

# Baryon acoustic oscillations reconstruction using convolutional neural networks

Tian-Xiang Mao <sup>1,2</sup>★, Jie Wang <sup>1,2</sup>★, Baojiu Li <sup>3</sup>, Yan-Chuan Cai <sup>4</sup>, Bridget Falck <sup>5</sup>, Mark Neyrinck <sup>6</sup> and Alex Szalay<sup>5</sup>

<sup>1</sup>National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK

<sup>4</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

<sup>5</sup>Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA

<sup>6</sup>Ikerbasque, the Basque Foundation for Science and Department of Physics, University of the Basque Country, E-48080 Bilbao, Spain

Accepted 2020 November 25. Received 2020 October 31; in original form 2020 February 24

## ABSTRACT

We propose a new scheme to reconstruct the baryon acoustic oscillations (BAO) signal, which contains key cosmological information, based on deep convolutional neural networks (CNN). Trained with almost no fine tuning, the network can recover large-scale modes accurately in the test set: the correlation coefficient between the true and reconstructed initial conditions reaches 90 per cent at  $k \leq 0.2 h\text{Mpc}^{-1}$ , which can lead to significant improvements of the BAO signal-to-noise ratio down to  $k \simeq 0.4 h\text{Mpc}^{-1}$ . Since this new scheme is based on the configuration-space density field in sub-boxes, it is local and less affected by survey boundaries than the standard reconstruction method, as our tests confirm. We find that the network trained in one cosmology is able to reconstruct BAO peaks in the others, i.e. recovering information lost to non-linearity independent of cosmology. The accuracy of recovered BAO peak positions is far less than that caused by the difference in the cosmology models for training and testing, suggesting that different models can be distinguished efficiently in our scheme. It is very promising that our scheme provides a different new way to extract the cosmological information from the ongoing and future large galaxy surveys.

**Key words:** cosmological parameters – dark energy – large-scale structure of Universe.

## 1 INTRODUCTION

Understanding observations and using them to constrain the nature of physics is a long-term task in modern cosmology, which requires both obtaining high-quality data and developing accurate data analysis methods. The baryon acoustic oscillations (BAOs), imprinted on large-scale structure, is a standard ruler in cosmology and play an important role in studying the cosmic expansion history, and the properties of dark energy.

BAOs arise from the coupling of baryons and photons in the early Universe (for a recent review see Weinberg et al. 2013). After recombination, this feature is imprinted in both the cosmic microwave background (CMB) and the matter distribution. CMB anisotropy measurements have provided highly precise constraints on the sound horizon at high redshift (Peebles & Yu 1970; Sunyaev & Zeldovich 1970, and for more recent results see Bennett et al. 2013; Planck Collaboration XIII 2016). In the case of galaxy clustering, the BAO feature imprinted in the form of a characteristic scale provides an absolute distance scale, and can be used to map the expansion history  $H(z)$ . It has been detected at redshift about  $0.1 \sim 0.8$  and  $2.5$  (for recent results and a summary of BAO measurements, see e.g. Alam et al. 2017) as a peak in the correlation function or as a harmonic

sequence of oscillations in the power spectrum. Fortunately, the relatively large scale of the BAO feature (about 145 Mpc) protects it from substantial non-linearity, making it straightforward to interpret, and a powerful tool for measuring the cosmological distance scale.

However, late-time non-linear evolution does broaden and shift the BAO peak in the correlation function, or, equivalently, dampen high- $k$  oscillations in the power spectrum, which decreases the accuracy and precision of the detection of the BAO signal (Meiksin, White & Peacock 1999; Angulo et al. 2005; Seo & Eisenstein 2005; Springel et al. 2005; Jeong & Komatsu 2006; Huff et al. 2007; Eisenstein, Seo & White 2007a; Angulo et al. 2008; Padmanabhan & White 2009; Seo et al. 2010; Mehta et al. 2011; Sherwin & Zaldarriaga 2012). Furthermore, a number of other effects can also introduce difficulties in BAO measurement, such as the survey boundary, galaxy bias, and redshift-space distortions. In order to correct this blurring caused by non-linear evolution, Eisenstein et al. (2007b) proposed a reconstruction method (hereafter ‘standard reconstruction’) by moving the galaxies back along large-scale bulk flows, which considerably enhances the BAO peak both in theory (Seo et al. 2008; Noh, White & Padmanabhan 2009; Padmanabhan, White & Cohn 2009; Seo et al. 2010; Mehta et al. 2011; Schmittfull et al. 2015; White 2015) and in observations (Padmanabhan et al. 2012; Xu et al. 2013; Anderson et al. 2014; Kazin et al. 2014; Ross et al. 2015; Beutler et al. 2016, 2017; Hinton et al. 2017).

\* E-mail: maotianxiang@bao.ac.cn (TXM); jie.wang@nao.cas.cn (JW)

Recently, motivated by the success of standard reconstruction and the current or upcoming observations (e.g. 2MASS, Skrutskie et al. 2006; 4MOST, de Jong et al. 2012; 6dF, Jones et al. 2009; SDSS, Alam et al. 2017; DES, Dark Energy Survey Collaboration 2016; PFS, Takada et al. 2014; DESI, DESI Collaboration 2016); *EUCLID*, Laureijs et al. 2011; LSST, Ivezić et al. 2008; Tianlai, Xu, Wang & Chen 2015; CHIME, Bandura et al. 2014; HIRAX, Newburgh et al. 2016; BINGO, Battye et al. 2016; and SKA, Godfrey et al. 2012), various other reconstruction methods have been proposed and gained broader applications (for a review, see Schmittfull, Baldauf & Zaldarriaga 2017). For example, Zhu, Pen & Chen (2016), Zhu et al. (2017) proposed a non-linear reconstruction technique based on iteratively solving the coordinate transform between the Lagrangian and Eulerian frames. It has been tested for dark matter density fields (Zhu et al. 2016, 2017), and it has been investigated with respect to Fisher information (Pan et al. 2017), BAO (Wang et al. 2017), biased tracers (Yu, Zhu & Pen 2017; Wang & Pen 2019), and redshift-space distortions (Zhu, Yu & Pen 2018). Schmittfull et al. (2017) described an iterative method to reconstruct the initial conditions, and Seljak et al. (2017), Feng, Seljak & Zaldarriaga (2018), Modi, Feng & Seljak (2018) converted the reconstruction to an optimization problem by forward modelling. For similar purposes, Shi, Cautun & Li (2018) proposed a multigrid relaxation algorithm and extended it for biased tracers (Birkin et al. 2019) and to remove redshift-space distortions from galaxy clustering (Wang, Li & Cautun 2020). Most of these methods achieve substantial improvements beyond the standard reconstruction and some other methods have been designed to gain more information in some specific cases (e.g. Kitaura 2013; Jasche & Wandelt 2013; Wang et al. 2013; Burden, Percival & Howlett 2015; Obuljen et al. 2017; Hada & Eisenstein 2018, 2019; Bos, Kitaura & van de Weygaert 2019; Kitaura et al. 2019; Leclercq et al. 2019; Sarpa et al. 2019).

Another potential approach to extract the BAO features from galaxy surveys is by using neural networks (hereafter networks; for some reviews see Lecun, Bengio & Hinton 2015; Goodfellow, Bengio & Courville 2016), which have been widely used in various fields in astronomy, such as gravitational lensing (Gupta et al. 2018; Morningstar et al. 2018, 2019; Li et al. 2019; Tewes et al. 2019; Springer et al. 2020), the Cosmic Microwave Background (Caldeira et al. 2019), neutral hydrogen (Shimabukuro & Semelin 2017; Rafieerantsoa, Andrianomena & Davé 2018; Gillet et al. 2019; Villanueva-Domingo & Villaescusa-Navarro 2020), constraining cosmological parameters (Ravanbakhsh et al. 2017; Schmelzle et al. 2017; Gupta et al. 2018; Mathuriya et al. 2018), large-scale structure classification (Aragon-Calvo 2019), and generation (Rodríguez et al. 2018) and structure formation (Lucie-Smith et al. 2018; Modi et al. 2018; Berger & Stein 2019; He et al. 2019; Lucie-Smith, Peiris & Pontzen 2019).

In this paper, we propose a new, network-based method to reconstruct the BAO signal from a dark matter density field. We use high-resolution  $N$ -body simulations, which provide all necessary information to construct our network model. In this case, we convert the reconstruction problem to a non-linear mapping from final non-linear density to initial linear density, by introducing a large number of parameters, which are optimized by feeding simulation data to the network for training.

This is an independent method from the ones commonly used in BAO analyses, and can be used to identify and understand potential modelling systematics in BAO measurements. Unlike methods based on perturbation theory, our method is more robust in regions close to the survey boundary, because it reconstructs the initial linear density from the local final non-linear density in configuration space. In

contrast, in Fourier space, the effect of the survey boundary is a broad window function that is convolved on to the density field, which can have a global impact on reconstruction. This effect becomes local in configuration space, reducing the impact on our reconstruction.

This paper is organized as follows. In Section 2 we describe our network model and the simulations used in this work. It is followed by results of the reconstruction in Section 3. We then discuss cosmology dependence and survey boundary effects in Section 4, and finally conclude in Section 5.

## 2 METHOD

In this section, we first describe the reconstruction problem by way of maximum-likelihood estimation and show that the reconstruction can be represented by the network. After that, we review traditional and convolutional neural networks and describe the network model we used in this work. Finally, we show the training process and describe the data set used in this paper.

