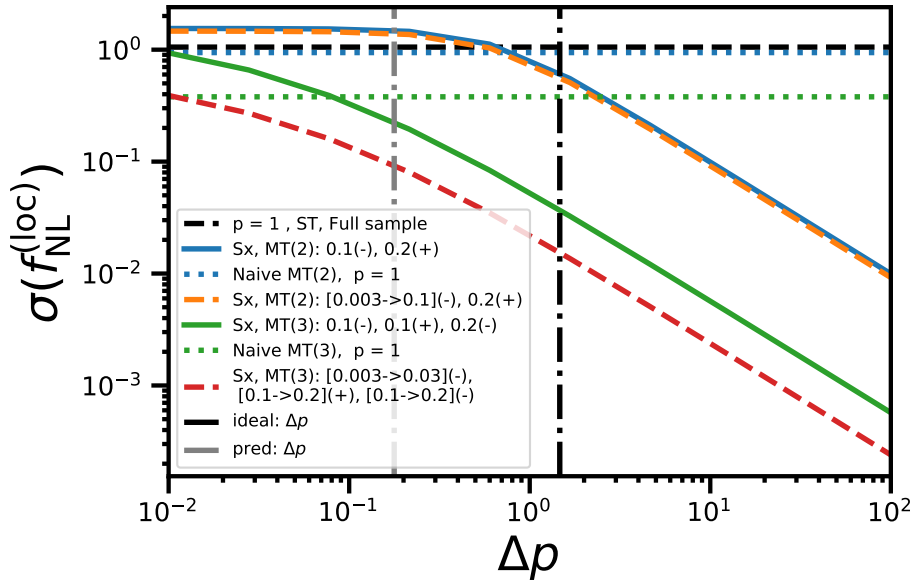
Figure 9. (*SPHEREx*): Similar to Figure 8, but for multi-tracer forecasts involving multiple SPHEREx redshift error samples. The black dashed line again shows the UMF single-tracer forecast. The solid blue line shows the two-tracer (2) multi-tracer (MT) forecast for the $\frac{\sigma_z}{1+z} = 0.01$ and $\frac{\sigma_z}{1+z} = 0.2$ redshift error samples, where the former has UMF parameter $p = 1 - \Delta p$ ("-") and the latter has $p = 1 + \Delta p$ ("+"). The orange dashed line shows the two-tracer MT forecast where the first tracer is the lowest redshift error sample ($\frac{\sigma_z}{1+z} = 0.003$) with $p = 1 - \Delta p$ and the second tracer is the combined sample of all other redshift error samples (with $p = 1 + \Delta p$). The green solid line shows the three-tracer (3) multi-tracer forecast for the three splits of the redshift error samples - the $\frac{\sigma_z}{1+z} = 0.01$ sample with $p = 1 - \Delta p$ (-), the $\frac{\sigma_z}{1+z} = 0.1$ sample with $p = 1 + \Delta p$ (+), and the $\frac{\sigma_z}{1+z} = 0.2$ sample with $p = 1 - \Delta p$ (-). The red dashed line shows another three-tracer (3) forecast using combined samples - the combined $\frac{\sigma_z}{1+z} = 0.003, 0.01$ samples with $p = 1 - \Delta p$ (-), the combined $\frac{\sigma_z}{1+z} = 0.03, 0.1, 0.2$ samples with $p = 1 + \Delta p$ (+), and the second half of the same sample with $p = 1 - \Delta p$ (-). Again, horizontal dash-dotted lines in gray and black show the approximate values of Δp for the learned b_ϕ for the mock DESI galaxy samples.

forecasted error. Further study of physical and practical limits on Δp will be the subject of future work.

We performed a similar exercise for SPHEREx in Figure 9, which shows forecasted $\sigma(f_{NL}^{\rm loc})$ as a function of Δp for several combinations of SPHEREx redshift error samples. In black dashed we again show the single-tracer forecast for $p = 1$ for the combined sample of all SPHEREx redshift error samples (using the most pessimistic redshift error). Multi-tracer forecasts for individual redshift error samples (labeled by $\frac{\sigma_z}{1+z} \in \{0.003, 0.01, 0.03, 0.1, 0.2\}$) where we assign splits on Δp are shown by the solid lines. We also show "naive" two-tracer forecasts when considering different redshift error samples using the same p (but these samples have different b). To take advantage of the high number density of SPHEREx, we also show multi-tracer forecasts for combinations of redshift error samples (for example, combining the two lowest redshift error samples to create one multi-tracer sample and combining the three highest redshift error samples to create the second multi-tracer sample) as the dashed lines. Here the "MT(i)" and ($\pm$) labels serve the same purpose as in Table 4, indicating how many tracers are used in the forecast and whether the top or bottom tertile of the Δp-split sample

is used.

We find similar results to Fig. 8, in that the multi-tracer forecasts improve over the single-tracer forecasts as Δp increases. First, we discuss the case where we use individual redshift error samples as the multi-tracer samples (solid lines). The naive forecasts (dotted lines) are also for the individual redshift samples, and should be compared to lines of the same color.

At low Δp the improvement of the two-tracer Δp multi-tracer forecast over the full single-tracer forecast (black-dashed) with the full sample disappears since both are dominated by the most pessimistic redshift error of $\frac{\sigma_z}{1+z} = 0.2$. This is in fact worse than the naive multi-tracer case (dotted blue line) due to the lower number density. Similar statements hold for the three-tracer case, though the three-tracer results show what is effectively only the high Δp regime of the 2-tracer case, and, we can see that the difference between the naive and Δp-split three-tracer forecasts is larger than in the two-tracer case.

