PkANN: Non-Linear Matter Power Spectrum Interpolation through Artificial Neural Networks

by

Shankar Agarwal, Ph.D

Submitted to the Department of Physics and Astronomy and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Committee:

Dr. Hume Feldman, Chairperson

Dr. Mikhail V. Medvedev

Dr. Danny Marfatia

Dr. Jack Shi

Dr. David Lerner

Date defended: 30 November 2012

The Dissertation Committee for Shankar Agarwal, Ph.D certifies that this is the approved version of the following dissertation:

PkANN: Non-Linear Matter Power Spectrum Interpolation through Artificial Neural Networks

Committee:

Dr. Hume Feldman, Chairperson

Dr. Mikhail V. Medvedev

Dr. Danny Marfatia

Dr. Jack Shi

Dr. David Lerner

Date approved: 30 November 2012

Abstract

We investigate the interpolation of power spectra of matter fluctuations using artificial neural networks (ANNs). We present a new approach to confront small-scale non-linearities in the matter power spectrum. 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 detail how an accurate interpolation of the matter power spectrum is achievable with only a sparsely sampled grid of cosmological parameters. We show that an optimally trained ANN, when presented with a set of cosmological parameters ($\Omega_{\rm m}h^2, \Omega_{\rm b}h^2, n_s, w_0, \sigma_8, \sum m_\nu$ and z), can provide a worst-case error ≤ 1 per cent (for redshift $z \leq 2$) fit to the non-linear matter power spectrum deduced through large-scale N-body simulations, for modes up to $k \leq 0.9\,h{\rm Mpc}^{-1}$. Our power spectrum interpolator, which we label 'PkANN', is designed to simulate a range of cosmological models including massive neutrinos and dark energy equation of state $w_0 \neq -1$. PkANN is accurate in the quasi-non-linear regime ($0.1\,h{\rm Mpc}^{-1} \leq k \leq 0.9\,h{\rm Mpc}^{-1}$) over the entire parameter space and marks a significant improvement over some of the current power spectrum calculators. The response of the power spectrum to variations in the cosmological parameters is explored using PkANN. Using a compilation of existing peculiar velocity surveys, we investigate the cosmic Mach number statistic and show that PkANN not only successfully accounts for the non-linear motions on small scales, but also, unlike N-body simulations which are computationally expensive and/or infeasible, it can be an extremely quick and reliable tool in interpreting cosmological observations and testing theories of structure-formation.

Acknowledgements

I would like to express my sincere gratitude to my advisor Dr. Hume A. Feldman for the continuous support of my Ph.D study and research. PkANN code was made possible in collaboration with Dr. Ofer Lahav, Dr. Filipe B. Abdalla and Dr. Shaun A. Thomas at the University College London, UK. I thank Dr. Salman Habib and Dr. Katrin Heitmann at the Argonne National Laboratory, University of Chicago, USA, for their insightful comments and suggestions on the PkANN code.

This work was supported by the National Science Foundation through TeraGrid resources provided by the National Center for Supercomputing Applications (NCSA) and by a grant from the Research Corporation. Computations described in this work were performed using the enzo code developed by the Laboratory for Computational Astrophysics at the University of California, San Diego (http://lca.ucsd.edu). I thank the users of yt (python-based package for analyzing cosmological datasets) and enzo for useful discussions and guidance towards running and analyzing simulations. I am grateful to Dr. Róman Scoccimarro at the New York University, USA, and the LasDamas (Large Suite of Dark Matter Simulations) collaboration for providing the LasDamas simulations suite.

Contents

List of Tables

List of Figures

1 Introduction

The growth of structure is a direct consequence of the primordial gravitational instability present in our Universe. Observations of galaxy clustering is a powerful tool to test structure-formation theories and complements the cosmic variance limited cosmic microwave background. This complementarity explains the large number of galaxy surveys in various stages of planning and construction 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 matter power spectrum (Fourier transform of the matter density field) 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 and Feldman, 2011] and analytical studies inspired from perturbation theory (PT) [Scoccimarro et al., 1999, Saito et al., 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 matter power spectrum over small scales and at low redshifts. There is frequently a choice to either exclude – 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 deal 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. On small scales ($\lambda \lesssim 60h^{-1}$Mpc, *i.e.* wavenumbers $k \equiv 2\pi/\lambda \gtrsim 0.1\,h\mathrm{Mpc}^{-1}$, where 1 Mpc $\approx$ 3.26 million light-years and h is the present-day normalized Hubble parameter in units of 100 km s^{-1} Mpc^{-1}), the matter power spectra estimated 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 et al. [2012]) to calibrate HALOFIT using N-body simulations to estimate 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 and Hannestad [2009a], Viel et al. [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 et al. [2009a]) to push further into the quasi-linear domain. Using high-resolution N-body simulations as reference, Carlson et al. [2009] have shown that although PT improves upon a linear description of the power spectrum on 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 per cent level is both redshift and cosmological model dependent. For cosmologies close to the 3-yr and 5-yr *Wilkinson Microwave Anisotropy Probe* (*WMAP*; Spergel et al. [2007], Komatsu

et al. [2009]) 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.

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 (i) extend the interpolation validity range to $z \leq 2$, (ii) incorporate massive neutrino cosmologies, and (iii) 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 a suite of N-body simulations spanning cosmologies close to the 7-yr *WMAP* [Komatsu et al., 2011] central parameters, we show that the ANN formalism enables a remarkable fit with a manageable number of simulations. The organization of chapters is as follows.

a Chapter 2: Numerical Simulations: The numerical methods employed in some recent cosmological studies of neutrinos are discussed. We introduce an alternative approach to incorporate neutrinos in N-body simulations. The convergence tests for the matter power spectrum calculated from N-body simulations are presented. The impact of massive neutrinos on the power spectrum of matter fluctuations is shown. The precision levels at which future galaxy surveys would need to measure the matter power spectrum in order to distinguish between the normal and inverted neutrino mass hierarchies is explored.

b Chapter 3: Developing PKANN – A Non-Linear Matter Power Spectrum Interpolator: The concept of machine learning is discussed. PKANN – an

artificial neural network framework, is developed to interpolate the power spectrum of matter fluctuations as a function of cosmological parameters. An improved Latin hypercube sampling of the underlying parameter space, which keeps the simulation number manageable and fitting accuracy high, is detailed.

c Chapter 4: Interpolating Matter Power Spectrum using PkANN: The matter power spectra estimated using PkANN are compared with the spectra computed directly from numerical simulations. The PkANN spectra are also compared to those estimated using a popular power spectrum calculator, the COSMIC EMULATOR. The sensitivity of the power spectrum to variations in the cosmological parameters is explored.

d Chapter 5: Estimating the Cosmic Mach Number using PkANN: The cosmic Mach number, a measure of the growth of cosmological structure, is reviewed. The statistical distribution of the Mach number is studied using numerical simulations of a *WMAP*-type cosmology, and compared with the Mach predictions based on PkANN spectra. Mach numbers are estimated for various galaxy peculiar velocity surveys using the maximum likelihood estimate (MLE) method.

2 Numerical Simulations

2.1 Prelude

In this chapter, we discuss the impact of massive neutrinos on the growth of large-scale structure in the Universe. We describe the methods employed to include massive neutrinos in numerical simulations in some recent studies. We present an alternate implementation of neutrinos in N-body simulations. The simulation volume and mass resolution tests are presented, with the intention of calculating the matter power spectrum at per cent level accuracy. We show the impact of massive neutrinos on the matter distribution through the matter power spectrum. The precision level at which future surveys would need to measure the matter power spectrum in order to distinguish between the normal and inverted neutrino mass hierarchies is discussed. We compare our results with the neutrino simulations performed by other groups.

2.2 Probing Structure Formation through Neutrinos

In the standard model of particle physics there are three types (flavors) of neutrinos: electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). Neutrino oscillation experiments [KamLAND, 2008, SNO, 2004] in the past decade indicate that at least two neutrino eigentstates have non-zero masses. The direct implication of massive neutrinos is a non-zero hot dark matter (HDM) contribution to the total energy density of the Universe. Being sensitive to the mass squared differences between the neutrino eigentstates, the oscillation experiments only provide a lower bound on the total neutrino mass. Mass splittings of $|\Delta m^2_{32}| = (2.43 \pm 0.13) \times 10^{-3}\,\mathrm{eV}^2$ and $\Delta m^2_{21} = (7.59 \pm 0.21) \times 10^{-5}\,\mathrm{eV}^2$ [Adamson et al., 2008, KamLAND, 2008] imply a lower limit for the sum of the neutrino masses to be 0.05 and 0.1 eV for the normal and inverted mass hierarchies [Otten and Weinheimer, 2008], respectively.

During the radiation era, matter perturbations on the sub-horizon scales grow logarithmically. The earlier a mode enters the horizon, the more it is suppressed due to

the decaying gravitational potentials. On the other hand, the super-horizon modes do not decay until they enter the horizon. As a result, the matter power spectrum turns over at a scale that corresponds to the one that entered the horizon at radiation–matter equality. Neutrinos with mass on the sub-eV scale behave as a hot component of the dark matter. Neutrinos stream out of high-density regions into low-density regions, thereby damping out small-scale density perturbations. Massive neutrinos, therefore, suppress the logarithmic growth of sub-horizon modes. Extremely low mass neutrinos become non-relativistic after the radiation era is over and the free-streaming damping of matter perturbations affects even those scales that were always outside the horizon during the radiation era.

The redshift-dependent free-streaming comoving wave number, $k_{\rm fs}$, is given by

$$k_{\rm fs}(z) = \sqrt{\frac{3}{2}} \frac{H(z)}{(1+z)v_{\rm th}}, \tag{2.1}$$

where $H(z)$ and $v_{\rm th}$ are the Hubble parameter and the neutrino thermal velocity, respectively. For relativistic neutrinos, the free-streaming comoving wave number shrinks in proportion to the comoving Hubble wave number (Eq. 2.1). After a neutrino eigentstate becomes non-relativistic, its thermal velocity decays as

$$\begin{aligned} v_{\rm th} &\approx \frac{3T_\nu}{m_\nu} = 3\left(\frac{4}{11}\right)^{1/3} \frac{T_\gamma^0(1+z)}{m_\nu} \\ &\approx 151(1+z)\left(\frac{1\,{\rm eV}}{m_\nu}\right) {\rm km/s}, \end{aligned} \tag{2.2}$$

where m_ν is the mass of a neutrino eigentstate in eV and the present-day photon temperature, T_γ^0, is 2.725 K [Komatsu et al., 2010].

Thus the free-streaming comoving wave number for non-relativistic neutrinos is given by

$$k_{\rm fs} \approx 0.81 \frac{\sqrt{\Omega_\Lambda + \Omega_{\rm m}(1+z)^3}}{(1+z)^2} \left(\frac{m_\nu}{1\,{\rm eV}}\right) h\,{\rm Mpc}^{-1}. \tag{2.3}$$

For a massive eigentstate, the redshift of non-relativistic transition ($m_\nu \approx 3T_\nu$) is given

by

$$1 + z_{\rm nr} \approx 1987 \left(\frac{m_\nu}{1\,{\rm eV}}\right). \tag{2.4}$$

After a neutrino eigentstate becomes non-relativistic, $k_{\rm fs}$ begins to grow as $k_{\rm fs} \propto (1 + z)^{-1/2}$. Thus, $k_{\rm fs}$ passes through a minimum, $k_{\rm nr}$, which can be shown to be (from Eq. 2.3)

$$k_{\rm nr} \approx 0.018 \left(\frac{m_\nu}{1\,{\rm eV}}\right)^{1/2} (\Omega_{\rm m} h^2)^{1/2}\,{\rm Mpc}^{-1}. \tag{2.5}$$

For modes with $k > k_{\rm fs}$, the neutrino density perturbations are erased. This weakens the gravitational potential wells and the growth of cold dark matter (CDM) perturbations is suppressed. Perturbations are free to grow again once their comoving wave numbers fall below $k_{\rm fs}$. Modes with $k < k_{\rm nr}$ are never affected by free-streaming and neutrino perturbations evolve like CDM perturbations. Baryon density perturbations, on the other hand, being pressure supported, can grow in amplitude only after photon decoupling. At the time of photon decoupling, baryons fall into the neutrino-damped dark matter potential wells. Thus, accurate measurements of the amplitude of clustering of matter in the Universe can provide strong upper bounds on the mass of neutrinos.

2.3 Implementing Neutrinos in N-body Simulations

Numerical studies of the effect of neutrinos on the matter distribution have been performed independently by [Brandbyge et al., 2008, Brandbyge and Hannestad, 2009b, 2010] and Viel et al. [2010]. Both groups choose similar cosmological parameters: ($\Omega_{\rm m} = 0.3, \Omega_{\rm b} = 0.05, \Omega_{\rm c} + \Omega_\nu = 0.25, \Omega_\Lambda = 0.7, h = 0.7, n_{\rm s} = 1$), a $512\,h^{-1}$Mpc box and an initial redshift for simulations, $z_{\rm i} = 49$. Brandbyge et al. [2008] and Brandbyge and Hannestad [2009b, 2010] use a weighted sum of the CDM+baryon transfer functions (since they do not have baryons in their simulations) to generate the initial conditions (ICs) for the CDM component using Zel'dovich Approximation (ZA; Zel'dovich [1970])+second-order Lagrangian perturbation theory (2LPT; Scoccimarro [1998]). The Viel et al. [2010] simulations include baryons and use ZA to generate ICs.

Both groups include neutrinos in their N-body simulations either as N-body particles, as a linear grid or use a hybrid method where neutrinos are treated as grid or particles depending on their thermal motion. In the grid-based implementation, the neutrino grid is evolved linearly and does not include the non-linear corrections. The particle-based implementation accounts for the non-linearities by including the coupling between the gravitational potential and neutrinos.

Brandbyge and Hannestad [2009b] (their fig. 1, middle panel) show that the error from neglecting non-linear neutrino perturbations at $z = 0$ is at most 1.25 per cent level at $k \sim 0.25\,h\mathrm{Mpc}^{-1}$ for $\Sigma m_\nu = 0.6\,\mathrm{eV}$. Also, the error between the grid and particle representations is shown to become smaller on small scales. Specifically, the two representations converge for $k \gtrsim 0.2\,h\mathrm{Mpc}^{-1}$. This is attributed to the fact that the neutrino white noise (due to the finite number of neutrino N-body particles) contribution to the matter power spectrum dominates only on ever smaller scales as the CDM perturbations grow at low redshifts. Viel et al. [2010] (their fig. 2, right panel) show that the non-linear correction at $z = 0$ may be as high as 6 per cent at $k \sim 1\,h\mathrm{Mpc}^{-1}$ for $\Sigma m_\nu = 0.6\,\mathrm{eV}$ and the agreement between the grid and particle representations begins to improve only at $k \gtrsim 1\,h\mathrm{Mpc}^{-1}$. The discrepancies between the results from the two groups worsens significantly when the above comparison is done at $z = 1$. These large discrepancies can not be explained solely due to the absence/presence of baryons or whether ZA or ZA+2LPT is used to generate the ICs since (i) the baryons closely trace the CDM distribution on scales $k \lesssim 1\,h\mathrm{Mpc}^{-1}$ and (ii) ZA or ZA+2LPT do not affect the final results significantly when the simulations start at a high redshift ($z_\mathrm{i} = 49$). The extent and the scale-dependence of non-linear neutrino corrections are topics of ongoing research.

2.4 Semi-Analytic Approach to Treat Neutrinos in N-body Simulations

Neutrinos in the mass range $0.05 < \Sigma m_\nu < 1\,\mathrm{eV}$ have present-day free-streaming scales $0.04 < k_{\rm fs} < 0.3\,h\mathrm{Mpc}^{-1}$ ($150 > \lambda_{\rm fs} > 20\,h^{-1}\mathrm{Mpc}$) and thermal velocities $3000 > v_{\rm th} > 450\,\mathrm{km/s}$ respectively. Such large thermal velocities would prevent neutrinos from clustering with CDM and baryons, thereby keeping the neutrino perturbations in the linear regime. As such, in our numerical simulations, we safely assume that the non-linear neutrino perturbations can be ignored and include the linear neutrino perturbations in the ICs only.

To generate the ICs for CDM particles and baryons, we use the publicly available codes CAMB [Lewis et al., 2000] and ENZO[1] [O'Shea et al., 2004, Norman et al., 2007] – an adaptive mesh refinement (AMR), grid-based hybrid code (hydro + N-Body) designed to simulate cosmological structure formation. We use the CAMB code to calculate the linear transfer functions for a given CDM+baryon+neutrino+Λ model. The linear density fluctuation field for CDM particles and baryons is then calculated from their transfer functions using ENZO. The initial positions and velocities for CDM particles and baryon velocities are calculated using the ZA. We do not include neutrinos in our simulations as N-body particles or as a linear grid. Neutrinos enter our simulations only as neutrino-weighted CDM and baryon transfer functions from CAMB.

The linear matter power spectrum $P_{\rm lin}$ can be calculated as the weighted average of the neutrino ($P^{\nu}_{\rm lin}$) and the combined CDM plus baryon($P^{\rm cb}_{\rm lin}$) linear spectra:

$$P_{\rm lin}(k) = \left((f^{\rm c} + f^{\rm b}) \sqrt{P^{\rm cb}_{\rm lin}(k)} + f^{\nu} \sqrt{P^{\nu}_{\rm lin}(k)} \right)^2 , \tag{2.6}$$

where the weights are $f^{\rm i} = \Omega_{\rm i}/\Omega_{\rm m}$ and $\Omega_{\rm m} = \Omega_{\rm b} + \Omega_{\rm c} + \Omega_\nu$. The CDM plus baryon

[1] http://lca.ucsd.edu/projects/enzo

power spectrum is

$$P_{\rm lin}^{\rm cb}(k) = (f^{\rm c} + f^{\rm b})^{-2} \left(f^{\rm c} \sqrt{P_{\rm lin}^{\rm c}(k)} + f^{\rm b} \sqrt{P_{\rm lin}^{\rm b}(k)} \right)^2 , \tag{2.7}$$

where $P_{\rm lin}^{\rm c}$ and $P_{\rm lin}^{\rm b}$ are the linear CDM and baryon power spectra respectively. Throughout this work, the subscripts 'lin' and 'nl' will indicate quantities in the linear and non-linear regimes, respectively. On smaller scales the matter perturbations have gone non-linear. So, the non-linear matter power spectrum $P_{\rm nl}$ becomes

$$P_{\rm nl}(k) = \left((f^{\rm c} + f^{\rm b}) \sqrt{P_{\rm nl}^{\rm cb}(k)} + f^{\nu} \sqrt{P_{\rm lin}^{\nu}(k)} \right)^2 , \tag{2.8}$$

where,

$$P_{\rm nl}^{\rm cb}(k) = (f^{\rm c} + f^{\rm b})^{-2} \left(f^{\rm c} \sqrt{P_{\rm nl}^{\rm c}(k)} + f^{\rm b} \sqrt{P_{\rm nl}^{\rm b}(k)} \right)^2 . \tag{2.9}$$

In Eq. 2.8, we calculate $P_{\rm nl}^{\rm cb}$ at redshift $z = 0$ from N-body simulations and combine it with $P_{\rm lin}^{\nu}$ at $z = 0$ as solved by the CAMB code to construct $P_{\rm nl}$. We do not account for the non-linear neutrino corrections in Eq. 2.8. Saito et al. [2009b] studied the non-linear neutrino perturbations using the higher-order perturbation theory (PT) to show that for low neutrino fractions ($f^{\nu} \lesssim 0.05$), the amplitude of the non-linear matter power spectrum increases by $\lesssim 0.01$ per cent at $k \sim 0.2\,h{\rm Mpc}^{-1}$ at $z = 3$ and by $\lesssim 0.15$ per cent at $k \sim 0.1\,h{\rm Mpc}^{-1}$ at $z = 0$. Since at $z = 0$, PT is expected to reproduce the N-body simulation results within 1 per cent – only up to $k \lesssim 0.1 - 0.15\,h{\rm Mpc}^{-1}$ (see Taruya et al. [2009]), the non-linear neutrino corrections at $z = 0$ may be somewhat larger on scales we probe in our simulations ($0.1 \leq k \leq 1\,h{\rm Mpc}^{-1}$) – the estimate of which requires multiple particle (CDM+baryon+neutrino) simulations.

2.5 N-body Simulations: Optimizing Boxsize and Number of Particles

We performed N-body simulations with the ENZO code. The code allows us to choose the geometry (box size, number of particles), the normalized densities of matter, baryon,

Box size (h^{-1}Mpc)	$N_{\rm cdm}$	$N_{\rm gas}$	Ω_ν
200	64^3	512^3	0.00
200	128^3	512^3	0.00
200	256^3	512^3	0.00
200	256^3	512^3	0.001
200	256^3	512^3	0.002
200	256^3	512^3	0.01
200	256^3	512^3	0.02
200	256^3	512^3	0.04
100	256^3	512^3	0.00
200	512^3	512^3	0.00
200	512^3	512^3	0.01
200	512^3	512^3	0.02
200	512^3	512^3	0.04

Table 1: Simulation parameters. All simulations were started at a redshift of $z_{\rm i} = 20$ and stopped at $z = 0$. We ran eight independent simulations for each row to suppress sampling variance.

CDM, neutrino and cosmological constant ($\Omega_{\rm m}, \Omega_{\rm b}, \Omega_{\rm c}, \Omega_\nu, \Omega_\Lambda$), the amplitude of fluctuation on $8\,h^{-1}$ Mpc scale: σ_8, the primordial spectral index: $n_{\rm s}$ and the initial redshift: $z_{\rm i}$. We kept AMR off (no adaptive mesh refinement) since it does not significantly affect the scales of interest. Throughout this chapter we work with the 7-yr *WMAP* data alone [Larson et al., 2011] central parameters: $\Omega_{\rm m} = 0.266$, $\Omega_{\rm b} = 0.044$, $\Omega_\Lambda = 0.734$, $h = 0.71$ and $n_{\rm s} = 0.963$. We vary Ω_ν such that $\Omega_{\rm c} + \Omega_\nu = 0.222$. The simulation parameters are listed in Table 1. In order to suppress sampling variance of the estimated power spectrum, for each row we ran eight simulations by changing the seed to generate the ICs.

First, we had to select an appropriate geometry (box size and the number of CDM/gas particles) for which the matter power spectrum converges at per cent level in the quasi-non-linear regime ($0.1 \lesssim k \lesssim 1\,h{\rm Mpc}^{-1}$). The largest mode that can fit in a $200\,h^{-1}$Mpc box is $k \sim 0.03\,h{\rm Mpc}^{-1}$ and the matter power spectrum is sufficiently linear on these scales. One can choose bigger volumes but unless the number of particles is also in-

creased accordingly, it leads to a poor mass resolution. Also, N-body simulations suffer from a discreteness problem that arises due to the finite number of macroparticles used to sample the matter distribution in the universe. Thus, given any theoretical cosmological model, the ICs are always undersampled.

The smallest scale for which the power spectrum can be resolved accurately is related to the Nyquist wavenumber, $k_{\rm Ny}$, given by:

$$k_{\rm Ny} = \frac{\pi (N_{\rm part})^{1/3}}{L_{\rm Box}}. \tag{2.10}$$

Given a combination of the number of particles and the box size, the power spectrum is dominated by shot noise for $k \gtrsim k_{\rm Ny}$. For $N_{\rm cdm} = 64^3$ particles in a $200\,h^{-1}$Mpc box, $k_{\rm Ny}$ is $1.01\,h$Mpc^{-1}, while the modes we aim to probe through simulations are $0.1 \lesssim k \lesssim 1\,h$Mpc^{-1}. Thus $N_{\rm cdm} = 64^3$ particles in a $200\,h^{-1}$Mpc box seems a reasonable combination to start with.

The number of gas particles fixes the root grid that determines the force resolution for the simulation. ENZO uses a particle mesh technique to calculate the gravitational potential on the root grid [O'Shea et al., 2005]. Forces are first computed on the mesh by finite-differencing the gravitational potential and then interpolated to the dark matter particle positions to update the particle's position and velocity information. This methodology requires that the root grid be at least twice as fine as the mean interparticle separation to obtain accurate forces down to the scale of the mean interparticle spacing. A coarse root grid renders the forces on the scale of the mean interparticle spacing, inaccurate.

Fig. 2.1 shows the matter power spectrum at $z = 0$ when $N_{\rm cdm} = 64^3$, 128^3, 256^3 and 512^3 particles are used to sample the ICs ($\Omega_\nu = 0$ for all four cases). Beyond the Nyquist wavenumbers, represented by vertical lines [64^3 – solid (red), 128^3 – long dash-dotted (green), 256^3 – dashed (blue) and 512^3 – long-dashed (cyan)], the power spectra become increasingly inaccurate due to particle shot noise contribution. Fig. 2.2 shows

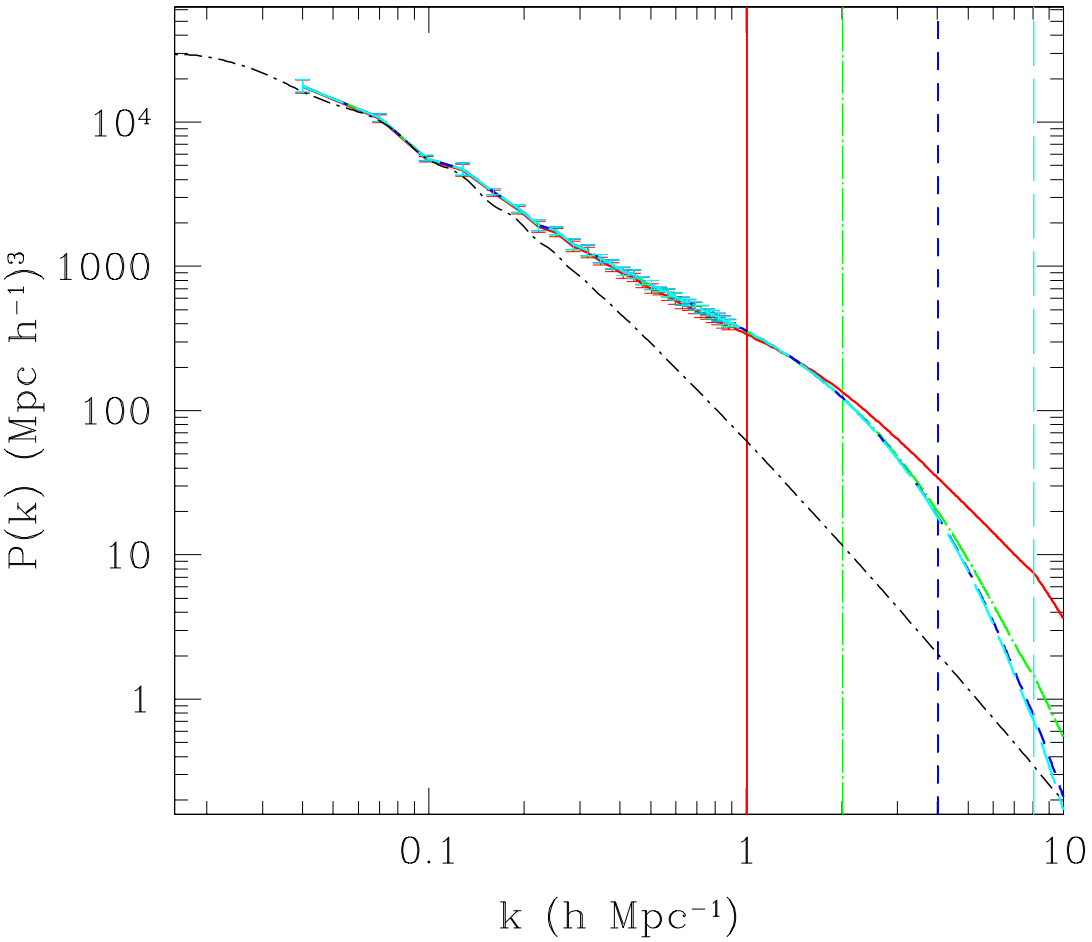

Figure 2.1: Matter power spectrum at $z = 0$ for undersampled ICs at $z_{\rm i} = 20$ with $N_{\rm cdm} = 64^3$ – solid (red), 128^3 – long dash-dotted (green), 256^3 – dashed (blue) and 512^3 – long-dashed (cyan). The vertical lines are the $k_{\rm Ny}$ wavenumbers for 64^3, 128^3, 256^3 and 512^3 CDM particles. Also plotted (dash-dotted line) is the linear theoretical power spectrum. For $k > k_{\rm Ny}$, particle shot noise dominates the true power spectrum.

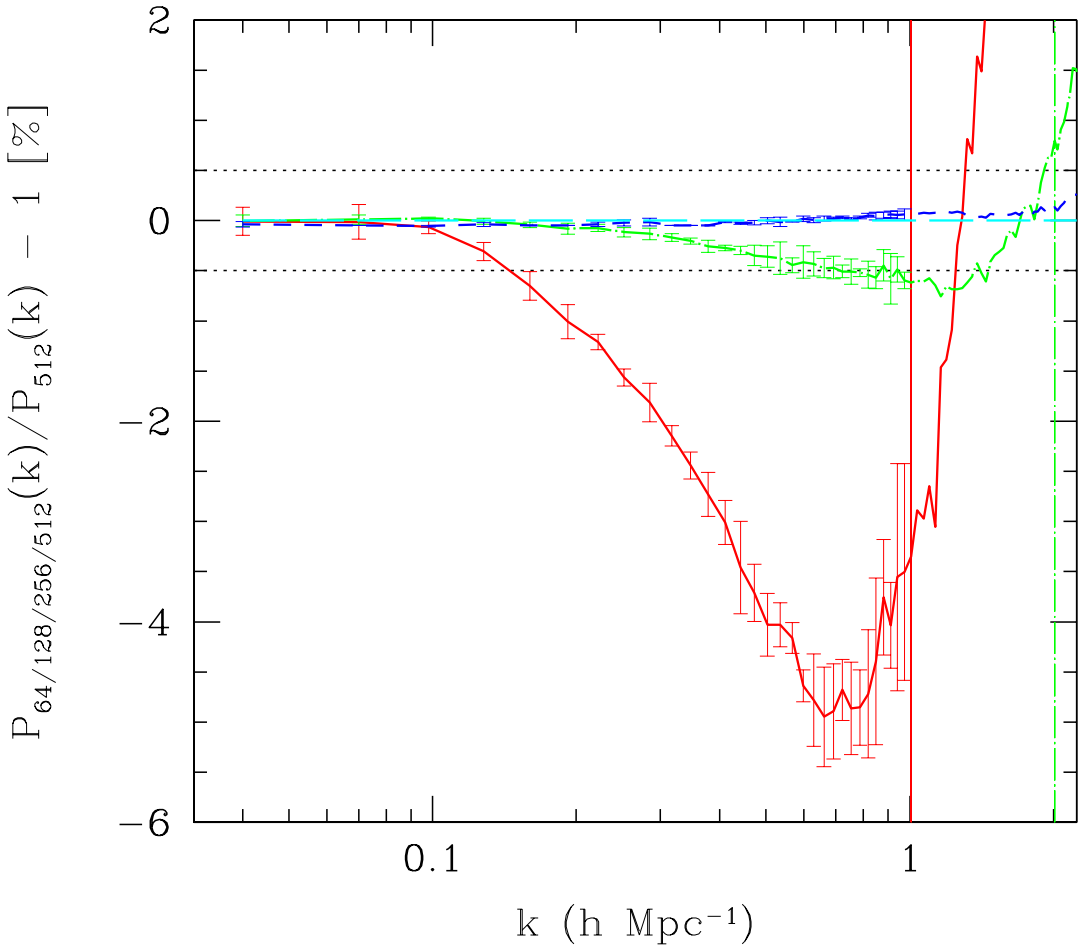

Figure 2.2: Same as Fig. 2.1 expressed as fractional suppression of the matter power spectrum at $z = 0$ when 64^3 – solid (red), 128^3 – long dash-dotted (green) and 256^3 – dashed (blue) CDM particles are used to sample the ICs w.r.t the case where 512^3 – long-dashed (cyan) CDM particles are used. $\Omega_\nu = 0$ for all four cases.The error bars correspond to eight simulations with different seeds for the ICs.

the fractional suppression of the matter power spectrum at $z = 0$. For $k \lesssim 1\,h\mathrm{Mpc}^{-1}$, the error due to undersampling the ICs is $\lesssim 5$ per cent for the 64^3 run, $\lesssim 0.5$ per cent for the 128^3 run and negligibly small for the 256^3 run. To keep the undersampling error at $k = 1\,h\mathrm{Mpc}^{-1}$ below 0.5 per cent, we narrowed down to a combination of $N_{\mathrm{cdm}} = 256^3$, $N_{\mathrm{gas}} = 512^3$ in a $200\,h^{-1}$Mpc box to investigate the effect of massive neutrinos on the matter power spectrum in the regime $0.1 \leq k \leq 1\,h\mathrm{Mpc}^{-1}$. Finally, we checked the smallest scales that are accurately resolved by the $200\,h^{-1}$Mpc box. Towards this, we ran eight simulations in a $100\,h^{-1}$Mpc box with $N_{\mathrm{cdm}} = 256^3$, $N_{\mathrm{gas}} = 512^3$. In Fig. 2.3, we plot the power spectrum from 100 and $200\,h^{-1}$Mpc boxes. The matter power spectrum from $100\,h^{-1}$Mpc box simulations begins to show excess power for $k \gtrsim 1\,h\mathrm{Mpc}^{-1}$. The non-linear evolution of perturbations on scales $k \gtrsim 1\,h\mathrm{Mpc}^{-1}$ is missed in the $200\,h^{-1}$Mpc box simulations. The spectrum from $200\,h^{-1}$Mpc box simulations shows convergence at per cent level for $k \lesssim 1\,h\mathrm{Mpc}^{-1}$ (Fig. 2.4).

2.6 Impact of Massive Neutrinos on Structural Growth

The contribution of massive neutrinos to the present-day critical energy density is given by:

$$\Omega_\nu = \frac{\Sigma m_\nu}{94.22 h^2}, \tag{2.11}$$

where Σm_ν is the sum of the masses of all neutrino eigentstates. In this section we consider four neutrino models: $\Omega_\nu = 0$, 0.01, 0.02 and 0.04 corresponding to $\Sigma m_\nu =$ 0, 0.475, 0.95 and 1.9 eV, respectively. We assume three degenerate neutrino eigentstates, so that $m_\nu = \Sigma m_\nu / 3$.

