

# Hierarchical summaries for primordial non-Gaussianities

M. S. Cagliari<sup>1,\*</sup>, A. Bairagi<sup>2</sup>, and B. Wandelt<sup>3,4,2,5</sup>

<sup>1</sup> Laboratoire d'Annecy de Physique Theorique (LAPTh), CNRS/USMB, 99 Chemin de Bellevue BP110, Annecy, F-74941 Annecy Cedex, France

<sup>2</sup> CNRS & Sorbonne Université, Institut d'Astrophysique de Paris (IAP), UMR 7095, 98 bis bd Arago, F-75014 Paris, France

<sup>3</sup> Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

<sup>4</sup> Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

<sup>5</sup> Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

Received 27 October 2025 / Accepted 15 April 2026

## ABSTRACT

The advent of Stage IV galaxy redshift surveys such as DESI and *Euclid* marks the beginning of an era of precision cosmology, with one key objective being the detection of primordial non-Gaussianities (PNG), which are potential signatures of inflationary physics. In particular, constraining the amplitude of local-type PNG, parametrised by  $f_{\text{NL}}$ , with  $\sigma_{f_{\text{NL}}} \sim 1$ , would provide a critical test of single-versus-multi-field inflation scenarios. While current large-scale structure and cosmic microwave background analyses have achieved  $\sigma_{f_{\text{NL}}} \sim 5\text{--}9$ , further improvements demand novel data compression strategies. We propose a hybrid estimator that hierarchically combines standard two-point and three-point statistics with a field-level neural summary, motivated by recent theoretical works that have indicated that such a combination is nearly optimal in effectively disentangling primordial from late-time non-Gaussianity. We employed PATCHNET, a convolutional neural network that extracts small-scale information from sub-volumes (i.e. patches) of the halo number density field, while large-scale information is retained via the power spectrum and bispectrum. Using QUIJOTE-PNG simulations, we evaluated the Fisher information of this combined estimator across various redshifts, halo mass cuts, and scale cuts. Our results demonstrate that the inclusion of patch-based field-level compression always enhances constraints on  $f_{\text{NL}}$ , reaching gains of 30–45% at low  $k_{\text{max}}$  ( $\sim 0.1 h \text{Mpc}^{-1}$ ) and of 15–25% when the standard summary statistics include  $k$  modes comparable to those probed by the patches ( $k_{\text{max}} \sim 0.4 h \text{Mpc}^{-1}$ ). This shows that, even in this configuration, information from beyond the bispectrum can be captured. This approach offers a computationally efficient and scalable pathway to tightening the PNG constraints with forthcoming survey data.

**Key words.** methods: analytical – methods: statistical – cosmological parameters – large-scale structure of Universe

## 1. Introduction

With the delivery of results and observations from stage IV surveys, such as the data release 1 of the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016; Adame et al. 2025) and the first observations by *Euclid* (Euclid Collaboration: Mellier et al. 2025), we have entered the era of precision large-scale structure (LSS) cosmology. The key objective of new-generation surveys, such as DESI and *Euclid*, but also SPHEREx (Crill et al. 2020), as well as the Vera C. Rubin Legacy Survey of Space and Time (LSST; LSST Science Collaboration 2009), is to stress test the standard  $\Lambda$ CDM cosmological model, at late times, measuring the dark energy equation of state, and at early times, detecting primordial features that would serve as the smoking gun of inflation. In particular, inflationary models predict small deviations from Gaussianity in the primordial density distribution, called primordial non-Gaussianities (PNG), whose footprint we ought to be able to observe in the late-time distribution of galaxies.

Different inflation models produce different types of PNG, as described in Baumann (2011). In the present work, we are interested in the PNG of local type, which we parametrised with  $f_{\text{NL}}$ . This parameter controls the amplitude of the non-linear relation between a mean-zero Gaussian field,  $\varphi$ , and the primordial

gravitational potential,  $\Phi_p = \varphi + f_{\text{NL}}(\varphi^2 - \langle \varphi^2 \rangle)$ . We know that local PNG should be very close to zero in the case of single-field inflation (Maldacena 2003; Creminelli & Zaldarriaga 2004; Cabass et al. 2017), whereas we see  $f_{\text{NL}} \sim \mathcal{O}(1)$  for multi-field inflation models (Senatore & Zaldarriaga 2012; Alvarez et al. 2014). Therefore, a detection or non-detection of local PNG with  $\sigma_{f_{\text{NL}}} \sim 1$  would rule out single-field inflation or constraint multi-field models, respectively. For this reason, intensive efforts have taken place in the cosmological community to measure and improve our constraining power over this observable using different probes.

Currently, the most stringent constraints on  $f_{\text{NL}}$  come from the measurements of the three-point function of the cosmic microwave background (CMB) by *Planck*, which are consistent with zero at  $\sigma_{f_{\text{NL}}} \sim 5$  (Planck Collaboration IX 2020). We expect the new generation of CMB surveys to tighten these bounds by a factor of 2, but without reaching  $\sigma_{f_{\text{NL}}} \sim 1$  (Abazajian et al. 2016). On the other hand, local PNG affect the late-time LSS both at the two- and three-point function level. In Fourier space, this corresponds to the power spectrum and the bispectrum (Dalal et al. 2008; Slosar et al. 2008; Desjacques & Seljak 2010). The most stringent bounds from LSS measurements come from the power spectrum analysis of the first data release of DESI, expressed as  $\sigma_{f_{\text{NL}}} \sim 9$  (Chaussidon et al. 2025). This corresponds to a 40% improvement over the previous power

\* Corresponding author: marina.cagliari@lapth.cnrs.fr

spectrum-based measurement that used the quasar sample of data release 16 of the extended Baryon Oscillation Spectroscopic Survey (Cagliari et al. 2024) and an improvement of over 50% with respect to previous measurements based on the power spectrum and bispectrum (e.g. D’Amico et al. 2025; Cabass et al. 2022). This offers a first glimpse of the constraining power of stage IV surveys. These results will improve even more when the complete surveys become available.