Turning now to the combined samples, we see that combining several redshift samples (with the appropriated weighted linear biases) leads to a reduction in shot noise[6] due to increased number density of the sample, and this gives lower $\sigma(f_{NL}^{\rm loc})$ when $\Delta p \to 0$ as seen in the difference between solid and dashed multi-tracer lines. We can also see from the red dashed line that the improvement over single-tracer is similar to the red dotted curve in Fig. 8, again suggesting that the uniform reduction in $\sigma(f_{NL}^{\rm loc})$ over the three-tracer single sample forecast (green solid line) is due to the increased number density of the combined samples. Essentially all of the lowest error multi-tracer forecasts come from the high redshift error samples (or combinations that use the highest redshift error), as these have the highest number density (roughly $2 - 3 \times 10^{-3}$ $[h^{-1} \ {\rm Mpc}]^{-3}$), largest redshift range (extend out to $z = 3$), and the measurement of $f_{NL}^{\rm loc}$ is relatively insensitive to these errors.

4.3 Discussion

We are leveraging information about b_ϕ through the portal of halo mass and concentration, but in galaxy survey data we only have access to observable galaxy properties, not the concentration itself. We have fixed the galaxy-halo connection (GHC) that maps the halo field to the galaxy field for a specific galaxy sample to that afforded by the IllustrisTNG galaxy formation model. Even assuming this fixed GHC, we will inevitably lose information about the halo mass and concentration when passing to simulated galaxies. For this reason, we report multi-tracer forecasts for both "ideal" and "predicted" b_ϕ for the ELG and LRG galaxy samples, where the predicted case captures the limitations of using only observable galaxy properties rather than host halo properties. The former refers to the (fictitious) scenario in which we could perfectly recover b_ϕ from the simulated galaxy sample based on mass and concentration, and the latter to what the ML model of Section 3 recovers. The "ideal" forecasts results should therefore be interpreted as quantifying the maximal b_ϕ information available from the given galaxy sample, while the "predicted" forecasts tell us what information about b_ϕ can be obtained by applying the ML predictions directly to DESI data (assuming the TNG galaxy formation model, and that the observable input features have actually been measured). Following this initial work, several strategies could be employed to improve the ML prediction of b_ϕ. The larger MilleniumTNG simulation [90] could be used, which would increase the size of the datasets used by a factor of almost 15. We also found, as expected, that results were

---

[6]The redshift error is unaffected, as we always choose the most pessimistic redshift error for a combined sample.

much improved in real space compared to redshift space, so adding a reconstruction procedure [91] could potentially improve results. Another option would be to use additional galaxy properties that are more challenging to model as input features, such as stellar half-mass radius, metallicity, or maximum value of spherically-averaged rotation curve.

We have made several simplifying assumptions in this work when applying the $b_\phi(b_1)$ results of Section 3 to the Fisher forecasts of this section. We have ignored the impact of fiber collisions, which, for real galaxies, would impact the ability to determine close pairs of galaxies and to compute the local environment statistics used for determining b_ϕ from observables. However, we expect that since local environment features are significantly less informative than the magnitude bands and stellar mass, including fiber collisions would have a minimal impact on our results. We used the concentration-LPNG bias relation of Ref. [35] which uses different simulations than those of IllustrisTNG itself, but since Ref. [35] uses logarithmized concentration, we do not expect a significant effect from this choice. We neglected the exact redshift dependence of the linear bias b, instead using the linear growth factor to account for the redshift dependence as in Ref. [52]. While here we only worked with halo concentration and mass, there is certainly no a priori reason to ignore halo properties such as age (e.g. vis a vis recent mergers or formation time [24, 34]), and future work should further address the connection of these properties with $f_{NL}^{\rm loc}$. The most significant limitation of this work is that we used only the IllustrisTNG galaxy formation model - further work should of course be done to consider alternative galaxy formation models to establish the robustness of (or lack thereof) the ability to calibrate $b_\phi(b)$ using observable galaxy properties.

While we were finishing this manuscript, the preprint of Ref. [92] appeared presenting a two-tracer investigation of two IllustrisTNG galaxy populations selected by various secondary properties (such as halo concentration and $g - r$ color). While the focus of our work and Ref. [92] are similar, here we focus on DESI ELG and LRG galaxies in redshift space, specific future survey forecasts, and report $\sigma(f_{NL}^{\rm loc})$, rather than the improvement over single-tracer information. We consider both Ref. [92] and our work to be complementary, as Ref. [92] provides a clear and simplified explanation of the results of Ref. [15] applied to secondary galaxy properties, while also characterizing systematic effects and priors on b_ϕ, which we largely ignore here. In the other direction, we have provided a significantly more realistic treatment of how splitting by b_ϕ on a secondary parameter (in the special case of halo concentration) can be used in a multi-tracer setting for current and upcoming spectroscopic surveys.

We also developed the machine learning tools described in Section 3, which incorporate several observable quantities to predict b_ϕ. Our ability to constrain $f_{NL}^{\rm loc}$ is, however, limited by the dependence of b_ϕ on concentration and mass, while Ref. [92] makes use of the dependence b_ϕ on quantities such as $(g - r)$ color directly through hydrodynamical Separate Universe simulations. It would be interesting to go beyond both our work here and the work of Ref. [92] to further explore the optimal choice of observable characteristics by which to split a tracer population to obtain maximally different b_ϕ values with ML methods using different galaxy formation models.

5   Conclusions

Galaxy surveys contain information about the field content of the inflationary universe through the influence of local Primordial non-Gaussianity on dark matter halo formation. To extract this information, we must assume a relationship between the number density of LSS tracers

and primordial fields, which is quantified by the LPNG bias parameter b_ϕ. In this work, we trained a machine learning model to connect this parameter to observable properties of DESI galaxies by way of host halo mass and concentration using simulated IllustrisTNG galaxies. We then used the learned relationship between the linear bias b_1 and LPNG bias b_ϕ for simulated galaxies to perform multi-tracer Fisher forecasts for the amplitude of LPNG $f_{NL}^{\rm loc}$ in DESI using galaxy power spectra. We also illustrated that for future LSS galaxy surveys that will target LPNG (MegaMapper, SPHEREx), there is potentially significant untapped information on $f_{NL}^{\rm loc}$ that can be extracted with the multi-tracer method when the samples can be split by host halo concentration.