### 2.1 Maximum-likelihood estimation

In general, the non-linear power spectrum of the final density field can be expressed as a sum of two parts (Crocco & Scoccimarro 2006):

$$P_{\text{nl}}(k) = G^2(k)P_{\text{ini}}(k) + P_{\text{mc}}(k), \quad (1)$$

where the  $P_{\text{ini}}(k)$  is the linear power spectrum,  $G(k)$  is the propagator, and  $P_{\text{mc}}(k)$  indicates the power spectrum from mode coupling. The propagator term encodes the memory of the initial conditions, and  $G(k) \rightarrow 1$  as  $k \rightarrow 0$ , indicating that the information on large scales is well preserved even at late times. On intermediate scales, mode coupling modulates large-scale information into smaller, non-linear, scales. In regions where shells cross, small-scale information is lost, and it cannot be estimated uniquely if only the final density field is provided. Since the BAO features are located on large and intermediate scales, shell crossing is not a major concern for BAO reconstruction. As we focus on BAO reconstruction in this paper, we shall, following the standard practice in the reconstruction literature, assume there is no shell crossing.

We suppose a parametric model  $f(\delta_{\text{f}}; \theta)$ , with parameter set  $\theta$ , that can predict the initial linear density field  $\delta_{\text{i}}$  above a certain length scale, given the corresponding final density field  $\delta_{\text{f}}$ . It can be written as

$$\delta_{\text{i}} = f(\delta_{\text{f}}; \theta), \quad (2)$$

and the corresponding conditional probability is  $P(\delta_{\text{i}}|\delta_{\text{f}}; \theta)$ . Under the assumption of independent and identically distributed (i.i.d.) data points, the likelihood is

$$\mathcal{L} = \prod_{k=1}^{N_{\text{p}}} P(\delta_{\text{i}}|\delta_{\text{f}}; \theta), \quad (3)$$

where  $N_{\text{p}}$  is total number of data points, i.e. pixels or cells where the density fields are evaluated, and the maximum-likelihood estimation of  $\theta$  is

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} \mathcal{L} = \arg \max_{\theta} \sum_k \log P(\delta_{\text{i}}|\delta_{\text{f}}; \theta). \quad (4)$$

If we assume

$$P(\delta_{\text{i}}|\delta_{\text{f}}; \theta) = \mathcal{N}(\delta_{\text{i}}; f(\delta_{\text{f}}; \theta), \sigma), \quad (5)$$

where the measured (or simulated)  $\delta_{\text{i}}$  is regarded as a data point sampled from a Gaussian distribution  $\mathcal{N}$ , the function  $f(\delta_{\text{f}}; \theta)$  gives

the prediction of the mean of  $\mathcal{N}$  and a fixed standard deviation  $\sigma$  is assumed to simplify the problem. In such conditions, the maximum-likelihood estimation is equivalent to minimizing the mean square error (MSE),

$$\hat{\theta}_{\text{ML}} = \arg \min_{\theta} \sum_{\theta}^k (f(\delta_f^k; \theta) - \delta_i^k)^2, \quad (6)$$

where the superscript  $k$  again indicates i.i.d. data points. We assume a Gaussian distribution of  $\mathcal{N}$  in equation (5) because there is no preferred distribution. The Gaussian distribution is a natural choice according to the central limit theorem, which applies when the number of data points is large.

In the reconstruction problem, it is not easy to estimate  $\theta_{\text{ML}}$  directly because we do not know the mathematical form of the model  $f(\delta_f; \theta)$ . A possibility is fitting the parameters  $\theta$  from observed data with the help of an optimization method. However, due to the complexity of the model and the huge number of parameters, traditional optimization methods, such as Markov chain Monte Carlo (MCMC) are not efficient. In this work, we tackle this difficulty by using gradient-based neural network to construct model and optimize parameters.

## 2.2 Artificial neural networks

Artificial neural networks (see some reviews, Lecun et al. 2015; Goodfellow et al. 2016) are suitable for solving problems with no known specific mathematical expressions. By constructing a non-linear parametric model, the network converts complex problems into non-convex optimization and optimizes the trainable parameters by gradient-descent based methods (e.g. stochastic gradient descent; Bottou 1998). The process of optimizing trainable parameters by feeding a series of data points into a fixed network architecture is called training.

A standard feed-forward neural network consists of multiple layers. Each layer performs a weighted linear combination of its inputs, followed by an element-wise non-linear activation function and a bias term. These weights and biases on all layers constitute the trainable parameters of the network.

For layer  $n$ , if we set the input vector as  $\mathbf{x}_{n-1}$ , weight matrix  $\mathbf{W}_n$  and bias vector  $\mathbf{b}_n$ , then the output of this layer is

$$\mathbf{x}_n = a(\mathbf{W}_n \mathbf{x}_{n-1} + \mathbf{b}_n). \quad (7)$$

Here the  $a$  denotes a non-linear activation function. In this paper, we use the rectified linear unit (ReLU; Nair & Hinton 2010) activation function. For the network, the output of one layer is the input of the next layer. By stacking a series of functions in equation (7), the network will have the potential to approximate the  $f(\delta_f; \theta)$  in equation (6) and the trainable parameters correspond to the parameter set  $\theta$ .

In deep learning, increasing the number of layers  $N$  always expands the capacity of the network, by enlarging the hypothesis space of solutions that the algorithm is able to choose from, although it may lead to difficulties in training. Once the network architecture is determined, the trainable parameters in the network will be optimized by minimizing a loss function. The loss function describes a kind of distance between the network prediction and the target value. In this paper, we choose MSE loss as the loss function.

## 2.3 Convolutional neural networks

Convolutional neural networks (hereafter CNNs, see e.g. LeCun et al. 1990; Krizhevsky, Sutskever & Hinton 2012) are well known in

processing visual imagery because of their shift-invariant property and reduced number of free parameters. In this study, we will perform the MSE estimation described in equation (6) using a 3D CNN.

CNNs replace the matrix-vector product  $\mathbf{W}_n \mathbf{x}_{n-1}$  in equation (7) by a sum of convolutions, the latter being more efficient and having fewer trainable parameters. Like in equation (7), we represent the output of the  $l$ -th kernel in layer  $n$  as

$$\mathbf{x}_n^l = a \left( \sum_{k=1}^k \mathbf{W}_n^l \otimes \mathbf{x}_{n-1}^k + \mathbf{b}_n^l \right), \quad (8)$$

where  $\otimes$  indicates the 3D convolution operation,  $\mathbf{W}_n^l$  denotes trainable convolutional kernels for layer  $n$ ,  $l$  indicates the  $l$ -th kernel in this layer, and  $k$  indicates the output corresponding to the  $k$ -th convolutional kernel in the previous layer, which is also called the  $k$ -th channel.

In addition to the convolutional layers, standard CNNs usually contain pooling layers (e.g. Krizhevsky et al. 2012). In our network, the pooling layers are replaced by a striding in 2 voxels per side in the convolutional calculation. For all convolutional kernels in this paper, we set their size as  $3 \times 3 \times 3$  voxels, except the last convolution layer whose kernel size is  $1 \times 1 \times 1$ . The detailed network architecture is shown in Table 1.

As described in equation (2), the input and output of the network are the final density field  $\delta_f$  and the initial density field  $\delta_i$ , respectively. To further reduce the computational and memory requirements in the training, we generate  $\delta_f$  and  $\delta_i$  fields in a small sub-box instead of the whole simulation box.  $\delta_f$  is generated in a cubic region with a length of  $76 \text{ Mpc } h^{-1}$  per side. For  $\delta_i$ , we choose the corresponding central region at initial time with a length of  $1.95 \text{ Mpc } h^{-1}$  per side. This is because as the structure evolves, the particles initially located in a small region can diffuse to a larger volume.<sup>1</sup> In other words, a big region with enough volume contains almost all information of its central sub-region at the initial time. More details about the data can be found in Section 2.5.

## 2.4 Training

Once the network architecture is fixed, training can help us optimize the random parameters to suitable values. In the training process, there are some hyper-parameters which should be selected, such as in the activation function  $a$  and the loss function. The search for the optimal hyperparameters is called fine-tuning, which requires us to train the network multiple times with different hyperparameters (Goodfellow et al. 2016). In this paper, we report our preliminary results of BAO reconstruction by using a neural network, for which the hyperparameters are selected roughly and we defer further fine-tuning to a future study. Below, we list the hyperparameters that are used:

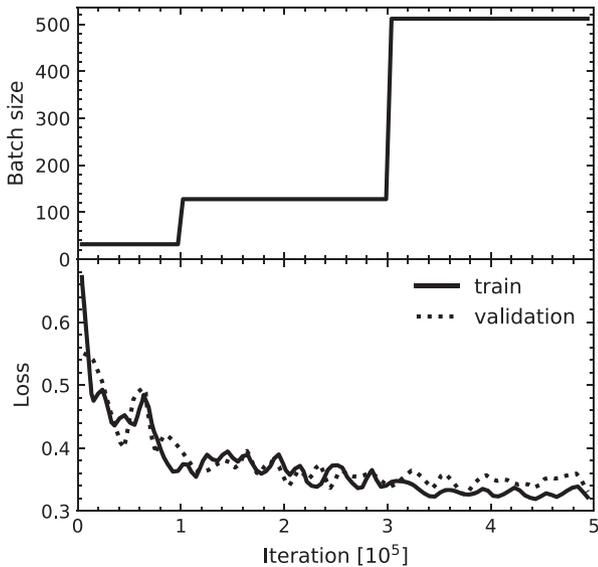
(i) We initialize the parameters following Jia et al. (2014), using the function `variance_scaling_initializer` in TensorFlow (Abadi et al. 2016). The xavier initialization (Glorot & Bengio 2010) also worked well in our test.