In Fig. 2.5 we show slices of the baryon density field at $z = 0$ extracted from $200\,h^{-1}$Mpc box with $N_{\mathrm{cdm}} = 256^3$, $N_{\mathrm{gas}} = 512^3$. The top panel is from a simulation without neutrinos, the middle and the bottom panels correspond to simulations with $\Omega_\nu = 0.02$ and 0.04 respectively. All slices are $200\,h^{-1}$Mpc wide. The slices show the

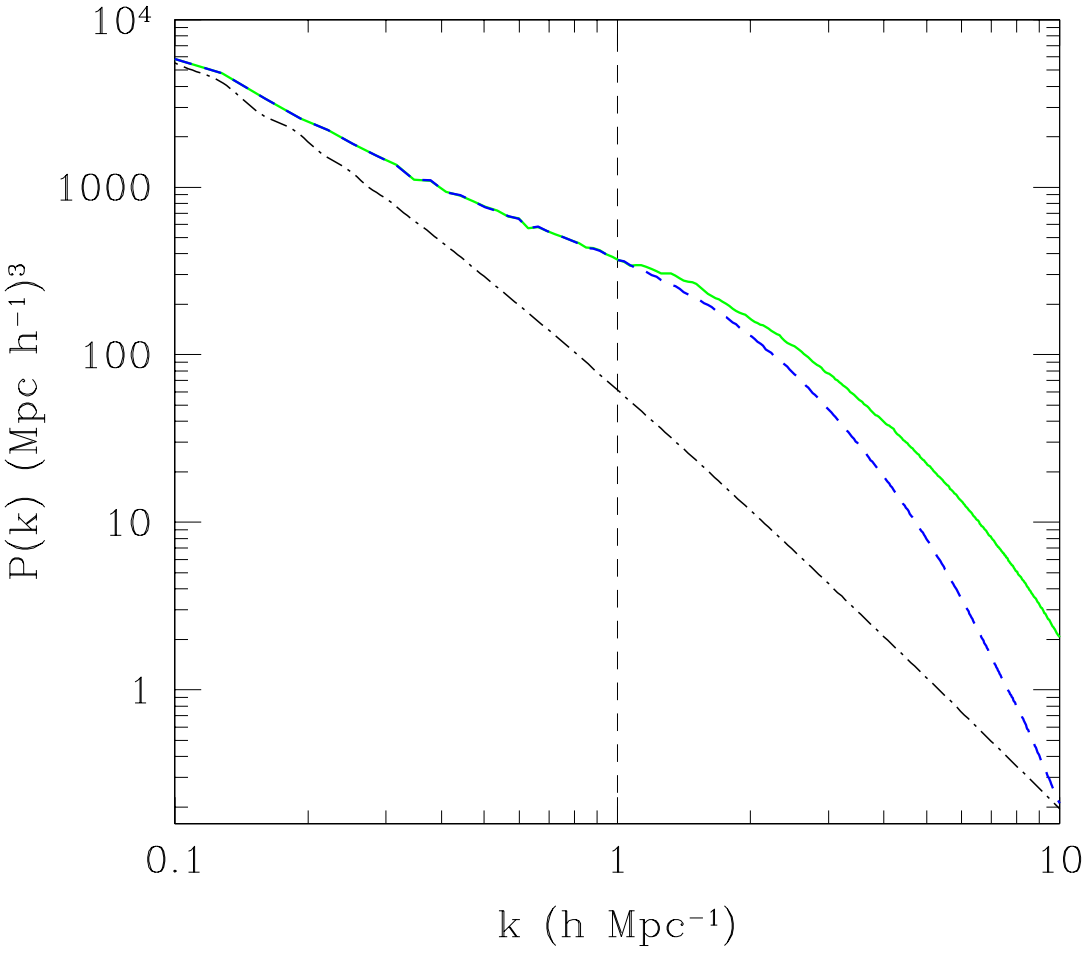

Figure 2.3: Matter power spectrum at $z = 0$ from $100\,h^{-1}$Mpc − solid (green) and $200\,h^{-1}$Mpc − dashed (blue) box simulations. The linear theory spectrum (dash-dotted line) is also shown. The vertical dashed line is the maximum wavenumber up to which the power spectrum from $200\,h^{-1}$Mpc box simulations can be trusted at per cent level.

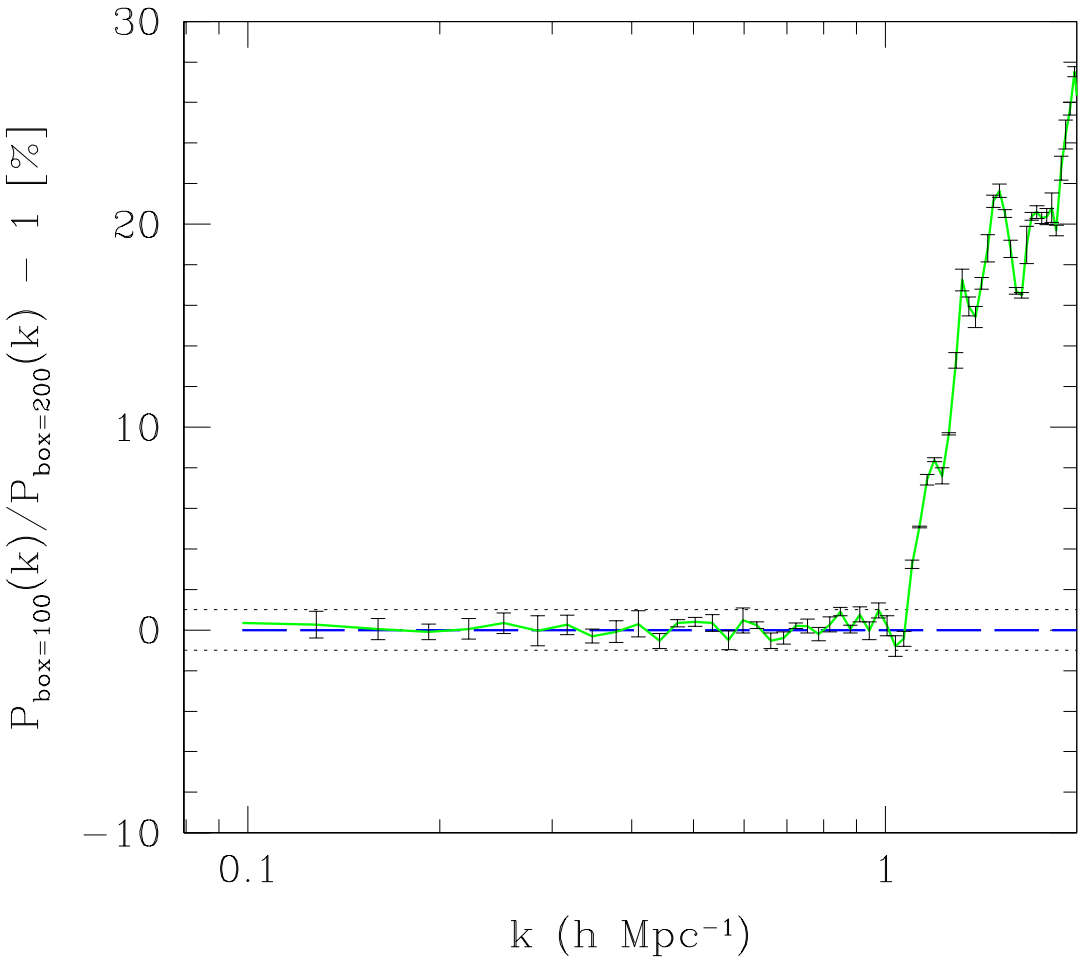

Figure 2.4: Same as Fig. 2.3 expressed as fractional suppression of the matter power spectrum at $z = 0$ as a function of the box size. Spectrum from $100\,h^{-1}$Mpc – solid (green) and $200\,h^{-1}$Mpc – dashed (blue) box agree at per cent level for $k \lesssim 1\,h$Mpc^{-1}.

baryonic mass averaged over the volume of a grid cell. Each grid cell in our simulations is $\sim 391\,h^{-1}$kpc.

As neutrinos become more massive, the suppression in the growth of density perturbations becomes clear by the relatively diffused density filaments. The baryon density fields in the middle and the bottom panels are less evolved relative to the massless neutrino (top panel) case. The gravitational potential wells are much deeper in the top panel. This is evident from the voids (dark blue regions) which are more underdense in the top panel compared to the voids in the lower panels. To quantify the difference between simulations with and without massive neutrinos, we measure the matter power spectrum by converting the positions of the CDM and gas particles into 512^3-point grids of densities using a Cloud-In-Cell (CIC) interpolation scheme. We do not compensate for the smoothing effect introduced by the CIC filtering since the smoothing affects scales that are close to the Nyquist wavenumber which for our choice of parameters ($N_{\rm gas}=512^3$, Box=$200\,h^{-1}$Mpc) is $k_{\rm Ny} = 8.04\,h{\rm Mpc}^{-1}$, while the quasi-non-linear modes of interest are $0.1 \lesssim k \lesssim 1\,h{\rm Mpc}^{-1}$. The density fields are fast Fourier transformed to calculate $P^{\rm b}_{\rm nl}(k)$ and $P^{\rm c}_{\rm nl}(k)$ – the non-linear power spectrum for baryons and CDM respectively. We then construct the non-linear matter power spectrum $P_{\rm nl}(k)$ at $z = 0$ using Eqs 2.8 and 2.9. To suppress sampling variance of the estimated $P_{\rm nl}(k)$, we take the average $P_{\rm nl}(k)$ from eight independent realizations.

Fig. 2.6 shows the matter power spectrum at $z = 0$ from simulations and linear theory (dash-dotted lines) as a function of neutrino mass for the four neutrino models: $\Omega_\nu = 0\,(\Sigma m_\nu = 0\,{\rm eV})$ – solid (red), $\Omega_\nu = 0.01\,(\Sigma m_\nu = 0.475\,{\rm eV})$ – long dash-dotted (green), $\Omega_\nu = 0.02\,(\Sigma m_\nu = 0.95\,{\rm eV})$ – dashed (blue) and $\Omega_\nu = 0.04\,(\Sigma m_\nu = 1.9\,{\rm eV})$ – long-dashed (cyan). The simulation spectra are significantly above the linear theory predictions at high k. The linear theory predictions break down for $k \gtrsim 0.1\,h{\rm Mpc}^{-1}$ ($\lambda \lesssim 60\,h^{-1}$Mpc). Also, as the total neutrino mass is increased (keeping the number of degenerate neutrino eigentstates fixed at three), the matter power spectrum is further suppressed. Since neutrino eigentstates with higher mass constitute a larger fraction of

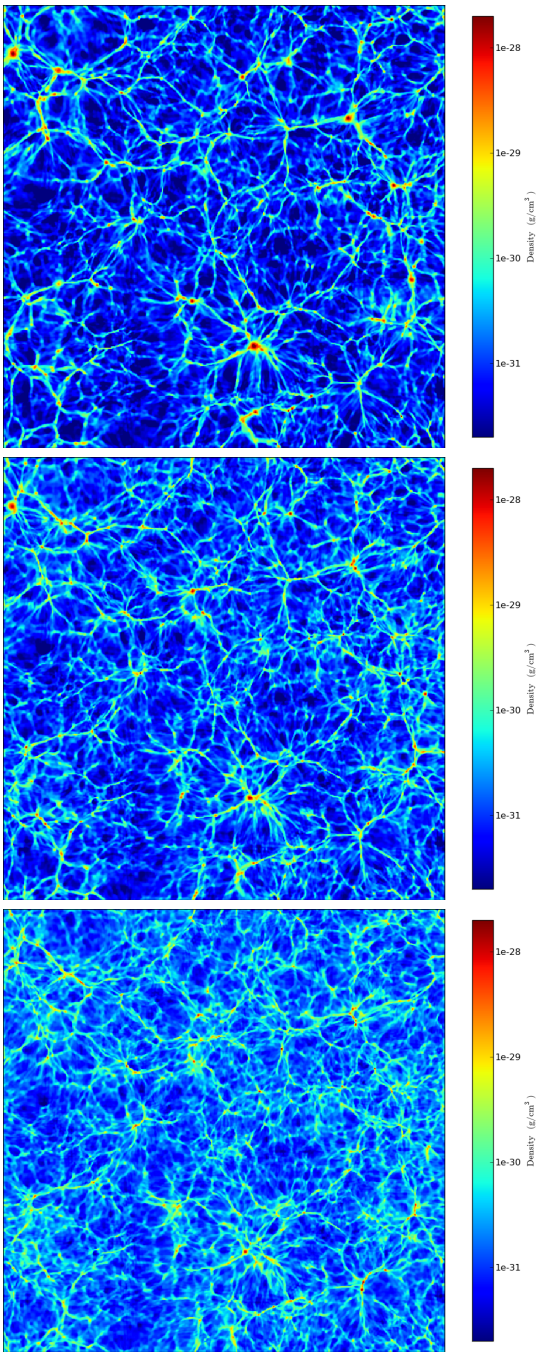

Figure 2.5: Slices of baryon density distribution. All slices are $200\,h^{-1}$Mpc wide and show the baryonic mass averaged over the volume of a grid cell. Each grid cell is $\sim 391\,h^{-1}$kpc. The top panel shows a simulation without neutrinos. The middle and the bottom panels are taken from simulations with $\Omega_\nu = 0.02\,(\Sigma m_\nu = 0.95\,\mathrm{eV})$ and $\Omega_\nu = 0.04\,(\Sigma m_\nu = 1.9\,\mathrm{eV})$. The baryon density fields in the middle and the bottom panels are less evolved relative to the no-neutrino (top panel) case. The simulations were run with $N_{\mathrm{cdm}} = 256^3$, $N_{\mathrm{gas}} = 512^3$. The density projections were made using YT: an analysis and visualization tool [Turk, 2008].

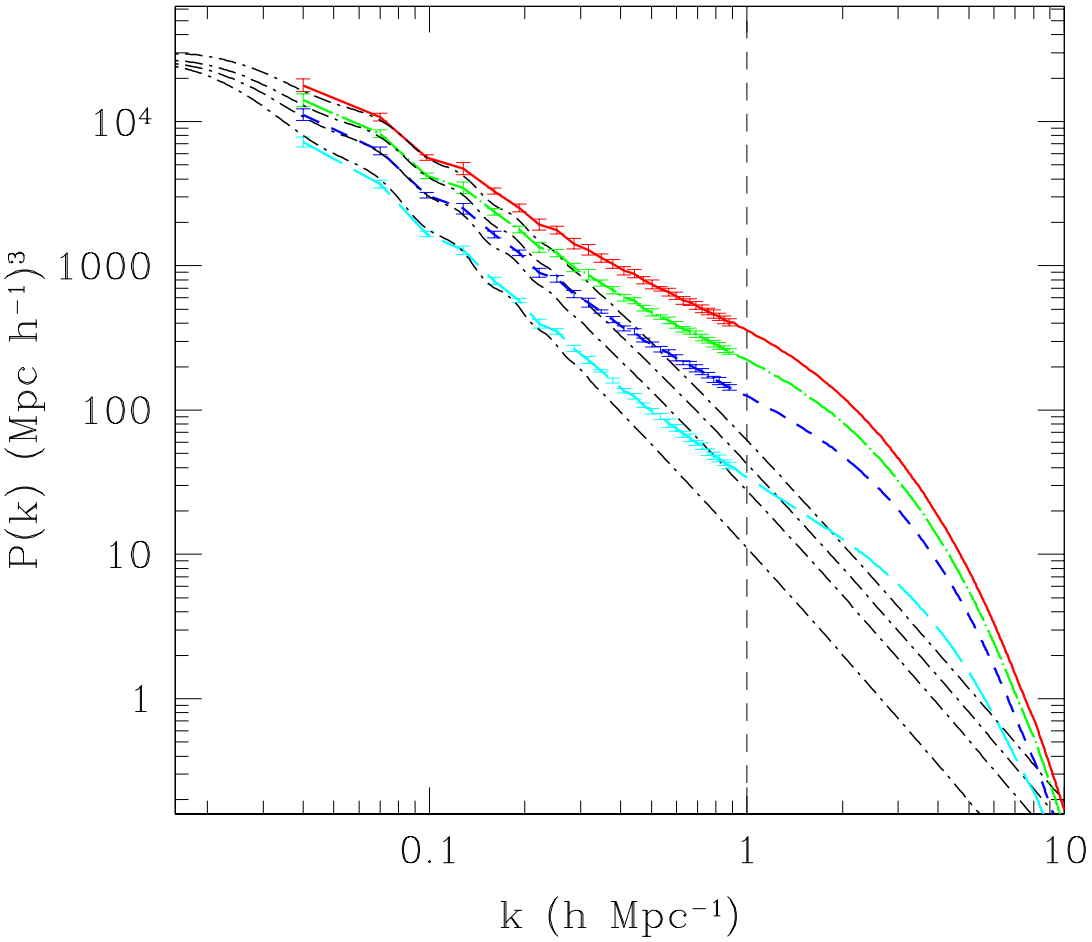

Figure 2.6: Matter power spectrum at $z = 0$ from simulations and linear theory (dash-dotted lines) as a function of neutrino mass. The four neutrino models are: $\Omega_\nu = 0\,(\Sigma m_\nu = 0\,\mathrm{eV})$ – solid (red), $\Omega_\nu = 0.01\,(\Sigma m_\nu = 0.475\,\mathrm{eV})$ – long dash-dotted (green), $\Omega_\nu = 0.02\,(\Sigma m_\nu = 0.95\,\mathrm{eV})$ – dashed (blue) and $\Omega_\nu = 0.04\,(\Sigma m_\nu = 1.9\,\mathrm{eV})$ – long-dashed (cyan). The vertical dashed line is the maximum wavenumber up to which the power spectra from $200\,h^{-1}$Mpc box simulations are valid at 1 per cent level.

the total energy density, they are more effective in damping small-scale power than low mass neutrinos.

In Fig. 2.7 we plot the fractional difference between the matter power spectra with and without massive neutrinos, from the simulations as well as the linear theory predictions. The linetypes for the spectra are the same as in Fig. 2.6. The linear theory predicts a nearly scale-independent suppression for $k \gtrsim 0.2\,h\mathrm{Mpc}^{-1}$. On the other other hand, the non-linear power spectra from the simulations show an enhanced suppression for $k \gtrsim 0.1\,h\mathrm{Mpc}^{-1}$. At $k \sim 1\,h\mathrm{Mpc}^{-1}$, the non-linear spectra are ~ 10 per cent more suppressed compared to the corresponding linear spectra.

2.7 Resolving Neutrino Mass Hierarchy from Numerical Simulations

The mass splittings of $|\Delta m^2_{32}| = (2.43 \pm 0.13) \times 10^{-3}\,\mathrm{eV}^2$ and $\Delta m^2_{21} = (7.59 \pm 0.21) \times 10^{-5}\,\mathrm{eV}^2$ [Adamson et al., 2008, KamLAND, 2008] allow for two possible neutrino mass hierarchies: normal ($m_3 > m_2 > m_1$) and inverted ($m_2 > m_1 > m_3$). For $\Sigma m_\nu > 0.4 - 0.5\,\mathrm{eV}$, all neutrino eigentstates are essentialy degenerate, the mass of each eigentstate being $m_\nu \approx \Sigma m_\nu/3$. However, for smaller Σm_ν, the individual eigentstate masses differ significantly in the normal and inverted hierarchies. The free-streaming comoving wave number, k_{nr}, is a function of the mass of each neutrino eigentstate (see Eqs 2.4 and 2.5). As the mass is increased, it becomes non-relativistic earlier and the free-streaming scale gets shorter. The mass dependence of k_{nr} means that the matter power spectrum is modified differently for eigentstates with different masses. This makes the matter power spectrum a powerful tool to distinguish between the normal and inverted hierarchies. In this section we discuss the precision levels above which the power spectrum from future galaxy surveys would be able to resolve between the two mass hierarchies.

The mass splittings of $|\Delta m^2_{32}| = (2.43 \pm 0.13) \times 10^{-3}\,\mathrm{eV}^2$ and $\Delta m^2_{21} = (7.59 \pm 0.21) \times 10^{-5}\,\mathrm{eV}^2$ imply that the lower bounds on the total neutrino mass are $\Sigma m_\nu = 0.05$ and $0.1\,\mathrm{eV}$ for the normal and inverted mass hierarchies respectively. We performed numerical simulations for $\Sigma m_\nu = 0.05$ and $0.1\,\mathrm{eV}$. For $\Sigma m_\nu = 0.05\,\mathrm{eV}$, we assumed 1 massive

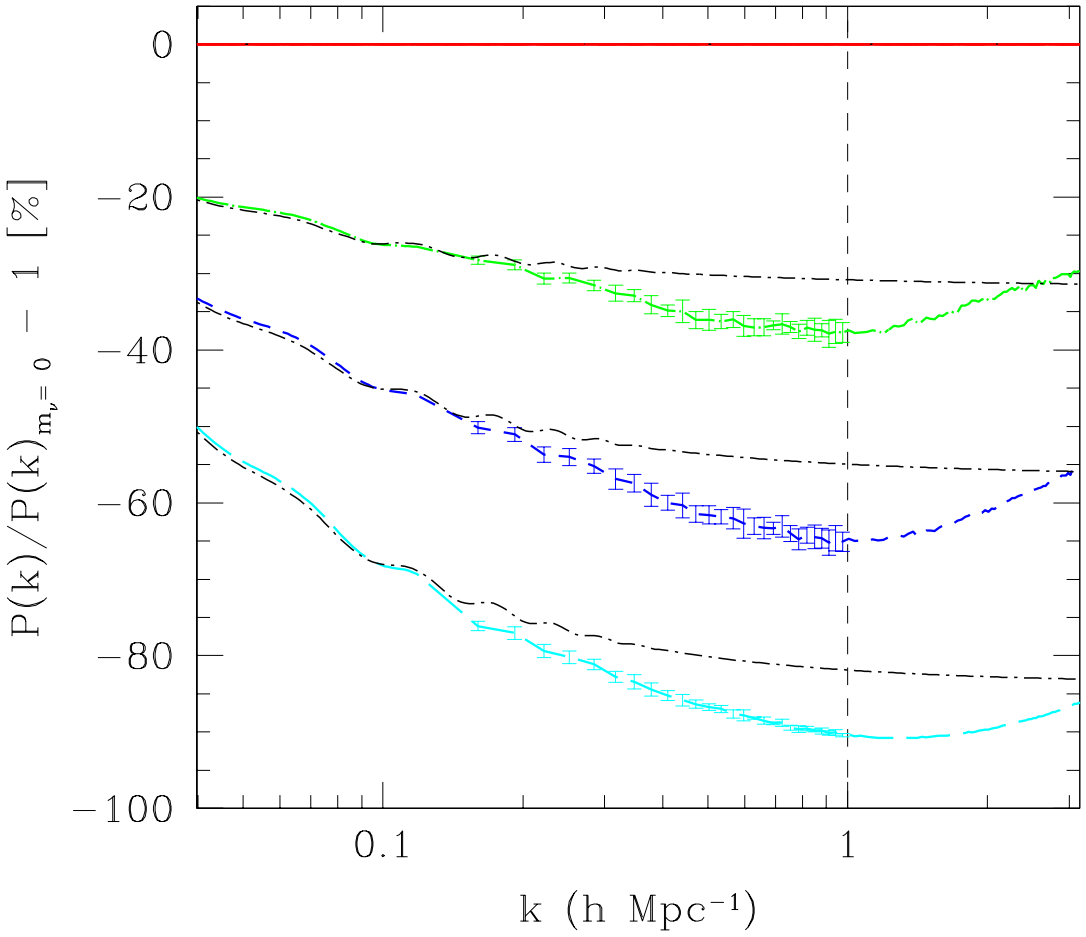

Figure 2.7: Fractional difference between the matter power spectra with and without massive neutrinos at $z = 0$, from the simulations and the linear theory predictions (dash-dotted lines). The four neutrino models are: $\Omega_\nu = 0$ ($\Sigma m_\nu = 0\,\mathrm{eV}$) – solid (red), $\Omega_\nu = 0.01$ ($\Sigma m_\nu = 0.475\,\mathrm{eV}$) – long dash-dotted (green), $\Omega_\nu = 0.02$ ($\Sigma m_\nu = 0.95\,\mathrm{eV}$) – dashed (blue) and $\Omega_\nu = 0.04$ ($\Sigma m_\nu = 1.9\,\mathrm{eV}$) – long-dashed (cyan). The error bars correspond to eight simulations with different seeds for the ICs.

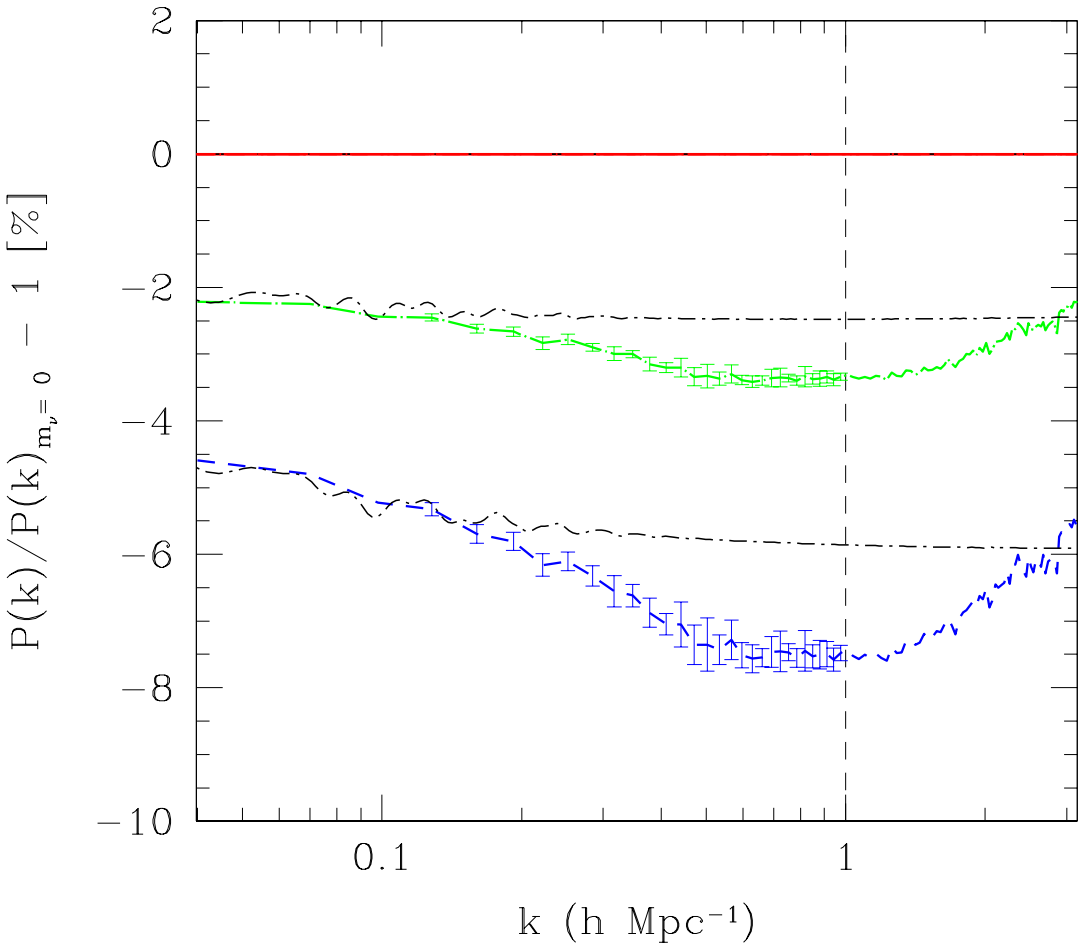

Figure 2.8: Same as Fig. 2.7, but for neutrino models with much lower neutrino mass: $\Omega_\nu = 0.001$ ($\Sigma m_\nu = 0.05\,\mathrm{eV}$) – long dash-dotted (green) and $\Omega_\nu = 0.002$ ($\Sigma m_\nu = 0.1\,\mathrm{eV}$) – dashed (blue).

and 2 massless eigentstates (mimicking the normal hierarchy). For $\Sigma m_\nu = 0.1\,\mathrm{eV}$, we assumed 2 massive and 1 massless eigentstate (mimicking the inverted hierarchy). In Fig. 2.8, we show the fractional suppression in the power spectrum for two neutrino models: $\Omega_\nu = 0.001$ ($\Sigma m_\nu = 0.05\,\mathrm{eV}$) – long dash-dotted (green) and $\Omega_\nu = 0.002$ ($\Sigma m_\nu = 0.1\,\mathrm{eV}$) – dashed (blue). The growth of structure formation is suppressed by as much as 3.5 per cent (7.5 per cent) at $k \sim 0.6\,h\mathrm{Mpc}^{-1}$ for the two models. The measurement errors in the power spectrum from future galaxy surveys are expected to be at the 1 per cent level. In case future surveys constrain $\Sigma m_\nu < 0.1\,\mathrm{eV}$ with sufficient precision, that would rule out the inverted mass hierarchy. The current constraint from the 7-yr *WMAP* data alone [Larson et al., 2011] is $\Sigma m_\nu < 1.3\,\mathrm{eV}$ (95 per cent CL). At this level, it is not possible to discriminate between the normal and inverted hierarchies since all eigentstates are essentially degenerate.

Next, we consider a scenario with $\Sigma m_\nu = 0.1\,\mathrm{eV}$, at which the difference between the normal and inverted hierarchies is most prominent. We ran N-body simulations in the following three ways: (i) ($N_{\mathrm{massive}} = 3$, $N_{\mathrm{degen}} = 3$) where N_{massive} is the number of massive eigentstates and N_{degen} is the degeneracy amongst the massive eigentstates. This combination corresponds to $m_\nu = \Sigma m_\nu/3 = 0.033\,\mathrm{eV}$; (ii) ($N_{\mathrm{massive}} = 2$, $N_{\mathrm{degen}} = 2$), this is the inverted hierarchy scenario with one massless and two equally massive eigentstates ($m_\nu \sim 0.05, 0.05, 0\,\mathrm{eV}$); (iii) ($N_{\mathrm{massive}} = 3$, $N_{\mathrm{degen}} = 2$), this is the normal hierarchy scenario with three massive eigentstates ($m_\nu \sim 0.056, 0.022, 0.022\,\mathrm{eV}$). Note that case (i) is meaningless at $\Sigma m_\nu = 0.1\,\mathrm{eV}$ given that $|\Delta m^2_{32}| = (2.43 \pm 0.13) \times 10^{-3}\,\mathrm{eV}^2$ and $\Delta m^2_{21} = (7.59 \pm 0.21) \times 10^{-5}\,\mathrm{eV}^2$. We include case (i) for illustrative purposes only.

In Fig. 2.9, we plot the matter power spectrum for cases (i), (ii) and (iii) divided by the spectrum for case (i). The linear theory predictions are shown by dash-dotted lines. Since non-linearities become important only for $k \gtrsim 0.1\,h\mathrm{Mpc}^{-1}$, we have plotted the theoretical power spectrum for $k < 0.1\,h\mathrm{Mpc}^{-1}$, calculated using the CAMB code. The suppression from simulations is $\sim 0.05 - 0.2$ per cent higher than the linear predictions. The inverted hierarchy - dashed line (green) shows excess power for wavenumbers $0.001 <$

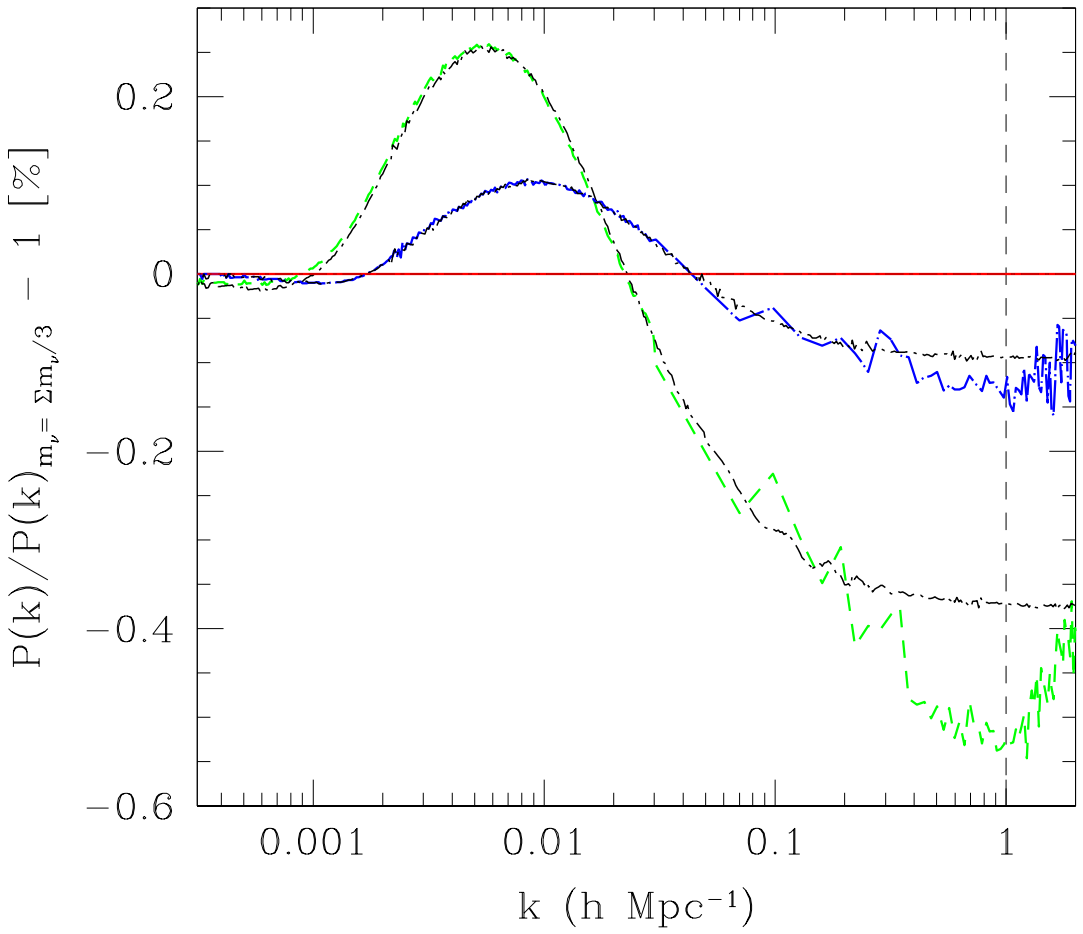

Figure 2.9: Matter power spectrum for normal – long dash-dotted line (blue) and inverted – dashed line (green) hierarchies divided by the matter power spectrum for $m_\nu = \Sigma m_\nu/3$ – solid line (red). The linear theory predictions are shown by dash-dotted lines. The neutrino model considered here is $\Sigma m_\nu = 0\,\mathrm{eV}$. The individual masses for the three eigentstates are ($m_\nu \sim 0.05, 0.05$ and $0\,\mathrm{eV}$) for the inverted hierarchy and ($m_\nu \sim 0.056, 0.022$ and $0.022\,\mathrm{eV}$) for the normal hierarchy. The inverted hierarchy shows more damping of small-scale power than the normal hierarchy.

