Cosmology with Large-scale Structure and Galaxy Flows

by

Morag Iona Scrimgeour

THE UNIVERSITY OF WESTERN AUSTRALIA

Presented for the degree of
Doctor of Philosophy
at the University of Western Australia

International Centre for Radio Astronomy Research,
School of Physics

November 2013
To my parents, Euan and Sylvia Scrimgeour
Declaration

This thesis comprises only my original work, except where indicated in the Preface, and contains no material that has been presented for a degree at this, or any other, university. Due acknowledgement has been made in the text to all other material used. Any contribution made to the research by others, with whom I have worked at UWA or elsewhere, is explicitly acknowledged in the thesis.

Morag I. Scrimgeour
Acknowledgements

There are many people who helped me throughout my PhD whom I would like to sincerely thank. First and foremost I would like to thank my PhD supervisors, Tamara Davis and Lister Staveley-Smith, who have both been incredibly helpful and supportive throughout my PhD. Tamara, thank you for our many Skype meetings, your inspiring enthusiasm for cosmology, and for many great visits to Brisbane! Lister, thank you for giving me the opportunity to do my PhD at UWA, and for all of your much-appreciated help and guidance throughout the last few years.

Thanks also to Chris Blake, whom I have also been fortunate to work with a great deal throughout my PhD. Your advice and support have been invaluable, and thank you also for some great visits to Melbourne. Also, thanks to Peter Quinn for inspiring me to join ICRAR, and for your support throughout my PhD.

A big thanks goes to everyone at the International Centre for Radio Astronomy Research (ICRAR), for making it such a great place to work over the past four years. Thanks to Jacinta Delhaize, Laura Hoppmann, Giovanna Zanardo, Stefan Westerlund, Toby Potter, Lee Kelvin, Mehmet Alpaslan, Shaun Hooper, Jonathan Diaz, Aaron Robotham and all the other students, staff and postdocs I have known during my time here. Particular thanks goes to Alan Duffy for being a great supervisor during my iVEC/ICRAR summer internship, and for all of your generous IDL and cosmology assistance. Thanks also to Renu Sharma, Leanne Goodsell, Kathy Kok and all the administration staff at ICRAR for assistance throughout my study. I would also like to thank everyone at the School of Physics for their help and support, particularly Ian McArthur, Paul Abbott and Jay Jay Jegathesan.

It was a great experience to be a part of the WiggleZ Cosmology team during my PhD, and I would like to thank Felipe Marin, Signe Riemer-Sørensen, Greg Poole, David Parkinson, Berian James, and Carlos Contreras for many helpful discussions over the last few years. It has also been great working with the 6dFGS team this past year, and I would like to thank Matthew Colless, Christopher Springob, Christina Magoulas and Heath Jones for contributing their time and knowledge to my research.

Thanks also to Brian Schmidt for taking time out of your busy schedule to work with me, and for your confidence in me in nominating me for the Malcolm McIntosh bankmecu scholarship. I would also like to thank Nicolas Regnault, Hume Feldman, and Richard Watkins for helpful e-mail correspondence from afar.

I am grateful for the financial support of a Jean Rogerson Postgraduate Scholarship from the University of Western Australia, and for further support from ICRAR. I would
also like to thank the University of Western Australia for providing travel funding to attend conferences and workshops. Thanks also to the ARC Centre of Excellence for All-sky Astrophysics (CAASTRO) for travel and financial support, and the Astronomical Society of Australia (ASA) for travel support. Furthermore I would like to thank bankmecu for a Malcolm McIntosh bankmecu Student Scholarship.

Most of all, I would like to thank my parents Euan and Sylvia Scrimgeour for your constant support, encouragement and assistance throughout my study, and for always believing in me.

And my biggest thanks go to my boyfriend Florian Beutler, for the never-ending encouragement, advice and inspiration you have given me throughout my PhD, and for making it an awesome four years.
Abstract

Understanding the large-scale structure of the Universe is a crucial test of the standard model of cosmology, Lambda Cold Dark Matter (ΛCDM). This thesis presents two different tests of the properties of large-scale structure, to test assumptions and predictions of ΛCDM.

The first is a test of the large-scale homogeneity of the Universe, a key assumption of ΛCDM that allows the Friedmann-Robertson-Walker (FRW) metric to be used as the description of space-time. Many aspects of ΛCDM directly rely on the assumption that the large-scale Universe is homogeneous and isotropic, including the postulation of dark energy as the cause of the observed cosmic acceleration, and the use of statistical measurements such as the power spectrum to constrain properties of the Universe. Although the isotropy of the cosmic microwave background (CMB) at \( z \approx 1100 \) provides good evidence for homogeneity, it is crucial the assumption be robustly tested at low-redshift using a 3D measurement.

To this end, we have made the largest-volume measurement to date of the transition to large-scale homogeneity in the galaxy distribution, using the WiggleZ Dark Energy Survey. WiggleZ is a spectroscopic redshift survey of \( \sim 200,000 \) blue galaxies in a cosmic volume of \( \sim 1 \, h^{-3} \) Gpc\(^3\), over 1000 deg\(^2\) of sky, and in a redshift range of \( 0.1 < z < 1 \). We measure the average amount of clustering as a function of scale, by measuring both the mean galaxy counts in spheres of radius \( r \), \( N(<r) \), and the fractal correlation dimension \( D_2(r) \) of the WiggleZ galaxies. With the large depth of WiggleZ, we make the first investigation of how the homogeneity scale evolves with cosmic epoch. The fractal dimension of the sample is within 1 per cent of homogeneity, or of \( D_2 = 3 \), for radii larger than \( 71 \pm 8 \, h^{-1} \) Mpc at \( z \approx 0.2 \), \( 70 \pm 5 \, h^{-1} \) Mpc at \( z \approx 0.4 \), \( 81 \pm 5 \, h^{-1} \) Mpc at \( z \approx 0.6 \), and \( 75 \pm 4 \, h^{-1} \) Mpc at \( z \approx 0.8 \). We also demonstrate the robustness of our results to selection function and edge effects, using a suite of fractal mock catalogues, and the GiggleZ \( N \)-body simulation. Our results are a strong confirmation of the ΛCDM predicted scale of homogeneity, and a strong consistency check of the FRW metric.

The next part of this thesis focuses on a test of the large-scale motions of galaxies, the ‘bulk flow’ in the local Universe, which is a probe of large-scale matter fluctuations. Several recent measurements have hinted at an anomalously high local bulk flow, which could indicate either a higher value of rms density fluctuations on 8 \( h^{-1} \) Mpc scales, \( \sigma_8 \), than from CMB constraints, an exotic scenario such as a ‘tilted universe,’ or that we exist in a very unlikely part of the Universe with a large local overdensity.
We make a new, improved measurement to test this, using the 6dF Galaxy Survey peculiar velocity sample (6dFGSv). This sample contains 8885 Fundamental Plane distances to a sample of early-type galaxies covering the whole Southern sky, with Dec < 0° and galactic latitude |b| > 10°, out to redshift $z = 0.054$. This sample provides the largest, most homogeneously-selected peculiar velocity sample to date. We measure the bulk flow of the sample using two different estimators: a simple maximum-likelihood estimate, and a more sophisticated ‘Minimum Variance’ estimator introduced by Watkins et al. (2009). We find a bulk flow of $259 \pm 54 \, \text{km} \, \text{s}^{-1}$ at a radius of $50 \, h^{-1} \, \text{Mpc}$ in the direction $(l = 317 \pm 16^\circ, b = 35 \pm 11^\circ)$, and $254 \pm 54 \, \text{km} \, \text{s}^{-1}$ at a radius of $70 \, h^{-1} \, \text{Mpc}$ in the direction $(l = 317 \pm 16^\circ, b = 35 \pm 11^\circ)$. The direction of our measurement is consistent with other measurements in the literature, and is also close to the direction of the Shapley Supercluster, suggesting this as a possible origin of at least part of the bulk flow. The magnitude of the bulk flow is consistent with several recent measurements that also find a low-amplitude bulk flow, and is consistent with a ΛCDM prediction within 95.5% confidence.

The final chapter of this thesis presents a prediction for how well the SkyMapper Transient and Supernova Survey (SMT) will measure the local bulk flow. SMT, currently running on the SkyMapper telescope, will observe $\sim 500$ Type Ia Supernovae (SNe Ia) over 5 years in the Southern sky. In its current configuration, SMT will measure the bulk flow at a scale of $150 \, h^{-1} \, \text{Mpc}$ with an accuracy of $\sim 90 \, \text{km} \, \text{s}^{-1}$. We also investigate how a general supernova survey can be optimised to measure the bulk flow, and find that an all-sky survey, with the same number of fields and supernovae of SMT, would have a 26% improvement in bulk flow accuracy compared to the SMT distribution. SkyMapper should provide an important new constraint on the bulk flow, and will be even more powerful if combined with a northern-sky supernova survey.
Preface

This thesis is composed of a series of published and in preparation papers, in compliance with the rules for PhD thesis submission from the Graduate Research School at the University of Western Australia. I am the primary author on these papers, and I confirm that the work in this thesis is my own. I have detailed the contribution of each of the co-authors to the papers, below. I confirm I have permission from the co-authors to reproduce the papers here.

Publications arisen from this thesis are:

1. **The WiggleZ Dark Energy Survey: the transition to large-scale cosmic homogeneity** (Chapter 2)
   arXiv:1205.6812
   • I am the primary author on this paper, and performed all of the data analysis and writing. The 5 lead co-authors supervised the data analysis, contributed scientific expertise, and helped edit the text. This was a WiggleZ core cosmology paper, and so the WiggleZ Dark Energy Survey core team are also co-authors, and appear alphabetically after the 6 lead authors. The core team was responsible for the survey observations and collection of the galaxy redshift data used for this analysis.

2. **The 6dF Galaxy Survey: the minimum variance bulk flow on 50–70 h$^{-1}$ Mpc scales** (Chapter 3)
   In preparation
   • I am the primary author on this paper, and performed all of the data analysis and writing. The peculiar velocity data used in this paper was collected and derived
by the 6dFGS team, who are co-authors, and their contribution is explicitly
acknowledged in the paper. The co-authors also provided supervision of the
data analysis, scientific guidance, and helped with editing of the text.

3. **Bulk Flow estimation for the SkyMapper Supernova and Transient Survey and general survey optimisation** (Chapter 4)
   M. I. Scrimgeour, T. Davis, B. P. Schmidt, N. Regnault, L. Staveley-Smith
   In preparation
   - I am the primary author on this paper, and performed all of the simulations,
simulated data analysis, and writing. My simulations incorporate observation
logs from a previous, unpublished simulation of SkyMapper by Nicolas Regnault,
which is referenced as Scalzo et al. (in preparation). The co-authors provided
supervision of the simulation design and analysis, and helped with editing of the
text.
## Contents

1 Introduction .................................................. 1

1.1 The Standard Model of Cosmology ......................... 3

1.1.1 General Relativity & Spacetime Geometry .................. 3

1.1.2 The Cosmological Principle and the Friedmann-Robertson-Walker metric .................. 4

1.2 Structure Formation ........................................... 11

1.2.1 Gravitational Instability and the Growth Rate .............. 11

1.2.2 The Correlation Function and Power Spectrum ............... 13

1.2.3 Peculiar Velocity Statistics ................................ 15

1.2.4 Inflation and the Primordial Power Spectrum ............... 18

1.3 Observational Cosmology .................................... 20

1.3.1 Cosmic Microwave Background ......................... 20

1.3.2 Type Ia Supernovae ........................................ 21

1.3.3 Peculiar Velocities ........................................ 24

1.4 Thesis Motivation and Overview ......................... 28

1.4.1 Testing large-scale homogeneity in the WiggleZ Dark Energy Survey .................. 28

1.4.2 Bulk Flow tests with SkyMapper and the 6 degree Field Galaxy Survey peculiar velocities .................. 30

2 Large-scale cosmic homogeneity in the WiggleZ Dark Energy Survey ............................................... 32

2.1 Abstract ................................................... 32

2.2 Introduction .................................................. 33
3.6 Defining $v$ from $P(v)$ ........................................... 93
  3.6.1 Statistical definitions of $v$ .................................. 95
  3.6.2 Obtaining an unbiased $v$ ...................................... 95
3.7 Velocity uncertainties and bulk flow weights ......................... 99
  3.7.1 The standard deviation of $P(v)$: correlation with $v$ ........ 99
  3.7.2 Bulk flow bias from $\sigma_{SDV}$ ............................. 99
  3.7.3 A solution: distance-dependent $\sigma_n$ ...................... 101
3.8 6dFGSv $\Lambda$CDM Mock Catalogues ................................ 104
  3.8.1 6dFGSv survey selection function .......................... 106
  3.8.2 Fundamental Plane fitting .................................... 107
  3.8.3 Mock sample algorithm ....................................... 107
  3.8.4 The GiggleZ Simulation ...................................... 110
  3.8.5 Determining 6dFGSv halo bias & mass ......................... 110
  3.8.6 Selecting the largest haloes .................................. 112
  3.8.7 The Mocks ................................................... 113
3.9 Results and Discussion ............................................ 115
  3.9.1 Bulk flow results ........................................... 117
  3.9.2 6dFGSv bulk flow distribution in $\Lambda$CDM mocks .......... 121
  3.9.3 Comparison with linear theory: 3D bulk flow ................ 124
  3.9.4 Comparison with linear theory: 1D bulk flow ................. 128
  3.9.5 Bulk flow in redshift shells .................................. 132
  3.9.6 Comparison with other results ................................ 134
  3.9.7 Comparison with method of Magoulas et al., (2014) .......... 134
  3.9.8 Implications for Cosmography ................................ 136
  3.9.9 Cosmological implications .................................... 137
3.10 Conclusion ...................................................... 138

4 Bulk Flow estimation for the SkyMapper Supernova and Transient Survey (SMT) and general survey optimisation 140
  4.1 Abstract ....................................................... 140
  4.2 Introduction .................................................... 141
  4.3 SkyMapper Supernova Survey Fields .............................. 144
  4.4 Survey Simulation ............................................. 145
1

Introduction

The standard model of cosmology, Lambda Cold Dark Matter (ΛCDM) is now very well established, and is supported to high precision by a wide range of cosmological observations. It has its origins in Einstein’s General Theory of Relativity (Einstein, 1916) which provided the first full description of gravity, connecting mass to energy and the curvature of spacetime. With Hubble’s discovery of the expansion of the Universe (Hubble, 1929) the Big Bang model of the Universe was established. Over the following decades, several more important observational discoveries were made, including the discovery of dark matter (Zwicky, 1933, 1937; Freeman, 1970; Rubin & Ford, 1970), the cosmic microwave background (CMB) radiation (Penzias & Wilson, 1965), the CMB anisotropies (Smoot et al., 1992), and the discovery of cosmic acceleration implying the existence of dark energy (Riess et al., 1998; Perlmutter et al., 1999). During this time there was a corresponding theoretical drive, leading to theoretical developments including Big Bang nucleosynthesis as the origin of heavy elements (Alpher & Herman, 1948), and inflation (Guth, 1981) as the cause of both the large-scale homogeneity and flatness of the Universe, and the primordial density perturbations. All this led to the formulation of the ΛCDM model, postulating a flat (zero-curvature) universe in which the energy density is dominated by dark energy in the form of Einstein’s cosmological constant Λ (∼ 68% of the Universe) and dark matter (∼ 28% of the Universe), with the fields of the standard model of particle physics comprising the remaining ∼ 4%. This model is now supported by a wide variety of observations, including the CMB, Type Ia supernovae, large-scale
galaxy clustering and gravitational lensing. Indeed, the years since the Wilkinson Microwave Anisotropy Probe (WMAP, Spergel et al., 2003), which made the first accurate measurement of the CMB anisotropies, are often claimed to be the ‘precision age of cosmology,’ implying the model is now accurate and all that remains is to make tighter constraints on its parameters.

However, despite the success of ΛCDM, our understanding of the Universe is far from complete. ΛCDM tells us about the apparent composition of the Universe, but tells us nothing about the nature of dark energy and dark matter, the two components believed to dominate the Universe’s evolution today. Explaining their nature has become one of the great challenges of modern physics.

There are many planned and ongoing cosmological surveys that are seeking to uncover the nature of dark energy and dark matter, or to find inconsistencies in the model, looking for signs of new physics and testing the predictions of ΛCDM. In this thesis we investigate two ways of testing the ΛCDM model, which could potentially give hints of departures from ΛCDM. The first is a consistency test of the large-scale homogeneity of the Universe, a key assumption of ΛCDM, using the WiggleZ Dark Energy Survey (Scrimgeour et al., 2012). A departure from homogeneity could potentially invalidate the use of the Friedmann-Robertson-Walker metric (see Section 1.1.2) and even explain the observed accelerating cosmic expansion (Wiltshire, 2007b; Räsanen, 2011).

The second test in this thesis is a measurement of the large-scale motions of galaxies in the nearby Universe, the so-called ‘bulk flow,’ which depends on the large scale modes of the matter power spectrum, a key prediction of ΛCDM. A number of observations have hinted at the presence of an unusually large bulk flow, which, if true, could be a sign that density perturbations are much larger than predicted by ΛCDM (Watkins et al., 2009), or an even more exotic scenario such as a ‘tilted universe’ (Turner, 1991; Kashlinsky et al., 2008; Ma et al., 2011).

In this chapter we first provide some theoretical background to modern cosmology, in particular General Relativity and the FRW metric, in Section 1.1. In Section 1.2 we describe the theory of structure formation, which is an important foundation for the tests in this thesis. In Section 1.3 we describe some of the key observational probes of cosmology that are of relevance for our work, and we present an overview of the two tests in this thesis in Section 1.4.
For further background, we refer the reader to Strauss & Willick (1995), Peacock (1999), Dodelson (2003), Frieman et al. (2008), Lyth & Liddle (2009), and Ruiz-Lapuente (2010).

1. The Standard Model of Cosmology

1.1 General Relativity & Spacetime Geometry

Modern cosmology has its beginnings at the start of the 20th century, with Einstein’s formulation of General Relativity (Einstein, 1916). General Relativity (GR) provides a description of gravity in terms of the geometric properties of space and time, or ‘space-time.’ It has been tested to extremely high precision, both in the weak field limit on the Earth (Pound & Snider, 1964) and solar system scales (Reynaud & Jaekel, 2009), and for stronger fields in binary pulsars within the galaxy (Hulse & Taylor, 1975).

In general relativity, an event in space-time is labelled by the four coordinates \( x^\mu \) of time and space, where conventionally \( \mu = (ct, x, y, z) \). Two events with separation \( dx^\mu \) are separated by the line element \( ds \), which is invariant under coordinate transformation:

\[
ds^2 = g_{\mu\nu}dx^\mu dx^\nu,
\]

where we use the Einstein convention of summing repeated indices, \( c \) is the speed of light, and \( g_{\mu\nu} \) is the metric tensor, a matrix that represents the distance between space-time points. It is a tensor since Equation 1.1 must be valid in all frames.

Einstein’s field equations relate the geometry of space-time to its mass-energy content:

\[
G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}.
\]

The Einstein tensor \( G_{\mu\nu} \) on the left-hand side represents the geometry of space-time. It is a function of the metric tensor, \( g_{\mu\nu} \), and its first two derivatives. The stress-energy tensor \( T_{\mu\nu} \) on the right hand side represents the matter and energy components of the Universe, such as particles, radiation, fields, and zero-point energies. \( G \) is the gravitational constant.

For a perfect fluid in a homogeneous, isotropic universe, the stress-energy tensor
1.1. THE STANDARD MODEL OF COSMOLOGY

would be

\[ T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \] (1.3)

The diagonal terms \( \rho \) and \( p \) define the energy density and pressure; it is symmetrical due to the assumed symmetry of the Universe (Peebles & Ratra, 2003).

It is possible to add an additional constant to the left-hand side of Equation 1.2, the cosmological constant \( \Lambda \):

\[ G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \] (1.4)

On the left side of the equation, it represents the curvature of empty space. However, if it is moved to the right-hand side then it appears like the energy-momentum tensor of the vacuum,

\[ T_{\mu\nu}^{\text{vac}} = -\frac{\Lambda c^4}{8\pi G} g_{\mu\nu}. \] (1.5)

If the vacuum has non-zero energy density and pressure, it turns out that the vacuum will have a negative pressure, causing a gravitational repulsion. Einstein first introduced this term to his equations to allow for a static solution, before it was found observationally that the Universe is expanding. However, since the discovery of cosmic acceleration, \( \Lambda \) has become the leading explanation for dark energy.

1.1.2 The Cosmological Principle and the Friedmann-Robertson-Walker metric

The starting point of the Standard Model of Cosmology is to find a solution to Einstein’s equations. This requires two important simplifying assumptions; firstly, that the stress-energy tensor can be averaged on large scales (Ellis & Buchert, 2005), and secondly, that the Universe has large-scale symmetry. This symmetry is invoked by the Cosmological Principle, which states that on large enough scales, matter in the Universe is homogeneous (i.e. has constant density) and isotropic (is the same in all directions) (Peacock, 1999).

An isotropic and homogeneous matter distribution is the simplest possible dis-
distribution, and allows a simple solution to the Einstein field equations in the form of
the Friedmann-Robertson-Walker (FRW) metric. For this reason, homogeneity and
isotropy was assumed by Einstein, Friedmann and other early theorists for developing
models of the Universe, before there was observational evidence for large-scale
homogeneity. This evidence has since been provided by both the isotropy of the
CMB (Smoot et al., 1991; Bennett et al., 1996) and measurements using galaxy
surveys (Hogg et al., 2005; also Scrimgeour et al., 2012), which forms Chapter 2
of this thesis).

For a homogeneous, isotropic universe the covariant line element between two
points is given by

\[ ds^2 = c^2 dt^2 - a^2(t) dl^2, \]  

(1.6)

where \( a(t) \) is the dimensionless cosmological scale factor and \( t \) is the proper time
measured by an observer in the Hubble flow (i.e. taking part in the Hubble expansion).
The comoving distance, \( l \), between any two points taking part in the Hubble
expansion remains constant with time. The proper distance corresponds to the in-
terval \( s \) along a slice of \( dt = 0 \). The scale factor \( a(t) \) is normalised to 1 at the present
day, and is zero at \( t = 0 \).

Written in spherical coordinates, this equation becomes the standard form of the
Friedman-Robertson-Walker (FRW) metric:

\[ ds^2 = c^2 dt^2 - R^2(t) \left[ d\chi^2 + S_k^2(\chi)(d\theta^2 + \sin^2 \theta d\phi^2) \right]. \]  

(1.7)

Here, \( R \) is the scale factor in dimensions of distance, \( R = R_0 a \), with \( R_0 = c(H_0 \sqrt{|\Omega_k|})^{-1} \)
the present-day value, and \( \chi \) is a dimensionless radial coordinate. The physical
comoving coordinate distance is \( d_C = R_0 \chi \), and \( \theta, \phi \) are the angular comoving
coordinates. \( S_k \) depends on the curvature parameter \( k \), an integration constant
of Einstein’s equations. For a flat (zero-curvature) universe, \( k = 0 \); for an open
(positive-curvature) universe, \( k = +1 \); for a closed (negative-curvature) universe,
We then have

\[ S_k(\chi) = \begin{cases} 
  \sin \chi & \text{if } k = +1, \\
  \chi & \text{if } k = 0, \\
  \sinh \chi & \text{if } k = -1.
\end{cases} \]  

(1.8)

The evolution of the Universe can be reduced to the evolution of the scale factor, \(a\). If the expansion is monotonic then \(a\) is related to cosmological redshift via

\[ 1 + z = \frac{a_0}{a(t)}, \]  

(1.9)

where \(a_0\) is the value of \(a\) at the present day, normalised to 1. The redshift of an object is the fractional shift in its emitted light, due to the expansion of space:

\[ z \equiv \frac{\nu_e}{\nu_o} - 1 = \frac{\lambda_o}{\lambda_e} - 1 \]  

(1.10)

where \(\nu_e\) and \(\lambda_e\) are the emitted frequency and wavelength, and \(\nu_o\) and \(\lambda_o\) are the observed values. A redshift occurs for a light ray travelling along a null geodesic \((ds^2 = 0)\) as the space through which it travels expands.

In an expanding homogeneous and isotropic universe, the physical distance \(l(t)\) between any two comoving points (i.e. points moving purely with the expansion) varies with world time \(t\) as

\[ l(t) \propto a(t). \]  

(1.11)

The expansion rate of the Universe is the rate of change of \(l(t)\),

\[ v = \frac{dl}{dt} = Hl, \]  

(1.12)

where

\[ H(t) = \frac{\dot{a}}{a} \]  

(1.13)

is the Hubble parameter. The present-day value of \(H\) is the Hubble constant, \(H_0\), normally expressed in terms of the dimensionless parameter \(h\), as \(H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}\). For small \(v\), Equation (1.12) is Hubble’s law, which gives the recession velocity due to the expansion at distance \(l\) from the observer.
In general relativity, for a homogeneous and isotropic universe the second derivative of the scale factor satisfies the equation

$$\ddot{a} = -\frac{4}{3} \pi G (\rho + \frac{3p}{c^2}) + \Lambda$$

(1.14)

where $G$ is Newton’s gravitational constant, and $\rho$ and $p$ are the mean mass density and pressure of the combined components of the Universe. (This equation is obtained from the Einstein field equation, Equation 1.2, by putting in the stress-energy tensor for a homogeneous, isotropic universe, Equation 1.3.) The combined pressure and density satisfy the first law of thermodynamics,

$$\dot{\rho} = -3 \frac{\dot{a}}{a} (\rho + p).$$

(1.15)

Taking the integral of Equation 1.14 gives the so-called Friedmann expansion equation,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8}{3} \pi G \sum \rho_i - \frac{kc^2}{a^2} + \frac{\Lambda}{3}$$

(1.16)

where the sum is over all components of the energy density of the Universe.

Each of the density components can be expressed as ratios of the physical density to the critical density, $\rho_{\text{crit}}$, which is defined as the density required for the Universe to have zero spatial curvature, i.e. a flat geometry. From Equation 1.16, the critical density can be calculated by setting $k$ and $\Lambda$ to zero, as

$$\rho_{\text{crit}}(t) = \frac{3H(t)^2}{8\pi G}.$$  

(1.17)

The ratio of the $i^{th}$ density component to the critical density is denoted $\Omega_i$. These density parameters sum to $\Omega$, the ratio of the mean density of the Universe to the critical density. If the Universe is at the critical density then $\Omega = 1$, and so we have

$$\Omega = \Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1.$$  

(1.18)

Here, $\Omega_m$ is the mean mass density of the Universe in non-relativistic matter, mainly baryonic matter and non-baryonic matter in the form of dark matter. $\Omega_r$ is the mean mass density of relativistic matter, mainly in the cosmic microwave background.
1.1. THE STANDARD MODEL OF COSMOLOGY

photons, and also neutrinos while they are relativistic. $\Omega_\Lambda$ is the present-day mass
density of dark energy in the form of $\Lambda$, while $\Omega_k$ is the effective mass density due
to the curvature of space. The dimensionless density parameters are \cite{peebles1993}:

$$\Omega = \frac{8\pi G \rho}{3H^2}, \quad \Omega_m = \frac{8\pi G \rho_m}{3H^2}, \quad \Omega_\Lambda = \frac{\Lambda c^2}{3H^2}, \quad \Omega_k = -\frac{k}{H^2 a^2}. \quad (1.19)$$

Assuming each component can be approximated as a perfect fluid, and so can
be completely characterised by its isotropic pressure $p$ and its energy density $\rho c^2$,
then its evolution can be described in terms of its equation of state,

$$w \equiv \frac{p}{\rho c^2}. \quad (1.20)$$

The equation of state determines the evolution of the energy density of each com-
ponent, as $\rho \propto V^{-(1+w)}$ for constant $w$, where $V$ is the volume containing the
component.

For non-relativistic matter, $w = 0$; for relativistic matter, $w = 1/3$; for curvature,
$w = -1/3$; and for $\Lambda$ (and also inflation), $w = -1$. We can therefore rewrite the
Friedmann equation as

$$H^2 = H_0^2 \left[ \Omega_{\Lambda,0} + \frac{\Omega_{k,0}}{a^2} + \frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{r,0}}{a^4} \right],$$

$$= H_0^2 \sum \Omega_x a^{-3(1+w)}, \quad (1.21)$$

which shows how the the expansion of the Universe over time depends on its curva-
ture and the density of its constituents. The 0 subscripts indicate the present-day
values. The exponents of the scale factor in the above equation tell us how the dif-
erent components evolve with the expansion. The matter density of a box decreases
proportionally to volume, or $L^3$ if $L$ is the length of the box, and hence the cosmic
matter density scales as $1/a^3$. The curvature density scales with area, and hence
$1/a^2$, while radiation scales as $1/a^4$. $\Omega_\Lambda$ does not scale with $a$, since the energy
density of $\Lambda$ remains a constant property of space.

As we can see, the assumption of cosmic homogeneity and isotropy is an ex-
tremely important and fundamental assumption for the standard model of cosmol-
ogy, as the metric that it produces forms the basis of all calculations within the
model. In particular, dark energy, which is now believed to make up the vast majority of the energy density of the Universe, is inferred from observations within this framework, i.e. with the assumption of homogeneity and isotropy. It is therefore crucial to test this assumption. We do this in Scrimgeour et al. (2012), Chapter 2 of this thesis, using the WiggleZ Dark Energy Survey, and make the most accurate test to date of large-scale cosmic homogeneity, providing an important consistency check of the FRW-based $\Lambda$CDM model.

**Cosmological distances**

In an expanding universe, we must be careful in specifying distances, since the distances between comoving points are constantly changing, and an observer looking out in distance is also looking back in time. (See Hogg, 1999, for a summary of distance measures in cosmology).

It is common to express distances in terms of the so-called ‘expansion function’ (Peebles, 1993)

$$E(z) = \frac{H(z)}{H_0} = \sqrt{\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}.$$  \hfill (1.22)

The dimensionless comoving coordinate $\chi$ that we introduced in Equation 1.7 is a useful basis for defining the different distance measures that we will use. It is given by

$$\chi(z) = \frac{c}{R_0} \int_0^z \frac{dz'}{H(z')} = \frac{c}{R_0H_0} \int_0^z \frac{dz'}{E(z')}.$$ \hfill (1.23)

When defining cosmological distances, two things must be considered: the redshift $z$ of the source with comoving coordinate $\chi(z)$, and the time $t$ at which the distance is evaluated, by using the scale factor at that time, $R(t)$.

The *proper distance* to an object at redshift $z$ is the distance along a constant time surface ($dt = 0$) between us and that object, defined as $D(t, z) = R(t)\chi(z)$. Throughout this thesis we will frequently use the *comoving distance*, $d_C = R_0\chi(z)$, the distance between two points in the Hubble flow that remains constant with time. It is equal to the proper distance at the present day. The radial comoving distance to an object that emits light at time $t_e$ (redshift $z$) and is observed today at time $t_o$.
1.1. THE STANDARD MODEL OF COSMOLOGY

is given by

\[ d_C(z) = R_0 \chi(z) = \int_0^{t_e} \frac{c dt}{a} = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \].