While we are in the process of gathering new data, we can attempt to improve our constraints on  $f_{\text{NL}}$  (or other cosmological parameters) by devising new estimators that retain more information than the standard approach. For example, the authors of Castorina et al. (2019) computed the optimal quadratic estimator to measure the cosmological  $f_{\text{NL}}$  signal from the LSS power spectrum multipoles. In addition, many works have combined the power spectrum and the bispectrum (Cagliari et al. 2025), utilising effective field theory (EFT; see Cabass et al. 2023, for a review) to extend the scale range we can probe with these statistics to smaller scales. Alternatively, we could use field-level algorithms that bypass the data compression into  $n$ -point statistics. These methods are based on either forward modelling (Andrews et al. 2023, 2026) or machine learning algorithms applied to the late-time galaxy distribution (Kvasiuk et al. 2025), while also reconstructing the initial conditions (Chen et al. 2025). In addition, they can also be applied to the CMB (Nagarajappa & Ma 2024). Several detailed simulation studies have explored the choice (or combinations) of summary statistics to extract PNG information (Coulton et al. 2023b; Jung et al. 2023a; Coulton et al. 2024; Jung et al. 2024) and possibly guide the community effort in this area of study.

The work of Giri et al. (2023) produced highly informative estimates of  $f_{\text{NL}}$  from the non-linear matter field by training a neural network to compute local estimates of  $\sigma_8$  and modelling auto- and cross-spectra of the  $\sigma_8$  field and the matter. The methodology relies on a bias model specifically for local non-Gaussianity, which neatly bypasses the need for running training simulations with non-Gaussian initial conditions. However, it is not clear how this approach might be generalised to other forms of primordial non-Gaussianity.

In this work, we propose a method that approximates global field-based analysis by hierarchically combining standard  $n$ -point statistics with a local, neural summary that extracts field-level information. The main idea is that we can use the power spectrum and bispectrum optimally to extract large-scale information, along with machine learning-based data compression to extract small-scale information in the non-linear regime. Given a large cosmological volume, we expect that on large scales the majority of the  $f_{\text{NL}}$  information will be encoded in the two- and three-point functions; on the other hand, on small scales, a field-level algorithm would be expected to extract more information than the power spectrum and bispectrum alone (Bairagi & Wandelt 2026). Aside from the application to non-Gaussianity, we can go beyond the dark matter field and focus on catalogues of halos. Neither of them are directly observable, but we note that the halo field is a better proxy than the dark matter field in terms of a signal that is achievable with observations.

This hierarchical combination of global statistics with local field-level analysis is theoretically well-motivated for the problem of disentangling non-linear evolution from primordial non-Gaussianity. Recent works have demonstrated that locality fundamentally protects primordial signals from contamination by late-time gravitational nonlinearities (Baumann & Green 2022). Specifically, primordial non-Gaussianity creates genuinely non-local correlations between widely separated points

(generated during inflation), while late-time gravitational evolution is constrained by causality to produce only local effects that cannot mimic these long-range correlations.

This locality principle has important implications for data analysis. A Fisher analysis has shown that higher order bias parameters become increasingly orthogonal to primordial signals in the perturbative regime and that is only quadratic nonlinearities that affect the map-level analysis and all higher order non-linearities end up decoupled (Baumann & Green 2022). This suggests that combining two-point and three-point statistics with local field-level information naturally incorporates the constraints from higher order correlations without requiring their explicit computation. The map-level approach thus captures the essential information content on primordial non-Gaussianity more efficiently than computing progressively higher order correlation functions.

In this spirit, our algorithm measures the power spectrum and bispectrum of the whole observational volume, while it runs a field-level analysis on patches. We analyse the patches with a convolutional neural network (CNN; LeCun et al. 1989), named PATCHNET (Bairagi & Wandelt 2026), which summarises the field-level information as a single number and then aggregates over all the patches to get a single-value compression. Then, the final summary statistic is the combination of the two- and three-point summaries with the field-level compression. This combination was shown in Bairagi & Wandelt (2026) to produce state-of-the-art information extraction for cosmological parameters in the non-linear regime.

On a practical level, this approach has the advantages of a machine learning field-level analysis without the shortcoming of being too computationally expensive. The network only acts on small sub-volumes of the full computational volume (dubbed ‘patches’) and it greatly reduces the number of simulations required to train the networks, compared with a full field-based analysis, since dividing the volume into patches greatly increases the training set size.

In this paper, we estimate the  $f_{\text{NL}}$  information content of this new estimator applied to the QUIJOTE-PNG dark matter halos (Coulton et al. 2023a) in a cube of  $1 \text{ Gpc}^3 h^{-3}$ . We perform a detailed study of the Fisher information of combining the power spectrum with local patches and the combined power spectrum, bispectrum, and patches as a function of redshift, halo mass cut, and scale cut for the standard summary statistics.

The main finding of this work is in line with the theoretical expectations, asserting that patch-based data compression consistently provides additional information compared to standard summary statistics. The most significant improvements occur at the lowest  $k_{\text{max}}$  for the  $n$ -point summary statistics, with gains ranging between 30% and 45%. Furthermore, when comparing the power spectrum and bispectrum constraints to those derived from patch-combined estimators, the patches consistently enhance the results, demonstrating they capture information beyond the three-point function.

The paper is organised as follows. In Sect. 2, we present the simulated data we employed in the analysis. In Sect. 3, we introduce the Fisher formalism and the  $n$ -point statistics we used, along with a description of the architecture of PATCHNET. We discuss our results in Sect. 4 and present our conclusions in Sect. 5.

