PkANN – I. Non-linear matter power spectrum interpolation through artificial neural networks

Shankar Agarwal[1,†], Filipe B. Abdalla[2,§], Hume A. Feldman[1,‡], Ofer Lahav[2,♭] & Shaun A. Thomas[2,⋆]

[1]*Department of Physics & Astronomy, University of Kansas, Lawrence, KS 66045, USA.*
[2]*Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK.*
emails: [†]*sagarwal@ku.edu,* [§]*fba@star.ucl.ac.uk,* [‡]*feldman@ku.edu,* [♭]*lahav@star.ucl.ac.uk,* [⋆]*sat@star.ucl.ac.uk*

ABSTRACT
We investigate the interpolation of power spectra of matter fluctuations using Artificial Neural Network (PkANN). We present a new approach to confront small-scale non-linearities in the power spectrum of matter fluctuations. This ever-present and pernicious uncertainty is often the Achilles' heel in cosmological studies and must be reduced if we are to see the advent of precision cosmology in the late-time Universe. We show that an optimally trained artificial neural network (ANN), when presented with a set of cosmological parameters ($\Omega_{\mathrm{m}} h^2, \Omega_{\mathrm{b}} h^2, n_s, w_0, \sigma_8, \sum m_\nu$ and redshift z), can provide a worst-case error ≤ 1 per cent (for $z \leq 2$) fit to the non-linear matter power spectrum deduced through N-body simulations, for modes up to $k \leq 0.7\,h\mathrm{Mpc}^{-1}$. Our power spectrum interpolator is accurate over the *entire* parameter space. This is a significant improvement over some of the current matter power spectrum calculators. In this paper, we detail how an accurate interpolation of the matter power spectrum is achievable with only a sparsely sampled grid of cosmological parameters. Unlike large-scale N-body simulations which are computationally expensive and/or infeasible, a well-trained ANN can be an extremely quick and reliable tool in interpreting cosmological observations and parameter estimation. This paper is the first in a series. In this method paper, we generate the non-linear matter power spectra using HALOFIT and use them as mock observations to train the ANN. This work sets the foundation for Paper II, where a suite of N-body simulations will be used to compute the non-linear matter power spectra at sub-per cent accuracy, in the quasi-non-linear regime ($0.1\,h\mathrm{Mpc}^{-1} \leq k \leq 0.9\,h\mathrm{Mpc}^{-1}$). A trained ANN based on this N-body suite will be released for the scientific community.

1 INTRODUCTION

Studying the growth of structure and the distribution of galaxies in our Universe is a potent method for understanding both fundamental physics and the cosmological model. Measurements of the large scale structure are capable of testing our theory of gravity (Amendola, Kunz, & Sapone 2008), distinguishing between dark energy models and probing the absolute neutrino mass scale (Thomas, Abdalla, & Lahav 2010). It is also an immense complement to the incrementally cosmic variance limited cosmic microwave background. This complementarity explains the overwhelmingly large number of galaxy surveys on the horizon that promise to refine or even alter our understanding of the cosmos, *e.g.* DES (The Dark Energy Survey Collaboration 2005), the Large Synoptic Survey Telescope (LSST) (Ivezic et al. 2008),

and the Baryon Oscillation Spectroscopic Survey (BOSS) (Eisenstein et al. 2011). These surveys promise to achieve high-precision measurements of galaxy power spectrum amplitudes and offer a possibility to improve constraints on cosmological parameters including dark energy and neutrino masses. However, with this promise comes a great technical and systematic difficulty.

Arguably the most ubiquitous problem in both galaxy clustering and weak lensing surveys is that as structures collapse they evolve from being linear, for which one can solve analytically, to non-linear, for which one cannot. Using N-body simulations (Heitmann et al. 2010; Agarwal & Feldman 2011) and analytical studies inspired from perturbation theory (PT) (Scoccimarro, Zaldarriaga, & Hui 1999; Saito, Takada, & Taruya 2008), the non-linear effects have been shown to be significant compared to the precision levels of future surveys. A consequence of this is the uncertainty in calculating the theoretical power spectra over smaller scales and at low redshift. There is frequently a choice to either ex-

clude – and therefore *waste* – the wealth of available and expensively obtained data, or to use an inaccurate procedure, which may *bias* and invalidate any measurement determined with anticipated precision.

At present there are several approaches to dealing with this fruitful yet frustrating regime. One is to use sophisticated N-body simulations commonly produced with codes such as ENZO (O'Shea et al. 2010) and GADGET (Springel 2005). The most popular non-linear prescription HALOFIT (Smith et al. 2003) is a semi-analytical fit and has been fantastically successful. However, with larger and ever-improving state of the art of N-body simulations, the non-linearities on smaller scales have been shown to be at levels higher than the ones that were used in calibrating HALOFIT. As such, on smaller scales ($k \geq 0.1\,h\mathrm{Mpc}^{-1}$) the matter power spectra predicted by HALOFIT do not match the high precision N-body results well enough. If we are to perform precision cosmology it is imperative to go far beyond the levels of precision offered by current analytical approximations. An obstacle to further progress in obtaining accurate fits to underlying spectra is the vast computational demand from detailed N-body simulations and a high dimensionality in the cosmological parameter space.

There have been attempts (see Bird, Viel, & Haehnelt 2012) to calibrate HALOFIT using N-body simulations to predict suppression of matter power spectrum for cosmological models with massive neutrinos. However, semi-analytical fits like HALOFIT will themselves become obsolete with near-future surveys that promise to reach per cent level of precision. Moreover, implementing neutrinos as particles in numerical simulations is a topic of ongoing research, with results (see Brandbyge & Hannestad 2009; Viel, Haehnelt, & Springel 2010) contradictory at a level (factor of ~ 5 or higher) that can not be justified as due to (non)-inclusion of baryonic physics.

An alternative procedure to tackle small-scale non-linearities is to use higher order PT (*e.g.* Saito et al. 2008; Nishimichi et al. 2009; Saito, Takada, & Taruya 2009) to push further into the quasi-linear domain. Using high-resolution N-body simulations as reference, Carlson, White, & Padmanabhan (2009) have shown that although PT improves upon a linear description of the power spectrum at large scales ($k \sim 0.04\,h\mathrm{Mpc}^{-1}$), it expectedly fails on smaller scales ($k \gtrsim 0.08\,h\mathrm{Mpc}^{-1}$). The range of scales where PT is reliable at the per cent level is both redshift and cosmological model dependent. For cosmologies close to WMAP best-fit parameters, Taruya et al. (2009) have shown that at redshift $z = 0$, the one-Loop standard perturbation sequence to the non-linear matter power spectrum is expected to converge with the N-body simulation results to within 1 per cent - only for scales $k \lesssim 0.09\,h\mathrm{Mpc}^{-1}$. With the measurements from surveys expected to be at 1 per cent level precision, these upcoming data sets create new challenges in analyses and need alternative ways to efficiently estimate cosmological parameters.

One might argue that a machine-learning approach to determine the non-linear response from varying parameter settings is a rather black-box approach that goes against the traditional approach to spectra: based on scientific understanding and physics. However, we view this direction as a pragmatic one: a new approach is urgent given the impending flood of new data from upcoming surveys, and in an age of supposed precision cosmology, we will be theory limited in this specific area. It is therefore crucial to strive towards per cent level precision in the determination of the non-linear power spectrum.