The *luminosity distance* \( d_L \) is the distance an object appears to have, due to its luminosity; it is defined by the relation between emitted flux \( f \) and observed luminosity \( L \),

\[ d_L \equiv \sqrt{\frac{L}{4\pi f}}. \]

Since light from distant objects is redshifted, this is not equal to the true comoving distance, and it is also affected by curvature. For a general universe, the luminosity distance to an object at redshift \( z \) is

\[ d_L = R(t) S_k(\chi)(1 + z). \]

The *transverse comoving distance*, \( d_M(z) \), relates the comoving distance between two objects at the same redshift but separated by an angle \( \delta \theta \), via \( d_M \delta \theta \). The transverse comoving distance is given by

\[
  d_M = \begin{cases}
  \frac{c}{H_0} \frac{1}{\sqrt{|\Omega_k|}} \sinh[\sqrt{|\Omega_k|} \frac{H_0}{c} d_C] & \text{if } \Omega_k > 0, \\
  d_C & \text{if } \Omega_k = 0, \\
  \frac{c}{H_0} \frac{1}{\sqrt{|\Omega_k|}} \sin[\sqrt{|\Omega_k|} \frac{H_0}{c} d_C] & \text{if } \Omega_k < 0.
  \end{cases}
\]

The *angular diameter distance* \( d_A \) is the ratio of an object’s physical transverse size \( r \), to its angular size \( \theta \) in radians. It is the distance measured by standard rulers, such as baryon acoustic oscillations, or galaxy sizes calibrated using the Fundamental Plane, and is given by

\[ d_A = \frac{r}{\theta} = \frac{d_M(z)}{1 + z}. \]

In order to calculate a cosmological volume, e.g. that of a galaxy survey, due to the expansion we must integrate an infinitesimal volume element, as a function of solid angle and redshift, over redshift. The *comoving volume* \( V_C \) is the volume measure in which the number density of objects in the Hubble flow stays constant with redshift. It is equal to the proper volume times three factors of the relative
scale factor between now and the maximum redshift of the volume, i.e. \((1+z)^3\), and it is equal to the proper volume at the present day. The comoving volume element with solid angle \(d\Omega\) (in steradians) and redshift interval \(dz\) is

\[
dV_C(z) = \frac{c}{H_0} \frac{(1+z)^2 d_A(z)^2}{E(z)} d\Omega dz.
\] (1.29)

For a flat universe, \(d_A = d_C/(1+z)\) and so the total comoving volume out to redshift \(z_{\text{max}}\) within solid angle \(\Omega\) is

\[
V_C = \left( \frac{c}{H_0} \right)^3 \Omega \int_0^{z_{\text{max}}} \frac{dz}{E(z)} \left( \int_0^z \frac{dz'}{E(z')} \right)^2.
\] (1.30)

### 1.2 Structure Formation

In the ΛCDM paradigm, the Universe began as an extremely hot and dense quark-gluon plasma, with a highly homogeneous density. The structure in the Universe began as quantum fluctuations in the plasma, which were seeded and expanded during inflation. As the Universe expanded and cooled, these density fluctuations grew via gravitational instability to form the large-scale structure visible today. These density fluctuations are a rich source of information about the nature of the primordial Universe, the cosmic expansion history and the nature of gravity.

#### 1.2.1 Gravitational Instability and the Growth Rate

The large-scale structure in the Universe today is far removed from the tiny fluctuations that existed in the primordial density field. These primordial perturbations grew as matter fell into overdense regions, under the gravitational instability paradigm. This process is described by the equations of mass continuity, force and gravitation. In an expanding universe, in proper coordinates, and for scales well within the horizon, these are \(^1\)

\(^1\)These equations ignore relativistic effects and assume pressure terms are negligible (Strauss & Willick [1995]).
1.2. STRUCTURE FORMATION

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (1.31)
\]

\[
\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} + \nabla \phi = 0, \quad (1.32)
\]

\[
\nabla^2 \phi = 4\pi G \rho, \quad (1.33)
\]

where \( \rho \) is the mass density field, \( \mathbf{v} \) is the velocity field, \( \phi \) is the gravitational potential and \( G \) is the gravitational constant. Equation (1.33) is the familiar Poisson equation. Expanding these equations to first order in linear perturbation theory and taking the time derivative of the continuity equation, the divergence of the velocity field can be directly related to the density field in comoving coordinates via (Strauss & Willick, 1995)

\[
\nabla \cdot \mathbf{v} = -a \frac{\partial \delta}{\partial t} = -a H_0 f \delta, \quad (1.34)
\]

where \( \delta \) is the dimensionless density contrast,

\[
\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}}, \quad (1.35)
\]

with \( \bar{\rho} \) the mean mass density, and \( f \) is the growth rate,

\[
f \equiv \frac{1}{H D} \frac{dD}{dt} = \frac{d\ln D}{d\ln a}, \quad (1.36)
\]

with \( D \) the linear growth factor,

\[
D(t) = \frac{\dot{a}}{a} \int_0^a \dot{a}^{-3} da. \quad (1.37)
\]

The growth rate \( f \) is a useful parameter for distinguishing different models of gravity, since it depends on the strength of the gravitational force. It is strongly dependent on the amount of gravitational matter in the Universe, and hence the matter density \( \Omega_m \), so is commonly parameterised as \( f(z) \simeq \Omega_m(z)^{\gamma} \), where \( \gamma \), the growth index, is a constant for a particular model (in GR, \( \gamma \approx 0.55 \), Linder, 2005).
Here, $\Omega_m(z)$ is the matter density at redshift $z$, given by

$$\Omega_m(z) = \frac{H_0^2}{H(z)^2} \Omega_m(z = 0)(1 + z)^3,$$

(1.38)

and $H_0^2/H(z)^2$ is calculated via Equation 1.22.

1.2.2 The Correlation Function and Power Spectrum

The two-point autocorrelation function, usually referred to as the correlation function, $\xi(r)$, is a powerful way of characterising the clustering of the galaxy and matter distributions as a function of scale. For a continuous density field $\delta(r)$, it is defined by

$$\xi(r) \equiv \langle \delta(x)\delta(x + r) \rangle,$$

(1.39)

where the averaging is over position $x$. If the Universe is isotropic then $\xi(r)$ is independent of the direction of $r$.

The galaxy distribution is a point process rather than a density field; for galaxies, the correlation function is defined as the excess probability of finding two galaxies in volumes $dV_1$ and $dV_2$, separated by distance $r$, above that expected for a random Poisson distribution (Peebles, 1980):

$$P(r) = \bar{\rho}^2[1 + \xi(r)]dV_1dV_2,$$

(1.40)

where $\bar{\rho}$ is the mean number density of galaxies.

The power spectrum $P(k)$ is the Fourier transform of the correlation function, and hence describes the Fourier-space clustering properties of density perturbations. The Fourier transform $\tilde{\delta}(k)$ of the density field $\delta(r)$, as a function of comoving coordinates $r$, is

$$\tilde{\delta}(k) = \int d^3r e^{-ik\cdot r}\delta(r), \quad \delta(r) = \frac{1}{(2\pi)^3} \int d^3k \tilde{\delta}(k)e^{ik\cdot r}.$$

(1.41)

In $\Lambda$CDM, the Universe is assumed to be homogeneous and isotropic, and the initial density perturbations laid down by inflation were close to Gaussian. This means that the statistical properties of $\tilde{\delta}(k)$ are independent of direction $\hat{k}$, and can be
1.2. STRUCTURE FORMATION

entirely described by their 2-point statistics. The power spectrum $P(k)$ therefore provides an adequate description, and is defined by

$$
\langle \tilde{\delta}(k)\tilde{\delta}(k') \rangle = (2\pi)^3 P(k)\delta_D^3(k - k'),
$$

(1.42)

where $\delta_D^3$ is a 3D Dirac-delta function and the averaging on the left hand side is over directions of $k$. The function $P(k)$ is technically a power spectral density and thus represents the power per unit volume in $k$-space [Bertschinger 1992, Strauss & Willick 1995]; it has units of volume cubed, usually $h^{-3} \text{Mpc}^3$.

In an isotropic universe, the power spectrum has no preferred direction, and the correlation function is simply related to the power spectrum via a Hankel transform,

$$
\xi(r) = \frac{V}{(2\pi)^3} \int P(k) j_0(kr) 4\pi k^2 dk,
$$

(1.43)

where $j_0(kr) = \sin(kr)/kr$ is the zeroth order spherical Bessel function and $V$ is the volume over which $\xi$ is averaged.

In general, the full matter distribution cannot be directly observed, only the distribution of baryonic matter, most commonly in the form of galaxies, or intergalactic neutral hydrogen. As we have seen, baryons only make up $\sim 4\%$ of matter, and they do not directly follow the clustering of dark matter. The relative clustering of galaxies ($\delta_g$) to dark matter ($\delta_m$) is described in terms of a bias parameter, $b$ [Kaiser 1984]. On linear scales this is normally assumed to be linear and scale independent, so that

$$
\delta_g(r) = b\delta_m(r).
$$

(1.44)

The relation between the correlation function or power spectrum of galaxies and dark matter is then

$$
\xi_g(r) = b^2 \xi_m(r), \quad P_g(k) = b^2 P_m(k).
$$

(1.45)

A common way of quantifying the density fluctuation field is in terms of the variance of the mass fluctuations in spherical shells of radius $R$,
$$\sigma_R^2 = \left\langle \left( \frac{\delta M}{M} \right)^2 \right\rangle_R$$
$$= \frac{1}{(2\pi)^3} \int d^3k P(k) \tilde{W}_{\text{TH}}^2(kR)$$
$$= \frac{V}{(2\pi)^3} \int k^2 P(k) \tilde{W}_{\text{TH}}^2(kR) dk,$$

where $\tilde{W}_{\text{TH}}^2(kR)$ is the Fourier transform of a top hat window function. The variance on scales of $8 h^{-1}$ Mpc, denoted $\sigma_8$, is commonly used to normalise the power spectrum, $P(k) \propto \sigma_8^2$. In practice, when measuring galaxy clustering, $\sigma_8$ is degenerate with $b$.

The window function accounts for the fact that a given survey only observes a finite volume, and acts to smooth, or filter, the density field. It acts via a convolution with the correlation function in real space, or a multiplication with the power spectrum in Fourier space. The top hat window function is equal to 1 out to a radius $R$ then drops to zero,

$$\tilde{W}_{\text{TH}}(kR) = \frac{3j_1(kR)}{kR},$$

where $j_1(x) \equiv (\sin x - x \cos x)/x^2$ is the first order spherical Bessel function. Another common function is the Gaussian window function,

$$\tilde{W}_G(kR) = \exp(-k^2 R^2/2).$$

### 1.2.3 Peculiar Velocity Statistics

Along with the clustering properties of the matter distribution, the velocities that arise from structure growth, so-called peculiar velocities, also provide a wealth of information. We refer the reader to Kaiser (1988); Górski (1988); Peebles (1993); Strauss & Willick (1995); Dodelson (2003); Li et al. (2012) for further background on peculiar velocity theory.

In linear theory, velocities are directly related to matter overdensities. From Equation 1.34 we can relate the velocity field to the density field, in proper coordi-
nates, by (Peebles 1993)

\[ v(\mathbf{r}) = \frac{H_0 f}{4\pi} \int d^3\mathbf{r}' \frac{\delta(\mathbf{r}') (\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3}. \]  

(1.49)

Hence, in linear theory, peculiar velocities are sensitive to the total matter distribution over a large range of scales. By measuring both the velocity field and the density field, it is thus possible to constrain the growth rate. However, in practice, since dark matter cannot be directly observed it is usually the galaxy distribution \( \delta_g(\mathbf{r}) \) which is measured (e.g. using galaxy redshift surveys). As we have seen, this is biased, \( \delta_g = b \delta \). So rather than constraining \( f \), this allows us to constrain the redshift-space distortion parameter \( \beta \),

\[ \beta = \frac{f(z)}{b}. \]  

(1.50)

So when comparing peculiar velocities with the galaxy distribution, Equation (1.49) becomes

\[ v(\mathbf{r}) = \frac{H_0 \beta}{4\pi} \int d^3\mathbf{r}' \frac{\delta_g(\mathbf{r}') (\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3}. \]  

(1.51)

A common way of analysing the peculiar velocity field is in terms of its multipoles (Haugboelle et al. 2007; Feldman et al. 2010). The most commonly studied multipole of the peculiar velocity field is the dipole, or bulk flow. The bulk flow is the average velocity of a given volume of space, with respect to a given frame of reference. Most commonly, the volume is taken to be a spherical region centred on us, and the bulk flow is defined with respect to the CMB. The bulk flow, so defined, is then given by (Clutton-Brock & Peebles 1981; Kaiser 1988)

\[ V(R) = \frac{3}{4\pi R^3} \int_{r=0}^{R} v(\mathbf{r}) d^3r, \]  

(1.52)

where \( R \) is the radius of the sphere in which the bulk flow is measured. In practice, the bulk flow is measured within a particular survey window function \( W(R) \), and is therefore a convolution of the velocity field with the window function,

\[ V(R) = v \otimes W(R) = \frac{1}{(2\pi)^3} \int v(\mathbf{k}) \tilde{W}^2(\mathbf{k}R) d^3k, \]  

(1.53)
1. INTRODUCTION

where $\widetilde{W}^2(kR)$ is the Fourier transform of $W(R)$.

The probability distribution function (PDF) of the bulk flow vector $V$ is given by

$$p(V)dV = p(V)dVp(\hat{n}_V)d\hat{n}_V,$$

(1.54)

where $\hat{n}_V$ is the unit vector in the direction of the bulk flow. If the Universe is statistically isotropic, then $\langle \hat{n}_V \rangle = 0$, and so the expectation value of the bulk flow at any location is zero, $\langle V \rangle = \langle V \cdot \hat{n}_V \rangle = 0$. However, the presence of statistical inhomogeneities on small scales (i.e. small-scale density perturbations) give rise to local bulk flows, with the distribution given above. The distribution of amplitudes depends on the amplitude of density perturbations, and hence on cosmological parameters such as $\Omega_m$ and $\sigma_8$, while the direction will depend on the local distribution of structure, which should be random below the scale of statistical homogeneity.

If the density field is close to Gaussian random, then the peculiar velocity field is expected to be Maxwellian. $N$-body simulations have shown this to be an accurate description (Li et al., 2012). For a Maxwell-Boltzmann distribution, the probability distribution function of the bulk flow amplitude $V$ is (Bahcall et al., 1994; Coles & Lucchin 1996)

$$p(V)dV = \sqrt{2\pi} \left( \frac{3}{\sigma_V^2} \right)^{3/2} V^2 \exp \left( -\frac{3V^2}{2\sigma_V^2} \right) dV,$$

(1.55)

where $\sigma_V$ is the root mean square velocity dispersion. For such a distribution the most likely (maximum likelihood) bulk flow amplitude is $V_{ML} = \sqrt{2/3} \sigma_V$, while the mean bulk flow amplitude is $\langle V \rangle = 2V_{ML}/\sqrt{\pi} = \sqrt{8/3\pi} \sigma_V$. The bulk flow amplitude is therefore entirely determined by $\sigma_V$. Analogous to Equation (1.46), $\sigma_V$ is determined for a window function of effective radius $R$ by

$$\sigma_V^2 \equiv \langle V^2(R) \rangle = \frac{H_0^2 f^2}{2\pi^2} \int P(k)\widetilde{W}^2(kR)dk.$$

(1.56)

We can see from Equations (1.46) and (1.56) that the velocity variance has two fewer powers of $k$ in the integrand than the mass variance, and so the velocity field on a given scale is sensitive to much larger scales than the density field (Strauss &
1.2. STRUCTURE FORMATION

Willick (1995). Note that Equation 1.56 is the 3D rms velocity, and should not be confused with the 1D rms velocity $\sigma_{V,i} = \sigma_V / \sqrt{3}$, given by

$$\sigma_{V,i}^2 = \frac{H_0^2 f^2}{2\pi^2} \int P(k) W_{ii}^2(kR) dk,$$

where $W_{ii}$ is the tensor window function for the $i$th coordinate direction.

1.2.4 Inflation and the Primordial Power Spectrum

The origin of the primordial density perturbations and the matter power spectrum is believed to have been inflation. Inflation (Guth, 1981; Linde, 1982; Albrecht & Steinhardt, 1982), a period of exponential expansion in the early Universe, was introduced in order to explain the extreme fine-tuning of the initial conditions necessary in the conventional Big Bang model. Two specific problems that it resolves are the horizon problem and the flatness problem, and it also explains the initial conditions of the matter power spectrum. Inflation consequently provides a mechanism that both explains, and predicts, large-scale homogeneity and isotropy, providing theoretical support for the Cosmological Principle. For an overview, see Baumann (2009) and Lyth & Liddle (2009).

The horizon problem concerns the fact that the Cosmic Microwave Background (CMB) is observed to be extremely isotropic, indicating that all points on it were in causal contact at the time of decoupling, yet in the standard Big Bang model they could not have been. It can be understood by considering the comoving particle horizon, $\tau$, the maximum distance a light ray can travel between time 0 and time $t$,

$$\tau \equiv \int_0^t \frac{dt'}{a(t')} = \int_0^a d\ln a \left( \frac{1}{aH} \right).$$

Here, $(aH)^{-1}$ is the comoving Hubble radius, beyond which objects recede faster than the speed of light. In standard Big Bang expansion without acceleration, $(aH)^{-1}$ grows monotonically, and $\tau$, or the fraction of the Universe in causal contact, increases with time,

$$\tau \propto a^{\frac{4}{3} (1+3w)}$$

where $w$ is the equation of state of the dominating component of the Universe. But
if \( \tau \) grows monotonically with time, then comoving scales entering the horizon today could not have been in causal contact at the time of decoupling.

The \textit{flatness problem} can be quantified by considering the Friedmann equation in Equation 1.16 rearranged in the form

\[
1 - \Omega_{\text{tot}}(a) = \frac{-k}{(aH)^2},
\]

(1.60)

where \( \Omega_{\text{tot}}(a) \equiv \rho(a)/\rho_{\text{crit}}(a) \), and \( \rho_{\text{crit}}(a) \equiv 3H(a)^2 \). For a flat universe \((k = 0)\), the total matter density is equal to the critical density, i.e. \( \Omega_{\text{tot}}(a) = 1 \). In the standard Big Bang model without acceleration, the comoving Hubble radius \((aH)^{-1}\) grows with time, and hence from Equation 1.60 the quantity \(|\Omega_{\text{tot}}(a) - 1|\) must diverge, meaning that \( \Omega_{\text{tot}} = 1 \) is an unstable fixed point. To explain the CMB observation that the Universe is extremely close to flat today, requires an extreme fine-tuning of \( \Omega_{\text{tot}} \) close to 1 in the primordial Universe.

Inflation explains the above two problems by postulating a phase of \textit{decreasing} Hubble radius \((aH)^{-1}\) in the early Universe. This solves the horizon problem, since it means that (with enough \( e \)-folds of inflationary expansion), large scales entering the comoving horizon today were inside the horizon before inflation. We can also see it solves the flatness problem, because if \((aH)^{-1}\) decreases in Equation 1.60 then the Universe is driven towards \( \Omega_{\text{tot}} = 1 \) rather than away from it. A period of decreasing Hubble radius \((aH)^{-1}\) implies an accelerated expansion \( \ddot{a} > 0 \), as can be seen from the relation

\[
\frac{d}{dt}(aH)^{-1} = \frac{-\ddot{a}}{(aH)^2}.
\]

(1.61)

Along with explaining the flatness and large-scale homogeneity of the Universe, inflation also provides a mechanism for generating the primordial density perturbations. To produce accelerated expansion requires a negative pressure source in the early Universe, which is most easily explained by a scalar field \( \phi \), called the \textit{inflaton}, whose potential energy \( V(\phi) \) dominates over the kinetic energy \( \frac{1}{2} \dot{\phi}^2 \). The scalar field rolls down its potential from a high to low energy state, during which inflation occurs. Quantum fluctuations in the background evolution of \( \phi \) cause local delays in the time that inflation ends, inducing relative density perturbations \( \delta \rho(x, t) \). As they form, the Hubble radius continues to decrease, and they sequentially leave the
horizon. This produces a power spectrum which has a scale dependence determined by the time dependence of the Hubble parameter, and which is described by

$$P(k) \propto k^{n_s},$$

(1.62)

where $n_s$ is the scalar spectral index,

$$n_s - 1 \equiv \frac{d \ln \Delta^2_s}{d \ln k},$$

(1.63)

and $\Delta_s$ is the dimensionless power spectrum of comoving curvature perturbations $R$, $\Delta^2_s = (k^3/2\pi^2)P_R(k)$. For scale-invariance, $n_s = 1$; the simplest models of inflation predict $n_s \approx 0.96$.

### 1.3 Observational Cosmology

The standard model of cosmology is now strongly supported by a wide variety of cosmological probes. However, in light of the issues raised by $\Lambda$CDM, in particular the unknown nature of dark matter and dark energy, ongoing observations are crucial for determining whether there is unknown physics yet to be discovered.

This thesis concerns two particular observational probes: galaxy clustering and peculiar velocities. In this section we introduce some of the key observational probes, starting with the CMB which is perhaps the most important probe so far in establishing $\Lambda$CDM. We then introduce Type Ia supernovae and peculiar velocities, and explain where the work in this thesis fits in to the ongoing observational effort.

#### 1.3.1 Cosmic Microwave Background

The cosmic microwave background (CMB) is one of the most important existing cosmological probes, and one of the only available sources of information about the early Universe. It was formed $\sim 380 \,000$ years after the Big Bang at the decoupling epoch, when photons last scattered off electrons, and these photons have free-streamed relatively undisturbed since that time.

Overall the CMB is extremely isotropic, with a blackbody spectrum of temperature 2.725 $K$, indicating that the Universe before that time was a homogeneous
plasma in thermal equilibrium, the strongest evidence for the hot Big Bang model. However, there existed small perturbations in the primordial plasma, and photons travelling out of underdense and overdense regions were gravitationally blueshifted and redshifted, respectively \citep{Sachs:1967er}. This effect is visible in the CMB photons as tiny ($O(10^{-5})$) anisotropies, shown in the top panel of Figure 1.1. They were first detected in 1992 by the Cosmic Background Explorer \citep[COBE,][]{Smoot:1992wd}, followed by a range of ground and balloon-based experiments, most significantly the BOOMERanG experiment \citep{Lange:2001rk,Netterfield:2002yq}. In 2003, the Wilkinson Microwave Anisotropy Probe \citep[WMAP,][]{Spergel:2003cb} began a new era of high-precision satellite-based CMB measurements, followed by the most recent measurement by the Planck satellite \citep{Planck:2013kta}, and complemented by new high-resolution ground-based measurements by the South Pole Telescope \citep{Keisler:2011hw} and the Atacama Cosmology Telescope \citep{Dunkley:2011ch}.

These anisotropies provide a wealth of cosmological information. The angular power spectrum $\ell(\ell + 1)C_\ell/2\pi$ as measured by Planck, along with the best-fitting $\Lambda$CDM model, is shown in the bottom panel of Figure 1.1. This constrains six parameters, which can be used to parameterise the $\Lambda$CDM model: the baryon density $\Omega_b h^2$, the cold dark matter (CDM) density $\Omega_c h^2$, the first BAO peak position $\theta_s$, the amplitude of initial fluctuations at $k = 0.05 \text{ Mpc}^{-1}$, $A_s$, the scalar index of primordial perturbations $n_s$, and the optical depth to reionisation, $\tau$. From these, six derived parameters can be obtained (making assumptions including flatness, no warm dark matter, and an effective number of neutrinos $N_{\text{eff}} = 3.04$): the baryon fraction $\Omega_b$, the CDM fraction $\Omega_c$, the cosmological constant fraction $\Omega_\Lambda$, the Hubble constant $H_0$, the amplitude of late-time fluctuations on $8 h^{-1} \text{ Mpc}$ scales $\sigma_8$, and the redshift of reionisation, $z_{\text{re}}$.

1.3.2 Type Ia Supernovae

Type Ia Supernovae (SNe Ia) have played an extremely important role in cosmology, since they are a type of ‘standard candle,’ an object with a universally calibratable intrinsic magnitude. The physics of SN Ia explosions is not well known, and several possible mechanisms have been proposed. One possibility is that they are a run-
1.3. OBSERVATIONAL COSMOLOGY

Figure 1.1: Top: the *Planck* first data release CMB map (with 3% of the sky replaced by a constrained Gaussian realisation). From Planck Collaboration et al. (2013). Bottom: The 2013 *Planck* CMB temperature power spectrum, from Planck collaboration et al. (2013). The green line shows the best-fit six parameter ΛCDM model spectrum. The error bars include cosmic variance, which is shown by the green shaded area around the model.
away thermonuclear explosion of a carbon-oxygen white dwarf star, which occurs when the white dwarf accretes matter from a companion star until it exceeds the critical Chandrasekhar mass of $\sim 1.4M_\odot$ (e.g. Mazzali et al., 2007). Type Ia SNe show a strong silicon absorption line, indicating the occurrence of carbon fusion. Whatever the details of the mechanism are, they give rise to a relatively uniform peak magnitude, enabling the use of SNe Ia as a distance indicator.

For SNe Ia, the observed apparent magnitude $m$ is related to the peak intrinsic magnitude $M$ by

$$m(z) = 5 \log d_L(z) + [M + 25 - 5 \log(H_0)].$$

(1.64)

The quantity within brackets is often denoted as a nuisance parameter, $\mathcal{M} \equiv M - 5 \log H_0 + 25$, which folds in both the uncertainty on the absolute magnitude, and the Hubble constant.

Type Ia SNe have a characteristic light curve, the plot of their absolute magnitude or luminosity as a function of time after the explosion. The luminosity initially increases rapidly for about 15 days, until it reaches a peak, after which it declines for the next 3 – 4 weeks. SNe Ia are found to have some dispersion in both their peak magnitude and light curve shape, which was initially problematic for making precise cosmological measurements. However, Pskovskii (1977) found that there is a strong correlation between the peak magnitude and the width of the light curve, which was later quantified by Phillips (1993). This allowed SNe Ia to be calibrated using a ‘light curve template.’ The first of these, introduced by Riess et al. (1996), was the multicolour light-curve shapes (MCLS) method, which accounted for the correlation of colour with light curve shape, and reduced the peak magnitude dispersion to 0.12 mag. The more recently introduced SALT (Guy et al., 2007) and SiFTO (Conley et al., 2008) light curve fitters empirically derive the colour-luminosity dependence.

The distance measurements provided by SNe Ia allow a direct geometrical measurement of the expansion history of the Universe, by mapping out the Hubble diagram. In 1998 the high-redshift Hubble diagram was measured for the first time by the High-$z$ Supernova Search Team (HIZ; Riess et al., 1998) and the Supernova Cosmology Project (SCP; Perlmutter et al., 1999), which gave the first clear evidence for the accelerating expansion of the Universe. Figure 1.2 shows the Hubble
diagram results from Riess et al. (1998), which favour a flat $\Omega = 0.76$ model over $\Omega = 0$ models. In combination with evidence from WMAP CMB data and galaxy redshift survey data, this established $\Lambda$CDM as the leading theory of cosmology.

As well as measuring the Hubble diagram, distance indicators such as SNe Ia enable local measurements of galaxy peculiar velocities, as we will explore in more detail in Section 1.3.3.

1.3.3 Peculiar Velocities

In recent years, peculiar velocities (PVs) have emerged as a promising cosmological probe. As we have seen, they are directly related to the total matter distribution, over a wide range of scales. They hence provide a way of probing the distribution of dark matter, and the nature of gravity.

In practice, we can only observe the line-of-sight component of a galaxy’s peculiar velocity, rather than the full 3D vector. The line-of-sight component of the peculiar velocity $v$ of a galaxy at position $r$ is given by

$$v_p \equiv v \cdot \hat{r} = c \left( \frac{z_{\text{obs}} - z_r}{1 + z_r} \right),$$

(1.65)

where $c$ is the speed of light, $z_{\text{obs}}$ is the observed redshift, measured spectroscopically and corrected to the CMB rest frame, and $z_r$ is the redshift corresponding to the real-space distance of the galaxy, or equivalently its cosmological recession velocity at real comoving distance $r$.

In order to measure a peculiar velocity, a redshift-independent measurement of the distance $r$ is required. There are various methods of measuring distances, using so-called distance indicators (DIs), which usually entail some form of standard candle or standard ruler. A standard candle is a type of object with a calibratable luminosity, such as Type Ia Supernovae or spiral galaxies following the Tully-Fisher relation (Tully & Fisher 1977), an empirical relation relating their intrinsic luminosity and velocity width. A standard ruler is an object whose size is known or can be calibrated, such as the radii of early-type galaxies calibrated using the Fundamental Plane relation (Dressler et al. 1987b), which relates their radius, velocity dispersion and luminosity. For a review of different distance indicators, see Jacoby.
1. INTRODUCTION

Figure 1.2: Hubble diagram for 10 high-redshift SNe from the High-z Supernova Search Team, from Riess et al. (1998). Top panel: the Hubble diagram for high- and low-redshift SNe with distances measured using the MLCS method. Overplotted are three cosmologies: low and high $\Omega_m$, with $\Omega_\Lambda = 0$, and the best fit for a flat cosmology, $\Omega_m = 0.24$ and $\Omega_\Lambda = 0.76$. Bottom panel: the difference between the data and models, showing favour for $\Omega_\Lambda = 0.76$. 

Ω M=0.24, Ω Λ =0.76
Ω M=0.20, Ω Λ =0.00
Ω M=1.00, Ω Λ =0.00
et al. (1992). Distance indicators typically have large uncertainties, which dominate the uncertainty on the peculiar velocity, and are usually a constant fraction of distance. These have posed the greatest challenge for peculiar velocity measurements to date. At distances we typically want to analyse PVs, say $z = 0.01$, a 20% distance uncertainty translates to an error of $\sim 600 \text{ km s}^{-1}$, significantly larger than the typical value of peculiar velocities, of order $300 \text{ km s}^{-1}$. This means that typically peculiar velocities can only be analysed statistically, and at low-redshift. Recently some alternative new methods have been proposed, which instead of using DIs use galaxy two-point correlation functions (Song et al., 2011) and fluctuations in galaxy magnitudes (Nusser et al., 2011; Abate & Feldman, 2012).

There have been a number of methods proposed in the literature to obtain cosmological information from peculiar velocities. These include their correlation function (Górski, 1988; Gordon et al., 2007), their dipole or bulk flow (Rubin et al., 1976; Kaiser, 1988; Feldman & Watkins, 1994; Watkins & Feldman, 1995), their angular power spectrum (Haugboelle et al., 2007; Hannestad et al., 2008), their 3D power spectrum (Scoccimarro, 2004; Burkey & Taylor, 2004; Jaffe & Kaiser, 1995; Abate & Erdoğdu, 2009; Macaulay et al., 2011), their covariance (Abate & Lahav, 2008) or their mean pairwise velocity (Ferreira et al., 1999; Bhattacharya et al., 2011).

