


This is the accepted manuscript made available via CHORUS. The article has been published as:

Robust neural network-enhanced estimation of local primordial non-Gaussianity

Utkarsh Giri, Moritz Münchmeyer, and Kendrick M. Smith

Phys. Rev. D **107**, L061301 — Published 20 March 2023

DOI: 10.1103/PhysRevD.107.L061301

Robust Neural Network-Enhanced Estimation of Local Primordial Non-Gaussianity

Utkarsh Giri* and Moritz Münchmeyer
University of Wisconsin-Madison, Madison, Wisconsin, USA

Kendrick M. Smith
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, CA

(Dated: December 12, 2022)

When applied to the non-linear matter distribution of the universe, neural networks have been shown to be very statistically sensitive probes of cosmological parameters, such as the linear perturbation amplitude σ_8. However, when used as a "black box", neural networks are not robust to baryonic uncertainty. We propose a robust architecture for constraining primordial non-Gaussianity f_{NL}, by training a neural network to locally estimate σ_8, and correlating these local estimates with the large-scale density field. We apply our method to N-body simulations, and show that $\sigma(f_{NL})$ is 3.5 times better than the constraint obtained from a standard halo-based approach. We show that our method has the same robustness property as large-scale halo bias: baryonic physics can change the normalization of the estimated f_{NL}, but cannot change whether f_{NL} is detected.

Keywords: non-Gaussianity, neural networks, large-scale structure

I. INTRODUCTION

Observations of cosmological perturbations on quasi-linear scales, with the cosmic microwave background (CMB) and large scale structure (LSS), have resulted in precise measurements of fundamental cosmological parameters. However, quasilinear scales contain only a small fraction of the theoretically accessible information. On smaller scales, traditional N-point correlation function analysis becomes sub-optimal, and inference methods such as forward modelling [1, 2] or machine learning [3–7] can provide stronger constraints.

Here, we will focus on inference of the important primordial physics parameter f_{NL}, which arises in multi-field models of inflation [8–11]. In such models, the primordial Bardeen potential $\Phi(\mathbf{x})$ can be parameterized as:

$$\Phi(\mathbf{x}) = \Phi_G(\mathbf{x}) + f_{NL}(\Phi_G(\mathbf{x})^2 - \langle\Phi_G^2\rangle) \tag{1}$$

where Φ_G is a Gaussian field and f_{NL} quantifies the level of non-Gaussianity. Constraining $f_{NL} < 1$ is a major goal of upcoming galaxy surveys [12].

Several statistical approaches for estimating f_{NL} in LSS have been proposed, including the squeezed bispectrum [13] and the scale-dependent power spectrum approach [14, 15] together with the idea of sample variance cancellation using a variety of probes [16–19].

In this paper, we will propose a neural network (NN) based approach to estimating f_{NL}. We will show that our NN-based analysis obtains significantly smaller error bars than an analysis based on large-scale matter and halo fields. This is perhaps unsuprising since a neural network with enough capacity, trained on enough simulations, should give *statistically* optimal parameter constraints.

However, *robustness* is a central challenge for neural networks. As is widely appreciated, simulations on small scales suffer from baryonic feedback uncertainties and are not reliable at the sub-percent level accuracy required to tighten parameter bounds. For example, an NN trained to measure σ_8 on one set of simulations will likely give incorrect results on different simulations or on real data [20, 21].

This problem is not unique to NN-based methods. In particular, the average dark matter halo density $\bar{n}_h$ is statistically a precise probe of σ_8, but it is not robust because $\bar{n}_h$ is sensitive to uncertain local physics. However, *anisotropy* in the halo field can be used to place robust constraints on f_{NL}. A famous result [14] states that if $f_{NL} \neq 0$, the halo bias $b_h(k)$ contains a characteristic $1/k^2$ term, which cannot be induced by local physics.

The idea of this paper is to replace the halo field by an NN-based local estimate $\pi(\mathbf{x})$ of σ_8. As in the halo case, $\bar{\pi}$ is not robust as an absolute measurement of σ_8, but the bias $b_\pi(k)$ contains a term proportional to f_{NL}/k^2, which can be used to place robust constraints on f_{NL}. We will show that this approach combines the statistical power of neural networks with the robustness of the traditional halo-based analysis.

II. FORMALISM

In an f_{NL} cosmology, the halo bias $b_h(k)$ contains a term proportional to f_{NL}/k^2 [14]. In the next few paragraphs, we review the derivation, in a language which will generalize to NN-based observables.