We summarize our major conclusions here:

- We find that when considering simulated IllustrisTNG galaxies in a realistic setting (with observed DESI number densities, and including redshift-space distortions), the ML method can only extract part of the information about halo concentration from observable properties. This illustrates the challenge of using (unobservable) halo properties for accessing information on LPNG. We may however be limited by the small size of IllustrisTNG training data, rather than by the predictive power of observables, in which case our predictions are conservative.

- Nevertheless, our ML method extracts predictions for b_ϕ that contain enough information to allow multi-tracer Fisher forecasts of $\sigma(f_{NL}^{\rm loc})$ to greatly improve upon single-tracer forecasts. For certain sub-sample selections of DESI ELGs and LRGs, we find reductions of $\sigma(f_{NL}^{\rm loc})$ that are frequently a factor of 50%, and up to a factor of 3.

- We also provide forecasts for the "ideal" case in which b_ϕ information is perfectly known for each halo. In that case, even further reductions in $\sigma(f_{NL}^{\rm loc})$ are possible - up to a factor of 5, motivating further improvement of the methods developed here.

- We argue that future spectroscopic surveys targeting PNG (such as SPHEREx and MegaMapper) could greatly improve their constraining power on LPNG by employing a strategy similar to the one presented here, and support this argument with UMF-based forecasts for these surveys using varying Δp. More careful investigation of the potential sub-sample multi-tracer gains in LPNG information from these surveys is required.

It will be interesting to further explore to what extent the forecasts we have presented are attainable in real data analyses, and to what extent it is physically possible to obtain very large values of Δb_ϕ in cosmologies that are consistent with observed data. While here we have only explored the case of galaxy multi-tracer two-point functions; the multi-tracer bispectrum [93], and cross-correlation with a matter tracer (e.g. via CMB lensing [10]) are potential directions in which to extend the methods presented here.

Acknowledgments

A $\quad F_{f^{\rm loc}_{NL} f^{\rm loc}_{NL}}$ expressions

We first reproduce the expressions for the 2-tracer case, where $\mathbf{C}^{(2)} = \frac{1}{V}\begin{pmatrix} \alpha^2 P_2 + \frac{1}{\bar{n}_1} & r\alpha P_2 \\ & P_2 + \frac{1}{\bar{n}_2} \end{pmatrix}$, and add $F_{\alpha P_2}$ to the expressions for $F_{\alpha\alpha}$ and $F_{P_2 P_2}$ of Ref. [15][7].

$$F_{\alpha\alpha} = \frac{\alpha^2 r^4 + r^2(X_2 + 1)\left(X_1 - 3\alpha^2\right) + 2\alpha^2(X_2 + 1)^2}{\left(\alpha^2\left(-r^2 + X_2 + 1\right) + X_1(X_2 + 1)\right)^2}$$

$$F_{P_2 P_2} = \frac{2\alpha^2 X_1\left(r^2(X_2 - 1) + 1\right) + \alpha^4\left(2r^4 - 2r^2(X_2 + 2) + X_2(X_2 + 2) + 2\right) + X_1^2}{2\left(\alpha^2 P_2\left(-r^2 + X_2 + 1\right) + P_2 X_1(X_2 + 1)\right)^2}$$

$$F_{\alpha P_2} = \frac{\alpha r^2 X_1 X_2 + \alpha^3\left(-r^2 + X_2 + 1\right)^2}{P_2\left(\alpha^2\left(-r^2 + X_2 + 1\right) + X_1(X_2 + 1)\right)^2},$$

where $X_i \equiv \frac{1}{\bar{n}_i P_2}$ quantifies the relative signal-to-noise.

In this work we used explicit symbolic expressions for the 3-tracer case, extending the expressions of Ref. [15]. The Fisher matrix element for $f^{\rm loc}_{NL}$ is given by the simple transformation:

$$F_{f^{\rm loc}_{NL} f^{\rm loc}_{NL}} = \sum_{\lambda,\lambda'} \frac{\partial\lambda}{\partial f^{\rm loc}_{NL}} F_{\lambda\lambda'} \frac{\partial\lambda'}{\partial f^{\rm loc}_{NL}} \tag{A.1}$$

where, in this work, $\lambda, \lambda' \in \{\alpha, \beta, P_2\}$. This reduces to the two-tracer expressions when β is neglected.

For the 3-tracer expressions, to make contact with the notation of Ref. [15], we can write eqn. 4.2 as

$$\mathbf{C}^{(3)} = \frac{1}{V}\begin{pmatrix} \alpha^2 P_2 + \frac{1}{\bar{n}_1} & r_{12}\alpha P_2 & r_{13}\alpha\beta P_2 \\ & P_2 + \frac{1}{\bar{n}_2} & r_{23}\beta P_2 \\ & & \beta^2 P_2 + \frac{1}{\bar{n}_3} \end{pmatrix} \tag{A.2}$$

The 3-tracer Fisher matrix element expressions for the multi-tracer parameters α, β, P_2 are then given by:

$$\begin{aligned}
N_{\alpha\alpha} =\; & \alpha^2 r_{12}^4(\beta^2 + X_3)^2 - 4\alpha^2\beta^2 r_{12}^3 r_{13} r_{23}(\beta^2 + X_3) \\
& + r_{12}^2(\alpha^2(\beta^4(2r_{13}^2(2r_{23}^2 + X_2 + 1) + 3(r_{23}^2 - X_2 - 1)) + \beta^2 X_3(2r_{13}^2(X_2 + 1) + 3r_{23}^2 - 6(X_2 + 1)) \\
& - 3(X_2 + 1)X_3^2) + X_1(\beta^2 + X_3)(\beta^2(-r_{23}^2 + X_2 + 1) + (X_2 + 1)X_3)) \\
& - 2\beta^2 r_{12} r_{13} r_{23}(\alpha^2(\beta^2(2r_{13}^2(X_2 + 1) + 3(r_{23}^2 - X_2 - 1)) - 3(X_2 + 1)X_3) + \beta^2 X_1(-r_{23}^2 + X_2 + 1) \\
& + X_1(X_2 + 1)X_3) + \beta^2(X_2 + 1)X_3(r_{13}^2(X_2 + 1)(X_1 - 3\alpha^2) + 4\alpha^2(-r_{23}^2 + X_2 + 1)) \\
& + \beta^4(\alpha^2 r_{13}^4(X_2 + 1)^2 - r_{13}^2(X_2 + 1)(r_{23}^2 - X_2 - 1)(X_1 - 3\alpha^2) \\
& + 2\alpha^2(-r_{23}^2 + X_2 + 1)^2) + 2\alpha^2(X_2 + 1)^2 X_3^2 \\
D_{\alpha\alpha} =\; & (\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12}r_{13}r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1)) \\
& + \beta^2 X_1(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2 \\
F_{\alpha\alpha} =\; & \frac{N_{\alpha\alpha}}{D_{\alpha\alpha}}
\end{aligned}$$

---

[7]We find a slight difference here in the expression for $F_{P_2 P_2}$ with respect to Ref. [15], though this does not affect any of the conclusions of Ref. [15] since they are all stated in the sample-variance limit, where this difference disappears.

$$N_{\alpha\beta} = \alpha\beta(-r_{12}r_{23} + r_{13}X_2 + r_{13})(\beta^2 r_{13}(\alpha^2(r_{12}^2 - 2r_{12}r_{13}r_{23}$$
$$+ (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1) + X_1(r_{23}^2 - X_2 - 1)) + \alpha^2 r_{13}X_3(r_{12}^2 - X_2 - 1)$$
$$+ X_1 X_3(-2r_{12}r_{23} + r_{13}X_2 + r_{13}))$$
$$D_{\alpha\beta} = (\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12}r_{13}r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1))$$
$$+ \beta^2 X_1(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2$$
$$F_{\alpha\beta} = \frac{N_{\alpha\beta}}{D_{\alpha\beta}}$$

$$N_{P_2\alpha} = \alpha(\alpha^2 r_{12}^4(\beta^2 + X_3)^2 - 4\alpha^2\beta^2 r_{12}^3 r_{13}r_{23}(\beta^2 + X_3) + r_{12}^2(2\alpha^2(\beta^4(r_{13}^2(2r_{23}^2 + X_2 + 1) + r_{23}^2 - X_2 - 1)$$
$$+ \beta^2 X_3((r_{13}^2 - 2)X_2 + r_{13}^2 + r_{23}^2 - 2) - (X_2 + 1)X_3^2) + X_1(\beta^2 r_{23}^2 X_3 + X_2(\beta^2 + X_3)^2))$$
$$- 2\beta^2 r_{12}r_{13}r_{23}(2\alpha^2(\beta^2((r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1) - (X_2 + 1)X_3) + X_1(\beta^2 X_2 + 2X_2 X_3 + X_3))$$
$$+ \beta^4(r_{13}^2 r_{23}^2 X_1 X_2 + \alpha^2(r_{13}^2(X_2 + 1) + r_{23}^2 - X_2 - 1)^2) + \beta^2(X_2 + 1)X_3(r_{13}^2(X_2 + 1)(X_1 - 2\alpha^2)$$
$$+ 2\alpha^2(-r_{23}^2 + X_2 + 1)) + \alpha^2(X_2 + 1)^2 X_3^2)$$
$$D_{P_2\alpha} = P_2(\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12}r_{13}r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1))$$
$$+ \beta^2 X_1(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2$$
$$F_{P_2\alpha} = \frac{N_{P_2\alpha}}{D_{P_2\alpha}}$$

$$N_{P_2P_2} = 2\alpha^2 X_1(\beta^4(2(r_{23}^2 - 1)(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1)$$
$$+ 2(r_{13}^2 - 1)(r_{23}^2 - 1)X_2 - (r_{13}^2 - 1)X_2^2) + \beta^2 X_3(2r_{12}^2(r_{23}^2 + X_2 - 1)$$
$$- 6r_{12}r_{13}r_{23}X_2 + r_{13}^2 X_2(X_2 + 2) + 2r_{23}^2(X_2 - 1) + 2) + X_3^2(r_{12}^2(X_2 - 1) + 1))$$
$$+ \alpha^4(\beta^4(4(r_{13}^2 - 1)X_2(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1) + 3(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1)^2$$
$$+ 2(r_{13}^2 - 1)^2 X_2^2) + 2\beta^2 X_3(2(r_{12}^2 - 1)(r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2 + r_{23}^2 - 1) + 2(r_{12}^2 - 1)(r_{13}^2 - 1)X_2$$
$$- (r_{13}^2 - 1)X_2^2) + X_3^2(2r_{12}^4 - 2r_{12}^2(X_2 + 2) + X_2(X_2 + 2) + 2))$$
$$+ X_1^2(2\beta^2 X_3(r_{23}^2(X_2 - 1) + 1) + \beta^4(2r_{23}^4 - 2r_{23}^2(X_2 + 2) + X_2(X_2 + 2) + 2) + X_3^2)$$
$$D_{P_2P_2} = 2P_2^2(\alpha^2(X_3(-r_{12}^2 + X_2 + 1) - \beta^2(r_{12}^2 - 2r_{12}r_{13}r_{23} + (r_{13}^2 - 1)X_2 + r_{13}^2 + r_{23}^2 - 1)) + \beta^2 X_1$$
$$(-r_{23}^2 + X_2 + 1) + X_1(X_2 + 1)X_3)^2$$
$$F_{P_2P_2} = \frac{N_{P_2P_2}}{D_{P_2P_2}}$$

We do not write the expressions for $F_{\beta P_2}$ or $F_{\beta\beta}$ since these are the same as $F_{\alpha P_2}, F_{\alpha\alpha}$ with interchanged arguments $\beta \longleftrightarrow \alpha$.