Another approach is to make a direct comparison of the peculiar velocity and density fields, using both a peculiar velocity survey and a galaxy redshift survey covering the same volume of space. This is done by using the density field inferred from the distribution of galaxies, to reconstruct the expected velocity field, with which the measured velocities are compared. Such methods include the POTENT method of Bertschinger & Dekel (1989); Dekel et al. (1990, 1999), the Wiener filter-based method of Erdoğdu et al. (2004) and the second-order Lagrangian perturbation theory-based method of Kitaura et al. (2012a, b). We saw in Equation 1.51 that this allows a constraint of the redshift-space distortion parameter $\beta$, as well as the gravitational instability paradigm itself. Hudson & Turnbull (2012) recently did this for a combination of the COMPOSITE peculiar velocity sample with the IRAS Point Source Catalog Redshift (PSCz) survey galaxy density field (Saunders et al., 2000), while Magoulas et al. (2013, in preparation) combined the 6dFGSv sample with the Two-Micron All-Sky Redshift Survey (2MRS Huchra et al., 2012; Erdoğdu et al. 2012, submitted).
Many peculiar velocity analyses have focused on measuring the local bulk flow, the large-scale motion of our local volume, which we defined in Equation 1.52. Since it is a large-scale measurement, it is the easiest statistic to measure with sparse, noisy data, and it is on linear scales making it straightforward to compare with theory. There has been a long history of bulk flow measurements starting in the 1970’s (see e.g. Kaiser 1988; Strauss & Willick 1995 for a review), although many early measurements were troubled by having small, noisy datasets. Some early measurements found a large local bulk flow (Rubin et al., 1976; Dressler et al., 1987a), while others found values consistent with predictions (Hart & Davies 1982; de Vaucouleurs & Peters 1984; Aaronson et al., 1986). More recently, larger and better-quality peculiar velocity datasets have come available, which have led to a surge of new measurements. Several of these also find evidence for an unusually large bulk flow, in conflict with ΛCDM predictions (Kashlinsky et al., 2008; Watkins et al., 2009; Feldman et al., 2010; Abate & Feldman, 2012; Lavaux et al., 2013), while others find a smaller bulk flow consistent with ΛCDM (Colin et al., 2011; Nusser & Davis, 2011; Osborne et al., 2011; Dai et al., 2011; Turnbull et al., 2012; Ma & Scott, 2013; Planck Collaboration et al., 2013b).

The CMB provides an alternative way of measuring peculiar velocities to DIs, via the kinetic Sunyaev-Zeldovich (kSZ) effect (Sunyaev & Zeldovich, 1980). The kSZ effect occurs when CMB photons scatter off hot electrons in the intra-cluster medium of galaxy clusters. If the overall cluster has a peculiar velocity with respect to the CMB, the scattered photons gain a Doppler shift due to the component of this velocity along the line of sight. There have been several recent peculiar velocity measurements of clusters using the kSZ effect (Kashlinsky et al., 2008; Hand et al., 2012; Lavaux et al., 2013; Planck Collaboration et al., 2013b). Kashlinsky et al. (2008) used WMAP three year data with an all-sky combination of X-ray cluster catalogues, and claimed to find a large, coherent bulk flow of $\sim 1000\,\text{km}\,\text{s}^{-1}$ out to $z \sim 0.1$, which they term the ‘dark flow’, in strong disagreement with ΛCDM. However, Keisler (2009) showed that they underestimated their uncertainties, significantly reducing the significance of their result. More recently, Planck Collaboration et al. (2013b) combined Planck CMB data with the Meta Catalogue of X-ray detected Clusters of galaxies (MCXC). They did not find evidence for the dark flow, finding an upper limit of $254\,\text{km}\,\text{s}^{-1}$ for a local volume of radius $2h^{-1}\text{Gpc}$, although
Atrio-Barandela (2013) disputed their result, claiming their uncertainties were overestimated. Planck cannot constrain the local bulk flow on scales up to $120 \, h^{-1} \, \text{Mpc}$, but it appears to rule out the large-scale bulk flow claimed by Kashlinsky et al. (2008, 2010), and theoretical consequences of it such as a tilted Universe.

1.4 Thesis Motivation and Overview

In this thesis, we present two different measurements of large-scale structure that can be used to test the ΛCDM cosmological model. The first is an investigation of the large-scale homogeneity of the Universe, a key assumption of ΛCDM, using the WiggleZ Dark Energy Survey. The second is an investigation of the bulk flow in the local Universe, using two different peculiar velocity datasets: the 6 degree Field Galaxy Survey peculiar velocity sample (6dFGSv), and the currently-running SkyMapper Supernova and Transient Survey (SMT).

1.4.1 Testing large-scale homogeneity in the WiggleZ Dark Energy Survey

In the first part of this thesis, Chapter 2, we present an analysis of the large-scale homogeneity of the galaxy distribution, using the WiggleZ Dark Energy Survey. WiggleZ (Drinkwater et al., 2010) is a large, spectroscopic redshift survey of $\sim 200,000$ blue, star-forming galaxies, in 1000 deg$^2$ of sky, out to redshift $z = 1$. It has a total volume of $\sim 1 \, h^{-3} \, \text{Gpc}^3$, making it ideal for a measurement of the large-scale homogeneity in the galaxy distribution.

This is an important consistency test of the ΛCDM model, because as we have seen, the assumption of large-scale homogeneity and isotropy is one of the most fundamental assumptions of the model. It plays an extremely important role in cosmology; it is used whenever galaxy redshifts are converted to distances, since this assumes the light has been travelling through an FRW metric, and it is necessary for many statistical tests. The commonly used two-point correlation function $\xi(r)$ and power spectrum $P(k)$ are only defined for a point distribution that is statistically homogeneous, and applying them to a survey smaller than the scale of homogeneity would give misleading results. In addition, aspects of the ΛCDM model, most
importantly dark energy, are inferred from observations under the assumption of homogeneity; it has been proposed that large-scale inhomogeneity could equally reproduce the observational signatures of accelerated expansion \cite{Wiltshire2007b, Rasanen2011}. Due to its importance, it is crucial that homogeneity be tested as rigorously as possible.

We present the largest-volume measurement to date of the transition to homogeneity using the WiggleZ galaxy distribution. We measure both the mean counts-in-spheres, $N(<r)$, following \cite{Hogg2005}, as well as the fractal correlation dimension $D_2(r)$. The correlation dimension is defined as

$$N(<r) \propto r^{D_2}, \quad (1.66)$$

where $N(<r)$ is the mean number of galaxies within distance $r$ of a galaxy, from which we have

$$D_2(r) = \frac{d \ln N(<r)}{d \ln r}. \quad (1.67)$$

From this, we can see that a homogeneous distribution of points will have $D_2 = 3$, while a clustered distribution will have $D_2 < 3$. By measuring the behaviour of $D_2(r)$ over a wide range of scales, from much smaller to much larger than the expected homogeneity scale, we investigate the transition to homogeneity, and compare it with a $\Lambda$CDM prediction.

One of the main challenges for homogeneity measurements is dealing with biases that occur at the survey edges \cite{Pan2002, Gabrielli2005}. We test this for our analysis using a suite of fractal mock catalogues, which incorporate the WiggleZ selection function, and have a range of input fractal dimensions $D_2$. We also make use of the GiggleZ $N$-body simulation \cite{Poole2014} with the WiggleZ selection imposed. By performing our analysis on these mocks, we show that there is no significant bias in our results over the relevant scales on which we detect homogeneity in WiggleZ.

The analysis we present here is the largest-volume measurement to date of the transition to homogeneity in the galaxy distribution, and the first to demonstrate it over a range of epochs (in the redshift range $0.1 < z < 0.9$). Our result, in strong agreement with a $\Lambda$CDM model prediction incorporating redshift-space distortions, provides a strong confirmation of the FRW-based $\Lambda$CDM model.
1.4.2 Bulk Flow tests with SkyMapper and the 6 degree Field Galaxy Survey peculiar velocities

In the second part of this thesis, we investigate the bulk flow of the local Universe, as measured by galaxy peculiar velocities. This is currently an area of much interest, since local bulk flow measurements are one of the few probes which currently appear to disagree with ΛCDM predictions. A large local bulk flow, in disagreement with the velocity dispersion predicted by ΛCDM, could have a number of implications. It could indicate a disagreement with certain parameters within ΛCDM – e.g. it suggests a higher local value of $\sigma_8$ than that measured from the CMB (Watkins et al., 2009). It could indicate the breakdown of the Cosmological Principle, if we happen to reside in an unlikely part of the Universe that happens to have a large local bulk flow. Or, it could also indicate some form of physics beyond ΛCDM. One possibility is a ‘tilted’ universe (Turner, 1991), in which pre-inflationary fluctuations in scalar fields such as the inflaton or axion, or pre-inflationary density perturbations, produce a dipole anisotropy in the CMB. This would mean that all galaxies fixed in the Hubble flow would have a uniform velocity in the CMB frame, and it would appear as a non-converging bulk flow that continues below the local volume. It could be that the large bulk flow measurements can be explained by an as-yet unknown systematic, but the explanation of new physics remains an intriguing possibility.

In Chapter 3, we use the 6 degree Field Galaxy Survey peculiar velocity sample (6dFGSv) to make a measurement of the local bulk flow on a scale of 70 $h^{-1}$ Mpc. We do this using two different estimators of the bulk flow; the Maximum Likelihood Estimate of Kaiser (1988), and the more optimised ‘Minimum Variance’ estimate of Watkins et al. (2009); Feldman et al. (2010). With a dataset of $\sim 9000$ Fundamental Plane peculiar velocities, 6dFGSv is the largest, most homogeneously selected peculiar velocity sample to date. As we show, it provides the most accurate measurement to date of the local bulk flow amplitude.

In Chapter 3 we present predictions for the accuracy with which the SkyMapper Supernova and Transient Survey (SMT) will measure the bulk flow. SMT, currently running on the SkyMapper telescope, aims to measure $\sim 500$ SNe over 5 years, which will provide distances that can be used to measure peculiar velocities. We perform a realistic simulation of SMT, and use the Minimum Variance bulk flow
We find that SMT in its current configuration, with 900 fields in the southern hemisphere, should measure the bulk flow to an accuracy of $\sim 140 \text{ km s}^{-1}$ on a scale of $\sim 150 h^{-1} \text{ Mpc}$. In combination with a northern-sky supernova survey, this would provide competitive new constraints on the local bulk flow. We also investigate how the properties of a generalised supernova survey, such as the sky coverage and magnitude uncertainty, affect the accuracy of the bulk flow measurement, as a guide for future surveys. We find that keeping the same number of fields, but distributing them isotropically over the whole sky, would increase the accuracy by 30%, while decreasing the magnitude uncertainty from 0.14 to 0.10 mag would give an 8% improvement.

We summarise and conclude in Chapter 5.
Large-scale cosmic homogeneity in the WiggleZ Dark Energy Survey


2.1 Abstract

We have made the largest-volume measurement to date of the transition to large-scale homogeneity in the distribution of galaxies. We use the WiggleZ survey, a spectroscopic survey of over 200,000 blue galaxies in a cosmic volume of $\sim 1 \, h^{-3} \text{Gpc}^3$. A new method of defining the ‘homogeneity scale’ is presented, which is more robust than methods previously used in the literature, and which can be easily compared between different surveys. Due to the large cosmic depth of WiggleZ (up to $z = 1$) we are able to make the first measurement of the transition to homogeneity over a range of cosmic epochs. The mean number of galaxies $N(<r)$ in spheres of comoving radius $r$ is proportional to $r^3$ within 1 per cent, or equivalently the fractal dimension of the sample is within 1 per cent of $D_2 = 3$, at radii larger than $71 \pm 8 \, h^{-1} \text{Mpc}$ at $z \sim 0.2$, $70 \pm 5 \, h^{-1} \text{Mpc}$ at $z \sim 0.4$, $81 \pm 5 \, h^{-1} \text{Mpc}$ at $z \sim 0.6$, and $75 \pm 4 \, h^{-1} \text{Mpc}$ at $z \sim 0.8$. We demonstrate the robustness of our results against selection function effects, using a ΛCDM $N$-body simulation and a suite of inhomogeneous fractal distributions. The results are in excellent agreement with both the ΛCDM $N$-body simulation and an analytical ΛCDM prediction. We can exclude a fractal
distribution with fractal dimension below $D_2 = 2.97$ on scales from $\sim 80\, h^{-1}\, \text{Mpc}$ up to the largest scales probed by our measurement, $\sim 300\, h^{-1}\, \text{Mpc}$, at 99.99 per cent confidence.

### 2.2 Introduction

One of the main assumptions of the standard theory of cosmology, ΛCDM (based on cold dark matter and a cosmological constant), is that the Universe is homogeneous and isotropic on large scales, and hence can be described by the Friedmann-Robertson-Walker (FRW) metric. ‘Homogeneous’ means that its statistical properties (such as density) are translationally invariant; ‘isotropic’ means it should be rotationally invariant. The Universe clearly deviates from this on small scales, where galaxies are clustered, but on large enough scales ($\gtrsim 100\, h^{-1}\, \text{Mpc}$ in ΛCDM), the distribution of matter is assumed to be ‘statistically homogeneous’ – i.e., the small-scale inhomogeneities can be considered as perturbations, which have a statistical distribution that is independent of position. However, this is merely an assumption, and it is important for it be accurately verified by observation. Over the last decade there has been a debate in the literature as to whether the Universe really is homogeneous, or whether it has a fractal-like structure extending to large scales. It is important to resolve this contention if we are to be justified in assuming the FRW metric.

In fact, although ΛCDM is based on the assumption of large-scale homogeneity and an FRW metric, inflation (which ΛCDM incorporates) actually predicts a certain level of density fluctuations on all scales. Inflation predicts that the primordial density power spectrum was close to scale-invariant. In the standard model, the scalar index $n_s$, which quantifies the scale-dependence of the primordial power spectrum, is close to 0.96 (Baumann 2009), while a scale-invariant power spectrum has $n_s = 1$. In this case, these density fluctuations induce fluctuations in the metric, $\delta \Phi$, which are virtually independent of scale, and are on the order of $\delta \Phi / c^2 \sim 10^{-5}$ (Peacock 1999). Since these perturbations are small, the FRW metric is still valid, but it means that we expect the Universe to have a gradual approach to large-scale homogeneity rather than a sudden transition.

The most important implication of inhomogeneity is the so-called ‘averaging
problem’ in General Relativity (GR). This arises when we measure ‘average’ quantities (such as the correlation function and power spectrum, and parameters such as the Hubble constant) over a spatial volume. In doing so we assume the volume is homogeneous and smooth, when it may not be. Since the Einstein equations are nonlinear, density fluctuations can affect the evolution of the average properties of the volume – this is known as the ‘backreaction mechanism’ (e.g. Buchert (2000), Ellis & Buchert (2005), Li & Schwarz (2007); see Räätänen (2011) for a summary). If we observe a quantity within such a volume, we need to take averaging into account to compare it with theory. It is therefore important to know how much inhomogeneity is present, in order to obtain meaningful results from averaged measurements (and so most, if not all, cosmological measurements).

Backreaction has also been proposed as an explanation of dark energy, which is believed to be a negative-pressure component of the Universe that drives the accelerated expansion. Some authors have suggested that instead of introducing exotic new forms of dark energy, or modifications to GR, we should revisit the fundamental assumptions of the ΛCDM model, such as homogeneity. If we assume that GR holds, but take inhomogeneities into account, it can be shown that backreaction can cause a global cosmic acceleration, without any additional dark energy component (see e.g. Schwarz, 2002; Kolb et al., 2005; Räätänen, 2006; Wiltshire, 2007a; Buchert, 2008; Räätänen, 2011). This effect appears to be too small to explain the observed acceleration, but highlights the importance of understanding the amount of inhomogeneity in the Universe.

Another, related, consequence of inhomogeneity is that it can affect the path travelled by light rays, and the calibration of clocks and rods of observers. It can therefore affect distance measurements, such as redshifts and luminosity distances (Wiltshire, 2009; Meures & Bruni, 2011). This has also been proposed as a possible explanation of the observed cosmic acceleration, although the effect appears to be on the order of only a few percent at \( z \sim 1 \) (Brouzakis et al., 2007).

Homogeneity is required by several important statistical probes of cosmology, such as the galaxy power spectrum and \( n \)-point correlation functions, in order for them to be meaningful. Applying these to a galaxy sample below the scale of homogeneity would be problematic, since if a distribution has no transition to homogeneity, it does not have a defined mean density, which is required to calculate
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK ENERGY SURVEY

and interpret these statistics. It is also not possible to model its cosmic variance, so the error in the measurements would be ill-defined, making it impossible to relate these statistics to a theoretical model. It is therefore important to quantify the scale on which the Universe becomes close enough to homogeneous to justify their use.

Large-scale homogeneity is already well supported by a number of different observations. In particular, the high degree of isotropy of the CMB (Fixsen et al., 1996) gives very strong support for large-scale homogeneity in the early Universe, at redshift $z \sim 1100$. The isotropy of the CMB also indicates the Universe has remained homogeneous, since there are no significant Integrated Sachs-Wolfe (ISW) effects distorting our view of the isotropic CMB (Wu et al., 1999). Other high-redshift evidence for homogeneity includes the isotropy of the X-Ray Background (XRB) (Peebles, 1993; Scharf et al., 2000), believed to be emitted by high-redshift sources, and the isotropy of radio sources at $z \sim 1$ (Blake & Wall, 2002).

However, these measurements of high-redshift isotropy do not necessarily imply homogeneity of the present Universe. If every point in the Universe is isotropic, then this implies the Universe is homogeneous; so if we accept the Copernican principle that our location is non-special, then the observed isotropy should imply homogeneity (Peacock, 1999). However most of these measurements (except the ISW effect) only tell us about the high-redshift Universe. We know that it has evolved to a clustered distribution since then, and it is possible that it could also have become anisotropic. It is also possible for the matter distribution to be homogeneous while the galaxy distribution is not, since the galaxy distribution is biased relative to the matter field (Kaiser, 1984) – although since galaxy bias is known to be linear on large scales (Coles, 1993; Scherrer & Weinberg, 1998), this seems unlikely. The ISW effect (Sachs & Wolfe, 1967) gives information about the low-redshift Universe, since it mostly probes the dark energy dominated era, $z \lesssim 1$ (Afshordi, 2004), but it is an integral over the line-of-sight and so does not give full 3D information.

Galaxy surveys are the only 3D probe of homogeneity in the nearby Universe, and a number of homogeneity analyses have been carried out with different surveys, with seemingly conflicting results. Most statistical methods used to measure homogeneity

---

1We note that the Copernican principle is not incompatible with an inhomogeneous Universe. It assumes only that our location is non-special, not that every location is the same (Joyce et al., 2000; Clifton et al., 2008; Sylos Labini et al., 2009).
have been based on the simple ‘counts-in-spheres’ measurement, that is, the number of galaxies \(N(<r)\) in spheres of radius \(r\) centred on galaxies, averaged over a large number of such spheres. This quantity scales in proportion to the volume \((r^3)\) for a homogeneous distribution, and homogeneity is said to be reached at the scale above which this holds. Hogg et al. (2005) applied this to the Sloan Digital Sky Survey (SDSS) luminous red galaxy (LRG) sample at \(z \sim 0.3\), and found the data became consistent with homogeneity at \(\sim 70 \, h^{-1}\) Mpc for this galaxy population (using a different method of determining the homogeneity scale than we do).

This measurement can also be extended to a fractal analysis. Fractal dimensions can be used to quantify clustering; they quantify the scaling of different moments of galaxy counts in spheres, which in turn are related to the \(n\)-point correlation functions. The most commonly used is the correlation dimension \(D_2(r)\), which quantifies the scaling of the 2-point correlation function, and is based on the counts-in-spheres, which scale as \(N(<r) \sim r^{D_2}\). One can also consider the more general dimensions \(D_q\), where \(q\) are different moments of the counts-in-spheres. Using fractal analyses, some researchers have found a transition from \(D_2 < 3\) to \(D_2 = 3\), at around \(70 – 150 \, h^{-1}\) Mpc (Martínez & Coles, 1994; Guzzo, 1997; Martínez et al., 1998; Scaramella et al., 1998; Amendola & Palladino, 1999; Pan & Coles, 2000; Kurokawa et al., 2001; Yadav et al., 2005; Sarkar et al., 2009), whereas other authors have found no such transition (Coleman & Pietronero, 1992; Pietronero et al., 1997; Sylos Labini et al., 1998; Joyce et al., 1999; Sylos Labini, 2011). However, many of the galaxy redshift surveys used in the above-mentioned works are too shallow, sparse, or have survey geometries too complicated, to give conclusive results.

In this work, we use the WiggleZ Dark Energy Survey (Drinkwater et al., 2010) to make a new measurement of the counts-in-spheres and correlation dimension, to test for the transition to homogeneity. WiggleZ provides a larger volume than previous surveys, making it ideal for a homogeneity measurement, and it covers a higher redshift range, allowing us to also investigate how homogeneity changes with cosmic epoch. It is not volume-limited, but we show that this does not significantly affect our measurement. The transition to homogeneity can be used as a test of a particular cosmological model, since we would expect it to differ for different cosmologies. In this work we test a ΛCDM model with best-fitting parameters from the Wilkinson
Microwave Anisotropy Probe (WMAP) data ([Komatsu et al., 2011](#)), which we refer to as $\Lambda$CDM+WMAP. We demonstrate the robustness of our measurement against systematic effects of the survey geometry and selection function, by repeating our analysis on both the GiggleZ $N$-body simulation and on a suite of inhomogeneous, fractal distributions.

Before we make any meaningful test of homogeneity, however, it is crucial to properly define what we mean by the so-called ‘scale’ of homogeneity. Since there is only a gradual approach to homogeneity, such a definition may be arbitrary. In the past, authors have defined the ‘scale of homogeneity’ as the scale where the data becomes consistent with homogeneity within 1-$\sigma$ ([e.g. Hogg et al., 2005](#), [Bagla et al., 2008](#), [Yadav et al., 2010](#)). However, this method has several disadvantages. It depends on the size of the error bars on the data, and hence on the survey size. A larger survey should have smaller error bars, and so will automatically measure a larger scale of homogeneity. It also depends on the bin spacing, and is susceptible to noise between data points. We therefore introduce a different, and more robust, method for determining homogeneity: we fit a smooth, model-independent polynomial curve to all the data points, and find where this intersects chosen values close to homogeneity.

Certain parts of our analysis require the assumption of a cosmological model and, implicitly, homogeneity (i.e. for converting WiggleZ redshifts to distances, correcting for the selection function, calculating the uncertainties using lognormal realisations, and finding the best-fitting bias). In these cases, we use an input $\Lambda$CDM cosmology with $h = 0.71$, $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.04482$, $\sigma_8 = 0.8$ and $n_s = 0.96$. Here, the Hubble constant is $H_0 = 100h$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_m$ is the mass density, $\Omega_\Lambda$ is the dark energy density, $\Omega_b$ is the baryon density, $\sigma_8$ is the root mean square mass variation within spheres of $8h^{-1}$ Mpc radius, and $n_s$ is the spectral index of the primordial power spectrum. This is the same fiducial cosmology used by [Blake et al., 2011a](#), and we use this for consistency. We discuss the implications of assuming a $\Lambda$CDM model on the results of our homogeneity measurement in Section 2.8.
2.3 The WiggleZ Dark Energy Survey

The WiggleZ Dark Energy Survey \cite{Drinkwater2010} is a large-scale spectroscopic galaxy redshift survey conducted at the 3.9m Anglo-Australian Telescope, and was completed in January 2011. It maps a cosmic volume of $\sim 1 \text{ Gpc}^3$ up to redshift $z = 1$, and has obtained 239,000 redshifts for UV-selected emission-line galaxies with a median redshift of $z_{\text{med}} = 0.6$. Of these, 179,599 are in regions contiguous enough to be used for our analysis. It covers $\sim 1000 \text{ deg}^2$ of equatorial sky in 7 regions, shown in \cite{Drinkwater2010} (their Figure 1).

The observing strategy and galaxy selection criteria of the WiggleZ survey are described in \cite{Blake2009, Drinkwater2010}. The selection function we use is described in \cite{Blake2010}. The targets were selected from Galaxy Evolution Explorer satellite (GALEX) observations matched with ground-based optical measurements, and magnitude and colour cuts were applied to preferentially select blue, extremely luminous high-redshift star-forming galaxies with bright emission lines.

The WiggleZ Survey offers several advantages for a new study of the scale of homogeneity. Its very large volume allows homogeneity to be probed on scales that have not previously been possible, and at a higher redshift; it probes a volume at $z > 0.5$ comparable to the SDSS LRG catalogue at $z < 0.5$. This allows us to make the first measurement of the change in the homogeneity scale over a range of cosmic epochs. We divide our sample into four redshift slices, $0.1 < z < 0.3$, $0.3 < z < 0.5$, $0.5 < z < 0.7$ and $0.7 < z < 0.9$. The sizes of the WiggleZ regions in each redshift slice are listed in Table 2.1; they sample scales well above the expected scale of homogeneity. The numbers of galaxies in each redshift slice are listed in Table 2.2.

We also benefit from having 7 regions distributed across the equatorial sky, which reduces the effect of cosmic variance. In addition, WiggleZ probes blue galaxies, whereas SDSS (which has obtained the largest-scale measurements of homogeneity to date) used Luminous Red Galaxies, and so it can also constrain any systematic effects introduced by the choice of tracer galaxy population \cite{Drinkwater2010}. Since blue galaxies are less biased, they are also more representative of the underlying matter distribution.

There are, however, several aspects of the survey that could potentially be detri-
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK ENERGY SURVEY

Table 2.1: Comoving dimensions of WiggleZ regions in each redshift slice, in units of $[h^{-1}\text{Mpc}]$. The dimensions correspond to the line-of-sight, RA, and Dec directions, respectively.

<table>
<thead>
<tr>
<th>Region</th>
<th>$0.1 &lt; z &lt; 0.3$</th>
<th>$0.3 &lt; z &lt; 0.5$</th>
<th>$0.5 &lt; z &lt; 0.7$</th>
<th>$0.7 &lt; z &lt; 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>00-hr</td>
<td>$551 \times 146 \times 222$</td>
<td>$505 \times 232 \times 353$</td>
<td>$459 \times 309 \times 471$</td>
<td>$417 \times 378 \times 575$</td>
</tr>
<tr>
<td>01-hr</td>
<td>$549 \times 127 \times 132$</td>
<td>$499 \times 202 \times 209$</td>
<td>$450 \times 269 \times 279$</td>
<td>$405 \times 329 \times 341$</td>
</tr>
<tr>
<td>03-hr</td>
<td>$549 \times 133 \times 129$</td>
<td>$499 \times 211 \times 206$</td>
<td>$450 \times 281 \times 274$</td>
<td>$405 \times 343 \times 335$</td>
</tr>
<tr>
<td>09-hr</td>
<td>$551 \times 221 \times 133$</td>
<td>$504 \times 351 \times 212$</td>
<td>$458 \times 467 \times 283$</td>
<td>$416 \times 571 \times 346$</td>
</tr>
<tr>
<td>11-hr</td>
<td>$553 \times 280 \times 145$</td>
<td>$509 \times 445 \times 231$</td>
<td>$466 \times 592 \times 308$</td>
<td>$426 \times 724 \times 376$</td>
</tr>
<tr>
<td>15-hr</td>
<td>$553 \times 295 \times 150$</td>
<td>$510 \times 468 \times 238$</td>
<td>$468 \times 623 \times 317$</td>
<td>$429 \times 762 \times 387$</td>
</tr>
<tr>
<td>22-hr</td>
<td>$550 \times 142 \times 143$</td>
<td>$500 \times 225 \times 228$</td>
<td>$452 \times 300 \times 303$</td>
<td>$408 \times 367 \times 371$</td>
</tr>
</tbody>
</table>

Table 2.2: Number of WiggleZ galaxies in each redshift slice.

<table>
<thead>
<tr>
<th>Redshift</th>
<th>Number of galaxies</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1 &lt; z &lt; 0.3$</td>
<td>25 187</td>
</tr>
<tr>
<td>$0.3 &lt; z &lt; 0.5$</td>
<td>45 698</td>
</tr>
<tr>
<td>$0.5 &lt; z &lt; 0.7$</td>
<td>70 191</td>
</tr>
<tr>
<td>$0.7 &lt; z &lt; 0.9$</td>
<td>38 523</td>
</tr>
</tbody>
</table>

WiggleZ has a complex window function, with a complicated edge geometry including holes in the angular coverage, and the spectroscopic completeness varies across the sky. In addition, the population properties of the galaxies are known to vary with redshift, due to the effects of Malmquist bias (since WiggleZ is a flux-limited survey), downsizing (the observed fact that the size of the most actively star-forming galaxies decreases with time, Cowie et al., 1996; Glazebrook et al., 2004), and the colour and magnitude selection cuts (Blake et al., 2010). This means that WiggleZ preferentially selects larger-mass, higher-luminosity galaxies at higher redshift. A consequence of this is that it is not possible to define volume-limited subsamples of WiggleZ. However, we can correct for these effects by using random catalogues (Section 2.4.1), which account for the survey selection function (Blake et al., 2010). We also divide the survey into four redshift-slices, reducing the amount of galaxy population evolution in any region. In addition, we show that our results are not biased by the assumption of homogeneity in the selection function corrections, using an $N$-body simulation and a suite of inhomogeneous fractal distributions, described in Section 2.7. We are therefore
confident that our result is not distorted by any features of the survey.

We convert the redshifts of the WiggleZ galaxies to comoving distances $d_c$, using

$$d_c(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} ,$$

(2.1)

where

$$E(z) = \frac{H(z)}{H_0} = \left[ \Omega_{m,0}(1 + z)^3 + \Omega_{\Lambda,0} \right]^{1/2} ,$$

(2.2)

and we use the fiducial ΛCDM parameter values listed in Section 1. To do this, we assume the FRW metric and ΛCDM. This is necessary for any homogeneity measurement, since we must always assume a metric in order to interpret redshifts. Therefore in the strictest sense this can only be used as a consistency test of ΛCDM. However, if we find the trend towards homogeneity matches the trend predicted by ΛCDM, then this is a strong consistency check for the model and one that an inhomogeneous distribution would find difficult to mimic. We discuss this further in Section 2.8.

### 2.4 Methodology

Here we describe our methodology for measuring the transition to homogeneity. We first calculate the mean counts-in-spheres $N(< r)$, then we find the fractal correlation dimension $D_2(r)$, from the slope of $N(r)$. Although they are closely related it is interesting to consider both, since the counts-in-spheres is the simplest measurement of homogeneity, whilst the correlation dimension provides direct information about the fractal properties of the distribution. We also describe our method of determining uncertainties, and our method of defining the ‘homogeneity scale’ $R_H$.

#### 2.4.1 Scaled counts-in-spheres $N(< r)$

The simplest test of homogeneity of a set of points is to find the average number $N(< r)$ of neighbouring points from any given point, up to a maximum distance $r$; if the distribution is homogeneous, then (for large enough $r$),

$$N(< r) \propto r^D ,$$

(2.3)
where $D$ is the ambient dimension (the number of dimensions of the space; for a homogeneous volume, $D = 3$).

We find $N(<r)$ for spheres centred on each of the WiggleZ galaxies, and correct for incompleteness by dividing by the number expected for a homogeneous distribution with the same level of completeness. (We show that this does not bias our final results). This is done by finding the mean $N(<r)$ about the coordinate position of the WiggleZ galaxy from 100 random catalogues, each with the same number density, window function and redshift distribution of the WiggleZ survey. The method of generating the random catalogues is described in [Blake et al. (2010)]. We then take the average over all the galaxies, to obtain the mean, scaled counts-in-spheres measurement $\mathcal{N}(<r)$:

$$
\mathcal{N}(<r) = \frac{1}{G} \sum_{i=1}^{G} \frac{N^i(<r)}{\frac{1}{R} \sum_{j=1}^{R} \rho_j N^i_j(<r)},
$$

(2.4)

where $G$ is the number of WiggleZ galaxies used as sphere centres, $R$ is the number of random catalogues, $N(<r)$ is the counts for WiggleZ galaxies, $N_R(<r)$ the counts for random galaxies (centred on the position of the $i^{th}$ WiggleZ galaxy), and $\rho_j \equiv n_W/n_{\text{rand},j}$ is the ratio of the total number of WiggleZ galaxies ($n_W$) to the number of random galaxies in the $j^{th}$ random catalogue ($n_{\text{rand},j}$). In our analysis, we have $G = n_W$, but this would not be the case if, for example, we excluded spheres near the survey edges.