Machine-learning techniques – in the form of Gaussian processes – have already been used as cosmological non-linear emulators in Habib et al. (2007), Schneider et al. (2008), Heitmann et al. (2009), Lawrence et al. (2010) and Schneider, Holm, & Knox (2011). Gaussian process modeling (see MacKay 1997; Rasmussen & Williams 2006 for a basic introduction to Gaussian processes) is a non-linear interpolation scheme that, after optimal learning, is capable of making predictions when queried at a suitable input setting. There are several advantages and disadvantages when using neural networks and Gaussian processes to interpolate data. From a practical point of view, a neural network compresses data into a small number of weight parameters, so a large number of simulations could be fitted into a small number of files whereas a Gaussian process has to carry a large matrix which can be of the order of the number of points used for training the Gaussian process. Heitmann et al. (2009) dealt with large matrices by using principal component analysis (PCA) to reduce their sizes to ones easily manipulated. Again from a practical point of view, usually Gaussian processes can do better than neural networks in the case of a small number of training points given that a neural network could be flexible enough to be misused and misfit the data. From a theoretical point of view, the two methods should fare equally especially as there are certain kernels used in Gaussian processes which are equivalent to the interpolation and fit one would have with neural networks. Overall, given the implementation, we believe that the two methods should produce equivalent results especially if the artificial neural network (ANN) procedure is trained by a larger number of simulations.

While using any machine-learning technique in lieu of N-body simulation output, it is critical that (i) the queried input setting not lie outside the input parameter ranges that are used during machine learning and (ii) the input parameter space must be sampled densely enough for the machine procedure to interpolate/predict accurately. Machine learning has been used in the fitting of cosmological functions (Auld et al. 2007; Fendt & Wandelt 2007; Auld, Bridges, & Hobson 2008) and photometric redshifts (Collister & Lahav 2004).

HALOFIT is accurate at the $5-10$ per cent level at best (see Heitmann et al. 2010). A far more accurate matter power spectrum calculator is the COSMIC EMULATOR (see Heitmann et al. 2009; Lawrence et al. 2010); although accurate at sub-per cent level, it makes predictions that are valid only for redshifts $z \leq 1$ and does not include cosmological models with massive neutrinos. In order to extend the interpolation validity range to $z \leq 2$, as well as improve the accuracy levels, we work on a new technique to fit results from cosmological N-body simulations using an ANN procedure with an improved Latin hypercube sampling of

the cosmological parameter space. Using HALOFIT spectra as mock *N*-body spectra, we show that the ANN formalism enables a remarkable fit with a manageable number of simulations.

The outline of this paper is as follows. We discuss the concept of machine learning, in particular the details of neural networks, in Section 2. The improved Latin hypercube sampling of the underlying parameter space, which keeps the simulation number manageable and fitting accuracy high, is detailed in Section 3. Our fitting results are included in Section 4. Finally, we conclude in Section 5.

2 MACHINE LEARNING - THE NEURAL NETWORK

Machine learning is associated with a series of algorithms that allow a computational unit to evolve in its behavior, given access to empirical data. The major benefit is the potential to automatically learn complex patterns. As a subset of artificial intelligence, machine learning has been used in a variety of applications ranging from brain-machine interfaces to the analyses of stock market. There exist a range of techniques (see *e.g.* Nilsson 2005) including genetic algorithms, decision tree learning and gaussian processes (as referenced above). In this work we focus on the neural network technique.

An ANN is simply an interconnection of neurons or *nodes* analogous to the neural structure of the brain. This can take a more specific form whereby the nodes are arranged in a series of layers with each node in a layer connected, with a weighting, to all other nodes in adjacent layers. This is often referred to as a multi-layer perceptron (MLP). In this case one can impart values onto the nodes of the first layer (called the *input* layer), have a series of *hidden* layers and finally receive information from the last layer (called the *output* layer). The configuration of nodes is often called the network's architecture and is specified from input to output as $N_i : N_1 : ... : N_n : N_o$. That is, a network with an architecture 7 : 49 : 50 has seven inputs, a single hidden layer with 49 nodes and finally 50 outputs. An extra node (called the *bias* node) is added to the input layer and connects to all the nodes in the hidden layers and the output layer (see Bishop 1995 for more details). The total number of weights N_W for a generic architecture $N_i : N_1 : ... : N_n : N_o$ can be calculated using the formula

$$N_W = N_i \cdot N_1 + \sum_{l=2}^{n} N_{l-1} \cdot N_l + N_n \cdot N_o + \sum_{l=1}^{n} N_l + N_o, \quad (1)$$

where the summation index l is over the hidden layers only. For a network with a single hidden layer, the second term on the right-hand side is absent. Specifically, the architecture 7 : 49 : 50 has a total of $7 \times 49 + 0 + 49 \times 50 + 49 + 50 = 2892$ weights, collectively denoted by **w**.

Each node (except the input nodes) is a neuron with an *activation*, $z_j = g(a)$, taking as its argument

$$a = \sum_i w_{ji} z_i, \quad (2)$$

where the sum is over all nodes i (of the previous layer) sending connections to the jth node (of the current layer). The activation functions are typically taken to be sigmoid functions such as $g(a) = 1/[1 + \exp(-a)]$. The sigmoid functions impart some degree of non-linearity to the neural network models. A network becomes overly non-linear if the weights **w** deviate significantly from zero. This drives the activation of the nodes to saturation. The number and size of the hidden layers add to the complexity of ANNs. For a particular input vector, the output vector of the network is determined by progressing sequentially through the network layers, from inputs to outputs, calculating the activation of each node.

Adjusting the weights **w** to get the desired mapping is called the *training* of the network. For matter power spectrum estimation, we use a training set of N-body simulations for which we have full information about the non-linear matter power spectra $P_{\rm nl}(k, z)$, as well as the underlying cosmological parameters: $\mathbf{I} \equiv (\Omega_m h^2, \Omega_b h^2, n_s, w_0, \sigma_8, \sum m_\nu)$, where $h, \Omega_m, \Omega_b, n_s, w_0, \sigma_8$ and $\sum m_\nu$ are the present-day normalized Hubble parameter in units of $100 \, {\rm km \, s^{-1} \, Mpc^{-1}}$, the present-day matter and baryonic normalized energy densities, the primordial spectral index, the constant equation of state parameter for dark energy, the amplitude of fluctuation on an $8 \, h^{-1}$ Mpc scale and the total neutrino mass, respectively.