The Jacobian factors for the transformation in eqn. A.1 for the parameters α and β are:

$$\frac{\partial\alpha}{\partial f_{NL}^{\rm loc}} = \alpha\left(\frac{b_{\phi,1}}{b_1} - \frac{b_{\phi,2}}{b_2}\right)\mathcal{M}^{-1} \tag{A.3}$$

$$\frac{\partial P_2}{\partial f_{NL}^{\rm loc}} = 2b_{\phi,2}\mathcal{M}^{-1}\frac{P_2}{b_2} \tag{A.4}$$

where we dropped the k-dependence of $\mathcal{M}$ and P_2 to be consistent with the previous expressions, and again the β expression is analogous to the α expression.

In the sample variance limit, at leading order in X_i, and assuming $b_1 = b_2 = b_3$, the ratio of the $f_{NL}^{\rm loc}$ Fisher information is:

$$\frac{F_{f_{NL}^{\rm loc}}^{(3)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2)}} = \frac{(b_{\phi,2} - b_{\phi,3})^2}{(b_{\phi,1} - b_{\phi,2})^2} + \frac{2X_2 \left(b_{\phi,1}^2 - 9b_{\phi,1}b_{\phi,2} + 7b_{\phi,1}b_{\phi,3} + 7b_{\phi,2}^2 - 5b_{\phi,2}b_{\phi,3} - b_{\phi,3}^2\right)}{3(b_{\phi,1} - b_{\phi,2})^2}$$

(A.5)

$$\begin{aligned}
\frac{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(3,P)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2,P)}} = \frac{F_{f_{NL}^{\rm loc}}^{(3)}}{F_{f_{NL}^{\rm loc}}^{(2)}} &- \frac{4b_{\phi,2}X_2}{3(b_{\phi,1} - b_{\phi,2})^4}[b_{\phi,1}^2(17b_{\phi,2} - 21b_{\phi,3}) \\
&+ 4b_{\phi,2}^2 \left(b_{\phi,2} - 2b_{\phi,3}\right) \\
&+ b_{\phi,1} \left(-8b_{\phi,2}^2 + 3b_{\phi,2}b_{\phi,3} + 13b_{\phi,3}^2\right)]
\end{aligned}$$

(A.6)

when considering only relative amplitude parameters, and when including P_2, respectively. These expressions illustrate that these contributions both approach $\frac{\Delta b_{\phi,23}}{\Delta b_{\phi,12}}$ as $X_2 \to 0$, but can be significantly different otherwise.

Similarly, when considering the ratio of the Fisher information when using P_2 as a parameter or not for two tracers, we have:

$$\frac{F_{f_{NL}^{\rm loc}}^{(2,P)}}{F_{f_{NL}^{\rm loc} f_{NL}^{\rm loc}}^{(2)}} = 1 + X_2 \frac{52b_{\phi,1}b_{\phi,2}}{3(b_{\phi,1} - b_{\phi,2})^2}.$$

(A.7)

Again, we see that in the sample variance limit the Fisher information is equivalent, but outside of the case where $X_2 \to 0$ the exact values of b_ϕ in each sample can drive the ratio above or below one (e.g. in the latter case if the values of $b_{\phi,1}, b_{\phi,2}$ have different signs).

We now make a few brief comments about these expressions, which may be useful for the interested reader of the main text. We note that even in the sample variance limit, it is clearly possible to have three tracers be less informative than two tracers - as this can occur when $(b_{\phi,2} - b_{\phi,3})^2 < (b_{\phi,1} - b_{\phi,2})^2$. To be consistent with the notation of Ref. [15] we have taken the two-tracer case to have LPNG biases $b_{\phi,1}$ and $b_{\phi,2}$, so care should be taken to interperet these expressions appropriately (and not to simply associate $b_{\phi,1}$ with b_ϕ^- or $b_{\phi,3}$ with b_ϕ^+). In Section 4.2, for future surveys we considered a symmetric Δp for the upper and lower tertiles, though in general this will not be the case. Eqn. A.6 also indicates that the exact departure from this symmetry may affect whether a 2-tracer of 3-tracer forecast is desirable. We also note that, in general, both the linear bias b of the sample and the number density $\bar{n}$ will spoil these statements as we leave the sample variance limit and b changes over the splits.

References

[1] P. D. Meerburg, D. Green, R. Flauger, B. Wallisch, M. C. D. Marsh, E. Pajer et al., *Primordial Non-Gaussianity*, BAAS **51** (2019) 107 [1903.04409].

[2] J. M. Bardeen, *Gauge-invariant cosmological perturbations*, Phys. Rev. D **22** (1980) 1882.

[3] M. Alvarez, T. Baldauf, J. R. Bond, N. Dalal, R. de Putter, O. Doré et al., *Testing Inflation with Large Scale Structure: Connecting Hopes with Reality*, arXiv e-prints (2014) arXiv:1412.4671 [1412.4671].

[4] A. Achúcarro, M. Biagetti, M. Braglia, G. Cabass, R. Caldwell, E. Castorina et al., *Inflation: Theory and Observations*, arXiv e-prints (2022) arXiv:2203.08128 [2203.08128].

[5] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, A&A 641 (2020) A9 [1905.05697].

[6] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold et al., *CMB-S4 Science Book, First Edition*, arXiv e-prints (2016) arXiv:1610.02743 [1610.02743].

[7] P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali, D. Alonso et al., *The Simons Observatory: science goals and forecasts*, J. Cosmology Astropart. Phys. 2019 (2019) 056 [1808.07445].

[8] R. de Putter and O. Doré, *Designing an inflation galaxy survey: How to measure σ $(f_{NL}) \sim 1$ using scale-dependent galaxy bias*, Phys. Rev. D 95 (2017) 123513 [1412.3854].

[9] N. Sailer, E. Castorina, S. Ferraro and M. White, *Cosmology at high redshift - a probe of fundamental physics*, J. Cosmology Astropart. Phys. 2021 (2021) 049 [2106.09713].