The random catalogue correction has the effect of reducing the scaling by the number of dimensions (i.e. $D=3$), so that for a homogeneous distribution $\mathcal{N}(<r)$ scales as

$$
\mathcal{N}(<r) \propto r^{3-3} = 1.
$$

(2.5)

In each WiggleZ region we make $\mathcal{N}(<r)$ measurements in spheres with 12 to 15 logarithmically-spaced radial bins (depending on the size of the region). We determine the large-scale cutoff in each region by calculating the mean volume in the selection function, $V(r)$, in a thin shell of mean radius $r$ enclosing a central galaxy. We illustrate this for the 15-hr $0.5 < z < 0.7$ region in Figure [2.1] and compare it with the ‘true’ volume of the shells. Below $\sim 2\, h^{-1}\, \text{Mpc}$ there is noise due to low resolution, but above this the two curves are very close. On large scales however,
2.4. METHODOLOGY

Figure 2.1: An illustration of the scales for which survey edge effects become important for the homogeneity measurement. Top panel: The mean volume $\bar{V}(r)$ in a thin shell of mean radius $r$ surrounding a WiggleZ galaxy within the selection function (black curve). We show this for the 15-hr 0.5 < $z$ < 0.7 region. The true volume of the shells is shown as a blue dashed line. The black curve deviates from the blue at large scales, where an increasing proportion of the shells extends outside the survey. Bottom panel: The ratio of the mean volume to the true volume, as a function of $r$. The grey dashed line indicates a ratio of 1.

an increasing proportion of shells surrounding galaxies go off the edge of the survey, and their volume within the survey decreases. On such scales, corrections for edge effects will become important. We take the large-scale cutoff of our homogeneity measurement at the radius of the maximum-volume shell. For the region in Figure 2.1 this corresponds to shells that are $\sim$ 20% complete. We show later that edge effects up to this scale do not impact our homogeneity measurement.

Our correction method using random catalogues maximises the use of the data, and accounts for incompleteness and the fact that WiggleZ is not volume-limited. However, it can potentially bias our result towards detecting homogeneity, since it
assumes homogeneity on the largest scales of the survey. It is equivalent to weighting each measurement by the volume of the sphere included within the survey, multiplied by an arbitrary mean density. So while the counts-in-spheres measurement should have the advantage of not assuming a mean density (Hogg et al., 2005), our correction method means that we do. Therefore $N(<r)$ should tend to 1 on the largest scales of the survey, regardless of whether homogeneity has been reached. However, for a distribution with homogeneity size smaller than the survey, $N(<r)$ should reach 1, and remain at 1, for a range of scales smaller than the survey scale.

We check the robustness of our method against effects of the selection function and correction method in Section 2.7 and show that our analysis is robust against this potential source of systematic error out to scales far greater than the homogeneity scale we measure.

There are other correction methods used in the literature. There is the so-called ‘exclusion’ or ‘deflation’ method (e.g. Coleman & Pietronero, 1992; Pan & Coles, 2000, 2002; Sylos Labini et al., 2009) which only considers central points that are surrounded by complete spheres within the survey. This therefore excludes as central points any galaxies within a certain distance from the survey edges. However, this does not make the best use of the data, since it excludes data and so reduces the volume of the sample.

There is also the so-called ‘angular correction’ model, which has been shown to be more optimal, by using all the available data without introducing a bias due to edge corrections (Pan & Coles, 2002). This corrects measurements in a sphere by the solid angle subtended by regions in the sphere that are outside the survey boundary. However, as they point out, this method is difficult to apply to surveys with a complicated geometry, especially if they contain holes, as WiggleZ does. Finally, another correction that minimises bias at the survey edges, but wastes little data, is the ‘Ripley’ estimator (Ripley, 1977; Martínez et al., 1998), which corrects the measurement in a sphere by the area of the sphere contained within the survey. Both the angular correction and the Ripley estimator assume isotropy of the samples. Due to the geometry of the WiggleZ survey, we choose to use a random catalogue correction, but make robustness tests to quantify any bias it may introduce to the results.

To demonstrate the robustness of our measurement against the method of cor-
recting for the selection function, we compare our method to an analysis using only complete spheres, with and without correcting for incompleteness, in Section 2.7.3 and show we obtain consistent results.

2.4.2 Correlation dimension \( D_2(r) \)

Fractal dimensions can be used to describe the clustering of a point distribution. There exists a general family of dimensions \( D_q \), the Minkowski-Bouligand dimensions, which describe the scaling of counts in spheres centred on points (see e.g. Borgani [1995] Martínez & Saar [2002] for a review). To completely characterise the clustering of our galaxy distribution we would need to consider all the moments \( q \) of the distribution (corresponding to combinations of \( n \)-point correlation functions). However, to identify the scale of homogeneity we consider only \( D_2 \), the ‘correlation dimension,’ which quantifies the scaling behaviour of the two-point correlation function \( \xi(r) \).

If we take a galaxy and count the number of other galaxies, \( N(<r) \), within a distance \( r \), then this quantity scales as

\[
N(<r) \propto r^{D_2},
\]

where \( D_2 \) is the fractal dimension of the distribution.

From this, the correlation dimension is defined as:

\[
D_2(r) \equiv \frac{d \ln N(<r)}{d \ln r}.
\]

Since we must correct each WiggleZ \( N(<r) \) measurement for completeness, obtaining the scaled quantity \( N(<r) \propto r^{D_2-3} \), we must calculate \( D_2(r) \) via

\[
D_2(r) = \frac{d \ln N(<r)}{d \ln r} + 3.
\]

For a homogeneous distribution, \( D_2 = 3 \). If \( D_2 < 3 \) then the distribution has a scale-invariant, fractal (and so inhomogeneous) clustering pattern. If \( D_2 > 3 \) the distribution is said to be ‘super-homogeneous’ and corresponds to a lattice-like distribution (Gabrielli et al., 2002). A power-law in \( 1 + \xi(r) \sim r^{-\gamma} \) has \( D_2 = 3 - \gamma \).
for $\xi \gg 1$.

Some previous works have found $D_2(r)$ simply by fitting a straight line to a log-log plot of $N(<r)$. This method can give a false indication of a fractal (Martínez et al., 1998); calculating $D_2(r)$ explicitly gives a more reliable measurement.

We note that our estimator for $N(<r)$, Equation 2.4, is essentially equivalent to $1 + \bar{\xi}_g(r)$, where (Hamilton, 1992)

$$\bar{\xi}(r) = \frac{3}{r^3} \int_0^r x^2 \xi(x) dx. \quad (2.9)$$

(This can be seen more clearly by rearrangement of the theoretical expression for $N(<r)$ given by Equation 2.19). Many measurements of $\xi_g(r)$ have been made with different galaxy surveys (e.g. Hawkins et al., 2003; Zehavi et al., 2005; Blake et al., 2011b; Beutler et al., 2011). On small scales the correlation function is well described by a power law,

$$\xi(r) = \left( \frac{r_0}{r} \right)^\gamma, \quad (2.10)$$

where $r_0 \approx 5 h^{-1}$ Mpc is the so-called clustering length, and $\gamma \approx 1.8$, depending on the galaxy population. On scales $\gtrsim 20 h^{-1}$ Mpc however, the correlation function is observed to turn over, consistent with large-scale homogeneity.

However, the correlation function cannot be used to test large-scale homogeneity, since the way it is determined from surveys depends on the mean galaxy density. Determinations of $\xi(r)$ commonly use the Landy-Szalay (Landy & Szalay, 1993) or Hamilton estimators (Hamilton, 1993), which compare the galaxy clustering to that of random catalogues of the same mean density as the survey.

Our $N(<r)$ estimator also compares the data to random catalogues, since we must correct for the selection function. Therefore $N(<r)$, like $\xi(r)$, does assume a mean density on the scale of the survey. However, our estimator is slightly different, as we correct each object separately rather than an average pair count.

The correlation dimension $D_2(r)$, on the other hand, measures the scaling of $N(<r)$, which is not affected by the assumption of the mean density. (This only affects the amplitude of $N(<r)$). Deviations from a volume-limited sample, which
require random-catalogue corrections, only cause second-order changes to \( D_2(r) \), whilst they would be leading-order in the raw correlation function. Therefore, \( D_2(r) \) is much more robust to both the assumed mean density and details of the selection function, making it the most reliable measure of homogeneity.

### 2.4.3 Lognormal realisations and covariance matrix

Lognormal realisations \( \text{(Coles & Jones, 1991)} \) are an important tool for determining uncertainties in galaxy surveys. A lognormal random field is a type of non-Gaussian random field, which can be used to model the statistical properties of the galaxy distribution, and simulate datasets with an input power spectrum. We have used 100 such realisations, generated using an input \( \Lambda \text{CDM} \) power spectrum with the fiducial parameters listed in Section \( \text{2.2} \). These are sampled with the survey selection function, to create 100 mock catalogues for each of the WiggleZ regions. We use these to calculate the full covariance and errors of our measurement. Jack-knife resampling does not permit enough independent regions within the survey volume to give a reliable estimate of the uncertainties.

We obtain \( N(<r) \) and \( D_2(r) \) for each of the lognormal realisations in the same way as for the WiggleZ data. The covariance matrix between radial bins \( i \) and \( j \) is given by

\[
C_{ij} = \frac{1}{n-1} \sum_{l=1}^{n} [x_{l}(r_{i}) - \overline{x(r_{i})}][x_{l}(r_{j}) - \overline{x(r_{j})}],
\]

(2.11)

where \( x(r) \) is \( N(<r) \) or \( D_2(r) \), the sum is over lognormal realisations \( l \), \( n \) is the total number of lognormal realisations and \( \overline{x(r)} = \frac{1}{n} \sum_{l=1}^{n} x_{l}(r) \). The diagonal values \( j = k \) give the variance, \( \sigma^2 \).

The correlation coefficient between bins \( i \) and \( j \) is given by

\[
r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}.\]

(2.12)

The correlation matrices for \( N(<r) \) and \( D_2(r) \) are shown in Figures 2.2 and 2.3 respectively, for the combined regions (see next section) in the \( 0.5 < z < 0.7 \) redshift slice.

It is noticeable that the \( N(<r) \) measurement is more correlated than \( D_2(r) \).
This is because $N(<r)$ is effectively an integral of $D_2(r)$, so its covariance is effectively a cumulative sum of that of $D_2(r)$. It can also be explained by the fact that $D_2(r)$ is the logarithmic slope of $N(<r)$, so it does not depend on the correlations between widely separated $N(<r)$ bins but rather the variations between neighbouring bins.

We note that the uncertainties calculated using lognormal realisations assume $\Lambda$CDM, and represent the variance we would expect to measure in a $\Lambda$CDM universe. It would not be possible to calculate uncertainties for a fractal universe, since a fractal has no defined cosmic variance as it has no defined mean density. However, calculating uncertainties this way gives a valid consistency check of $\Lambda$CDM.

### 2.4.4 Combining WiggleZ regions

In each of the 4 redshift slices, we combine the measurements from each of the 7 WiggleZ regions using inverse-variance weighting. The combined measurements $x_{\text{comb}}$ are given by

$$x_{\text{comb}}(i) = \frac{\sum_n w_n(i) x_n(i)}{\sum_n w_n(i)},$$

(2.13)

where $n$ are the 7 WiggleZ regions, $x_n(i)$ is the measurement in the $i^{\text{th}}$ radial bin in the $n^{\text{th}}$ region, and $w_n(r) = 1/C_n(i,i)$ is a weight function. $C_n(i,j)$ is the covariance matrix of the $n^{\text{th}}$ region, and the combined covariance matrix is calculated by

$$C_{\text{comb}}(i,j) = \frac{\sum_n C_n(i,j) w_n(i) w_n(j)}{\sum_n w_n(i) \sum_n w_n(j)},$$

(2.14)

We show $C_{\text{comb}}(i,j)$ for the $N(<r)$ and $D_2(r)$ measurements in the $0.5 < z < 0.7$ redshift slice, in terms of the correlation matrix $C_{ij}/\sqrt{C_{ii}C_{jj}}$, in Figures 2.2 and 2.3.

### 2.4.5 Likelihood ($\chi^2$) & parameter fitting

The $\chi^2$ of our model fit to the data is given by

$$\chi^2 \equiv \sum_{i=1}^{\text{Rhino}} \sum_{j=1}^{\text{Rhino}} [x_{\text{th}}(r_i) - x_{\text{obs}}(r_i)] C_{ij}^{-1} [x_{\text{th}}(r_j) - x_{\text{obs}}(r_j)],$$

(2.15)
2.4. METHODOLOGY

Figure 2.2: The correlation matrix $C_{ij}/\sqrt{C_{ii}C_{jj}}$ for the $N(r)$ measurement, obtained from lognormal realisations, for the $0.5 < z < 0.7$ redshift slice.

Figure 2.3: The correlation matrix $C_{ij}/\sqrt{C_{ii}C_{jj}}$ for the $D_2(r)$ measurement, obtained from lognormal realisations, for the $0.5 < z < 0.7$ redshift slice.
where \( x(r) \) is \( \mathcal{N}(<r) \) or \( D_2(r) \), \( n_{\text{bins}} \) is the number of radial bins, \( x_{\text{th}} \) is the theoretical model and \( x_{\text{obs}} \) is the measured value. \( C^{-1} \) is the inverse of the covariance matrix.

### 2.4.6 A new method of defining the ‘homogeneity scale’ \( R_H \)

As previously mentioned, the definition of a ‘homogeneity scale’ is somewhat arbitrary since we expect the Universe to have a gradual approach to homogeneity. Recently Bagla et al. (2008) and Yadav et al. (2010) proposed that the homogeneity scale be defined where the measured fractal dimension \( D_q \) becomes consistent with the ambient dimension within the \( 1\sigma \) statistical uncertainty, \( \sigma_{\Delta D_q} \). They derived an approximation for \( D_q(r) \) and \( \sigma_{\Delta D_q} \), in the limit of weak clustering, for a given correlation function, and showed that both scale the same way, and so the homogeneity scale stays constant, with bias and epoch. This definition is therefore beneficial as it is not arbitrary, and is robust to the tracer galaxy population. However, in deriving \( \sigma_{\Delta D_q} \) they considered only shot noise, and cosmic variance from variance in the correlation function, while ignoring contributions from the survey geometry and selection function. Any real survey will have these contributions to the statistical uncertainty, which cannot be separated from the variance due to the correlation function alone. The value derived from this definition in a real survey is therefore difficult to interpret in a meaningful way, i.e. one that allows comparison with theory, or with different surveys of differing volume and selection function.

We therefore introduce a different method for determining a ‘homogeneity scale’ \( R_H \), which is easier to compare with theory and between surveys. Our method is to fit a smooth, model-independent polynomial to the data, and find the scale at which this intercepts a chosen value, or ‘threshold,’ close to homogeneity. This scale is then defined as the homogeneity scale \( R_H \). For example, \( R_H \) could be the value of \( r \) at which the polynomial intercepts a line 1% from \( \mathcal{N}(<r) = 1 \), or \( D_2(r) = 3 \) (see Figure 2.4 for an illustration). The uncertainty is found using the 100 lognormal realisations. The homogeneity scale measured this way does not depend directly on the survey errors (although the uncertainties on \( R_H \) do), and is less susceptible to noise in the data, making this a preferable method that allows comparisons between different surveys. It also allows easy comparison between the data and a given model, e.g. \( \Lambda \)CDM, and we can check that the data converges to
2.4. METHODOLOGY

Figure 2.4: Illustration of our method of defining the homogeneity scale, $R_H$, shown here for the $D_2(r)$ measurement. We first fit a model-independent polynomial (red curve) to the data (black data points). We then find where this intercepts a chosen value close to homogeneity, e.g. 1% from homogeneity, $D_2 = 2.97$ (dotted grey line). This gives us $R_H$. We find the uncertainty in $R_H$ from the root mean square variance of 100 lognormal realisations (pink curves).

$N = 1$ or $D_2 = 3$ as expected for a homogeneous distribution, by choosing a range of thresholds approaching homogeneity (see Figure 2.9 in Section 2.6.2).

We can also take this further and construct a likelihood distribution for the homogeneity scale, as described in Section 2.6.3.

Although our choice of $D_2$ threshold is arbitrary, by choosing the same threshold for different surveys we obtain an $R_H$ value that can be meaningfully compared, and that can be easily compared to a theoretical model. Our choice of threshold may be limited by the amount of noise in the data, however. For instance, we can measure an intercept 1% away from homogeneity for the WiggleZ data, but cannot measure 0.1% in two of the redshift slices, due to noise (the data do not come this close to homogeneity, although they are consistent with it within the uncertainties) – see Figure 2.9. Nonetheless, we can easily choose a threshold that is possible given our data, and use this to compare with a model. The Baryon Acoustic Oscillation
(BAO) feature can also potentially affect the appropriate choice of intercept value as it causes a small distortion in \( D_2(r) \) – we discuss this in Section 2.8.

For the main results in this paper we have chosen a threshold of 1% away from homogeneity, since this is about the closest threshold to homogeneity we can measure, considering the noisiness of the data.

## 2.5 ΛCDM model prediction of \( \mathcal{N}(<r) \) and \( D_2(r) \)

In this section we derive theoretical ΛCDM predictions for the counts-in-spheres and correlation dimension. This allows us to compare our measurements of the transition to homogeneity to the predictions of a ΛCDM model that fits WMAP data.

### 2.5.1 \( \mathcal{N}(<r) \) and \( D_2(r) \)

For a particular galaxy population, we can calculate the mean counts-in-spheres \( \mathcal{N}(<r) \) from the 2-point matter correlation function predicted by ΛCDM. The 2-point correlation function \( \xi(r) \) is defined as the excess probability above random of finding two objects in volumes \( dV_1 \) and \( dV_2 \), separated by distance \( r \) (Peebles 1980):

\[
P(r) = \bar{\rho}^2 [1 + \xi(r)] dV_1 dV_2,
\]

where \( \bar{\rho} \) is the mean number density.

The mean number of galaxies surrounding a random galaxy up to distance \( r \) is found by integrating the correlation function

\[
\mathcal{N}(<r) = \bar{\rho} \int_0^r [1 + b^2 \xi(r')] 4\pi r'^2 dr',
\]

where \( b \) is the galaxy bias, relating the clustering of a particular galaxy population to the underlying dark matter distribution. Note that ΛCDM assumes large-scale homogeneity, and indeed we must assume large-scale homogeneity in order for a mean density \( \bar{\rho} \) to be defined.

We obtain our model correlation function by transforming a ΛCDM matter power spectrum \( P_{\delta\delta}(k) \) generated using CAMB (Lewis et al. 2000). Since we make our
measurements in redshift space, we first convert $P_{\delta\delta}(k)$ to the redshift-space galaxy power spectrum $P_g^s(k)$ (described in the next section), then convert this to the redshift-space galaxy correlation function $\xi_g(s)$, where $s$ denotes distance in redshift-space. Since we use the angle-averaged power spectrum (assuming the power spectrum is isotropic), we do not need to integrate the angular part of the $k$-space integral, and so use a spherical Hankel transform rather than a Fourier transform to obtain $\xi_g(s)$:

$$\xi_g(s) = \frac{1}{2\pi^2} \int P_g^s(k) \frac{\sin ks}{ks} k^2 dk. \quad (2.18)$$

To compare with our WiggleZ measurement (Equation 2.4), where we correct for incompleteness, we divide our counts-in-spheres prediction by the number that would be expected for a random distribution, i.e. $\bar{\rho}_g^4 \pi r^3$:

$$N(< r) = \frac{3}{4\pi r^3} \int_0^r [1 + \xi_g(s)] 4\pi s^2 ds. \quad (2.19)$$

We calculate the model $D_2(r)$ by simply applying Equation 2.7 to our model $N(< r)$.

### 2.5.2 Redshift-space distortions and nonlinear velocity damping

Here we describe how we implement redshift-space distortions in our analytical model. In practice, we measure the positions of galaxies in redshift-space, which are affected by redshift-space distortions. These are due to the peculiar velocities of galaxies along the line of sight, which add to the measured redshifts and perturb the inferred galaxy positions. This anisotropic effect creates anisotropy in the observed redshift-space galaxy power spectrum $P_g^s(k, \mu)$, and can be modelled by multiplying (convolving in configuration space) the real-space matter power spectrum by an angle-dependent function $F(k, \mu)$:

$$P_g^s(k, \mu) = b^2 F(k, \mu) P_{\delta\delta}(k). \quad (2.20)$$

There are two forms of redshift-space distortion of relevance to our measurement, which we find are necessary for a good fit to the data:
On large, linear scales ($\gtrsim 20 \, h^{-1} \, \text{Mpc}$) the bulk infall of galaxies towards over-densities creates an enhancement in the observed power spectrum along the line of sight. This can be modelled by the linear Kaiser formula \cite{Kaiser1987},

$$P_g^*(k, \mu) = b^2 \left(1 + \frac{f \mu^2}{b}\right)^2 P_{\delta\delta, L}(k)$$

$$= b^2 (1 + \beta \mu^2)^2 P_{\delta\delta, L}(k),$$

where $P_{\delta\delta, L}(k)$ is the linear matter power spectrum, $f$ is the growth rate of structure and $\beta = f/b$ is the redshift-space distortion parameter.

Note that this formula assumes a perturbed FRW universe with small real-space density perturbations $|\delta(r)| \ll 1$.

On quasilinear scales ($10 \lesssim s \lesssim 20 \, h^{-1} \, \text{Mpc}$) the peculiar velocities resulting from the scale-dependent growth of structure distort the shape of the power spectrum, via a scale-dependent damping effect. A common way of modelling this is the ‘streaming model’ \cite{Peebles1980, Fisher1995, Hatton&Cole1998}, which combines the linear theory Kaiser formula with a velocity streaming term. We choose to use a Lorentzian term, $F = [1 + (k \sigma_p \mu)^2]^{-1}$, for an exponential velocity probability distribution function, since Blake et al. \cite{Blake2011} found this to be a good fit to the WiggleZ power spectrum for $k < 0.1 \, h^{-1} \, \text{Mpc}^{-1}$. This gives us the so-called ‘dispersion model’ \cite{Peacock&Dodds1994} for the full redshift-space power spectrum:

$$P_g^*(k, \mu) = b^2 \frac{(1 + \beta \mu^2)^2}{1 + (k \sigma_p \mu)^2} P_{\delta\delta}(k).$$

Here, $\sigma_p$ is the pairwise velocity dispersion along the line of sight. Both $\sigma_p$ and $\beta$ are parameters that must be fitted to the data. We use the values obtained by Blake et al. \cite{Blake2011} in each redshift slice from fits to the WiggleZ two-dimensional power spectrum – these are listed in Table 2.3.

We note that the streaming model is motivated by virialised motions of particles within haloes, on much smaller scales - the so-called ‘Finger of God’ effect at $\lesssim 2 \, h^{-1} \, \text{Mpc}$. However, it is heuristic in nature and can also describe
Table 2.3: Values of the redshift space distortion parameter \( \beta \), and the pairwise velocity dispersion \( \sigma_p \), used in our modelling of nonlinear redshift-space distortion effects. These values were obtained by Blake et al. (2011a) from fits to the WiggleZ two-dimensional power spectrum.

<table>
<thead>
<tr>
<th>Redshift</th>
<th>( \beta )</th>
<th>( \sigma_p ) [h km/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 &lt; ( z &lt; 0.3 )</td>
<td>0.69</td>
<td>346</td>
</tr>
<tr>
<td>0.3 &lt; ( z &lt; 0.5 )</td>
<td>0.73</td>
<td>275</td>
</tr>
<tr>
<td>0.5 &lt; ( z &lt; 0.7 )</td>
<td>0.60</td>
<td>275</td>
</tr>
<tr>
<td>0.7 &lt; ( z &lt; 0.9 )</td>
<td>0.51</td>
<td>86</td>
</tr>
</tbody>
</table>

physical scales of tens of \( h^{-1} \) Mpc. Blake et al. (2011a) apply it this way by fitting for \( \sigma_p \) on these scales, rather than on Finger-of-God scales. We find that including it gives a significant improvement of our model fit to data.

To obtain the angle-averaged redshift-space galaxy power spectrum \( P_g^s(k) \) we need to convert the full \( F(k, \mu) \) to \( F(k) \), which we do by integrating over \( \mu \):

\[
F(k) = \int_{\mu=0}^{1} \frac{(1 + \beta \mu^2)^2}{1 + (k \sigma_v \mu)^2} d\mu. \tag{2.23}
\]

The angle-averaged redshift-space galaxy power spectrum is then

\[
P_g^s(k) = b^2 F(k) P_{\delta\delta}(k). \tag{2.24}
\]

Equation 2.22 normally assumes a linear matter power spectrum; however, we choose to use a nonlinear \( P_{\delta\delta}(k) \), calculated by CAMB using the HALOFIT code (Smith et al., 2003), since Blake et al. (2011a) found this gave a better fit to the data.

### 2.5.3 Correcting the WiggleZ data for galaxy bias

The amplitude of the galaxy correlation function is affected by galaxy bias and redshift-space distortions, and the shape is affected by nonlinear damping. Therefore, these also affect the amplitude and shape of \( N(< r) \), as well as the measured scale of homogeneity. It is possible to correct the data for bias, and so determine the
homogeneity scale for the underlying matter distribution, by assuming a particular model, i.e. our ΛCDM + WMAP model, fitting for bias by minimising the $\chi^2$ value, and correcting the data for this.

In our analysis, we only consider linear galaxy bias. This relates the galaxy correlation function $\xi_g(r)$ to the matter correlation function $\xi_m(r)$ through $\xi_g(r) = b^2 \xi_m(r)$. Since we are only interested in large scales, we do not consider scale-dependent bias that may occur on small scales.

We fix the redshift space distortion parameter $\beta$ and velocity dispersion $\sigma_p$ to the values listed in Table 2.3. We then obtain our corrected measurements $\mathcal{N}_{\text{biasfree}}(< r)$,

$$
\mathcal{N}_{\text{biasfree}}(< r) = \frac{\mathcal{N}(< r) - 1}{b^2} + 1,
$$

and we calculate the bias-corrected correlation dimension $D_{2,\text{biasfree}}(r)$ from this, using Equation 2.7.

Since we assume a ΛCDM model to fit for bias, we cannot make a model-independent measurement of transition to homogeneity of the underlying matter distribution. However, it is still interesting to look at the variation of the homogeneity scale of the matter distribution with redshift, assuming ΛCDM + WMAP.

We note that our measurement of the homogeneity scale of the WiggleZ galaxies is independent of the ΛCDM modelling shown in this section. This modelling is only done so that we can show that the measurement is consistent with that expected from ΛCDM.

2.6 Results

2.6.1 $\mathcal{N}(< r)$ and $D_2(r)$

The $\mathcal{N}(< r)$ measurements in each of the four redshift slices are shown in Figure 2.5. The data are compared with a ΛCDM+WMAP model (described in Section 2.5). For each successive redshift slice, the reduced $\chi^2$ is $[0.57, 0.91, 0.69, 1.1]$. The first two redshift slices have 14 data bins from 12.5 to 251 $h^{-1}$ Mpc, while the last two have 15 data bins from 12.5 to 316 $h^{-1}$ Mpc. The data is consistent with a monotonically
decreasing function, so we can fit a polynomial and find where this intercepts a chosen threshold, as per our definition of homogeneity.

The intercepts of the polynomial fit with $N = 1.01$ (1% away from homogeneity), which we define as the homogeneity scale $R_H$, are shown as red error bars. The errors were determined from the 100 lognormal realisations, and correspond to the square-roots of the diagonal elements of their covariance matrix. The $R_H$ values and their errors are shown in Figure 2.5 and listed in Table 2.4 along with the values for the ΛCDM model, which are in good agreement.

The $D_2(r)$ measurements in each of the redshift slices are shown in Figure 2.6 along with a ΛCDM+WMAP model with best-fitting bias. In each redshift slice the data range and degrees of freedom are the same as for the $N(<r)$ measurement. The reduced $\chi^2$ values in each redshift slice are $[0.83, 0.90, 0.74$ and $0.98]$. In each case a polynomial is fitted to the data. The homogeneity scales measured where these intercept 1% away from homogeneity, $D_2 = 2.97$, are listed in Table 2.4 and are also in excellent agreement with the ΛCDM values.

### 2.6.2 Effect of bias and $\sigma_8(z)$ on $R_H$

The homogeneity scale of the model galaxy distribution, measured at 1% from homogeneity, will depend on the amplitude and shape of the correlation function, and so on galaxy bias $b$ and $\sigma_8(z)$. We would also expect it to depend on redshift, since $\sigma_8(z)$ increases over time. These two parameters are in fact completely degenerate in the $N(<r)$ and $D_2(r)$ measurements.

Figure 2.7 shows how our ΛCDM $N(<r)$ and $D_2(r)$ models vary with bias, at $z = 0.2$, for fixed $\sigma_8(z = 0) = 0.8$. Larger bias means a larger amplitude of clustering, so that both curves are steeper on small scales. This means that the models reach homogeneity at larger radii for higher bias. This can be understood qualitatively, since highly biased galaxies are more clustered together than less biased galaxies, so we must go to larger scales before we reach a homogeneous distribution.

Figure 2.8 shows how the homogeneity scale $R_H$ of the galaxy distribution varies with bias, for different intercept values approaching homogeneity, for the $N(<r)$ and $D_2(r)$ models at $z = 0.2$ with fixed $\sigma_8(z = 0) = 0.8$. For a particular intercept value, larger bias again gives a larger homogeneity scale. Since $N(<r)$ and $D_2(r)$
Figure 2.5: Scaled counts-in-spheres $N(<r)$ for the combined WiggleZ data in each of the four redshift slices (black error bars). A $\Lambda$CDM model with best-fitting bias $b^2$ is shown in blue. A 5$^{th}$-degree polynomial fit to the data is shown in red. The red errorbar and label show the homogeneity scale $R_H$ for the galaxy distribution, measured by the intercept of the polynomial fit with 1.01 (1% away from homogeneity), with the error given by lognormal realisations. This scale is consistent with the $\Lambda$CDM intercept with 1.01, labelled in blue.
### Table 2.4: Measured values of the homogeneity scale $R_H$ (where data intercepts 1% of the homogeneity value).

<table>
<thead>
<tr>
<th>$D = 0.1$</th>
<th>$D = 0.3$</th>
<th>$D = 0.5$</th>
<th>$D = 0.7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z &lt; 0.3$</td>
<td>72 ± 43</td>
<td>78 ± 44</td>
<td>74 ± 44</td>
</tr>
<tr>
<td>$z &lt; 0.5$</td>
<td>72 ± 42</td>
<td>78 ± 42</td>
<td>74 ± 42</td>
</tr>
<tr>
<td>$z &lt; 0.7$</td>
<td>72 ± 41</td>
<td>78 ± 41</td>
<td>74 ± 41</td>
</tr>
<tr>
<td>$z &lt; 0.9$</td>
<td>72 ± 40</td>
<td>78 ± 40</td>
<td>74 ± 40</td>
</tr>
</tbody>
</table>

$N_{\delta}$ values shown are all for the galaxy distribution, except for the bias-corrected data, which are for the underlying matter distribution. $\delta = 1, 2, 3$ values directly assume a $\Lambda$CDM+WMAP model.