If $f_{NL} \neq 0$, the locally observed amplitude of short-wavelength modes is a function $\sigma_8^{\rm loc}(\mathbf{x})$ of position. On large scales $\mathbf{k}_L \to 0$, anisotropy in $\sigma_8^{\rm loc}$ is related to the potential Φ by:

$$\frac{\sigma_8^{\rm loc}(\mathbf{k}_L)}{\bar{\sigma}_8} = 2 f_{NL}\, \Phi(\mathbf{k}_L) \tag{2}$$

* Correspondence email address: ugiri@wisc.edu

The coupling between large and small scales described by Eq. (2) arises from the term $f_{NL}\Phi_G^2$ in Eq. (1), which mixes scales. For a formal derivation of (2), see [11, 22].

On large scales, the local halo abundance δ_h is sensitive to both $\sigma_8^{\rm loc}$ and the matter overdensity δ_m:

$$\delta_h(\mathbf{x}) = b_h^G \delta_m(\mathbf{x}) + \frac{1}{2} b_h^{NG} \log\left(\frac{\sigma_8^{\rm loc}(\mathbf{x})}{\bar{\sigma}_8}\right) + (\text{noise}) \quad (3)$$

The first term $b_h^G\delta_m$ is the usual (Gaussian) halo bias, and the second term $b_h^{NG}\log(\sigma_8^{\rm loc})$ is a non-Gaussian bias term. The coefficient b_h^{NG} is the derivative of the halo density $\bar{n}_h$ with respect to the cosmological parameter σ_8 [22–25]:

$$b_h^{NG} = 2\frac{\partial \log \bar{n}_h}{\partial \log \sigma_8} \quad (4)$$

(Sometimes the approximation $b_h^{NG} \approx 2\delta_c(b_h^G - 1)$ is used, but we will not use it in this paper.)

On linear scales, the density field δ_m and potential Φ are related via the Fourier-space Poisson equation $\delta_m(\mathbf{k}) = \alpha(\mathbf{k}, z)\Phi(\mathbf{k})$ [26]. Here, $\alpha(k, z)$ is given by

$$\alpha(k, z) \equiv \frac{2k^2 T(k) D(z)}{3\Omega_m H_0^2} \quad (5)$$

Combining Eqs. (2)–(5), we derive the NG halo bias:

$$\delta_h(\mathbf{k}_L) = b_h(k_L)\delta_m(\mathbf{k}_L) + (\text{Poisson noise}) \quad (6)$$

where:

$$b_h(k) = b_h^G + b_h^{NG}\frac{f_{NL}}{\alpha(k, z)} \quad (7)$$

In this paper, we generalize the preceding results as follows. First we note that they do not depend on any specific properties of halos, other than the local halo abundance $n_h(\mathbf{x})$ being sensitive to $\sigma_8^{\rm loc}(\mathbf{x})$. Any field $\pi(\mathbf{x})$ which is derived from the nonlinear density field, in a reasonably local way, should have the same property. We propose constructing a field $\pi(\mathbf{x})$ using a neural network trained to maximize sensitivity to $\sigma_8^{\rm loc}$.

We assume that our input data consists of the nonlinear density field $\delta_m(\mathbf{x})$ as a 3-d pixelized map at $\sim$2 Mpc resolution without noise. Thus, our constraints on f_{NL} should be interpreted as information content in principle, given complete information at a specified resolution. In future work, we plan to apply our approach to simulated galaxy catalogs, which are more representative of real data.

We define a field $\pi(\mathbf{x})$ by applying a CNN to $\delta_m(\mathbf{x})$ (Figure 1). The CNN has a small receptive field ($\sim$18 h^{-1} Mpc), so that the output field $\pi(\mathbf{x})$ is fairly local in the input field $\delta_m(\mathbf{x})$. As described in §III, we train the CNN so that $\pi(\mathbf{x})$ is an estimate of $\sigma_8^{\rm loc}(\mathbf{x})$ with low statistical noise.

Following the logic above for halos, we make the following predictions for the behavior of $\pi(\mathbf{x})$ on large scales. First, we predict that the matter-π and π-π power spectra are given by:

$$P_{m\pi}(k) = b_\pi(k) P_{mm}(k) \quad (8)$$

$$P_{\pi\pi}(k) = b_\pi(k)^2 P_{mm}(k) + N_{\pi\pi} \quad (9)$$

where the linear bias $b_\pi(k)$ is the sum of Gaussian (constant) and non-Gaussian terms:

$$b_\pi(k) = b_\pi^G + b_\pi^{NG}\frac{f_{NL}}{\alpha(k, z)} \quad (10)$$

We also predict that the non-Gaussian bias b_π^{NG} is related to the σ_8 dependence of the mean π-field:

$$b_\pi^{NG} = 2\frac{\partial \bar{\pi}}{\partial \log \sigma_8} \quad (11)$$

Finally, we predict that the noise $N_{\pi\pi}$ defined in (9) is constant in k. In §IV, we will verify these predictions and show that they lead to strong constraints on f_{NL}.