[10] M. Schmittfull and U. Seljak, *Parameter constraints from cross-correlation of CMB lensing with galaxy clustering*, Phys. Rev. D 97 (2018) 123540 [1710.09465].

[11] U. Giri, M. Münchmeyer and K. M. Smith, *Robust Neural Network-Enhanced Estimation of Local Primordial Non-Gaussianity*, arXiv e-prints (2022) arXiv:2205.12964 [2205.12964].

[12] A. Moradinezhad Dizgah, M. Biagetti, E. Sefusatti, V. Desjacques and J. Noreña, *Primordial non-Gaussianity from biased tracers: likelihood analysis of real-space power spectrum and bispectrum*, J. Cosmology Astropart. Phys. 2021 (2021) 015 [2010.14523].

[13] L. R. Abramo, J. V. D. Ferri and I. L. Tashiro, *Fisher matrix for multiple tracers: the information in the cross-spectra*, J. Cosmology Astropart. Phys. 2022 (2022) 013 [2112.01812].

[14] S. Jolicoeur, R. Maartens and S. Dlamini, *Constraining primordial non-Gaussianity by combining next-generation galaxy and 21 cm intensity mapping surveys*, arXiv e-prints (2023) arXiv:2301.02406 [2301.02406].

[15] U. Seljak, *Extracting Primordial Non-Gaussianity without Cosmic Variance*, Phys. Rev. Lett. 102 (2009) 021302 [0807.1770].

[16] P. McDonald and U. Seljak, *How to evade the sample variance limit on measurements of redshift-space distortions*, J. Cosmology Astropart. Phys. 2009 (2009) 007 [0810.0323].

[17] S. Ferraro and K. M. Smith, *Using large scale structure to measure f_{NL} , g_{NL} and τ_{NL}*, Phys. Rev. D 91 (2015) 043506 [1408.3126].

[18] G. M. Bernstein and Y.-C. Cai, *Cosmology without cosmic variance*, MNRAS 416 (2011) 3009 [1104.3862].

[19] O. Darwish, S. Foreman, M. M. Abidi, T. Baldauf, B. D. Sherwin and P. D. Meerburg, *Density reconstruction from biased tracers and its application to primordial non-Gaussianity*, Phys. Rev. D 104 (2021) 123520 [2007.08472].

[20] T. Mergulhão, H. Rubira, R. Voivodic and L. R. Abramo, *The effective field theory of large-scale structure and multi-tracer*, J. Cosmology Astropart. Phys. 2022 (2022) 021 [2108.11363].

[21] Y. Wang, G.-B. Zhao, C. Zhao, O. H. E. Philcox, S. Alam, A. Tamone et al., *The clustering of the SDSS-IV extended baryon oscillation spectroscopic survey DR16 luminous red galaxy and emission-line galaxy samples: cosmic distance and structure growth measurements using multiple tracers in configuration space*, MNRAS 498 (2020) 3470 [2007.09010].

[22] Y. Wang and G.-B. Zhao, *A brief review on cosmological analysis of galaxy surveys with multiple tracers*, Research in Astronomy and Astrophysics **20** (2020) 158 [2009.03862].

[23] N. Dalal, O. Doré, D. Huterer and A. Shirokov, *Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects*, Phys. Rev. D **77** (2008) 123514 [0710.4560].

[24] A. Slosar, C. Hirata, U. Seljak, S. Ho and N. Padmanabhan, *Constraints on local primordial non-Gaussianity from large scale structure*, J. Cosmology Astropart. Phys. **2008** (2008) 031 [0805.3580].

[25] B. Leistedt, H. V. Peiris and N. Roth, *Constraints on Primordial Non-Gaussianity from 800 000 Photometric Quasars*, Phys. Rev. Lett. **113** (2014) 221301 [1405.4315].

[26] E.-M. Mueller, M. Rezaie, W. J. Percival, A. J. Ross, R. Ruggeri, H.-J. Seo et al., *The clustering of galaxies in the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Primordial non-Gaussianity in Fourier Space*, arXiv e-prints (2021) arXiv:2106.13725 [2106.13725].

[27] E. Castorina, N. Hand, U. Seljak, F. Beutler, C.-H. Chuang, C. Zhao et al., *Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space*, J. Cosmology Astropart. Phys. **2019** (2019) 010 [1904.08859].

[28] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonović and M. Zaldarriaga, *Constraints on multifield inflation from the BOSS galaxy survey*, Phys. Rev. D **106** (2022) 043506 [2204.01781].

[29] G. D'Amico, M. Lewandowski, L. Senatore and P. Zhang, *Limits on primordial non-Gaussianities from BOSS galaxy-clustering data*, arXiv e-prints (2022) arXiv:2201.11518 [2201.11518].

[30] A. Barreira, G. Cabass, F. Schmidt, A. Pillepich and D. Nelson, *Galaxy bias and primordial non-Gaussianity: insights from galaxy formation simulations with IllustrisTNG*, J. Cosmology Astropart. Phys. **2020** (2020) 013 [2006.09368].

[31] A. Barreira, *On the impact of galaxy bias uncertainties on primordial non-Gaussianity constraints*, J. Cosmology Astropart. Phys. **2020** (2020) 031 [2009.06622].

[32] A. Barreira, *Predictions for local PNG bias in the galaxy power spectrum and bispectrum and the consequences for f_{NL} constraints*, J. Cosmology Astropart. Phys. **2022** (2022) 033 [2107.06887].

[33] A. Barreira, *Can we actually constrain f_{NL} using the scale-dependent bias effect? An illustration of the impact of galaxy bias uncertainties using the BOSS DR12 galaxy power spectrum*, J. Cosmology Astropart. Phys. **2022** (2022) 013 [2205.05673].

[34] B. A. Reid, L. Verde, K. Dolag, S. Matarrese and L. Moscardini, *Non-Gaussian halo assembly bias*, J. Cosmology Astropart. Phys. **2010** (2010) 013 [1004.1637].