## 2. Simulations

The QUIJOTE simulations (Villaescusa-Navarro et al. 2020) are a set of N-body simulations run with GADGET-III, which was

originally developed for the Aquarius project (Springel et al. 2008) and is an improved version of the GADGET-II code (Springel et al. 2005). They are specifically designed to test machine learning applications on cosmological data. They start from initial conditions at  $z = 127$  generated with the 2LPTIC code<sup>1</sup> and simulate five redshift snapshots ( $z = 0, 0.5, 1, 2,$  and  $3$ ) with a co-moving volume of  $1000^3 \text{ Mpc}^3 h^{-3}$  and a fiducial resolution of  $512^3$  particles<sup>2</sup>. The main data products of the QUIJOTE simulations are the particle snapshots over which two halo finders were run. For this work, we used the halo catalogues produced with the friend-of-friends (FoF; Davis et al. 1985) halo finder.

The original QUIJOTE dataset spans over a range of  $\Lambda$ CDM and  $\nu$ wCDM cosmologies. Recently, a new set of simulations has become available, the QUIJOTE-PNG (Coulton et al. 2023a) simulations, which cover four different types of PNG: LSS orthogonal, CMB orthogonal, equilateral, and local PNG. The initial conditions of the QUIJOTE-PNG simulation were generated with the 2LPTPNG code (Coulton et al. 2023a)<sup>3</sup>. For this work, we only used the datasets related to local PNG. For the training of PATCHNET, we used the 1000  $f_{\text{NL}}$  local Latin hypercube simulations. In this set,  $f_{\text{NL}}$  uniformly spans the interval  $(-300, 300)$ . Then, we used 5000 realisations of the fiducial cosmology for the Fisher information study, where  $f_{\text{NL}} = 0$ . Finally, we also used the two sets of 500 simulations with  $f_{\text{NL}}$  displaced from the fiducial value. These simulations have  $f_{\text{NL}} = -50$  or  $f_{\text{NL}} = 50$  and are used to estimate the derivative of the summary statistics as a function of  $f_{\text{NL}}$ . For all these simulations, except for  $f_{\text{NL}}$ , all the other cosmological parameters are fixed to the fiducial values,  $\Omega_m = 0.3175$ ,  $\Omega_b = 0.049$ ,  $h = 0.6711$ ,  $n_s = 0.9624$ ,  $\sigma_8 = 0.834$ ,  $w = -1$ , and  $M_\nu = 0 \text{ eV}$ .

Considering the redshift range probed by current surveys (e.g. the Euclid Wide Survey; Euclid Collaboration: Scaramella et al. 2022; Euclid Collaboration: Mellier et al. 2025), we are mainly interested in the results at  $z = 1$ , which should better represent the clustering observed by this experiment. In addition, we studied the redshift dependence of the algorithm performance at the other four redshifts available in the Quijote dataset,  $z = 0, 0.5, 2,$  and  $3$ . We tested two scenarios: the first is the case where the same mass cut is applied at all redshifts (i.e. we selected halos with  $M_h > 3.12 \times 10^{13} M_\odot h^{-1}$ ) and the second case is where the number density is constant at different redshifts. In this latter case, at every redshift we chose a halo mass cut,  $M_{\text{halo}} > M_{\text{min}}$ , that would fix the number density of the halos in the fiducial cosmology to the number density at  $z = 3$  ( $n(z) \sim 5 \times 10^{-6} h^3 \text{ Mpc}^{-3}$ ), which is the least populated snapshot. We present the threshold masses,  $M_{\text{min}}$ , in Table 1.

### 3. Methods

#### 3.1. Fisher Information

It is possible to evaluate the amount of information and the constraining power of an estimator through the Fisher information analysis. The Fisher information matrix is defined as

$$F_{ij} = \left( \frac{\partial \mathbf{x}}{\partial \theta_i} \right)^T \mathbf{C}^{-1} \left( \frac{\partial \mathbf{x}}{\partial \theta_j} \right), \quad (1)$$

where  $\mathbf{x}$  is a summary statistic of the optimal unbiased estimator of the parameters  $\theta_i$  and  $\theta_j$ , and  $\mathbf{C}^{-1}$  is the inverse of the covari-

<sup>1</sup> <https://cosmo.nyu.edu/roman/2LPT/>

<sup>2</sup> High-resolution  $1024^3$ -particle and low-resolution  $256^3$ -particle simulations are available.

<sup>3</sup> <https://github.com/dsjameson/2LPTPNG>

**Table 1.** Threshold masses at different redshifts.

Redshift	$M_{\text{min}} [M_\odot h^{-1}]$
3.0	0.0
2.0	$3.47 \times 10^{13}$
1.0	$1.05 \times 10^{14}$
0.5	$1.91 \times 10^{14}$
0.0	$3.47 \times 10^{14}$

**Notes.** The values of  $M_{\text{min}}(z)$  keep the number density constant ( $n(z) \sim 5 \times 10^{-6} h^3 \text{ Mpc}^{-3}$ ) at the different redshifts in the fiducial cosmology.