$k < 0.02\,h\mathrm{Mpc}^{-1}$ and an enhanced suppression of ~ 0.5 per cent at $k \sim 1\,h\mathrm{Mpc}^{-1}$ relative to case (i). This can be explained by the fact that in case (ii) $\Sigma m_\nu = 0.1\,\mathrm{eV}$ is shared equally between two eigentstates, while in case (i) $\Sigma m_\nu = 0.1\,\mathrm{eV}$ is shared equally between three eigentstates. Each eigentstate is more massive in case (ii), thereby making the free-streaming length shorter compared to that in case (i). Higher mass neutrinos are better at wiping out small-scale perturbations and their shorter free-streaming length implies that the spatial extent of damping is limited.

Another factor contributing to the appearance of Fig. 2.9 is a shift in the radiation–matter equality redshift. Higher mass neutrinos become non-relativistic at higher redshifts and start contributing to Ω_m before low mass neutrinos do. This shifts the radiation–matter equality epoch to a higher redshift and reduces the scale corresponding to the one that entered the horizon at radiation–matter equality. The modes entering the horizon after radiation–matter equality grow linearly (as opposed to logarithmically during the radiation era) which contributes to the excess power [compare dashed (green) and solid (red) lines in Fig. 2.9] for wavenumbers $0.001 < k < 0.02\,h\mathrm{Mpc}^{-1}$. The same reasoning can be applied to the normal hierarchy – long dash-dotted line (blue). At $\Sigma m_\nu = 0.1\,\mathrm{eV}$, precision better than 0.5 per cent would be needed in measuring the matter power spectrum to discriminate between the normal and inverted hierarchies. For $\Sigma m_\nu > 0.2\,\mathrm{eV}$ all eigentstates become degenerate, this would make it extremely difficult for a future survey to resolve the two hierarchies.

2.8 Comparison: Semi-Analytic versus Full Numerical Treatment

In this section we compare the estimated overall suppression of the matter power spectrum due to massive neutrinos from our N-body simulations with the results obtained by Brandbyge et al. [2008] and Viel et al. [2010]. In linear theory, the suppression of the matter power spectrum amplitude is approximately given by $\Delta P/P \sim -8 f_\nu$ [Hu et al., 1998]. Numerical simulations, however, show that the neutrino suppression is enhanced in the non-linear regime ($k \gtrsim 0.1\,h\mathrm{Mpc}^{-1}$). In Fig. 2.10 we plot

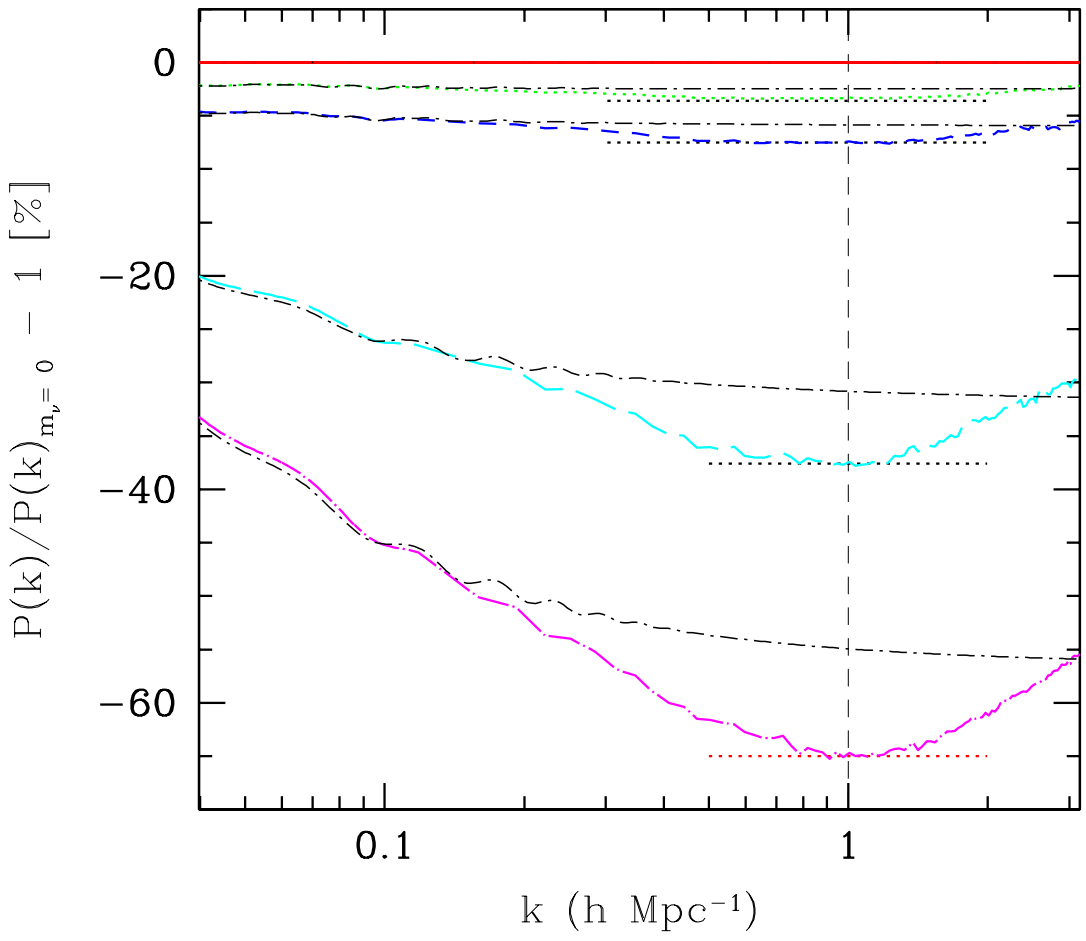

Figure 2.10: Fractional difference between the matter power spectra with and without massive neutrinos at $z = 0$, from numerical simulations and linear theory predictions (dash-dotted lines). The four neutrino models are: $\Omega_\nu = 0.001$ ($\Sigma m_\nu = 0.05\,\mathrm{eV}$) – dotted (green), $\Omega_\nu = 0.002$ ($\Sigma m_\nu = 0.1\,\mathrm{eV}$) – dashed (blue), $\Omega_\nu = 0.01$ ($\Sigma m_\nu = 0.475\,\mathrm{eV}$) – long-dashed (cyan) and $\Omega_\nu = 0.02$ ($\Sigma m_\nu = 0.95\,\mathrm{eV}$) – long dash-dotted (magenta). The maximum relative suppression of $\Delta P/P \sim -10 f_\nu$ is shown as short horizontal dotted lines. The horizontal (red) dotted line for $\Sigma m_\nu = 0.95\,\mathrm{eV}$ is at $\Delta P/P \sim -8.6 f_\nu$.

the fractional difference between the matter power spectra with and without massive neutrinos at $z = 0$, from numerical simulations as well as linear theory predictions (dash-dotted lines) for four neutrino models: $\Omega_\nu = 0.001\,(\Sigma m_\nu = 0.05\,\mathrm{eV})$ – dotted (green), $\Omega_\nu = 0.002\,(\Sigma m_\nu = 0.1\,\mathrm{eV})$ – dashed (blue), $\Omega_\nu = 0.01\,(\Sigma m_\nu = 0.475\,\mathrm{eV})$ – long-dashed (cyan) and $\Omega_\nu = 0.02\,(\Sigma m_\nu = 0.95\,\mathrm{eV})$ – long dash-dotted (magenta). We found a maximum non-linear suppression of $\Delta P/P \sim -10 f_\nu$ for neutrino masses $\Sigma m_\nu = 0.05, 0.1, 0.475\,\mathrm{eV}$. Although we ran our simulations with a slightly different set of cosmological parameters, Brandbyge et al. [2008] measured $\Delta P/P \sim -9.8 f_\nu$ for $\Sigma m_\nu \leq 0.6\,\mathrm{eV}$ while Viel et al. [2010] reported $\Delta P/P \sim -9.5 f_\nu$ at $z = 0$. For $\Sigma m_\nu = 0.95\,\mathrm{eV}$, we get $\Delta P/P \sim -8.6 f_\nu$ while Viel et al. [2010] reported $\Delta P/P \sim -8 f_\nu$ for $\Sigma m_\nu = 1.2\,\mathrm{eV}$. The scale at which the suppression turns over, $k_{\rm nr}$, moves from $k_{\rm nr} \sim 0.6 - 0.7\,h\mathrm{Mpc}^{-1}$ for $\Sigma m_\nu = 0.05\,\mathrm{eV}$ to $k_{\rm nr} \sim 1\,h\mathrm{Mpc}^{-1}$ for $\Sigma m_\nu = 0.95\,\mathrm{eV}$. The turnover may be related to the non-linear collapse of structures as discussed in Brandbyge et al. [2008] who reported $k_{\rm nr} \sim 1\,h\mathrm{Mpc}^{-1}$.

2.9 Matter Power Spectrum Error Estimates

In our N-body simulations, we have implemented neutrinos in the ICs only. Neutrino-weighted CDM and baryon transfer functions from CAMB were used to generate the ICs for CDM particles and baryons. To construct $P_{\rm nl}(k)$ at $z = 0$, we used Eq. 2.8. We calculated $P_{\rm nl}^{\rm cb}$ from N-body simulations and combined it with $P_{\rm lin}^{\nu}$ at $z = 0$ as solved by the CAMB code. This methodology introduces errors in the estimated matter power spectrum for two reasons: (i) the linear neutrino perturbations were taken into account only at the initial ($z_{\rm i} = 20$) and the final ($z = 0$) redshifts. There is no feedback from the neutrinos on to the CDM component in our N-body simulations. (ii) the non-linear evolution of neutrino perturbations was not accounted for in our N-body simulations. While the extent of non-linear neutrino corrections to the matter power spectrum is still being studied, we use Brandbyge et al. [2008] and Brandbyge and Hannestad [2009b] to estimate the errors in our N-body spectra. Brandbyge and Hannestad [2009b] describe

the linear neutrino density on a grid and evolve this density forward in time using linear theory. The neutrino contribution is added to the CDM component when calculating the gravitational forces. Thus, the linear neutrino component is accounted for recursively over the redshift range over which the matter power spectrum is to be evolved. Brandbyge et al. [2008] (their fig. 7, left panel) show that the matter power spectrum is underesolved by ~ 3 per cent for $\Sigma m_\nu \leq 0.6\,\mathrm{eV}$ on scales $k \geq 0.2\,h\mathrm{Mpc}^{-1}$ when the neutrino grid is neglected. Accordingly, our matter power spectrum estimates are expected to be underesolved by roughly $\lesssim 4, 1$ and 0.1 per cent for $\Sigma m_\nu = 0.95, 0.475$ and $0.1\,\mathrm{eV}$, respectively, for $k \gtrsim 0.2\,h\mathrm{Mpc}^{-1}$ at $z = 0$. Fig. 1 in Brandbyge and Hannestad [2009b] shows that the power is further suppressed by ~ 5 per cent for $\Sigma m_\nu \leq 1.2\,\mathrm{eV}$ at $k \approx 0.2 - 0.3\,h\mathrm{Mpc}^{-1}$ when the neutrino non-linearities are neglected. Overall, we estimate our N-body spectrum errors to be $\lesssim 5, 1.5$ and 0.1 per cent for $\Sigma m_\nu = 0.95, 0.475$ and $0.1\,\mathrm{eV}$, respectively, for $k \gtrsim 0.2\,h\mathrm{Mpc}^{-1}$ at $z = 0$.

2.10 Summary

In this chapter we simulated the matter power spectrum at $z = 0$ in order to study how massive neutrinos impact structure formation. The most important factors in obtaining an accurate power spectrum are (i) the Nyquist wavenumber, which depends on the simulation box size and the number of particles and (ii) the force resolution, which depends on the size of the root grid. Above the Nyquist wavenumber, the power spectrum is dominated by shot noise. For modes up to $k \lesssim 1\,h\mathrm{Mpc}^{-1}$, we found that $N_{\mathrm{cdm}} = 256^3$ in a $200\,h^{-1}\mathrm{Mpc}$ box is enough to keep the sampling errors at per cent level. We used a root grid of $N_{\mathrm{gas}} = 512^3$, which is twice as fine as N_{cdm}, to accurately calculate the gravitational forces down to the scale of the mean interparticle spacing. We showed that neutrinos with mass $\sim 0.5\,\mathrm{eV}$ or less, can be treated with linear theory since the errors due to neglecting non-linear neutrino perturbations are at sub-per cent level.

3 Developing PkANN – A Non-Linear Matter Power Spectrum Interpolator

3.1 Prelude

Achieving high-precision measurements of galaxy power spectrum from numerical simulations is computationally expensive and time consuming. Exploring the cosmological parameter space through a brute force application of simulations can be not only challenging, but in some cases impossible given the computing resources available. In this chapter we will develop the formalism for estimating the non-linear matter power spectrum using Artificial Neural Networks (ANN). As we discuss in the next chapter, the ANN technique is extremely fast and, more importantly, accurate way to determine the fully non-linear power spectrum.

3.2 Machine-Learning

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 of machine learning 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 [Jenatton et al., 2011, Pedregosa et al., 2012] to the analyses of stock market [Ghosh, 2011, Hurwitz and Marwala, 2012].

Fig. 3.1 shows a skeleton of a machine-learning network. Using a suitable training set (input parameters for which data is available), the machine-learning algorithm is trained to learn a parameterization. With this parameterization the network is capable of reproducing (as closely as possible) the output, when queried with input parameter settings that are part of the training set. The trained network can now be presented with new settings of the input parameters (for which one does not have any prior data) and by using the same parameterization learnt during the training process, the network makes predictions.

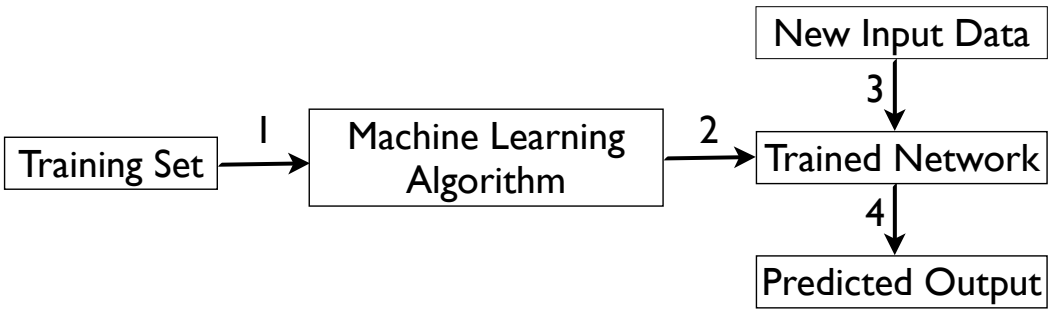

Figure 3.1: Steps 1 and 2: A machine-learning network learns to parameterize the output, for the input patterns that form the training set. Steps 3 and 4: The trained network is capable of making predictions when presented with input parameter settings. The queried input settings must lie within the parameter ranges of the patterns in the training set.

When using any machine-learning technique to predict the outcome, 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.

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.

There exist a range of techniques (see *e.g.* Nilsson [2005]) including genetic algorithms, decision tree learning,neural networks and Gaussian processes. Machine-learning techniques have been used in the fitting of cosmological functions [Auld et al., 2007, Fendt and Wandelt, 2007, Auld et al., 2008] and photometric redshifts [Collister and Lahav, 2004]. Gaussian processes have already been used as cosmological non-linear emulators [Habib et al., 2007, Schneider et al., 2008, Heitmann et al., 2009, Lawrence

et al., 2010, Schneider et al., 2011]. Gaussian process modeling (see MacKay [1997], Rasmussen and 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 ANN procedure is trained using a larger number of simulations. In this work we focus on the neural network technique.

3.3 Artificial Neural Networks

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 weight, 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 $N_{in} : N_1 : N_2 : ... : N_n : N_{out}$. That is, a network with an architecture $4 : 9 : 5 : 7$ has 4 inputs, two hidden layers with 9 and 5 nodes respectively, and finally 7 outputs. An extra node (called the *bias* node) is added to the input layer as well as to each of the hidden layers. The bias nodes are added in order to compensate for the difference between the mean of the output vector of the network and the mean of the output vector of training set patterns (for details, refer Bishop [1995]). Each bias node connects to all the nodes in the next layer. Note that the counts $N_{in}, N_1, N_2, ..., N_n$ do not include the bias nodes. The output layer has no bias node. The total number of connections (also called the *weights*) N_W for a generic architecture $N_{in} : N_1 : N_2 : ... : N_n : N_{out}$ can be calculated using the formula

$$N_W = N_{in} \cdot N_1 + \sum_{l=2}^{n} N_{l-1} \cdot N_l + N_n \cdot N_{out} + \sum_{l=1}^{n} N_l + N_{out}, \tag{3.1}$$

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. As an example, the architecture $7 : 49 : 50$ has a total of $7 \times 49 + 0 + 49 \times 50 + 49 + 50 = 2892$ weights, which we call the weight vector $\mathbf{w}$.

In Fig. 3.2, we show a typical ANN architecture (left-hand panel) and the formulae to calculate the node activations (right-hand panels). In the network configuration depicted, there are N_{in} input parameters/features $(x_1, ..., x_i)$, a single hidden layer with N_1 nodes $(z_1, ..., z_j)$, and N_{out} output parameters/features $(y_1, ..., y_k)$. The bias nodes in the input and hidden layers are x_0 and z_0, respectively.

Each node in the lth hidden layer is a neuron with an *activation*, $z_j \equiv g(a_j)$, taking as its argument

$$a_j = \sum_{i=0} w_{ji} z_i, \tag{3.2}$$

where the sum is over all nodes i (including the bias node) of the previous layer sending connections to the jth node (barring the bias node) of the current layer. Note that for

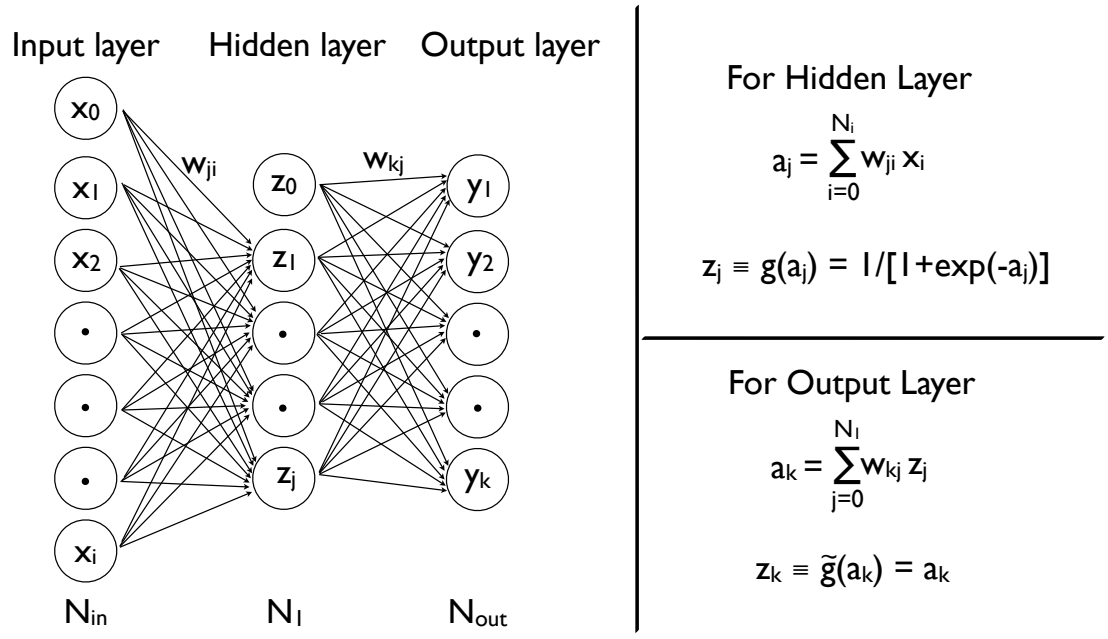

Figure 3.2: A typical ANN architecture (left-hand panel) with node activation formulae for the hidden and output layers (right-hand panels).

networks with a single hidden layer (as in Fig. 3.2), z_i in Eq. 3.2 would correspond to the input parameters x_i. The activation functions are typically taken to be sigmoid functions such as $g(a_j) = 1/[1 + \exp(-\mathrm{a_j})]$. Since the range of $g(a_j)$ is from 0 to 1, it allows the output of the neurons to be interpreted as the probability that any specific neuron will 'fire' when presented with an input parameters setting. The sigmoid functions impart some degree of non-linearity to the neural network models. A network becomes overly non-linear if the weights $\mathbf{w}$ deviate significantly from zero. This drives the activation $g(a_j)$ of the nodes to saturation. The number and size of the hidden layers add to the complexity of ANNs. The activation of all bias nodes is permanently set to a value of 1 and during network training, the bias parameters (namely, w_{j0} and w_{k0} in Fig. 3.2 left-hand panel) are adjusted so as to minimize the difference between the mean prediction for the network and the mean of the outputs of the training set patterns.

The activation $y_k \equiv \tilde{g}(a_k)$ for neurons in the output layer is usually taken to be a_k,

i.e. $\tilde{g}(a_k) = a_k$, with a_k being the weighted sum of all nodes in the final hidden layer,

$$a_k = \sum_{j=0} w_{kj} z_j. \tag{3.3}$$

For a particular input vector $(x_1, ..., x_i)$, the output vector $(y_1, ..., y_k)$ 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 $\mathbf{w}$ 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

$$\mathbf{I} \equiv (\Omega_{\rm m} h^2, \Omega_{\rm b} h^2, n_s, w_0, \sigma_8, \sum m_\nu)$$

where $h, \Omega_{\rm m}, \Omega_{\rm b}, n_s, w_0, \sigma_8$ and $\sum m_\nu$ are the present-day normalized Hubble parameter in units of 100 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 parameterization to arbitrary accuracy by training the weights $\mathbf{w}$. This is done by minimizing a suitable *cost function* (for derivation, see Appendix 7.1),

$$\chi^2(\mathbf{w}) \;=\; \frac{1}{2} \sum_{t=1}^{T} \sum_{k_i \in \{k\}} [P_{\rm nl}^{\rm ANN}(k, z|\mathbf{w}, \mathbf{I}_t) - P_{\rm nl}(k, z|\mathbf{I}_t)]^2 \tag{3.4}$$

with respect to the weights $\mathbf{w}$. The sum t is over all the cosmologies $\mathbf{I}_t$ in the training set. Remember that the matter power spectrum is a function of scale $k\,(h\mathrm{Mpc}^{-1})$. We sample the matter spectrum at discreet values between $0.006\,h\mathrm{Mpc}^{-1} \leq k \leq 1\,h\mathrm{Mpc}^{-1}$ and assign the sampled spectrum to the output nodes of the neural network. The

discreet values of scale k form the set $\{k\}\,h\mathrm{Mpc}^{-1}$. In Eq. 3.4, the sum k_i is over all the nodes in the output layer, with each node sampling the matter power spectrum at some specific scale, $k\,(h\mathrm{Mpc}^{-1})$. $P_{\rm nl}(k,z|\mathbf{I})$ is the non-linear matter power spectrum for the specific cosmology $\mathbf{I}$, computed using N-body simulations. Given the weights $\mathbf{w}$, $P_{\rm nl}^{\rm ANN}(k,z|\mathbf{w},\mathbf{I})$ is the ANN's predicted power spectrum for the $\mathbf{I}$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 et al., 2000]. As such, weighing Eq. 3.4 by $P_{\rm lin}(k,z)$ gives,

$$\chi^2(\mathbf{w}) = \frac{1}{2}\sum_{t=1}^{T}\sum_{k_i\in\{k\}}\left[\frac{P_{\rm nl}^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t) - P_{\rm nl}(k,z|\mathbf{I}_t)}{P_{\rm lin}(k,z|\mathbf{I}_t)}\right]^2 \tag{3.5}$$

$$= \frac{1}{2}\sum_{t=1}^{T}\sum_{k_i\in\{k\}}\left[R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t) - R(k,z|\mathbf{I}_t)\right]^2. \tag{3.6}$$

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 $\mathbf{w}$, $R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I})$ in Eq. 3.6 is the network's prediction of the ratio $R(k,z|\mathbf{I})$ for the specific cosmology $\mathbf{I}$. The predicted non-linear spectrum $P_{\rm nl}^{\rm ANN}(k,z|\mathbf{w},\mathbf{I})$ in Eq. 3.5 is recovered by multiplying $R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I})$ by the corresponding linear spectrum $P_{\rm lin}(k,z|\mathbf{I})$.

We ran N-body simulations over a range of cosmological parameters with the ENZO code. We include radiative cooling of baryons using an analytical approximation [Sarazin and 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. 3.3, 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 (dash-dotted), (ii) N-body (dotted), (iii) HALOFIT (dashed) and (iv) COSMIC EMULATOR (solid). The vertical dashed line at $k = 0.8\,h\mathrm{Mpc}^{-1}$ is the highest k up to which our N-body power spectra agree with the COSMIC EMULATOR 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 ($k \gtrsim 0.9\,h\mathrm{Mpc}^{-1}$), our numerical simulations lack the force resolution required to calculate power spectrum at sub-per cent level. 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.

We use the one-Loop standard PT as implemented by Saito et al. [2008] for estimating the matter power spectrum up to $k \leq 0.085\,h\mathrm{Mpc}^{-1}$ and stitch it with the non-linear power spectrum from numerical simulations. The stitched spectrum is sampled at 50 k-values between $0.006\,h\mathrm{Mpc}^{-1} \leq k \leq 1\,h\mathrm{Mpc}^{-1}$. In Fig. 3.4, we show the stitched-and-sampled N-body power spectrum (dotted) that we use as $P_{\rm nl}(k, z)$ for ANN training. This stitch-and-sample procedure is repeated for each cosmology $\mathbf{I}_t$ in the training set to get all $P_{\rm nl}(k, z|\mathbf{I}_t)$.

In Eq. 3.6, optimizing the weights $\mathbf{w}$ so as to minimize $\chi^2(\mathbf{w})$ 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*

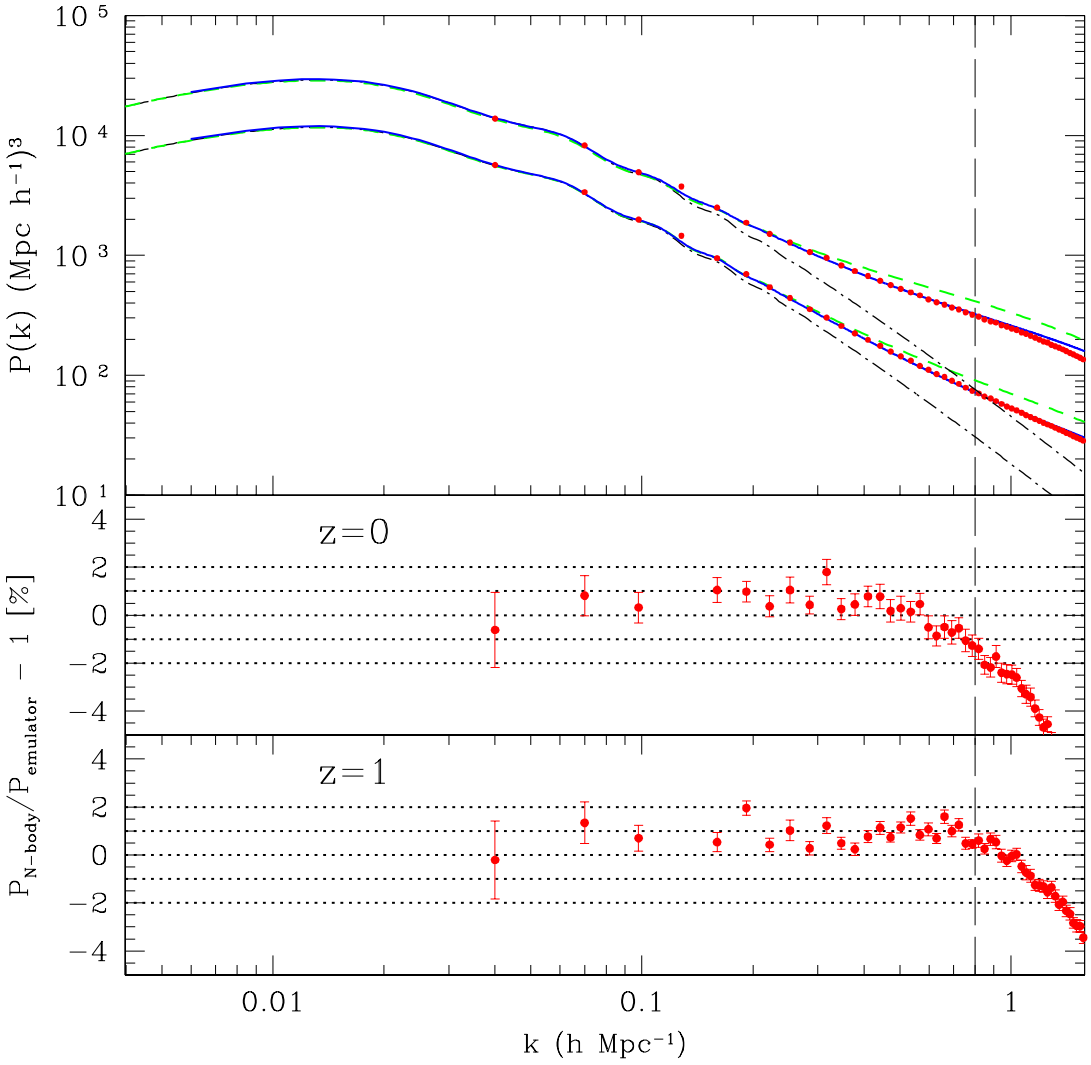

Figure 3.3: *Top:* matter power spectrum evaluated at redshift $z = 0$ (upper set) and $z = 1$ (lower set). At each redshift, the spectrum is calculated using (i) linear theory (dash-dotted), (ii) N-body (dotted), (iii) HALOFIT (dashed) and (iv) COSMIC EMULATOR (solid, see Lawrence et al. [2010]). The cosmological parameters are: $\mathbf{I} \equiv (0.13, 0.0224, 0.986, -1.23, 0.72, 0)$ with $h = 0.8$. Our N-body power spectra agree with the COSMIC EMULATOR at per cent level for $k \leq 0.8\,h\mathrm{Mpc}^{-1}$. *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 $z = 1$.

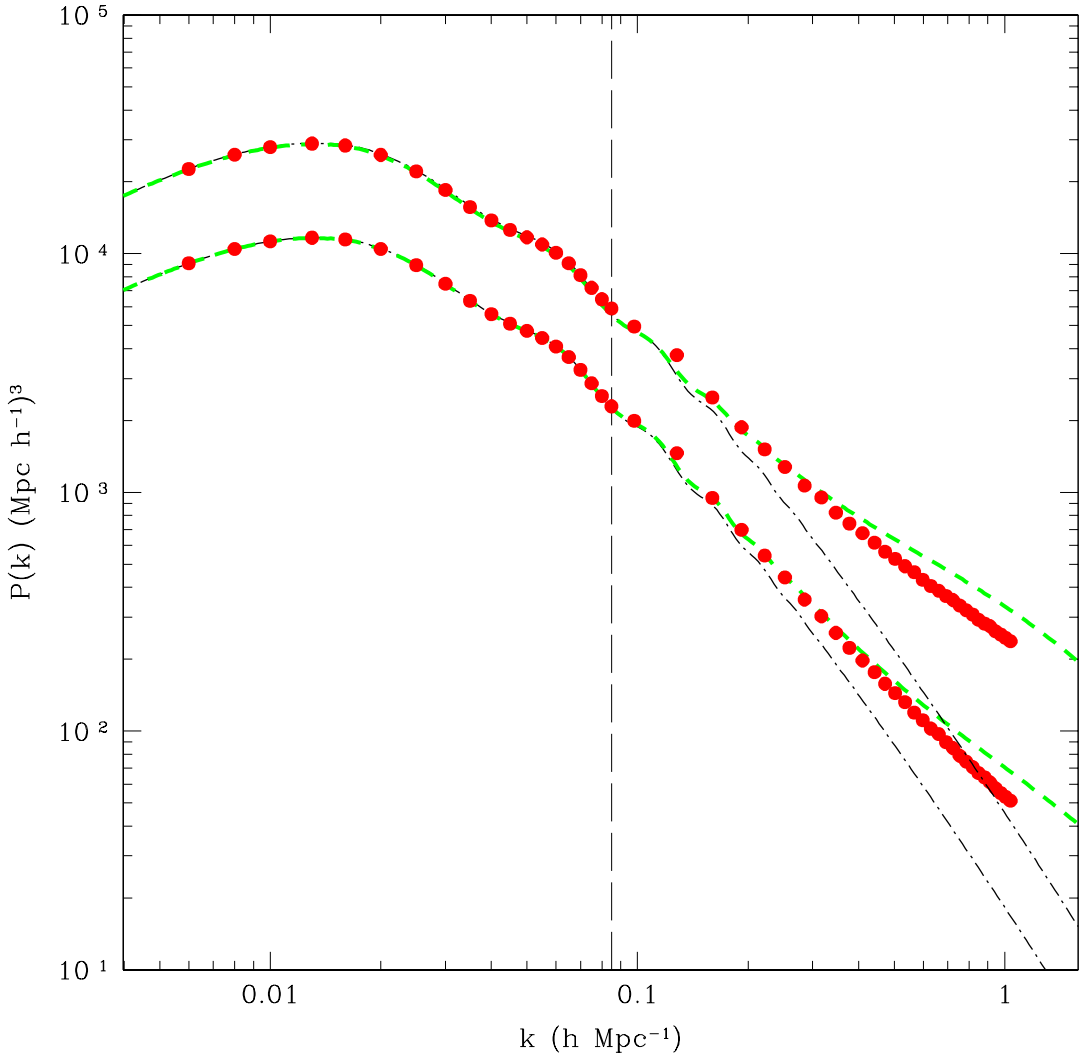

Figure 3.4: Linear theory, HALOFIT and N-body spectra from Fig. 3.3, top panel are re-plotted – with the only difference that on scales $k \leq 0.085\,h\mathrm{Mpc}^{-1}$ the N-body spectrum is approximated by the one-Loop standard PT. The stitched spectrum is then sampled at 50 k-values between $0.006\,h\mathrm{Mpc}^{-1} \leq k \leq 1\,h\mathrm{Mpc}^{-1}$ and used as $P_{\rm nl}(k,z)$ for ANN training.

included in the training set. This usually indicates (i) an overly simple network architecture (very few hidden layer nodes), (ii) very sparsely or 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 $\chi^2_Q(\mathbf{w})$ is added to the cost function $\chi^2(\mathbf{w})$,

$$\chi^2_Q(\mathbf{w}) = \frac{\xi}{2}||\mathbf{w}||^2, \tag{3.7}$$

where $||\mathbf{w}||^2$ is the quadratic sum of all the weights. The regularization parameter ξ controls the degree of regularization (smoothing) of a network's predictions. Thus, the overall cost function which is presented to the ANN for minimization with respect to the weights $\mathbf{w}$ is,

$$\chi^2_C(\mathbf{w}) = \frac{1}{2}\sum_{t=1}^{T}\sum_{k_i\in\{k\}}\left[R^{\mathrm{ANN}}(k,z|\mathbf{w},\mathbf{I}_t) - R(k,z|\mathbf{I}_t)\right]^2 + \frac{\xi}{2}||\mathbf{w}||^2. \tag{3.8}$$

To minimize $\chi^2_C(\mathbf{w})$ with respect to the weights $\mathbf{w}$, we use an iterative quasi-Newton algorithm (Appendix 7.2) that involves evaluating the first-order derivative (gradient) of the cost function. See Appendix 7.3 for the derivation of the gradient. The quasi-Newton algorithm also involves information about the inverse of the Hessian (second-order derivative) matrix which we approximate using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (see Appendix 7.4. For details, see Bishop [1995]).