III. NEURAL NETWORK

Architecture: Our neural network uses a fully convolutional, sliding-window architecture with a total of 16433 parameters (Figure 1). The network takes the 3D matter density field $\delta_m(\mathbf{x})$ from an N-body simulation, and produces an output field $\pi(\mathbf{x})$ with the same resolution as the input. We use small convolution kernels, including several layers with $(1 \times 1 \times 1)$ kernels, so that the total receptive field of the network will be small (18 h^{-1} Mpc). The size of the receptive field limits the scales which the neural network can use for estimating π and thus enforces locality. We leave systematic exploration of neural network architecture to future work.

Simulations: We want to train the network so that its output field $\pi(\mathbf{x})$ is an optimal estimate of σ_8. To do this, we need a training set of N-body simulations with multiple values of σ_8. We use the `s8_p` and `s8_m` datasets from the `Quijote` simulations [27], with $\sigma_8 = 0.849$ and 0.819 respectively. The remaining cosmological parameters are $\Omega_{\rm m} = 0.3175$, $\Omega_{\rm b} = 0.049$, $h = 0.6711$, $n_s = 0.9624$, and $w = -1$. Each dataset contains 400 collisionless simulations. Each simulation has 512^3 particles and volume $(1\ h^{-1}{\rm Gpc})^3$. For each simulation, we inpaint particles from the $z = 0$ snapshot on a 512^3 3D mesh using the Cloud-in-Cell algorithm implemented in `nbodykit`. This produces a voxelized 3D matter density field $\delta_m(\mathbf{x})$, which we save to disk for neural network training. We require only two values of σ_8 in the training data rather than a continuum, because the variance of $\pi(\mathbf{x})$ per receptive field is much larger than the difference between the two σ_8 values.

Loss function and optimizer: For each simulation, let $\pi(\mathbf{x})$ be the CNN output, and let $\sigma_8^{\rm true}$ be the value of σ_8

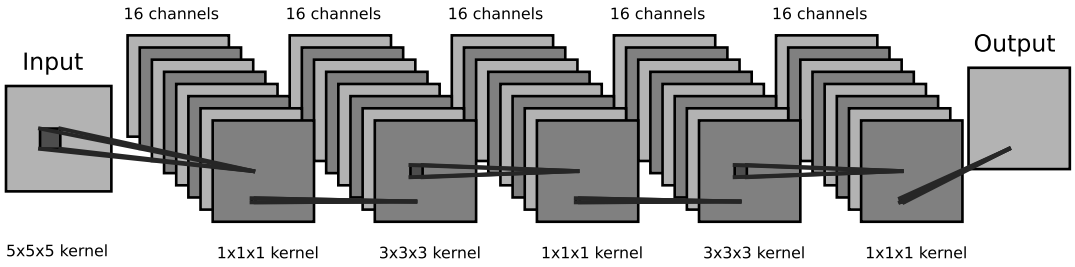


Figure 1. A 2D schematic representation of our CNN architecture. The input is the nonlinear 3-d density field $\delta_m(\mathbf{x})$ from an N-body simulation, and we train the network so that the output 3-d field $\pi(\mathbf{x})$ is an estimate of σ_8. The total receptive field size is $(9 \times 9 \times 9)$ voxels, equivalent to $(18\ h^{-1}\mathrm{Mpc})^3$. Each convolution except the last is followed by a ReLU activation function. Each grey square represents a logical array of size $(512 \times 512 \times 512)$ with periodic boundary conditions. However, as an implementation detail to reduce GPU memory usage, we divide the simulation volume into slightly overlapping subvolumes which can be processed independently.

in the simulation. We define the loss function:

$$\mathcal{L} = \left[\left(\frac{1}{N_{\rm voxels}} \sum_{\rm voxels\ \mathbf{x}} \pi(\mathbf{x})\right) - \sigma_8^{\rm true}\right]^2 \qquad (12)$$

Intuitively, minimizing $\mathcal{L}$ should produce an output field $\pi(\mathbf{x})$ which is an optimal estimate of σ_8, by minimizing the difference between $\sigma_8^{\rm true}$ and the spatially averaged π-field.

We use the Adam optimizer [28] with a learning rate of (5×10^{-5}) to minimize the loss function. The learning rate is reduced by a factor of 0.7 whenever the loss fails to register any improvement for 5 successive epochs of training. The architecture is implemented in `PyTorch` [29] and uses `PyTorch-lightning` [30] for high level interfacing with mixed precision training [31].

We find that the overall normalization W and additive bias b of the NN are slow to converge, so as a final training step, we fix all parameters except (W, b), and minimize the loss (12). This minimization can be done exactly in a single epoch, since the loss is a quadratic function of (W, b).