ance matrix of the summary statistics,  $\mathbf{x}$ . The expected variance of parameter  $i$  is  $\sigma_i^2 = (F^{-1})_{ii}$ . Therefore, to estimate the constraining power of a given summary statistic, we first need to compute the covariance matrix of the estimator and its derivatives with respect to the parameters of interest, which in this study is only  $f_{\text{NL}}$ . We estimated both  $\mathbf{C}^{-1}$  and  $\frac{\partial \mathbf{x}}{\partial f_{\text{NL}}}$  numerically. We used 5000 QUIJOTE simulations in the fiducial cosmology to estimate the covariance matrix,  $\hat{\mathbf{C}}^{-1}$ , which we also corrected with the Hartlap factor (Hartlap et al. 2007),

$$\mathbf{C}^{-1} = \frac{n_s - n_x - 2}{n_s - 1} \hat{\mathbf{C}}^{-1}, \quad (2)$$

to account for the non-infinite number of data we use to estimate the covariance matrix. In Eq. (2),  $n_s$  is the number of simulations used to compute the covariance and  $n_x$  is the length of the summary statistic. Finally, we estimate the derivatives numerically from the finite difference,

$$\frac{\partial \mathbf{x}}{\partial f_{\text{NL}}} = \frac{\bar{\mathbf{x}}(f_{\text{NL}}^{\text{fid}} + \delta f_{\text{NL}}) - \bar{\mathbf{x}}(f_{\text{NL}}^{\text{fid}} - \delta f_{\text{NL}})}{2 \delta f_{\text{NL}}}, \quad (3)$$

where,  $\bar{\mathbf{x}}(f_{\text{NL}}^{\text{fid}} \pm \delta f_{\text{NL}})$  are the mean summary statistics of the simulations with  $f_{\text{NL}} = \pm 50$ . Then, as  $f_{\text{NL}}^{\text{fid}} = 0$ ,  $\delta f_{\text{NL}} = 50$ <sup>4</sup>.

#### 3.2. Summary statistics: Power spectrum and Bispectrum

In this work, we aim to combine large- and small-scale information to improve the constraining power over local  $f_{\text{NL}}$  measurements. A first solution is to use effective field theory to increase the minimum scale that we can probe with standard summary statistics. The alternative approach we explored is to use field-level data compression based on machine learning to extract small-scale information, while the large-scale information still comes from standard summary statistics.

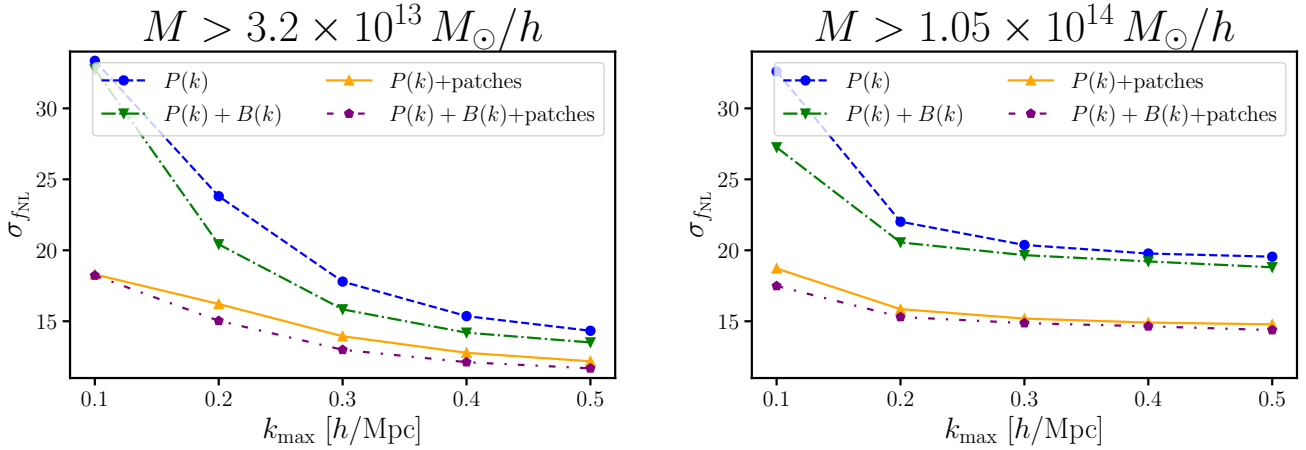
The summary statistics we utilised are the power spectrum and the bispectrum. Given the halo over-density,  $\delta_h$ , the power spectrum corresponds to the two-point correlation function in Fourier space and we define it as

$$\langle \delta_h(\mathbf{k}) \delta_h^*(\mathbf{k}') \rangle = (2\pi)^3 \delta^{\text{D}}(\mathbf{k} - \mathbf{k}') P(k), \quad (4)$$

where  $\delta^{\text{D}}$  is the 3D Dirac's delta function. Similarly, we can define the halo bispectrum, which is the three-point correlation function in Fourier space, as

$$\langle \delta_h(\mathbf{k}_1) \delta_h(\mathbf{k}_2) \delta_h(\mathbf{k}_3) \rangle = (2\pi)^3 \delta^{\text{D}}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3). \quad (5)$$

<sup>4</sup> To test convergence, we run the power spectrum Fisher analysis varying the number of simulations used to compute the numerical derivatives. We verified that the results are stable from 300 simulations.



**Fig. 1.** Expected uncertainty for the different summary statistics at  $z = 1$  for two halo-mass cuts as a function of the power spectrum and bispectrum  $k_{\text{max}}$ . In both cases, the PATCHNET information (patches) improves the constraining power of the power spectrum and the power spectrum combined with bispectrum analyses.

To measure the power spectra and bispectra of the simulated boxes, we used the public codes `Pylians3` (Villaescusa-Navarro 2018) and `pySpectrum` (Scoccimarro 2015; Hahn et al. 2020), respectively. We measured the power spectra on a  $512^3$  grid from the fundamental wave number of the box,  $k_f \sim 0.006 h \text{Mpc}^{-1}$ , to  $k_{\text{max}} \sim 0.8 h \text{Mpc}^{-1}$  with a  $k_f$  linear binning; on the other hand, we used a  $256^3$  grid for the bispectrum, measured with a  $3 k_f$  linear binning from  $k_{\text{min}} = 3 k_f$  to  $k_{\text{max}} \sim 0.5 h \text{Mpc}^{-1}$ .