Note: $R_H$ values shown are all for the galaxy distribution, except for the bias-corrected data, which are for the underlying matter distribution. $\delta = 1, 2, 3$ values directly assume a $\Lambda$CDM+WMAP model.
Figure 2.6: Same as for Figure 2.5 but for the correlation dimension $D_2(r)$. The $D_2(r)$ measurements for the combined WiggleZ data in each of the four redshift slices are shown as black error bars. A ΛCDM model with best-fitting bias $b_2$ is shown in blue. A 5th-degree polynomial fit to the data is shown in red. The red errorbar and label show the homogeneity scale $R_H$ measured by the intercept of the polynomial fit with 2.97 (1% away from homogeneity), with the error given by lognormal realisations. This scale is consistent with the ΛCDM intercept with 2.97, labelled in blue.
2.6. RESULTS

Figure 2.7: The effect of bias on a ΛCDM \( N(<r) \) model (left) and \( D_2(r) \) model (right) at \( z = 0.2 \). Increasing bias increases the value of \( N(<r) \) on small scales, and decreases the value of \( D_2(r) \) on small scales, and produces a larger homogeneity scale, as seen by the intercepts of the curves with 1% of homogeneity (\( N = 1.01 \) and \( D_2 = 2.97 \), red dotted lines).

approach homogeneity asymptotically, the intercept jumps to higher values when we consider an intercept value closer to homogeneity. This image also shows the mapping between \( N(<r) \) and \( D_2(r) \) homogeneity values for intercepts at 1%, 0.1% and 0.001% away from homogeneity. They are not identical since they are slightly different methods, but give similar results. This plot illustrates that there are many potential ways to define homogeneity, and so it is important to make consistent measurements between surveys in order for them to be comparable with each other and with theory.

Our bias-corrected \( N(<r) \) and \( D_2(r) \) measurements are listed in Table 2.4. The errors on the bias-corrected data were determined by applying a bias correction to each of the 100 lognormal realisations individually, and recalculating the covariance matrix. Since the bias correction aims to set the bias of all the realisations to \( b = 1 \), it lowers the overall variance, and so the error bars are slightly smaller than for the pure data. These measurements give our measured homogeneity scale for the matter distribution, assuming ΛCDM. We find that this scale increases with redshift, as expected in ΛCDM. However, since we are assuming ΛCDM, which has \( \sigma_8(z) \) increasing with time, this is not a model-independent result.
As already mentioned, the effect of bias on the homogeneity scale of galaxies is degenerate with the amplitude of the correlation function, $\sigma_8(z)$, since the correlation function at redshift $z$ depends on a combination of these, $\xi(r, z) \propto b^2 \sigma_8(z)^2$. So far we have assumed a fixed value of $\sigma_8(z = 0)$. But we can also make predictions independent of $\sigma_8$, by finding how $R_H$ changes as a function of the combination $b^2 \sigma_8(z)^2$. This is shown in Figure 2.9. We also show the WiggleZ results, which we have plotted for the best-fit $b^2 \sigma_8(z)^2$ value in each redshift slice. This increases with redshift, so the data points from left to right go from low to high redshift. The WiggleZ results are in very good agreement with the $\Lambda$CDM+WMAP predictions. We show, for comparison, the values obtained from defining the homogeneity scale as where the data comes within 1σ of homogeneity. These values have much greater stochasticity than those from our method of fitting a smooth curve to many data points, and do not give informative results in this plane.

We see that the model $R_H - b^2 \sigma_8(z)^2$ curves are monotonically increasing. Since we expect $\sigma_8(z)$ in $\Lambda$CDM to grow over time due to growth of structure, we would therefore also expect the homogeneity scale to increase over time, for galaxies with fixed bias.

For the WiggleZ data, however, the measured homogeneity scale does not appear to decrease with redshift. This is explained by the fact that the WiggleZ galaxies have increasing bias with redshift, assuming a $\Lambda$CDM growth rate. As explained previously, this is understood to be due to the effects of Malmquist bias and downsizing on the selection of the WiggleZ galaxy population, and the colour and magnitude cuts. This counteracts the effect of decreasing $\sigma_8$ with redshift. As can be seen in Figure 2.9 within the $b^2 \sigma_8(z)^2$ range of the data we would not expect a significant change in $R_H$, measured at 1% from homogeneity, with redshift.

### 2.6.3 Likelihood analysis for homogeneity scale

Rather than trying to measure the scale of homogeneity directly, it can be informative to consider the likelihood that the Universe has reached homogeneity by a certain scale. We can construct a probability distribution function (PDF) for the homogeneity scale, by combining the likelihood distribution of the data with that of the definition of the homogeneity scale. This gives the probability that homogeneity
2.6. **RESULTS**

![Figure 2.8](image)

Figure 2.8: $R_H$ for ΛCDM $\mathcal{N}(r)$ (green) and $D_2(r)$ (purple) models at $z = 0.2$ with differing bias $b^2$. Each curve corresponds to $R_H$ evaluated at a different threshold – 1%, 0.1% or 0.01% away from homogeneity (from bottom to top, labelled).

is reached at a certain scale $r$.

We could apply a likelihood analysis to either our $\mathcal{N}(< r)$ or $D_2(r)$ measurements. However, it would arguably be invalid for highly correlated data, such as the $\mathcal{N}(< r)$ measurement, since the different contributions to the probability distribution would be correlated, and we do not correct for this. The $D_2(r)$ measurement is much less correlated (see Figures 2.2 and 2.3). Also, as we have explained, $D_2(r)$ is the most robust measurement of homogeneity, so we have chosen this for our likelihood analysis.

The likelihood distribution on the data is simply given by the mean and variance of the lognormal realisations in each bin. These give the expected variance of a ΛCDM distribution sampled with the WiggleZ selection function, so take into account both cosmic variance and shot noise. Here we assume they are Gaussian-distributed by virtue of the central limit theorem. However we also consider the true distributions provided by the lognormal realisations – this gives similar results but with larger errors, see Appendix A.
Figure 2.9: Homogeneity scale $R_H$ as a function of $b^2\sigma_8(z)^2$, as predicted by the $\Lambda$CDM model, for different thresholds approaching homogeneity (10%, 1% and 0.1% from homogeneity, coloured curves from bottom to top), for $N(<r)$ (top) and $D_2(r)$ (bottom). The corresponding WiggleZ results are shown as error bars of corresponding colour (the errors are found using lognormal realisations). The $b^2\sigma_8(z)^2$ values of the data increase with redshift slice (so the data points from left to right are from low to high redshift). Not all redshift slices have measurements at 0.1% (blue) since the data do not reach this value in those slices. The definition of homogeneity used by previous authors is where the data comes within 1$\sigma$ of homogeneity – we show these scales as black diamonds. It is clear that this definition has much greater stochasticity than our definition, which fits a smooth curve to many data points. Indeed, this approach gives quite uninformative results in this plane and cannot be compared to the model prediction.
2.6. RESULTS

At each \( r \), the data therefore provide a probability distribution \( p_{D}[D_2(r)] \), which we model as Gaussians, as:

\[
p_{D}[D_2(r)] = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(D_2(r) - \mu)^2}{2\sigma^2}},
\]

where \( \mu \) is the WiggleZ value of \( D_2 \) at radius \( r \) and \( \sigma \) is the root mean square variance given by the lognormal realisations. See Figure 2.10 for an illustration.

We also expect there to be a likelihood distribution on the \( D_2(r) \) value we would measure for a perfectly homogeneous distribution, due to cosmic variance and shot noise caused by our selection function. We can represent this by a likelihood distribution on the homogeneity scale – this would not be a simple delta-function at \( D_2(r) = 3 \), but would have some spread. This distribution should be one-sided – that is, we don’t expect to measure \( D_2(r) > 3 \), only \( D_2(r) \leq 3 \), if we have a distribution that approaches homogeneity.\(^2\) We might expect it to be represented by the variance in the random catalogues, which are essentially homogeneous distributions sampled with the WiggleZ selection function. However, the same sources of noise are also present in the lognormal realisations, so we would potentially double-count errors if we also used lognormal realisations to determine the likelihood distribution of the data. This means we cannot easily determine the true variance in the value of \( D_2(r) \) we would expect to measure for a homogeneous distribution, independently of the variance of the data.

For this reason, we choose to find the likelihood distribution of the data reaching 1% of homogeneity. That is, we assume the likelihood distribution on the homogeneity scale, \( p_{H}[D_2(r)] \), is a delta-function at \( D_2 = 2.97 \):

\[
p_{H}[D_2(r)] = \delta[D_2(r) - 2.97].
\]

We can then construct the cumulative probability distribution function \( P(R_H \leq r) \), which gives the probability that the homogeneity scale has been reached at or before scale \( r \), from:

\(^2\)In some cases the WiggleZ data does fall below \( N = 1 \) or above \( D_2 = 3 \); this can be explained as the effect of shot noise introduced by the selection function rather than a physical effect, as shown in our comparison with the GiggleZ simulation in Section 2.7.2.
Figure 2.10: The probability distributions $p_D[D_2(r)]$ for each of the $r$ bins in the $0.5 < z < 0.7$ redshift slice. Blue-to-red gradient indicates small to large radius, from 12.5 to 316 $h^{-1}$ Mpc. The area of each distribution above $D_2(r) = 2.97$ (dotted line) gives the probability that homogeneity has been reached within 1% by that value of $r$.

\[
P(R_H \leq r) = \int_{-\infty}^{\infty} p_D[D_2(r)] \left( \int_{-\infty}^{D_2(r)} p_H(x)dx \right) dD_2(r) = \int_{2.97}^{\infty} p_D[D_2(r)]dD_2(r). \tag{2.28}
\]

That is, the probability of having reached homogeneity is the area of the likelihood distribution of the data that falls at or above $D_2 = 2.97$.

This cumulative probability is calculated at each scale $r$. We can then find the PDF for the homogeneity scale, $p(R_H)$, from

\[
p(R_H) = \frac{dP(R_H \leq r)}{dr}. \tag{2.29}
\]

The PDFs for the homogeneity scale, for WiggleZ galaxies in each redshift slice,
2.7 ROBUSTNESS OF HOMOGENEITY MEASUREMENT

Figure 2.11: Probability distributions for the scale of homogeneity, \( p(R_H) \), for WiggleZ galaxies in each of the four redshift slices. The homogeneity scale is defined as the scale where the data reaches values 1% away from \( D_2 = 3 \), i.e. \( D_2 = 2.97 \). These are shown in Figure 2.11. We have interpolated between the data bins in order to obtain smoother PDFs. We find the most probable \( R_H \) values from the mean of the distributions. These are all between 70 and 81 \( h^{-1} \) Mpc, and are listed in Table 2.4. They represent the most probable scale at which the galaxy distribution reaches 1% of homogeneity. We also list the values found for the bias-corrected data, which give the most probable homogeneity scales for the matter distribution, assuming \( \Lambda CDM+WMAP \).

2.7 Robustness of homogeneity measurement

In this section we address several issues that could potentially influence our measurement of the homogeneity scale, and perform tests to ensure the robustness of our results. We base our tests on the 15-hr \( 0.5 < z < 0.7 \) region, which is the largest and most populated WiggleZ sub-region, but the results are applicable to the entire survey.
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK ENERGY SURVEY

2.7.1 Fractal model test of selection function and boundary effects

A major potential source of bias in our results is the method used to correct for edge effects and the selection function of the survey. We have used 100 random catalogues to correct each individual WiggleZ measurement, as described in Section 2.4.1. However, this method can potentially bias homogeneity measurements, since it weights measurements by the volume of spheres of radius $r$ included in the survey, and so assumes a homogeneous distribution outside the survey (e.g. Coleman & Pietronero 1992, Sylos Labini et al. 2009). It is therefore important to check that this is not imposing any distortion in our measured correlation dimension, so producing a ‘false relaxation’ to homogeneity.

To test this we apply our correction method to a range of fractal distributions of known correlation dimension. This has been done previously by a number of works, e.g. Lemson & Sanders (1991), Provenzale et al. (1994) and Pan & Coles (2002). This allows us to check that our method returns the correct input correlation dimension up to the largest scales we measure in the survey, and to quantify any distortion that may occur.

We generate our fractal distributions using the $\beta$-model, a simple self-similar cascading model (see e.g. Castagnoli & Provenzale 1991). This method starts with a cube of side $L_0$ and splits it up into $M$ smaller cubes of side $L_0/n$ (we take $n = 2$, so $M = 8$). Each subcube is then assigned a probability $p$ of surviving to the next iteration. This is repeated for a certain number of iterations $k$, and the resultant set of survived points is taken as the final distribution. In the limit of an infinite number of iterations, this produces a monofractal with a correlation dimension given by

$$D_2 = \lim_{k \to \infty} \frac{\log (p M)^k}{\log n^k} = \frac{\log p M}{\log n}. \tag{2.30}$$

Our procedure for using $\beta$-models to test our analysis method is as follows:

1. We choose a range of $D_2$ values (2.7, 2.8, 2.9, 2.95 and 2.97), and for each we generate 100 fractal galaxy distributions, with a boxsize of $(L_0 h^{-1} \text{Gpc})^3$, where $L_0$ is the length of the longest side of the WiggleZ 15-hr 0.5 $< z <$ 0.7 selection function grid (623.5 $h^{-1} \text{Mpc}$).
2.7. ROBUSTNESS OF HOMOGENEITY MEASUREMENT

2. We then sample each distribution with the WiggleZ selection function for the 15-hr \(0.5 < z < 0.7\) region. We normalise the resulting distribution to give the same number of points as WiggleZ galaxies in this region. This gives us fractal mock catalogues, and we then measure \(D_2\) for these in the same way as for the WiggleZ data, correcting the counts-in-spheres measurements with random catalogues. This gives a result that is influenced by both the WiggleZ selection function, and our correction method.

Figure 2.12 shows the mean \(N(<r)\) and \(D_2(r)\) results for the different fractal distributions up to \(D_2 = 2.95\), with the WiggleZ selection function and correction method applied. Even up to \(D_2 = 2.95\) they are clearly distinguishable from ΛCDM and the WiggleZ data. The \(D_2\) values measured are consistent with the input values up to at least 200 \(h^{-1}\) Mpc, well above the homogeneity scales measured for the data. These results indicate that the WiggleZ selection function and correction method do not have a significant effect on the measured correlation dimension up to the scales we measure for homogeneity.

It is noticeable that the size of the error bars gets smaller for models with larger \(D_2\). This is a real effect in the model; for larger \(D_2\), more boxes survive at each iteration, so there are a smaller number of possible configurations for the final distribution, resulting in lower variance for a box of a particular volume.

To quantify how well we can exclude fractal models, we fit a line of constant \(D_2\) to each set of fractal data, over the range \([80, 300] h^{-1}\) Mpc (shown in Figure 2.12). This gives us the best-fit \(D_2\) value we would expect to measure for each fractal distribution over this range, taking into account bias from the selection function. We then find the formal probability of these values fitting the WiggleZ data. Doing this, we find we can exclude a fractal dimension of \(D_2(r) = [2.9, 2.95, 2.97]\) at the [19,6,4]-\(\sigma\) level. In other words, we can exclude fractal distributions with dimension \(D_2(r) < 2.97\) at over 99.99% confidence on scales from 80 to 300 \(h^{-1}\) Mpc.

Our results agree with those of Lemson & Sanders (1991), Provenzale et al. (1994) and Pan & Coles (2002) who also find that for samples on scales larger than the homogeneity scale, boundary corrections do not have a significant effect on the analysis. A further check would be to test different types of fractal model other than the β-model, but we leave this for future work. We also note that we still assume
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK ENERGY SURVEY

Figure 2.12: Fractal model comparisons of the measured correlation dimension from the WiggleZ 15-hr region, $0.5 < z < 0.7$ redshift slice. The WiggleZ data (black error bars) and ΛCDM model (blue curve) are compared with several different β-models with different fractal dimension ($D_2 = 2.7$, 2.8, 2.9 and 2.95), which have been sampled with the 15-hr selection function and analysed in the same way as the WiggleZ data (coloured error bars). The uncertainties shown are the error-in-the-mean of 100 fractal realisations. The input fractal dimensions are shown as dotted lines with corresponding colours. The best-fit $D_2(r)$ values fit over the range $r = [80, 300] h^{-1}$ Mpc are shown as solid lines.

an FRW metric in our fractal analysis; an improved consistency check would be to calculate the actual metric in these fractal models (see Section 2.8), but this is beyond the scope of this paper.

2.7.2 Comparison with the GiggleZ $N$-body simulation

We have also compared our results with a ΛCDM cosmological $N$-body simulation. This allows us to check that our analytical ΛCDM+WMAP model (incorporating Kaiser and streaming models for redshift-space distortions) is consistent, over the relevant scales, with a full simulation including nonlinear effects. It also provides a further test of selection function effects, since we can show that homogeneity measurements of a ΛCDM distribution are not distorted when the WiggleZ selection function is applied.

The GiggleZ (Giga-parsec WiggleZ) simulation [Poole et al. 2014] is a suite
2.7. ROBUSTNESS OF HOMOGENEITY MEASUREMENT

of dark matter $N$-body simulations run at Swinburne University of Technology, designed for theoretical analyses of the WiggleZ dataset. It was run using a modified version of the $N$-body code GADGET-2 (Springel et al., 2001) using a WMAP-5 cosmology. We use the main simulation, which has a volume of $(1000 \, h^{-1} \, \text{Mpc})^3$ and $2160^3$ particles of mass $7.5 \times 10^9 h^{-1} M_\odot$.

Halo finding for GiggleZ was performed using Subfind (Springel et al., 2001), which utilises a friends-of-friends (FoF) algorithm to identify coherent overdensities of particles and a substructure analysis to determine bound overdensities within each FoF halo. For our analysis, we rank-order the resulting Subfind substructure catalogues by their maximum circular velocity ($V_{\text{max}}$) and select a contiguous subset of 250 000 halos (selected to yield a number density comparable to the survey) over a range of $V_{\text{max}}$ chosen to yield a bias comparable to that of WiggleZ. We use a catalogue at a redshift of $z = 0.593$, corresponding to the mid redshift of the WiggleZ $0.5 < z < 0.7$ redshift slice.

We add the effect of redshift-space distortions by shifting the positions of the haloes according to their line-of-sight peculiar velocities. That is, the comoving position of each halo relative to an observer, $x$, is shifted by a vector $\Delta x$,

$$\Delta x = \frac{x}{|x|} \frac{(1 + z)}{H(z)} v_{\text{rad}},$$

where $v_{\text{rad}} = (x \cdot v)/|x|$ is the radial velocity of the halo along the observer’s line-of-sight, and we place the observer at the same coordinates relative to the GiggleZ box as for the WiggleZ selection function grid.

We then calculate $N(<r)$ and $D_2(r)$ using two different methods:

1. Using the full GiggleZ box. We correct the measurement using a random distribution within the same volume box, with 100 times the number of galaxies as the GiggleZ sample.

2. Applying the WiggleZ 15-hr $0.5 < z < 0.7$ selection function to GiggleZ. This creates a mock WiggleZ survey containing 10 830 galaxies. We correct the measurement using the random catalogues used for the WiggleZ data. This allows us to see the effects induced purely by the selection function.

The results are shown in Figure 2.13. The full GiggleZ dataset is very con-
consistent with both the WiggleZ data and the ΛCDM model. The consistency with the model indicates that the implementation of redshift-space distortions in the model, described in Section 2.5.2, has a good level of accuracy, to scales as small as \( \sim 20 \, h^{-1} \, \text{Mpc} \). The deviation of the smallest-scale bin is a resolution effect, since the GiggleZ catalogue is sparser than WiggleZ (with 10,830 galaxies vs. 17,928 WiggleZ galaxies in the same region and redshift slice). The GiggleZ results both with, and without, the WiggleZ selection function are also consistent within the size of the WiggleZ error bars, showing that the selection function and correction method do not have a significant effect. It can be seen that the selection function causes a few data points in the \( \mathcal{N}(<r) \) plot to go below 1, and in the \( D_2(r) \) plot to go above 3. This shows that adding shot noise can produce this apparent unphysical effect, explaining why this is also seen in some of the WiggleZ results (Figures 2.5 and 2.6).
2.7. ROBUSTNESS OF HOMOGENEITY MEASUREMENT

2.7.3 Comparison of different correction methods

To further demonstrate the robustness of our correction method, we illustrate the results obtained for two alternative correction methods – firstly, using only complete spheres for the counts-in-spheres measurements (the ‘exclusion’ method mentioned in Section 2.4.1), but still correcting for the selection function, and secondly, using no correction for the selection function. By using the exclusion method we do not have to deal with survey edge effects; however, since WiggleZ is not volume-limited, it is still necessary to use random catalogues to correct for the selection function. That is, we recalculate Equation 2.4 but with \( G \) equal to the number of WiggleZ galaxies at the centre of complete spheres of radius \( r \). We show this result in Figure 2.14 for the 15-hr \( 0.5 < z < 0.7 \) region, as red error bars, where the uncertainty is calculated using our lognormal realisations. We compare this to the result using our correction method (black error bars). The red error bars are consistent with the black error bars, but show more scatter and have higher noise, which increases for larger radius, reflecting the lower statistics where fewer spheres contribute to the measurement. The measurements must also be cut off at a lower radius, since not enough larger spheres fit inside the survey region.

We can also compare the result without any selection function correction, which illustrates the necessity of correcting for having a non-volume-limited sample. However, WiggleZ contains holes in the angular coverage, which are independent of assumptions about completion, so we must still take these into account. We therefore normalise each \( N_i(r) \) measurement by the volume within the selection function included in that sphere. So we calculate:

\[
N_{\text{no-corr}}(r) = \frac{1}{G_{\text{complete}}(r)} \sum_{\text{complete spheres } i} \frac{N_i(r)}{V_{\text{SF},i}} \times \bar{\rho},
\]

where \( G_{\text{complete}}(r) \) is the number of WiggleZ galaxies at the centre of complete spheres of radius \( r \), \( V_{\text{SF},i} \) is the volume within the selection function of the \( i \)th galaxy, and \( \bar{\rho} = n_W/V_{\text{SF, tot}} \) is the mean density of WiggleZ, i.e. the number of WiggleZ galaxies divided by the total volume in the selection function, excluding holes. We show this in Figure 2.14 as purple error bars, where the uncertainty is again calculated using our lognormal realisations (so the uncertainty assumes
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK ENERGY SURVEY

Figure 2.14: Illustration of the result using different correction methods, for the 15-hr $0.5 < z < 0.7$ region. Black error bars show the result using our correction method with random catalogues. The red error bars are obtained when only complete spheres are used in the counts-in-spheres measurement, while still correcting for the selection function. The purple error bars show the result for complete spheres, with no selection function correction. In each case the uncertainties are calculated using lognormal realisations. The red and purple error bars are shifted slightly to the right for clarity.

$\Lambda$CDM). Although there is a vertical offset in the $N(< r)$ plot, caused by the selection function, the $D_2(r)$ results are remarkably similar to those when correcting for the selection function, though again with more noise. There is still a clear transition towards values close to $D_2 = 3$ on large scales. The offset in $N(< r)$ can be attributed to a selection effect: since we use only complete spheres we weight the measurement towards the central part of the redshift range, where the completeness is highest and so the number density of WiggleZ galaxies is highest. Since $D_2(r)$ is the slope of $N(< r)$, it does not depend on the number density itself and so is more robust when summing over a varying selection function.

2.8 Discussion

Our WiggleZ $N(< r)$ and $D_2(r)$ results show a very strong agreement with an FRW-based $\Lambda$CDM model. However, one of the strongest arguments against previous
homogeneity measurements is that the method of correcting for survey selection functions, such as using random catalogues as we do, can distort the data, producing a ‘false relaxation’ to homogeneity. We have therefore tested this by applying the WiggleZ selection function to a range of inhomogeneous fractal models and a ΛCDM N-body simulation. We have shown that there is no significant impact on our homogeneity measurement, up to at least 200 $h^{-1}$ Mpc. In addition, we have compared the results from our correction method to an analysis using only complete spheres, with and without correcting for the selection function, and have shown that these are consistent, further demonstrating the robustness of our result. We can rule out fractals with fractal dimension up to $D_2 = 2.97$, on scales between 80 and 300 $h^{-1}$ Mpc, at over 99.99% confidence.

We can also be confident that our result is robust to any assumptions in modelling the WiggleZ survey selection function, since Blake et al. (2010) showed that even extreme variations in modelling the angular completeness produce only $\sim 0.5\sigma$ shifts in estimates of the power spectrum. Changes in the parameterisation of the redshift distribution were shown to cause larger deviations, but only at scales $> 200 h^{-1}$ Mpc, which are well above the scales on which we measure homogeneity.

Our result is a very good consistency check of ΛCDM. However, it is not independent of the assumption of the FRW metric. A complication for all homogeneity measurements is that we can only observe galaxies on our light cone, and not on spatial surfaces. Maartens (2011) points out that it is therefore not possible to make a homogeneity measurement without making some assumptions, such as the FRW metric (to convert redshifts to distances) and the Cosmological Principle. Indeed, we must always assume an FRW-based model, ΛCDM in our case, to convert redshifts and angles to distance coordinates. This is also problematic considering that inhomogeneities produce perturbations in the FRW metric, so can potentially distort distance measurements by affecting the paths travelled by light rays (e.g. Wiltshire 2009; Meures & Bruni 2011). However, it seems highly unlikely that our measurements would so closely agree with FRW-based ΛCDM (in both the amplitude and shape of the $N(<r)$ and $D_2(r)$ curves), if the distribution were, actually, inhomogeneous up to the largest scales probed, and we had incorrectly assumed an FRW metric. We have tested our method using a fractal model, and shown that this gives a completely different form of these curves. Our assumption of FRW also
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK ENERGY SURVEY

seems reasonable, since we know that distance measurements from Type Ia supernovae fit the Hubble diagram well to first order up to $z \sim 1.4$ (Conley et al., 2011; Davis et al., 2011), so that any perturbations due to inhomogeneities can only be a second- or higher-order effect. All this means that we can take our results to be a strong consistency check of FRW-based ΛCDM.

It would be possible to test this further by making isotropy measurements in thin redshift shells, over a range of redshifts, without converting them to distances. We could also directly test the assumption of FRW, by calculating the non-FRW metrics of our fractal models and testing our analysis on these. We could also calculate the effects of metric perturbations for ΛCDM, including the effects of backreaction, as in e.g. Buchert (2000), Li & Schwarz (2007), Wiltshire (2007b) and Behrend et al. (2008). This would allow us to test the effect of incorrectly assuming an FRW metric. Alternatively, several possible consistency tests of homogeneity and the Copernican Principle that do not assume FRW have been suggested by e.g. Clarkson et al. (2008); Shafieloo & Clarkson (2010); Maartens (2011); Heavens et al. (2011). However, we leave these suggestions for future work.

An alternative way of defining the ‘homogeneity scale’ was recently suggested by Bagla et al. (2008) and Yadav et al. (2010), as where the measured fractal dimension $D_q$ becomes consistent with the ambient dimension within the 1σ statistical uncertainty, $\sigma_{\Delta D_q}$. They make predictions for a ΛCDM model, by deriving an approximation for $D_q$ and $\sigma_{\Delta D_q}$, given a particular correlation function, and showing how these scale with sphere size. Using this, they predict that the ‘true’ homogeneity scale in ΛCDM is $260 h^{-1}$ Mpc. They also predict that $D_q$ and $\sigma_{\Delta D_q}$ scale the same way with the correlation function, and so their definition of homogeneity does not change with bias or redshift. Their definition is therefore beneficial, as it is robust to the tracer galaxy population. It is also not arbitrary, and indeed the scale above which the fractal dimension is consistent with homogeneity within cosmic variance is arguably a physically meaningful scale to define as homogeneous, since above this scale the distribution cannot be distinguished from a homogeneous one. However, as we have pointed out (Section 2.4.6), it is difficult to apply their definition to a real measurement, since their approximation for $\sigma_{\Delta D_q}$ only accounts for the variance of the correlation function (in the limit of weak clustering) and shot noise, but ignores errors due to survey geometry and the selection function.
2.8. DISCUSSION

These additional contributions mean that real errors will always be larger, and so will always measure a smaller homogeneity scale than the ‘true’ one. Since these errors will be different for different surveys, and cannot be separated out from the variance in the underlying correlation function, homogeneity measurements made in this way cannot be easily compared between different surveys, or with theory. The benefit of our method of defining the homogeneity scale, even if we have to make an arbitrary choice about the value of $\Delta D_q$ we accept for homogeneity, is that it can be used to easily compare the results from different surveys and with a theoretical model.

A homogeneity scale below $100\, h^{-1}\, \text{Mpc}$ may also seem to contradict the fact that the correlation function has a known feature, the Baryon Acoustic Oscillation (BAO) peak at $\sim 105\, h^{-1}\, \text{Mpc}$ (Eisenstein et al., 2005; Percival et al., 2010; Beutler et al., 2011; Blake et al., 2011c). However, the BAO peak has only a small impact on the counts-in-spheres statistic. It is visible in our $\Lambda$CDM prediction for $D_2(r)$ (which is more sensitive to it than $N(< r)$, since it is a differential measurement), as a small dip at just over $100\, h^{-1}\, \text{Mpc}$ (see Figure 2.7). It means that the $D_2(r)$ curve does not increase monotonically around this scale, so we must be careful if attempting to measure an intercept with homogeneity that lies close to this. The magnitude of the distortion due to the BAO peak, $\Delta D_2$, is of order $\sim 0.01$ for $b \sim 1$ so does not affect our homogeneity scale measured at $D_2 = 2.97$. However, it is more significant for more highly-biased tracers (as also pointed out by Bagla et al., 2008), and for a highly-biased population, such as LRGs, it may be necessary to measure the homogeneity scale at a $D_2$ value closer to 3, to avoid this region.

It has been pointed out (e.g. Sylos Labini & Pietronero, 2010) that measurements of large structures in galaxy surveys are seemingly at odds with a homogeneity scale below $100\, h^{-1}\, \text{Mpc}$. Previous surveys have detected structures on scales much larger than this (Geller & Huchra, 1989; de Lapparent et al., 1989). The largest observed structure in the Universe, the Sloan Great Wall, is $320\, h^{-1}\, \text{Mpc}$ long (Gott et al., 2005), appears inconsistent with the existence of a homogeneity scale below $100\, h^{-1}\, \text{Mpc}$. However, this is not incompatible with our results, since we show the scale where the data has $D_2 \geq 2.97$, and it is not impossible to have fluctuations on scales larger than this. Also, these large structures, including the Sloan Great Wall, are usually filamentary, whereas we have measured a volume statistic which
2. LARGE-SCALE COSMIC HOMOGENEITY IN THE WIGGLEZ DARK
ENERGY SURVEY

averages over fluctuations.

2.9 Conclusion

We have measured the large-scale transition to homogeneity in the distribution of galaxies, using the WiggleZ survey, in four redshift bins between $z = 0.1$ and $z = 0.9$. We measured the mean, scaled counts-in-spheres $N(< r)$ and the correlation dimension, $D_2(r)$, and found these to be in excellent agreement with a ΛCDM model with WMAP parameters, including redshift space distortions. We also presented a new, model-independent method for determining the homogeneity scale $R_H$ from data. This involves fitting a polynomial curve to the data, and finding where this intercepts chosen values close to homogeneity. This is a more reliable method than finding where the data comes within $1\sigma$ of homogeneity, since it does not depend directly on the size of the error bars and is less susceptible to noise. It also allows a direct comparison between data and theory, and between different surveys of differing bias and redshift.