The penalty term $\chi^2_Q(\mathbf{w})$, usually a quadratic sum of the weights, prevents them from becoming too large during the training process by penalizing in proportion to the sum. After having initialized ξ, its value is re-estimated during the training process iteratively. For the update formula, see Appendix 7.5. For its derivation, see Bishop [1995].

Starting with randomly assigned weights $\mathbf{w}$, their values are re-estimated iteratively,

making sure that each iteration proceeds in a direction that lowers the cost function $\chi^2_C(\mathbf{w})$. In order to avoid over-fitting to the training set, after each iteration to the weights, Eq. 3.8 is also calculated for what is known in neural network parlance as a validation set. The validation set for our application of neural networks, is a small set of simulations with known $\mathbf{I} \equiv (\Omega_{\rm m}h^2, \Omega_{\rm b}h^2, n_s, w_0, \sigma_8, \sum m_\nu)$ and $P_{\rm nl}(k,z)$. The final weights $\mathbf{w}_f$ are chosen such as to give the best fit (minimum $\chi^2_C(\mathbf{w})$) to the validation set. The network training is considered finished once $\chi^2_C(\mathbf{w})$ is minimized with respect to the validation set. The trained network can now be used to predict $P_{\rm nl}(k,z)$ 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.

It has been shown (see Hornik [1991], Ito [1991], Bishop [1995]) that networks with a single hidden layer are capable of making arbitrarily accurate approximation to a function and its derivatives. As such, for PKANN's architecture, we restrict our analysis to single-hidden layer with sigmoidal activations and output nodes with linear ($\tilde{g}(a_k) = a_k$) activations.

As can be seen in Fig. 3.3, the HALOFIT predictions are in error by as much as 50 per cent on small scales. Our intention is to use this neural network technique to iterpolate 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.

3.4 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

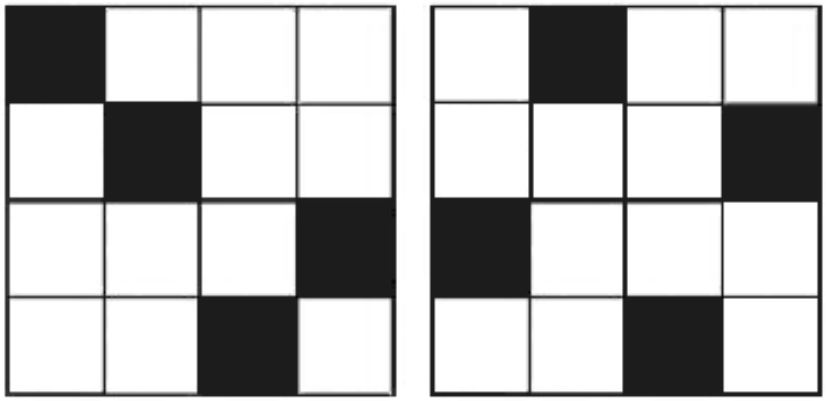

Figure 3.5: *Left:* An example of a Latin hypercube distribution. Every interval dx and dy is sampled; however each row and column are sampled only once. *Right:* an improved Latin hypercube where the distribution is more evenly spread through the space. Each subspace is equally sampled and there are no voids or clusters as in the left-hand panel (bottom left and right corners, respectively).

space with only 10 grid intervals, this quickly escalates to 10^6 points. The problem is exacerbated because N-body simulations are computationally expensive. 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. 3.5. 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 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 $\mathrm{d}x = (b-a)/n$ 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. 3.5. 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.

3.4.1 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 2. The limits on this six-dimensional parameter space are chosen so as to include the 7-yr *WMAP*+BAO+H_0 [Komatsu et al., 2011] constraints (see Table 2).

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

Cosmological parameters	Lower value	Upper value	7-yr *WMAP*+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 $w = -1$.

Table 2: 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 7-yr *WMAP*+BAO+H_0 constraints at 68 per cent CL.

degenerate. The effective number of neutrino species is fixed at $N_{\rm eff} = 3.04$. We derive the Hubble parameter h (for derivation, see Appendix 7.6) using the 7-yr *WMAP*+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.

Using Table 2 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. 3.6.

As can be seen in Fig. 3.5, 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 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. 3.6. The performance of a neural network can severely degrade near the parameter boundaries.

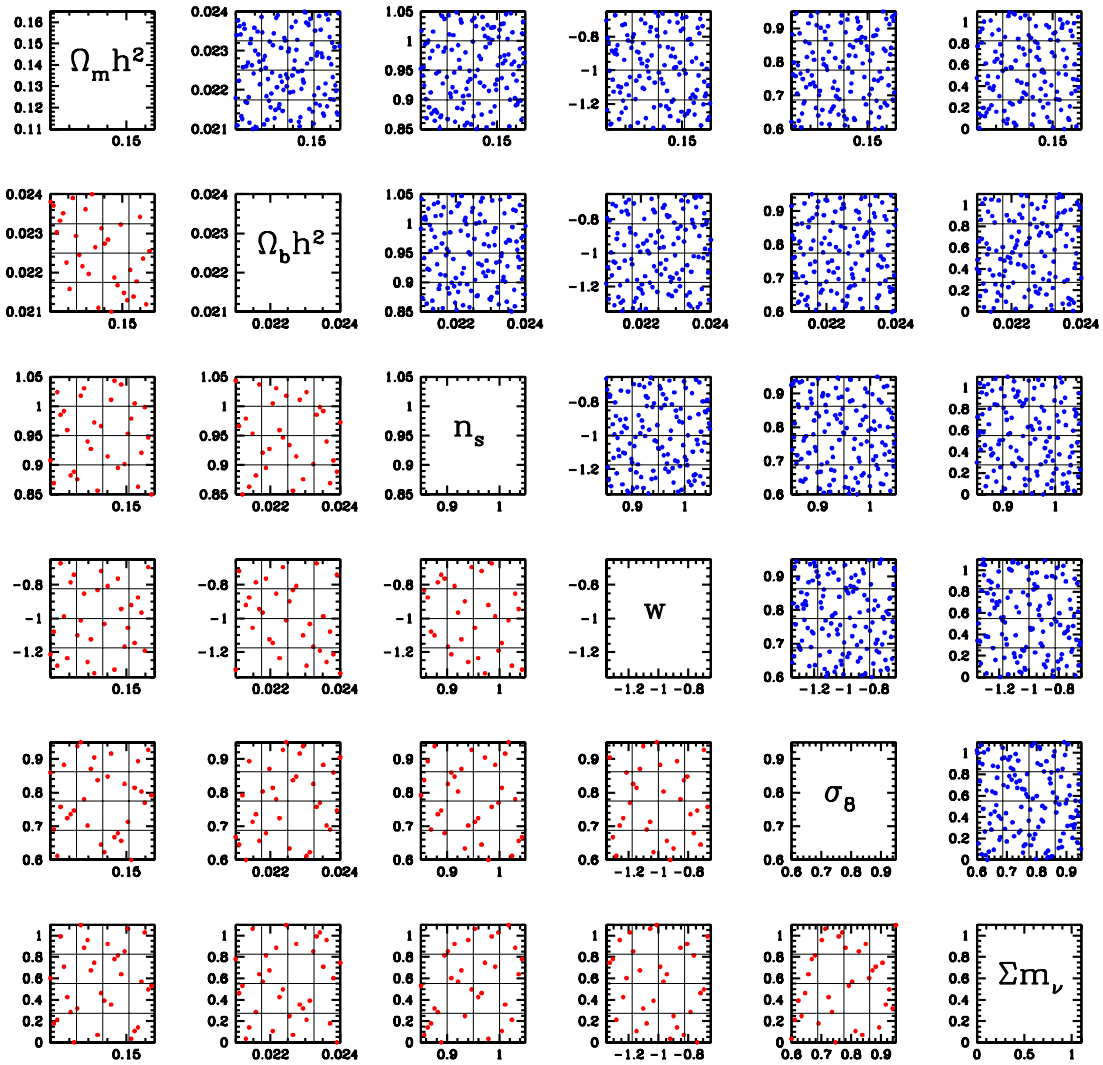

Figure 3.6: Upper triangle: ANN training set with 130 viable cosmologies, in a six-dimensional parameter space. Lower triangle: ANN validation set with 32 viable cosmologies, in a six-dimensional parameter space. See Table 2 for the parameter priors used to generate the training and validation sets.

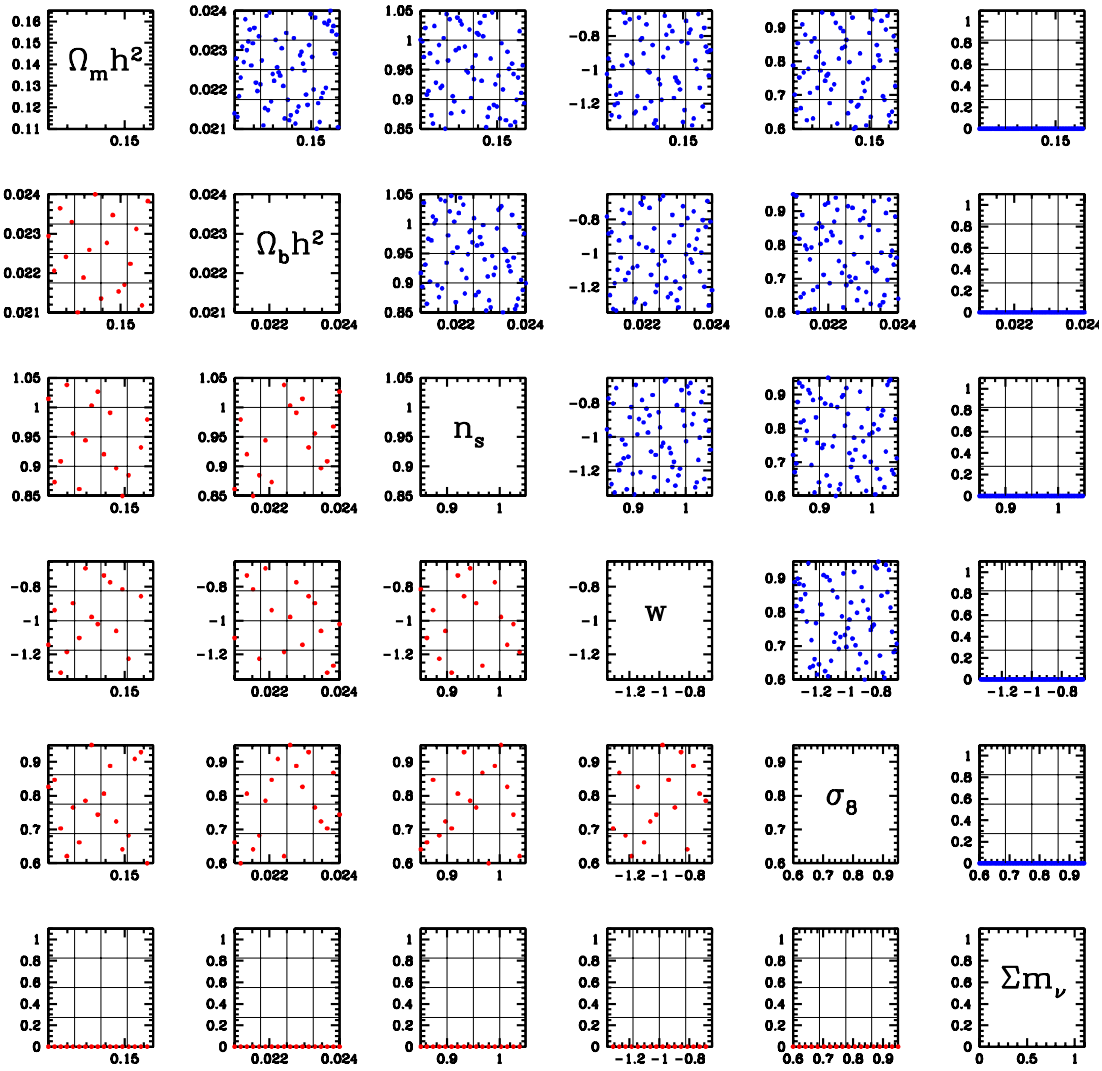

Figure 3.7: Upper triangle: extending the ANN training set (upper triangle in Fig. 3.6) with 70 cosmologies with $\sum m_\nu = 0$. Lower triangle: extending the ANN validation set (lower triangle in Fig. 3.6) with 18 cosmologies with $\sum m_\nu = 0$.

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 $\sum m_\nu \gtrsim 0$.

Adding several additional simulations at the parameter boundary may not be a computationally feasible solution to the problem due to the multi-dimensionality of the parameter space. Instead we propose to use a nested hypercube with $6 - 1 = 5$ dimensions. We fixed $\sum m_\nu = 0$ and varied the rest of the parameters over their aforementioned limits. We extended the ANN training and validation sets to include this five-dimensional hyperplane. Towards this, we generated 70 (for training) and 18 (for validation) cosmologies with $\sum m_\nu = 0$. Fig. 3.7 shows the five-dimensional hyperplane.

3.5 Summary

In this chapter we developed PkANN – a $N_{in} : N_{hidden} : N_{out}$ neural network, for the purposes of interpolating the non-linear matter power spectrum. The overall size of our training and validation sets is $(130 + 70 = 200)$ and $(32 + 18 = 50)$ cosmologies, respectively. A high number of cosmologies is preferable as it improves the sampling of the parameter space and lowers the error in the predicted power spectrum. However, the available computing resources limit the number of simulations one can possibly run. With our training and validation set sizes, we were able to keep the error in the predicted power spectrum at sub-per cent level. In the subsequent chapters, we will test PkANN's accuracy against results obtained directly from numerical simulations.

4 Interpolating Matter Power Spectrum using PkANN

4.1 Prelude

In chapter 2, we developed an artificial neural network framework (PkANN) to confront small-scale non-linearities in the power spectrum of matter fluctuations. In this chapter, we test the precision with which PkANN can predict the non-linear matter power spectrum.

4.2 Comparing PkANN's Performance against Numerical Simulations

We selected the combination $7:N_{hidden}:50$ as our PkANN architecture, where N_{hidden} (number of nodes in the hidden layer) was varied from 7 to 98, in steps of 7. The number of inputs were fixed at 7, corresponding to $\mathbf{I} \equiv (\Omega_{\rm m}h^2, \Omega_{\rm b}h^2, n_s, w_0, \sigma_8, \sum m_\nu)$ including redshift z. As discussed in Sec. 2.4, we use the CAMB code to calculate the CDM, baryon and neutrino transfer functions for the specific cosmology $\mathbf{I}$. The ICs for CDM particles and baryons are then generated from their transfer functions using ENZO. The non-linear matter power spectrum $P_{\rm nl}(k)$ is constructed using Eqs 2.8 and 2.9. To suppress sampling variance of the estimated $P_{\rm nl}(k)$, we take the average $P_{\rm nl}(k)$ from ten independent realizations.

Note that we do not sample the redshift in the Latin hypercube but instead evaluate $P_{\rm nl}(k,z)$ at 111 redshifts between $z=0$ and $z=2$ from numerical simulations, using Eqs 2.8 and 2.9. As we discussed in Sec. 3.3, we extend the range of our spectra to $k = 0.006\,h{\rm Mpc}^{-1}$ by using the one-loop standard PT Saito et al. [2008]. We estimate the matter power spectrum up to $k \leq 0.085\,h{\rm Mpc}^{-1}$ using the one-loop standard PT and stitch it with $P_{\rm nl}(k,z)$. The stitched spectrum is then sampled at 50 k-modes between $0.006\,h{\rm Mpc}^{-1} \leq k \leq 1\,h{\rm Mpc}^{-1}$. Since our training and validation sets have $(130+70)$ and $(32+18)$ cosmologies, respectively, we calculated $P_{\rm nl}(k,z)$ for each cosmology, at 111 redshifts. These $P_{\rm nl}(k,z)$ are scaled by their respective linear spectra $P_{\rm lin}(k,z)$ (see Eqs 2.6 and 2.7), before being fed to the neural network. Thus, the overall size N_T of

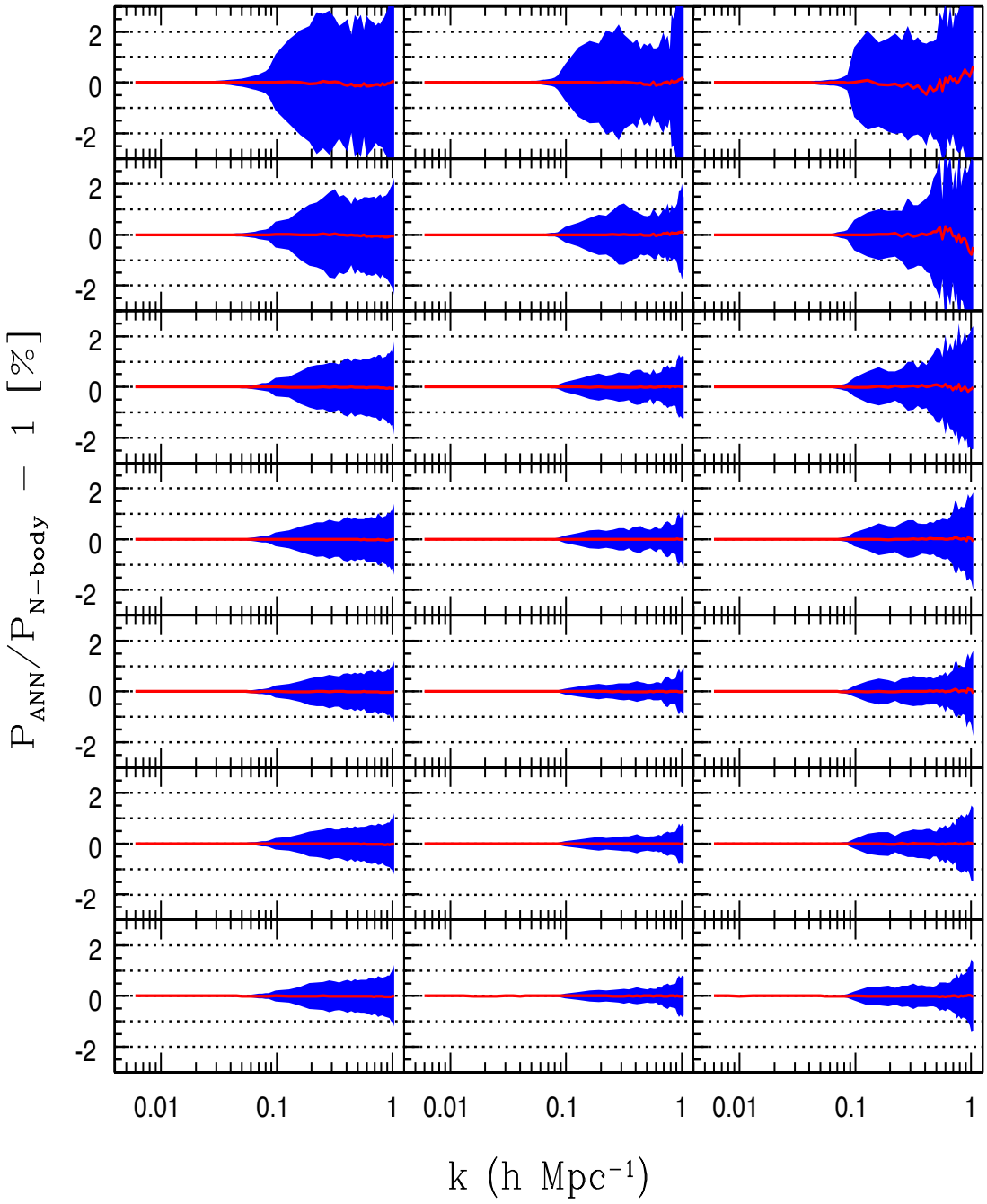

Figure 4.1: Percentage error at redshift $z = 0$ (left-hand panel), $z = 1$ (middle panel) and $z = 2$ (right-hand panel) between the predicted non-linear power spectrum (using PkANN) and the true underlying spectrum (using N-body simulations) for 200 training set cosmologies shown in the upper triangles of Figs. 3.6 and 3.7. The shaded region contains the middle 99.73% (3σ) of the residuals. The rows (from top to bottom) correspond to $N_{hidden} = 14 - 98$ in increments of 14. The mean error over all 200 cosmologies is shown by a solid line – an indicator of any bias in the ANN training scheme.

the training set that we train our ANN with is $N_T = 200 \times 111 = 22,200$. 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 $\chi^2_C(\mathbf{w})$ (see Eq. 3.8) is minimized with respect to the cosmologies in the validation set.

In Fig. 4.1, 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 $P_{\rm nl}^{\rm ANN}(k,z)$ predictions over all 16 ANN committee members. The rows correspond to $N_{hidden} = 14-98$ (from top to bottom) in increments of 14. 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 solid line in each panel, to get an idea about any systematics in our ANN training scheme. With $N_{hidden} = 70$ and higher, the ANN predictions at redshifts $z = 0$ and $z = 1$, on *all* scales, are within ± 1 per cent of the N-body 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 $k \leq 0.8\,h{\rm Mpc}^{-1}$, after which the performance degrades to ± 1.5 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 (solid line in Fig. 4.1), 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 all cosmological settings *within* the parameter priors (see Table 2) the non-linear power spectrum at $z \leq 2$ is correctly predicted within ± 1 per cent up

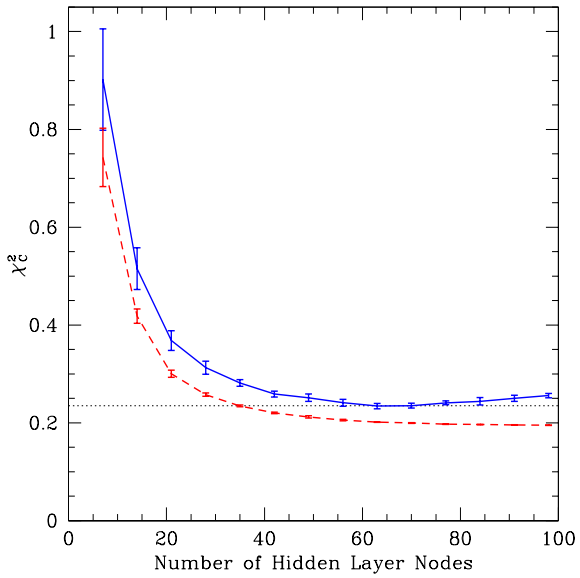

Figure 4.2: The residual error $\chi^2_C(\mathbf{w})$ (see Eq. 3.8) evaluated as a function of the number of nodes in the hidden layer, N_{hidden}. The error is a monotonically decreasing function for the training set (dashed line) while for the validation set (solid line), it starts increasing beyond $N_{hidden} = 70$ indicating that the generalizing ability of the neural network is best with $N_{hidden} = 70$. The error bars correspond to the spread in $\chi^2_C(\mathbf{w})$ for the 16 ANN committee members.

to $k \leq 0.8\,h\mathrm{Mpc}^{-1}$, 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 ($\sim$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. 10 from Lawrence et al. [2010] with our Fig. 4.1, we see that the ANN implementation performs better on all scales and redshifts.

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. 4.1, 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

Cosmological parameters	Lower value	Upper value	7-yr *WMAP*+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 $w = -1$.

Table 3: 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 2 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 7-yr *WMAP*+BAO+H_0 constraints at 68 per cent CL.

presented with a new dataset. The validation set helps in controlling the complexity of a network, as we discussed earlier after Eq. 3.8. In Fig. 4.2, we show the residual cost function $\chi^2_C(\mathbf{w})$ (see Eq. 3.8) evaluated as a function of the number of nodes in the hidden layer, N_{hidden}. The residual error is a monotonically decreasing function for the training set (dashed line) while for the validation set (solid line), it increases beyond $N_{hidden} = 70$. The performance of the trained ANNs as a function of N_{hidden} units, over the cosmologies in the validation set, is shown in Fig. 4.3. Increasing N_{hidden} beyond 70 increases the error marginally, indicating that $N_{hidden} = 70$ saturates the generalizing ability of our network.

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 3 for the parameter limits of the testing set. We show the testing set in Fig. 4.4, 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. 4.5. Increasing N_{hidden} beyond 70 does not contribute to a significant error reduction on the testing set, confirming our assessment

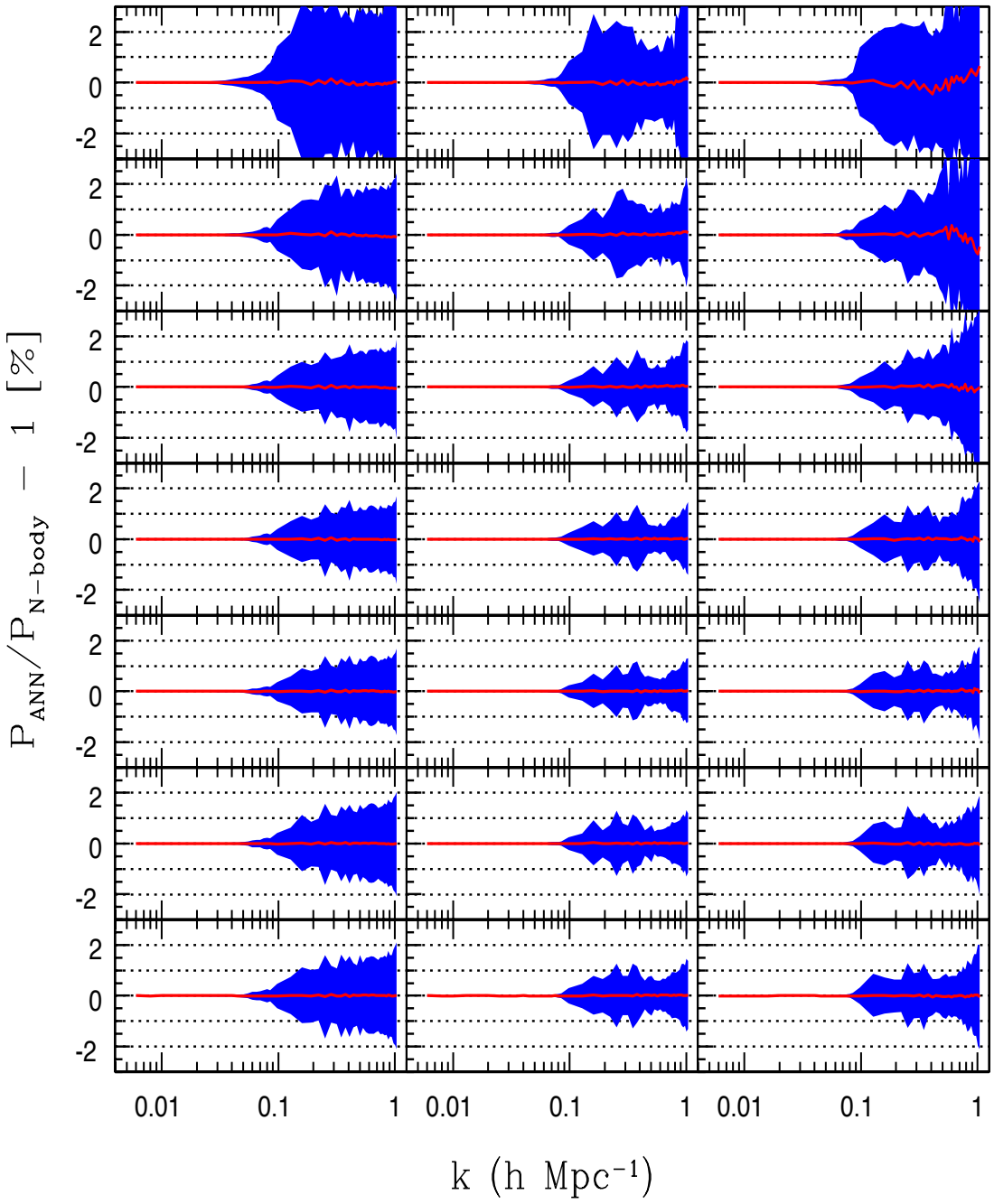

Figure 4.3: Similar to Fig. 4.1, using 50 validation set cosmologies. The validation set is shown in the lower triangles of Figs 3.6 and 3.7.

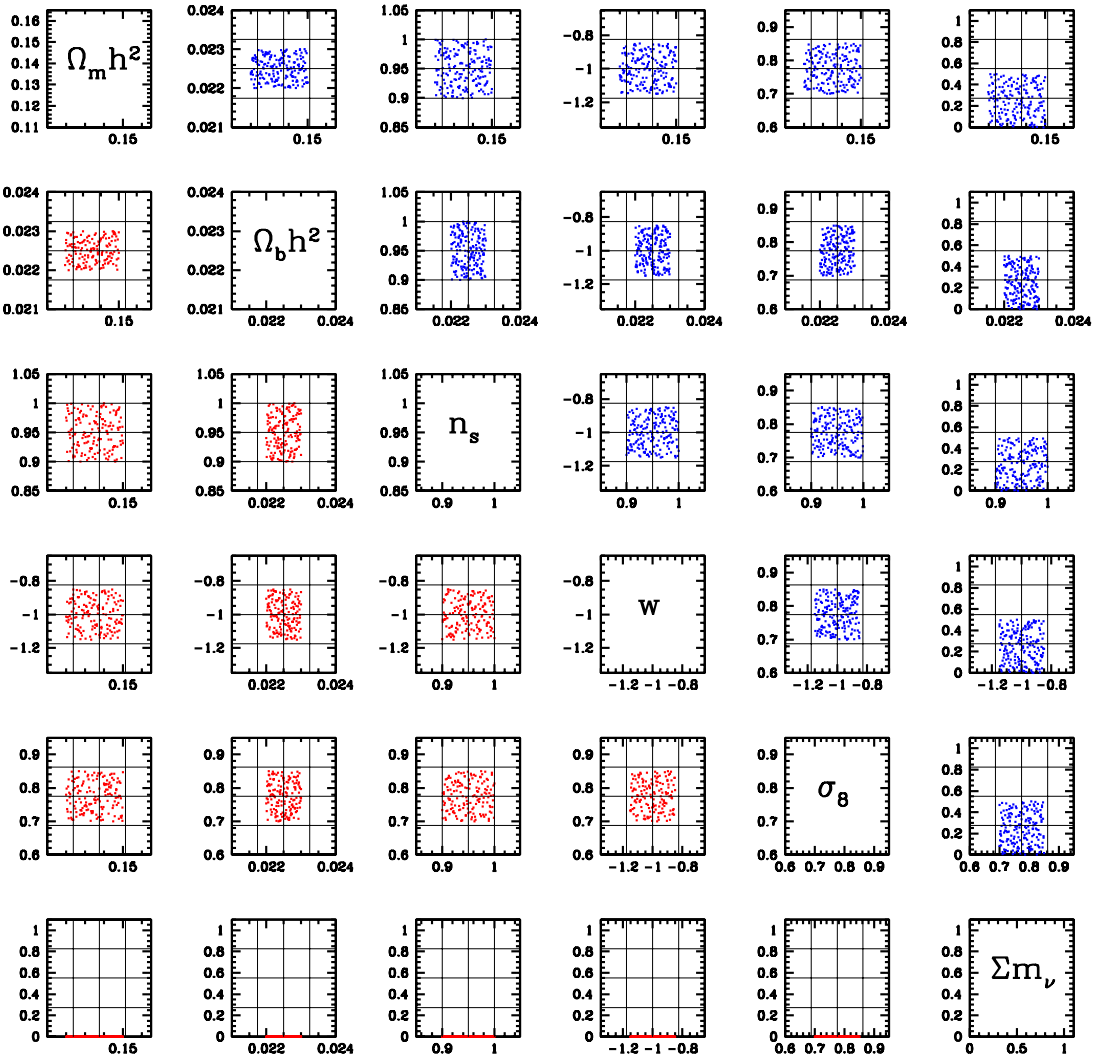

Figure 4.4: Upper triangle: ANN testing set with 180 cosmologies with $\sum m_\nu > 0$. Lower triangle: extending the ANN testing set with 150 cosmologies with $\sum m_\nu = 0$. See Table 3 for the parameter priors used to generate the testing set.