(ii) The learning rate is a hyperparameter that controls how much we adjust the trainable parameters based on the loss gradient. Reducing the learning rate helps us to depress the gradient noise, which makes the network tend to converge to a local or global minimum. However, the gradient noise can also be beneficial in

<sup>1</sup>The opposite can also happen, but our sub-box volume choices automatically account for such situations.

**Table 1.** The network architecture. Our network consists of seven convolutional layers and one fully connected layer. Here, the kernel size shows the shape of convolutional kernels in each convolutional layer. The output shape describes the output size of each layer. For convolutional layers, each dimension means [batch size, depth, height, width, channels]. For the fully connected layer, each dimension indicates [batch size, channels]. In the layer before fully connected, we average the output of the conv7 layer in dimensions of performing convolution, which can also be seen as an average pooling layer. All convolutional layers are followed by a ReLU (Nair & Hinton 2010) activation function in our network.

| Layer           | Kernel size | Output shape           | Stride    | Activation function |
|-----------------|-------------|------------------------|-----------|---------------------|
| input           | None        | (None, 39, 39, 39, 1)  | None      | None                |
| conv1           | (3, 3, 3)   | (None, 20, 20, 20, 32) | (2, 2, 2) | ReLU                |
| conv2           | (3, 3, 3)   | (None, 20, 20, 20, 32) | (1, 1, 1) | ReLU                |
| conv3           | (3, 3, 3)   | (None, 10, 10, 10, 64) | (2, 2, 2) | ReLU                |
| conv4           | (3, 3, 3)   | (None, 10, 10, 10, 64) | (1, 1, 1) | ReLU                |
| conv5           | (3, 3, 3)   | (None, 5, 5, 5, 128)   | (2, 2, 2) | ReLU                |
| conv6           | (3, 3, 3)   | (None, 5, 5, 5, 128)   | (1, 1, 1) | ReLU                |
| conv7           | (1, 1, 1)   | (None, 5, 5, 5, 128)   | (1, 1, 1) | ReLU                |
| Mean            | None        | (None, 128)            | None      | None                |
| Fully connected | None        | (None, 1)              | None      | None                |



**Figure 1.** The batch size (upper panel) and loss functions (lower panel). The loss functions are normalized by the variance of initial conditions, cf. equation (9) and are shown in solid and dotted lines for training and validation set, respectively. The loss functions in both the training and the validation sets decrease with training progress, and there is no obvious overfitting.

some cases, such as helping to escape ‘sharp minima’ (Smith & Le 2017). In general, the algorithm calculates the gradient in a mini-batch, i.e. by computing the gradient against more than one training data point, but less than the full data set, at each iteration. As shown in Smith et al. (2017), in the training, decreasing the learning rate is usually equivalent to increasing the minibatch size (hereafter, batch size). Therefore, we fixed the learning rate to 0.0001 and changed the batch size as shown in Fig. 1.

(iii) Batch normalization has become a part of the standard toolkits recently for accelerating and improving the training of deep network by reducing the internal covariate shift (Ioffe & Szegedy 2015). However, since batch normalization uses the mean and variance values in the mini-batch, it is not suitable for small or non-i.i.d. mini-batch training (Ioffe 2017). In our task, on the one hand, we use sub-box density fields (see Section 2.5) when training the network

which are non-i.i.d. data sets. On the other hand, the memory usage of 3D convolutions limits the batch size of our network. Therefore, we remove all batch normalization layers in our network.

(iv) We use the moment-based Adam Optimizer (Kingma & Ba 2014) in this work.

In the training, the loss function is the most important indicator. It can be used to monitor the network’s ability and overfitting. Here, overfitting means the network is trained to work so well on the training set that it works poorly on data it has not seen before. The training set is the data set fed into the network and used to calculate the gradients for updating trainable parameters. In the training, the data that the network ‘has not seen before’ is called the validation set. The validation set will also be fed into the network, but only its loss will be used to indicate overfitting or not. When the network is overfitted, the loss of training set reduces but the validation loss increases instead.

In Fig. 1, we show the batch size and loss function in training. In the top panel, the batch size increases gradually with the training progress to reduce the gradient noise. In the bottom panel, the loss functions of the training set and validation set are represented by black and red solid lines, respectively. The loss function shown here is normalized by the variance of initial conditions, in other words

$$\frac{\text{MSE}}{\sigma^2} = \frac{\sum^k (f(\delta_i^k; \theta) - \delta_i^k)^2}{\sum^k (\bar{\delta}_i - \delta_i^k)^2}, \quad (9)$$

where  $\bar{\delta}_i$  denotes mean of the initial density field. In this case, the loss will be 1 if the network predicts initial density only by its mean value. We find the losses in both the training set and the validation set decrease gradually in the training, and there is no obvious overfitting. Thus, we do not use regularizations such as  $L2$  regularization or dropouts (Srivastava et al. 2014) in our network, aimed to avoid overfitting. We tested adding residual architectures (He et al. 2015) to the network as well, but found no significant improvement; this may be because our network is not very deep. Therefore, we did not use residual layers in our model.

## 2.5 Data set

The data set in this study is based on the INDRA simulations (Falck et al, in preparation), a suite of  $N$ -body simulations (512 runs) evolved from different initial conditions using L-Gadget (Springel 2005),

each with  $1024^3$  dark matter particles in a periodic cube of  $1 h^{-1} \text{Gpc}$  on a side. The cosmological parameters in these simulations are taken to be the best-fitting parameters of WMAP7 (Komatsu et al. 2011):  $\Omega_m = 0.272$ ,  $\Omega_\Lambda = 0.728$ ,  $\Omega_b = 0.045$ ,  $h = 0.704$ ,  $\sigma_8 = 0.81$ , and  $n_s = 0.967$ , where  $\Omega_m$ ,  $\Omega_\Lambda$ , and  $\Omega_b$  are respectively the present-day density parameters for matter, cosmological constant, and baryons;  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$ ,  $\sigma_8$  is the rms matter density fluctuations at  $z = 0$ , and  $n_s$  the index of the primordial power spectrum of the density perturbations.

In total, 24 simulations were used to build the data set, equally split between training, validation, and test sets. We define the snapshot  $z = 10$  as the initial condition and the snapshot  $z = 0$  as the final condition. Here, we choose the initial time arbitrarily: if we instead define the initial time as a redshift higher than 10, we only need to retrain the network with the data at that corresponding redshift.

For each simulation, we assign the dark matter particles on to a  $512^3$  grid using Piecewise Cubic Spline (PCS; see e.g. Chaniotis & Poulidakos 2004), and smooth the initial density field with a  $3 \text{ Mpc } h^{-1}$  Gaussian filter. The network is designed to input  $39^3$  cells and output one value for reducing computational complexity. Here, the input is a cubic region of the final density field, and the output is the predicted density  $\delta_i$  in the central  $(i, j, k) = (20, 20, 20)$  cell of this  $39^3$  cube. Since the side length of each cell is  $1.95 \text{ Mpc } h^{-1}$ , the input is a cubic sub-box with side length  $76 \text{ Mpc } h^{-1}$ .

For the training and validation sets, we separate each simulated final density field into sub-boxes in stride of 16 grids per side. Thus, in both the training and validation sets, we generate 32 768 sub-boxes per simulation and 262 144 sub-boxes in total. When training, we use all sub-boxes in the training set but randomly select 4096 sub-boxes from the validation set to monitor the overfitting.

To further enlarge the training set, we augment each sub-box with six different rotations and eight different axis-reflections to expand the training set by a factor of 48 (Ravanbakhsh et al. 2017). Note that expanding the training set with more simulations is an alternative possibility. However, the sub-box data of each simulation occupies a huge storage space (about 14 GB) in our method, thus expanding the data set by using more simulations would have caused storage pressure; data augmentation is more efficient. Additionally, since overfitting is not an urgent problem as shown in Fig. 1, more simulations are not necessary in our study.

Besides the training set and the validation set, an independent test set is also needed to test the final results of our model, because even though the validation set makes no contribution to the gradient, it is used to choose the hyperparameters. Unlike the training and validation sets, we do not use the test set to measure the loss. In the test results, the network is seen as a complex convolutional kernel and we convolve it on the whole final density field. In this way, we get the corresponding reconstructed density field. We note that all results in Section 3 are calculated from eight simulations in a test set with the same cosmology, but different initial conditions compared to the training and validation sets.

## 3 RESULTS

### 3.1 Density maps and probability density functions

Visualization of density maps can provide intuition of the quality of the reconstruction. In Fig. 2, we show the density maps of the initial condition  $\delta_i$ , final condition  $\delta_f$ , reconstruction  $\delta_r$ , and the residual between reconstruction and the initial condition,  $\text{RES} \equiv \delta_r - \delta_i$ , respectively. To show these density maps clearly, we linearly

extrapolate the corresponding density contrasts  $\delta$  to  $z = 0$  using the linear growth factor  $D_+$  and perform a  $4 \text{ Mpc } h^{-1}$  Gaussian smoothing on all density fields. The projection depth of all slices is  $1.95 \text{ Mpc } h^{-1}$ . As shown in the residual map,  $\delta_r$  is almost identical to the initial density  $\delta_i$ .