Given the training set, the network can be used to learn some parametrization to arbitrary accuracy by training the weights **w**. This is done by minimizing a suitable *cost function*,

$$\chi^2 = \sum_{t=1}^{T} \sum_{k=1}^{c} \frac{[P_{\rm nl}^{\rm ANN}(k, z, \mathbf{w}, \mathbf{I}_t) - P_{\rm nl}(k, z, \mathbf{I}_t)]^2}{\sigma(k, z, \mathbf{I}_t)^2} \quad (3)$$

with respect to the weights **w**. The sum t is over all the cosmologies $\mathbf{I}_t$ in the training set. The sum k is over all the nodes in the output layer. Note that each output node samples the matter power spectrum at some specific scale, $k \, (h{\rm Mpc}^{-1})$. $P_{\rm nl}(k, z, \mathbf{I})$ is the true non-linear matter power spectrum for the specific cosmology **I**. In this paper, we use HALOFIT's approximation for $P_{\rm nl}(k, z)$. In Paper II N-body simulations will be used to calculate $P_{\rm nl}(k, z)$. Given the weights **w**, $P_{\rm nl}^{\rm ANN}(k, z, \mathbf{w}, \mathbf{I})$ is the ANN's predicted power spectrum for the **I**th cosmology. In our fitting procedure, we work with the ratio of the non-linear to linear power spectrum, namely $R(k, z) \equiv P_{\rm nl}(k, z)/P_{\rm lin}(k, z)$, where $P_{\rm lin}(k, z)$ is calculated using CAMB (Lewis, Challinor, & Lasenby 2000). As such, weighting the numerator in Eq. 3 by $\sigma = P_{\rm lin}(k, z)$ gives,

$$\chi^2 = \sum_{t=1}^{T} \sum_{k=1}^{c} \left[\frac{P_{\rm nl}^{\rm ANN}(k, z, \mathbf{w}, \mathbf{I}_t) - P_{\rm nl}(k, z, \mathbf{I}_t)}{P_{\rm lin}(k, z, \mathbf{I}_t)} \right]^2 \quad (4)$$

$$= \sum_{t=1}^{T} \sum_{k=1}^{c} [R_{\rm ANN}(k, z, \mathbf{w}, \mathbf{I}_t) - R(k, z, \mathbf{I}_t)]^2. \quad (5)$$

The ratio $R(k, z)$ is a flatter function and gives better performance, particularly at higher redshifts where the ratio tends to 1. Given the weights **w**, $R_{\rm ANN}(k, z, \mathbf{w}, \mathbf{I})$ in Eq. 5 is the network's prediction of the ratio $R(k, z, \mathbf{I})$ for the specific cosmology **I**. The predicted non-linear spec-

trum $P_{\mathrm{nl}}^{\mathrm{ANN}}(k, z, \mathbf{w}, \mathbf{I})$ in Eq. 4 is recovered by multiplying $R_{\mathrm{ANN}}(k, z, \mathbf{w}, \mathbf{I})$ by the corresponding linear spectrum $P_{\mathrm{lin}}(k, z, \mathbf{I})$.

We ran N-body simulations over a range of cosmological parameters with the publicly available adaptive mesh refinement (AMR), grid-based hybrid (hydro+N-body) code ENZO[1] (Norman et al. 2007; O'Shea et al. 2010). We include radiative cooling of baryons using an analytical approximation (Sarazin & White 1987) for a fully ionized gas with a metallicity of 0.5 $M_\odot$. The cooling approximation is valid over the temperature range from $10^4 - 10^9$ K. Below 10^4 K, the cooling rate is effectively zero. However, we do not account for metal-line cooling, supernova (SN) feedback or active galactic nucleus (AGN) feedback. It is worth mentioning here that van Daalen et al. (2011) have shown that the inclusion of AGN feedback can reproduce the optical and X-ray observations of groups of galaxies, and decrease the power relative to dark matter-only simulations at $z = 0$, ranging from 1 per cent at $k \approx 0.4\,h\mathrm{Mpc}^{-1}$ to as much as 10 per cent at $k \approx 1\,h\mathrm{Mpc}^{-1}$. As such, understanding and including the effects of baryonic physics in numerical simulations will be critical to predicting the non-linear matter power spectrum at sub-per cent level. Further, the ANN prescription we are using in this paper could also be used for fitting these kinds of baryonic effects by introducing additional parameters beyond the cosmological ones, especially since gasdynamical runs are much more expensive than dark matter-only simulations.

In Fig. 1, top panel, we show the power spectrum for a cosmological model $\mathbf{I} \equiv$ (0.13, 0.0224, 0.986, -1.23, 0.72, 0 eV), with $h = 0.8$. The spectrum is evaluated at redshift $z = 0$ (upper set) and $z = 1$ (lower set). At each redshift, the power spectrum is calculated using (i) linear theory (dot-dashed), (ii) N-body (dots), (iii) HALOFIT (dash) and (iv) COSMIC EMULATOR (solid). The vertical dash line at $k = 0.9\,h\mathrm{Mpc}^{-1}$ is the highest k upto which our N-body power spectrum is accurate at per cent level. We average over 10 realizations of the initial power spectrum to suppress the scatter in the N-body results. On smaller scales, our numerical simulations do not have the force resolution to give accurate results. The ratio of the N-body spectrum to COSMIC EMULATOR's prediction is shown in the middle ($z = 0$) and bottom ($z = 1$) panels. The error bars correspond to the scatter in the N-body results.

In this method paper, we match the linear theory power spectrum to the non-linear power spectrum from simulations at $k = 0.09\,h\mathrm{Mpc}^{-1}$. In Paper II we intend to use the one-Loop standard PT as implemented by Saito et al. (2008) for estimating the matter power spectrum upto $k \leq 0.085\,h\mathrm{Mpc}^{-1}$ and stitch it with the N-body spectrum. The stitched spectrum will be sampled at 50 k-values between $0.006\,h\mathrm{Mpc}^{-1} \leq k \leq 1\,h\mathrm{Mpc}^{-1}$. In Fig. 2, we show the stitched-and-sampled N-body power spectrum (solid dots) which we will use as $P_{\mathrm{nl}}(k, z)$ for ANN training in Paper II. The HALOFIT spectrum, sampled at the same k-values,
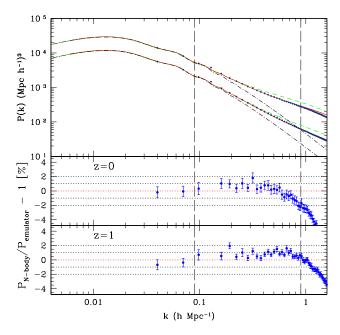
[1] http://lca.ucsd.edu/projects/enzo

Figure 1. *Top:* matter power spectrum evaluated at redshift $z = 0$ (upper set) and $z = 1$ (lower set). At each redshift, the spectrum is calculated using (i) linear theory (dot-dashed), (ii) N-body (dots), (iii) HALOFIT (dashed) and (iv) COSMIC EMULATOR (solid, see Lawrence et al. 2010). The cosmological parameters are: $\mathbf{I} \equiv$ $(0.13, 0.0224, 0.986, -1.23, 0.72, 0)$ with $h = 0.8$. In this method paper, we use the linear matter power spectrum for scales $k \leq 0.09\,h\mathrm{Mpc}^{-1}$. Due to the lack of force resolution on small scales, our N-body power spectrum is accurate at per cent level for $k \leq 0.9\,h\mathrm{Mpc}^{-1}$. *Middle:* The ratio of the N-body spectrum to COSMIC EMULATOR's prediction at $z = 0$. The error bars correspond to the scatter in the N-body results. The horizontal dotted lines denote $\pm 2, \pm 1$ and 0 per cent error. *Bottom:* The same as the middle panel, at $z = 1$.

is also shown (open circles) and is used as $P_{\mathrm{nl}}(k, z)$ in this method paper.