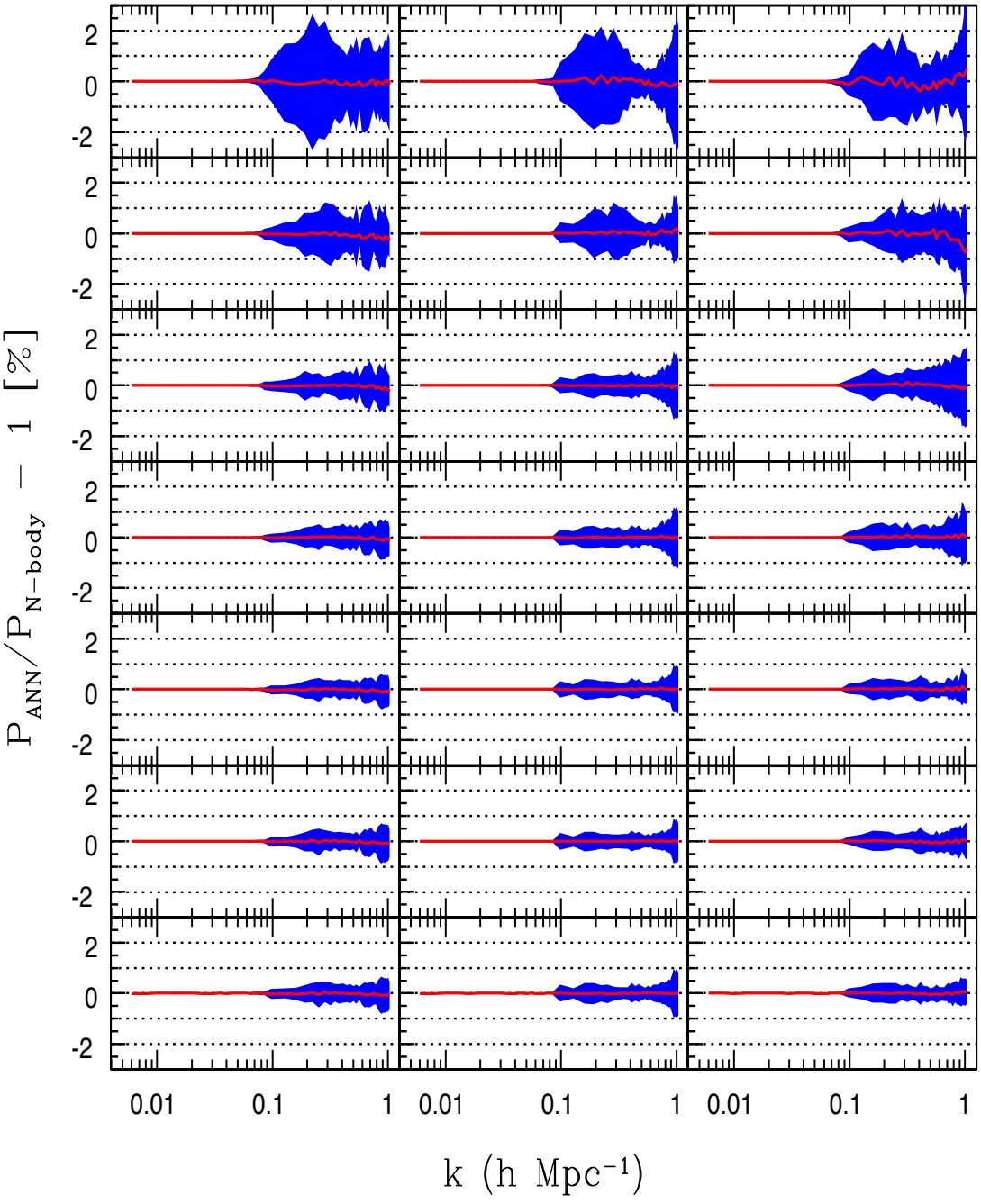

Figure 4.5: Similar to Fig. 4.1, using 330 testing set cosmologies shown in Fig. 4.4.

that $N_{hidden} = 70$ saturates the generalizing ability of the network. With $N_{hidden} = 70$, the ANN prediction for *every* cosmology, at *all* redshifts $z \leq 2$, is within ± 0.5 per cent of the N-body power spectra up to $k \leq 0.9\,h\mathrm{Mpc}^{-1}$. As expected, the ANN performs exceedingly well within from the boundaries of the restricted parameter space (see Table 3). 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.

4.3 Exploring Cosmological Parameter Space with PkANN

Having built the power spectrum interpolator, we now study the behavior of the power spectrum as a function of the cosmological parameters. In Fig. 4.6, we show variations in the power spectrum at redshift $z = 0$ (top row), $z = 1$ (middle row), $z = 2$ (bottom row). At each redshift, $\Omega_{\rm m}h^2$ is varied between its minimum and maximum value (Table 3, columns 2 and 3) while $\Omega_{\rm b}h^2, n_s, w_0, \sigma_8$ are fixed at their central values. We fix $\sum m_\nu = 0$ since we want to compare our PkANN predictions with the COSMIC EMULATOR, which is not trained for massive neutrino cosmologies. The left-hand panels show natural logarithm of the ratio of the power spectra with different $\Omega_{\rm m}h^2$ to the base power spectrum. The base power spectrum corresponds to the central values: $\Omega_{\rm m}h^2 = 0.135, \Omega_{\rm b}h^2 = 0.0225, n_s = 0.95, w_0 = -1, \sigma_8 = 0.775$, with $\sum m_\nu = 0$. The absolute power spectra are shown in the right-hand panels. Within each panel, the power spectra (from top to bottom) correspond to increasing $\Omega_{\rm m}h^2$. Higher $\Omega_{\rm m}h^2$ reduces the large-scale normalization of the power spectrum significantly. Accurate measurements of the power spectrum amplitude on large scales can help improve the constraints on $\Omega_{\rm m}h^2$. The PkANN predictions (dotted) agree well with the COSMIC EMULATOR (solid lines). Note that for redshift $z = 2$, we only show the PkANN predictions since the COSMIC EMULATOR can make predictions only up to $z = 1$.

In Figs 4.7 – 4.10, we vary $\Omega_{\rm b}h^2, n_s, w_0$ and σ_8, respectively. The power spectra trends from minimum to maximum values are as follows: top to bottom (n_s and w_0)

and bottom to top ($\Omega_{\rm b}h^2$ and σ_8). At $z = 0$, except σ_8, all other parameters affect the power spectrum predominantly on large scales ($\sim k < 0.1\,h{\rm Mpc}^{-1}$). Using power spectrum measurements to further improve $\Omega_{\rm b}h^2$ constraints is going to be challenging since $\Omega_{\rm b}h^2$ affects the power spectrum at not more than $\sim 5\%$ level within the range we consider. Reducing uncertainties in the other parameters using small-scale data would be difficult unless one measures the power spectrum at higher redshifts where almost all parameters leave discernible imprints.

4.4 Summary

We have successfully reconstructed the non-linear matter power spectrum using PkANN. It is worth stressing here that this method 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.

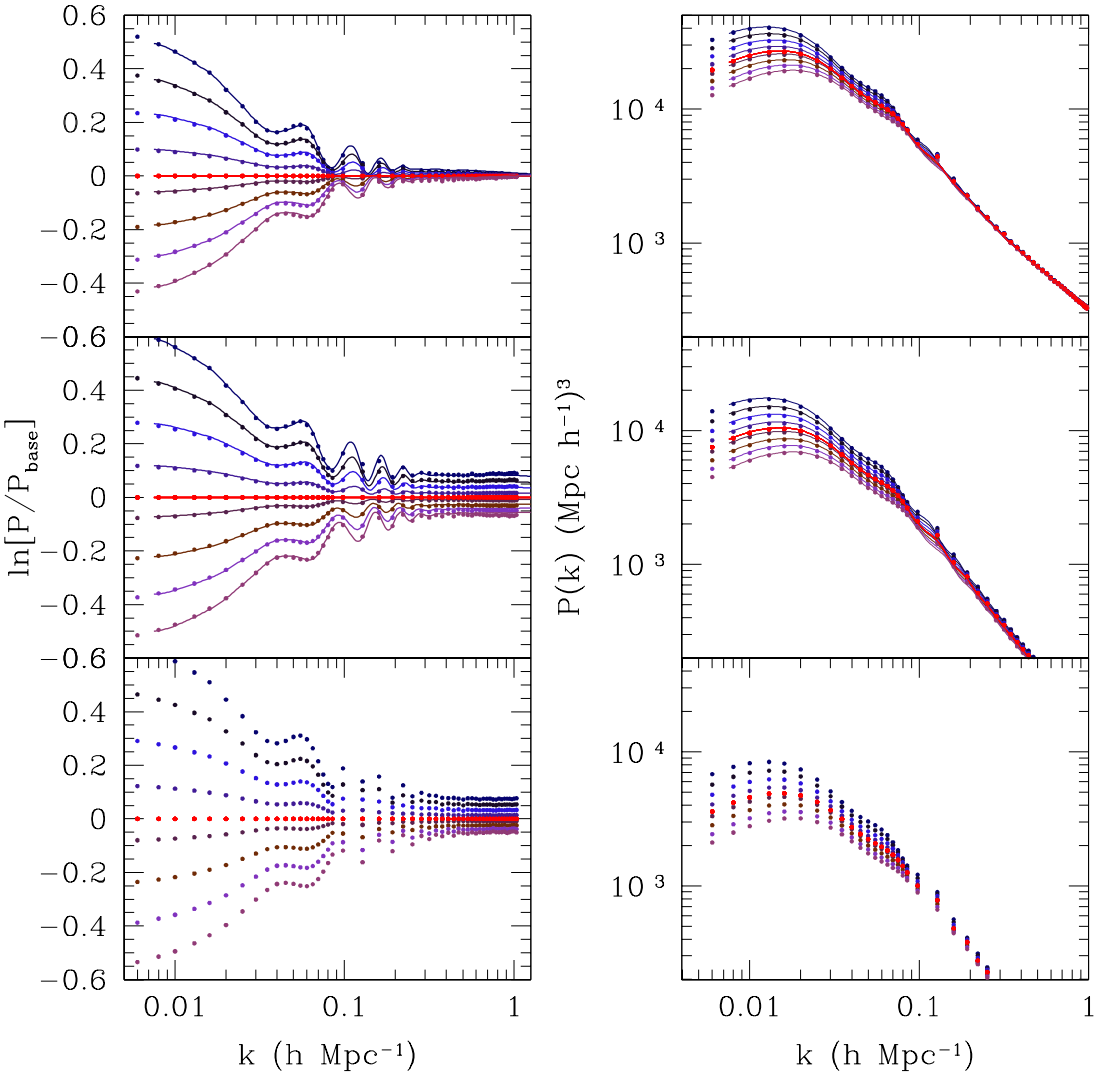

Figure 4.6: Variations in the power spectrum at redshift $z = 0$ (top row), $z = 1$ (middle row), $z = 2$ (bottom row). Parameter $\Omega_{\rm m}h^2$ is varied between its minimum and maximum value (Table 3, columns 2 and 3) while $\Omega_{\rm b}h^2, n_s, w_0, \sigma_8$ are fixed at their central values. $\sum m_\nu = 0$ to facilitate comparison with the COSMIC EMULATOR. The left-hand panels show natural logarithm of the ratio of the power spectra with different $\Omega_{\rm m}h^2$ to the base power spectrum. The base power spectrum corresponds to $\Omega_{\rm m}h^2 = 0.135, \Omega_{\rm b}h^2 = 0.0225, n_s = 0.95, w_0 = -1, \sigma_8 = 0.775$, with $\sum m_\nu = 0$. The absolute power spectra are shown in the right-hand panels. Within each panel, the power spectra (from top to bottom) correspond to increasing values of $\Omega_{\rm m}h^2$. PKANN predictions (dotted) are within 0.2% of the COSMIC EMULATOR spectra (solid lines). At $z = 2$, only PKANN predictions are shown.

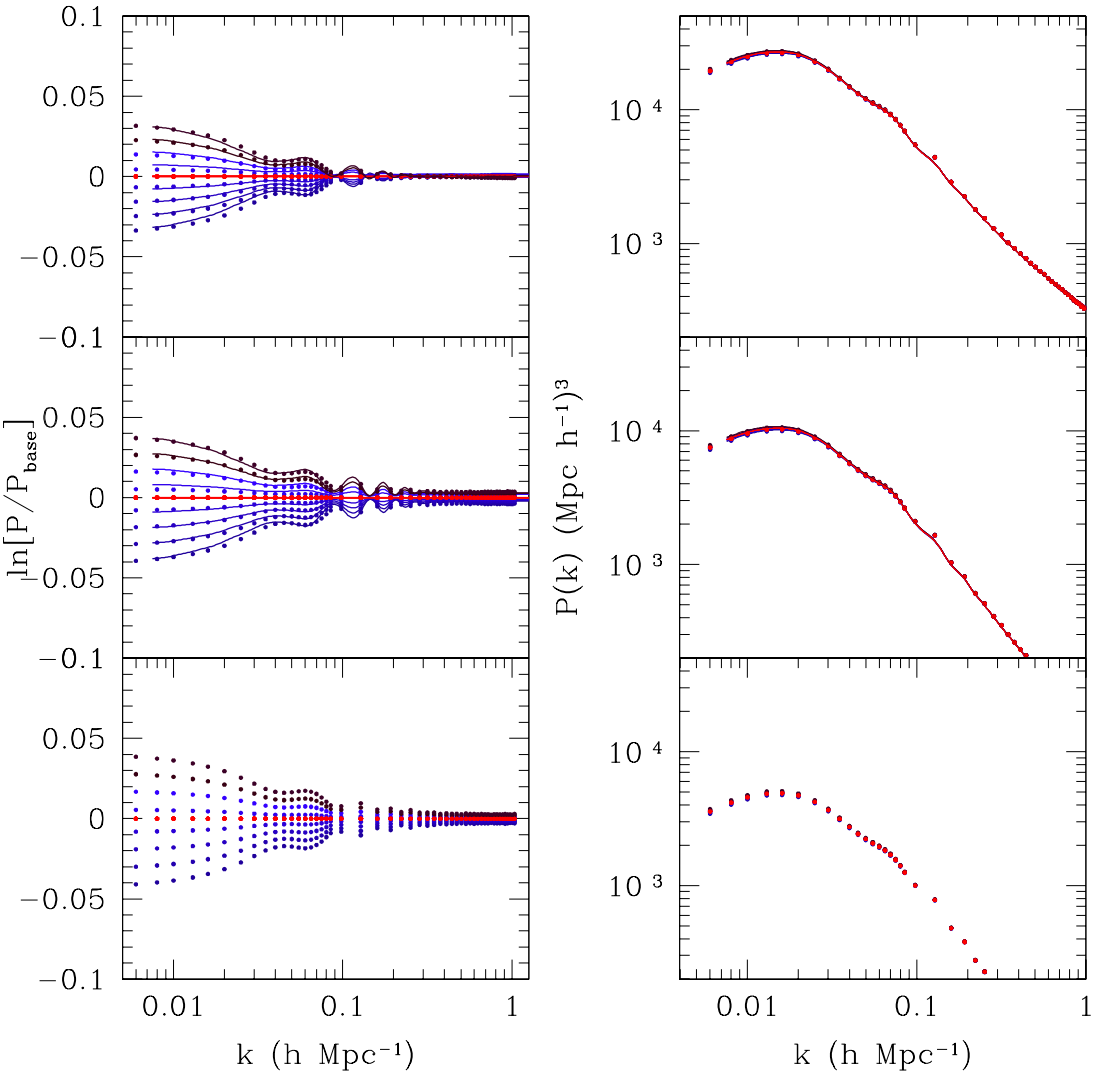

Figure 4.7: Similar to Fig. 4.6, but for a range of $\Omega_{\rm b}h^2$ values. Within each panel, the power spectra from bottom to top correspond to increasing $\Omega_{\rm b}h^2$ values.

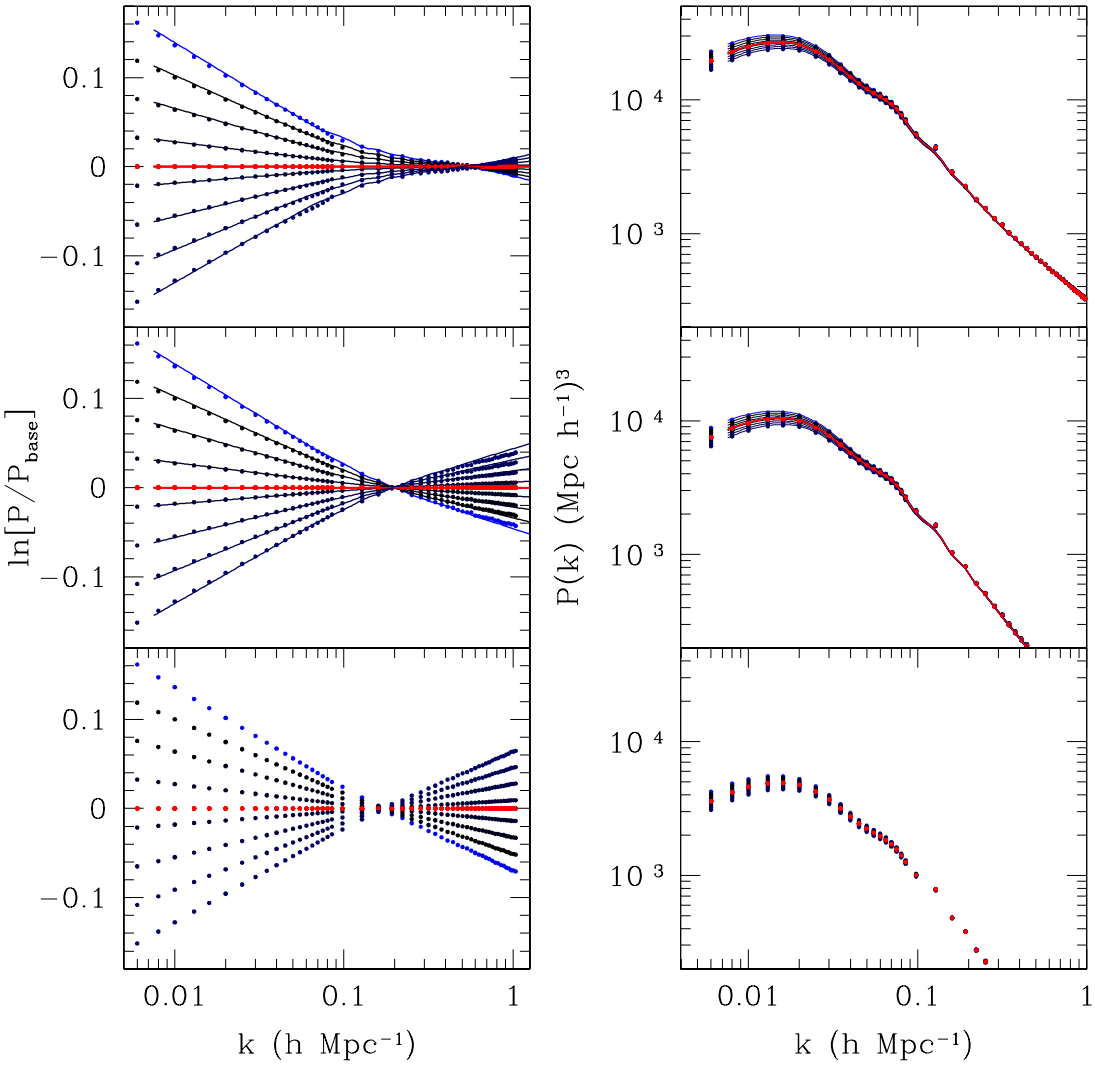

Figure 4.8: Similar to Fig. 4.6, but for a range of n_s values. Within each panel, the power spectra from top to bottom correspond to increasing n_s values.

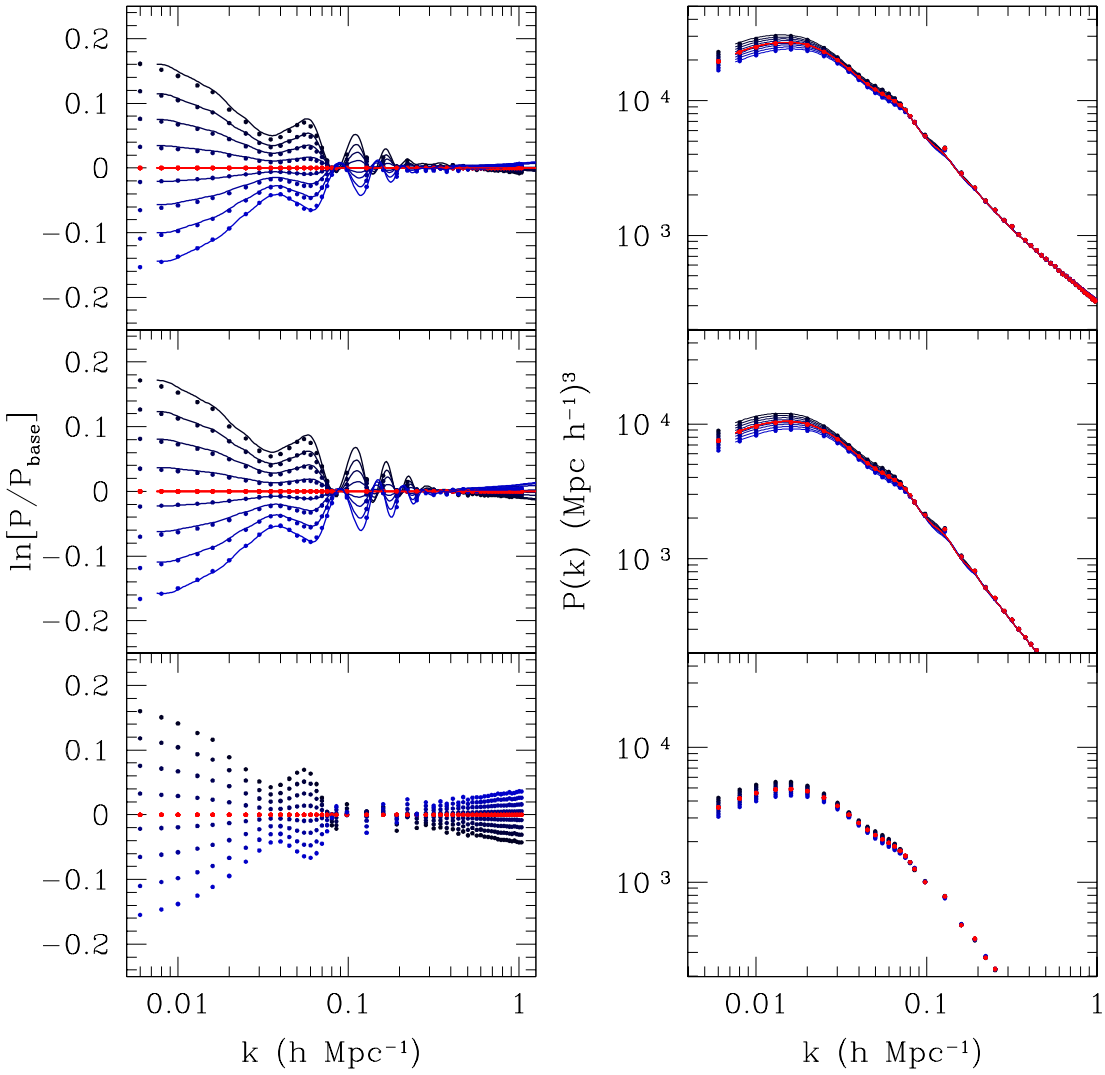

Figure 4.9: Similar to Fig. 4.6, but for a range of w_0 values. Within each panel, the power spectra from top to bottom correspond to increasing w_0 values.

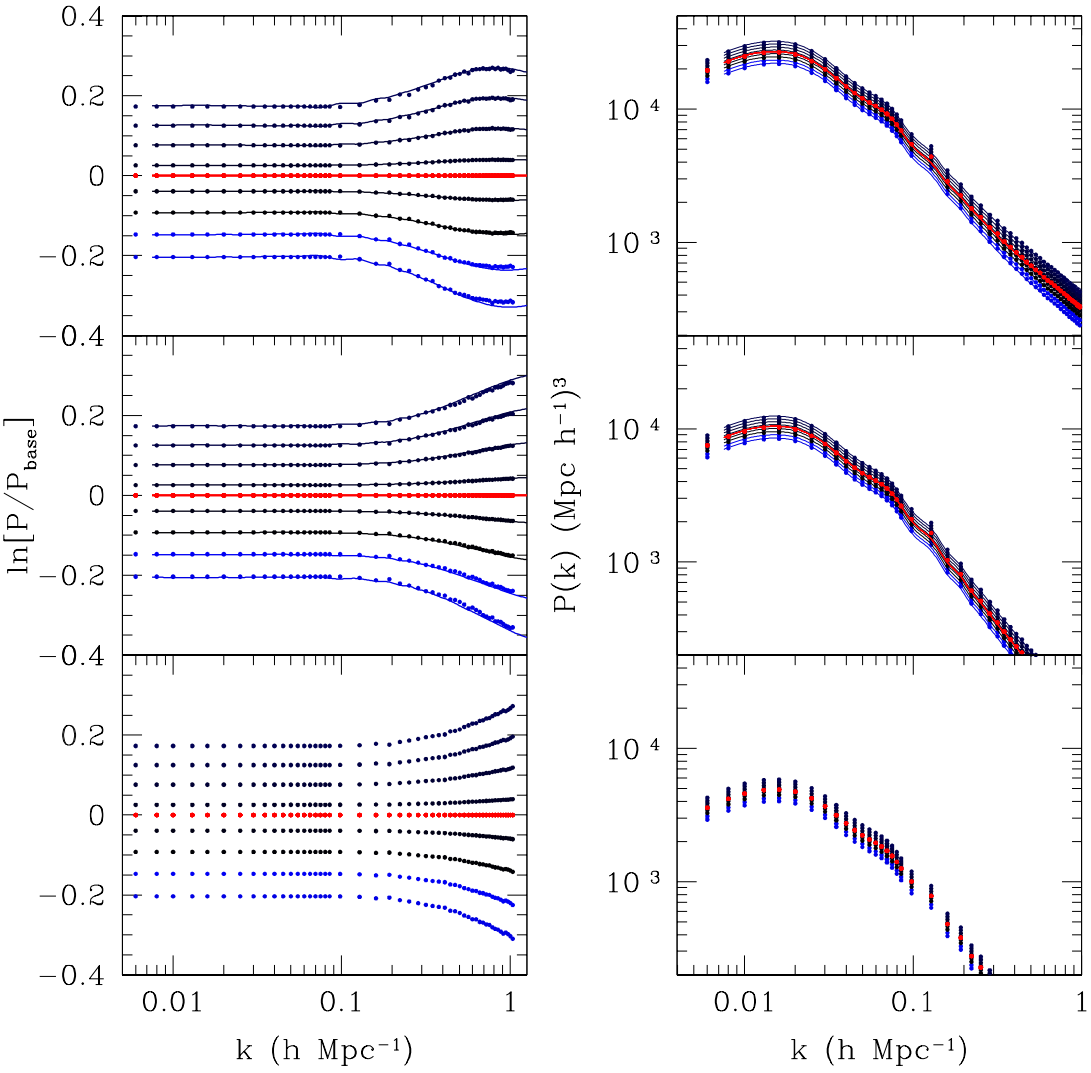

Figure 4.10: Similar to Fig. 4.6, but for a range of σ_8 values. Within each panel, the power spectra from bottom to top correspond to increasing σ_8 values.

5 Estimating the Cosmic Mach Number using PkANN

5.1 Prelude

Peculiar velocities are a sensitive probe of the underlying large-scale matter density fluctuations in our Universe. In particular, large, all-sky surveys of the peculiar velocities of galaxies or clusters of galaxies can provide important constraints on cosmological parameters. However, studies of peculiar velocities suffer from several drawbacks, including (i) the presence of small-scale, non-linear flows, such as infall into clusters, can potentially bias analyses which typically rely on linear theory, (ii) sparse, non-uniform sampling of the peculiar velocity field can lead to aliasing of small-scale power on to large scales and bias due to heavier sampling of dense regions, (iii) large measurement uncertainties of individual peculiar velocity measurements, particularly for distant galaxies or clusters, make it necessary to work with large surveys in order to extract meaningful constraints.

Ostriker and Suto [1990] introduced a dimensionless statistic of the cosmological structure - the cosmic Mach number, as a way to measure the warmth/coldness of the velocity field on some scale R. Specifically, one measures the bulk flow $\mathbf{u}(\mathbf{x_0}; R)$ of a region of size R centered at $\mathbf{x_0}$, as well as the velocity dispersion $\sigma(\mathbf{x_0}; R)$ of the objects within this region. The ratio of the bulk flow to the velocity dispersion, namely $(|\mathbf{u}(\mathbf{x_0}; R)|^2/\sigma^2(\mathbf{x_0}; R))^{1/2} \equiv M(\mathbf{x_0}; R)$, is the cosmic Mach number. The ensemble average over $\mathbf{x_0}$ gives the statistic $M(R)$. Since both $|\mathbf{u}(\mathbf{x_0}; R)|^2$ and $\sigma^2(\mathbf{x_0}; R)$ scale equally by the amplitude of the matter density perturbation, the statistic M is independent (at least in linear approximation) of the normalization of the matter power spectrum.

In linear theory, given the cosmological parameters, M can be readily calculated and compared with its measured value from the peculiar velocity field catalogues. However, comparing theoretical predictions with observations is not straightforward: (i) one has to correct for the small-scale non-linearities in observations as well as take into account the fact that observations represent only a discreet sample of the continuous velocity field. This can be remedied by smoothing the velocity field on a suitable scale r_s

($\sim 5h^{-1}$Mpc, since on larger scales the matter density field is expected to be linear), before estimating the quantities $\mathbf{u}(\mathbf{x_0}; R)$ and $\sigma(\mathbf{x_0}; R)$. However, any residual non-linearity in the observed field can still bias the M estimates; (ii) Non-uniform, noisy and sparse sampling of the peculiar velocity field can lead to aliasing of small-scale power onto larger scales. When making comparisons with theory, one has to carefully take into account the selection function and the noise of the real dataset.

Over the last couple of decades, the statistic M has been investigated: Ostriker and Suto [1990] used linear theory and Gaussian selection function to show that standard Cold Dark Matter (sCDM) model is inconsistent (predicts M almost twice the observed value) with observations at $\sim 95\%$ CL; Suto et al. [1992], using Tophat and Gaussian selection functions, studied the distribution of M using N-body simulations to rule out the sCDM scenario at 99% CL; Strauss et al. [1993] took into account the selection function of real surveys and extracted mocks from numerical simulations over a range of cosmologies including sCDM and tilted CDM (scalar spectral index, $n_s \neq 1$) among others, to reject the sCDM model at 94% CL; Ma et al. [2012] explored the potential of using M in distinguishing cosmological models, including modified gravity and massive neutrino cosmologies.

In this chapter, (i) we estimate the cosmic Mach number for various galaxy peculiar velocity datasets; (ii) we investigate how likely it is to get these Mach values in a ΛCDM universe. To achieve this, we study the statistical distribution of the expected Mach number by extracting mocks of the real catalogues from numerical simulations of a ΛCDM universe. We show that a ΛCDM universe with 7-yr *WMAP* type cosmology is consistent with the Mach observations at 2σ CL; (iii) we further show that our M estimates for the mocks are not biased by the selection function of the mocks. Towards this, we extract dense and nearly-isotropic distributions with a Gaussian profile $f(r) \propto e^{-r^2/2R^2}$ with $R = 10{-}100h^{-1}$Mpc. We show that the Mach numbers estimated from the mocks are very similar to the values based on Gaussian profiles (of similar depth R as the mocks); (iv) we use the non-linear matter power spectrum interpolation scheme PkANN

to check if we can avoid N-body simulations completely and predict $M(R)$ by only using PkANN's prediction for the non-linear power spectrum. This is crucial because high-resolution hydrodynamic N-body simulations are computationally expensive and extremely time consuming. Exploring the parameter space with numerical simulations with reasonable computing resources and time might not be possible. A full use of a statistic like M can only be realized with a prescription for the non-linear matter power spectrum.

It is worth mentioning here the reason for our choice of a Gaussian profile $f(r)$ over, for example, a Tophat filter in our analysis. A Tophat filter gets significant contribution from smaller (than Tophat size R) scales, which are contaminated by non-linearities at low redshifts. Bulk flow calculated using a Tophat filter can be compared with expectations from linear theory *only* if the observed velocity field is reasonably dense and uniform, so that the small-scale systematics average out. However, observations typically are sparse and non-uniform with large uncertainties. This leads to aliasing of small-scale power on to large scales, making comparison with theory difficult. A Gaussian filter, on the other hand, gets very little contribution from small scales, thereby making comparison with linear theory meaningful.

5.2 The Cosmic Mach Number

Given a peculiar velocity field $\mathbf{v(x)}$, one can calculate the bulk flow, which represent the net streaming motion of a region in some direction relative to the background Hubble expansion. The bulk flow $\mathbf{u}(\mathbf{x_0}; R)$ of a region of size R centered at $\mathbf{x_0}$ can be defined as

$$\mathbf{u}(\mathbf{x_0}; R) \quad = \quad \int d\mathbf{x}\ \mathbf{v}(\mathbf{x}) F(|\mathbf{x} - \mathbf{x_0}|, R), \tag{5.1}$$

where $F(|\mathbf{x} - \mathbf{x_0}|, R)$ is the filter used to average the velocity field $\mathbf{v(x)}$ on a characteristic scale R. Although Tophats and Gaussian filters are the preferred choices,

$F(|\mathbf{x}-\mathbf{x_0}|,R)$ can be designed to mimic the selection function of the real datasets. This is useful when dealing with datasets whose selection function depends strongly on the position in the sky. In Fourier space, Eq. 5.1 can be written as

$$\mathbf{u}(\mathbf{x_0};R) \quad = \quad \int d\mathbf{k}\ \mathbf{v}(\mathbf{k})W(\mathbf{k},R)e^{-i\mathbf{k}\cdot\mathbf{x_0}}, \tag{5.2}$$

where $\mathbf{v}(\mathbf{k})$ and $W(\mathbf{k},R)$ are the Fourier transforms of the peculiar velocity field $\mathbf{v}(\mathbf{x})$ and the filter $F(|\mathbf{x}-\mathbf{x_0}|,R)$, respectively.