Validation: Our NN has been trained so that the spatially averaged π field $\bar{\pi} = V_{\rm box}^{-1} \int_{\mathbf{x}} \pi(\mathbf{x})$ is an estimate of σ_8 with lowest possible noise. In the top panel of Figure 2, we verify this statement, by evaluating $\bar{\pi}$ on a test set of 100+100 simulations with $\sigma_8 \in \{0.819, 0.849\}$. We see that the network recovers the correct value of σ_8, and that the NN obtains better statistical separation between σ_8 values than counting halos (bottom panel).

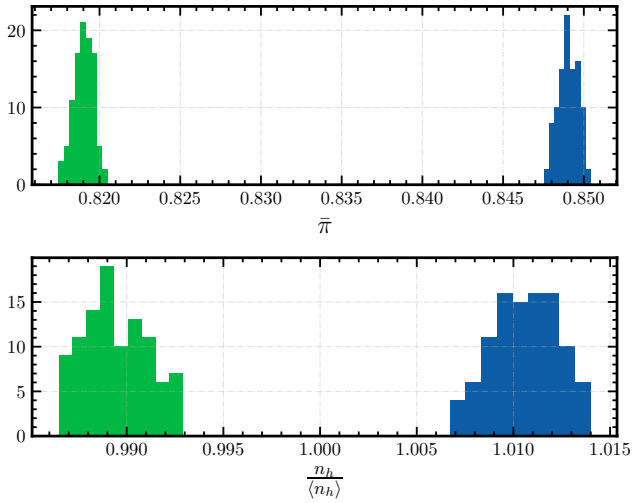


Figure 2. Estimating σ_8 using the neural network from Figure 1. *Top panel.* Histogrammed NN estimates $\bar{\pi} = V_{\rm box}^{-1} \int_{\mathbf{x}} \pi(\mathbf{x})$ on a test set of simulations with $\sigma_8 = 0.819$ (green) and $\sigma_8 = 0.849$ (blue). *Bottom panel.* Histogrammed halo counts from the same test set, showing worse statistical separation between σ_8 values than the NN.

IV. ESTIMATING f_{NL}

In this section, we apply our neural network to simulations with $f_{NL} \neq 0$. We use a test set of 10 N-body simulations with $\sigma_8 = 0.834$, $f_{NL} = 250$, and non-Gaussian initial conditions generated using the Zeldovich approximation.

Power spectra: We next verify the predictions in Eqs. (8)–(11) for the large-scale power spectra $P_{m\pi}$, $P_{\pi\pi}$ in an f_{NL} cosmology. First, we note that since $\bar{\pi} = \sigma_8$ (Fig. 2), our prediction (11) for the non-Gaussian bias b_π^{NG} is:

$$b_\pi^{NG} = 2\frac{\partial \bar{\pi}}{\partial \log \sigma_8} = 2\sigma_8 \qquad (13)$$

and so our prediction (10) for the total bias $b_\pi(k)$ is:

$$b_\pi(k) = b_\pi^G + 2\sigma_8 \frac{f_{NL}}{\alpha(k,z)} \qquad (14)$$

In the the top panel of Fig. 3, we compare the bias model (14) to the empirical bias obtained from cross-correlating π and δ_m in k-bins, and find good agreement. In the bottom panel of Fig. 3, we verify the prediction that the noise power spectrum N_π defined in Eq. (9) is constant in k, by plotting the power spectrum of the residual field $\epsilon(\mathbf{k}) = \pi(\mathbf{k}) - b_\pi(k)\delta_m(\mathbf{k})$.