In Eq. 5, optimizing the weights $\mathbf{w}$ in order to minimize χ^2 generates an ANN that predicts the power spectrum very well for the specific cosmologies in the training set. However, such a network might not make accurate predictions for cosmologies *not* included in the training set. This usually indicates (i) an overly simple network architecture (very few hidden layer nodes), (ii) very sparsely/poorly sampled parameter space and/or (iii) a highly complex non-linear mapping that actually over-fits to the noise on the training dataset. In order to generate smoother network mappings that generalize better when presented with new cosmologies that are not part of the training set, a penalty term χ_Q^2 is added to the cost function χ^2,

$$\chi_Q^2 = \alpha \sum_{i,j} w_{ij}^2, \qquad (6)$$

where w_{ij} is the weight connecting the jth node to the ith node of the next layer. χ_Q^2, usually a quadratic sum of the weights, prevents them from becoming too large during the training process, by penalizing in proportion to the sum. The hyperparameter α controls the degree of regularization
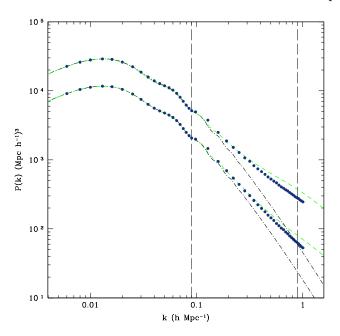
Figure 2. Linear theory, HALOFIT and N-body spectra from Fig. 1, left-hand panel are re-plotted – with the only difference that on scales $k \leq 0.09\,h\mathrm{Mpc}^{-1}$, the HALOFIT and N-body spectra are approximated by linear theory. The stitched spectra are then sampled at 50 k-values between $0.006\,h\mathrm{Mpc}^{-1} \leq k \leq 1\,h\mathrm{Mpc}^{-1}$. In this paper, we use the HALOFIT spectrum as $P_{\mathrm{nl}}(k, z)$ for ANN training. In Paper, II the N-body spectrum will be used along with the one-Loop standard PT.

of the network's non-linearity. After having initialized α, its value itself is re-estimated during the training process iteratively. See Bishop (1995) for more details.

Thus, the overall cost function which is presented to the ANN for minimization with respect to the weights $\mathbf{w}$ is,

$$\chi_C^2 = \sum_{t=1}^{T} \sum_{k=1}^{c} [R_{\mathrm{ANN}}(k, z, \mathbf{w}, \mathbf{I}_t) - R(k, z, \mathbf{I}_t)]^2 + \alpha \sum_{i,j} w_{ij}^2. \tag{7}$$

To minimize the cost function χ_C^2 w.r.t. the weights $\mathbf{w}$, we use an iterative quasi-Newton algorithm which involves evaluating the inverse Hessian matrix

$$H_{ij} = \frac{\partial^2 \chi_C^2}{\partial w_i w_j}. \tag{8}$$

Specifically, we employ the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method to approximate the inverse Hessian matrix (see Bishop 1995, for details). Starting with randomly assigned weights $\mathbf{w}$, their values are re-estimated iteratively, assuring that each iteration proceeds in a direction that lowers the cost function χ_C^2. After each iteration to the weights, Eq. 7 is also calculated for what is known in neural network parlance as a validation set, in order to avoid over-fitting to the training set. The validation set for our application of neural networks, is a small set of simulations with known $\mathbf{I} \equiv (\Omega_{\mathrm{m}}h^2, \Omega_{\mathrm{b}}h^2, n_s, w_0, \sigma_8, \sum m_\nu)$ and $P_{\mathrm{nl}}(k, z)$. The final weights $\mathbf{w}_f$ are chosen such as to give the best fit (minimum χ_C^2) to the validation set. The network training

is considered finished once χ_C^2 is minimized with respect to the validation set. The trained network can now be used to predict $P_{\mathrm{nl}}(k, z)$ for new cosmologies. In practice, a number of networks are trained that start with an alternative random configuration of weights. The trained networks are collectively called a committee of networks and subsequently give rise to better performance. The final output is usually given by averaging over the outputs of the committee members.

The ANN technique has been used successfully in empirical photometric redshift estimation with the ANNZ (Collister & Lahav 2004) package. ANNZ learns an effective parameterization of redshift as a function of broad-band photometric colours by training on a representative set of galaxies that have both photometric and spectroscopic information. This has been shown to be more successful than template or synthetic-based methods (Abdalla et al. 2011; Thomas, Abdalla, & Lahav 2011). In this work we use an MLP similar to the original ANNZ engine.

Our intention is to use this neural network technique to learn the non-linear matter power spectrum as a function of cosmological parameters by training on N-body simulations. This natural fitting procedure removes both the effort and unnecessary potential bias that results from invoking ultimately imperfect sets of fitting equations such as the HALOFIT. As can be seen in Fig. 2, the HALOFIT predictions are in error by as much as 50 per cent on small scales. As we will discuss in Section 4, the ANN technique is extremely fast and, more importantly, accurate.

3 LATIN HYPERCUBE PARAMETER SAMPLING

In order to fit a set of parameters optimally one strives to sample them as finely and as evenly as possible. However, a regularly spaced grid with N sampling intervals along one dimension and d parameters scales as N^d. For a six-dimensional parameter space with only 10 grid intervals, this quickly escalates to 10^6 points. The problem is exacerbated because an N-body simulation is a computationally expensive activity. To further compound this issue, each parameter configuration needs to be simulated over multiple realizations to beat down simulation (sample) variance. An alternate approach could be to interpolate the fitting function over a selection of randomly distributed points throughout the parameter space. However, this is prone to statistical clustering and will lead to a degradation of the machine-learning fit for parameters or regions affected by it. In order to circumvent these problems, we select parameters distributed on a Latin hypercube.

A square grid is said to be populated as a Latin square if, and only if, there is exactly *one* sample in each row and each column of the square. This is illustrated clearly in Fig. 3. A similar sampling scheme was developed first by Leonhard Euler who indexed the samples with Latin characters, motivating the name 'Latin square'. A Latin hypercube is a generalization of Euler's Latin square to a higher dimensional parameter space and is an example of a stratified
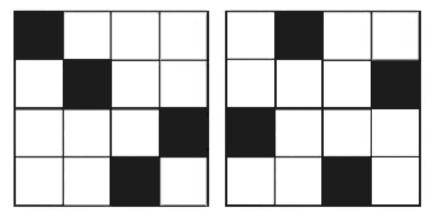
Figure 3. *Left:* An example of a Latin hypercube distribution. Every interval dx and dy is sampled; however each row and column are sampled only once. *Right:* an improved Latin hypercube where the distribution is more evenly spread through the space. Each subspace is equally sampled and there are no voids or clusters as in the left-hand panel (bottom left and right corners, respectively).

sampling technique. This ensures that each and every segment/interval along a parameter axis is sampled with high resolution without a vast number of points. That is, one can sample a d-dimensional space with n simulations and have all parameters evaluated along every d$x = (b - a)/n$ increment, where b and a are the upper and lower limits of the parameter, respectively. Therefore, it is independent of d. However, a random implementation of a Latin hypercube algorithm can still lead to statistically under-sampled regions. An example of this can be seen in Fig. 3. Each panel shows a random implementation of Latin hypercube algorithm. In both panels, the square is partitioned into four subspaces. The left-hand panel has voids (and clusters) in two of its subspaces. The right-hand panel has each subspace equally sampled (while still obeying the Latin hypercube definition) and represents an improved Latin hypercube sampling. In this case the sample space is partitioned into equally probable subspaces and the variance in the pairwise separation of the sampled points is minimized.