In linear theory of structure formation, at low redshifts, the velocities are related to the matter overdensities via

$$\mathbf{v}(\mathbf{k}) \quad = \quad ifH_0\delta(\mathbf{k})\frac{\mathbf{k}}{k^2}, \tag{5.3}$$

where $\delta(\mathbf{k})$ is the Fourier transform of the overdensity field $\delta(\mathbf{x})$. The linear growth rate factor f can be approximated as $f=\Omega_m^{0.6}$. Thus, the velocity power spectrum $P_v(k)$ is proportional to the matter power spectrum $P(k)$ at low redshifts,

$$P_v(k) \quad = \quad (H_0 f)^2\frac{P(k)}{k^2}. \tag{5.4}$$

Using Eq. 5.2 and Eq. 5.4, the mean-squared bulk value of $\mathbf{u}(\mathbf{x_0};R)$ can be shown to be

$$\sigma_v^2(R) \equiv <\mathbf{u}^2(\mathbf{x_0};R)> = \frac{H_0^2\Omega_{\rm m}^{1.2}}{2\pi^2}\int dk\ P(k)W^2(kR), \tag{5.5}$$

where the average is taken over all spatial positions $\mathbf{x_0}$.

The squared velocity dispersion within a region of size R centered at $\mathbf{x_0}$ can be similarly defined as

$$\sigma^2(\mathbf{x_0};R) = \int d\mathbf{x}\ |\mathbf{v}(\mathbf{x})|^2F(|\mathbf{x}-\mathbf{x_0}|,R) - |\mathbf{u}(\mathbf{x_0};R)|^2. \tag{5.6}$$

In Fourier space, the ensamble average of Eq. 5.6 over $\mathbf{x_0}$ becomes

$$\sigma^2(R) \equiv < \sigma^2(\mathbf{x_0}; R) > = \frac{H_0^2 \Omega_{\rm m}^{1.2}}{2\pi^2} \int dk\ P(k) \left(1 - W^2(kR)\right). \tag{5.7}$$

Using Eq. 5.5 and Eq. 5.7, the cosmic Mach number can now be defined as

$$M(R) \equiv < M^2(\mathbf{x_0}; R) >^{1/2} = \left(\frac{\sigma_v^2(R)}{\sigma^2(R)}\right)^{1/2}. \tag{5.8}$$

As discussed in literature [Ostriker and Suto, 1990, Suto et al., 1992, Strauss et al., 1993], the cosmic Mach number is essentially a measure of the shape of the matter power spectrum: The *rms* bulk flow $\sigma_v(R)$ gets most of its contribution from scales larger than R, while the velocity dispersion $\sigma(R)$ is a measure of the strength of velocities on scales smaller than R and gets most contribution from small scales. Furthermore, the statistic M is expected to be independent of the matter power spectrum normalization – at least on large scales, where the perturbations are still well described by linear theory and affect both $\sigma_v^2(R)$ and $\sigma^2(R)$ equally. M can be a powerful tool to test not only the ΛCDM scenario, but also a wide range of cosmologies including models with massive neutrinos. Massive neutrinos suppress the matter power spectrum in a scale dependent way, thereby altering the velocity dispersion much more prominently than the bulk flow. Mach number M provides an easy to interpret technique to distinguish between various cosmological models.

5.3 N-body Simulations

In order to study the statistical distribution of M, we extract mock surveys from the 41 numerical realizations of a ΛCDM universe. The N-body simulation we use in our analysis is Large Suite of Dark Matter Simulations (LasDamas, hereafter LD) (McBride et al. [2009]; McBride et al. [2011] in prep[2]). The LD simulation parameters are: $\Omega_{\rm m} = 0.25$, $\Omega_{\rm b} = 0.04$, $\Omega_\Lambda = 0.75$, $h = 0.7$, $\sigma_8 = 0.8$, $n_{\rm s} = 1.0$ and $L_{\rm Box} = 1h^{-1}$Gpc for the

[2]http://lss.phy.vanderbilt.edu/lasdamas

Cosmological parameters	LD-*Carmen*
Matter density, $\Omega_{\rm m}$	0.25
Cosmological constant density, Ω_Λ	0.75
Baryon density, $\Omega_{\rm b}$	0.04
Hubble parameter, h (100 km s^{-1} Mpc^{-1})	0.7
Amplitude of matter density fluctuations, σ_8	0.8
Primordial scalar spectral index, $n_{\rm s}$	1.0
Simulation design parameters	
Simulation box size on a side (h^{-1}Mpc)	1000
Number of CDM particles	1120^3
Initial redshift, z	49
Particle mass, $m_{\rm p}$ (10^{10} $h^{-1}M_\odot$)	4.938
Gravitational force softening length, f_ϵ (h^{-1}kpc)	53

Table 4: The cosmological parameters and the design specifications of the LD-*Carmen* simulations.

matter, baryonic and cosmological constant normalized densities, the Hubble parameter, the amplitude of matter density fluctuations, the primordial scalar spectral index and the simulation box size, respectively. The LD simulations is a suite of 41 independent realizations of dark matter N-body simulations named *Carmen* and have information at $z = 0.13$. Using the Ntropy framework [Gardner et al., 2007], bound groups of dark matter particles (halos) are identified with a parallel friends-of-friends (FOF) code [Davis et al., 1985]. The cosmological parameters and the design specifications of the LD-*Carmen* are listed in Table 4.

We extract 100 mock catalogues from each of the 41 LD-*Carmen* boxes, for a total of 4100 mocks. The mocks are randomly centered inside the boxes. They are extracted to mimic the radial distribution of the real catalogues (described in Sec. 5.4.1), as closely as possible.

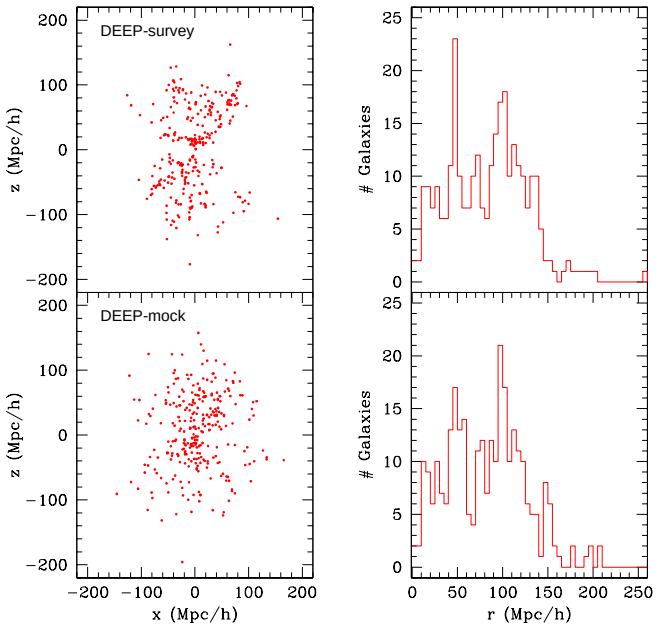

Figure 5.1: Top row: DEEP catalogue (left) and its radial distribution (right). Bottom row: DEEP mock catalogue (left) and its radial distribution (right).

5.4 Peculiar Velocity Catalogues

5.4.1 Real Catalogues

We use a compilation of five galaxy peculiar velocity surveys to study the Mach statistic. This compilation, which we label 'DEEP', includes 103 Type-Ia Supernovae (SNIa) [Tonry et al., 2003], 70 Spiral Galaxy Culsters (SC) Tully-Fisher (TF) clusters [Giovanelli et al., 1998, Dale et al., 1999a], 56 Streaming Motions of Abell Clusters (SMAC) fundamental plane (FP) clusters [Hudson et al., 1999, 2004], 50 Early-type Far galaxies (EFAR) FP clusters [Colless et al., 2001] and 15 TF clusters [Willick, 1999]. In all, the DEEP catalogue consists of 294 data points. In Fig. 5.1, top row, we show the DEEP catalogue (left-hand panel) and its radial distribution (right-hand panel). The bottom row shows a typical mock extracted from the LD simulations. The procedure to extract mocks is described in Sec. 5.4.2.

5.4.2 Mock Catalogues

Inside the N-body simulation box, we first select a point at random. Next, we extract a mock realization of the real catalogue by imposing the constraint that the mock should have a similar radial distribution to the real catalogue. The right-hand panels in Fig. 5.1 show the radial distribution of galaxies in the DEEP catalogue (top) and its mock (bottom). We do not constrain the mocks to have the same angular distribution as the real catalogue for two reasons: (i) the LD simulation boxes are not dense enough to give us mocks that are exact replicas of the real catalogue, (ii) the objects in a real survey are typically weighted depending only on their velocity errors. Consequently, even though the real catalogue and its mocks have similar radial profiles, their angular distribution differ considerably, with the mocks having a relatively featureless angular distribution. To make the mocks more realistic, we impose a $10^{\rm o}$ latitude zone-of-avoidance cut.

Using the angular position $\{\hat{r}_x, \hat{r}_y, \hat{r}_z\}$, the true radial distance d_s from the mock center and the peculiar velocity vector $\mathbf{v}$, we calculate the true line-of-sight peculiar velocity v_s and the redshift $cz = d_s + v_s$ for each mock galaxy, all in km s^{-1}. We then perturb the true radial distance d_s of the mock galaxy with a velocity error drawn from a Gaussian distribution of width equal to the corresponding real galaxy's velocity error, e. Thus, $d_p = d_s + \delta_d$, where d_p is the perturbed radial distance of the mock galaxy (in km s^{-1}) and δ_d is the velocity error drawn from a Gaussian of width e. The mock galaxy's measured line-of-sight peculiar velocity v_p is then assigned to be $v_p = cz - d_p$, where cz is the redshift we found above. This procedure ensures that the weights we assign to the mock galaxies are similar to the weights of the real galaxies.

5.5 The Maximum Likelihood Estimate Method

One of the most common weighting scheme used in analysis of the bulk flow is the maximum likelihood estimate (hereafter MLE) method, obtained from a maximum likelihood analysis introduced by Kaiser [1988]. The motion of galaxies is modeled as being due to a streaming flow with Gaussian distributed measurement uncertainties. Given a pe-

culiar velocity survey, the MLE estimate of its bulk flow is obtained from the likelihood function

$$L[u_i|\{S_n, \sigma_n, \sigma_*\}] = \prod_n \frac{1}{\sqrt{\sigma_n^2 + \sigma_*^2}} \exp\left(\frac{-\frac{1}{2}(S_n - \hat{r}_{n,i}u_i)^2}{\sigma_n^2 + \sigma_*^2}\right), \tag{5.9}$$

where $\hat{\mathbf{r}}_n$ is the unit position vector of the nth galaxy, σ_n is the measurement uncertainty of the nth galaxy and σ_* is the 1-D velocity dispersion accounting for smaller-scale motions. The three components of the bulk flow u_i can be written as weighted sum of the measured radial peculiar velocities of a survey

$$u_i = \sum_n w_{i,n} S_n, \tag{5.10}$$

where S_n is the radial peculiar velocity of the nth galaxy of a survey, and $w_{i,n}$ is the weight assigned to this velocity in the calculation of u_i. Throughout this paper, subscripts i, j and k run over the 3 components of the bulk flow, while subscripts m and n run over the galaxies. Maximizing the likelihood given by Eq. 5.9, gives the three components of the bulk flow u_i with the MLE weights

$$w_{i,n} = \sum_{j=1}^{3} A_{ij}^{-1} \frac{\hat{r}_{n,j}}{\sigma_n^2 + \sigma_*^2}, \tag{5.11}$$

where

$$A_{ij} = \sum_n \frac{\hat{r}_{n,i}\hat{r}_{n,j}}{\sigma_n^2 + \sigma_*^2}. \tag{5.12}$$

The 1-D velocity dispersion σ_* is $1/\sqrt{3}$ of the 3-D velocity dispersion (see Eq. 5.7) which we aim to ultimately measure. Since the weights $w_{i,n}$ (and u_i) are themselves a function of σ_*, we converge on to the MLE estimate for σ_* iteratively. See Strauss et al. [1993] for a discussion on how to estimate the best-fit u_i and σ_* iteratively.

The effective depth of a survey can be roughly estimated by a weighted sum $\sum w_n r_n / \sum w_n$ of the radial distances r_n of the survey objects, where $w_n = 1/(\sigma_n^2 + \sigma_*^2)$. This weighting scheme has been used by Ma, Ostriker, and Zhao [2012] in their analyses of peculiar

velocity datasets. A drawback of using weights $w_n = 1/(\sigma_n^2 + \sigma_*^2)$ in estimating the depth of a survey is that while the weights w_n take into account the measurement noise σ_n, they do not make any corrections for the survey geometry. A better estimate of the effective depth can be made by looking at the survey window functions W_{ij}^2. Window function gives an idea of the scales that contribute to the bulk flow estimates. Ideally, the window function should fall quickly to zero for scales smaller than that being studied. This ensures that the bulk flow estimates are minimally biased from small-scale non-linearities.

Armed with the MLE weights $w_{i,n}$ from Eq. 5.11, the angle-averaged tensor window function $W_{ij}^2(k)$ (equivalent to $W^2(kR)$ of Eq. 5.5) can be constructed (for details, see Feldman et al. [2010]) as

$$\begin{aligned} W_{ij}^2(k) &= \sum_{m,n} w_{i,m} w_{j,n} \int \frac{d^2\hat{k}}{4\pi} \left(\hat{\mathbf{r}}_m \cdot \hat{\mathbf{k}}\right) \left(\hat{\mathbf{r}}_n \cdot \hat{\mathbf{k}}\right) \\ &\quad \times \exp\left(ik\ \hat{\mathbf{k}} \cdot (\mathbf{r}_m - \mathbf{r}_n)\right). \end{aligned} \tag{5.13}$$

The diagonal elements W_{ii}^2 are the window functions of the bulk flow components u_i. The window function gives an idea of the scales that contribute to the bulk flow estimates. Ideally, the window function should fall quickly to zero for scales smaller than that being studied. This ensures that the bulk flow estimates are minimally biased from small-scale non-linearities. See Watkins, Feldman, and Hudson [2009] for the window functions of the bulk flow components for a range of surveys.

Having constructed the survey window functions W_{ii}^2, the effective depth of the survey can be defined to be the one for which W_{ii}^2 is a close match to the window function for an idealized survey. In order to construct the ideal window functions, we first imagine an idealized survey containing radial velocities that well sample the velocity field in a region. This survey consists of a large number of objects, all with zero measurement uncertainty. The radial distribution of this idealized survey is taken to be a Gaussian profile of the form $f(r) \propto e^{-r^2/2R^2}$, where R gives a measure of the depth

of the survey. This idealized survey has easily interpretable bulk flow components that are not affected by small-scale aliasing and which reflect the motion of a well-defined volume.

The MLE weights of an ideal, isotropic survey consisting of N' exact radial velocities $v_{n'}$ measured at randomly selected positions $\mathbf{r}'_{n'}$ are

$$w'_{i,n'} = \sum_{j=1}^{3} A^{-1}_{ij} \frac{\hat{r}'_{n',j}}{N'}, \tag{5.14}$$

where

$$A_{ij} = \sum_{n'=1}^{N'} \frac{\hat{r}'_{n',i}\hat{r}'_{n',j}}{N'}. \tag{5.15}$$

Similar to Eq. 5.13, the window functions ${}^{I}W^2_{ij}$ for an idealized survey of scale R can be constructed as

$$\begin{aligned} {}^{I}W^2_{ij}(k;R) &= \sum_{m,n} w'_{i,m'} w'_{j,n'} \int \frac{d^2\hat{k}}{4\pi} \left(\hat{\mathbf{r}}'_m \cdot \hat{\mathbf{k}}\right) \left(\hat{\mathbf{r}}'_n \cdot \hat{\mathbf{k}}\right) \\ &\quad \times \exp\left(ik\ \hat{\mathbf{k}} \cdot (\mathbf{r}'_m - \mathbf{r}'_n)\right). \end{aligned} \tag{5.16}$$

In Fig. 5.2, left-hand panel, we show the diagonal window functions W^2_{ii} (see Eq. 5.13) of the bulk flow components calculated using MLE weights (see Eq. 5.11) for the DEEP catalogue. The x, y, z components are dot-dashed, short-dashed and long-dashed lines, respectively. Also shown are the ideal window functions ${}^{I}W^2_{ij}$ (see Eq. 5.16) for scales $R = 10 - 40h^{-1}$Mpc (in $5h^{-1}$Mpc increments), the window functions being narrower for larger scales. Comparing the DEEP and the ideal window functions gives the DEEP catalogue an effective depth of $\sim R = 35h^{-1}$Mpc. We note that the weighted sum $\sum w_n r_n / \sum w_n$ gives the DEEP catalogue a depth of 59 h^{-1}Mpc, an over-estimation by nearly 70%. Estimating the survey depth correctly is crucial when it comes to comparing the survey bulk flow with theoretical predictions. One might have a high-quality survey

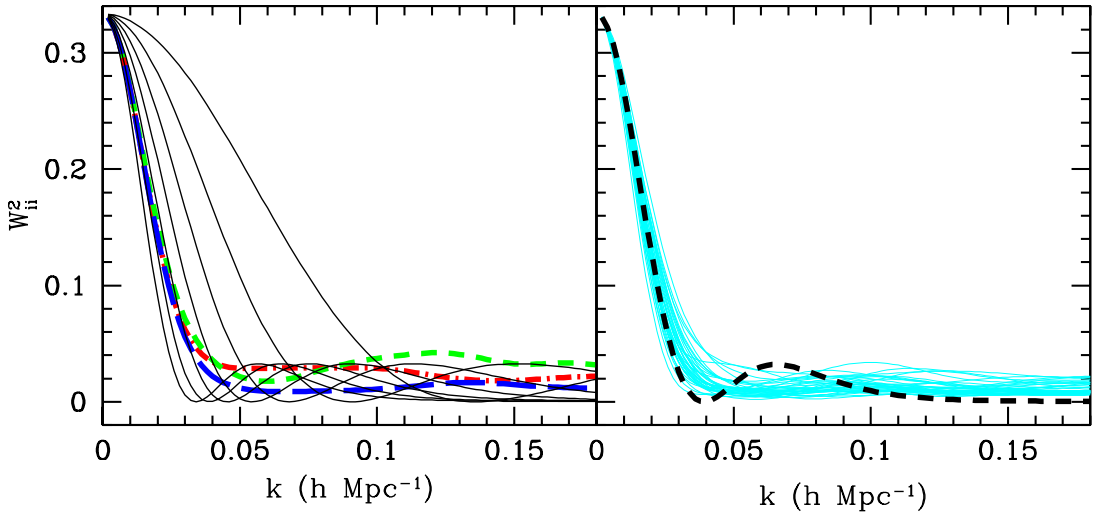

Figure 5.2: The window functions W_{ii}^2 of the bulk flow components calculated using MLE weights for the DEEP catalogue (left-hand panel). The x, y, z components are dot-dashed, short-dashed and long-dashed lines, respectively. The solid lines are the ideal window functions ${}^{I}W_{ij}^2$ for scales $R = 10 - 40h^{-1}$Mpc (in $5h^{-1}$Mpc increments), the window functions being narrower for larger scales. The window functions for a subset of 4100 DEEP mocks are shown in the right-hand panel (solid lines). The characteristic depth of the DEEP catalogue and its mocks is $\sim R = 35h^{-1}$Mpc (dashed line, right-hand panel).

but a poorly estimated depth can introduce substantial errors when comparing with theory. Throughout our work, we define the characteristic depth R of a survey as the one estimated from its window functions. The right-hand panel shows the window functions for a subset of 4100 DEEP mocks (solid lines). The fact that the mock window functions are nearly centered on the $\sim R = 35h^{-1}$Mpc ideal window, shows that our procedure for mock extraction works well.

5.6 Cosmic Mach Number Statistics

5.7 Mach statistics for DEEP mocks

For each of the 4100 DEEP mock realizations, using the MLE weighting scheme (Sec. 5.5), we estimated the bulk flow moments $\{u_x, u_y, u_z\}$, the velocity dispersion σ and the cosmic Mach number M. In Fig. 5.3, we show the probability distribution for the 4100 DEEP mocks: bulk flow u (left-hand panel), dispersion σ (middle panel) and

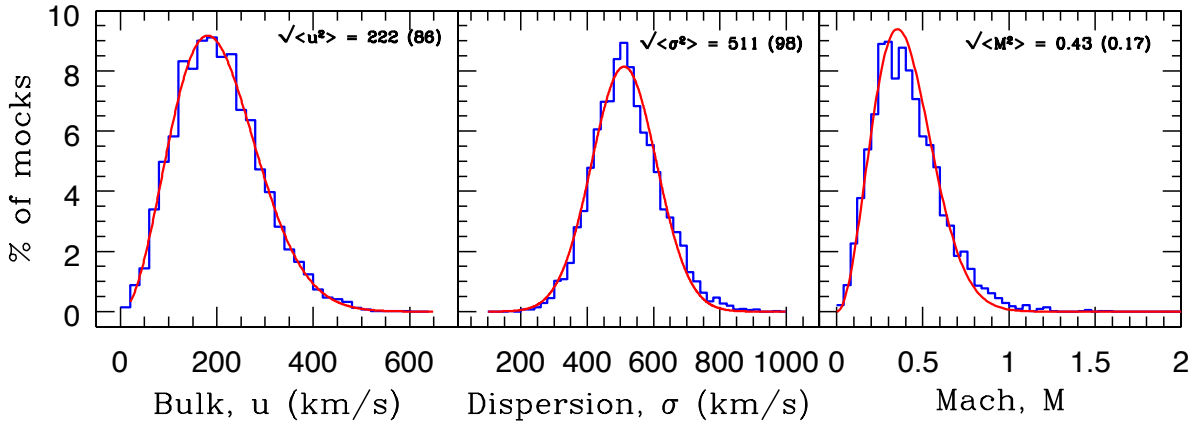

Figure 5.3: Histograms showing the normalized probability distribution for the 4100 DEEP mocks: bulk flow u (left-hand panel), dispersion σ (middle panel) and the cosmic Mach number M (right-hand panel). We also superimpose the best-fitting Maxwellian (for bulk and Mach) and Gaussian (for dispersion) distributions with the same widths as the corresponding histograms. The *rms* values and the 1σ CL intervals are mentioned within each panel. These results correspond to the LD cosmology.

the cosmic Mach number M (right-hand panel). We found the *rms* bulk flow to be $\sigma_v = 222 \pm 86$ km s^{-1}with a velocity dispersion of $\sigma = 511 \pm 98$ km s^{-1}. Together this implies $M = 0.43 \pm 0.17$ at 1σ CL. Since the DEEP mocks have a characteristic depth of $R = 35h^{-1}$Mpc, we can say that for the LD cosmology, the expected Mach number on scales of $R = 35h^{-1}$Mpc is $M = 0.43 \pm 0.17$.

5.8 Mach statistics for Gaussian mocks

In order to find the expected Mach number as a function of scale R for the LD cosmology, we went to the same central points for each of the 4100 DEEP mocks and computed the weighted average of the velocities of all the galaxies in the simulation box, the weighting function being $e^{-r^2/2R^2}$. We repeated this for a range of scales between $R = 10 - 100h^{-1}$Mpc in increments of $5h^{-1}$Mpc. In Fig. 5.4, we show the expected values for the bulk, dispersion and Mach number (dashed line) together with their 1σ CL intervals. The corresponding values for the 4100 DEEP mocks are shown by a solid circle at the characteristic scale $R = 35h^{-1}$Mpc.

The expected bulk ($\sigma_v = 234 \pm 94$ km s^{-1}), dispersion ($\sigma = 517 \pm 56$ km s^{-1})

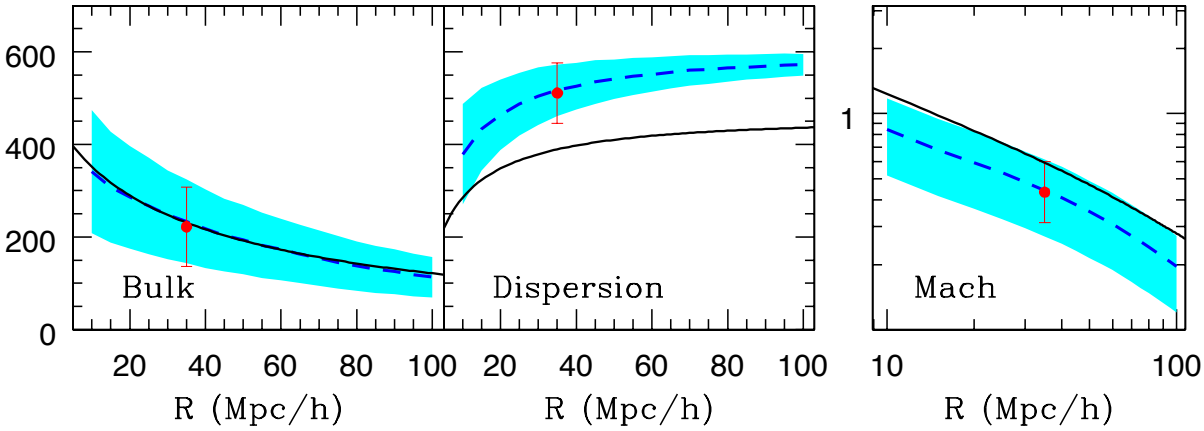

Figure 5.4: The *rms* values of the bulk flow (left-hand panel), dispersion (middle panel) and the cosmic Mach number (right-hand panel) are plotted as a function of scale R. In each panel, the dashed line corresponds to measurements from the Gaussian mocks. The shaded region is the 1σ CL interval for the Gaussian mocks. The solid circle at $R = 35h^{-1}$Mpc is for the DEEP mocks. The error bar is the statistical variance calculated from the 4100 DEEP mocks. Linear theory predictions are shown by solid line. These results correspond to the LD cosmology.

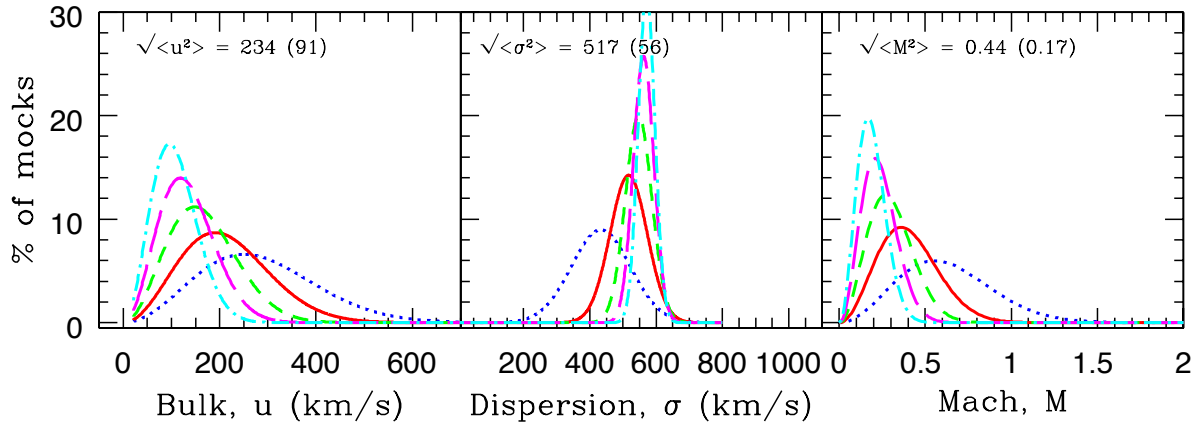

Figure 5.5: The same as Fig. 5.3, but for Gaussian window with $R = 15h^{-1}$Mpc (dotted), $R = 35h^{-1}$Mpc (solid), $R = 55h^{-1}$Mpc (short-dashed), $R = 75h^{-1}$Mpc (long-dashed), $R = 95h^{-1}$Mpc (dot-dashed). For clarity, instead of the histograms, only the best-fitting Maxellian/Gaussian distributions with the same widths as the corresponding histograms are shown. The *rms* values and the 1σ CL intervals for $R = 35h^{-1}$Mpc are mentioned within each panel, and are in good agreement with the corresponding values for the DEEP mocks (shown in Fig. 5.3). Table 5 summarizes the results for Gaussian widths $R = 10 - 100\ h^{-1}$Mpc.

and Mach number ($\sigma = 0.44 \pm 17$) for Gaussian window with $R = 35h^{-1}$Mpc are in excellent agreement with the corresponding values for the DEEP mocks. This shows that the DEEP catalogue probes scales up to $\sim R = 35h^{-1}$Mpc, and not $R = 59h^{-1}$Mpc as one would have inferred from $\sum w_n r_n / \sum w_n$ using the weights $w_n = 1/(\sigma_n^2 + \sigma_*^2)$.

Linear theory predictions for the LD cosmology are shown by the solid lines in Fig. 5.4. The onset of non-linear growth in structure formation at low redshifts boosts the velocity dispersion on all scales, causing linear theory to over-predict the Mach values.

The probability distributions for u, σ and M are plotted in Fig. 5.5 for a range of Gaussian widths R. For clarity, we only show scales $R = 15$, 35, 55, 75 and $95h^{-1}$Mpc. For reference, results for scales $R = 10 - 100h^{-1}$Mpc are summarized in Table 5 as well.

As expected, the *rms* bulk flow (dispersion) is a declining (increasing) function of scale R (see Figs. 5.4 and 5.5). This can be readily understood from the ideal window functions in Fig. 5.2. Larger scales have narrower window functions in the Fourier space. Only the small scale modes ($k \propto 1/R$) contribute to the *rms* bulk flow integral in Eq. 5.5, resulting in smaller bulk flow on larger scales. The dispersion integral (see Eq. 5.7) gets most of its contribution from higher $k-$values ($k > 1/R$) and gradually increases with narrower windows. Similar histogram trends were found by Suto et al. [1992] from numerical simulations of a CDM universe.

5.9 Mach statistics for other mocks

In addition to testing the Gaussian mocks against the DEEP mocks (see Fig. 5.4), we compared the Gaussian mocks with the mocks (4100 each) of the SBF (Surface Brightness Fluctuations) [Tonry et al., 2001], ENEAR (Early-type Nearby Galaxies) [da Costa et al., 2000, Bernardi et al., 2002, Wegner et al., 2003], SFI++, SNIa and SC peculiar velocity surveys. Note that the SC and SNIa surveys are also part of our DEEP compilation. The SFI++ (Spiral Field I-band) catalogue [Masters et al., 2006, Springob et al., 2007, 2009] is the densest and most complete peculiar velocity survey

R (h^{-1}Mpc)	$\sqrt{< u^2 >}$ (km s^{-1})	$\sqrt{< \sigma^2 >}$ (km s^{-1})	$\sqrt{< M^2 >}$
10	341 ± 133	379 ± 108	0.85 ± 0.33
15	308 ± 120	433 ± 89	0.68 ± 0.27
20	286 ± 111	464 ± 76	0.59 ± 0.23
25	267 ± 104	487 ± 68	0.53 ± 0.21
30	248 ± 96	504 ± 62	0.48 ± 0.19
35	234 ± 91	517 ± 56	0.44 ± 0.17
40	218 ± 85	526 ± 50	0.41 ± 0.16
45	204 ± 79	535 ± 47	0.38 ± 0.15
50	194 ± 75	541 ± 43	0.35 ± 0.14
55	182 ± 71	547 ± 40	0.33 ± 0.13
60	173 ± 67	551 ± 37	0.31 ± 0.12
65	163 ± 63	556 ± 35	0.29 ± 0.11
70	154 ± 60	560 ± 33	0.27 ± 0.11
75	145 ± 57	562 ± 31	0.26 ± 0.10
80	137 ± 53	565 ± 29	0.24 ± 0.09
85	130 ± 51	567 ± 27	0.23 ± 0.09
90	125 ± 48	569 ± 26	0.22 ± 0.08
95	118 ± 46	571 ± 25	0.21 ± 0.08
100	113 ± 44	572 ± 23	0.20 ± 0.07

Table 5: The *rms* values of the bulk flow (2nd column), velocity dispersion (3rd column) and cosmic Mach number (4th column) together with their 1σ CL intervals for Gaussian windows with width R (1st column). These values are calculated for the LD cosmology (for the LD parameters, see Table 4).

of field spirals to date. We use data from Springob et al. [2009]. The sample consists of 2720 TF field galaxies (SFI++f) and 736 groups (SFI++g).