We summarise our results as follows:

1. Our $N(< r)$ and $D_2(r)$ results show a very strong agreement with a FRW-based ΛCDM-WMAP model incorporating redshift-space distortions. They show a clear transition from an inhomogeneous, clustered distribution on small scales, to a homogeneous one on large scales. This transition matches that of the ΛCDM model. We have thereby conducted a very stringent consistency check of ΛCDM.

2. If we define the ‘homogeneity scale’ $R_H$ as the scale where the data become consistent with homogeneity within 1%, then from a likelihood analysis of $D_2(r)$ we measure $R_H$ to be $71 \pm 8 h^{-1}$ Mpc at $z \sim 0.2$, $70 \pm 5 h^{-1}$ Mpc at $z \sim 0.4$, $81 \pm 5 h^{-1}$ Mpc at $z \sim 0.6$, and $75 \pm 4 h^{-1}$ Mpc at $z \sim 0.8$. These values are consistent with those of the ΛCDM-WMAP model with best-fitting bias, of $R_H = [76, 78, 81, 78] h^{-1}$ Mpc.

3. We find that the homogeneity scale of our ΛCDM-WMAP model increases with clustering amplitude $b(z)\sigma_8(z)$. For a population with fixed bias, we
therefore predict the homogeneity scale to grow over time, since \( \sigma_8(z) \) increases due to growth of structure. The bias of the WiggleZ galaxies increases with redshift, so the measured \( R_H \) values do not change with redshift. If we correct our data for bias, assuming \( \Lambda \)CDM, then we measure a homogeneity scale for the matter distribution that increases over time, consistent with our \( \Lambda \)CDM+WMAP model.

4. The WiggleZ results are in excellent agreement with those of the GiggleZ \( N \)-body simulation incorporating redshift-space distortions. It is also in excellent agreement with our analytic \( \Lambda \)CDM+WMAP model, showing that our model for redshift-space distortions is accurate down to scales as small as \( 20 \, h^{-1} \) Mpc.

5. We can exclude a fractal with fractal dimension up to \( D_2 = 2.97 \), on scales between \( \sim 80 \, h^{-1} \) Mpc, and the largest scales probed by our measurement, \( \sim 300 \, h^{-1} \) Mpc, at 99.99% confidence.

6. By applying our analysis to the GiggleZ simulation, as well as a suite of fractal distributions of differing fractal dimension, we have shown that our result is not significantly distorted by the WiggleZ selection function and our method of correcting for it. We also show that we obtain consistent results even using different correction methods, i.e. using only complete spheres for the measurement, with and without correcting for incompleteness. This therefore confirms the reliability and robustness of our results.
3

The 6dF Galaxy Survey: the Minimum Variance Bulk Flow on 50 $- 70 \, h^{-1} \text{Mpc}$ Scales

Scrimgeour, M. I. et al., in preparation

3.1 Abstract

We measure the bulk flow of the local Universe using the 6dF Galaxy Survey peculiar velocity sample (6dFGSv), the largest and most homogeneous peculiar velocity sample to date. 6dFGSv is a Fundamental Plane sample of $\sim 10^4$ peculiar velocities covering the whole southern hemisphere for galactic latitude $|b| > 10^\circ$, out to redshift $z = 0.054$. We apply the ‘Minimum Variance’ bulk flow weighting method introduced by Watkins et al. (2009), which allows us to make a robust measurement of the bulk flow on scales of 50 and 70 $h^{-1}$Mpc. We investigate and correct for potential bias due to the lognormal velocity uncertainties, and verify our method by constructing $\Lambda$CDM 6dFGSv mock catalogues incorporating the survey selection function. For 50 $h^{-1}$Mpc scales we find a bulk flow amplitude of $U = 259 \pm 54$ km s$^{-1}$ in the direction $(l, b) = (317 \pm 16, 35 \pm 11)$, and for 70 $h^{-1}$Mpc scales we find $U = 254 \pm 54$ km s$^{-1}$, also in the direction $(l, b) = (317 \pm 16, 35 \pm 11)$. Our results are in agreement with other recent measurements of the direction of the bulk flow,
and our measured amplitude is consistent with a ΛCDM prediction.

### 3.2 Introduction

Galaxy peculiar velocities are one of the only probes of large-scale structure in the nearby Universe, and are gaining interest as a promising cosmological probe. Peculiar velocities are the motions of galaxies caused by gravity, separate from the isotropic Hubble expansion. They are usually measured statistically via redshift-space distortions (Kaiser, 1987; Peacock et al., 2001; Tegmark et al., 2004; Guzzo et al., 2008) but can also be measured directly. The line-of-sight component of the peculiar velocity $v$ of a galaxy at position $r$ is given by

$$ v \equiv v \cdot \hat{r} = c \left( \frac{z_{\text{obs}} - z_r}{1 + z_r} \right) $$  \hspace{1cm} (3.1)

where $c$ is the speed of light, $z_{\text{obs}}$ is the observed redshift, measured spectroscopically and corrected to the CMB restframe, and $z_r$ is the redshift corresponding to the real-space comoving distance $r$ of the galaxy.\(^1\)

In order to measure a peculiar velocity, a redshift-independent measurement of the distance $r$ is required. There are various methods of measuring distances, using so-called ‘distance indicators,’ which usually entail some form of standard candle (e.g. Type Ia supernovae or the Tully-Fisher relation) or standard ruler (e.g. galaxy sizes calibrated by the Fundamental Plane). Distance estimators typically have large uncertainties, which dominate the uncertainty on the peculiar velocity. These uncertainties are usually a constant fraction of distance.

Observations of the peculiar velocity field $v(r)$ are useful, since in the linear regime $v(r)$ is directly related to the density field $\delta(r)$, via (Peebles, 1980)

$$ v(r) = \frac{H_0 f}{4\pi} \int d^3r' \frac{\delta(r')(r' - r)}{|r' - r|^3}, $$ \hspace{1cm} (3.2)

where $f \equiv d\ln D/d\ln a$ is the present-day growth rate of cosmic structure (in terms

\(^1\)Equation 3.1 is often approximated in the literature as $v = c z_{\text{obs}} - H_0 D$, where $H_0$ is the Hubble constant and $D$ is the proper distance to the galaxy. However, this is only accurate for $z \ll 0.1$ (Harrison, 1993; Davis & Lineweaver, 2004; Davis & Scrimgeour, 2014).
of the linear growth factor $D$ and cosmic scale factor $a$), and

$$\delta(r) = \frac{\rho(r) - \bar{\rho}}{\bar{\rho}}$$

with $\bar{\rho}$ the average density of the Universe. Peculiar velocity measurements therefore allow us to trace the total matter distribution, including dark matter, without the complication of galaxy bias, and over a large range of scales. They also probe the nature of gravity through the growth rate $f$.

There are different ways of using peculiar velocity information to test the cosmological model. A commonly used statistic is the dipole, or ‘bulk flow’ of the velocity field, which corresponds to the average velocity of galaxies within a given volume with respect to the CMB, and is given by

$$U(r) = \frac{3}{4\pi r^3} \int_{x=0}^{r} v(x) d^3x$$

where $r$ is the radius of the sphere in which the bulk flow is measured, and $v(x)$ is the peculiar velocity field.

One aim of measuring the bulk flow is to understand the motion of the Local Group (LG) with respect to the CMB, of $627 \pm 22$ km s$^{-1}$ towards $l = 276 \pm 3^\circ$, $b = 30 \pm 2^\circ$ (Kogut et al., 1993). In the gravitational instability model of linear theory, this is expected to be due to the presence of a large nearby overdensity, and should converge to the CMB dipole at large enough distance. However, attempts to reconstruct the CMB dipole using the density field have been inconsistent. Some studies have suggested that it is necessary to go to scales of at least that of the Shapley Supercluster at $150 h^{-1}$ Mpc to recover the dipole motion (Kocevski & Ebeling, 2006; Muñoz & Loeb, 2008; Lavaux et al., 2010) while Erdoğan et al. (2006a,b) suggest only $\sim 30\%$ of the motion is due to structures beyond $50 h^{-1}$ Mpc.

Bulk flow measurements have seen much interest in the literature, but have also had a history of producing conflicting results. Some of the earliest measurements gave indications of apparently large bulk flows (Rubin et al., 1976; Dressler et al., 1987a; Lynden-Bell et al., 1988), while others found values consistent with predictions (Hart & Davies, 1982; de Vaucouleurs & Peters, 1984; Aaronson et al., 1986) – see Kaiser (1988) and Strauss & Willick (1995) for a review of early measurements.
More recently, an increase in the amount and quality of peculiar velocity data has led to a surge of new measurements. Again, some of these appear to find evidence of an unusually large bulk flow \cite{Kashlinsky2008, Watkins2009, Feldman2010, Abate2012, Lavaux2013}, while others find results consistent with the ΛCDM prediction \cite{Colin2011, Nusser2011, Osborne2011, Dai2011, Turnbull2012, Ma2013, Planck2013b}.

Some reported detections of unusually large bulk flows have been directly challenged. \cite{Kashlinsky2008} claimed to find a large dipole in the WMAP kinetic Sunyaev-Zel’dovich (kSZ) effect, indicating a bulk flow of 600-1000 km s\(^{-1}\) out to \(z \sim 0.1\), while \cite{Keisler2009} showed their uncertainties were underestimated, reducing the significance of their result. \cite{Watkins2009} combined several different peculiar velocity catalogues, and used a ‘minimum variance’ bulk flow estimator to find a bulk flow of 407 km s\(^{-1}\) on a scale of 50 \(h^{-1}\) Mpc, while \cite{Ma2013} repeated their analysis using a hyperparameter method to combine the surveys, along with a different choice of velocity dispersion parameter, and found a smaller bulk flow consistent with ΛCDM. Hence, although a large bulk flow remains an intriguing possibility, it is also possible it could be attributed to unaccounted-for systematic or statistical errors in existing measurements.

In this work we aim to shine new light on this controversy, by using peculiar velocity data from the 6-degree Field Galaxy Survey (6dFGS, \cite{Jones2004, Magoulas2012}) to make a new estimate of the local bulk flow. This dataset is the largest, most homogeneously derived peculiar velocity sample to date, with 8885 Fundamental Plane distances.

This chapter is structured as follows. In Section \ref{sec:6dfgs} we describe the 6dFGSv peculiar velocity sample. In Section \ref{sec:method} we explain how we derive peculiar velocity probability distributions from the derived Fundamental Plane logarithmic distances. In Section \ref{sec:weight} we describe the two different weighting methods we use to estimate the bulk flow. In Section \ref{sec:velocity} we investigate the best way of defining the peculiar velocity of each galaxy from the velocity probability distributions, to avoid bias in the estimated bulk flow. In Section \ref{sec:uncertainty} we describe how to estimate the velocity uncertainties in a way that does not bias the bulk flow estimation. In Section \ref{sec:mock} we describe our ΛCDM-based 6dFGSv mock catalogues. We present and discuss our
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 $H^{-1}$ MPC SCALES

results in Section 3.9 and conclude in Section 3.10.

Throughout this work we assume a ΛCDM cosmology with parameters from the Planck first data release, of $\Omega_m = 0.3175$, $\Omega_\Lambda = 0.6825$, $\sigma_8 = 0.8344$, and $H_0 = 100h$ km s$^{-1}$ Mpc$^{-1}$ with $h = 0.67$. We only use this cosmology when converting between distance and redshift, and for comparing our bulk flow results with the ΛCDM predicted velocity dispersion. Since 6dFGSv is at low redshift ($\leq 0.0537$) the results are only weakly dependent on the values of the cosmological parameters we assume. The uncertainties on these parameters are also significantly smaller than the uncertainties on our measurement, and so we fix these parameters throughout this work, since varying them should have a negligible effect.

3.3 6dFGS peculiar velocity sample

The 6-degree field Galaxy Survey (6dFGS) is a combined redshift and peculiar velocity survey of almost the whole southern hemisphere, performed using the Six-Degree Field (6D) multi-fibre spectrograph on the UK Schmidt Telescope from May 2001 to January 2006 (Jones et al., 2004, 2006, 2009). The survey covers galactic latitudes $|b| > 10^\circ$ out to a redshift of $z \sim 0.15$. The redshift survey (6dFGSz) contains 125,071 near-infrared (NIR) and optically selected spectroscopic galaxy redshifts, over 17,000 deg$^2$ and with a median redshift of 0.053. Targets were selected in the $JHK$ bands from the 2MASS Extended Source Catalog (2MASS XSC; Jarrett et al., 2000), with secondary samples in the $bJ$ and $r_F$ bands.

The peculiar velocity sample, denoted 6dFGSv (Campbell, 2009; Campbell et al., 2014), is a subset of 8885 bright, early-type galaxies for which distances were derived using the Fundamental Plane (FP) relation. This sample was drawn from $\sim 11,000$ galaxies in 6dFGSz with measured FP data, in the form of velocity dispersions and photometric scalelengths (Campbell et al., 2014). The sample was selected by requiring good redshift quality ($Q = 3 - 5$), $J$-band magnitude $J < 13.75$, redshifts less than 16,500 km s$^{-1}$ (or $z < 0.054$), and velocity dispersions larger than $\sigma_0 \geq 112$ km s$^{-1}$, with high signal-to-noise ($S/N > 5\AA^{-1}$).

The redshift distribution of 6dFGSv compared to that of the parent $J$-band 6dFGSz sample is shown in Figure 3.1. The fitting of the FP, and the selection cuts applied to obtain the FP and peculiar velocity samples, are described in detail
Figure 3.1: Redshift distribution of the 6dFGS peculiar velocity sample (6dFGSv, solid red histogram) compared to the parent J-band spectroscopic sample (6dFGSz, black line histogram). The vertical dashed line shows the redshift cut imposed on the velocity sample.

in (Magoulas et al., 2012, hereafter M12). The derivation of the FP distances and peculiar velocities, and correction for Malmquist bias and other selection effects, is described in (Springob et al., 2014, hereafter S14).

### 3.4 Calculating peculiar velocity probability distributions \( P(v) \) for 6dFGSv

The output of the FP peculiar velocity derivation for 6dFGSv (S14) is a probability distribution for the ‘logarithmic distance ratio’ for each galaxy, \( \eta \), defined by

\[
\eta = \log_{10}(D_z/D_r),
\]

(3.5)

where \( D_z \) is the co-moving distance in the fiducial \( \Lambda \)CDM cosmology corresponding to the observed redshift \( z \), while \( D_r \) is the co-moving distance corresponding to the
3. **THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 H⁻¹ Mpc SCALES**

angular diameter distance inferred from the Fundamental Plane.

Instead of obtaining \( \eta \) as a single value with an uncertainty, S14 derive the full posterior probability distributions \( P(\eta) \), in order to retain all the available information resulting from the selection cuts on the FP. These probability distributions are close to Gaussian in log distance, with a small skew due to the different selection effects and bias corrections, as described in S14.

We can convert \( P(\eta) \) into a probability distribution of peculiar velocity, \( P(v) \), via the relation

\[
P(v) = P(\eta) \frac{1}{D_r \ln(10)} \frac{dD_r}{dz_r} \frac{(1 + z_r)^2}{c(1 + z)}
\]

where \( z_r \) is the redshift corresponding to \( D_r \) in the assumed cosmology, and \( \ln \) is the natural logarithm. (See Appendix B for the derivation of this relation).

For each galaxy, we calculate \( D(z) \) via

\[
D(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}
\]

where

\[
E(z) = \frac{H(z)}{H_0} = \left[ \Omega_m,0 (1 + z)^3 + \Omega_{\Lambda,0} \right]^{1/2}
\]

and we use the fiducial \( \Lambda \)CDM parameter values listed in Section 3.2.

For each value of \( P(\eta) \), we calculate \( D_r(\eta, D_z) \) from Equation 3.5, then \( z_r(D_r) \) from the inverse of Equation 3.7. The derivative \( dD_r/dz_r \) is calculated numerically and evaluated at \( z_r \). To calculate the peculiar velocity \( v \) corresponding to \( \eta \) we use the relation

\[
(1 + z) = (1 + z_r)(1 + z_p)
\]

where \( z \) is the total, observed redshift, \( z_r \) is the redshift corresponding to \( D_r \) in the assumed cosmology, and \( z_p = v/c \) is the peculiar velocity redshift. From this we have

\[
v(\eta, z) = c \left( \frac{z - z_r(\eta, z)}{1 + z_r(\eta, z)} \right)
\]

which is a relation that must be calculated numerically. This relation is illustrated for the 6dFGSv sample in Figure 3.2.

The resulting velocity probability distribution \( P(v) \) is illustrated for a single 6dFGSv galaxy in Figure 3.3. The log distance probability distributions are close to
Gaussian, with some skew introduced by the selection cuts applied to the sample, and so the peculiar velocity probability distributions are roughly lognormal. We show on this plot four different ways of defining the velocity $v$ from the distribution, which we will explain in Section 3.6.

The distribution of the mean $\eta$ values, $\langle \eta \rangle = \int_{-\infty}^{\infty} \eta P(\eta) d\eta$, for 6dFGSv is shown in Figure 3.4. The mean over the total sample is $\langle \eta \rangle = (4.6 \pm 1.2) \times 10^{-3}$.

We describe the derivation of the velocity uncertainties, $\sigma_v$, in Section 3.7.

### 3.5 Bulk flow estimators

The bulk flow is the average velocity of a given volume of space, usually taken to be a spherical region centred on us, as defined by Equation 3.4. In practice, however, we can never perfectly sample the velocity field. Peculiar velocity samples are typically sparse, with complicated geometries and large uncertainties, and so the survey window function must be taken into account. Additionally, we only observe the line-of-sight component of the peculiar velocities rather than the full 3D vectors. As shown by Kaiser (1988), this means that while the bulk flow with 3D information would be $\mathbf{U} = (1/N) \sum v_n$, the solution with line-of-sight information would be $\mathbf{U} = (3/N) \sum (v_n \cdot \mathbf{\hat{r}}_n) \mathbf{\hat{r}}_n$.

There have been different methods suggested in the literature to measure the bulk flow for a given peculiar velocity dataset. These include a maximum likelihood estimate (Dressler et al., 1987b; Kaiser, 1988), a comparison with the density field (Bertschinger et al., 1990; Dekel et al., 1999; Willick & Strauss, 1998; Turnbull et al., 2012), reconstruction of the velocity field using random velocity fields based on a velocity power spectrum (Nusser & Davis, 2011) and most recently, a ‘Minimum Variance’ weighting method (Watkins et al., 2009; Feldman et al., 2010).

In the next two subsections we give an overview of the maximum likelihood estimate (MLE) and the Minimum Variance (MV) method, each of which we will apply to the 6dFGSv data.

These methods calculate the bulk flow vector in Cartesian coordinates, and require the positions of the galaxies to be in Cartesian coordinates. We choose to do our analysis in both Galactic Cartesian coordinates, and Equatorial Cartesian coordinates (with the $z$ axis pointing towards the north pole). Since the 6dFGSv survey
3. **THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW**

**ON 50 – 70 H\(^{-1}\) MPC SCALES**

---

Figure 3.2: ‘Direct’ peculiar velocity \(v\) from Equation 3.10 as a function of \(\langle \eta \rangle\), with \(\eta = \log_{10}(D_z/D_r)\), for the 6dFGSv sample, colour-coded by redshift distance \(D_z\). The \(v(\eta)\) relation is monotonic for a given \(D_z\), and is increasingly nonlinear for increasing \(D_z\).

---

Figure 3.3: A typical \(P(v)\) distribution for 6dFGSv, for an imagined galaxy at the mean redshift of 6dFGSv, and having the mean \(\eta\) uncertainty, \(\sigma_\eta\), of the sample, but with \(\langle \eta \rangle \equiv 0\). The red long-dashed line is the mean of \(P(v)\), the magenta dot-dashed line is the maximum likelihood, the cyan solid line is the median and the black short-dashed line is the direct conversion of \(\langle \eta \rangle\) to \(v\), which is almost identical to the median. Since \(\langle \eta \rangle\) is zero, the peculiar velocity of the galaxy should be zero, but only the direct \(v\) has this value.
3.5. **BULK FLOW ESTIMATORS**

![Figure 3.4: Distribution of the expectation values of the $P(\eta)$ distributions, $\langle \eta \rangle$, for the 6dFGSv sample, with $\eta = \log_{10}(D_z/D_r)$. The dashed grey line shows $\eta = 0$.](image)

only covers the southern hemisphere, we expect systematic bias from the sky coverage to be largest in the north-south direction, so doing the analysis in Equatorial coordinates should make it easier to see any such bias.

### 3.5.1 Maximum Likelihood estimate

The MLE has traditionally been the most common technique used to measure the bulk flow. We consider here the MLE using inverse variance weighting of Kaiser (1988). Given a sample of $N$ objects at positions $r_{n,i}$, each having a measured line-of-sight velocity $v_n$ with uncertainty $\sigma_n$, the ideal method is to do a maximum likelihood fit to a 4-parameter model comprised of a 3-component bulk flow $u_i$ and small-scale velocities parameterised by their 1D velocity dispersion, $\sigma_*$. The likelihood function is then

$$L(u_i, \sigma_*) = \prod_n \frac{1}{\sqrt{\sigma_n^2 + \sigma_*^2}} \exp \left( -\frac{1}{2} \frac{(v_n - \hat{r}_{n,i}u_i)^2}{\sigma_n^2 + \sigma_*^2} \right). \quad (3.11)$$

This likelihood assumes that both the small-scale velocities and the observational uncertainties are Gaussian distributed.
3. **THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 H\(^{-1}\) MPC SCALES**

In practice, it is not possible to model the velocity on all scales, since current models are only accurate for large, linear scales, and break down on nonlinear scales. Since the bulk flow should be relatively insensitive to the small-scale velocities, it is possible to assume a fixed value for \(\sigma_*\). In this case, the MLE solution for the bulk flow components \(u_i\) is (Kaiser, 1988):

\[
    u_i = A_{ij}^{-1} \sum_n \frac{\hat{r}_{n,j} v_n}{\sigma_n^2 + \sigma_*^2}
\]

where

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

The MLE weight for each individual peculiar velocity is then

\[
    w_{i,n} = \sum_j A_{ij}^{-1} \frac{\hat{r}_{n,j}}{\sigma_n^2 + \sigma_*^2}
\]

and the uncertainties on the bulk flow components are

\[
    \sigma_i = \sqrt{\sum_n (w_{i,n} \sigma_n)^2}.
\]

This solution makes the following assumptions:

1. the observational errors, \(\sigma_n\), are Gaussian
2. linear theory holds, so \(v_n << H_0 r_n\)
3. we can neglect uncertainty in \(r_n\)
4. that \(u_i\) is fairly insensitive to small-scale velocities, and that \(\sigma_*\), which will be strongly influenced by nonlinear flows, can be fixed at a given value.

In practice, nearly all of these assumptions will be violated to some extent. The observational errors on \(v\) are *not* Gaussian, \(v \sim 20 – 30\% \text{ of } H_0 r\), and linear theory does not strictly apply, since \(\sigma_* \sim 300 \text{ km s}^{-1}\) is comparable to the expected bulk flow amplitude on the scales we measure. However, we do not expect these to have a significant impact on our measurement.
3.5. **BULK FLOW ESTIMATORS**

3.5.2 Minimum Variance method

Although the MLE is simple to perform and should give the best estimator of a sparse sample, it has several disadvantages. It will have a complex window function dependent on the geometry and uncertainties of a particular peculiar velocity survey, making it difficult to compare between surveys and with theory. It is also density-weighted rather than volume-weighted, as it tends to upweight high-density regions where galaxies are more likely to be measured, and downweight low density regions. Finally, because it down-weights galaxies with larger uncertainties, i.e. the more distant galaxies, the MLE tends to be dominated by the most nearby galaxies in the sample and so minimises the scale on which the bulk flow is measured.

The ‘Minimum Variance’ (MV) method of Watkins et al. (2009) (hereafter WFH09) and Feldman et al. (2010) is an extension of the MLE method, which constructs a more optimal set of weights that allow a volume-weighted measurement of the bulk flow to be made with a specified window function. This is achieved by determining weights \( w_{i,n} \) for each galaxy that minimise the variance between the bulk flow measured by the sample, and the bulk flow that would be measured by an ‘ideal’ survey, with the specified window function. In their case, they choose this to be a perfectly sampled, all-sky Gaussian survey with ‘ideal’ radius \( R_I \).

The MV weights are calculated from

\[
\mathbf{w}_i = (\mathbf{G} + \lambda \mathbf{P})^{-1} \mathbf{Q}_i, \tag{3.16}
\]

where \( i \) denotes the three bulk flow components. \( \mathbf{P} \) is the \( k = 0 \) limit of the angle-averaged window function, \( \mathbf{Q}_i \) incorporates information about the input ideal window function, \( \lambda \) is a Lagrange multiplier, and \( \mathbf{G} \) is the covariance matrix of the individual peculiar velocities, given by

\[
G_{nm} = \langle v_n v_m \rangle = \delta_{nm} (\sigma_n^2 + \sigma_z^2) + \frac{f(\Omega_m)^2 H_0^2}{2 \pi^2} \int P(k) f_{mn}(k) dk, \tag{3.17}
\]

where \( H_0 \) is the Hubble constant, \( f \sim \Omega_m^{0.55}(z) \) is the growth rate of cosmic structure, and \( f_{mn}(k) \) is the angle-averaged window function. The first term is the noise term,
while the second part is the cosmic variance, or ‘geometrical’ term, and incorporates the power spectrum of a given cosmological model. We give further details of how the weights are calculated in Appendix C; also see WFH09 and Feldman et al. (2010).

The MV bulk flow is then

$$u_i = \sum_n w_{i,n} v_n.$$  \hspace{1cm} (3.18)

Following WFH09 we also choose a Gaussian survey as our ideal survey, using two different ideal radii: (1) $R_I = 50\ h^{-1}\text{Mpc}$ for comparison with WFH09; and (2) $R_I = 70\ h^{-1}\text{Mpc}$. We choose the latter since it is close to the ‘Maximum Likelihood Estimate depth’ of 6dFGSv, which is calculated via

$$d_{\text{MLE}} = \frac{\sum r_n w_n}{\sum w_n},$$  \hspace{1cm} (3.19)

where the MLE weights are $w_n = 1/(\sigma_n^2 + \sigma_z^2)$. We find it to be $\sim 72\ h^{-1}\text{Mpc}$ for 6dFGSv.

The ideal Gaussian survey will have a radial density profile given by

$$\rho(r) \propto \exp(-r^2/2R_I^2),$$  \hspace{1cm} (3.20)

and its radial number distribution is

$$n(r) \propto r^2 \exp(-r^2/2R_I^2).$$  \hspace{1cm} (3.21)

We plot $n(r)$ for our two ideal surveys in Figure 3.5 along with the number distribution of 6dFGSv for comparison. The 6dFGSv sample has a cutoff at $160\ h^{-1}\text{Mpc}$ corresponding to $z = 0.054$, so we also apply this to our ideal surveys.

The ideal survey used by WFH09 is an all-sky survey, since the dataset they used was all-sky; in the case of 6dFGSv, we only have half the sky. We discuss the effect of partial sky coverage on our measurement in Section 3.9.
3.5. **BULK FLOW ESTIMATORS**

3.5.3 $\sigma_*$ estimation

The 1D velocity dispersion parameter $\sigma_*$, as previously mentioned, accounts for small-scale random motions. The value of $\sigma_*$ should only affect the bulk flow measurement very weakly. It should have the strongest effect on nearby galaxies, for which the velocity errors are smallest, but in the Minimum Variance method where these are down-weighted by the $R_I = 50h^{-1}$Mpc window function, $\sigma_*$ should still have only a small effect on the measured bulk flow (Feldman et al., 2010).

Different surveys will have different values for $\sigma_*$, depending on the properties of the galaxy sample. Ma et al. (2011) performed a fit for $\sigma_*$ to several different galaxy surveys (including SN, Tully-Fisher and FP velocities), and for ENEAR, a survey of 698 FP distances to early-type galaxies, they find $\sigma_* \sim 280$ km s$^{-1}$. We might expect $\sigma_*$ for the 6dFGSv galaxies to be similar. However, for 6dFGSv, about one third of the galaxies have cluster redshifts, potentially reducing the amount of
small-scale, nonlinear motion, so we might expect the effective $\sigma_* \to$ to be lower.

In the case of 6dFSGv, small-scale velocities need to be accounted for in the fitting of the FP, since the fitting is done assuming each galaxy is at its redshift distance, and so velocities add to the scatter of the plane. A value of $\sigma_* = 300 \text{ km s}^{-1}$ is accounted for in the fitting of the FP by M12 and S14, and is effectively subtracted from the uncertainty in the $P(\eta)$ distribution for each galaxy. This means we need to ‘add back in’ this uncertainty in our bulk flow weights. Johnson et al. (2014) perform a fit to $\sigma_*$ for the 6dFSGv sample and find it to peak at zero. However, $\sigma_*$ also acts to regularise the bulk flow weights, to prevent galaxies with low error dominating the results, so assuming a zero $\sigma_*$ is not ideal. We find that varying $\sigma_*$ from 0 to 250 km s$^{-1}$ has little effect on our results, changing the MV bulk flow on the order of $\sim 2\%$. We therefore fix $\sigma_* = 250 \text{ km s}^{-1}$ for our analysis.

As a consistency test, we investigate the effect of $\sigma_*$ on the MLE bulk flow for 6dFSGv, for a range of $\sigma_*$ values from 0 to 1000 km/s. The results, and the percentage difference in $u_i$ caused by varying $\sigma_*$, are shown in Figure 3.6. We find that the MLE bulk flow amplitude is affected by up to 6% as $\sigma_*$ is varied from 0 to 300 km s$^{-1}$, and the individual components $u_i$ can vary up to 10%. However, this change only has a very small effect on our measurement, and its level of agreement with $\Lambda$CDM.

### 3.6 Defining $v$ from $P(v)$

Once we have converted the $P(\eta)$ distributions to $P(v)$ distributions for each of the 6dFSGv galaxies from Section 3.4, the question is how to use these to calculate the bulk flow. In order to use the MLE or MV methods, we need to know a single peculiar velocity value for each galaxy. However, the fact that the $\eta \to v$ conversion is nonlinear, meaning that the velocity probability distributions $P(v)$ are non-Gaussian, poses a significant difficulty for determining this value.

There are two main problems:

1. Having $\langle \eta \rangle = 0$ does not mean that $\langle v \rangle = 0$. The zeropoint of the 6dFGS Fundamental Plane was calibrated by setting $\langle \eta \rangle = 0$ for a strip of galaxies encircling the equator, with Dec $> -20 \text{ (Magoulas et al. 2013, in preparation).}$
3.6. DEFINING $v$ FROM $P(v)$

Figure 3.6: Top: MLE bulk flow components $u_i$ for 6dFGS$v$, as a function of $\sigma_*$, in Equatorial Cartesian coordinates. The coloured lines are for $u_x$ (green), $u_y$ (blue) and $u_z$ (red), as labelled. Bottom: the percentage change $\Delta u_i$ in $u_i$, when $\sigma_*$ is varied from 0. The coloured lines are the same as the top panel, and the black dashed line shows the change in the bulk flow amplitude, $\Delta |u|$. 

94
However, this automatically imposes a mean peculiar velocity \( \langle v \rangle \neq 0 \).

2. Even for a sample with \( \langle v \rangle = 0 \), the non-Gaussian distributions mean that \( P(v) \) has a tail towards higher peculiar velocity, and so the peculiar velocities will still be potentially biased.