Since the introduction of the Latin hypercube sampling technique (McKay et al. 1979), the procedure has become common in computer science, uncertainty analysis and engineering emulation (where simulation of complex machinery is overwhelmingly time consuming). Similarly, variations of the Latin hypercube sampling technique have been implemented in cosmological analyses before, e.g., Habib et al. (2007), Heitmann et al. (2009), Schneider et al. (2011) and references therein. In this paper, we use the improved Latin hypercube technique to set up the cosmological models to be used for ANN training.

3.1 Setting up an improved Latin hypercube for cosmological parameters

We varied six cosmological parameters $\mathbf{I} \equiv (\Omega_{\mathrm{m}}h^2, \Omega_{\mathrm{b}}h^2, n_s, w_0, \sigma_8, \sum m_\nu)$ between the limits specified in Table 1. The limits on this six-dimensional parameter space are chosen so as to include the WMAP

7-year+BAO+H_0 (Komatsu et al. 2011) constraints (see Table 1).

Throughout this paper, we only consider spatially flat models with the present-day CMB temperature $T_\gamma{}^0 = 2.725K$. We also assume that all massive neutrino species are degenerate. The effective number of neutrino species is fixed at $N_{\mathrm{eff}} = 3.04$. We derive the Hubble parameter h using the WMAP 7-year+BAO constraint on the acoustic scale $\pi d_{ls}/r_s = 302.54$, where d_{ls} is the distance to the last scattering surface and r_s is the sound horizon at the redshift of last scattering. We derive h as follows.

(i) For a given $\Omega_b h^2$ and $\Omega_m h^2$, compute the redshift of the last scattering surface, z_{ls}, using the fit proposed by Hu & Sugiyama (1996):

$$z_{ls} = 1048 \left[1 + \frac{0.00124}{(\Omega_b h^2)^{0.738}} \right] \left[1 + b_1(\Omega_m h^2)^{b_2} \right] \quad (9)$$

$$b_1 = \frac{0.0783}{(\Omega_b h^2)^{0.238}} \left[1 + 39.5(\Omega_b h^2)^{0.763} \right]^{-1} \quad (10)$$

$$b_2 = \frac{0.560}{1 + 21.1(\Omega_b h^2)^{1.81}} \quad (11)$$

(ii) For a given $\Omega_b h^2$, $\Omega_m h^2$ and $\sum m_\nu$, choose a value for h and compute its evolution, $h(a)$. Here we follow section 3.3 from Komatsu et al. (2011), which includes the effect of massive neutrinos on $h(a)$:

$$h(a) = h\sqrt{\frac{\Omega_b + \Omega_c}{a^3} + \frac{\Omega_\gamma}{a^4}\left[1 + \frac{\Omega_\nu}{\Omega_\gamma}\right] + \frac{\Omega_\Lambda}{a^{3(1+w_0)}}} \quad (12)$$

$$\frac{\Omega_\nu}{\Omega_\gamma} = N_{\mathrm{eff}} \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} F(y) \quad (13)$$

$$F(y) = \frac{120}{7\pi^4} \int_0^\infty \frac{x^2\sqrt{x^2 + y^2}}{e^x + 1}dx, \quad (14)$$

Table 1. The six cosmological parameters and their ranges, used in generating the ANN training and validation sets. This six-dimensional parameter space is sampled using the improved Latin hypercube technique (see text for details). The last column shows the corresponding WMAP 7-year+BAO+H_0 constraints at 68 per cent CL.

Cosmological parameters	Lower value	Upper value	WMAP 7-year+BAO+H_0^a
$\Omega_{\rm m} h^2$	0.110	0.165	0.1352 ± 0.0036
$\Omega_{\rm b} h^2$	0.021	0.024	0.02255 ± 0.00054
$n_{\rm s}$	0.85	1.05	0.968 ± 0.012
w_0	-1.35	-0.65	-1.1 ± 0.14
σ_8	0.60	0.95	0.816 ± 0.024
$\sum m_\nu$ (eV)	0	1.1	$< 0.58^b$

Note. ^{a}Komatsu et al. (2011); ^{b}95 per cent CL for $w = -1$.

where

$$y \equiv \frac{m_\nu a}{T_\nu{}^0}$$

$$T_\nu{}^0 = \left(\frac{4}{11}\right)^{1/3} T_\gamma{}^0$$

$$\Omega_\gamma = \frac{2.4706 \times 10^{-5}}{h^2}\left(\frac{T_\gamma{}^0}{2.725}\right)^4.$$

$T_\nu{}^0$ is the present-day neutrino temperature and Ω_γ is the present-day normalized photon energy density. Given $\sum m_\nu$, the function $F(y)$ calculates the contribution of neutrinos to the radiation energy density at scale factor a.

(iii) Using $h(a)$ from step (ii), compute the comoving sound horizon $r_s(z)$ at the last scattering redshift z_{ls}:

$$r_s(z_{ls}) = \frac{c}{\sqrt{3}} \int_{a=0}^{1/(1+z_{ls})} \frac{da}{a^2 h(a)\sqrt{1 + (3\Omega_b/4\Omega_\gamma)a}}. \quad (15)$$

(iv) Using $r_s(z_{ls})$ from step (iii), together with the WMAP 7-year+BAO constraint on the acoustic scale $\pi d_{ls}/r_s = 302.54$, compute the comoving distance to the last scattering surface, d_{ls}:

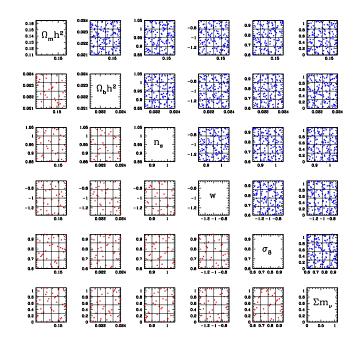$$d_{ls} = \frac{302.54}{\pi} r_s(z_{ls}). \quad (16)$$

(v) Using $h(a)$ from step (ii), compute the comoving distance to the surface of last scattering $\chi(z_{ls})$:

$$\chi(z_{ls}) = c \int_{1/(1+z_{ls})}^{a=1} \frac{da}{a^2 h(a)}. \quad (17)$$

(vi) Compare results from steps (iv) and (v). Minimize the difference $|d_{ls} - \chi(z_{ls})|$ by varying h in step (ii) and re-estimating steps (ii)-(v).

Using Table 1 as the parameter priors, we sampled this six-dimensional parameter space with an improved Latin hypercube technique. We generated 130 cosmologies to be used as the ANN training set and another 32 cosmologies for the validation set. We show the training set (upper triangle) and the validation set (lower triangle) in Fig. 4.