To further quantify the quality of the reconstruction, we show the probability density functions (PDFs) in Fig. 3. The dashed, dotted-dashed, and solid lines indicate the PDFs for  $\delta_i$ ,  $\delta_f$ , and  $\delta_r$ , respectively. As in Fig. 2, we linearly extrapolate the corresponding density fields to  $z = 0$  and perform a  $4 \text{ Mpc } h^{-1}$  Gaussian smoothing. We find that, compared with the final condition, the PDF of the reconstruction is much closer to the initial conditions, and have a shape that is closer to Gaussian.

### 3.2 Transfer function

The transfer function can be used to quantify the discrepancy between the power spectra of the initial condition and reconstruction. It is defined as

$$T(k) = \sqrt{P_r(k)/P_i(k)}, \quad (10)$$

where  $P_r(k)$  and  $P_i(k)$  are respectively the power spectra of  $\delta_r$  and  $\delta_i$ .

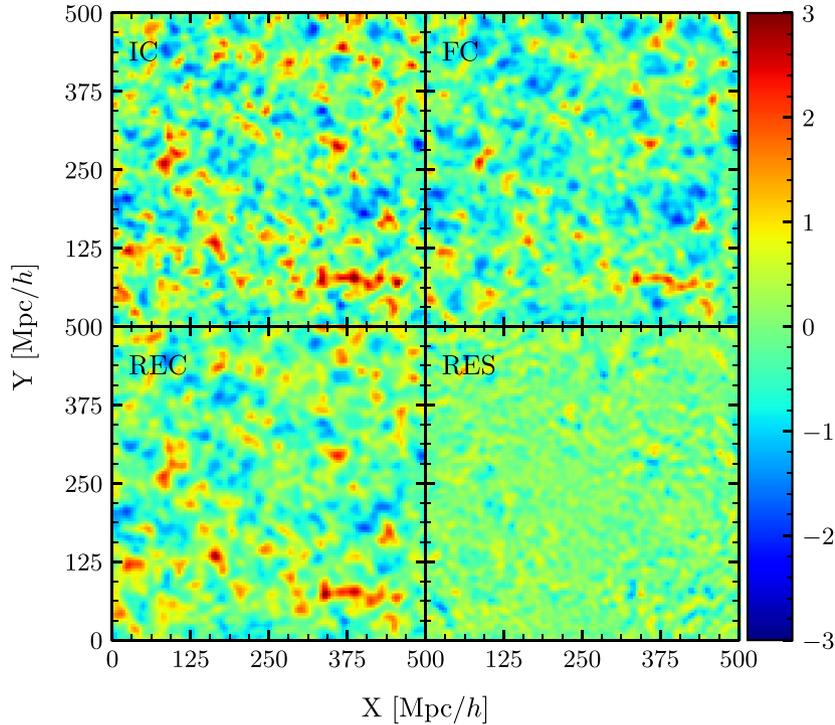
Fig. 4 shows the transfer function averaged over eight simulations. The transfer function decays towards small scales, which is as expected since the complicated small-scaling clustering features are harder to reproduce. We also note that the reconstructed density field is slightly biased on the largest scales, where the transfer function  $T$  is larger than 1. To test if this result is due to sample variance, we have measured the standard deviations over the eight simulations, and found that for all scales except the first  $k$  bin shown in Fig. 4, the scatters are smaller than 5%. Even for the first  $k$  bin which suffers most from cosmic variance, the standard deviation is only  $\sim 1.2$  per cent. Therefore this bias seems indeed to be systematic. A possible reason for this is that our method reconstructs the initial conditions from a small sub-box volume, which lacks larger scale information. Fortunately, these scales on which the bias occurs are not important in BAO reconstruction, since they are nearly unaffected by non-linear evolution and do not need reconstruction. Furthermore, if we can measure this bias in simulations and it turns out to be stable in different cases, we can calibrate the density fluctuation by transfer function, as is widely used in other reconstruction methods (Schmittfull et al. 2017; Seljak et al. 2017; Zhu et al. 2017).

### 3.3 Correlation coefficient

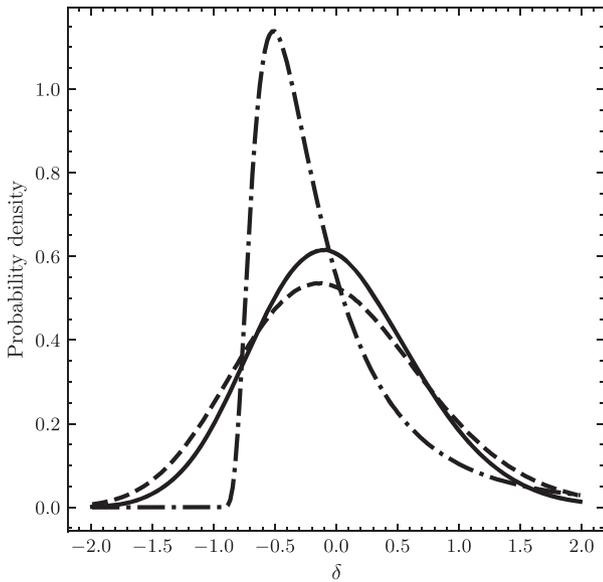
The correlation coefficient between two fields describes their correlation in Fourier phases. It is defined as

$$r(k) = P_{12}(k)/\sqrt{P_1(k)P_2(k)}, \quad (11)$$

where  $P_1$  and  $P_2$  are the auto power spectra of the two fields, and  $P_{12}$  their cross-power spectrum. In Fig. 5, we show the correlation coefficient between the initial condition and reconstruction with the black solid curve, and use the dotted-dashed curve to denote the correlation between the initial and final conditions. We can see that the reconstruction increases  $r$  to  $\sim 0.5$  at  $k = 0.4 h \text{ Mpc}^{-1}$ . The restoration of the information larger than this scale is enough to recover the BAO signal, because the BAO peaks at  $k \gtrsim 0.4 h \text{ Mpc}^{-1}$  are weaker than 1 per cent (see Fig. 6), not currently detectable in observations.

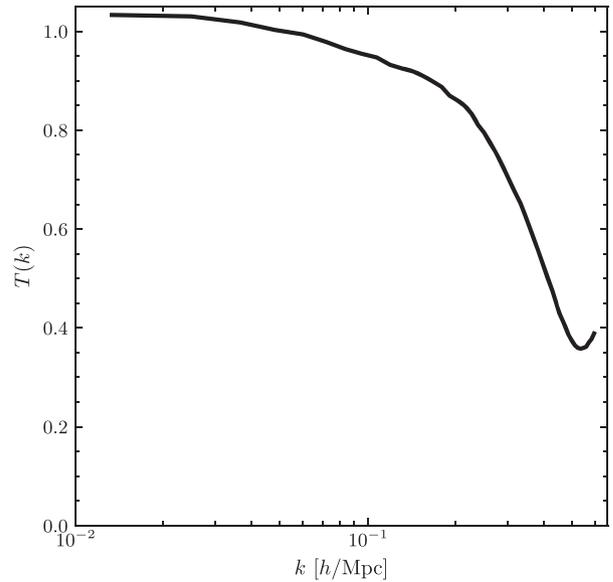


**Figure 2.** Density maps. We show the density map in a  $1.95 \text{ Mpc } h^{-1}$  slice of IC: initial condition  $\delta_i$ , FC: final condition  $\ln(\delta_f + 1)$ , REC: reconstruction  $\delta_r$  and RES: residual between reconstruction and initial condition  $\delta_r - \delta_i$ . For clarity, we have linearly extrapolated the corresponding density contrast  $\delta$  to  $z = 0$  using the linear growth factor, and smoothed all density fields with a Gaussian filter with  $\sigma = 4 \text{ Mpc } h^{-1}$ .



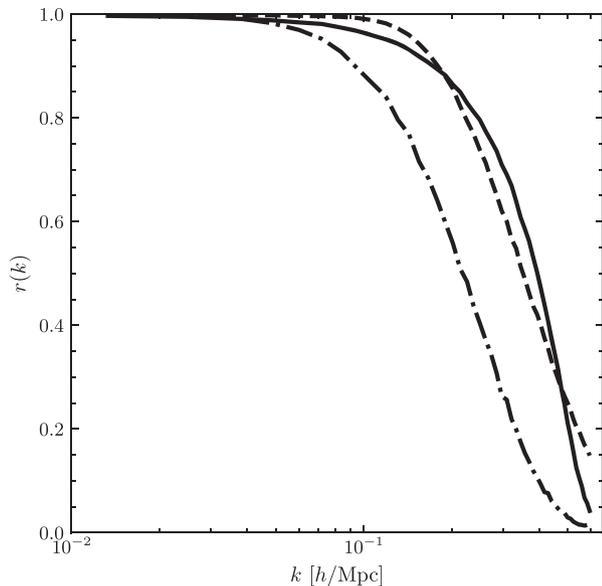
**Figure 3.** Probability density functions. The dashed, dotted-dashed, and solid lines indicate the PDFs of  $\delta_i$ ,  $\delta_f$ , and  $\delta_r$ . We have linearly extrapolated the initial and reconstructed density fields to  $z = 0$  with the linear growth factor, and perform a  $4 \text{ Mpc } h^{-1}$  Gaussian smoothing as in Fig. 2. Unlike  $\delta_f$ , the PDF of  $\delta_r$  is close to  $\delta_i$ .