### 3.6.1 Statistical definitions of \( v \)

To measure the bulk flow, we want to make use of all the available information in the \( P(\eta) \) and \( P(v) \) distributions; however, the estimators we use require a single input velocity value. We therefore consider four different ways of defining the ‘best’ velocity \( v \), illustrated in Figure 3.3 for an individual galaxy:

- the mean of \( P(v) \), \( \langle v \rangle \equiv \int_{-\infty}^{\infty} vP(v)dv \)
- the median of \( P(v) \)
- the maximum likelihood (M.L.) of \( P(v) \)
- the direct conversion of the mean of \( P(\eta) \), \( \langle \eta \rangle \), to \( v \) via Equation 3.10 (i.e. without computing \( P(v) \)). We denote this the “direct \( v \).”

The statistics of each of these definitions of \( v \) for the dataset – their mean, standard deviation, minimum and maximum values – are listed in Table 3.1. Here, every galaxy is equally weighted, with all weights being unity. We also show the distribution of velocities for each of these definitions in Figure 3.7. They are all similar, except that the ML \( v \) is systematically shifted to larger values.

### 3.6.2 Obtaining an unbiased \( v \)

The “direct \( v \)” – i.e. converting the mean of the Gaussian-distributed observable to a velocity – is the method commonly used in the literature. This seems suboptimal, since it does not make use of the full information in the \( P(v) \) distributions - however, it is in fact the least biased of the four methods, since the mean, median and maximum likelihood of \( P(v) \) will all give a non-zero peculiar velocity for a zero \( \eta \).

We first look at the case of the mean \( v \). In Figure 3.8 we illustrate what \( P(v) \) looks like for a galaxy at the mean redshift of 6dFGSv, where we have set \( \langle \eta \rangle \equiv 0 \),
3.6. DEFINING \( v \) FROM \( P(v) \)

Figure 3.7: Distribution of \( v \) values for 6dFGSv, for each of the four definitions we use: the mean, median and ML of \( P(v) \) (black, red and magenta), and the direct conversion of \( \langle \eta \rangle \) to \( v \) (blue). The median and direct \( v \) histograms lie almost on top of each other; the ML \( v \) histogram is shifted to larger values of \( v \) than the other definitions.

Figure 3.8: Illustration of the \( P(v) \) distributions obtained for a galaxy at the mean redshift of 6dFGSv, with \( \langle \eta \rangle \equiv 0 \) and \( \sigma_\eta \) set to the minimum (blue), maximum (red) or mean (purple) \( \sigma_\eta \) value of the total 6dFGSv sample, labelled. The value of \( \langle v \rangle \) is negative in each case, also labelled, and indicated by the vertical lines.
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 $H^{-1}$ MPC SCALES

Table 3.1: Statistics of the 6dFGSv peculiar velocities, depending on definition of $v$. For each definition we show the mean over the whole sample (2nd column), the standard deviation (S.D.) of the velocities (3rd column), minimum and maximum velocity (4th and 5th columns).

<table>
<thead>
<tr>
<th>$v$ definition</th>
<th>mean (km s$^{-1}$)</th>
<th>S.D. (km s$^{-1}$)</th>
<th>min (km s$^{-1}$)</th>
<th>max (km s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct $v$</td>
<td>-124</td>
<td>2715</td>
<td>-18373</td>
<td>8442</td>
</tr>
<tr>
<td>mean $v$</td>
<td>-540</td>
<td>2821</td>
<td>-19302</td>
<td>8196</td>
</tr>
<tr>
<td>median $v$</td>
<td>-117</td>
<td>2727</td>
<td>-18786</td>
<td>8459</td>
</tr>
<tr>
<td>ML $v$</td>
<td>707</td>
<td>2565</td>
<td>-16723</td>
<td>8963</td>
</tr>
</tbody>
</table>

but with $\sigma_\eta$ varying from the minimum to the maximum $\sigma_\eta$ in the dataset. Since Equation 3.6 is nonlinear, $P(v)$ is non-gaussian even though $P(\eta)$ is close to gaussian. The mean peculiar velocity, $\langle v \rangle$, is shown for each distribution and is always negative, even though $\langle \eta \rangle \equiv 0$.

If every 6dFGSv galaxy had $\langle \eta \rangle \equiv 0$, then the mean sample peculiar velocity would be negative, which would appear as a spurious bulk flow towards the north pole, since we only have half the sky.

The mean, median and maximum likelihood of $P(v)$ all have this problem. In Figure 3.9 we show the distribution of mean, median and ML $v$ values for 6dFGSv, if we set $\langle \eta \rangle \equiv 0$ for each galaxy. Even if $\eta \equiv 0$, the shape and spread of the resulting probability distribution $P(v)$ means that the mean, median and M.L. are all nonzero and have a dispersion caused by the differing $\sigma_v(\sigma_\eta, z, D_r)$ of each galaxy. The median has the lowest dispersion and is closest to the true value - it should be equal to zero, since the probability distribution is equal either side of zero, but due to numerical errors it is not identical to zero.

However, the direct $v$ will not suffer this problem, and would appear as a delta function at $v = 0$ in this figure. We therefore choose the direct $v$ as the least biased definition of velocity, and we use this for all our main results in this work.
3.6. DEFINING $v$ FROM $P(v)$

Figure 3.9: Distribution of the mean (magenta), median (black) and maximum likelihood (ML, blue) values of peculiar velocity $v$ for 6dFGSv, derived from the peculiar velocity distributions $P(v)$, if we first set the mean of $P(\eta)$, $\langle \eta \rangle \equiv 0$ for each galaxy. The mean is systematically negative; the ML is systematically positive; while only the median is distributed around zero.
3.7 Velocity uncertainties and bulk flow weights

The way in which the velocity uncertainties $\sigma_n$ are determined can have an important impact on the measured bulk flow, since the weights depend on the uncertainties. Caution is required, since if the velocity uncertainties are derived in a way that makes them dependent on the velocities themselves, this can give rise to a systematic bias in the bulk flow measurement.

This is an important effect that has not been previously addressed in the literature. It is more important for surveys like 6dFGSv that are at relatively high redshift and have large uncertainties, but its effect should be noted in any bulk flow measurement.

3.7.1 The standard deviation of $P(v)$: correlation with $v$

A naïve approach to estimating the error on individual velocity measurements, $\sigma_n$, is to calculate the standard deviation of $P(v)$, which we denote $\sigma_n^{SDv}$, via the standard formula

$$\sigma_n^{SDv} = \left( \int_{-\infty}^{\infty} v^2 P(v) dv - \bar{v}^2 \right)^{1/2}. \quad (3.22)$$

However, since the $\eta \rightarrow v$ conversion is nonlinear, the width of $P(v)$ depends on the value of $\eta$ (or $v$), and so $\sigma_n^{SDv}$ is correlated with $v$ at a given redshift, as we show in Figure 3.10. Negative peculiar velocities have, on average, larger uncertainties than positive peculiar velocities. This is because the nonlinear relation between $v$ and $\eta$, which we saw in Figure 3.2, is steeper for more negative $\eta$, or equivalently, the derivative $dv/d\eta$ is larger, as we illustrate in Figure 3.11. For more negative $\eta$, a given interval in $\eta$ maps to a larger interval in $v$, and so $P(v)$ becomes wider, producing a larger $\sigma_n^{SDv}$.

3.7.2 Bulk flow bias from $\sigma_n^{SDv}$

Since the MLE and MV bulk flow weights are inversely proportional to $\sigma_n$ (see Equations 3.14 and 3.16), we see from Figure 3.10 that positive peculiar velocities will be up-weighted relative to negative peculiar velocities. This distorts the amplitude and direction of the measured bulk flow. For 6dFGSv, the effect is
3.7. VELOCITY UNCERTAINTIES AND BULK FLOW WEIGHTS

Figure 3.10: The correlation between the standard deviation $\sigma_{n}^{SDv}$ of $P(v)$, and direct peculiar velocity $v$, colour-coded by redshift. For any given redshift, there is a close-to-linear relation between $\sigma_{n}^{SDv}$ and $v$, with negative peculiar velocities having larger uncertainties on average.

Figure 3.11: The derivative $dv/d\eta$ as a function of $\eta$, for three different values of constant distance $D_z$: 10, 70 and 150 $h^{-1}$ Mpc, as labelled.
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 $H^{-1}$ MPC SCALES

particularly significant, since the survey only covers half the sky, and so any radial systematic in the velocities will translate to a spurious bulk flow component in the north-south direction. Since galaxies with positive peculiar velocity (negative Equatorial $z$-component) are up-weighted relative to galaxies with a negative peculiar velocity (positive $z$-component), this results in a large, spurious bulk flow component towards the south pole.

We can see the effect of this by calculating the bulk flow in individual redshift bins, using $\sigma_n^{SDv}$ as the velocity uncertainty. The MLE bulk flow components are shown, as a function of redshift, in Figure 3.12. The $x$ and $y$ components are fairly constant with redshift, but the $z$ component increases in amplitude with redshift, indicating a radial-dependent systematic. We see this is not produced by a ‘true’ bulk flow, as the radial dependence remains when we randomly rearrange the peculiar velocities within each bin. We also note that the average radial peculiar velocity in each bin, given by the dashed line, is close to zero, indicating the large $z$-component is not real. However, the measured $u_z$ is very close in magnitude to the mean absolute radial velocity, since $u_z$ is dominated by the positive velocities in the $z$ direction that are not cancelled out as there are no galaxies in the north.

In Table 3.2, we list the MLE and MV bulk flow results we obtain for 6dFGSv, when using $\sigma_n = \sigma_n^{SDv}$. While the $x$ and $y$ values appear more reasonable, there is a very large, negative $z$-component of the bulk flow, caused by the bias explained above.

3.7.3 A solution: distance-dependent $\sigma_n$

We avoid this problem by defining the peculiar velocity uncertainties $\sigma_n$ in a way that does not depend on $v$ or $\eta$, for the purposes of calculating the bulk flow weights. We do this by making a linear fit to $\sigma_n^{SDv}$ as a function of distance. We plot $\sigma_n^{SDv}$ versus distance in Figure 3.13 - overall this is linearly proportional, but there is a dependence on $v$, almost orthogonal to redshift. By taking the linear best fit we remove this dependence, essentially taking the uncertainty a galaxy would have if it had zero peculiar velocity. For our bulk flow estimation, we therefore use the
3.7. VELOCITY UNCERTAINTIES AND BULK FLOW WEIGHTS

Figure 3.12: Top: The MLE bulk flow components \((u_x, u_y, u_z)\) and magnitude \(|U|\), for the 6dFGSv sample, in Equatorial Cartesian coordinates, binned in redshift bins of \(dz = 0.01\). The redshift bins are shown by the grey lines. The dashed black line is the mean radial peculiar velocity in each redshift bin, and the dot-dashed line is the mean absolute radial peculiar velocity. Bottom: the same, but with the radial peculiar velocities in each bin randomly shuffled before calculating the BF.
Table 3.2: Bulk flow results for the MLE and MV estimators, using the standard deviation of $P(v)$, $\sigma^p$, as the peculiar velocity uncertainties $\sigma_n$. Columns are the bulk flow vectors $(u_x, u_y, u_z)$ in Equatorial Cartesian coordinates, and magnitudes $|U|$. The derived bulk flow uncertainties $\delta_u$ are shown in the last row. For the MLE, the noise uncertainties are shown; for the MV, the full noise + cosmic variance uncertainties are shown, with noise-only uncertainties in parentheses. The MV uses an ideal Gaussian window function with $R_I = 70 \, h^{-1}$ Mpc, and a cutoff at $< 160 \, h^{-1}$ Mpc.

<table>
<thead>
<tr>
<th>$v$ definition</th>
<th>MLE</th>
<th>MV ($R_I = 70 , h^{-1}$ Mpc, $&lt; 160 , h^{-1}$ Mpc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_x$ (km s$^{-1}$)</td>
<td>$-159 \pm 46$</td>
<td>$-177 \pm 34$</td>
</tr>
<tr>
<td>$u_y$ (km s$^{-1}$)</td>
<td>$-1285 \pm 99$</td>
<td>$-1285 \pm 49$</td>
</tr>
<tr>
<td>$u_z$ (km s$^{-1}$)</td>
<td>$-1285 \pm 128$</td>
<td>$-1285 \pm 138$</td>
</tr>
<tr>
<td>$</td>
<td>U</td>
<td>$ (km s$^{-1}$)</td>
</tr>
<tr>
<td>$\delta_u$ (km s$^{-1}$)</td>
<td>$+36$</td>
<td>$+51$</td>
</tr>
</tbody>
</table>

Note: All values are shown with uncertainties in parentheses.
3.8  6dFGSv ΛCDM MOCK CATALOGUES

![Figure 3.13: The correlation between the standard deviation \( \sigma_n^{SDv} \) of \( P(v) \), and redshift distance \( D_z \). The colour gradient shows the corresponding peculiar velocity. A linear best-fit to the points is shown in black.](image)

The correlation between the standard deviation \( \sigma_n^{SDv} \) of \( P(v) \), and redshift distance \( D_z \). The colour gradient shows the corresponding peculiar velocity. A linear best-fit to the points is shown in black.

The uncertainty on the bulk flow measurement calculated using this new \( \sigma_n \) remains similar to that using \( \sigma_n^{SDv} \), as can be seen by comparing Tables 3.2 and 3.3.

This removes the correlation of the weights with velocity, as shown in Figure 3.14, where we show the \( z \)-component MLE weights \( w_{z,n} \), as a function of peculiar velocity for both definitions of \( \sigma_n \). The weights are clearly larger for positive velocities in the first case, but this trend goes away for the new uncertainties.

3.8  6dFGSv ΛCDM Mock Catalogues

In order to test possible systematics in our bulk flow measurement arising from the survey selection function, we apply our bulk flow analysis to random ΛCDM mock catalogues of 6dFGSv, incorporating the survey selection function. To create...
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 $H^{-1}$ MPC SCALES

Figure 3.14: The Equatorial Cartesian $z$-component MLE weights $w_{i,z}$, as a function of peculiar velocity, for (top) peculiar velocity uncertainties as the standard deviation of $P(v)$, i.e. $\sigma_n = \sigma_v^{SDv}$, and (bottom) for uncertainties as $\sigma_n = 0.324H_0D_z$. The colour scale indicates the redshift distance. A trend is clearly visible in the top plot, but goes away in the bottom plot.
ACDM peculiar velocity mocks, we need to make use of an N-body simulation, which provides both positions and velocities of galaxies. In this section we describe how we determine the selection function of 6dFGSv, and use this to generate mock catalogues using the GiggleZ N-body simulation. This selection function also allows for the creation of random catalogues for clustering analysis. The selection function \( W(x) \) is a function indicating the expected number density of 6dFGSv galaxies at a position \( x \), due to the different selection criteria of the sample. These can be both angular- and redshift- dependent.

### 3.8.1 6dFGSv survey selection function

M12 and S14 describe the selection process used to obtain the 6dFGSv sample of 8885 galaxies from the full 6dFGS redshift sample of \( \sim 125,000 \) galaxies (‘6dFGSz’). They first select galaxies suitable for fitting the FP, by selecting galaxies with reliable redshifts (with redshift quality \( Q = 3 - 5 \)) and redshifts less than 16 500 km s\(^{-1}\) (or \( z < 0.0537 \)), above which a key spectral feature used to measure velocity dispersion is shifted out of the wavelength range. They then morphologically select early-type (E/S0) galaxies, by matching the observed spectra to template galaxy spectra. This produced a sample of \( \sim 20,000 \) galaxies.

These \( \sim 20,000 \) galaxies then had their velocity dispersions measured using the Fourier cross-correlation technique ([Campbell] 2009). Of these, galaxies with a high signal-to-noise (S/N > 5 A\(^{-1}\)), and velocity dispersions larger than the instrumental resolution limit (\( s \geq 2.05, \text{ or } \sigma_0 \geq 112 \text{ km s}^{-1} \)) were selected, to produce a ‘Fundamental Plane sample’ of 11,287 galaxies. This sample, with both spectroscopic measurements from 6dFGS and photometric measurements from 2MASS in the \( J, H \) and \( K \) bands, was used by M12 for the fitting of the FP parameters.

Finally, the peculiar velocity sample 6dFGSv was obtained from the FP sample after several further cuts. A stricter redshift limit of \( cz < 16120 \) (\( z < 0.054 \)) was imposed in the CMB frame, along with further magnitude cuts of \( J \leq 13.65, H \leq 12.85 \) and \( K \leq 12.55 \), to maintain high completeness over the sky. Further galaxies were removed after a visual inspection, and a velocity dispersion \( \chi^2 \) cut, to obtain the final peculiar velocity sample of 8885 galaxies.
3.8.2 Fundamental Plane fitting

Here we introduce the FP terminology we will use in making the mocks. The FP relation can be written in logarithmic units as

\[ r = as + bi + c, \] (3.24)

where \( r \equiv \log R_e \), \( s \equiv \log \sigma_0 \) and \( i \equiv \log \langle I_e \rangle \), where \( R_e \) is the effective radius in units of \( h^{-1} \) kpc, \( \sigma_0 \) is the central velocity dispersion in units of kms\(^{-1}\), and \( \langle I_e \rangle \) is the mean surface brightness, in units of L\(_{\odot}\) pc\(^{-2}\). The coefficients \( a \) and \( b \) are the slopes of the plane and \( c \) is the offset of the plane. M12 use logarithms of base 10.

M12 determine the FP parameters for 6dFGSv using a maximum-likelihood fit to a 3D Gaussian model. The FP can be described either in terms of the observational parameters \((r, s, i)\), or in terms of the three unit vectors corresponding to the axes of the 3D Gaussian describing the galaxy distribution. M12 refer to these as ‘FP-space’ and ‘v-space’ respectively. The model can then be described by eight parameters: \( \{a, b, \bar{r}, \bar{s}, \bar{i}, \sigma_1, \sigma_2, \sigma_3\} \), where \((\bar{r}, \bar{s}, \bar{i})\) define the centre of the 3D Gaussian in FP-space and \((\sigma_1, \sigma_2, \sigma_3)\) are the dispersion of the Gaussian along each of the three axes in v-space. The offset of the FP can be calculated as \( c = \bar{r} - a\bar{s} - b\bar{i} \).

3.8.3 Mock sample algorithm

We create mock ΛCDM realisations of the 6dFGSv dataset for a given set of FP parameters \( \{a, b, c, \bar{r}, \bar{s}, \bar{i}, \sigma_1, \sigma_2, \sigma_3\} \), derived for the 6dFGS data by M12. We use the following steps to generate the mock 6dFGSv catalogue (some of which are the same as in that paper; we credit code from M12 for those parts of the algorithm):

1. For a clustered mock, start by drawing haloes from an N-body simulation in a mass range equivalent to the 6dFGS elliptical galaxies, i.e. pick haloes that match the bias of 6dFGS (this should effectively be a cut in morphological type). Or, for a random mock, generate a starting number of random galaxy positions.
Angular & Redshift cuts

2. Define the location of the observer, and calculate RA, Dec, true comoving distance $D_r$ and radial peculiar velocity $v$ for each galaxy. Also calculate the true and observed redshifts $z_t$, $z$, using Equation 3.9.

3. Only include haloes within hard angular cuts $\text{Dec} < 0^\circ$ and Galactic latitude $|b| > 10^\circ$.

4. Impose a redshift cut of $cz < 16120 \text{ km s}^{-1}$.

5. Normalise by applying a random subsampling to obtain the number of galaxies in the 6dFGS parent redshift sample.

Magnitude & Velocity Dispersion cuts

6. For each galaxy, draw values for $v_1$, $v_2$ and $v_3$ at random from a 3D Gaussian with standard deviations $\sigma_1$, $\sigma_2$ and $\sigma_3$ as listed in Table 3 of M12. We use the J-band values as the J-band has the smallest photometric errors.

7. Transform these values from the $\mathbf{v}$-space (principle axes) coordinate system to the $\{r, s, i\}$-space (observed parameters) coordinate system using the inverse of Equation 6 in M12, with the specified FP slopes ($a$ and $b$) and FP mean values ($\bar{r}$, $\bar{s}$ and $\bar{i}$). This gives the true Fundamental Plane parameters $(r_t, s_t, i_t)$ for the simulated galaxies.

8. Re-order each set of $(r_t, s_t, i_t)$ parameters in descending order of luminosity, 

$$\log L = l = 2r + i,$$

and assign them to the haloes in descending order of maximum circular velocity $V_{\text{max, sub}}$.

9. Use the comoving distance $D_r$ of each galaxy from the observer to determine the angular radius $\theta$ from the physical radius $r_t$, by calculating the angular diameter distance $D_A$:

$$D_A \equiv \frac{r_t}{\theta} = \frac{D_r}{1 + z_{\text{true}}},$$

108
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 H⁻¹ MPC SCALES

(this relation is true for Ω_k = 0, see Hogg 1999). Then θ is obtained from

\[ \log \theta = \log r_t - \log D_A. \]  (3.27)

10. Determine the true apparent magnitude \( m_t \) from the angular radius \( \theta \) and the degraded surface brightness \( i \) using

\[ m_t = \langle \mu_e \rangle - 2.5 \log[2\pi \theta^2], \]  (3.28)

where \( \langle \mu_e \rangle = M - 2.5i + 21.57 \), where \( M = 3.67 \) for the J-band. The surface brightness \( i \) is first degraded by ‘de-correcting’ for K-correction and surface brightness dimming.

11. Obtain the correlated measurement uncertainties in \( r, s, i \), \( (\epsilon_r, \epsilon_s, \epsilon_i) \), from the magnitude \( m_t \), using the matrix in Equation 13 of M12.

12. Add these measurement errors to the \( \{r, s, i\} \) obtain the observed values \( \{r_o, s_o, i_o\} \) for each galaxy.

13. Only include galaxies with velocity dispersion \( s_o > \log(116 \text{ km/s}) \). (Cut for instrumental resolution).

14. Determine the observed magnitude \( m_o \) using the observed values \( r_o \) and \( i_o \).

15. Keep the galaxy if the observed magnitude \( m_o \) is brighter than the faint limit for the velocity sample (\( J \leq 13.65 \)).

16. Use the selection function described in Jones et al. (2006) to determine the angular completeness of the 6dFGS spectroscopic follow-up, given the (RA, Dec, \( m_o \)) values for each galaxy. Sub-sample the galaxies with this probability.

17. Apply a random subsampling to account for cuts in signal-to-noise (S/N) ratio and \( R \).
3.8.4 The GiggleZ Simulation

In order to generate ΛCDM clustered 6dFGSv mocks, we apply our mock sample algorithm to the GiggleZ (Giga-parsec WiggleZ) simulation. GiggleZ (Poole et al., 2014) is a suite of dark matter $N$-body simulations run at Swinburne University of Technology, designed for theoretical analyses of the WiggleZ data set. GiggleZ has a WMAP-5 cosmology with $(\Omega_\Lambda, \Omega_m, \Omega_b, h, \sigma_8, n) = (0.727, 0.273, 0.0456, 0.705, 0.812, 0.960)$, with the initial conditions constructed to yield a CAMB (Lewis et al., 2000) power spectrum at a starting redshift of $z = 49$ using the Zeldovich approximation (Zel’dovich, 1970; Buchert, 1992).

We use the GiggleZ Main simulation, which contains $2160^3$ dark matter particles in a periodic box of side $1h^{-1}$ Gpc. The particle mass is $7.5 \times 10^9 h^{-1} M_\odot$, which allows bound systems with masses $\gtrsim 1.5 \times 10^{11} h^{-1} M_\odot$ to be resolved. This produces haloes with clustering bias $b$ from near unity to greater than 2.

Halo finding for GiggleZ was performed using subfind (Springel et al., 2001), which utilises a friends-of-friends (FoF) algorithm to identify coherent overdensities of particles and a substructure analysis to determine bound overdensities within each FoF halo. The resulting subfind substructure catalogues are rank-ordered by their maximum circular velocity ($V_{\text{max,sub}}$) as a proxy for halo size.

3.8.5 Determining 6dFGSv halo bias & mass

In order to generate our mocks using a GiggleZ halo catalogue, we need to find the mass range of haloes that corresponds to the masses of the 6dFGSv galaxies.

To do this, we first determine the bias of the 6dFGSv sample by calculating several different clustering statistics of the sample - the power spectrum $P(k)$, correlation function $\xi(r)$, angular power spectrum $C_l$ and angular correlation function $w(\theta)$. In order to do this we generate a random catalogue of $1 \times 10^6$ galaxies, using our 6dFGSv selection function. From the average best-fit bias of the clustering statistics, we find a bias of $b \sim 1.7$ for the mock catalogue. The power spectrum of 6dFGSv is shown in Figure 3.15 and has a best-fit bias of $1.54 \pm 0.04$.

Once we know the 6dFGSv bias, we then determine the bias of each GiggleZ halo group using the dark matter and group correlation functions ($\xi_{DM}$ and $\xi_G$) computed by Contreras et al. (2013). The ratio of these provides the bias of the
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 − 70 H⁻¹ MPC SCALES

Figure 3.15: The power spectrum $P(k)$ of the 6dFGSv sample. The red line shows a $\Lambda$CDM model with a Kaiser factor to model redshift-space distortion, and best-fitting bias of $1.54 \pm 0.04$. Credit: Chris Blake.

Figure 3.16: Bias values $b(r)$ for the 45 GiggleZ groups at $z = 0$. Dashed grey lines show the range in which we determine the mean bias $b$. 

111
halo group, via

\[ b^2(r) = \frac{\xi_G(r)}{\xi_{DM}(r)}. \] (3.29)

We plot \( b(r) \) for each of the 45 GiggleZ halo groups at \( z = 0 \) in Figure 3.16. On small scales, the bias is affected by cosmic variance, while on scales approaching the simulation size the number of pair counts decreases, again causing variance. Hence the most reliable range to measure \( b \) is \( 10h^{-1}\text{Mpc} \lesssim r \lesssim 120h^{-1}\text{Mpc} \). We take the bias of the halo group as the average within this range. The halo group with the nearest-value bias to 6dFGSv then provides the halo mass range of 6dFGSv.

In the case that the existing halo groupings do not match the halo bias we require, we can select GiggleZ haloes by applying a minimum cut in maximum subhalo circular velocity \( (V_{\text{max,sub}}) \). Since we need a subsample with a number density matching the 6dFGS data, we use a fast iterative method to identify the value of \( V_{\text{max,sub}} \) that produces the required number of haloes. (A straightforward sorting of the GiggleZ subhaloes by circular velocity would be far more computationally time-consuming, due to the large number of subhaloes - 9907707). This iterative method works by:

1. Start by calculating the mean \( V_{\text{max,sub}} \) for all the subhaloes in GiggleZ, \( V_{\text{mean}} \).
2. Set this as the initial \( V_{\text{max,sub}} \) cutoff value, \( V_{\text{cutoff}} \). Also start with a given increment size for \( V_{\text{cutoff}} \), \( dV \) (which we set initially to 200 km/s).
3. Find the number \( N_{\text{sub, incl}} \) of subhaloes that have \( V_{\text{max,sub}} > V_{\text{mean}} \) and so are included by the cut.
4. If \( N_{\text{sub, incl}} \) is too small, decrease \( V_{\text{cutoff}} \) by \( dV \).
5. If \( N_{\text{sub, incl}} \) is too large, increase \( V_{\text{cutoff}} \) by \( dV \).
6. Halve the increment size, i.e. \( dV_{\text{new}} = dV/2 \).
7. Repeat steps (iii) to (vi) until the desired \( N_{\text{sub, incl}} \) is reached.

### 3.8.6 Selecting the largest haloes

In practice, our mock sample algorithm in section 3.8.3, applied directly to the GiggleZ high circular velocity subsample we derive in section 3.8.5, does not produce
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW
ON 50 – 70 $H^{-1}$ MPC SCALES

Figure 3.17: Histogram showing the maximum circular velocity $V_{\text{max,sub}}$ of the dark matter haloes we select from GiggleZ, before (black) and after (red/blue) the cuts dependent on $\log M$ (blue) or $\log L$ (red). The blue and red histograms are almost identical. The histograms shown are the sum of 8 independent GiggleZ mocks.

We need to make sure that the cuts we apply preferentially remove low-luminosity, low-mass haloes, rather than random haloes. Therefore, whenever we require a random subsampling or renormalisation, we only remove haloes with the lowest $V_{\text{max,sub}}$.

In Figure 3.17 we plot histograms of $V_{\text{max,sub}}$ of the dark matter haloes in 8 mocks, before and after applying cuts that depend on halo mass or luminosity. It is clear that the majority of haloes with low $V_{\text{max,sub}}$ are removed by the cuts, leaving only the high-$V_{\text{max,sub}}$ haloes. The histogram does not depend on whether the sampling is ordered by mass or luminosity.

3.8.7 The Mocks

With the size of GiggleZ and 6dFGSv, we can generate 20 independent mocks using GiggleZ. We illustrate the sky distribution of one of these in Figure 3.19 along with that of the 6dFGSv sample for comparison, and a random mock using the selection
Figure 3.18: The relation between log $L$ and log $M$ from random $\{r, s, i\}$ parameters generated for 43,000 haloes.
function we developed above.

It is clear by eye that the clustering in the Millennium Simulation and GiggleZ mocks is still less than in the 6dFGSv data sample, and indeed we find the mocks do not reproduce the bias of $b \sim 1.7$ we find in the data, even after we have selected only the largest haloes in the simulation. This shows that selecting haloes based on mass alone is not sufficient to reproduce the high bias. The haloes that are selected for the sample have $v_{\text{max}} \gtrsim 300 \text{ km s}^{-1}$, which have a bias of $b \sim 1.2$. To achieve the number density of 6dFGS, we require $\sim 9000$ galaxies in half the sky out to a redshift of $z = 0.054$, a density of $n \sim 1 \times 10^{-3} h^{-3} \text{Mpc}^3$. However, to obtain a bias of $b \sim 1.7$ requires galaxies with a $v_{\text{max}} \gtrsim 400 \text{ km s}^{-1}$, but this only gives a number density of $2.5 \times 10^{-4} h^{-3} \text{Mpc}^3$ (which is consistent with LRGs).

A probable reason for the high bias of 6dFGSv is due to the morphological cut. Since it preferentially selects early-type galaxies, it will preferentially select galaxies in groups and clusters. Our simple method of populating the GiggleZ dark matter haloes with galaxies cannot account for the morphological selection. We plan to achieve this in future work by applying a halo occupation distribution (HOD) model to GiggleZ, such as that computed by Beutler et al. (2013). For the purposes of our present analysis however, the bulk flow properties of the mock catalogues do not depend strongly on the bias of the tracers, and so these low-bias mocks are sufficient.

We show the mean and variance of the redshift distribution of our 16 mocks, compared with 6dFGSv, in Figure 3.20. The data appears low in the highest redshift bins ($0.04 < z < 0.05$), although this could possibly be attributed to cosmic variance. Again, the large-scale bulk flow properties of the mocks will not depend strongly on the exact shape of the redshift distributions.

### 3.9 Results and Discussion

We present in this section the bulk flow results of our 6dFGSv analysis, for the MV and MLE estimators, along with the bulk flow results for our ΛCDM mocks. We then compare our results to a ΛCDM prediction, firstly by comparing the 3D bulk flow amplitude we measure with that predicted by ΛCDM, and secondly by doing a full comparison with each of the three 1D bulk flow components to obtain constraints.
Figure 3.19: **Top:** Sky distribution of full 6dFGSv dataset from Chris Springob (8986 galaxies), in Galactic longitude ($l$) and latitude ($b$), plotted in Mollweide projection. **Middle:** Clustered J-band mock (8901 galaxies), based on the GiggleZ simulation. Fields with completeness $0 > c > 60\%$ have been set to $c = 60\%$. **Bottom:** Random J-band mock (8901 galaxies). Fields with completeness $0 > c > 60\%$ have been set to $c = 60\%$. 116
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 H\(^{-1}\) MPC SCALES

Figure 3.20: The mean redshift distribution of our 20 GiggleZ mocks (black histogram), along with the standard deviation (blue histograms), compared to 6dFGSv (grey histogram).