As can be seen in Fig. 3, a major advantage of improved Latin hypercube sampling technique is the relatively uniform coverage it provides. This is, of course, highly useful for training a machine-learning algorithm. As with any interpolation mechanism, one hopes that the neural network

Figure 4. Upper triangle: ANN training set with 130 viable cosmologies, in a six-dimensional parameter space. Lower triangle: ANN validation set with 32 viable cosmologies, in a six-dimensional parameter space. See Table 1 for the parameter priors used to generate the training and validation sets.

can generalize from what it has learned to new and slightly different input data (in this case cosmological parameters). In reality, the response will be uncertain in poorly trained areas. Therefore, the caveat with our sampling will reside near the edges of the parameter hypercube. A parameter value that we might want emulated may not be encapsulated within the hypervolume of a simulated, and therefore trained, point. This can be understood with reference to Fig. 4. The performance of a neural network can severely degrade near the parameter boundaries. The solution is simply to choose prior ranges that are marginally wider than those of real interest. The allowance could easily be found empirically by projecting the hypercube realizations. The real problem in cosmology therefore arises when one has a parameter that is physically bounded, an example being the neutrino mass $\sum m_\nu \gtrsim 0$.
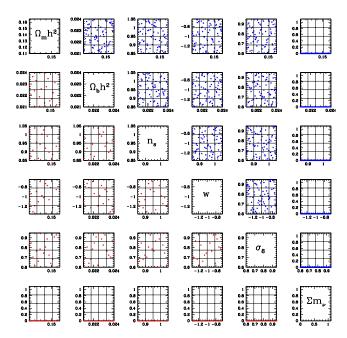
Figure 5. Upper triangle: extending the ANN training set (upper triangle in Fig. 4) with 70 cosmologies with $\sum m_\nu = 0$. Lower triangle: extending the ANN validation set (lower triangle in Fig. 4) with 18 cosmologies with $\sum m_\nu = 0$.

4   ANALYSIS AND RESULTS

We tested the precision with which our neural network can predict the non-linear matter power spectrum. We selected the combination $7 : N_{hidden} : 50$ as our network architecture, where N_{hidden} (number of nodes in the hidden layer) was varied from 7 to 56, in steps of 7. The number of inputs were fixed at 7, corresponding to $\mathbf{I} \equiv (\Omega_m h^2, \Omega_b h^2, n_s, w_0, \sigma_8, \sum m_\nu)$ including redshift z. Note that we do not sample the redshift in the Latin hypercube but instead evaluate $P_{nl}(k, z)$, at 111 redshifts between $z = 0$ and $z = 2$, using the existing prescription HALOFIT coupled with the CAMB software. These HALOFIT-generated spectra serve as mock N-body spectra. We sampled $P_{nl}(k, z)$ at 50 points between $0.006 \, h\mathrm{Mpc}^{-1} \le k \le 1.0 \, h\mathrm{Mpc}^{-1}$. Since our training and validation sets have $(130+70)$ and $(32+18)$ cosmologies, respectively, we calculated $P_{nl}(k, z)$ for each cosmology, at 111 redshifts. These $P_{nl}(k, z)$ are scaled by their respective linear spectra $P_{lin}(k, z)$, before being fed to the

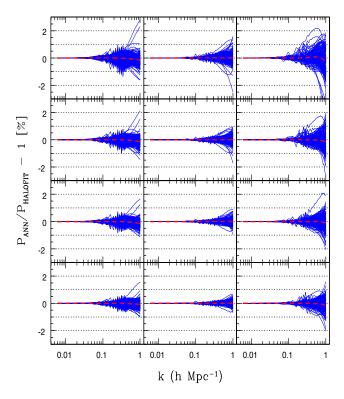**Figure 6.** *Left-hand column:* percentage error at redshift $z = 0$, between the predicted non-linear power spectrum (using ANN) and the true underlying mock (using HALOFIT) for 200 training set cosmologies shown in the upper triangles of Figs. 4 and 5. The profile is continuous as the 50 output values have been spline interpolated over the functional range. The rows (from top to bottom) correspond to $N_{hidden} = 28 - 49$ in increments of 7. The mean error over all 200 cosmologies is shown by a dashed line – an indicator of any bias in the ANN training scheme. *Middle-column:* the same as the left-hand column, at redshift $z = 1$. *Right-hand column:* the same as the left-hand column, at redshift $z = 2$.

neural network. Thus, the overall size of the training set that we train our ANN with is $200 \times 111 = 22,200$. Likewise, we have $50 \times 111 = 5,550$ patterns in the validation set. We trained a committee of 16 ANNs at each N_{hidden} setting. The weights $\mathbf{w}$ for each ANN were randomly initialized (the random configuration being different for each ANN). The weights are allowed to evolve until χ^2_C (see Eq. 7) is minimized with respect to the cosmologies in the validation set.

In Fig. 6, we show the performance of the trained ANNs with varying N_{hidden} units, when presented with each of the 200 cosmologies in the training set. Note that we average the $P_{nl}^{\mathrm{ANN}}(k, z)$ predictions over all 16 ANN committee members. The rows correspond to $N_{hidden} = 28 - 49$ (from top to bottom) in increments of 7. The columns (from left to right) correspond to $z = 0, 1, 2$. The mean error over all 200 cosmologies in the training set is shown by a dashed line in each panel, to get an idea about any systematics in our ANN training scheme. With $N_{hidden} = 49$, the ANN predictions at redshifts $z = 0$ and $z = 1$, on *all* scales, are within ± 1 per cent of the HALOFIT power spectra. Although

we show results at $z = 0$ and $z = 1$, we have checked that the predictions are 1 per cent level for all $z \leq 1$. Predictions are at the 1 per cent level even up to redshift $z = 2$ for $k \leq 0.7\,h\mathrm{Mpc}^{-1}$, after which the performance degrades to ± 2 per cent. We have checked and confirmed that the worst-performing cosmologies correspond to the parameter settings in which at least four of the six cosmological parameters are at their boundary values.

As we mentioned earlier, this fitting procedure will be less accurate near the boundaries of the parameter ranges because some parameter configurations may not be encapsulated within the volume of a training point. This also explains why the ANN performance is better at $z = 1$ – the mid-point of the redshift range. Looking at the bias (dashed line in Fig. 6), we see that the distribution of errors in the ANN predictions is centered on zero, indicating that our interpolations are not biased. A negligible bias, and the fact that for *every* cosmological setting *within* the parameter priors (see Table 1) the non-linear power spectrum at $z \leq 2$ is correctly predicted within ± 1 per cent up to $k \leq 0.7\,h\mathrm{Mpc}^{-1}$, demonstrates the stability of our ANN strategy. This marks a remarkable improvement over the currently popular interpolation scheme – the COSMIC EMULATOR, which has a significant number (~ 50 per cent) of cosmological models with errors at $\sim 0.5 - 1$ per cent level. We note, however, that the COSMIC EMULATOR, based on Gaussian processes, is able to achieve sub-per cent accuracy with only 37 distinct cosmologies while in the ANN scheme we use a suite of around 200 cosmologies. Comparing Fig. 9 from Heitmann et al. (2009) with our Fig. 6, we see that the ANN implementation performs better on all scales and redshifts.