As comparison, we also show the correlation between the initial conditions and the standard reconstruction with the dashed curve. Here, the standard reconstruction is performed using ‘Nbodykit’ (Hand et al. 2018), with Gaussian smoothing on a scale of

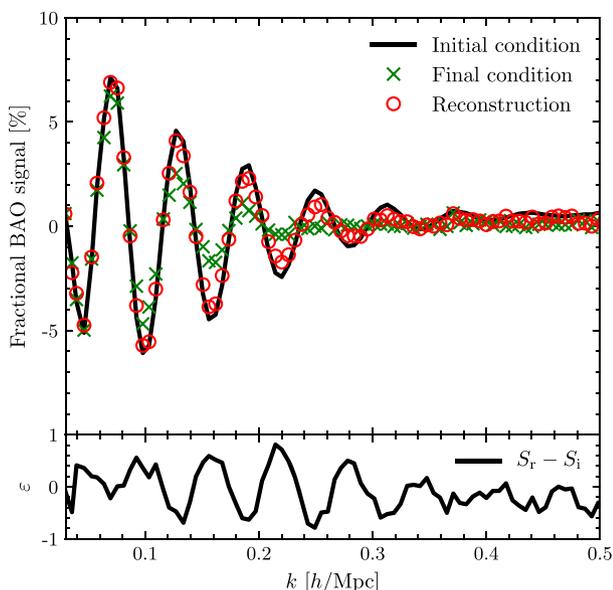


**Figure 4.** The mean transfer function over eight simulations. For most points, the standard deviations are smaller than 5% compared to the transfer function, and so they are not shown in the figure.

$20 \text{ Mpc } h^{-1}$ . On scales larger than  $k = 0.2 \text{ hMpc}^{-1}$ , the standard reconstruction method works slightly better, which again could be because the limited sub-box size in our method removes the clustering information on those large scales. On scales between  $k = 0.2 \text{ hMpc}^{-1}$  and  $0.4 \text{ hMpc}^{-1}$ , where the BAO signal is still strong, our method is better correlated with the initial conditions.



**Figure 5.** The correlation coefficient between the initial condition and the reconstruction (the black solid line). The reconstructed density field is about 90 per cent correlated with the initial density at  $k \simeq 0.2 \text{ hMpc}^{-1}$  and about 50 per cent correlated at  $k \simeq 0.4 \text{ hMpc}^{-1}$ . As a comparison, the correlation between the initial condition and standard reconstruction is shown as the dashed line. Compared with the standard reconstruction, our result is more correlated with the ground truth between  $k \simeq 0.2 \text{ hMpc}^{-1}$  and  $0.5 \text{ hMpc}^{-1}$ . On scales larger than  $0.2 \text{ hMpc}^{-1}$ , the standard reconstruction works better, possibly because of the small sub-box size in our method which means that the large-scale information is not used. The dotted-dashed line shows the correlation between the initial and final conditions.



**Figure 6.** Fractional BAO signals, equation (12), in the power spectrum. To reduce the cosmic variance, we run a pair of simulations from two initial conditions with the same initial random seed but generated by initial power spectra with and without BAO wiggles. *Upper panel:* the black solid line, green cross points and red hollow circles indicate, respectively, the fractional BAO signals in the initial conditions, final conditions and reconstruction. Compared with the final nonlinear case, reconstruction increases the signal-to-noise ratio until  $k \simeq 0.4 \text{ hMpc}^{-1}$ . *Lower panel:* the difference of fractional BAO signals measured from the initial conditions ( $S_i$ ) and the reconstruction ( $S_r$ ).

### 3.4 BAO signal

To directly test the quality of our reconstruction of the BAO signal, we show the fractional BAO signal in Fig. 6. The latter is defined as

$$S = (P_{\text{wiggles}} - P_{\text{nowiggles}}) / P_{\text{nowiggles}}, \quad (12)$$

where the subscripts ‘wiggles’ and ‘nowiggles’ denote simulations evolved from initial power spectra with and without BAO wiggles (Vlah, White & Aviles 2015). These simulation pairs have the same initial random seed, which helps to cancel the cosmic variance in the fractional BAO signal (Schmittfull et al. 2017). In the top panel, the black solid curve is the fractional BAO signal calculated from the initial conditions, in which we can clearly see a series of BAO peaks. However, for the final conditions, shown by the green cross points, the BAO peaks are broadened, which means that the signal-to-noise ratio of the peaks decreases because of non-linear evolution, especially on scales  $0.2\text{--}0.4 \text{ hMpc}^{-1}$ . After reconstruction, shown by the red hollow circles, the signal-to-noise ratio of BAO peaks is improved, until about  $k = 0.4 \text{ hMpc}^{-1}$ . We also show the difference between initial conditions and our reconstruction in the bottom panel. On all scales, we find the differences of fractional BAO signals measured from the initial condition and the reconstruction are smaller than 1 per cent. This indicates our reconstruction succeeds in removing the effect of non-linear evolution, and recovering the BAO wiggles in the initial conditions. As Fig. 6 shows, we can recover the peak around  $k = 0.32 \text{ hMpc}^{-1}$  and partly the peak around  $k = 0.38 \text{ hMpc}^{-1}$ .

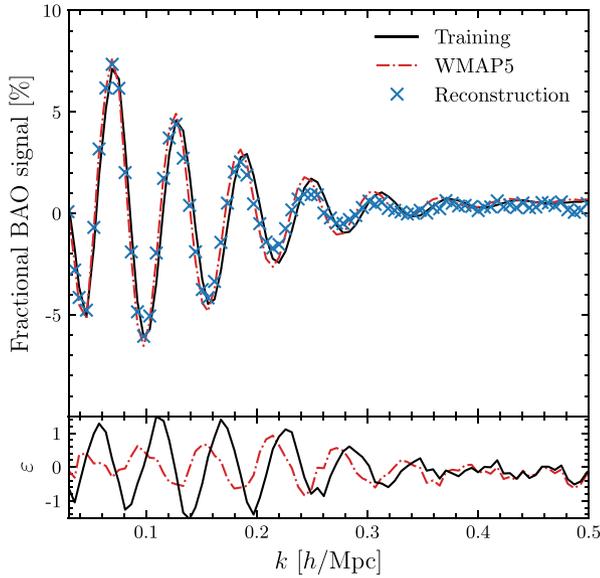
## 4 DISCUSSION

### 4.1 Cosmology dependence

In this study, we train our network using simulations with specific cosmological parameters, which introduces a cosmological dependence into our model. In this subsection, we check this dependence in more detail.

To check the cosmology dependence, we run another two pairs of simulations with a  $500 \text{ Mpc } h^{-1}$  box size from initial conditions with and without BAO wiggles like in Fig. 6. These simulations use different cosmological parameters from the training set: in the training set, the INDRA simulation use the best-fitting cosmological parameters of  $WMAP + BAO + H_0$  based on the 7-yr *WMAP* results, while in these new simulations, we choose the best-fitting parameters based on the *WMAP*-only data in *WMAP5* (Hinshaw et al. 2009) and *WMAP9* (Bennett et al. 2013), to increase the difference from the training sample. For clarity, in the discussion below about cosmology dependence, we call the universe of the training set the ‘training cosmology’ and use ‘truth cosmology’ to indicate both new simulations of *WMAP5* and *WMAP9* in the corresponding comparisons. In the following comparisons, we train our model in the training cosmology, but use it to reconstruct the BAO signal in the two truth cosmologies.

The results from the *WMAP5* cosmology are shown in Fig. 7. In the upper panel, the black solid and red dotted-dashed lines indicate the fractional BAO signals of the initial conditions in the training and truth cosmologies, respectively. The reconstruction in the truth cosmology is shown by the blue cross points. We find our reconstruction is closer to the truth cosmology than to the training cosmology on almost all scales. This indicates that, instead of ‘remembering’ the BAO signal of the training set, the network model has indeed successfully learned the relation  $\delta_i = f(\delta_i; \theta)$ , which is model-independent and allows it to reconstruct the initial condition for general cosmologies. In the lower panel, the black



**Figure 7.** Cosmology dependence. The network is trained in the training (*WMAP7*) cosmology, while the reconstruction is applied in the truth (*WMAP5*) cosmology. In the upper panel, the black solid and red dot-dashed lines indicate the fractional BAO signals of the initial conditions in the training and truth cosmologies, respectively. The cyan crosses show the reconstruction in the truth cosmology. Although the network is trained by simulations in the training cosmology, we find the reconstructed signal is closer to the truth cosmology than the signal in the training cosmology. In the lower panel, the black solid and red dotted-dashed lines show the differences between the reconstruction and initial conditions in training and truth cosmologies, respectively.

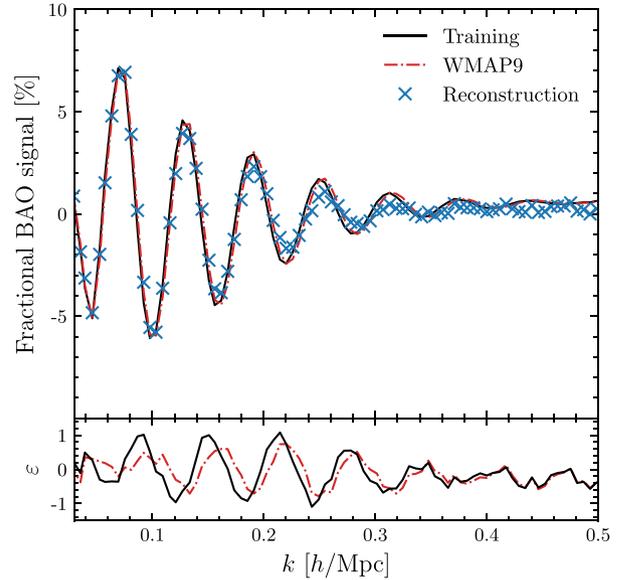
solid and red dotted-dashed lines show the differences between the reconstruction and the initial condition in the training and truth cosmologies, respectively. It is interesting to note that the differences are actually smaller for the truth cosmology than for the cosmology upon which the network has been trained.