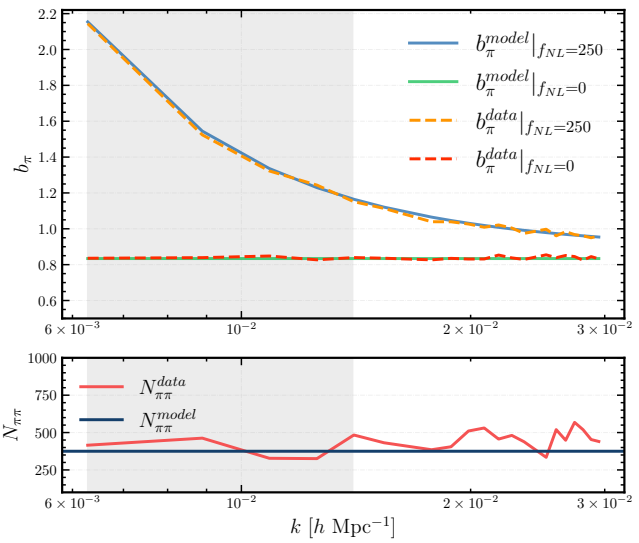


Figure 3. *Top panel.* Bias model (14) for the neural network output field $\pi(\mathbf{x})$, compared to the empirical bias $P_{m\pi}(k)/P_{mm}(k)$ from simulation, for $f_{NL} \in \{0, 250\}$. *Bottom panel.* Power spectrum of the residual field $\epsilon(\mathbf{k}) = \pi(\mathbf{k}) - b_\pi(k)\delta_m(\mathbf{k})$ compared to a best-fit constant N_π. Throughout this figure, best-fit model parameters (b_π^G, N_π) are obtained from the MCMC pipeline described in §IV, with $k_{\rm max} = 0.014$ $h\,{\rm Mpc}^{-1}$ (shown as the shaded region).

We emphasize that simulations with $f_{NL} \neq 0$ were never seen during the training process. The predictions in Eqs. (8)–(11) for power spectra in an f_{NL} cosmology (in particular the prediction $b_\pi^{NG} = 2\sigma_8$) are based entirely on the NN response to varying σ_8, and general considerations of locality. Therefore, the verification of these predictions is a strong test of our formalism.

MCMC pipeline: Now that our model for the power spectra $P_{m\pi}$, $P_{\pi\pi}$ has been verified, we develop an MCMC pipeline that combines large-scale modes of $\pi(\mathbf{k})$ and $\delta_m(\mathbf{k})$ in a joint analysis.

The Gaussian likelihood for our data vector $\mathcal{D} = [\delta_m, \pi]$ given model parameters $\Theta = (f_{NL}, b_\pi^G, N_{\pi\pi})$ is:

$$\mathcal{L}(\Theta|\mathcal{D}) \propto \prod_k \frac{1}{\sqrt{{\rm Det}\, C(k)}} \exp\left(-\frac{\mathcal{D}(\mathbf{k})^\dagger C(k)^{-1} \mathcal{D}(\mathbf{k})}{2V}\right)$$

where the 2×2 covariance matrix $C(k)$ is:

$$C(k) = \begin{bmatrix} P_{mm}(k) & P_{m\pi}(k) \\ P_{m\pi}(k) & P_{\pi\pi}(k) \end{bmatrix} = \begin{bmatrix} P_{mm}(k) & b_\pi(k) P_{mm}(k) \\ b_\pi(k) P_{mm}(k) & b_\pi(k)^2 P_{mm}(k) + N_{\pi\pi} \end{bmatrix} \quad (15)$$

with $b_\pi(k)$ given by Eq. (14). We truncate the likelihood at $k_{\rm max} = 0.014$ $h\,{\rm Mpc}^{-1}$. The posterior is defined using flat priors over a reasonable range for the model parameters Θ. We sample the posterior using affine-invariant sampling implemented in `emcee`[32] to obtain constraints on f_{NL}.

To compare the neural network to a traditional halo based analysis, we also run our MCMC pipeline using the halo field $\delta_h(\mathbf{k})$ instead of the NN-derived field $\pi(\mathbf{k})$. The only change is that we replace $b_\pi^{NG} = 2\sigma_8$ by the non-Gaussian halo bias b_h^{NG}, which we measure in simulations using Eq. (4).

MCMC results: We begin by analysing N-body simulations with Gaussian initial conditions *i.e.* $f_{NL} = 0$. We jointly analyze 100 `fiducial` Quijote simulations by multiplying together their posteriors before sampling. In the right panel of Fig. 4, we show f_{NL} constraints from a joint analysis of the large-scale matter density $\delta_m(\mathbf{k})$ and the NN-derived field $\pi(\mathbf{x})$. In the left panel, we show a similar analysis using $\delta_m(\mathbf{k})$ and the *halo* field $\delta_h(\mathbf{k})$. In both cases, the result is consistent with $f_{NL} = 0$ as expected. However, the neural network gives an f_{NL} error which is 3.5 times better than the halo based analysis!

In the left panel of Figure 4, we used a single halo field consisting of all halos with ≥ 20 particles ($M_{\rm min} = 1.3 \times 10^{13}\ h^{-1}\, M_\odot$). We checked that if narrow halo mass bins are used with optimal weighting, the Fisher forecasted error $\sigma(f_{NL})$ is only 25% better than the single-bin case.

In Fig. 5, we show f_{NL} constraints from a joint MCMC analysis of 10 simulations with $f_{NL} = 250$. We can see that for these non-Gaussian simulations, the correct value of f_{NL} is recovered, and the NN improvement over halos is just as good as in the $f_{NL} = 0$ case.

Robustness: To frame the issue of robustness concretely, imagine that the small-scale astrophysics in the real universe is slightly different from the training set. How will our constraints on f_{NL} be affected?