In Fig. 5.6, left-hand panels, we show the window functions W_{ii}^2 of the bulk flow components for the SBF, ENEAR, SFI++g, SNIa, SFI++f, DEEP and SC catalogues (top to bottom row, respectively). The right-hand panels show the window functions for a subset of the corresponding mocks. Comparing the window functions of the real catalogues with those of the ideal ones (solid lines in the left-hand panels), we estimate the characteristic depths of the SBF, ENEAR, SFI++g, SNIa, SFI++f, DEEP and SC catalogues to be R = 10, 19, 20, 23, 30, 35 and $40h^{-1}$Mpc, respectively. The window functions for these depths are shown in the right-hand panels (dashed lines). It is

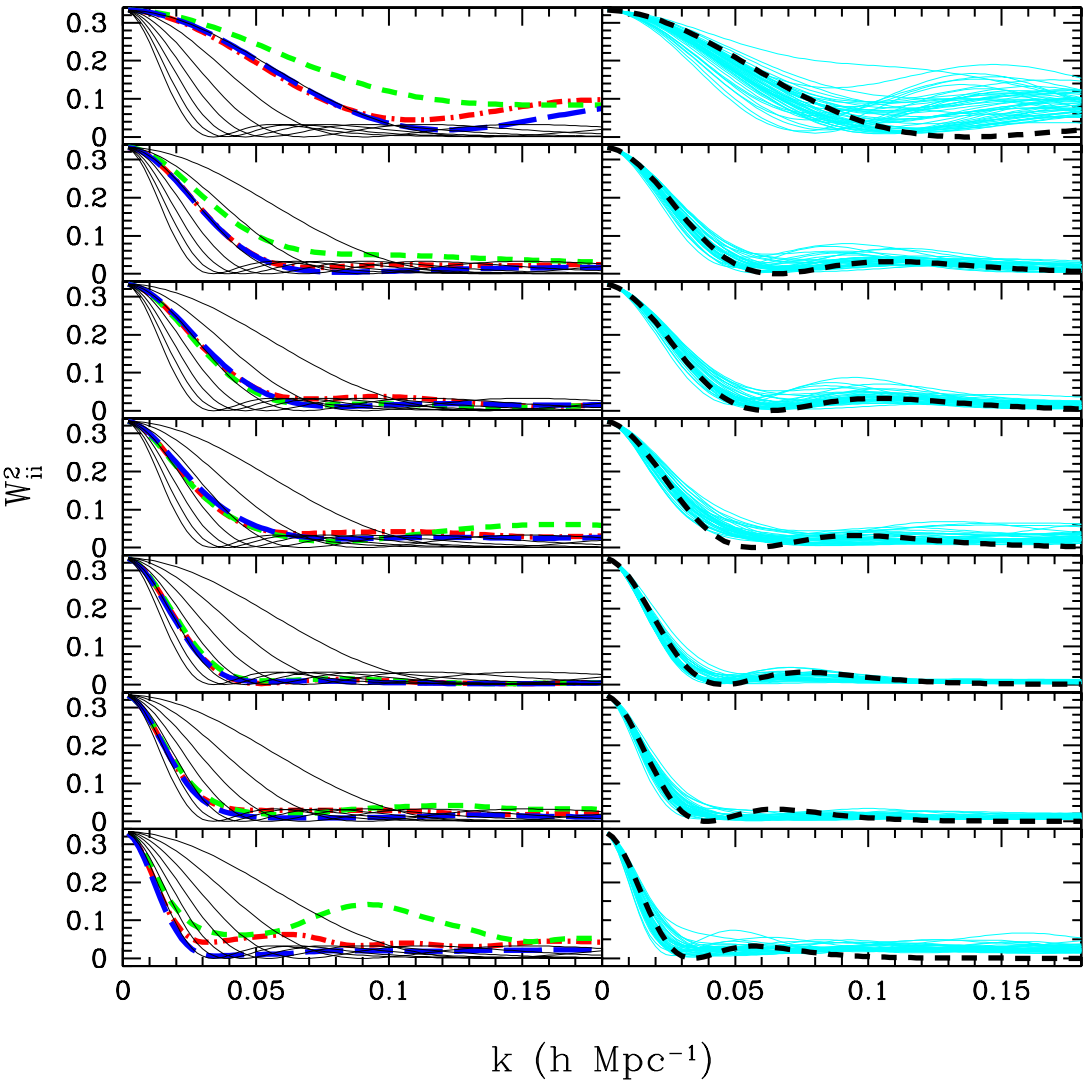

Figure 5.6: Similar to Fig. 5.2, for the SBF, ENEAR, SFI++g, SNIa, SFI++f, DEEP and SC catalogues (top to bottom row, respectively).

Mock Survey	R	$\frac{\sum w_n r_n}{\sum w_n}$	$\sqrt{< u^2 >}$	$\sqrt{< \sigma^2 >}$	$\sqrt{< M^2 >}$
	(h^{-1}Mpc)	(h^{-1}Mpc)	(km s^{-1})	(km s^{-1})	
SBF	10	19	322 ± 125	415 ± 100	0.74 ± 0.29
ENEAR	19	34	262 ± 102	490 ± 104	0.53 ± 0.21
SFI++g	20	35	280 ± 101	473 ± 66	0.59 ± 0.18
SNIa	23	42	275 ± 107	465 ± 73	0.58 ± 0.21
SFI++f	30	52	240 ± 86	510 ± 81	0.47 ± 0.15
DEEP	35	59	222 ± 86	511 ± 65	0.43 ± 0.17
SC	40	75	227 ± 88	485 ± 43	0.47 ± 0.15

Table 6: Peculiar velocity statistics for various surveys (1st column). For each survey, 4100 mocks were extracted from the LD cosmology (for the LD parameters, see Table 4). The characteristic depth R (2nd column) of the mock catalogues is estimated from the effective width of their window functions shown in Fig. 5.6. For reference, the error-weighted depth $\sum w_n r_n / \sum w_n$ where $w_n = 1/(\sigma_n^2 + \sigma_*^2)$, is mentioned in the 3rd column. The *rms* values of the bulk flow (4th column), velocity dispersion (5th column) and cosmic Mach number (6th column) are mentioned together with their 1σ CL intervals.

worth mentioning here that if one defines the characteristic depth of a survey to be the error-weighted depth $\sum w_n r_n / \sum w_n$ where $w_n = 1/(\sigma_n^2 + \sigma_*^2)$, these numbers change to 19, 34, 35, 42, 52, 59 and $75h^{-1}$Mpc, respectively. The characteristic depth R (based on the window functions in Fig. 5.6) and the error-weighted depth $\sum w_n r_n / \sum w_n$ are mentioned in columns 2 and 3, respectively, of Table 6. The surveys are arranged in order of increasing characteristic depth R. The expected values for the bulk, dispersion and Mach number and their 1σ CL intervals for the mocks are summarized in columns $4 - 6$.

Similar to Fig. 5.4, we show results for the SBF, ENEAR, SFI++g, SNIa, SFI++f and SC mocks in Fig. 5.7. Except for the SBF and SC catalogues, the results for the other catalogues are a close match to their Gaussian counterparts. Our SBF mocks are deeper than the real SBF survey because the LD simulations are not dense enough to extract mocks with depths less than $\sim R = 12h^{-1}$Mpc. This explains why the SBF window functions for the mocks (see Fig. 5.6, first row, right-hand panel) are narrower than the one for the SBF's depth of $R = 10h^{-1}$Mpc. Narrower window functions

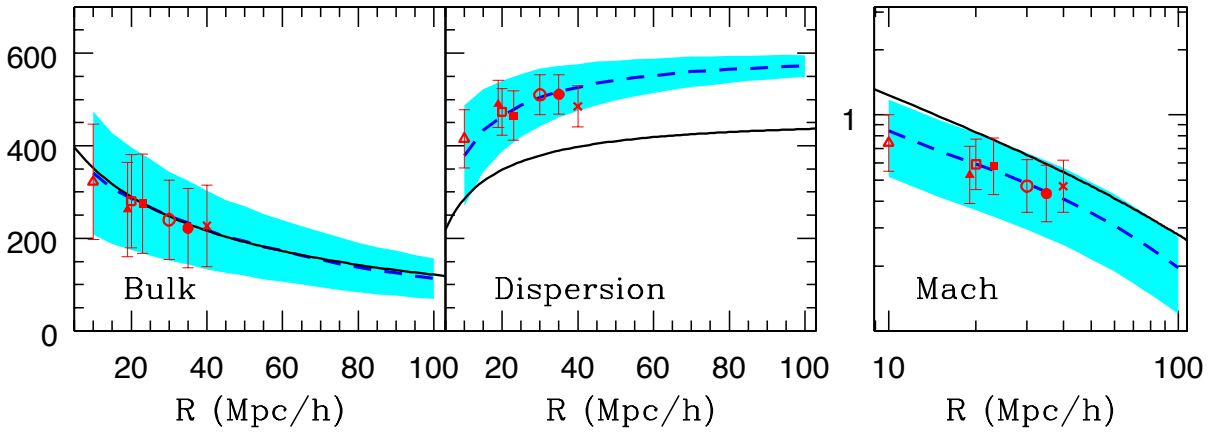

Figure 5.7: Similar to Fig. 5.4, including results for the SBF (open triangle), ENEAR (solid triangle), SFI++g (open square), SNIa (solid square), SFI++f (open circle), DEEP (solid circle) and SC (cross) mocks. The DEEP compilation includes the SC, SNIa, SMAC, EFAR and Willick surveys.

decrease (increase) our bulk flow (dispersion) estimates for the SBF mocks. For the SC mocks, the bulk flow (dispersion) gets excess (suppressed) contribution from smaller scales due to the extended tails of the window functions (see Fig. 5.6, row seven). The SC catalogue, with only 70 clusters, does not have a good sky coverage. The DEEP compilation, however, has a much better sky coverage, and the results (see Fig. 5.7, solid circle) match those from $R = 35h^{-1}$Mpc Gaussian mocks. We have included the results for the SBF and SC catalogues to specifically show that if the selection function of the real survey is not properly modeled, the predictions (in our case, based on Gaussian selection function) can be misleading.

For reasonably dense and well sampled velocity surveys, like DEEP, SFI++f, and SFI++g, a close match between the mock and the Gaussian results shows that the Mach analysis for such catalogues is not overly sensitive to the selection functions of the individual mocks. As such, one can skip the step of extracting mock realizations of the observations from N-body simulations, and simply use Mach predictions based on Gaussian selection function $e^{-r^2/2R^2}$ with R set to the characteristic depth of the survey being studied.

5.10 Moving Beyond N-Body Simulations: Mach Predictions Using PkANN

In Sec. 5.6, we showed that for velocity surveys with low contamination from small scales, reasonably accurate predictions for the Mach number can be made by extracting mocks having a Gaussian radial profile $e^{-r^2/2R^2}$, R being the characteristic depth of the survey being studied.

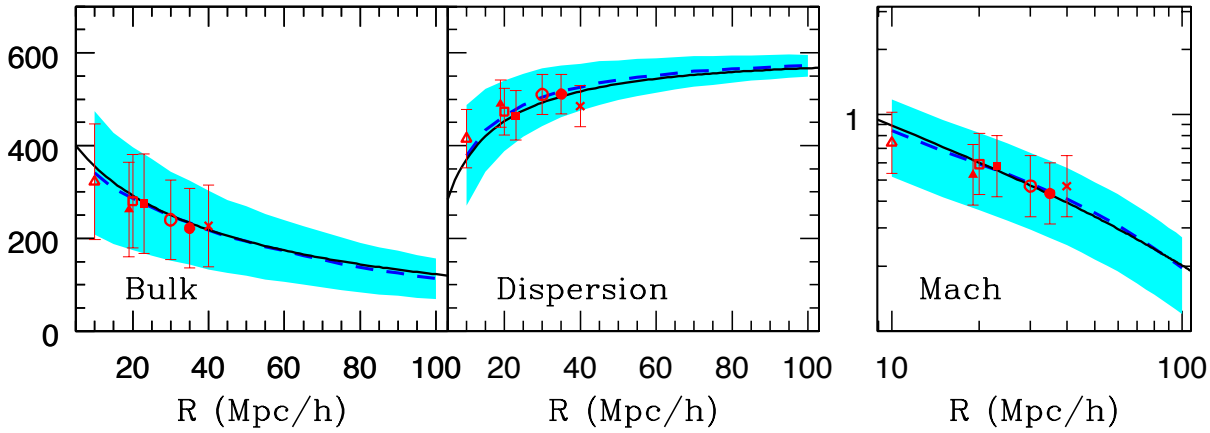

Figure 5.8: Similar to Fig. 5.7, but instead of showing linear theory predictions, we plot predictions based on the non-linear matter power spectrum for the LD cosmology estimated using PkANN.

Another simplification in the Mach analysis one can hope to achieve is to be able to predict $M(R)$ as a function of scale R without resorting to N-body simulations. Running high-resolution N-body simulations, even in the restricted parameter space around 7-yr *WMAP* [Komatsu et al., 2011] central parameters, is beyond present day computing capabilities. It would be much easier and faster to explore the parameter space using a prescription for the matter power spectrum, and using Eq. 5.5 and Eq. 5.7 to predict the cosmic Mach number. So far, this has been possible by using linear theory. However, for the linear theory results to be applicable, as mentioned in Sec. 5, one needs to correct for the non-linearities in the observed velocity field. Any residual non-linearity can still bias the Mach predictions.

In this section, we attempt to predict $M(R)$ using PkANN – an interpolation scheme

to predict the non-linear matter power spectrum up to $k \lesssim 0.9\,h\mathrm{Mpc}^{-1}$ between redshifts $z = 0 - 2$. Although PκANN accuracy worsens (starts under-predicting the non-linear spectrum for $k \gtrsim 0.9\,h\mathrm{Mpc}^{-1}$), we do not attempt to correct this by smoothing the velocity field over the relevant spatial scale.

In Fig. 5.8, we replace linear theory predictions shown in Fig. 5.7, with the ones calculated using PκANN for the LD cosmology. PκANN does an excellent job on all scales, showing it can substitute numerical simulations for the purpose of calculating the Mach number given a set of cosmological parameters. Although, we have shown PκANN's performance for only the LD cosmology, it is expected to perform satisfactorily for cosmologies around 7-yr *WMAP* central parameters for which PκANN has been specifically trained.

5.11 Mach Number Estimates From Real Catalogues

Ma et al. [2012] measured Mach number for four peculiar velocity surveys (SBF, EN-EAR, SNIa and SFI++f) and found that the ΛCDM model with 7-yr *WMAP* parameters is mildly consistent with the Mach number estimates for these four surveys at 3σ CL. However, as the authors mention in their work, their estimates are based on using linear approximation for the power spectrum. Given the fact that at low redshifts structure formation has gone non-linear on small scales, it is necessary to consider non-linearities when making theoretical predictions. Comparing Figs. 5.7 and 5.8 (middle panels), one can see that dispersion is significantly boosted by non-linearities, lowering the Mach predictions (third panels) by 1σ level.

Further, they work with Tophat window functions in their analysis. A Tophat filter assumes a volume-limited survey with a sharp edge in real space. However, the number density of objects sampled in a real survey typically fall at large distances. Real surveys thus have a narrower depth than what a Tophat would suggest. The sharp edge of a Tophat creates extended tails in k-space. Since it is the small scale modes that are most contaminated by non-linearities at low redshifts, a Tophat filter leads to aliasing

of small-scale power onto larger scales. As such, a Tophat filter is not a good choice if one wants to isolate the contribution from small scales.

It is worth mentioning here that using a Gaussian window function $W^2(kR) = e^{-K^2R^2}$ over-damps the high-k tails associated with a Tophat. The reason being that we only observe the line-of-sight component of the velocity field, whereas the equations presented in Sec. 5.2 are based on the full 3D velocity measurements. The line-of-sight component extends the tails of the survey window functions in k-space (see Grinstein et al. 1987, Kaiser 1988). This is the reason why in our analysis, we do not use $W^2(kR) = e^{-K^2R^2}$; instead, we compute the ideal window functions using only the line-of-sight information (see Eq. 5.16). The extended tails of the ideal window functions can be seen in Fig. 5.6 and should be contrasted against $W^2(kR) = e^{-K^2R^2}$.

Ma et al. [2012] estimated the characteristic depth of these surveys using $\sum w_n r_n / \sum w_n$ where $w_n = 1/(\sigma_n^2 + \sigma_*^2)$. Specifically, they found depths of 16.7, 30.5, 30.7 and $50.5h^{-1}$Mpc for the SBF, ENEAR, SNIa and SFI++f, respectively. However, from Fig. 5.6 and Table 6 (rows one, two, four and five), we show that these surveys probe scales of $\sim R = 10$, 19, 23 and $30h^{-1}$Mpc, respectively. Using linear theory with Tophat filters, and neglecting the survey window functions while estimating the effective depths, makes the bulk flow (and any derived) statistic highly complicated to interpret.

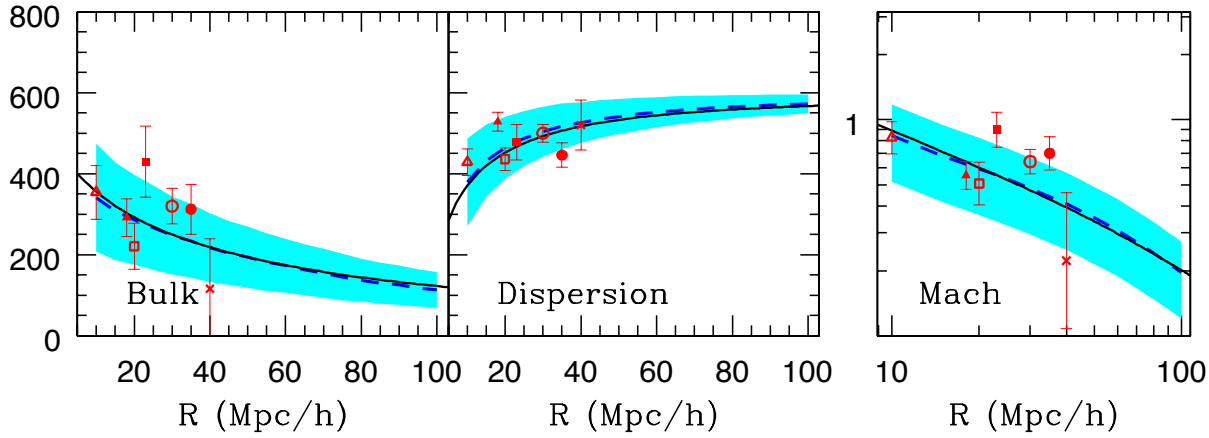

Figure 5.9: Similar to Fig. 5.8, but instead of showing the Mach numbers for the mocks, we plot the Mach numbers for the real surveys. The error bars are calculated using the radial distance uncertainties.

Real Survey	R	u	σ	M
	(h^{-1}Mpc)	(km s^{-1})	(km s^{-1})	
SBF	10	354 ± 66	428	0.83 ± 0.15
ENEAR	19	292 ± 46	528	0.55 ± 0.09
SFI++g	20	221 ± 57	755	0.29 ± 0.08
SNIa	23	430 ± 87	478	0.90 ± 0.18
SFI++f	30	320 ± 44	754	0.42 ± 0.06
DEEP	35	312 ± 61	446	0.70 ± 0.14
SC	40	116 ± 123	520	0.22 ± 0.23

Table 7: Similar to Table 6, but for the real data. The quoted errors are calculated using the radial distance uncertainties.

In Sec. 5.9, we used numerical simulations to study the Mach statistic for SBF, ENEAR, SFI++g, SNIa, SFI++f, DEEP and SC mocks. In this section, we calculate the Mach number using the real catalogues themselves. The results are shown in Fig. 5.9 and summarized in Table 6, columns $7-9$. We find the Mach observations lie within $\sim 1.5\sigma$ interval for a ΛCDM universe with LD parameters. The high uncertainty in the Mach number for the SC catalogue is attributed to its poor sky coverage.

5.12 Summary

The estimates of bulk flow and dispersion on scale R are subject to observational errors stemming from the accuracy levels of distance indicators used, and the survey geometry. Typically, the velocity power spectrum is smoothed using Tophat and Gaussian filters, with results depending on the exact smoothing procedure used. Often, bulk flow results are quoted and inferences drawn about our cosmological model, without paying much attention to the survey window functions which are very useful in determining the scales that contribute to quantities derived from peculiar velocities. A statistic such as the cosmic Mach number can be a useful tool to test theories of structure formation, provided the observational uncertainties are accounted for.

In this paper, we studied the statistical distribution of Mach number by extracting mock realizations of the real peculiar velocity catalogues from numerical simulations

of a ΛCDM universe with 7-yr *WMAP* type cosmology. We showed that the Mach number estimates from the real catalogues agree with the expectations for a ΛCDM universe at $\sim 1.5\sigma$ level. We checked if our Mach expectations derived from mock surveys were biased by the selection function effects: we extracted realizations with a Gaussian profile $f(r) \propto e^{-r^2/2R^2}$ and found no significant change to our Mach values for the mock surveys.

We presented an alternative method to study the cosmic Mach number – by using a prescription for the non-linear matter power spectrum, instead of running time-consuming and computationally-intensive numerical simulations. Non-linear power spectrum interpolators like PKANN offer tremendous leverage over numerical simulations, by being able to explore the parameter space quickly. The role of such interpolating schemes in the study of quantities derived from peculiar velocities needs further investigation.

6 Conclusions

The advent of the era of precision cosmology poses a serious 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, curvature 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 non-linear approximations. 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 7-yr *WMAP* central values. Parameters that are likely to reside by some hard physical prior, such as the neutrino mass, can be successfully brought under the realm of ANNs by sprinkling extra simulations in the corresponding (*e.g.* $\sum m_\nu = 0$) hyper-plane. PkANN is the first power spectrum calculator capable of predicting the non-linear matter power spectrum at sub-per cent level up to redshift $z = 2$ for a range of cosmological models, including massive neutrinos.

As a potential use of PkANN, we studied the cosmic Mach number statistic and found excellent agreement with results obtained directly from the LasDamas numerical simulations. While high-resolution numerical simulations can take millions of hours of computing, the PkANN code estimates the non-linear matter power spectrum in less than a second. However, one must be careful to use the PkANN code *only* for the parameters and ranges that have been simulated and trained with PkANN.

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. We intend to release an automated PkANN function for the scientific community. 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.

7 Appendix

7.1 PkANN Cost Function

For training the PkANN neural network to predict the matter power spectrum, we consider a training set consisting of cosmological models for which we have full information about the non-linear matter power spectra $P_{\rm nl}$ (computed from N-body simulations) as a function of scale k and redshift z, as well as the underlying cosmological parameters: $\mathbf{I} \equiv (\Omega_{\rm m}h^2, \Omega_{\rm b}h^2, n_s, w_0, \sigma_8, \sum m_\nu)$. The joint likelihood of getting the set of matter power spectra $\{P_{\rm nl}(z;\mathbf{I}_t)\}$ for all cosmologies $\mathbf{I}_t$ in the training set is

$$\begin{aligned} Ł\left[\{P_{\rm nl}(z;\mathbf{I}_t)\}\right] &= \prod_{t=1}^{T} p[P_{\rm nl}(z;\mathbf{I}_t)] \\ &= \prod_{t=1}^{T} p[P_{\rm nl}(z|\mathbf{I}_t)]\ p[\mathbf{I}_t], \end{aligned} \tag{7.1}$$

where $p[P_{\rm nl}(z|\mathbf{I}_t)]$ is to be interpreted as the conditional probability of getting spectrum $P_{\rm nl}(z)$ *given* cosmology $\mathbf{I}_t$, while $p[\mathbf{I}_t]$ is the unconditional probability that the cosmological parameters $\mathbf{I}$ take a particular setting of $\mathbf{I}_t$. The sum t is over all the cosmologies $\mathbf{I}_t$ in the training set. We can take the product of the individual probabilities since each model $\mathbf{I}_t$ is drawn independently from the cosmological parameter space.

The weights $\mathbf{w}$ of the PkANN network are chosen (iteratively during network training) so as to minimize the negative logarithm of the likelihood Ł (which is equivalent to maximizing Ł),

$$\chi^2 \;=\; -\ln Ł \;=\; \sum_{t=1}^{T} \ln p[P_{\rm nl}(z|\mathbf{I}_t)] + \sum_{t=1}^{T} \ln p[\mathbf{I}_t]. \tag{7.2}$$

If the power spectrum is sampled at different values of scale k (the k-modes being

represented by the set $\{k\}\,h\mathrm{Mpc}^{-1}$), we can write $p[P_{\rm nl}(z|\mathbf{I}_t)]$ as

$$p[P_{\rm nl}(z|\mathbf{I}_t)] \;=\; \prod_{k_i\in\{k\}} p[P_{\rm nl}(k,z|\mathbf{I}_t)], \tag{7.3}$$

where the product k_i is over all the scales that form the set $\{k\}\,h\mathrm{Mpc}^{-1}$, and we have assumed that $P_{\rm nl}(k,z|\mathbf{I}_t)$ have independent distributions.

To suppress sampling uncertainties in the power spectrum $P_{\rm nl}(k,z|\mathbf{I}_t)$, the numerical simulation code is run multiple times with different seeds while keeping the underlying cosmological model $\mathbf{I}_t$ fixed. Assuming $P_{\rm nl}(k,z|\mathbf{I}_t)$ has Gaussian distribution about the true power spectrum $P^{\rm Tr}_{\rm nl}(k,z|\mathbf{I}_t)$ with variance σ^2, we can write the probability that a numerical run would give $P_{\rm nl}(k,z|\mathbf{I}_t)$ as

$$p[P_{\rm nl}(k,z|\mathbf{I}_t)] \;=\; \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-\left[P^{\rm Tr}_{\rm nl}(k,z|\mathbf{I}_t)-P_{\rm nl}(k,z|\mathbf{I}_t)\right]^2}{2\sigma^2}}. \tag{7.4}$$

N-body codes give larger variance σ^2 on scales comparable to the simulation volume since the density field on these scales can only be sampled fewer times. However, to simplify the PkANN training algorithm, in Eq. 7.4 we have assumed that the variance σ^2 is independent of the scale k and model $\mathbf{I}_t$.

Since the aim of developing PkANN is to model the true spectrum $P^{\rm Tr}_{\rm nl}(k,z|\mathbf{I}_t)$ by making an optimal choice for the network weights $\mathbf{w}$, we replace $P^{\rm Tr}_{\rm nl}(k,z|\mathbf{I}_t)$ in Eq. 7.4 by the ANN prediction $P^{\rm ANN}_{\rm nl}(k,z|\mathbf{w},\mathbf{I}_t)$ to get

$$p[P_{\rm nl}(k,z|\mathbf{I}_t)] \;=\; \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-\left[P^{\rm ANN}_{\rm nl}(k,z|\mathbf{w},\mathbf{I}_t)-P_{\rm nl}(k,z|\mathbf{I}_t)\right]^2}{2\sigma^2}}. \tag{7.5}$$

Inserting Eq. 7.5 into Eq. 7.3, we get

$$p[P_{\rm nl}(z|\mathbf{I}_t)] \;=\; \frac{1}{(2\pi\sigma^2)^{N_{out}/2}} e^{\frac{-\sum_{k_i\in\{k\}}\left[P^{\rm ANN}_{\rm nl}(k,z|\mathbf{w},\mathbf{I}_t)-P_{\rm nl}(k,z|\mathbf{I}_t)\right]^2}{2\sigma^2}}, \tag{7.6}$$

where N_{out} is the number of k-modes in the set $\{k\}$. Remember that, by construction,

N_{out} is also the number of nodes in the output layer of the PkANN network. Using Eq. 7.6, we can now write Eq. 7.2 as

$$\begin{aligned}\chi^2(\mathbf{w}) &= \frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{k_i\in\{k\}}\left[P_{\rm nl}^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t)-P_{\rm nl}(k,z|\mathbf{I}_t)\right]^2 \\ &\quad - T\ln\left[\frac{1}{(2\pi\sigma^2)^{N_{out}/2}}\right]+\sum_{t=1}^{T}\ln p[\mathbf{I}_t]. \end{aligned} \tag{7.7}$$

We can drop the terms that do not depend on the weights $\mathbf{w}$, since these terms merely scale $\chi^2(\mathbf{w})$ without altering its location in the weight-space. Thus, the cost function for the purposes error minimization can be written as

$$\chi^2(\mathbf{w}) = \frac{1}{2}\sum_{t=1}^{T}\sum_{k_i\in\{k\}}\left[P_{\rm nl}^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t)-P_{\rm nl}(k,z|\mathbf{I}_t)\right]^2. \tag{7.8}$$

7.2 Quasi-Newton Method

Quasi-Newton method, used for finding stationary points (local maxima and minima) of a function, assumes that the function can be approximated by a quadratic in the region around a stationary point. Taylor expanding the PkANN cost function $\chi^2_C(\mathbf{w})$ (see Eq. 3.8) around some point $\mathbf{w}_0$ in the weight space and retaining terms up to second-order, we get

$$\chi^2_C(\mathbf{w}) = \chi^2_C(\mathbf{w}_0)+(\mathbf{w}-\mathbf{w}_0)^T\mathbf{g}_{\mathbf{w}_0}+\frac{1}{2}(\mathbf{w}-\mathbf{w}_0)^T\mathbf{H}_{\mathbf{w}_0}(\mathbf{w}-\mathbf{w}_0), \tag{7.9}$$

where the superscript T stands for the transpose and $\mathbf{g}_{\mathbf{w}_0}$ is defined to be the gradient of χ^2_C evaluated at $\mathbf{w}_0$,

$$\mathbf{g}_{\mathbf{w}_0}\equiv\bigtriangledown\chi^2_C\big|_{\mathbf{w}_0}. \tag{7.10}$$

$\mathbf{H}_{\mathbf{w}_0}$ is a symmetric $N_W \times N_W$ Hessian matrix (evaluated at $\mathbf{w}_0$) with elements

$$H_{ij}\big|_{\mathbf{w}_0} \equiv \left.\frac{\partial^2 \chi^2_C}{\partial w_i \partial w_j}\right|_{\mathbf{w}_0}, \tag{7.11}$$

where N_W (see Eq. 3.1) is the total number of nodes in the network. Note that in Eq. 7.11, instead of referencing the weights by the relevant nodes they connect to, for the sake of clarity we refer to the weights with a single subscript running from $1 - N_W$.

Taking the gradient of Eq. 7.9 gives the local approximation for the gradient itself,

$$\mathbf{g}_{\mathbf{w}} \;=\; \mathbf{g}_{\mathbf{w}_0} + \mathbf{H}_{\mathbf{w}_0}(\mathbf{w} - \mathbf{w}_0). \tag{7.12}$$

To find the stationary point around $\mathbf{w}_0$, one sets $\mathbf{g}_{\mathbf{w}}$ in Eq. 7.12 to zero, thereby giving the *Newton step*,

$$\mathbf{w} \;=\; \mathbf{w}_0 - \mathbf{H}^{-1}_{\mathbf{w}_0}\mathbf{g}_{\mathbf{w}_0}. \tag{7.13}$$

Since the cost function $\chi^2_C(\mathbf{w})$ is not an exact quadratic function, the Newton step of Eq. 7.13 does not point to the local minimum around $\mathbf{w}_0$. As such, we apply Eq. 7.13 iteratively, and if the Hessian matrix is positive definite (*i.e.* all of its eigenvalues are positive), then each successive Newton step moves closer to the local minimum. If the initial choice of the weights $\mathbf{w}$ happens to be around a local maximum of $\chi^2_C(\mathbf{w})$, then the Hessian matrix is not positive definite and the cost function may increase with each Newton step.

One can apply some modifications to the Newton method that guarantee convergence towards a local minimum, irrespective of the initial choice of the weights. Instead of taking a step in the *Newton direction* ($-\mathbf{H}^{-1}\mathbf{g}$), one proceeds in a *quasi-Newton direction* ($-\mathbf{G}\mathbf{g}$),

$$\mathbf{w} \;=\; \mathbf{w}_0 - \lambda_{\mathbf{w}_0}\mathbf{G}_{\mathbf{w}_0}\mathbf{g}_{\mathbf{w}_0}, \tag{7.14}$$

where matrix $\mathbf{G}$ represents an approximation to the inverse of the Hessian $\mathbf{H}^{-1}$, and λ is the size of the step taken along the quasi-Newton direction $-\mathbf{Gg}$. The step size λ is allowed to vary with each iteration to the weights. Its value is determined by proceeding in the direction $-\mathbf{Gg}$ until the minimum of the cost function is found along $-\mathbf{Gg}$. Thus, in Eq. 7.14, $\lambda_{\mathbf{w_0}}$ is such that the gradient of χ^2_C at $\mathbf{w}$ (namely, $\mathbf{g_w}$) vanishes along the direction $-\mathbf{G}_{\mathbf{w}_0}\mathbf{g}_{\mathbf{w}_0}$,

$$(-\mathbf{G}_{\mathbf{w}_0}\mathbf{g}_{\mathbf{w}_0})^T\,\mathbf{g}_{\mathbf{w}} \;=\; 0. \tag{7.15}$$

The quasi-Newton algorithm involves taking a a series of steps τ of Eq. 7.14, which can be written as

$$\mathbf{w}_{\tau+1} \;=\; \mathbf{w}_\tau - \lambda_{\mathbf{w}_\tau}\mathbf{G}_{\mathbf{w}_\tau}\mathbf{g}_{\mathbf{w}_\tau}, \tag{7.16}$$

with the step size $\lambda_{\mathbf{w}_\tau}$ for the τth step being such that

$$(-\mathbf{G}_{\mathbf{w}_\tau}\mathbf{g}_{\mathbf{w}_\tau})^T\,\mathbf{g}_{\mathbf{w}_{\tau+1}} \;=\; 0. \tag{7.17}$$

At each step of the algorithm, $\mathbf{G}$ is constructed to be positive definite, ensuring that the direction $-\mathbf{Gg}$ proceeds towards a local minimum of the cost function. To construct $\mathbf{G}$, we use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (see Appendix 7.4).

7.3 PkANN Cost Function Gradient

The overall cost function which is presented to the ANN for minimization with respect to the weights $\mathbf{w}$ is given by (see Eq. 3.8),

$$\chi^2_C(\mathbf{w}) \;=\; \frac{1}{2}\sum_{t=1}^{T}\sum_{k_i\in\{k\}}\left[R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t) - R(k,z|\mathbf{I}_t)\right]^2 + \frac{\xi}{2}||\mathbf{w}||^2. \tag{7.18}$$

We now derive the expression for the derivative with respect to the weights $\mathbf{w}$.

PKANN's network architecture is $N_{in} : N_1 : N_{out}$ with two layers of adaptive weights. The first layer of weights w_{ji} connect the input layer nodes $(x_0, x_1, ..., x_i)$ to the hidden nodes $(z_1, ..., z_j)$. Note that the hidden bias node activation z_0 is permanently fixed at 1 and therefore, does not receive any connections from the input layer. The activation of each hidden node is $z_j \equiv g(a_j)$, taking as its argument

$$a_j \;=\; \sum_{i=0}^{N_{in}} w_{ji} x_i, \tag{7.19}$$

where the sum is over all input nodes i (including the input bias) sending connections to the jth hidden node (barring the hidden bias node).

PKANN's hidden nodes have sigmoidal activations $g(a_j) = 1/[1 + \exp(-\mathrm{a_j})]$. The second layer of weights w_{kj} connect the hidden nodes $(z_0, z_1, ..., z_j)$ to the network outputs $(y_1, ..., y_k)$. The output nodes have linear activations $y_k = a_k$, with a_k being the weighted sum of all hidden nodes,

$$a_k \;=\; \sum_{j=0}^{N_1} w_{kj} z_j. \tag{7.20}$$

PkANN has two layers of adaptive weights and we will consider the cost function derivatives separately for the two layers.