Fig. 8 is similar to Fig. 7, but shows a second test of cosmology dependence using *WMAP9* as the truth cosmology. Since the cosmological parameters of the training and truth cosmologies are very close in this case, there is only a slight discrepancy of the BAO signals in those two cosmologies. We again find that the reconstruction is closer to the truth cosmology, although there are only very small discrepancy between the two, especially on scales between 0.1 and  $0.3 h\text{Mpc}^{-1}$ .

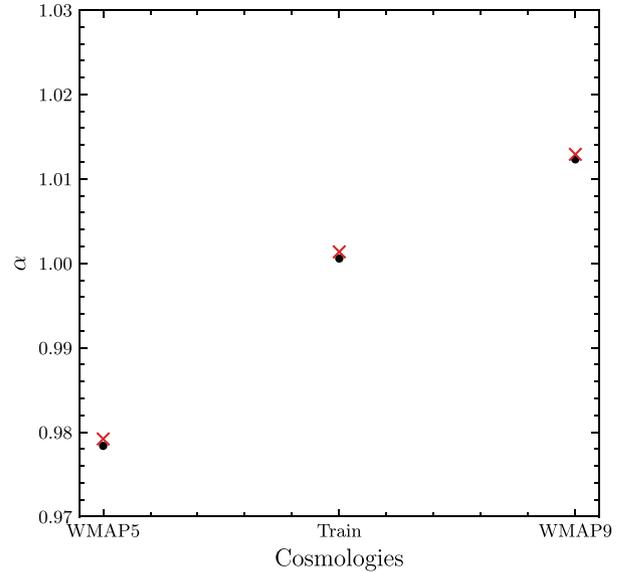
To quantify whether this method can be used to distinguish between different cosmologies, we define a scale dilation parameter  $\alpha$  that is used to adjust the location of the BAO peaks, as

$$S(k) = S_i(k/\alpha), \quad (13)$$

where the  $S$  is the fractional BAO signal defined in equation (12) and  $S_i$  indicates the initial fractional BAO signal in the training cosmology. Since the definition of the fractional BAO signal has removed most of the cosmic variance and the reconstruction has removed non-linear damping, the parameter  $\alpha$  will show how much the peak location of  $S$  is shifted with respect to  $S_i$ . We fit  $\alpha$  with  $S$  being the fractional BAO signal of the reconstruction and initial condition respectively, for the training, *WMAP5* and *WMAP9* cosmologies. In Fig. 9, the black points indicate the  $\alpha$  fitted from the initial conditions and the red cross points are fitted from the reconstruction. Note that  $\alpha = 1$  by definition for the initial condition of the training cosmology (the second black



**Figure 8.** The same as Fig. 7, but with the *WMAP9* cosmology as the truth cosmology. Since the training and truth cosmologies are so close, there are only very small differences between the BAO signals of the training, truth, and reconstruction at  $k \lesssim 0.2 h\text{Mpc}^{-1}$ , although the reconstruction does agree better with the truth cosmology as expected.



**Figure 9.** The scale dilation parameter  $\alpha$ , cf. equation (13), which quantifies the shift of the BAO peak position in a given density field with respect to the peak positions in the initial condition of the training cosmology. The  $x$ -axis shows the three different cosmologies used in our tests, and the black points and red crosses indicate the best-fitting  $\alpha$  values from the initial condition and reconstruction, respectively. The difference between the BAO peak positions in the initial condition and the reconstruction for both truth cosmologies (*WMAP5* and *WMAP9*) is much smaller than the difference between its locations in the different cosmologies. This is consistent with the results shown in Figs 7 and 8.

dot). There is a (very) slight shift of the reconstructed BAO peaks even in the training cosmology ( $\alpha > 1$  for the second red cross), and this shift seems to be the same for the reconstruction results of the two truth cosmologies, indicating again that the trained

network has negligible cosmology dependence.<sup>2</sup> More importantly, this shift is much smaller than the difference in  $\alpha$  that is due to the underlying cosmology (which is at the per cent level), suggesting that different models can be distinguished between our reconstruction method.

It should be noted that, to overcome the cosmology dependence completely, we should train the network with simulations in a series of cosmological parameters as done by Ravanbakhsh et al. (2017). A more detailed analysis of this is beyond the scope of this paper and will be left for future work.

## 4.2 Boundary effects

The discussion so far has been in the idealized case of dark matter, with no boundary effects. BAO measurement in galaxy survey data faces other complications, such as redshift-space distortions, shot noise, galaxy bias, and survey boundaries. Most perturbation based reconstruction algorithms estimate the displacement field in Fourier space. In this case, the irregular survey boundary and incomplete information near it disturbs the reconstruction results as far as 100 Mpc  $h^{-1}$  from the boundary (Zhu et al. 2020).

In order to deal with the irregular survey boundary, one can build a forward model that includes the boundary, and optimize them together (see e.g. Seljak et al. 2017; Feng et al. 2018; Modi et al. 2018). If we only focus on the effects of survey boundary, a forward-modelling method is a suitable choice because the boundary is fixed and we know it very well. However, limited by computing power and memory, the forward model usually cannot be too complicated. In practice, a limited forward model may be hard to generalize, and there is a trade-off between variance and bias, e.g. much effort is required to understand the bias of the forward model and make it close to the actual problem. The survey boundary is easy to handle in forward modelling, though not all observational effects are.

Since our network model estimates the initial conditions from a sub-box instead of the global density field, we expect that survey boundaries should have a small impact in this method. In this subsection, we check such a boundary dependence of our reconstruction (hereafter, boundary reconstruction) by assuming a survey boundary like the Apple logo, which provides a couple of separated irregular regions. For comparison, we use ‘full reconstruction’ to indicate the reconstruction that use all information in the box (i.e. with periodic boundaries).

Fig. 10 shows the density maps of (a) the full reconstruction, (b) the boundary reconstruction, and (c) the residual between them. All these maps are shown in a 1.95 Mpc  $h^{-1}$  slice. For the boundary reconstruction, we fill all cells outside the assumed survey boundary with the cosmic mean density during reconstruction. For clarity, we define a boundary distance  $d_b$  which describes the nearest distance to the survey boundary of each cell, which will be helpful for quantifying the impact of survey boundary in reconstruction. In our method, the area not affected by the survey boundary (hereafter, critical boundary) is circled by the black dashed line in the full and boundary reconstruction maps. The  $d_b$  of critical boundary ranges from 37 to 52.3 Mpc  $h^{-1}$ , depending on the position in the map, because of the sub-box length used in reconstruction. In the residual panel, from outside to inside, the dashed lines indicate  $d_b$  as 0, 26.2, 52.3, and 78.5 Mpc  $h^{-1}$ , respectively. Only a slight discrepancy between full

and boundary reconstruction maps is found in the area  $0 < d_b < 26.2$  Mpc  $h^{-1}$ .

In Fig. 11, we show the ratio of correlation coefficients between boundary and full reconstruction in areas of  $0 < d_b < 26.2$  Mpc  $h^{-1}$ ,  $26.2 < d_b < 52.3$  Mpc  $h^{-1}$ , and  $52.3 < d_b < 78.5$  Mpc  $h^{-1}$ , respectively. Note that to calculate the correlation coefficients in these three areas, we mask out the rest of the field. The solid curves indicate the results of our method, and the dashed curves are for the standard reconstruction method. To implement boundary reconstruction in the standard method, we replace the matter distribution beyond the survey boundary with a random catalogue. As  $d_b$  increases, in both methods, the correlation coefficients of the boundary reconstruction are less and less affected by the survey boundary. In the upper panel, where the boundary distance is smaller than 26.2 Mpc  $h^{-1}$ , the correlation coefficient ratios in both methods decline quickly, reaching 90 per cent at  $k \simeq 0.2$   $h$ Mpc $^{-1}$ . In the middle panel, where  $d_b$  is close to the critical boundary, the difference between boundary and full reconstructions in our method is significantly reduced down to  $k \simeq 0.46$   $h$ Mpc $^{-1}$ . When we calculate the correlation coefficient in region of  $52.3 < d_b < 78.5$  Mpc  $h^{-1}$ , the ratio between boundary and full reconstruction of our method is equal to one because this region is located entirely inside the critical boundary (lower panel).

While boundary reconstruction based on the standard reconstruction method works quite well, especially for  $d_b < 26.2$  Mpc  $h^{-1}$ , our method still offers better consistency between boundary and full reconstructions in the region of  $26.2 < d_b < 78.5$  Mpc  $h^{-1}$ . Thus, at a small price of a degraded performance (compared with standard reconstruction) in the outermost layer near the survey boundary – where both methods perform rather poorly anyway – our method leads to improved reconstruction results further away from the survey edge, where boundary effects are present in the standard method.