### 3.3. Field level compression with PATCHNET

We employed field-level data compression, using a CNN to extract small-scale information following the approach of Bairagi & Wandelt (2026). The network does not analyse the whole simulated box; it only analyses a sub-cube, which we dubbed a ‘patch’. This network, called PATCHNET, offers two main advantages in its approach. First, it reduces the memory requirement for the CNN while maintaining the pixel (or better voxel) resolution, as it processes a smaller volume. Second, the number of training data greatly increases as we can cut out a set of  $N_{\text{sb}}$  patches from one simulated box to train the network, effectively increasing the amount of training data by a factor of  $N_{\text{sb}}$ . As we will explain below, we trained PATCHNET to compress each patch into one informative statistic (i.e. the value of  $f_{\text{NL}}$ ) and we then aggregated these statistics by averaging across the  $N_{\text{sb}}$  patches of a given simulated box. Finally, we concatenated this aggregated information with the power spectrum or the combination of the power spectrum and bispectrum.

PATCHNET is characterised by the following architecture: first, there are three blocks composed of a convolutional layer and a 3D average pooling layer. Their output is flattened and processed by three dense layers before the final output. The convolutional layers, as well as the average pooling, have a  $3 \times 3 \times 3$  kernel with a stride of 1 and 0 padding. The first convolutional layer has eight filters in terms of output and the filters are increased by a factor of 4 in the following layers. After flattening, the latent features are compressed into 512 neurons and subsequently halved in terms of their number by the other dense layers. Both the convolutional and dense layers have a LeakyReLU (Maas et al. 2013) with a negative slope of 0.5 as activation function. We

used a modified version of the mean squared error as a loss function,

$$\text{Loss}(\mathbf{y}, \hat{\mathbf{y}}) = \sum_i \frac{1}{n_b} \left( \sum_j \frac{1}{n_{\text{sb}}} y_{j,i} - \sum_j \frac{1}{n_{\text{sb}}} \hat{y}_{j,i} \right)^2, \quad (6)$$

where  $\mathbf{y}$  represents the outputs of PATCHNET,  $\hat{\mathbf{y}}$  the labels ( $f_{\text{NL}}$ ),  $n_b$  is the batch size (i.e. the number of simulations we loaded), and  $n_{\text{sb}}$  is the number of patches we use for each simulation. We used  $n_b = 32$  and  $n_{\text{sb}} = 64$ ; therefore, a batch actually contains  $n_b \times n_{\text{sb}}$  different inputs.

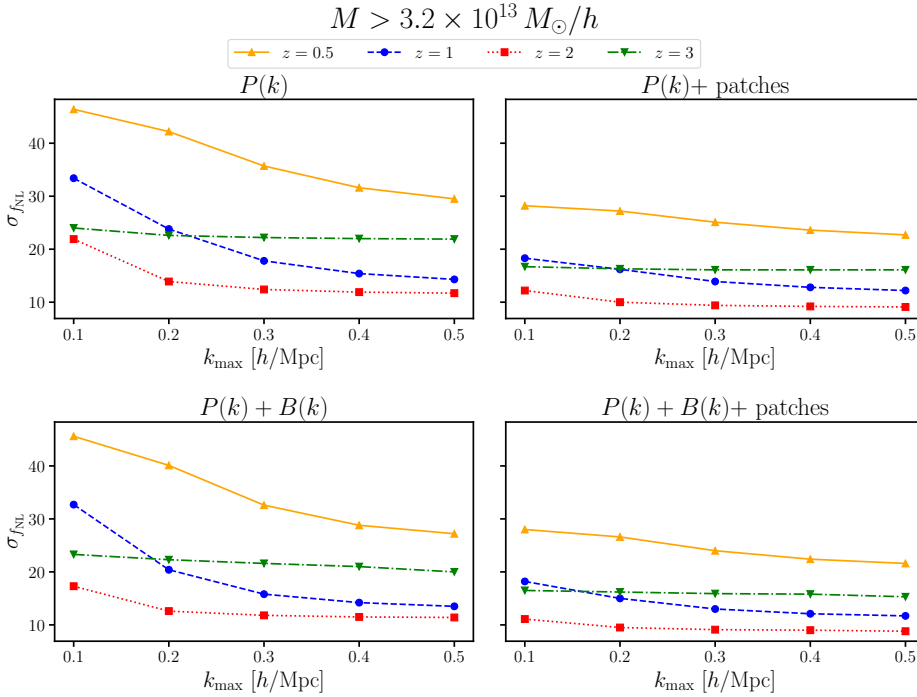
Before inputting the halo catalogues to the network, we pre-processed them. First, we divided the 1000 simulations with varying  $f_{\text{NL}}$  into the training, validation, and test sets, which we selected based on a standard split of 75%, 15%, and 10% of the data, respectively. Then, we built the patches from the simulated halo catalogue density contrast, which we computed by painting the halos on a  $128^3$  grid with the cloud in cell (CIC) algorithm. We note that during the painting process, we did not use any halo mass information. We divided the 3D number density contrast into  $16^3$  non-overlapping cubes<sup>5</sup>. This resulted in 512 patches for each simulation that have a physical dimension of  $125^3 \text{Mpc}^3 h^{-3}$  and a resolution of  $8 \text{Mpc} h^{-1}$ . In Fourier space, these scales correspond to  $k_{\text{min}} \sim 0.05 h \text{Mpc}^{-1}$  and  $k_{\text{max}} \sim 0.4 h \text{Mpc}^{-1}$ . Finally, before feeding the density contrast patches to the network, we centred and scaled them with the mean and standard deviations computed from the density contrast of the fiducial cosmology realisations. We note that we computed these two quantities for each redshift and mass cut configuration (see details in Sect. 2).

## 4. Results

In this section, we present and discuss the results of the Fisher information analysis. In particular, we compare the information content of the power spectrum and bispectrum with their combination with the patch information.