In order to check the performance of our trained ANNs over parameter configurations that do not touch the Latin hypercube, we generated a testing set of 330 cosmologies (of which 150 have $\sum m_\nu = 0$). See Table 2 for the parameter limits of the testing set.

A testing set serves another crucial purpose. Increasing the number of nodes in the hidden layer increases the flexibility of a neural network. An increasingly complex network can make extremely accurate predictions on the training set. This is evident from Fig. 6, where the prediction over the training set becomes progressively better (from top to bottom) with increasing N_{hidden} units. However, such complex networks can adversely affect their generalizing ability when presented with a new dataset. Measuring the performance of a neural network on an independent dataset (known as a testing set) as a function of N_{hidden} units helps in controlling its complexity. We show the testing set in Fig. 7, with the lower triangle corresponding to the 150 cosmologies with $\sum m_\nu = 0$.

The performance of the trained ANNs as a function of N_{hidden} units, over the cosmologies in the testing set, is shown in Fig. 8. Increasing N_{hidden} from 42 to 49 reduces the error marginally. We have checked that increasing N_{hidden} beyond 49 does not contribute to any further error reduction on the testing set, indicating that $N_{hidden} = 49$ saturates the generalizing ability of the network. With $N_{hidden} = 49$, the ANN prediction for every cosmology, on *all* scales, at
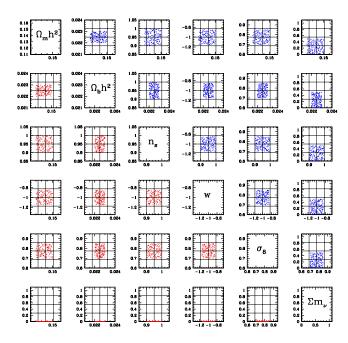
Figure 7. Upper triangle: ANN testing set with 180 cosmologies with $\sum m_\nu > 0$. Lower triangle: extending the ANN testing set with 150 cosmologies with $\sum m_\nu = 0$. See Table 2 for the parameter priors used to generate the testing set.

redshifts $z \leq 1$, is within ± 0.5 per cent of the HALOFIT power spectra. For $1 < z \leq 2$, in the high k range, the performance degrades slightly to ± 0.8 per cent. As expected, the ANN performs exceedingly well away from the boundaries of the parameter ranges. It is quite remarkable that our ANN scheme is capable of making predictions at sub-per cent level, especially on the testing set that is *not* a part of the ANN training process.
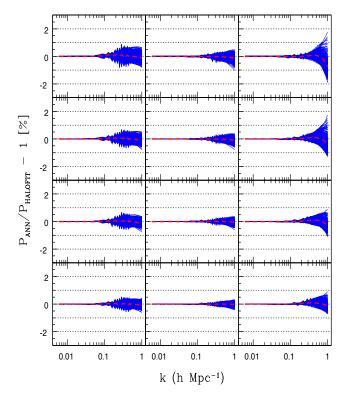
In Fig. 9, left panel, we show the non-linear power spectrum at redshifts $z = 0$ and $z = 1$ predicted by our trained ANN (solid dots) for the same cosmology that was used to generate the HALOFIT power spectrum of Fig. 2, namely $\mathbf{I} \equiv (0.13, 0.0224, 0.986, -1.23, 0.72, 0\text{ eV})$ with $h = 0.8$. N_{hidden} is fixed at 49, as discussed above. For comparison, the HALOFIT power spectra are re-plotted as starred points. The prediction errors at $z = 0$ and $z = 1$ are shown in the middle and the right-hand panels, respectively. The ANN predictions are well within ± 0.5 per cent over the scales of interest.

We reiterate that this method for reconstructing the non-linear power spectrum will only function for the parameters and ranges that have been simulated and trained with PKANN. The intention of this study is to provide a technique for high precision fits in the concordance model for the oncoming generation of surveys. This should therefore act as a safety mechanism as it demonstrates that the range of validity has been breached, as often occurs with blind application of other fits. In Paper II, we will put our ANN formalism to test using matter power spectra calculated from N-body simulations. On mildly non-linear scales $(0.1\,h\mathrm{Mpc}^{-1} < k < 1.0\,h\mathrm{Mpc}^{-1})$ the power spectrum from

Table 2. The six cosmological parameters and their ranges, used in generating the ANN testing set. This six-dimensional parameter space is sampled using the improved Latin hypercube technique (see the text for details). The parameter ranges are chosen so as to avoid the boundaries of the parameter space. See Table 1 for the parameter boundaries. Note that the lower bound on neutrino mass is still set at zero, since neutrinos are physically bound ($\sum m_\nu \gtrsim 0$). The last column shows the WMAP 7-year+BAO+H_0 constraints at 68 per cent CL.

Cosmological parameters	Lower value	Upper value	WMAP 7-year+BAO+H_0^a
$\Omega_{\rm m}h^2$	0.120	0.150	0.1352 ± 0.0036
$\Omega_{\rm b}h^2$	0.022	0.023	0.02255 ± 0.00054
$n_{\rm s}$	0.90	1.00	0.968 ± 0.012
w_0	-1.15	-0.85	-1.1 ± 0.14
σ_8	0.70	0.85	0.816 ± 0.024
$\sum m_\nu$ (eV)	0	0.50	$< 0.58^b$

Note. ^{a}Komatsu et al. (2011); ^{b}95 per cent CL for $w = -1$.

Figure 8. The same as Fig. 6, using 330 testing set cosmologies shown in Fig. 7.

N-body simulations is expected to vary smoothly as a function of cosmological parameters (Heitmann et al. 2006, 2009; Agarwal & Feldman 2011). As such, our ANN interpolation scheme that we have shown here to work well on HALOFIT power spectra is expected to perform satisfactorily on *N*-body spectra as well.

5   CONCLUSIONS

The advent of the era of precision cosmology poses an immense challenge to theoretical physics. The upcoming generation of surveys has the potential to breach per cent level of accuracy. Such high-precision data will improve our constraints on cosmological parameters including dark energy and neutrino masses. Efficiently dealing with this impending flood of precise data on ever smaller scales and lower redshifts requires that we move on from linear theory as well as any imperfect sets of fitting equations. Although numerical simulations are capable of achieving the levels of precision required by the near-future surveys, the high dimensionality of the cosmological parameter space renders their brute force usage impractical.

We have introduced a unique approach to coping with non-linearities in the matter power spectra in cosmology. By employing a multi-layer perceptron neural network together with an improved Latin hypercube parameter sampling technique, we have demonstrated that the non-linear spectrum can be reconstructed from a full set of Λ cold dark matter parameters to better than 1 per cent over the parameter space spanning 3σ confidence level around the WMAP 7-year central values. Parameters that are likely to reside by some hard physical prior, such as the neutrino mass, can be successfully brought under the realms of ANNs by sprinkling of extra simulations in the corresponding (*e.g.* $\sum m_\nu = 0$) hyper-plane.