Since the parameters $(b_\pi^G, b_\pi^{NG}, N_\pi)$ are sensitive to small-scale physics, their values will differ slightly from the training set. For b_π^G and N_π, this is harmless since we marginalize these parameters in our MCMC anyway.

For b_π^{NG}, we note that in the formalism from §II, the parameters b_π^{NG} and f_{NL} only appear in the combination $(b_\pi^{NG} f_{NL})$. Therefore, a small change in b_π^{NG} is equivalent to a change in the normalization of f_{NL} – it cannot "fake" a detection of nonzero f_{NL}. Physically, this is because local physics cannot generate a term in the bias $b_\pi(k)$ proportional to $1/k^2$. This is qualitatively similar to the familar case of halo counts.

Our method does depend on having a rough estimate for b_π^{NG} based on training data. As a check, we estimated the cosmological parameter dependence of b_π^{NG} using the Quijote "latin hypercube" simulations, which vary cosmological parameters over wide ranges. We find that $\bar{\pi}$ is well modelled by a quadratic polynomial in $(\Omega_m, \Omega_b, h, n_s, \sigma_8)$. Using this quadratic model, we find that if cosmological parameters are varied within Planck+BAO 2σ errors [33], the change in $b_\pi^{NG} = 2(\partial\bar{\pi}/\partial \log \sigma_8)$ is $\leq 1\%$. In future work, we hope to extend this analysis to study dependence of b_π^{NG} on subgrid physics.

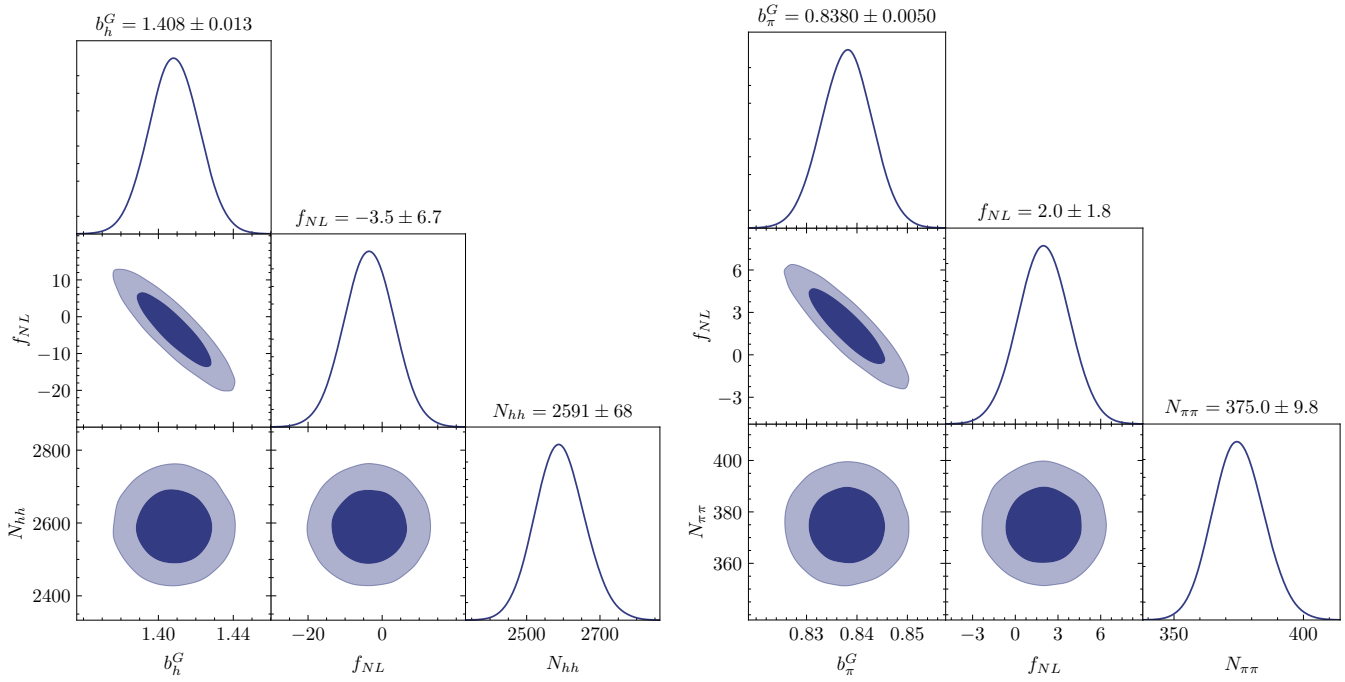


Figure 4. MCMC posteriors on f_{NL} and nuisance parameters (either (b_h^G, N_{hh}) or $(b_\pi^G, N_{\pi\pi})$) from joint analysis of 100 Quijote simulations with $f_{NL} = 0$. *Left.* Traditional halo based analysis using large-scale modes of the matter field $\delta_m(\mathbf{k})$ and halo field $\delta_h(\mathbf{k})$. *Right.* Neural network based analysis using $\delta_m(\mathbf{k})$ and the NN output field $\pi(\mathbf{k})$. The neural network reduces the error bar on f_{NL} by a factor $\sim$3.5.