Figure 1 shows the results of the Fisher information analysis at redshift  $z = 1$  for two mass cuts as a function of

<sup>5</sup> This choice is conservative and produces an underestimation of the information content. Indeed, the authors of Bairagi & Wandelt (2026) showed that the use of overlapping patches increases information by 15% to 20%.



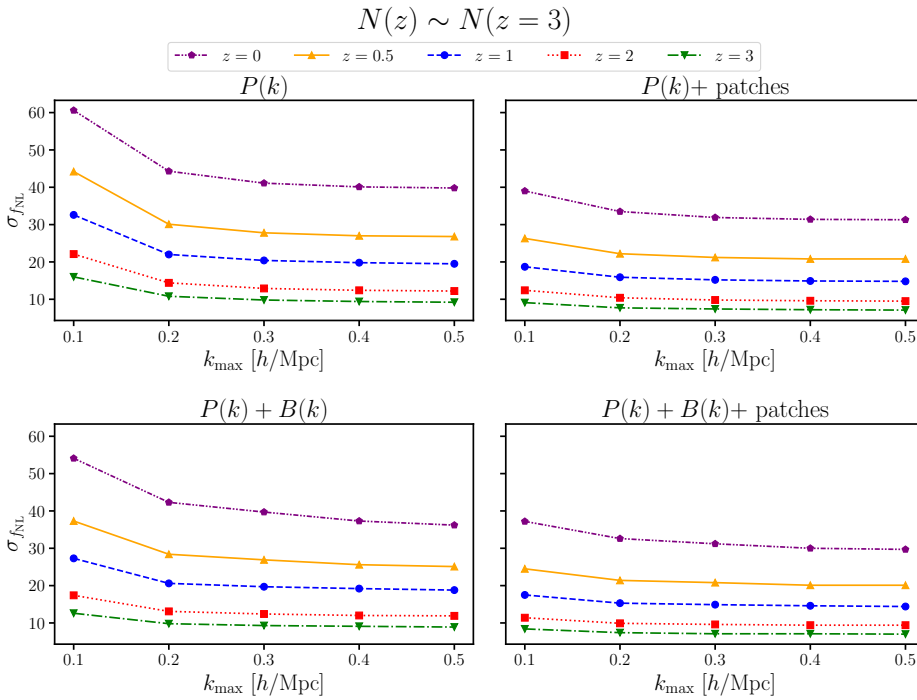
**Fig. 2.** Expected uncertainty for the different summary statistics as a function of the power spectrum and bispectrum  $k_{\max}$  at different redshift snapshots. At all redshifts, we applied the same halo-mass cut,  $M_h > 3.2 \times M_\odot h^{-1}$ . The patch information adds constraining power at every redshift.

the  $k_{\max}$  of the power spectrum and bispectrum. Adding patch information always produces tighter bounds than the standard summary statistics alone. Moreover, we also observe an improvement in the combination of the power spectrum and the patches (solid orange line with triangular markers) with respect to the combined power spectrum and bispectrum (dash-dotted green line with reverse triangular markers), showing that PATCHNET extracts information not captured by the two- and three-point functions. This is also backed up by the fact that even though the improvement due to the patch information diminishes when  $k_{\max}$  increases, as the minimum scale probed by the standard summary statistics gets closer to the minimum scale of the patches ( $k_{\max}^{\text{patch}} \sim 0.4 h \text{Mpc}^{-1}$ ), we still observe an improvement when  $k_{\max}^{\text{patch}} < k_{\max}^{P,B}$ , which indicates once again that the patches use information beyond the three-point function. Nevertheless, the bispectrum still brings in a small amount of information that the network cannot extract, as the combination of the three statistics (see the sparse dash-dotted purple line with pentagonal markers) is always slightly lower than the combination of the power spectrum and patches alone. This behaviour is not unexpected as the bispectrum has access to the very large-scale squeezed triangle configuration, which contains non-Gaussian information that the patches cannot access due to their limited size.

Comparing the left and right panels, we see that the overall information content of the smaller mass cut ( $M_{\min} = 3.2 \times 10^{13} M_\odot h^{-1}$  in the left panel) is higher than the larger mass cut ( $M_{\min} = 1.05 \times 10^{14} M_\odot h^{-1}$  in the right panel). This is because when the mass cut is smaller, the halo field has a lower shot noise and contains more information. However, the improvement related to the patch information is greater in the case of the larger mass cut for  $k_{\max} \geq 0.2 h \text{Mpc}^{-1}$ . This suggests that in terms of real objects, this method remains effective for the rarer, more highly biased objects that usually constitute the large volume required for  $f_{\text{NL}}$  measurements. The dependence of the performance on the bias would need to be explored with a more detailed tracer-specific study.

We present the constraining power of the different statistics as a function of redshift in Figs. 2 and 3. As shown in Fig. 2, we kept the  $M_{\min}$  constant, while Fig. 3 shows how we varied it as a function of redshift to keep the number density of the catalogues equal to the halo number density at  $z = 3$ . The feature that is common to both plots is that the addition of the patch information always makes the expected error at  $k_{\max} = 0.1 h \text{Mpc}^{-1}$  comparable to (if not better than) the constraint on the standard summary statistics alone at  $k_{\max} = 0.5 h \text{Mpc}^{-1}$ . This is an interesting result, considering that modelling the power spectrum and bispectrum becomes more involved as we enter the non-linear regime,  $k_{\max} > 0.1 h \text{Mpc}^{-1}$ , requiring a large number of simulations (or EFT) to model the observed data. The use of patch information can instead leverage the numerical simulation model while requiring a smaller simulation budget than neural estimators acting on the full field. Nevertheless, to provide a definitive statement about the advantage of the PatchNet approach over EFT modelling, we would need to perform additional tests. In particular, we would have to study the robustness of the algorithm against different halo finders and on non-linear and baryonic physics.