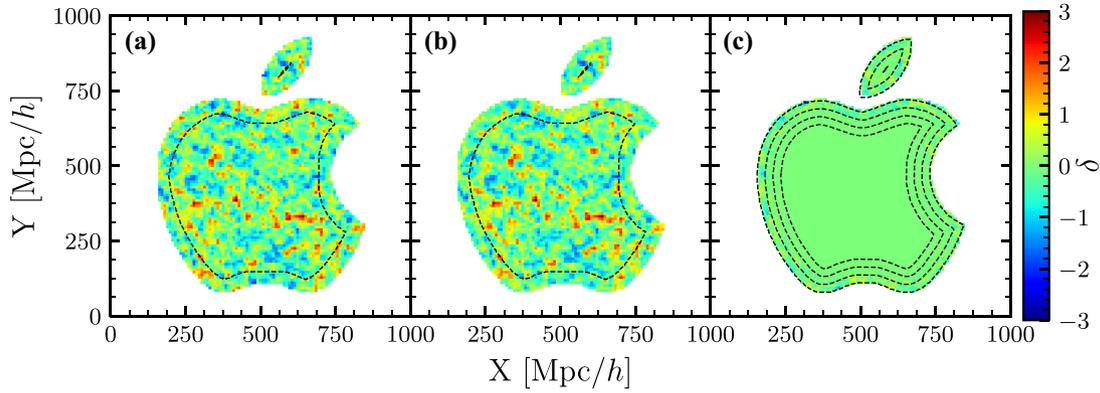
## 5 CONCLUSION

We present a new method of BAO reconstruction based on deep convolutional neural networks, and report its first results when applied to simulated dark matter density fields. The objective of reconstruction is to undo the bulk motions of matter, which could dampen and broaden the BAO peaks that are present in the primordial matter density field. Therefore, an indicator of its performance is the phase correlation coefficient between the initial and reconstructed density fields. Our method can lead to a correlation coefficient of about 90 per cent at  $k = 0.2$   $h$ Mpc $^{-1}$  and 50 per cent at  $k = 0.4$   $h$ Mpc $^{-1}$ . For the fractional BAO signal, the reconstruction can improve the signal-to-noise ratio down to  $k = 0.4$   $h$ Mpc $^{-1}$ , extending the range of scales at which the power spectrum matches linear theory by a factor about 2 compared to final condition.

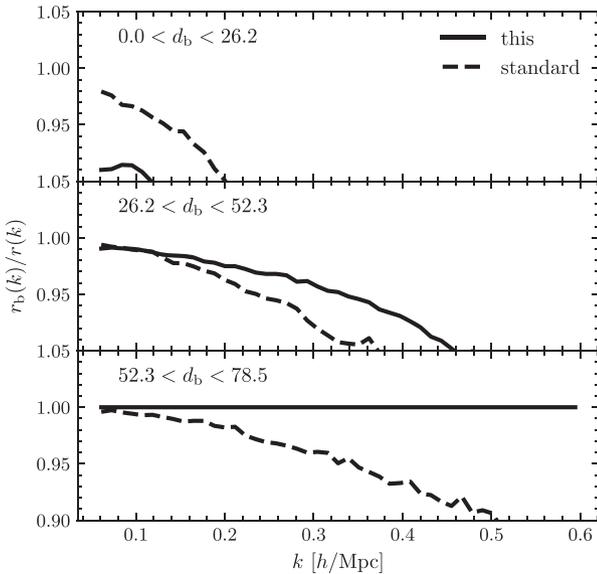
Since the network is trained by simulations with specific cosmological parameters, we have checked the cosmology dependence by applying the trained model to two different cosmologies. We do not find evidence for cosmology dependence, and the method seems insensitive to the training cosmology. We also demonstrate that it can distinguish the different cosmologies considered. However, we caution that the cosmologies on which the method is tested are both rather close to the training cosmology. In future, we will further this analysis by using a wider parameter range in the training and test sets.

Because its input data is the non-linear density field in cubic sub-boxes (which we have chosen to have a side length of 76 Mpc  $h^{-1}$ ), this new method is by design robust against boundary effects since areas inside the critical boundary ( $37 \sim 52.3$  Mpc  $h^{-1}$  from the survey edge) are not affected at all. This is an advantage over the standard reconstruction method, because survey boundaries can

<sup>2</sup>The shift itself could be a sign that reconstruction cannot fully remove the broadening of the BAO peaks caused by non-linear structure formation.



**Figure 10.** (a) The density map of the full reconstruction. (b) The density map of the boundary reconstruction. (c) The residual between the full and boundary reconstructions. In all panels, only the areas located in the assumed survey boundary are shown. In Panels (a) and (b), the black dashed lines indicate the critical boundary, while in Panel (c), from outside to inside, the dashed lines indicate the positions at which  $d_b$  is 0, 26.2, 52.3, and 78.5  $\text{Mpc } h^{-1}$ , respectively. With  $d_b$  decreasing, the residual becomes increasingly stronger due to the missed information outside the boundary.



**Figure 11.** The ratio between the correlation coefficients from the boundary and full reconstructions. The solid and dashed lines indicate our method and the standard reconstruction method, respectively. From top to bottom, different ranges of the boundary distance  $d_b$  are used to define the region that is used to calculate the correlation coefficient, as indicated in the legends. These correspond to the three regions between neighbouring dashed lines in the right-hand panel of Fig. 10.

substantially impact BAO reconstruction in galaxy surveys. Our tests show that, compared with the standard reconstruction method, the new method improves the consistency between boundary and full reconstructions in the region of  $26.2 < d_b < 78.5 \text{ Mpc } h^{-1}$  from the survey edge.

In this paper, we have tested our new scheme in dark-matter-only simulations, and found that it can accurately remove non-linear effects on scales larger than  $k = 0.4 \text{ hMpc}^{-1}$ , and enable us to recover the BAO wiggles up to  $k = 0.32 \text{ hMpc}^{-1}$ . In the future, we will test this method by applying it to density field reconstruct using galaxy surveys (Wang et al. 2009). We expect that using all available galaxies in a survey can put optimally tight BAO constraints on the underlying cosmological model.

## ACKNOWLEDGEMENTS

We thank Houjun Mo and Huiyuan Wang for very useful discussions. We also thank the anonymous reviewer for his/her many suggestions for improving this paper. JW acknowledges the support from the National Natural Science Foundation of China (NSFC) grant 11873051. This work is supported by Astronomical Big Data Joint Research Center, co-founded by National Astronomical Observatories, Chinese Academy of Sciences and Alibaba Cloud.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Abadi M. et al., 2016, preprint (arXiv:1603.04467)  
 Alam S. et al., 2017, *MNRAS*, 470, 2617  
 Anderson L. et al., 2014, *MNRAS*, 441, 24  
 Angulo R. E., Baugh C. M., Frenk C. S., Bower R. G., Jenkins A., Morris S. L., 2005, *MNRAS*, 362, L25  
 Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, *MNRAS*, 383, 755  
 Aragon-Calvo M. A., 2019, *MNRAS*, 484, 5771  
 Bandura K. et al., 2014, in Stepp L. M., Gilmozzi R., Hall H. J., eds, Proc. SPIE Conf. Ser. Vol. 9145, Ground-based and Airborne Telescopes V. SPIE, Bellingham, p. 914522  
 Battye R. et al., 2016, preprint (arXiv:1610.06826)  
 Bennett C. L. et al., 2013, *ApJS*, 208, 20  
 Berger P., Stein G., 2019, *MNRAS*, 482, 2861  
 Beutler F., Blake C., Koda J., Marín F. A., Seo H.-J., Cuesta A. J., Schneider D. P., 2016, *MNRAS*, 455, 3230  
 Beutler F. et al., 2017, *MNRAS*, 464, 3409  
 Birkin J., Li B., Cautun M., Shi Y., 2019, *MNRAS*, 483, 5267  
 Bos E. G. P., Kitaura F.-S., van de Weygaert R., 2019, *MNRAS*, 488, 2573  
 Bottou L., 1998, in Saad D., ed., Online Learning and Neural Networks, Online Algorithms and Stochastic Approximations, Cambridge University Press, Cambridge, UK, available at: <http://leon.bottou.org/papers/bottou-98x>  
 Burden A., Percival W. J., Howlett C., 2015, *MNRAS*, 453, 456  
 Caldeira J., Wu W. L. K., Nord B., Avestruz C., Trivedi S., Story K. T., 2019, *Astron. Comput.*, 28, 100307  
 Chaniotis A. K., Poulidakos D., 2004, *J. Comput. Phys.*, 197, 253