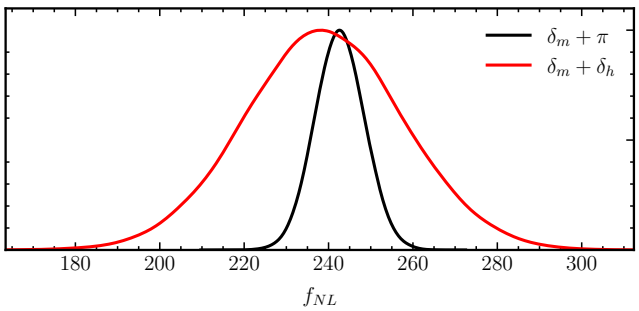


Figure 5. MCMC posterior on f_{NL} from joint analysis of 10 N-body simulations with $f_{NL} = 250$, using either the matter+halo fields (red), or matter+π fields (black), where $\pi(\mathbf{x})$ is the NN output field. The 1-d f_{NL} likelihoods are marginalized over nuisance parameters (either (b_h^G, N_{hh}) or $(b_\pi^G, N_{\pi\pi})$).

V. CONCLUSION AND OUTLOOK

In this letter we have demonstrated that the statistical power of neural networks can be combined with the idea of $(1/k^2)$ non-Gaussian bias to arrive at a robust measurement of f_{NL} from the matter distribution. Unlike forward modelling approaches which are difficult at strongly non-linear scales, our approach can use information from very small scales and still remain robust.

Our main next step will be to quantify to what extent the method presented here can improve f_{NL} constraints from realistic galaxy surveys, rather than the matter field. Machine learning based σ_8 constraints from simulated galaxy distributions have been examined in [34] (using a halo occupation distribution), [20] (using the hydrodynamic CAMELS simulations) and [35] (using a semi-analytic galaxy formation model). In particular, [20] highlighted the problem that different baryonic subgrid models lead to inconsistent results, which is precisely the issue our method is designed to overcome for f_{NL}. Recently, [36] introduced a non-linear estimator based on the Wavelet Scattering Transform (WST) and even applied it to BOSS data [37] to extract cosmological parameters including σ_8. The WST behaves similarly to a neural network and thus the claimed improvements in σ_8 suggest that our f_{NL} method could also work well for galaxies. We will investigate this question in detail in upcoming work.

A straightforward generalization of our method is to other scale-dependent biases, such as those induced by the trispectrum g_{NL} parameter [38], neutrino masses [39], and isocurvature perturbations [40]. More generally, our approach of using a neural network as a local probe may generalize to other observables which are large-scale modulations of local non-Gaussian fields or cross-correlations. This is a common setup in cosmology, often exploited for quadratic estimators.

ACKNOWLEDGEMENTS

Part of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. MM acknowledges support from DOE grant DE-SC0022342. KMS was supported by an NSERC Discovery Grant and a CIFAR fellowship. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. Perimeter Institutes's HPC system "Symmetry" was used to perform some of the analysis presented in the letter. We have extensively used several python libraries including `numpy`[41], `matplotlib`[42], `CLASS`[43], `getdist`[44] and `SciencePlots`[45].