Looking forward, our ANN procedure can be readily employed for a variety of cosmological tasks such as fitting halo mass functions obtained through high resolution *N*-body simulations. Moreover, mixed datasets such as the matter power spectra and the halo mass functions can be combined and presented to a neural network as the training set. An ANN trained with such a heterogeneous dataset would be capable of cosmological parameter estimation when presented with the combined observations of the matter power spectrum and the measured halo mass function. The implementation of our technique avoids complex calculations and, through the execution of only the neural network weights, is extremely fast (predictions take less than a second). An automated PκANN function will be released with our program of *N*-body results in a companion paper shortly. Beyond this we hope that with our method a collaborative effort could reduce non-linear error to only uncertainty in the *N*-body simulation's baryon interactions.
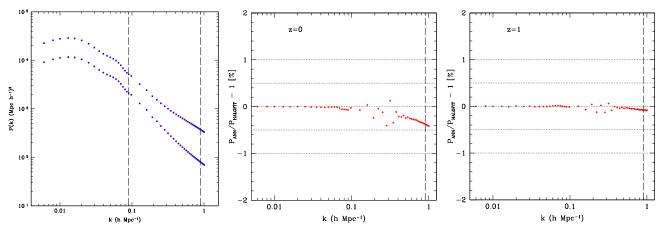
Figure 9. *Left:* ANN predictions (solid dots) at $z = 0$ (upper) and $z = 1$ (lower) for the HALOFIT power spectra of Fig. 2 (re-plotted here as starred points – difficult to see due to the excellent ANN predictions). The cosmological model is $\mathbf{I} \equiv$ (0.13, 0.0224, 0.986, -1.23, 0.72, 0 eV) with $h = 0.8$. N_{hidden} is fixed at 49, as discussed in the text. The ratio of ANN predictions to the HALOFIT spectra is shown in the other panels. *Middle:* percentage error in ANN predictions, at $z = 0$. *Right:* percentage error in ANN predictions, at $z = 1$. The ANN predictions are within ± 0.5 per cent over the scales of interest.

6 ACKNOWLEDGEMENTS

REFERENCES

Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, MNRAS, 417, 1891

Agarwal S., Feldman H. A., 2011, MNRAS, 410, 1647

Amendola L., Kunz M., Sapone D., 2008, JCAP, 4, 13

Auld T., Bridges M., Hobson M. P., 2008, MNRAS, 387, 1575

Auld T., Bridges M., Hobson M. P., Gull S. F., 2007, MNRAS, 376, L11

Bird S., Viel M., Haehnelt M. G., 2012, MNRAS, 2175

Bishop C. M., 1995, Neural Networks for Pattern Recognition. Oxford Univ. Press, New York

Brandbyge J., Hannestad S., 2009, JCAP, 5, 2

Carlson J., White M., Padmanabhan N., 2009, PRD, 80, 043531

Collister A. A., Lahav O., 2004, PASP, 116, 345

Eisenstein D. J., Weinberg D. H., Agol E., Aihara H., Allende Prieto C., Anderson S. F., Arns J. A., Aubourg É., Bailey S., Balbinot E., et al., 2011, AJ, 142, 72

Fendt W. A., Wandelt B. D., 2007, ApJ, 654, 2

Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, PRD, 76, 083503

Heitmann K., Higdon D., Nakhleh C., Habib S., 2006, ApJL, 646, L1

Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, ApJ, 705, 156

Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104

Hu W., Sugiyama N., 1996, ApJ, 471, 542

Ivezic Z., Tyson J. A., Acosta E., Allsman R., Anderson S. F., Andrew J., Angel R., Axelrod T., Barr J. D., Becker A. C., Becla J., Beldica C., Blandford R. D., Bloom J. S., Borne K., Brandt W. N., Brown M. E., Bullock J. S., Burke D. L., Chandrasekharan S., Chesley S., Claver C. F., Connolly A., Cook K. H., Cooray A., Covey K. R., Cribbs C., Cutri R., Daues G., Delgado F., Ferguson H., Gawiser E., Geary J. C., Gee P., Geha M., Gibson R. R., Gilmore D. K., Gressler W. J., Hogan C., Huffer M. E., Jacoby S. H., Jain B., Jernigan J. G., Jones R. L., Juric M., Kahn S. M., Kalirai J. S., Kantor J. P., Kessler R., Kirkby D., Knox L., Krabbendam V. L., Krughoff S., Kulkarni S., Lambert R., Levine D., Liang M., Lim K., Lupton R. H., Marshall P., Marshall S., May M., Miller M., Mills D. J., Monet D. G., Neill D. R., Nordby M., O'Connor P., Oliver J., Olivier S. S., Olsen K., Owen R. E., Peterson J. R., Petry C. E., Pierfederici F., Pietrowicz S., Pike R., Pinto P. A., Plante R., Radeka V., Rasmussen A., Ridgway S. T., Rosing W., Saha A., Schalk T. L., Schindler R. H., Schneider D. P., Schumacher G., Sebag J., Seppala L. G., Shipsey I., Silvestri N., Smith J. A., Smith R. C., Strauss M. A., Stubbs C. W., Sweeney D., Szalay A., Thaler J. J., Vanden Berk D., Walkowicz L., Warner M., Willman B., Wittman D., Wolff S. C., Wood-Vasey W. M., Yoachim P., Zhan H., for the LSST Collaboration, 2008, arXiv:0805.2366v2

Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., Nolta M. R., Page L., Spergel D. N., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Odegard N., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2011, ApJS, 192, 18

Lawrence E., Heitmann K., White M., Higdon D., Wagner

C., Habib S., Williams B., 2010, ApJ, 713, 1322

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

MacKay D. J. C., 1997, Gaussian processes – a replacement for supervised neural networks? Lecture notes for a tutorial at NIPS 1997

McKay M. D., Beckman R. J., Conover W. J., 1979, Technometrics, 21, 239

Nilsson N. J., 2005, Machine learn., 56, 387

Nishimichi T., Shirata A., Taruya A., Yahata K., Saito S., Suto Y., Takahashi R., Yoshida N., Matsubara T., Sugiyama N., Kayo I., Jing Y., Yoshikawa K., 2009, PASJ, 61, 321

Norman M. L., Bryan G. L., Harkness R., Bordner J., Reynolds D., O'Shea B., Wagner R., 2007, arXiv:0705.1556

O'Shea B. W., Bryan G., Bordner J., Norman M. L., Abel T., Harkness R., Kritsuk A., 2010, Astrophysics Source Code Library, 10072

Rasmussen C. E., Williams C., 2006, Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA

Saito S., Takada M., Taruya A., 2008, Physical Review Letters, 100, 191301

—, 2009, PRD, 80, 083528

Sarazin C. L., White III R. E., 1987, ApJ, 320, 32

Schneider M. D., Holm Ó., Knox L., 2011, ApJ, 728, 137

Schneider M. D., Knox L., Habib S., Heitmann K., Higdon D., Nakhleh C., 2008, PRD, 78, 063529

Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311

Springel V., 2005, MNRAS, 364, 1105

Taruya A., Nishimichi T., Saito S., Hiramatsu T., 2009, PRD, 80, 123503

The Dark Energy Survey Collaboration, 2005, arXiv:astro-ph/0510346

Thomas S. A., Abdalla F. B., Lahav O., 2010, Physical Review Letters, 105, 031301

—, 2011, MNRAS, 412, 1669

van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649

Viel M., Haehnelt M. G., Springel V., 2010, Journal of Cosmology and Astro-Particle Physics, 6, 15