[35] T. Lazeyras, A. Barreira, F. Schmidt and V. Desjacques, *Assembly bias in the local PNG halo bias and its implication for $f_{\rm NL}$ constraints*, arXiv e-prints (2022) arXiv:2209.07251 [2209.07251].

[36] V. Desjacques, D. Jeong and F. Schmidt, *Large-scale galaxy bias*, Phys. Rep. **733** (2018) 1 [1611.09787].

[37] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques and F. Schmidt, *Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity*, MNRAS **468** (2017) 3277 [1611.04901].

[38] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, *Linear response to long wavelength fluctuations using curvature simulations*, J. Cosmology Astropart. Phys. **2016** (2016) 007 [1511.01465].

[39] E. Sirko, *Initial Conditions to Cosmological N-Body Simulations, or, How to Run an Ensemble of Simulations*, ApJ **634** (2005) 728 [`astro-ph/0503106`].

[40] N. Y. Gnedin, A. V. Kravtsov and D. H. Rudd, *Implementing the DC Mode in Cosmological Simulations with Supercomoving Variables*, ApJS **194** (2011) 46 [`1104.1428`].

[41] C. Wagner, F. Schmidt, C. T. Chiang and E. Komatsu, *Separate universe simulations.*, MNRAS **448** (2015) L11 [`1409.6294`].

[42] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, *Galaxy bias and non-linear structure formation in general relativity*, J. Cosmology Astropart. Phys. **2011** (2011) 031 [`1106.5507`].

[43] Y. Li, W. Hu and M. Takada, *Super-sample covariance in simulations*, Phys. Rev. D **89** (2014) 083519 [`1401.0385`].

[44] P. McDonald, *Toward a Measurement of the Cosmological Geometry at z ~2: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, ApJ **585** (2003) 34 [`astro-ph/0108064`].

[45] L. Dai, E. Pajer and F. Schmidt, *On separate universes*, J. Cosmology Astropart. Phys. **2015** (2015) 059 [`1504.00351`].

[46] D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel et al., *The IllustrisTNG simulations: public data release*, Computational Astrophysics and Cosmology **6** (2019) 2 [`1812.05609`].

[47] A. Pillepich, D. Nelson, L. Hernquist, V. Springel, R. Pakmor, P. Torrey et al., *First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies*, MNRAS **475** (2018) 648 [`1707.03406`].

[48] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, MNRAS **475** (2018) 676 [`1707.03397`].

[49] D. Nelson, A. Pillepich, V. Springel, R. Weinberger, L. Hernquist, R. Pakmor et al., *First results from the IllustrisTNG simulations: the galaxy colour bimodality*, MNRAS **475** (2018) 624 [`1707.03395`].

[50] J. P. Naiman, A. Pillepich, V. Springel, E. Ramirez-Ruiz, P. Torrey, M. Vogelsberger et al., *First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium*, MNRAS **477** (2018) 1206 [`1707.03401`].

[51] F. Marinacci, M. Vogelsberger, R. Pakmor, P. Torrey, V. Springel, L. Hernquist et al., *First results from the IllustrisTNG simulations: radio haloes and magnetic fields*, MNRAS **480** (2018) 5113 [`1707.03396`].

[52] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen et al., *The DESI Experiment Part I: Science,Targeting, and Survey Design*, arXiv e-prints (2016) arXiv:1611.00036 [`1611.00036`].

[53] M. Davis, G. Efstathiou, C. S. Frenk and S. D. M. White, *The evolution of large-scale structure in a universe dominated by cold dark matter*, ApJ **292** (1985) 371.

[54] D. Anbajagane, A. E. Evrard and A. Farahi, *Baryonic imprints on DM haloes: population statistics from dwarf galaxies to galaxy clusters*, MNRAS **509** (2022) 3441 [`2109.02713`].

[55] J. F. Navarro, C. S. Frenk and S. D. M. White, *A Universal Density Profile from Hierarchical Clustering*, ApJ **490** (1997) 493 [`astro-ph/9611107`].

[56] B. Hadzhiyska, S. Tacchella, S. Bose and D. J. Eisenstein, *The galaxy-halo connection of emission-line galaxies in IllustrisTNG*, MNRAS **502** (2021) 3599 [`2011.05331`].

[57] S. Yuan, B. Hadzhiyska, S. Bose and D. J. Eisenstein, *Illustrating galaxy-halo connection in the DESI era with ILLUSTRISTNG*, MNRAS **512** (2022) 5793 [`2202.12911`].

[58] B. Hadzhiyska, L. Hernquist, D. Eisenstein, A. M. Delgado, S. Bose, R. Kannan et al., *The MillenniumTNG Project: Refining the one-halo model of red and blue galaxies at different redshifts*, arXiv e-prints (2022) arXiv:2210.10068 [2210.10068].

[59] R. Zhou, B. Dey, J. A. Newman, D. J. Eisenstein, K. Dawson, S. Bailey et al., *Target Selection and Validation of DESI Luminous Red Galaxies*, AJ **165** (2023) 58 [2208.08515].

[60] A. Raichoor, J. Moustakas, J. A. Newman, T. Karim, S. Ahlen, S. Alam et al., *Target Selection and Validation of DESI Emission Line Galaxies*, arXiv e-prints (2022) arXiv:2208.08513 [2208.08513].

[61] R. Zhou, J. A. Newman, Y.-Y. Mao, A. Meisner, J. Moustakas, A. D. Myers et al., *The clustering of DESI-like luminous red galaxies using photometric redshifts*, MNRAS **501** (2021) 3309 [2001.06018].

[62] P. Villanueva-Domingo, F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, F. Marinacci, D. N. Spergel et al., *Inferring Halo Masses with Graph Neural Networks*, ApJ **935** (2022) 30 [2111.08683].