Finally, the authors of Jung et al. (2023b) have shown that there is a high information content related to  $f_{\text{NL}}$  in the halo mass function. As some halo mass information is retained in the density contrast, PATCHNET may be just learning the relation between the halo mass function and  $f_{\text{NL}}$ , making the field-level analysis superfluous. To check whether this is indeed the case, we trained PATCHNET keeping the number of objects in the catalogue fixed. The difference between this analysis and the one with constant density is that in the latter, the mass cut was chosen at the fiducial cosmology (with  $f_{\text{NL}} = 0$ ) and used for all the other simulated cosmologies. Therefore, there was a variation in the number density as a function of  $f_{\text{NL}}$ . On the other hand, for this test, we kept the number of objects fixed even when we vary the value of  $f_{\text{NL}}$ . In doing so, we removed the halo mass function information. Even in this configuration, PATCHNET learns a relationship between the density contrast field and  $f_{\text{NL}}$ . In this case,



**Fig. 3.** Expected uncertainty for the different summary statistics at fixed number density over redshifts as a function of the power spectrum and bispectrum  $k_{\max}$ . At each redshift, we selected a halo-mass cut that has in the fiducial cosmology the same number density of  $z = 3$  (see Table 1). As in Fig. 2, the patch information improves the constraining power of the standard summary statistics.

the Pearson correlation coefficient is  $r_{w/o \text{ HMF}} \sim 0.78$ , which we compared with  $r_{w \text{ HMF}} \sim 0.99$  for the standard configuration. This shows that the HMF helps improve the precision of the network, as  $r_{w \text{ HMF}} > r_{w/o \text{ HMF}}$ . Nevertheless, the network can still extract PNG information even when we remove the HMF from the field, as  $r_{w/o \text{ HMF}} \gg 0.5$ .

## 5. Conclusions

In this work, we propose an alternative approach to EFT up to large  $k_{\max}$  or more standard ML-based field-level algorithms to analyse galaxy redshift surveys. The algorithm combines the large-scale information encoded in the  $n$ -point summary statistics, such as the power spectrum or bispectrum, with the small-scale information obtained by averaging over the compressed field-level information of small patches of the simulated box, the PATCHNET approach (Bairagi & Wandelt 2026). The main advantage of this approach is that the analysis of small patches is computationally lighter than analysing the whole box; additionally, by cutting the simulated boxes into patches, we can greatly increase the number of training examples for the PATCHNET (Bairagi et al. 2025). We estimated the Fisher information of this combined summary statistics to extract local PNG information.

We tested the algorithm on the FoF halo catalogues of the QUIJOTE-PNG simulations. We compared its performance with the more standard analyses based on the power spectrum or the combination of the power spectrum and bispectrum. We performed this comparison for different values of  $k_{\max}$  for the  $n$ -point statistics and studied it as a function of redshift and for different halo mass cuts.

Our preliminary study shows that the patch-based data compression always provides additional information relative to the standard summary statistics. We observed the greatest improvements when the power spectrum and the bispectrum were truncated at  $k_{\max} = 0.1 h \text{ Mpc}^{-1}$ , where the small-scale structure in the patches provides the most complementary information. At this scale (depending on the redshift, mass cut, and the

statistics combination), the improvement varies between 30% and 45%. At  $k_{\max} = 0.4 h \text{ Mpc}^{-1}$ , where the patches and the standard summary statistics get to the same minimum scales, the improvement falls to a more moderate range between 15% and 25%, but it is still present. Moreover, when we compare the power spectrum plus bispectrum bounds with the patch-combined bounds (both power spectrum with patches and power spectrum plus bispectrum with patches), the patches once again always bring an improvement, proving that they contain information beyond the three-point function. Nevertheless, we also observed that combining the bispectrum with the PATCHNET compression also provides an improvement, demonstrating that the large-scale bispectrum encodes information that the patches end up missing due to their limited size. In this regard, it is possible that the use of larger volumes (thereby increasing access to the squeezed bispectrum information) could actually reduce the improvement observed when adding the patch information, especially in the case of the local  $f_{\text{NL}}$ . On the other hand, local higher order information should help disentangle the non-linear coupling due to non-linear gravity from primordial non-Gaussianity (Baumann & Green 2022). The use of larger volumes will also allow us to test the effect super-sample covariance has on the method, even though we expect this effect to be subdominant at the patch-level, as the patches are already sub-volumes of the bigger simulation box. We would already be observing the effect of super-sample covariance.

A first extension to this work would be to test our method's performance for other forms of PNG and the effect of marginalising over cosmological parameters. Second, we need to perform robustness tests to check whether the method has the same performance with different and more accurate halo finders. Third, to fully understand the potential of the patch-based approach, it would be of interest to compare its information content with other alternative summary statistics; for instance, a wavelet scattering transform (e.g. Peron et al. 2024) or a marked power spectrum (e.g. Marinucci et al. 2025). In addition, to prepare this approach for applications to real data, we will need to introduce

a galaxy model in the simulations and add the lightcone, the survey mask, and systematic effects to the simulations. We plan to study these issues in future work.

*Acknowledgements.* We thank Francisco Villaescusa-Navarro, William Coulton, and Michele Liguori for their feedback on the first version of this manuscript. MSC thanks ‘Fondazione Angelo della Riccia’ for financial support. The work of MSC is supported by the Agence Nationale de la Recherche (ANR) grant n. ANR-23-CPJ1-0160-01. AB acknowledges support from the Simons Foundation as part of the Simons Collaboration on Learning the Universe. The Flatiron Institute is supported by the Simons Foundation. All neural training and post-analysis have been done on IAP’s Infinity cluster. The authors thank Stéphane Rouberol for his efficient management of this facility.