- Croce M., Scoccimarro R., 2006, *Phys. Rev. D*, 73, 063520
- Dark Energy Survey Collaboration, 2016, *MNRAS*, 460, 1270
- de Jong R. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, *Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV*. SPIE, Bellingham, p. 84460T
- DESI Collaboration, 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
- Eisenstein D. J., Seo H.-J., White M., 2007a, *ApJ*, 664, 660
- Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007b, *ApJ*, 664, 675
- Feng Y., Seljak U., Zaldarriaga M., 2018, *J. Cosmol. Astropart. Phys.*, 7, 043
- Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2019, *MNRAS*, 484, 282
- Glorot X., Bengio Y., 2010, in Teh Y. W., Titterton M., eds, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Understanding the difficulty of training deep feedforward neural networks, JMLR Workshop and Conference Proceedings, Chia Laguna Resort, Sardinia, Italy*, p. 249
- Godfrey L. E. H. et al., 2012, *Publ. Astron. Soc. Aust.*, 29, 42
- Goodfellow I., Bengio Y., Courville A., 2016, *Deep Learning*. MIT Press, Cambridge, MA
- Gupta A., Matilla J. M. Z., Hsu D., Haiman Z., 2018, *Phys. Rev. D*, 97, 103515
- Hada R., Eisenstein D. J., 2018, *MNRAS*, 478, 1866
- Hada R., Eisenstein D. J., 2019, *MNRAS*, 482, 5685
- Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, *AJ*, 156, 160
- He K., Zhang X., Ren S., Sun J., 2015, preprint ([arXiv:1512.03385](https://arxiv.org/abs/1512.03385))
- He S., Li Y., Feng Y., Ho S., Ravanbakhsh S., Chen W., Póczos B., 2019, *Proc. Natl. Acad. Sci.*, 116, 13825
- Hinshaw G. et al., 2009, *ApJS*, 180, 225
- Hinton S. R. et al., 2017, *MNRAS*, 464, 4807
- Huff E., Schulz A. E., White M., Schlegel D. J., Warren M. S., 2007, *Astropart. Phys.*, 26, 351
- Ioffe S., 2017, *Advances in Neural Information Processing Systems*. Curran Associates, Inc., Red Hook, NY, USA, p. 1945
- Ioffe S., Szegedy C., 2015, preprint ([arXiv:1502.03167](https://arxiv.org/abs/1502.03167))
- Ivezić Ž et al., 2008, preprint ([arXiv:0805.2366](https://arxiv.org/abs/0805.2366))
- Jasche J., Wandelt B. D., 2013, *MNRAS*, 432, 894
- Jeong D., Komatsu E., 2006, *ApJ*, 651, 619
- Jia Y., Shelhamer E., Donahue J., Karayev S., Long J., Girshick R., Guadarrama S., Darrell T., 2014, preprint ([arXiv:1408.5093](https://arxiv.org/abs/1408.5093))
- Jones D. H. et al., 2009, *MNRAS*, 399, 683
- Kazin E. A. et al., 2014, *MNRAS*, 441, 3524
- Kingma D. P., Ba J., 2014, preprint ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Kitaura F.-S., 2013, *MNRAS*, 429, L84
- Kitaura F.-S., Ata M., Rodriguez-Torres S. A., Hernandez-Sanchez M., Balaguera-Antolínez A., Yepes G., 2019, preprint ([arXiv:1911.00284](https://arxiv.org/abs/1911.00284))
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds, *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Red Hook, NY, US, p. 1097
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Leclercq F., Enzi W., Jasche J., Heavens A., 2019, *MNRAS*, 490, 4237
- LeCun Y., Boser B. E., Denker J. S., Henderson D., Howard R. E., Hubbard W. E., Jackel L. D., 1990, in Touretzky D., ed., *Advances in neural information processing systems*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 396
- Lecun Y., Bengio Y., Hinton G., 2015, *Nature*, 521, 436
- Li R., Shu Y., Su J., Feng H., Zhang G., Wang J., Liu H., 2019, *MNRAS*, 482, 313
- Lucie-Smith L., Peiris H. V., Pontzen A., Lochner M., 2018, *MNRAS*, 479, 3405
- Lucie-Smith L., Peiris H. V., Pontzen A., 2019, *MNRAS*, 490, 331
- Mathuriya A. et al., 2018, preprint ([arXiv:1808.04728](https://arxiv.org/abs/1808.04728))
- Mehta K. T., Seo H.-J., Eckel J., Eisenstein D. J., Metchnik M., Pinto P., Xu X., 2011, *ApJ*, 734, 94
- Meiksin A., White M., Peacock J. A., 1999, *MNRAS*, 304, 851
- Modi C., Feng Y., Seljak U., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 028
- Morningstar W. R., Hezaveh Y. D., Perreault Levasseur L., Blandford R. D., Marshall P. J., Putzky P., Wechsler R. H., 2018, preprint ([arXiv:1808.00011](https://arxiv.org/abs/1808.00011))
- Morningstar W. R. et al., 2019, *ApJ*, 883, 14
- Nair V., Hinton G. E., 2010, *Proceedings of the 27th international conference on machine learning (ICML-10)*, Omnipress, Madison, WI, USA, p. 807
- Newburgh L. B. et al., 2016, in Hall H. J., Gilmozzi R., Marshall H. K., eds, *Proc. SPIE Conf. Ser. Vol. 9906, Ground-based and Airborne Telescopes VI*. SPIE, Bellingham, p. 99065X
- Noh Y., White M., Padmanabhan N., 2009, *Phys. Rev. D*, 80, 123501
- Obuljen A., Villaescusa-Navarro F., Castorina E., Viel M., 2017, *J. Cosmol. Astropart. Phys.*, 9, 012
- Padmanabhan N., White M., 2009, *Phys. Rev. D*, 80, 063508
- Padmanabhan N., White M., Cohn J. D., 2009, *Phys. Rev. D*, 79, 063523
- Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, *MNRAS*, 427, 2132
- Pan Q., Pen U.-L., Inman D., Yu H.-R., 2017, *MNRAS*, 469, 1968
- Peebles P. J. E., Yu J. T., 1970, *ApJ*, 162, 815
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Rafieferantsoa M., Andrianomena S., Davé R., 2018, *MNRAS*, 479, 4509
- Ravanbakhsh S., Oliva J., Fromenteau S., Price L. C., Ho S., Schneider J., Poczos B., 2017, preprint ([arXiv:1711.02033](https://arxiv.org/abs/1711.02033))
- Rodríguez A. C., Kacprzak T., Lucchi A., Amara A., Sgier R., Fluri J., Hofmann T., Réfrégier A., 2018, *Comput. Astrophys. Cosmol.*, 5, 4
- Ross A. J., Samushia L., Howlett C., Percival W. J., Burden A., Manera M., 2015, *MNRAS*, 449, 835
- Sarpa E., Schimd C., Branchini E., Matarrese S., 2019, *MNRAS*, 484, 3818
- Schmelzle J., Lucchi A., Kacprzak T., Amara A., Sgier R., Réfrégier A., Hofmann T., 2017, preprint ([arXiv:1707.05167](https://arxiv.org/abs/1707.05167))
- Schmittfull M., Feng Y., Beutler F., Sherwin B., Chu M. Y., 2015, *Phys. Rev. D*, 92, 123522
- Schmittfull M., Baldauf T., Zaldarriaga M., 2017, *Phys. Rev. D*, 96, 023505
- Seljak U., Aslanyan G., Feng Y., Modi C., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 009
- Seo H.-J., Eisenstein D. J., 2005, *ApJ*, 633, 575
- Seo H.-J., Siegel E. R., Eisenstein D. J., White M., 2008, *ApJ*, 686, 13
- Seo H.-J. et al., 2010, *ApJ*, 720, 1650
- Sherwin B. D., Zaldarriaga M., 2012, *Phys. Rev. D*, 85, 103523
- Shi Y., Cautun M., Li B., 2018, *Phys. Rev. D*, 97, 023505
- Shimabukuro H., Semelin B., 2017, *MNRAS*, 468, 3869
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Smith S. L., Le Q. V., 2017, preprint ([arXiv:1710.06451](https://arxiv.org/abs/1710.06451))
- Smith S. L., Kindermans P.-J., Ying C., Le Q. V., 2017, preprint ([arXiv:1711.00489](https://arxiv.org/abs/1711.00489))
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V. et al., 2005, *Nature*, 435, 629
- Springer O. M., Ofek E. O., Weiss Y., Merten J., 2020, *MNRAS*, 491, 5301
- Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, *J. Mach. Learn. Res.*, 15, 1929
- Sunyaev R. A., Zeldovich Y. B., 1970, *Ap&SS*, 7, 3
- Takada M. et al., 2014, *PASJ*, 66, R1
- Tewes M., Kuntzer T., Nakajima R., Courbin F., Hildebrandt H., Schrabback T., 2019, *A&A*, 621, A36
- Villanueva-Domingo P., Villaescusa-Navarro F., 2020, preprint ([arXiv:2006.14305](https://arxiv.org/abs/2006.14305))
- Vlah Z., White M., Aviles A., 2015, *J. Cosmol. Astropart. Phys.*, 9, 014
- Wang X., Pen U.-L., 2019, *ApJ*, 870, 116
- Wang H., Mo H. J., Jing Y. P., Guo Y., van den Bosch F. C., Yang X., 2009, *MNRAS*, 394, 398
- Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, *ApJ*, 772, 63
- Wang X., Yu H.-R., Zhu H.-M., Yu Y., Pan Q., Pen U.-L., 2017, *ApJ*, 841, L29
- Wang Y., Li B., Cautun M., 2020, *MNRAS*, 497, 3451

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87  
White M., 2015, *MNRAS*, 450, 3822  
Xu X., Cuesta A. J., Padmanabhan N., Eisenstein D. J., McBride C. K., 2013, *MNRAS*, 431, 2834  
Xu Y., Wang X., Chen X., 2015, *ApJ*, 798, 40  
Yu Y., Zhu H.-M., Pen U.-L., 2017, *ApJ*, 847, 110  
Zhu H.-M., Pen U.-L., Chen X., 2016, preprint ([arXiv:1609.07041](https://arxiv.org/abs/1609.07041))

Zhu H.-M., Yu Y., Pen U.-L., Chen X., Yu H.-R., 2017, *Phys. Rev. D*, 96, 123502  
Zhu H.-M., Yu Y., Pen U.-L., 2018, *Phys. Rev. D*, 97, 043502  
Zhu H.-M., White M., Ferraro S., Schaan E., 2020, *MNRAS*, 494, 4244

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