[1] U. Seljak, G. Aslanyan, Y. Feng, and C. Modi, JCAP **12**, 009 (2017), arXiv:1706.06645 [astro-ph.CO].
[2] J. Jasche and B. D. Wandelt, MNRAS **432**, 894 (2013), arXiv:1203.3639 [astro-ph.CO].
[3] S. Ravanbakhsh, J. Oliva, S. Fromenteau, L. C. Price, S. Ho, J. Schneider, and B. Poczos, (2017), arXiv:1711.02033 [astro-ph.CO].
[4] F. Villaescusa-Navarro *et al.*, (2021), arXiv:2109.09747 [astro-ph.CO].
[5] A. Lazanu, JCAP **09**, 039 (2021), arXiv:2106.11061 [astro-ph.CO].
[6] F. Villaescusa-Navarro *et al.*, (2021), arXiv:2109.10360 [astro-ph.CO].
[7] H. J. Hortua, in *35th Conference on Neural Information Processing Systems* (2021) arXiv:2112.11865 [astro-ph.CO].
[8] A. Linde and V. Mukhanov, Physical Review D **56**, R535 (1997).
[9] G. Dvali, A. Gruzinov, and M. Zaldarriaga, Physical Review D **69** (2004), 10.1103/physrevd.69.023505.
[10] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Physics Reports **402**, 103 (2004).
[11] M. Biagetti, Galaxies **7**, 71 (2019).
[12] M. Alvarez *et al.*, (2014), arXiv:1412.4671 [astro-ph.CO].
[13] A. Moradinezhad Dizgah, M. Biagetti, E. Sefusatti, V. Desjacques, and J. Noreña, JCAP **05**, 015 (2021), arXiv:2010.14523 [astro-ph.CO].
[14] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008), arXiv:0710.4560 [astro-ph].
[15] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, JCAP **08**, 031 (2008), arXiv:0805.3580 [astro-ph].
[16] U. Seljak, Phys. Rev. Lett. **102**, 021302 (2009), arXiv:0807.1770 [astro-ph].
[17] K. M. Smith, M. S. Madhavacheril, M. Münchmeyer, S. Ferraro, U. Giri, and M. C. Johnson, (2018), arXiv:1810.13423 [astro-ph.CO].
[18] M. Münchmeyer, M. S. Madhavacheril, S. Ferraro, M. C. Johnson, and K. M. Smith, Phys. Rev. D **100**, 083508 (2019), arXiv:1810.13424 [astro-ph.CO].
[19] U. Giri and K. M. Smith, (2020), arXiv:2010.07193 [astro-ph.CO].
[20] P. Villanueva-Domingo and F. Villaescusa-Navarro, (2022), arXiv:2204.13713 [astro-ph.CO].
[21] F. Villaescusa-Navarro *et al.*, Astrophys. J. **915**, 71 (2021), arXiv:2010.00619 [astro-ph.CO].
[22] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, Journal of Cosmology and Astroparticle Physics **2008**, 031 (2008).
[23] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, Journal of Cosmology and Astroparticle Physics **2011**, 031 (2011).
[24] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rept. **733**, 1 (2018), arXiv:1611.09787 [astro-ph.CO].
[25] M. Biagetti, T. Lazeyras, T. Baldauf, V. Desjacques, and F. Schmidt, Monthly Notices of the Royal Astronomical Society **468**, 3277 (2017).
[26] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).
[27] F. Villaescusa-Navarro *et al.*, Astrophys. J. Suppl. **250**, 2 (2020), arXiv:1909.05273 [astro-ph.CO].
[28] D. P. Kingma and J. Ba, arXiv e-prints , arXiv:1412.6980 (2014), arXiv:1412.6980 [cs.LG].
[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," (2019).
[30] W. Falcon *et al.*, GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning **3**, 6 (2019).
[31] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, and H. Wu, "Mixed precision training," (2017).
[32] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publications of the Astronomical Society of the Pacific **125**, 306 (2013).
[33] N. Aghanim *et al.* (Planck), Astron. Astrophys. **641**, A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)], arXiv:1807.06209 [astro-ph.CO].
[34] M. Ntampaka, D. J. Eisenstein, S. Yuan, and L. H. Garrison, The Astrophysical Journal **889**, 151 (2020).
[35] L. A. Perez, S. Genel, F. Villaescusa-Navarro, R. S. Somerville, A. Gabrielpillai, D. Anglés-Alcázar, B. D. Wandelt, and L. Y. A. Yung, (2022), arXiv:2204.02408 [astro-ph.GA].
[36] G. Valogiannis and C. Dvorkin, (2021), arXiv:2108.07821 [astro-ph.CO].
[37] G. Valogiannis and C. Dvorkin, (2022), arXiv:2204.13717 [astro-ph.CO].
[38] K. M. Smith, S. Ferraro, and M. LoVerde, Journal of Cosmology and Astroparticle Physics **2012**, 032 (2012).
[39] C.-T. Chiang, W. Hu, Y. Li, and M. Loverde, Phys. Rev. D **97**, 123526 (2018), arXiv:1710.01310 [astro-ph.CO].
[40] A. Barreira, G. Cabass, D. Nelson, and F. Schmidt, JCAP **02**, 005 (2020), arXiv:1907.04317 [astro-ph.CO].

[41] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Nature **585**, 357 (2020).
[42] J. D. Hunter, Computing in Science & Engineering **9**, 90 (2007).
[43] D. Blas, J. Lesgourgues, and T. Tram, Journal of Cosmology and Astroparticle Physics **2011**, 034 (2011).
[44] A. Lewis, (2019), arXiv:1910.13970 [astro-ph.IM].
[45] J. D. Garrett, (2021), 10.5281/zenodo.4106649.