## References

- Abazajian, K. N., Adshead, P., Ahmed, Z., et al. 2016, arXiv e-prints [arXiv:1610.02743]
- Adame, A. G., Aguilar, J., Ahlen, S., et al. 2025, *JCAP*, 2025, 028
- Alvarez, M., Baldauf, T., Bond, J.R., et al. 2014, arXiv e-prints [arXiv:1412.4671]
- Andrews, A., Jasche, J., Lavaux, G., & Schmidt, F. 2023, *MNRAS*, 520, 5746
- Andrews, A., Jasche, J., Lavaux, G., et al. 2026, A&A, in press, <https://doi.org/10.1051/0004-6361/202553802>
- Bairagi, A., & Wandelt, B. 2026, *JCAP*, 2026, 028
- Bairagi, A., Wandelt, B., & Villaescusa-Navarro, F. 2025, A&A, 703, A301
- Baumann, D. 2011, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, 523
- Baumann, D., & Green, D. 2022, *JCAP*, 2022, 061
- Cabass, G., Pajer, E., & Schmidt, F. 2017, *JCAP*, 2017, 003
- Cabass, G., Ivanov, M. M., Philcox, O. H. E., Simonović, M., & Zaldarriaga, M. 2022, *Phys. Rev. D*, 106, 043506
- Cabass, G., Ivanov, M. M., Lewandowski, M., Mirbabayi, M., & Simonović, M. 2023, *Phys. Dark Universe*, 40, 101193
- Cagliari, M. S., Castorina, E., Bonici, M., & Bianchi, D. 2024, *JCAP*, 2024, 036
- Cagliari, M. S., Barberi-Squarotti, M., Pardede, K., Castorina, E., & D’Amico, G. 2025, *JCAP*, 2025, 043
- Castorina, E., Hand, N., Seljak, U., et al. 2019, *JCAP*, 2019, 010
- Chaussidon, E., Yèche, C., de Mattia, A., et al. 2025, *JCAP*, 2025, 029
- Chen, X., Padmanabhan, N., & Eisenstein, D. J. 2025, *JCAP*, 2025, 055
- Coulton, W. R., Villaescusa-Navarro, F., Jamieson, D., et al. 2023a, *ApJ*, 943, 64
- Coulton, W. R., Villaescusa-Navarro, F., Jamieson, D., et al. 2023b, *ApJ*, 943, 178
- Coulton, W. R., Abel, T., & Banerjee, A. 2024, *MNRAS*, 534, 1621
- Creminelli, P., & Zaldarriaga, M. 2004, *JCAP*, 2004, 006
- Crill, B. P., Werner, M., Akeson, R., et al. 2020, in *Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave*, eds. M. Lystrup, & M. D. Perrin, *SPIE Conf. Ser.*, 11443, 114430I
- Dalal, N., Doré, O., Huterer, D., & Shirokov, A. 2008, *Phys. Rev. D*, 77, 123514
- D’Amico, G., Lewandowski, M., Senatore, L., & Zhang, P. 2025, *Phys. Rev. D*, 111, 063514
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371
- DESI Collaboration (Aghamousa, A., et al.) 2016, arXiv e-prints [arXiv:1611.00036]
- Desjacques, V., & Seljak, U. 2010, *Adv. Astron.*, 2010, 908640
- Euclid Collaboration (Scaramella, R., et al.) 2022, A&A, 662, A112
- Euclid Collaboration (Mellier, Y., et al.) 2025, A&A, 697, A1
- Giri, U., Münchmeyer, M., & Smith, K. M. 2023, *Phys. Rev. D*, 107, L061301
- Hahn, C., Villaescusa-Navarro, F., Castorina, E., & Scoccimarro, R. 2020, *JCAP*, 2020, 040
- Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
- Jung, G., Karagiannis, D., Liguori, M., et al. 2023a, *ApJ*, 948, 135
- Jung, G., Ravenni, A., Baldi, M., et al. 2023b, *ApJ*, 957, 50
- Jung, G., Ravenni, A., Liguori, M., et al. 2024, *ApJ*, 976, 109
- Kvasiuk, Y., Münchmeyer, M., & Smith, K. 2025, *Phys. Rev. D*, 112, 023540
- LeCun, Y. 1989, in *Generalization and Network Design Strategies*, eds. R. Pfeifer, Z. Schreter, F. Fogelman, & L. Steels (Elsevier)
- LSST Science Collaboration (Abell, P. A., et al.) 2009, arXiv e-prints [arXiv:0912.0201]
- Maas, A.L., Hannun, A.Y., & Ng, A.Y. 2013 Proceedings of the 30th International Conference on Machine Learning (ICML), 30
- Maldacena, J. 2003, *J. High Energy Phys.*, 2003, 013
- Marinucci, M., Jung, G., Liguori, M., et al. 2025, *JCAP*, 2025, 036
- Nagarajappa, C. G., & Ma, Y.-Z. 2024, *MNRAS*, 529, 3289
- Peron, M., Jung, G., Liguori, M., & Pietroni, M. 2024, *JCAP*, 2024, 021
- Planck Collaboration IX. 2020, A&A, 641, A9
- Scoccimarro, R. 2015, *Phys. Rev. D*, 92, 083532
- Senatore, L., & Zaldarriaga, M. 2012, *J. High Energy Phys.*, 2012, 24
- Slosar, A., Hirata, C., Seljak, U., Ho, S., & Padmanabhan, N. 2008, *JCAP*, 2008, 031
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, *MNRAS*, 391, 1685
- Villaescusa-Navarro, F. 2018, Pylians: Python libraries for the analysis of numerical simulations, Astrophysics Source Code Library [record ascl:1811.008]
- Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020, *ApJS*, 250, 2