7.3.1 Gradient w.r.t. First Layer Weights

Taking the gradient of Eq. 7.18 with respect to a first layer weight w_{ji}, we get

$$\frac{\partial \left[\chi^2_C(\mathbf{w})\right]}{\partial w_{ji}} \;=\; \sum_{t,\{k\}} \left[R^{\mathrm{ANN}}(k, z|\mathbf{w}, \mathbf{I}_t) - R(k, z|\mathbf{I}_t)\right] \frac{\partial R^{\mathrm{ANN}}}{\partial w_{ji}} \;+\; \xi w_{ji}. \tag{7.21}$$

Since $R^{\mathrm{ANN}}(k, z|\mathbf{w}, \mathbf{I}_t) = a(k, z|\mathbf{w}, \mathbf{I}_t)$ (see Eq. 7.20) for the output nodes, we get

$$\frac{\partial \left[\chi^2_C(\mathbf{w})\right]}{\partial w_{ji}} \;=\; \sum_{t,\{k\}} \left[R^{\mathrm{ANN}}(k, z|\mathbf{w}, \mathbf{I}_t) - R(k, z|\mathbf{I}_t)\right] \frac{\partial a^t_k}{\partial w_{ji}} \;+\; \xi w_{ji}, \tag{7.22}$$

where we have introduced the shorthand notation $a_k^t \equiv a(k, z|\mathbf{w}, \mathbf{I}_t)$. Using Eq. 7.20 for a_k together with sigmoidal activation for z_j, we get

$$\begin{aligned} \frac{\partial a_k^t}{\partial w_{ji}} &= \sum_{j'=0}^{N_1} w_{kj'} \frac{\partial z_{j'}^t}{\partial w_{ji}} \\ &= \sum_{j'=0}^{N_1} w_{kj'} \frac{\partial g(a_{j'}^t)}{\partial a_{j'}^t} \frac{\partial a_{j'}^t}{\partial w_{ji}}. \end{aligned} \tag{7.23}$$

For sigmoidal activation functions, it is straightforward to show that

$$\frac{\partial g(a_j^t)}{\partial a_j^t} = g(a_j^t)\left(1 - g(a_j^t)\right). \tag{7.24}$$

Inserting Eq. 7.24 into Eq. 7.23, we get

$$\frac{\partial a_k^t}{\partial w_{ji}} = \sum_{j'=0}^{N_1} w_{kj'} g_{j'}^t \left(1 - g_{j'}^t\right) \frac{\partial a_{j'}^t}{\partial w_{ji}}. \tag{7.25}$$

Differentiating Eq. 7.19 with respect to the weights w_{ji}, we get

$$\begin{aligned} \frac{\partial a_{j'}^t}{\partial w_{ji}} &= \sum_{i'=0}^{N_{in}} x_{i'}^t \frac{\partial w_{j'i'}}{\partial w_{ji}} \\ &= \sum_{i'=0}^{N_{in}} x_{i'}^t \delta_{ii'} \delta_{jj'} \;=\; x_i^t \delta_{jj'}. \end{aligned} \tag{7.26}$$

Inserting Eq. 7.26 into Eq. 7.25, we get

$$\begin{aligned} \frac{\partial a_k^t}{\partial w_{ji}} &= \sum_{j'=0}^{N_1} w_{kj'} g_{j'}^t \left(1 - g_{j'}^t\right) x_i^t \delta_{jj'} \\ &= w_{kj} g_j^t \left(1 - g_j^t\right) x_i^t. \end{aligned} \tag{7.27}$$

From Eqs. 7.22 and 7.27, we get our final equation for the derivative of the PkANN

cost function with respect to the first layer of adaptive weights w_{ji} to be

$$\frac{\partial \left[\chi^2_C(\mathbf{w})\right]}{\partial w_{ji}} = \sum_{t,\{k\}} \left[R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t) - R(k,z|\mathbf{I}_t)\right] w_{kj} g^t_j \left(1-g^t_j\right) x^t_i \; + \; \xi w_{ji}. \quad (7.28)$$

7.3.2 Gradient w.r.t. Second Layer Weights

Taking the gradient of Eq. 7.18 with respect to a second layer weight w_{kj}, we get

$$\frac{\partial \left[\chi^2_C(\mathbf{w})\right]}{\partial w_{kj}} \;=\; \sum_{t,\{k'\}} \left[R^{\rm ANN}(k',z|\mathbf{w},\mathbf{I}_t) - R(k',z|\mathbf{I}_t)\right] \frac{\partial R^{\rm ANN}}{\partial w_{kj}} \; + \; \xi w_{kj}. \quad (7.29)$$

Since $R^{\rm ANN}(k',z|\mathbf{w},\mathbf{I}_t) = a(k',z|\mathbf{w},\mathbf{I}_t)$ (see Eq. 7.20) for the output nodes, we get

$$\frac{\partial \left[\chi^2_C(\mathbf{w})\right]}{\partial w_{kj}} \;=\; \sum_{t,\{k'\}} \left[R^{\rm ANN}(k',z|\mathbf{w},\mathbf{I}_t) - R(k',z|\mathbf{I}_t)\right] \frac{\partial a^t_{k'}}{\partial w_{kj}} \; + \; \xi w_{kj}, \quad (7.30)$$

where as before, we use the shorthand notation $a^t_{k'} \equiv a(k',z|\mathbf{w},\mathbf{I}_t)$. From Eq. 7.20, we get

$$\begin{aligned} \frac{\partial a^t_{k'}}{\partial w_{kj}} \;&=\; \sum_{j'=0}^{N_1} \frac{\partial w_{k'j'}}{\partial w_{kj}} z^t_{j'} \\ &=\; \sum_{j'=0}^{N_1} \delta_{kk'}\delta_{jj'} z^t_{j'} \;=\; \delta_{kk'} z^t_j \end{aligned} \quad (7.31)$$

Inserting Eq. 7.31 into Eq. 7.30, we get our final equation for the derivative of the PkANN cost function with respect to the second layer of adaptive weights w_{kj} to be

$$\begin{aligned} \frac{\partial \left[\chi^2_C(\mathbf{w})\right]}{\partial w_{kj}} \;&=\; \sum_{t,\{k'\}} \left[R^{\rm ANN}(k',z|\mathbf{w},\mathbf{I}_t) - R(k',z|\mathbf{I}_t)\right] \delta_{kk'} z^t_j \; + \; \xi w_{kj} \\ &=\; \sum_{t} \left[R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t) - R(k,z|\mathbf{I}_t)\right] z^t_j \; + \; \xi w_{kj} \end{aligned} \quad (7.32)$$

For any choice of weights $\mathbf{w}$, the network output vector $R^{\rm ANN}(k,z|\mathbf{w},\mathbf{I}_t)$ is determined for each cosmology $\mathbf{I}_t$ in the training set, by progressing sequentially through the

network layers, from inputs to outputs, calculating the activation of each node. Having calculated the activations and network outputs for all cosmologies, it is straightforward to compute the derivatives in Eqs. 7.28 and 7.32.

7.4 BFGS Approximation for Inverse-Hessian Matrix

In order to minimize the PKANN cost function $\chi^2_C(\mathbf{w})$ (see Eq. 3.8) with respect to the weights $\mathbf{w}$, the weights are first randomly initialized to $\mathbf{w}_0$ and then updated iteratively using Eq. 7.16.

Updating the weights involves estimating $\mathbf{G}$ – an approximation to the inverse Hessian matrix $\mathbf{H}^{-1}$. The inverse Hessian $\mathbf{H}^{-1}$ evaluated at $\mathbf{w_0}$ is approximated by a $N_W \times N_W$ identity matrix (*i.e.* $\mathbf{G}_{\mathbf{w}_0} = \mathbf{I}$). Following our discussion in Appendix 7.2, the weight vector is updated to $\mathbf{w}_1$ as

$$\mathbf{w}_1 \;=\; \mathbf{w}_0 - \lambda_{\mathbf{w}_0}\mathbf{g}_{\mathbf{w}_0} \tag{7.33}$$

by stepping a distance $\lambda_{\mathbf{w_0}}$ in the quasi-Newton direction $-\mathbf{g}_{\mathbf{w}_0}$. Note that the gradient $\mathbf{g}_{\mathbf{w}_0}$ is computed using Eqs. 7.28 and 7.32. The step size $\lambda_{\mathbf{w}_0}$ is such that the gradient of χ^2_C at $\mathbf{w}_1$ (namely, $\mathbf{g}_{\mathbf{w}_1}$) vanishes along the direction $-\mathbf{g}_{\mathbf{w}_0}$,

$$-\,\mathbf{g}^T_{\mathbf{w}_0}\mathbf{g}_{\mathbf{w}_1} \;=\; 0. \tag{7.34}$$

To make any further updates in the weight space, one needs to evaluate $\mathbf{H}^{-1}_{\mathbf{w}_1}$. The inverse Hessian, being a $N_W \times N_W$ matrix, can be computationally expensive to calculate exactly for networks with $N_W \gtrsim 1000$ connections. We employ the BFGS method to approximate $\mathbf{H}^{-1}_{\mathbf{w}_1}$ by $\mathbf{G}_{\mathbf{w}_1}$. In general, for the $(\tau+1)$ step, the approximation $\mathbf{G}_{\mathbf{w}_{\tau+1}}$ is

$$\mathbf{G}_{\mathbf{w}_{\tau+1}} \;=\; \mathbf{G}_{\mathbf{w}_\tau} + \frac{1}{S_1}\left[\left(1+\frac{S_2}{S_1}\right)\mathbf{a}\mathbf{a}^T - \mathbf{a}\mathbf{b}^T\mathbf{G}_{\mathbf{w}_\tau} - \mathbf{G}_{\mathbf{w}_\tau}\mathbf{b}\mathbf{a}^T\right], \tag{7.35}$$

where we use the following definitions for the vectors ($\mathbf{a}$ and $\mathbf{b}$) and the scalars (S_1 and

S_2),

$$\begin{aligned} \mathbf{a} &= \mathbf{w}_{\tau+1} - \mathbf{w}_{\tau} \\ \mathbf{b} &= \mathbf{g}_{\mathbf{w}_{\tau+1}} - \mathbf{g}_{\mathbf{w}_{\tau}} \\ S_1 &= \mathbf{a}^T\mathbf{b} \\ S_2 &= \mathbf{b}^T\mathbf{G}\mathbf{b} \end{aligned} \tag{7.36}$$

At each step, the BFGS method makes increasingly more accurate approximations for $\mathbf{G}$. Moreover, since $\mathbf{G}$ is positive definite (by construction), the $\chi^2_C(\mathbf{w})$ minimization algorithm is guaranteed to converge to a local minimum.

7.5 Regularization Parameter ξ

In situations where the training data is noisy, controlling the complexity of a network is crucial to avoid overfitting and underfitting issues. An overly complex network may fit the noise in the training data. On the other hand, a very simple network may not be able to capture the signal in a dataset, leading to underfitting. Both overfitting and underfitting lead to models with low predictive performance. One of the methods employed to regularize the complexity of a neural network is to train the network by minimizing a cost function that includes a penalty term $\chi^2_Q(\mathbf{w})$ (*e.g.* see Eq. 3.7).

Small (large) values of the regularization parameter ξ lead to complex (simple) networks. Since the optimum value for ξ is not known a priori, its value is initialized randomly, and updated iteratively by the cost minimization algorithm.

Here, we only present the updating rule for ξ. For its derivation, refer Bishop [1995]. The PkANN cost function (Eq. 3.8) can be written as

$$\chi^2_C(\mathbf{w}) = \beta \left[\frac{1}{2} \sum_{t=1}^{T} \sum_{k_i \in \{k\}} \left[R^{\mathrm{ANN}}(k, z|\mathbf{w}, \mathbf{I}_t) - R(k, z|\mathbf{I}_t) \right]^2 + \frac{\alpha}{2\beta} ||\mathbf{w}||^2 \right], \tag{7.37}$$

where α and β are the regularization parameters with $\xi \equiv \alpha/\beta$ and $\beta = 1$. For the

purposes of cost minimization, the overall scale factor β is irrelevant and the degree of regularization depends only on the ratio $\xi \equiv \alpha/\beta$. For networks where the number of training patterns N_T far exceeds the number of weights N_W, Bishop [1995] derives the following updating rules for α and β,

$$\alpha_{\tau+1} = \frac{N_W}{||\mathbf{w}_\tau||^2} \tag{7.38}$$

$$\beta_{\tau+1} = \frac{N_T}{\chi^2(\mathbf{w}_\tau)}, \tag{7.39}$$

where $\chi^2(\mathbf{w})$ (see Eq. 3.6) is the sum of squares of residuals for the training data. Thus, we update ξ as

$$\xi_{\tau+1} = \frac{N_W}{N_T}\frac{\chi^2(\mathbf{w}_\tau)}{||\mathbf{w}_\tau||^2}. \tag{7.40}$$

From Eq. 7.40, we see that for sufficiently complex networks ($N_W >> 1$) with lots of training data ($N_T >> N_W$), the parameter $\xi << 1$. It shows that underfitting and overfitting issues can be avoided by simply choosing network architectures that satisfy conditions: (i) $N_W >> 1$ and (ii) $N_T >> N_W$. However, both these conditions can put tremendous load on the computing resources. In situations where the computing time is at a premium, a penalty term is used to achieve a balance between computing load and desired prediction accuracy of the neural network.

7.6 Exact Calculation of Hubble Parameter h

(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 and Sugiyama [1996]:

$$z_{ls} = 1048\left[1+\frac{0.00124}{(\Omega_b h^2)^{0.738}}\right]\left[1+b_1(\Omega_m h^2)^{b_2}\right] \tag{7.41}$$

$$b_1 = \frac{0.0783}{(\Omega_b h^2)^{0.238}}\left[1+39.5(\Omega_b h^2)^{0.763}\right]^{-1} \tag{7.42}$$

$$b_2 = \frac{0.560}{1+21.1(\Omega_b h^2)^{1.81}} \tag{7.43}$$

(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 accounts for the effect of massive neutrinos on $h(a)$:

$$h(a) = h\sqrt{\frac{\Omega_b+\Omega_c}{a^3}+\frac{\Omega_\gamma}{a^4}\left[1+\frac{\Omega_\nu}{\Omega_\gamma}\right]+\frac{\Omega_\Lambda}{a^{3(1+w_0)}}} \tag{7.44}$$

$$\frac{\Omega_\nu}{\Omega_\gamma} = N_{\rm eff}\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}F(y) \tag{7.45}$$

$$F(y) = \frac{120}{7\pi^4}\int_0^\infty \frac{x^2\sqrt{x^2+y^2}}{e^x+1}dx, \tag{7.46}$$

where

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

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

$$\Omega_\gamma = \frac{2.4706\times10^{-5}}{h^2}\left(\frac{{T_\gamma}^0}{2.725}\right)^4.$$

${T_\nu}^0$ 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}:

$$r_s(z_{ls}) \;=\; \frac{c}{\sqrt{3}} \int_{a=0}^{1/(1+z_{ls})} \frac{da}{a^2 h(a) \sqrt{1 + (3\Omega_b/4\Omega_\gamma)a}}. \tag{7.47}$$

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

$$d_{ls} = \frac{302.54}{\pi} r_s(z_{ls}). \tag{7.48}$$

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

$$\chi(z_{ls}) \;=\; c \int_{1/(1+z_{ls})}^{a=1} \frac{da}{a^2 h(a)}. \tag{7.49}$$

(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).

8 References

References

Matthew Turk. Analysis and visualization of multi-scale astrophysical simulations using python and numpy. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 46 – 50, Pasadena, CA USA, 2008.

E. Lawrence, K. Heitmann, M. White, D. Higdon, C. Wagner, S. Habib, and B. Williams. The coyote universe. iii. simulation suite and precision emulator for the nonlinear matter power spectrum. *ApJ*, 713:1322–1331, April 2010. doi: 10.1088/0004-637X/713/2/1322.

The Dark Energy Survey Collaboration. The dark energy survey. *arXiv:astro-ph/0510346*, October 2005.

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

vestri, J. A. Smith, R. C. Smith, M. A. Strauss, C. W. Stubbs, D. Sweeney, A. Szalay, J. J. Thaler, D. Vanden Berk, L. Walkowicz, M. Warner, B. Willman, D. Wittman, S. C. Wolff, W. M. Wood-Vasey, P. Yoachim, H. Zhan, and for the LSST Collaboration. LSST: from Science Drivers to Reference Design and Anticipated Data Products. *arXiv:0805.2366v2*, May 2008.

D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, and et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, 142:72, sep 2011. doi: 10.1088/0004-6256/142/3/72.

K. Heitmann, M. White, C. Wagner, S. Habib, and D. Higdon. The coyote universe. i. precision determination of the nonlinear matter power spectrum. *ApJ*, 715:104–121, May 2010. doi: 10.1088/0004-637X/715/1/104.

S. Agarwal and H. A. Feldman. The effect of massive neutrinos on the matter power spectrum. *MNRAS*, 410:1647–1654, January 2011.

R. Scoccimarro, M. Zaldarriaga, and L. Hui. Power Spectrum Correlations Induced by Nonlinear Clustering. *ApJ*, 527:1–15, December 1999. doi: 10.1086/308059.

S. Saito, M. Takada, and A. Taruya. Impact of massive neutrinos on the nonlinear matter power spectrum. *Physical Review Letters*, 100(19):191301, May 2008. doi: 10.1103/PhysRevLett.100.191301.

B. W. O'Shea, G. Bryan, J. Bordner, M. L. Norman, T. Abel, R. Harkness, and A. Kritsuk. Enzo: AMR Cosmology Application. *Astrophysics Source Code Library*, page 10072, October 2010.

V. Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364:1105–1134, December 2005. doi: 10.1111/j.1365-2966.2005.09655.x.

R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, 341:1311–1332, June 2003. doi: 10.1046/j.1365-8711.2003.06503.x.

S. Bird, M. Viel, and M. G. Haehnelt. Massive neutrinos and the non-linear matter power spectrum. *MNRAS*, page 2175, January 2012. doi: 10.1111/j.1365-2966.2011.20222.x.

J. Brandbyge and S. Hannestad. Grid based linear neutrino perturbations in cosmological n-body simulations. *JCAP*, 5:2, May 2009a. doi: 10.1088/1475-7516/2009/05/002.

M. Viel, M. G. Haehnelt, and V. Springel. The effect of neutrinos on the matter distribution as probed by the intergalactic medium. *Journal of Cosmology and Astro-Particle Physics*, 6:15–+, June 2010. doi: 10.1088/1475-7516/2010/06/015.

T. Nishimichi, A. Shirata, A. Taruya, K. Yahata, S. Saito, Y. Suto, R. Takahashi, N. Yoshida, T. Matsubara, N. Sugiyama, I. Kayo, Y. Jing, and K. Yoshikawa. Modeling Nonlinear Evolution of Baryon Acoustic Oscillations: Convergence Regime of N-body Simulations and Analytic Models. *PASJ*, 61:321–, February 2009.

S. Saito, M. Takada, and A. Taruya. Nonlinear power spectrum in the presence of massive neutrinos: Perturbation theory approach, galaxy bias, and parameter forecasts. *PRD*, 80(8):083528, October 2009a. doi: 10.1103/PhysRevD.80.083528.

J. Carlson, M. White, and N. Padmanabhan. Critical look at cosmological perturbation theory techniques. *PRD*, 80(4):043531, August 2009. doi: 10.1103/PhysRevD.80.043531.

D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe

(WMAP) Observations: Implications for Cosmology. *ApJS*, 170:377–408, June 2007. doi: 10.1086/513700.

E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. *ApJS*, 180:330–376, February 2009. doi: 10.1088/0067-0049/180/2/330.

A. Taruya, T. Nishimichi, S. Saito, and T. Hiramatsu. Nonlinear evolution of baryon acoustic oscillations from improved perturbation theory in real and redshift spaces. *PRD*, 80(12):123503–+, December 2009.

K. Heitmann, D. Higdon, M. White, S. Habib, B. J. Williams, E. Lawrence, and C. Wagner. The coyote universe. ii. cosmological models and precision emulation of the nonlinear matter power spectrum. *ApJ*, 705:156–174, November 2009. doi: 10.1088/0004-637X/705/1/156.

E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological interpretation. *ApJS*, 192:18, February 2011.

KamLAND. Precision measurement of neutrino oscillation parameters with kamland. *Physical Review Letters*, 100:221803, 2008. URL `doi:10.1103/PhysRevLett.100.221803`.

SNO. Measurement of the total active 8b solar neutrino flux at the sudbury neutrino observatory with enhanced neutral current sensitivity. *Physical Review Letters*, 92: 181301, 2004. URL `doi:10.1103/PhysRevLett.92.181301`.

P. Adamson, C. Andreopoulos, K. E. Arms, R. Armstrong, D. J. Auty, D. S. Ayres, B. Baller, P. D. Barnes, G. Barr, W. L. Barrett, B. R. Becker, A. Belias, R. H. Bernstein, D. Bhattacharya, M. Bishai, A. Blake, G. J. Bock, J. Boehm, D. J. Boehnlein, D. Bogert, C. Bower, E. Buckley-Geer, S. Cavanaugh, J. D. Chapman, D. Cherdack, S. Childress, and B. C. Choudhary. Measurement of neutrino oscillations with the minos detectors in the numi beam. *Phys. Rev. Lett.*, 101(13):131802, Sep 2008. doi: 10.1103/PhysRevLett.101.131802.

E. W. Otten and C. Weinheimer. Neutrino mass limit from tritium beta decay. *Reports on Progress in Physics*, 71:086201, 2008. URL `http://www.citebase.org/abstract?id=oai:arXiv.org:0909.2104`.

E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *ArXiv e-prints*, January 2010.

J. Brandbyge, S. Hannestad, T. Haugbølle, and B. Thomsen. The effect of thermal neutrino motion on the non-linear cosmological matter power spectrum. *Journal of Cosmology and Astro-Particle Physics*, 8:20–+, August 2008. doi: 10.1088/1475-7516/2008/08/020.

J. Brandbyge and S. Hannestad. Grid based linear neutrino perturbations in cosmological N-body simulations. *Journal of Cosmology and Astro-Particle Physics*, 5:2–+, May 2009b. doi: 10.1088/1475-7516/2009/05/002.

J. Brandbyge and S. Hannestad. Resolving cosmic neutrino structure: a hybrid neutrino N-body scheme. *Journal of Cosmology and Astro-Particle Physics*, 1:21–+, January 2010. doi: 10.1088/1475-7516/2010/01/021.

Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89, March 1970.

R. Scoccimarro. Transients from initial conditions: a perturbative analysis. *MNRAS*, 299:1097–1118, October 1998. doi: 10.1046/j.1365-8711.1998.01845.x.

A. Lewis, A. Challinor, and A. Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *ApJ*, 538:473–476, August 2000. doi: 10.1086/309179.

B. W. O'Shea, G. Bryan, J. Bordner, M. L. Norman, T. Abel, R. Harkness, and A. Kritsuk. Introducing Enzo, an AMR Cosmology Application. *ArXiv Astrophysics e-prints*, March 2004.

M. L. Norman, G. L. Bryan, R. Harkness, J. Bordner, D. Reynolds, B. O'Shea, and R. Wagner. Simulating Cosmological Evolution with Enzo. *arXiv:0705.1556*, May 2007.

S. Saito, M. Takada, and A. Taruya. Nonlinear power spectrum in the presence of massive neutrinos: Perturbation theory approach, galaxy bias, and parameter forecasts. *PRD*, 80(8):083528–+, October 2009b. doi: 10.1103/PhysRevD.80.083528.

D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *ApJS*, 192:16, February 2011. doi: 10.1088/0067-0049/192/2/16.

B. W. O'Shea, K. Nagamine, V. Springel, L. Hernquist, and M. L. Norman. Comparing AMR and SPH Cosmological Simulations. I. Dark Matter and Adiabatic Simulations. *ApJS*, 160:1–27, September 2005. doi: 10.1086/432645.

W. Hu, D. J. Eisenstein, and M. Tegmark. Weighing Neutrinos with Galaxy Surveys. *Physical Review Letters*, 80:5255–5258, June 1998. doi: 10.1103/PhysRevLett.80.5255.

R. Jenatton, A. Gramfort, V. Michel, G. Obozinski, E. Eger, F. Bach, and B. Thirion. Multi-scale Mining of fMRI data with Hierarchical Structured Sparsity. *arXiv:1105.0363*, May 2011.

F. Pedregosa, A. Gramfort, G. Varoquaux, B. Thirion, C. Pallier, and E. Cauvet. Improved brain pattern recovery through ranking approaches. *arXiv:1207.3520*, July 2012.

A. Ghosh. Comparative study of Financial Time Series Prediction by Artificial Neural Network with Gradient Descent Learning. *arXiv:1111.4930*, November 2011.

E. Hurwitz and T. Marwala. Common Mistakes when Applying Computational Intelligence and Machine Learning to Stock Market modelling. *arXiv:1208.4429*, August 2012.

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

T. Auld, M. Bridges, M. P. Hobson, and S. F. Gull. Fast cosmological parameter estimation using neural networks. *MNRAS*, 376:L11–L15, March 2007. doi: 10.1111/j.1745-3933.2006.00276.x.

W. A. Fendt and B. D. Wandelt. Computing high accuracy power spectra with pico. *ApJ*, 654:2, December 2007.

T. Auld, M. Bridges, and M. P. Hobson. Cosmonet: fast cosmological parameter estimation in non-flat models using neural networks. *MNRAS*, 387:1575–1582, July 2008. doi: 10.1111/j.1365-2966.2008.13279.x.

A. A. Collister and O. Lahav. Annz: Estimating photometric redshifts using artificial neural networks. *PASP*, 116:345–351, April 2004.

S. Habib, K. Heitmann, D. Higdon, C. Nakhleh, and B. Williams. Cosmic calibration: Constraints from the matter power spectrum and the cosmic microwave background. *PRD*, 76(8):083503–+, October 2007. doi: 10.1103/PhysRevD.76.083503.

M. D. Schneider, L. Knox, S. Habib, K. Heitmann, D. Higdon, and C. Nakhleh. Simulations and cosmological inference: A statistical model for power spectra means and covariances. *PRD*, 78(6):063529–+, September 2008. doi: 10.1103/PhysRevD.78.063529.

M. D. Schneider, Ó. Holm, and L. Knox. Intelligent design: On the emulation of cosmological simulations. *ApJ*, 728:137–+, February 2011. doi: 10.1088/0004-637X/728/2/137.

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

Carl E. Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. URL `http://www.gaussianprocess.org/gpml/`.

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

C. L. Sarazin and R. E. White, III. Steady state cooling flow models for normal elliptical galaxies. *ApJ*, 320:32–48, September 1987. doi: 10.1086/165522.

M. P. van Daalen, J. Schaye, C. M. Booth, and C. Dalla Vecchia. The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology. *MNRAS*, 415:3649–3665, August 2011. doi: 10.1111/j.1365-2966.2011.18981.x.

Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257, 1991. ISSN 0893-6080. doi: 10.1016/0893-6080(91)90009-T. URL `http://www.sciencedirect.com/science/article/pii/089360809190009T`.

Yoshifusa Ito. Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. *Neural Networks*, 4(3):385 – 394, 1991. ISSN 0893-6080. doi: 10.1016/0893-6080(91)90075-G. URL `http://www.sciencedirect.com/science/article/pii/089360809190075G`.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, May 1979. doi: doi:10.2307/1268522.

J. P. Ostriker and Y. Suto. The Mach number of the cosmic flow - A critical test for current theories. *ApJ*, 348:378–382, January 1990. doi: 10.1086/168247.

Y. Suto, R. Cen, and J. P. Ostriker. Statistics of the cosmic Mach number from numerical simulations of a cold dark matter universe. *ApJ*, 395:1–20, August 1992. doi: 10.1086/171626.

M. A. Strauss, R. Cen, and J. P. Ostriker. The cosmic Mach number - Direct comparisons of observations and models. *ApJ*, 408:389–402, May 1993. doi: 10.1086/172596.

Y.-Z. Ma, J. P. Ostriker, and G.-B. Zhao. Cosmic Mach Number: a sensitive probe for the growth of structure. *JCAP*, 6:026, June 2012. doi: 10.1088/1475-7516/2012/06/026.

C. McBride, A. Berlind, R. Scoccimarro, R. Wechsler, M. Busha, J. Gardner, and F. van den Bosch. LasDamas Mock Galaxy Catalogs for SDSS. In *BAAS*, volume 41, page 425.06, January 2009.

J. P. Gardner, A. Connolly, and C. McBride. A Framework for Analyzing Massive Astrophysical Datasets on a Distributed Grid. In R. A. Shaw, F. Hill, and D. J. Bell, editors, *Astronomical Data Analysis Software and Systems XVI*, volume 376 of *Astronomical Society of the Pacific Conference Series*, page 69, October 2007.

M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, May 1985. doi: 10.1086/163168.

J. L. Tonry, B. P. Schmidt, B. Barris, P. Candia, P. Challis, A. Clocchiatti, A. L. Coil, A. V. Filippenko, P. Garnavich, C. Hogan, S. T. Holland, S. Jha, R. P. Kirshner, K. Krisciunas, B. Leibundgut, W. Li, T. Matheson, M. M. Phillips, A. G. Riess, R. Schommer, R. C. Smith, J. Sollerman, J. Spyromilio, C. W. Stubbs, and N. B. Suntzeff. Cosmological Results from High-z Supernovae. *ApJ*, 594:1–24, September 2003.

R. Giovanelli, M. P. Haynes, J. J. Salzer, G. Wegner, L. N. da Costa, and W. Freudling. The Motions of Clusters of Galaxies and the Dipoles of the Peculiar Velocity Field. *AJ*, 116:2632–2643, December 1998. doi: 10.1086/300652.

D. A. Dale, R. Giovanelli, M. P. Haynes, L. E. Campusano, and E. Hardy. Seeking the Local Convergence Depth. V. Tully-Fisher Peculiar Velocities for 52 Abell Clusters. *AJ*, 118:1489–1505, October 1999a.

M. J. Hudson, R. J. Smith, J. R. Lucey, D. J. Schlegel, and R. L. Davies. A Large-scale Bulk Flow of Galaxy Clusters. *ApJL*, 512:L79–L82, February 1999.

M. J. Hudson, R. J. Smith, J. R. Lucey, and E. Branchini. Streaming motions of galaxy clusters within 12 000 km s^{-1} – V. The peculiar velocity field. *MNRAS*, 352:61–75, July 2004.

M. Colless, R. P. Saglia, D. Burstein, R. L. Davies, R. K. McMahan, and G. Wegner. The peculiar motions of early-type galaxies in two distant regions - VII. Peculiar velocities and bulk motions. *MNRAS*, 321:277–305, February 2001.

J. A. Willick. The Las Campanas Observatory/Palomar 10,000 Kilometer per Second Cluster Survey. II. Constraints on Large-Scale Streaming. *ApJ*, 522:647–660, September 1999.

N. Kaiser. Theoretical implications of deviations from Hubble flow. *MNRAS*, 231: 149–167, March 1988.

H. A. Feldman, R. Watkins, and M. J. Hudson. Cosmic flows on 100 h^{-1} Mpc scales: standardized minimum variance bulk flow, shear and octupole moments. *MNRAS*, 407:2328–2338, October 2010. doi: 10.1111/j.1365-2966.2010.17052.x.

R. Watkins, H. A. Feldman, and M. J. Hudson. Consistently large cosmic flows on scales of $100h^{-1}$Mpc: a challenge for the standard ΛCDM cosmology. *MNRAS*, 392: 743–756, January 2009. doi: 10.1111/j.1365-2966.2008.14089.x.

J. L. Tonry, A. Dressler, J. P. Blakeslee, E. A. Ajhar, A. B. Fletcher, G. A. Luppino, M. R. Metzger, and C. B. Moore. The SBF Survey of Galaxy Distances. IV. SBF Magnitudes, Colors, and Distances. *ApJ*, 546:681–693, January 2001. doi: 10.1086/318301.

L. N. da Costa, M. Bernardi, M. V. Alonso, G. Wegner, C. N. A. Willmer, P. S. Pellegrini, C. Rité, and M. A. G. Maia. Redshift-Distance Survey of Early-Type Galaxies. I. Sample Selection, Properties, and Completeness. *AJ*, 120:95–109, July 2000.

M. Bernardi, M. V. Alonso, L. N. da Costa, C. N. A. Willmer, G. Wegner, P. S. Pellegrini, C. Rité, and M. A. G. Maia. Redshift-Distance Survey of Early-Type Galaxies. I. The ENEARc Cluster Sample. *AJ*, 123:2990–3017, June 2002.

G. Wegner, M. Bernardi, C. N. A. Willmer, L. N. da Costa, M. V. Alonso, P. S. Pellegrini, M. A. G. Maia, O. L. Chaves, and C. Rité. Redshift-Distance Survey of Early-Type Galaxies: Spectroscopic Data. *AJ*, 126:2268–2280, November 2003. doi: 10.1086/378959.

K. L. Masters, C. M. Springob, M. P. Haynes, and R. Giovanelli. SFI++ I: A New I-Band Tully-Fisher Template, the Cluster Peculiar Velocity Dispersion, and H_0. *ApJ*, 653:861–880, December 2006. doi: 10.1086/508924.

C. M. Springob, K. L. Masters, M. P. Haynes, R. Giovanelli, and C. Marinoni. SFI++. II. A New I-Band Tully-Fisher Catalog, Derivation of Peculiar Velocities, and Data Set Properties. *ApJS*, 172:599–614, October 2007. doi: 10.1086/519527.

C. M. Springob, K. L. Masters, M. P. Haynes, R. Giovanelli, and C. Marinoni. Erratum: "SFI++ II: A New I-Band Tully-Fisher Catalog, Derivation of Peculiar Velocities and Data Set Properties" (2007, ApJS, 172, 599). *ApJS*, 182:474–475, May 2009. doi: 10.1088/0067-0049/182/1/474.

B. Grinstein, H. D. Politzer, S.-J. Rey, and M. B. Wise. Interpretation of large-scale deviations from the Hubble flow. *ApJ*, 314:431–438, March 1987. doi: 10.1086/165076.

W. Hu and N. Sugiyama. Small-scale cosmological perturbations: an analytic approach. *ApJ*, 471:542, November 1996.