[63] S. Agarwal, R. Davé and B. A. Bassett, *Painting galaxies into dark matter haloes using machine learning*, MNRAS **478** (2018) 3410 [1712.03255].

[64] B. P. Moster, T. Naab, M. Lindström and J. A. O'Leary, *GalaxyNet: connecting galaxies and dark matter haloes with deep neural networks and reinforcement learning in large volumes*, MNRAS **507** (2021) 2115 [2005.12276].

[65] X. Xu, S. Kumar, I. Zehavi and S. Contreras, *Predicting halo occupation and galaxy assembly bias with machine learning*, MNRAS **507** (2021) 4879 [2107.01223].

[66] A. M. Delgado, D. Wadekar, B. Hadzhiyska, S. Bose, L. Hernquist and S. Ho, *Modelling the galaxy-halo connection with machine learning*, MNRAS **515** (2022) 2733 [2111.02422].

[67] N. S. M. de Santi, N. V. N. Rodrigues, A. D. Montero-Dorta, L. R. Abramo, B. Tucci and M. C. Artale, *Mimicking the halo-galaxy connection using machine learning*, MNRAS **514** (2022) 2463 [2201.06054].

[68] R. Stiskalek, D. J. Bartlett, H. Desmond and D. Anbajagane, *The scatter in the galaxy-halo connection: a machine learning analysis*, MNRAS **514** (2022) 4026 [2202.14006].

[69] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, arXiv e-prints (2016) arXiv:1603.02754 [1603.02754].

[70] S. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, arXiv e-prints (2017) arXiv:1705.07874 [1705.07874].

[71] S. Ferraro and M. J. Wilson, *Inflation and Dark Energy from spectroscopy at z > 2*, BAAS **51** (2019) 72 [1903.09208].

[72] O. Doré, J. Bock, M. Ashby, P. Capak, A. Cooray, R. de Putter et al., *Cosmology with the SPHEREX All-Sky Spectral Survey*, arXiv e-prints (2014) arXiv:1412.4872 [1412.4872].

[73] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont et al., *Planck 2015 results. XIII. Cosmological parameters*, A&A **594** (2016) A13 [1502.01589].

[74] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, MNRAS **475** (2018) 676 [1707.03397].

[75] N. Hamaus, U. Seljak and V. Desjacques, *Optimal constraints on local primordial non-Gaussianity from the two-point statistics of large-scale structure*, Phys. Rev. D **84** (2011) 083509 [1104.2321].

[76] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant*, Nature **281** (1979) 358.

[77] R. Zhou, B. Dey, J. A. Newman, D. J. Eisenstein, K. Dawson, S. Bailey et al., *Target Selection and Validation of DESI Luminous Red Galaxies*, AJ **165** (2023) 58 [2208.08515].

[78] N. Mostek, A. L. Coil, M. Cooper, M. Davis, J. A. Newman and B. J. Weiner, *The DEEP2 Galaxy Redshift Survey: Clustering Dependence on Galaxy Stellar Mass and Star Formation Rate at z ~1*, ApJ **767** (2013) 89 [1210.6694].

[79] G. Cabass, M. M. Ivanov, O. H. E. Philcox, M. Simonovic and M. Zaldarriaga, *Constraining Single-Field Inflation with MegaMapper*, arXiv e-prints (2022) arXiv:2211.14899 [2211.14899].

[80] M. J. Wilson and M. White, *Cosmology with dropout selection: straw-man surveys & CMB lensing*, J. Cosmology Astropart. Phys. **2019** (2019) 015 [1904.13378].

[81] S. Bhattacharya, K. Heitmann, M. White, Z. Lukić, C. Wagner and S. Habib, *Mass Function Predictions Beyond ΛCDM*, ApJ **732** (2011) 122 [1005.2239].

[82] B. Diemer, *COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos*, ApJS **239** (2018) 35 [1712.04512].

[83] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes et al., *Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality*, ApJ **688** (2008) 709 [0803.2706].

[84] J. L. Tinker, B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes et al., *The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests*, ApJ **724** (2010) 878 [1001.3162].

[85] M. Tegmark, A. N. Taylor and A. F. Heavens, *Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?*, ApJ **480** (1997) 22 [astro-ph/9603021].

[86] T. Baldauf, U. Seljak, R. E. Smith, N. Hamaus and V. Desjacques, *Halo stochasticity from exclusion and nonlinear clustering*, Phys. Rev. D **88** (2013) 083507 [1305.2917].

[87] D. Ginzburg and V. Desjacques, *Shot noise in multitracer constraints on f_{NL} and relativistic projections: Power spectrum*, MNRAS **495** (2020) 932 [1911.11701].

[88] A. Slosar, *Optimal weighting in f_{NL} constraints from large scale structure in an idealised case*, J. Cosmology Astropart. Phys. **2009** (2009) 004 [0808.0044].

[89] E. Castorina, Y. Feng, U. Seljak and F. Villaescusa-Navarro, *Primordial Non-Gaussianities and Zero-Bias Tracers of the Large-Scale Structure*, Phys. Rev. Lett. **121** (2018) 101301 [1803.11539].

[90] R. Pakmor, V. Springel, J. P. Coles, T. Guillet, C. Pfrommer, S. Bose et al., *The MillenniumTNG Project: The hydrodynamical full physics simulation and a first look at its galaxy clusters*, arXiv e-prints (2022) arXiv:2210.10060 [2210.10060].

[91] D. J. Eisenstein, H.-J. Seo, E. Sirko and D. N. Spergel, *Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak*, ApJ **664** (2007) 675 [astro-ph/0604362].

[92] A. Barreira and E. Krause, *Towards optimal and robust f_{NL} constraints with multi-tracer analyses*, arXiv e-prints (2023) arXiv:2302.09066 [2302.09066].

[93] D. Yamauchi, S. Yokoyama and K. Takahashi, *Multitracer technique for galaxy bispectrum: An application to constraints on nonlocal primordial non-Gaussianities*, Phys. Rev. D **95** (2017) 063530 [1611.03590].