...on \(\Omega_m\) and \(\sigma_8\). We also compare our method and results to that of Magoulas et al. (2014), and discuss the difference in our approaches to constraining the bulk flow.

3.9.1 Bulk flow results

We have calculated the bulk flow for 6dFGSv, for the two different bulk flow estimators described in Section 3.5:

1. The Minimum Variance (MV) estimate, using two different ideal surveys: (1) a Gaussian survey with effective radius \(R_I = 50 h\) Mpc, (2) a Gaussian survey of radius \(R_I = 70 h\) Mpc. To each ideal survey we apply a cut off at 160 \(h\) Mpc, the survey limit.

2. The Maximum Likelihood Estimate (MLE)

The results are presented in Table 3.3. We show both Galactic Cartesian coordinates and Equatorial Cartesian coordinates, since for the latter, having only the southern hemisphere will impact only the \(z\)-component (i.e. north-south direction), making any systematic effects clearer to distinguish. The uncertainties quoted are the
noise uncertainties, with the cosmic variance in parentheses. The cosmic variance is predicted for a given ΛCDM power spectrum, as we discuss further in Section 3.9.4.

Comparing the MLE and MV bulk flow components in Table 3.3 to those in Table 3.2 shows that estimating the velocity uncertainties as the linear fit $\sigma_n = 0.324H_0D_z$ removes the large, negative $z$-component of the bulk flow we found when using the standard deviation of $P(v)$. For the MV estimator with $R_I = 50h^{-1}$ Mpc, we find a bulk flow amplitude of $|U| = 259 \pm 54$ km s$^{-1}$ in the direction $(l, b) = (317 \pm 16^\circ, 35 \pm 11^\circ)$, and for $R_I = 70h^{-1}$ Mpc, we find a total bulk flow amplitude of $|U| = 254 \pm 54$ km s$^{-1}$ in the same direction.

For the MLE, we find a total bulk flow of $|U| = 280 \pm 46$ km s$^{-1}$ in the direction $(l, b) = (185 \pm 29^\circ, 7 \pm 9^\circ)$, which is not consistent with the direction of the MV results. The difference is largest in the Equatorial $z$ direction, and it is clear why from looking at the window function $W_{ii}^2$ of the different estimators, calculated from Eq. 5.35 in Figure 3.21. While the $x$ and $y$ window functions are similar for all the estimators, the $z$ window function is less compact for the MLE, giving more weight to smaller scales.

The MLE is much more sensitive to the window function of the survey than the MV, since the MV always upweights the same scale, while the scale of the MLE depends on the number of galaxies, their distribution, and their uncertainties. The survey covers only half the range of scales in the Equatorial $z$ direction (i.e. the north-south direction) than the $x$ and $y$ directions (i.e. east-west), and so smaller scales contribute to the MLE bulk flow in the $z$ direction. There is a lot of variance throughout the 6dFGSv velocity field, as shown by S14, so a difference in window function can give quite a large difference in the bulk flow, as we see.

We show the sky positions of our MV bulk flow measurements, using both ideal survey radii, in Figure 3.22. The measurement by WHF09 and the CMB dipole are shown for comparison. Our result is in a similar region of the sky to the WFH09 result. We also show the 6dFGSv result found by Magoulas et al. (2014), who measure the 6dFGSv bulk flow using an alternative method - doing a maximum likelihood fit to the bulk flow in logarithmic distance units $\eta$, by incorporating a model bulk flow into the FP fitting. Our results are consistent in direction with that of Magoulas et al. (2014), who find a direction of $(l, b) = (308 \pm 9^\circ, 20 \pm 11^\circ)$. Our amplitude is 1.4σ smaller, however; they find an amplitude of $|U| = 366 \pm$
### Table 3.3: Bulk flow results for the MV and MLE estimators, assuming peculiar velocity uncertainties $\sigma_n = 0.324 H_0$.

<table>
<thead>
<tr>
<th>Galactic coordinates</th>
<th>MV ($R_I = 50 h^{-1}$ Mpc)</th>
<th>MV ($R_I = 70 h^{-1}$ Mpc)</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>U</td>
<td>$</td>
<td>$u_x$</td>
</tr>
<tr>
<td>259 ± 54 (101)</td>
<td>155 ± 59 (105) &amp; -145 ± 65 (113) &amp; 149 ± 54 (102) &amp; 317 ± 16 &amp; 35 ± 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>254 ± 54 (102)</td>
<td>153 ± 59 (105) &amp; -143 ± 65 (112) &amp; 144 ± 54 (102) &amp; 317 ± 16 &amp; 35 ± 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>291 ± 39 (139)</td>
<td>66 ± 47 (135) &amp; 52 ± 43 (173) &amp; 279 ± 39 (136) &amp; 38 ± 29 &amp; 73 ± 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>291 ± 39 (139)</td>
<td>66 ± 47 (135) &amp; 52 ± 43 (173) &amp; 279 ± 39 (136) &amp; 38 ± 29 &amp; 73 ± 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>259 ± 54 (101)</td>
<td>155 ± 59 (105) &amp; -145 ± 65 (113) &amp; 149 ± 54 (102) &amp; 317 ± 16 &amp; 35 ± 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>254 ± 54 (102)</td>
<td>153 ± 59 (105) &amp; -143 ± 65 (112) &amp; 144 ± 54 (102) &amp; 317 ± 16 &amp; 35 ± 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>291 ± 39 (139)</td>
<td>66 ± 47 (135) &amp; 52 ± 43 (173) &amp; 279 ± 39 (136) &amp; 38 ± 29 &amp; 73 ± 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>291 ± 39 (139)</td>
<td>66 ± 47 (135) &amp; 52 ± 43 (173) &amp; 279 ± 39 (136) &amp; 38 ± 29 &amp; 73 ± 9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The bulk flow results are for the MV and MLE estimators, assuming peculiar velocity uncertainties $\sigma_n = 0.324 H_0$. The MV methods use an ideal Gaussian window function, with radius $R_I = 50 h^{-1}$ Mpc or $R_I = 70 h^{-1}$ Mpc, and with a cutoff at $160 h^{-1}$ Mpc. The uncertainties quoted are noise, with the cosmic variance uncertainty in parentheses.
3.9. RESULTS AND DISCUSSION

Figure 3.21: The window functions $W_{ii}^2$, from Eq. 5.35 of the bulk flow components for 6dFGSv, for each of our three estimators: the MV estimate with $R_I = 50\, h^{-1}\text{Mpc}$ or $R_I = 70\, h^{-1}\text{Mpc}$, and the MLE method. The $x, y, z$ components, here in Equatorial Cartesian coordinates, are the solid green, dashed blue, and solid red lines, respectively.

---

120
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 H⁻¹ Mpc SCALES

58 km s⁻¹. We will discuss the advantages and disadvantages of our two methods in Section 3.9.7.

We also show on this plot the position of the Shapley Supercluster (l = 312°, b = 31°). Our measurement is very close to the direction of Shapley, and consistent within the angular uncertainties. Unlike the WFH09 result, which probed the entire sky, 6dFGSv will be dominated by southern-sky structures, and so it is not surprising that our measurement is closer to Shapley. Also, the 6dFGSv number density of galaxies peaks beyond 100 h⁻¹ Mpc, incorporating part of Shapley, so this survey selection criteria itself will likely cause the Shapley region to dominate our bulk flow results.

In this figure, we also show the bulk flow results from our 20 GiggleZ-based mock catalogues, which we will discuss further in Section 3.9.2.

In Table 3.4, we list the 6dFSGv bulk flow results for peculiar velocity defined as the mean, median and ML of P(v) for each galaxy. Some differences are noticeable between the results for different definitions of v. The results for the median v are consistent between the methods, and similar to the result for the direct v in Table 3.3, which would be expected since it gives a similar estimate of v. However, the mean and ML v results show much more variation, and are not consistent within the uncertainties; this can also be explained by the bias in these definitions shown in Figure 3.9. We show these results here for completeness; however, they cannot be considered accurate without some form of correction.

3.9.2 6dFGSv bulk flow distribution in ΛCDM mocks

We can determine the expected distribution of bulk flows for 6dFSGv in a ΛCDM universe, given its window function, using our N-body simulation-based mock catalogues. We calculate the bulk flow amplitude of each of the 20 mocks, using the MV method with \( R_I = 50h^{-1} \) Mpc, and present their histogram in Figure 3.23. We also show the corresponding result from 6dFGS, along with its 1σ noise uncertainty, from Table 3.3. The result from the data is above average, but within the expected range of the mocks. Five of the mocks, or 25%, lie above our result, while 75% lie below.

The direction and amplitude of the bulk flow measured in each of these mocks is
Figure 3.22: The 6dFGSv bulk flow results in this work, compared with other bulk flow measurements and nearby superclusters. The figure shows Galactic longitude ($l$) and latitude ($b$), in an Aitoff projection. The diameter of the circles indicates the amplitude of the measured bulk flow; the diameters of the inner and outer circles show the $1\sigma$ confidence interval of the magnitude. The error bars show the $1\sigma$ angular uncertainty on the direction of the bulk flow. Our MV result for $R_I = 50h^{-1}$ Mpc is shown by the red circle, and our MV result for $R_I = 70h^{-1}$ Mpc is shown by the black circle (which mostly overlaps the red circle). Our result using the MLE method is shown in dark green. The result by Magoulas et al. (2014) is shown in blue, and is consistent with the direction of our measurement. We also show in orange the measurement by WFH09. The 6dFGSv galaxies are shown in grey for reference. The four largest local superclusters are shown in magenta: the Shapley Supercluster, Hydrus-Centaurus (HC), Horologium-Reticulum (HR) and Perseus-Pisces (PP). The South Pole is also shown in magenta for reference. The CMB dipole is indicated as the red and blue stars (with red the direction of the dipole), and the direction of the Local Group motion (from Kogut et al., 1993) is shown by the black star. The cyan circles show the distribution of bulk flows measured in our 20 GiggleZ 6dFGSv mock catalogues - since these are from simulations they have no measurement uncertainties.
Table 3.4: The bulk flow results for the MV and MLE estimators, using three different definitions of $v$: the mean, median and maximum likelihood (ML) of the $P(v)$ distributions. All results are in Equatorial Cartesian coordinates.

<table>
<thead>
<tr>
<th></th>
<th>$\parallel{50}$</th>
<th>$\parallel{70}$</th>
<th>$\parallel{110}$</th>
<th>$\delta_u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV</td>
<td>$-183 \pm 52$</td>
<td>$-147 \pm 52$</td>
<td>$76 \pm 52$</td>
<td>$-123 \pm 46$</td>
</tr>
<tr>
<td></td>
<td>$-195 \pm 55$</td>
<td>$-146 \pm 55$</td>
<td>$91 \pm 55$</td>
<td>$-123 \pm 55$</td>
</tr>
<tr>
<td></td>
<td>$-181 \pm 56$</td>
<td>$-126 \pm 56$</td>
<td>$137 \pm 56$</td>
<td>$-123 \pm 56$</td>
</tr>
<tr>
<td></td>
<td>$-191 \pm 59$</td>
<td>$-127 \pm 59$</td>
<td>$161 \pm 59$</td>
<td>$-123 \pm 59$</td>
</tr>
<tr>
<td></td>
<td>$-175 \pm 69$</td>
<td>$-83 \pm 69$</td>
<td>$58 \pm 69$</td>
<td>$-123 \pm 69$</td>
</tr>
<tr>
<td></td>
<td>$-180 \pm 71$</td>
<td>$-88 \pm 71$</td>
<td>$65 \pm 71$</td>
<td>$-123 \pm 71$</td>
</tr>
<tr>
<td>MLE</td>
<td>$-200 \pm 48$</td>
<td>$-217 \pm 48$</td>
<td>$594 \pm 48$</td>
<td>$-123 \pm 48$</td>
</tr>
<tr>
<td></td>
<td>$-179 \pm 55$</td>
<td>$-163 \pm 55$</td>
<td>$169 \pm 55$</td>
<td>$-123 \pm 55$</td>
</tr>
<tr>
<td></td>
<td>$-187 \pm 60$</td>
<td>$-172 \pm 60$</td>
<td>$275 \pm 60$</td>
<td>$-123 \pm 60$</td>
</tr>
<tr>
<td></td>
<td>$-193 \pm 65$</td>
<td>$-182 \pm 65$</td>
<td>$315 \pm 65$</td>
<td>$-123 \pm 65$</td>
</tr>
</tbody>
</table>

Maximum Likelihood Estimate

Mean

Median

ML

Any systematic bias in the polar direction will therefore mainly affect the z-components.

Table 3.4: The bulk flow results for the MV and MLE estimators, using three different definitions of $v$: the mean, median and maximum likelihood (ML) of the $P(v)$ distributions. All results are in Equatorial Cartesian coordinates.
Figure 3.23: Histogram of the bulk flow amplitudes $|U|$ in our 20 GiggleZ-based 6dFGSv mocks, for the MV estimator with $R_I = 50h^{-1}\text{Mpc}$. The vertical dashed line shows the corresponding value obtained for the data, using the direct $v$ definition of velocity, and the grey shaded area indicates the $1\sigma$ noise uncertainty in the measurement.

shown in Figure 3.22 This is a useful test to see whether the survey window function can bias the direction of the measured bulk flow. We see that the directions of the mocks appear fairly random and isotropic. There are more in the northern sky (13) than the southern sky (7), but this does not appear particularly significant. It is also possible the mocks may share large-scale modes, since they all lie within the same Gpc volume, so it could be they are not completely independent.

3.9.3 Comparison with linear theory: 3D bulk flow

Since the bulk flow amplitude is sensitive to the large-scale modes of the matter power spectrum, the measured bulk flow can be compared with the predicted value for a given cosmological model. If the Universe is statistically homogeneous and isotropic, then the expected mean bulk flow at any location is zero. The rms variance of the bulk flow amplitude, however, is cosmologically interesting, since it depends
on the power spectrum, as well as the scale and window function in which it is measured.

We compare our 6dFGSv MV bulk flow amplitude results, for both our ideal surveys, to a ΛCDM linear-theory prediction in Figure 3.24. This prediction is the most likely bulk flow amplitude, $V_{\text{ML}}(R)$, which depends on the rms velocity dispersion, $\sigma_V$. The rms velocity dispersion is given by

$$\sigma_V^2(R) \equiv \langle V(R)^2 \rangle = \frac{H_0^2 f^2}{2\pi^2} \int_{k=0}^{\infty} dk P(k) \tilde{W}(k; R)^2,$$

(3.30)

where $P(k)$ is the matter power spectrum, and $\tilde{W}(k; R)$ is the Fourier Transform of the window function, $W(R)$, at effective radius $R$. In this plot, we use an all-sky Gaussian window function, $\tilde{W}_G = \exp\left(-k^2 R^2 / 2\right)$.

The expected bulk flow velocity $V(R)$ can be predicted from $\sigma_V$, assuming the peculiar velocity field is Maxwellian, which it should be if the density field is Gaussian random. For a Maxwellian distribution, the probability distribution function of the bulk flow amplitude $V$ is (Bahcall et al., 1994; Coles & Lucchin, 1996)

$$p(V) dV = \sqrt{\frac{2}{\pi}} \left( \frac{3}{\sigma_V^3} \right)^{3/2} V^2 \exp\left( -\frac{3V^2}{2\sigma_V^2} \right) dV.$$  

(3.31)

For such a distribution the most likely (maximum likelihood) bulk flow amplitude is $V_{\text{ML}} = \sqrt{2/3} \sigma_V$, while the mean bulk flow amplitude is $\langle V \rangle = 2 V_{\text{ML}} / \sqrt{\pi} = \sqrt{8/3\pi} \sigma_V$.

In Figure 3.24 we plot $V_{\text{ML}}$ along with the upper and lower 1σ and 2σ confidence levels as the dark and light grey shaded regions, found from integrating Equation 3.31. These confidence levels correspond to the variance ranges $V_{\text{ML}} + 0.419 \sigma_V$ (1σ) and $V_{\text{ML}} + 0.891 \sigma_V$ (2σ). To calculate $\sigma_V$, we use a ΛCDM matter power spectrum, generated using CAMB (Lewis et al., 2000) with nonlinear evolution calculated using HALOFIT (Smith et al., 2003), and with the parameters listed in Section 3.2.

A caveat should be made for this plot, since comparing bulk flow results from surveys with such a prediction is problematic due to the window function of the surveys. Each of the surveys have their own way of defining the effective scale of the measurement given their window function, which may differ between surveys,
3.9. RESULTS AND DISCUSSION

Figure 3.24: The 6dFGSv bulk flow results in this work (red error bars) for the MV method with radius $R_I = 50$ and $70 \, h^{-1}\text{Mpc}$, compared to a $\Lambda$CDM prediction. These results are plotted at ‘effective radii’, corresponding to the radius of a full sphere with the same volume as the half-sky measurement, to show the variance we actually expect. The red arrows show how we have shifted the points from the measured radii $R_I$ to the effective radii $R_{\text{eff}}$. The black solid curve is a linear-theory $\Lambda$CDM prediction for an all-sky Gaussian window function. The dark grey and light grey regions show the 68.3% and 95.5% confidence levels, assuming a Maxwellian distribution of velocities. Other recent measurements are shown in in blue – these are: Lavaux et al. (2013) (L13), WFH09 (W09), Turnbull et al. (2012) (T12), Colin et al. (2011) (C11), Planck Collaboration et al. (2013b) (P13), Dai et al. (2011) (D11) and Magoulas et al. (2014) (M14). We also plot the M14 result at the effective radius of a full sphere. All error bars are $1\sigma$, while the two Planck arrows are the 95% upper limits.
and may not correspond to the position in which they are plotted here with respect to the theoretical model. A selection function tends to reduce the effective scale of a survey, which increases $\sigma_V$ and $V$ for that survey (Li et al. 2012). However, simulations show that the PDFs of bulk flows depend primarily on $\sigma_V$, and not on the type of window function, and so assuming an effective radius for the window function used in the model (e.g. a Gaussian in our case) that reproduces the same $\sigma_V$ as the survey window function should allow a comparison at that scale (Li et al. 2012). We have not done this in this plot, and note that there is some uncertainty on the true effective scale of the different surveys.

Since 6dFGSv only covers half the sky, we would expect our measurements at given radius $R_I$ to have more cosmic variance than predicted by the full-sky model at this radius. Conversely, we could consider 6dFGSv to be at a smaller effective radius. We therefore plot our 6dFGSv MV results at ‘effective radii’ $R_{\text{eff}}$ accounting for the fact that 6dFGSv covers only half the sky. For each of the $R_I = 50$ and $R_I = 70\,h^{-1}\text{Mpc}$ results, we calculate the radius of a full sphere with the same volume as the half-sky measurement, i.e.

$$R_{\text{eff}} = \left(\frac{R_I^3}{2}\right)^{1/3}.$$  \hfill (3.32)

This gives effective radii of $R_{\text{eff}} = (39.7, 55.6)\,h^{-1}\text{Mpc}$ for the $R_I = (50, 70)\,h^{-1}\text{Mpc}$ measurements. We plot arrows showing how we have shifted the measurements from the starting radius $R_I$ (the start of the arrows) to the new effective radii (the tips of the arrows). However, since 6dFGSv is not a perfectly sampled hemisphere, we might expect the effective radii to be even smaller than the $R_{\text{eff}}$ we calculate.

From Figure 3.24 we see that once shifted to the effective radii, both the 6dFGSv $R_I = 50\,h^{-1}\text{Mpc}$ and $R_I = 70\,h^{-1}\text{Mpc}$ bulk flow results appear to be consistent within 68.3% confidence with the theoretical prediction.

The uncertainty on the effective radii of previous surveys may mean that those that showed higher than predicted bulk flows could have been comparing to theory at too large a radius, without accounting for how the window function reduced the effective volume of the survey. It would be illuminating to recalculate the effective radii of these surveys to investigate this; we leave this for future work.
3.9. RESULTS AND DISCUSSION

3.9.4 Comparison with linear theory: 1D bulk flow

Unlike the 3D bulk flow amplitude, the 1D bulk flow components should be Gaussian-distributed, making them more useful for a robust test of ΛCDM. The 1D rms velocity variance is given for a particular survey by the covariance matrix of the bulk flow moments, $R_{ij} = \langle u_i u_j \rangle$:

$$R_{ij} = \sum_n w_{i,n} w_{j,n} (\sigma_n^2 + \sigma_*^2) + \frac{f^2 H_0^2}{2\pi^2} \int_0^\infty dk W_{ij}^2(k) P(k). \quad (3.33)$$

The first term is the shot noise component for the survey, with $w_{i,n}$ the weight for the $n^{th}$ galaxy in the $i^{th}$ direction (either a MLE or MV weight in our case), $\sigma_n$ the velocity measurement uncertainty, and $\sigma_*$ is the small-scale velocity dispersion. The second term is the cosmic variance component, with $f \sim \Omega_m^{0.55}(z)$ the growth rate of structure, $P(k)$ the matter power spectrum, and $W_{ij}(k)$ is the tensor window function for the $i^{th}$ and $j^{th}$ bulk flow components. This therefore depends on the survey geometry, the measurement noise, and the matter power spectrum. It is very similar to the rms velocity variance $\sigma^2_V$ in Equation 3.30 except for the addition of the noise component, and the cosmic variance component $R_{ij}^{(c)}$ contains the tensor window function $W_{ij}(k)$. (We previously defined $\sigma_*$ as the 1D velocity variance; this is in principle the average variance over all scales, which we assumed to be equal to $\sim 250 \text{ km s}^{-1}$. Here, however, we are looking at the variance as a function of scale.)

The deviation from zero of the observed bulk flow components $u_i$ can be directly compared with the predicted dispersion, by calculating the $\chi^2$ for the three moments, given by

$$\chi^2 = \sum_{i,j} u_i R_{ij}^{-1} u_j, \quad (3.34)$$

where $i$ and $j$ both go from 1 to 3 to specify the bulk flow components, $u_i$ and $u_j$ are the measured bulk flow components, and $R_{ij}$ is the covariance matrix of the moments for a specified set of cosmological parameters.

The $R_{ij}$ matrix is dominated by the cosmic variance term (typically of order $\sim 100 \text{ km s}^{-1}$, while the noise term is typically $\sim 40 \text{ km s}^{-1}$). Since the bulk flow depends on large-scale density fluctuations, $R_{ij}$ will be most sensitive to the amplitude and shape of the power spectrum. The power spectrum amplitude is parameterised
by the rms density fluctuations in spheres of $8h^{-1}$ Mpc radius, $\sigma_8$, while the shape is parameterised by the shape parameter, $\Gamma$, which on large scales can be approximated by $\Gamma = \Omega_m h$. The dependence on $\Omega_m$ also comes into the $f(\Omega_m)^2$ factor. We therefore follow Watkins et al. (2009) in using the bulk flow to constrain a combination of $\Omega_m$ and $\sigma_8$ – in our case, we fix $h$ to the best-fit value from Planck, $h = 0.67$.

In Figure 3.25 we present the constraints obtained for our MV 6dFGSv bulk flow measurements, with $R_I = 50$ or $R_I = 70h^{-1}$ Mpc, compared to our fiducial Planck parameters. There is a degeneracy between $\sigma_8$ and $\Omega_m$, since a lower $\sigma_8$ requires a lower $\Omega_m$ to produce the same bulk flow; or, for fixed $\sigma_8$, lower values of $\Omega_m$ lead to a larger bulk flow. This is because if $\sigma_8$ is fixed, then a lower $\Omega_m$ requires a larger power spectrum amplitude on large scales to allow this normalisation. However, since a lower $\Omega_m$ also decreases the growth rate $f(\Omega_m)$, these two effects partially cancel, and so the bulk flow does not have much constraining power on $\Omega_m$.

Our measurement is much more sensitive to $\sigma_8$, but it can only give a lower bound, since for any $\sigma_8$ there is no lower bound on the expected bulk flow. The bulk flow measurement is larger than the $\Lambda$CDM prediction, and clearly favours a high $\sigma_8$, though it is consistent with the Planck value within 2$\sigma$. For the Planck parameters, we find $\chi^2 = 4.23$ for our MV $R_I = 50$ measurement, and $\chi^2 = 5.21$ for our MV $R_I = 70$ measurement, for 3 degrees of freedom.

WFH09 find that their bulk flow measurement requires a very high $\sigma_8$ of $\sim 1.7$, with lower 95 and 99 per cent limits of 1.109 and 0.878, respectively. To compare to their result, following their method we calculate the likelihood of $\sigma_8$, fixing all other parameters to their Planck values, using

$$L(\Theta) \propto \frac{1}{\sqrt{|R|}} \exp \left( \sum_{i,j} -\frac{1}{2} u_i R^{-1}_{ij} u_j \right),$$

(3.35)

where $\Theta$ is a set of parameters, in our case $\sigma_8$. This is similar to Equation 3.34 except that it is normalised by the factor $\ln |R|$, which down-weights models with high cosmic variance.

The results are shown in Figure 3.26. Our results favour a high value of $\sigma_8$, but we do not find a significant disagreement with $\Lambda$CDM. For the MV $R_I = 50$ measurement, we find $\sigma_8 = 1.14^{+1.23}_{-0.51}$ (68.27% C.L.), and for $R_I = 70$, we find
3.9. RESULTS AND DISCUSSION

Figure 3.25: The $\chi^2$-based confidence levels on $\Omega_m$ and $\sigma_8$, obtained from the 6dFGSv bulk flow measurement. Top: for the MV bulk flow with $R_I = 50\, h^{-1}\, \text{Mpc}$. Bottom: for the MV bulk flow with $R_I = 70\, h^{-1}\, \text{Mpc}$. The cross indicates the best fit values found by Planck, used as the fiducial values in this work.
The diagonal elements of $R_{ij}$ provide the expected 1D rms bulk flow variance $\sigma_{V,i}^2$ in each of the three directions $i$, for a given survey window function and input model power spectrum. Following Watkins et al. (2009), we list $\sigma_{V,i} = \sqrt{R_{ii}}$ in Table 3.5 for 6dFGSv, for the MLE and MV estimators.

We also show in this table the 1D and 3D rms velocities calculated from our 20 GiggleZ-based mocks, using the MV estimator with $R_I = 50 \, h^{-1} \text{Mpc}$. We would expect these to closely agree with the 6dFGSv results for $R_I = 50$, since the mocks should reproduce the window function of the data, and this is roughly true. The last two rows of Table 3.5 show the analytic prediction for an all-sky Gaussian window function with radius 50 or 70 $h^{-1}$ Mpc, by evaluating Equation (3.30) at these values of $R$. Since the volume of such a window function is a factor of 2 larger than 6dFGSv, we would expect its 3D velocity dispersion to be smaller than for 6dFGSv; however, it is only slightly smaller. It is also larger, surprisingly, than for our half-sky mocks.
3.9. RESULTS AND DISCUSSION

Table 3.5: Comparison of the expected 1D rms velocity \( \sigma_{V,i} \), and 3D rms velocity \( \sigma_V \), for 6dFGSv, our GiggleZ-based 6dFGSv mocks, and for theory. We use Equatorial Cartesian coordinates, and assume a ΛCDM cosmology with parameters listed in Section 3.2.

<table>
<thead>
<tr>
<th>Source &amp; ( R_I ) (h(^{-1}) Mpc)</th>
<th>( \sigma_{V,x} ) (km s(^{-1}))</th>
<th>( \sigma_{V,y} ) (km s(^{-1}))</th>
<th>( \sigma_{V,z} ) (km s(^{-1}))</th>
<th>( \sigma_V ) (km s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>6dFGSv(^a)</td>
<td>MLE (( R_I \sim 70 ))</td>
<td>122</td>
<td>122</td>
<td>193</td>
</tr>
<tr>
<td>MV, ( R_I = 50 )</td>
<td>95</td>
<td>100</td>
<td>122</td>
<td>101</td>
</tr>
<tr>
<td>MV, ( R_I = 70 )</td>
<td>95</td>
<td>100</td>
<td>123</td>
<td>102</td>
</tr>
<tr>
<td>Mocks(^b)</td>
<td>MV, ( R_I = 50 )</td>
<td>129</td>
<td>114</td>
<td>120</td>
</tr>
<tr>
<td>Theory(^c)</td>
<td>( W_G, R_I = 50 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( W_G, R_I = 70 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>177</td>
</tr>
</tbody>
</table>

\(^a\)\( \sigma_{V,i} = \sqrt{R_{ii}} \), from Eq. 5.36, and \( \sigma_V^2 = J R_{ij} J^T \), where J is the Jacobian (see Section C.5 for details). Includes both noise and cosmic variance.

\(^b\)All calculated from root mean square of bulk flow components of 20 mocks. Includes both noise and cosmic variance.

\(^c\)Calculated from Eq. 3.30. Includes noise only and assumes a full-sky window function.

of the same radius. In future work we will investigate the velocity dispersion in the mocks further - it could be it is underestimated since the 20 mocks are all in the same Gpc\(^3\) volume, and so may share large-scale modes.

Since the sky sampling is worst in the north-south direction, we expect the largest dispersion in the \( z \) direction, which is what we see for the data. However, for the mocks, the largest dispersion is actually in the \( x \) direction. This could again be due to the proximity of the mocks within the GiggleZ box, since they are packed closely in the \( z \) direction and so may share more modes in this direction.

3.9.5 Bulk flow in redshift shells

It is interesting to look at how the bulk flow varies as a function of redshift. In Figure 3.27 we plot the MLE bulk flow, split into redshift bins of \( \Delta z = 0.01 \). The results are noisy, but the amplitude of the bulk flow seems to be fairly constant up to the maximum redshift of \( z = 0.054 \). This is what would be expected if the source of the bulk flow is an overdensity more distant than the scales measured.
Figure 3.27: The MLE bulk flow for 6dFGSv in redshift shells of width $\Delta z = 0.01$. The coloured lines show the Equatorial $(x, y, z)$ components (labelled), while the black line shows the bulk flow magnitude. The number of galaxies in each redshift shell is $\{75, 813, 1371, 2563, 2802, 1261\}$. The error bars indicate the noise uncertainty in each bin.
3.9. RESULTS AND DISCUSSION

3.9.6 Comparison with other results

We compare our bulk flow result to other recent measurements in the literature, in Table 3.6. Our result is one of the most precise to date, thanks to the large number of galaxies in 6dFGSv. Our MV result of $259 \pm 54 \text{ km s}^{-1}$ at $R_I = 50 h^{-1} \text{ Mpc}$ is a significantly lower amplitude than that of WFH09 at the same scale, despite the fact that the 6dFGSv survey volume is smaller than the COMPOSITE sample that they use, and so might be expected to have more cosmic variance. The level of disagreement between our result and WFH09, not accounting for this volume difference, is $1.56\sigma$. Our measurement also does not appear to support the high-redshift $600 - 1000 \text{ km s}^{-1}$ measurement of Kashlinsky et al. (2008), although since their scale is much larger we cannot directly rule it out.

Our result is consistent with a growing number of recent measurements that find a lower bulk flow amplitude on scales of around $50 - 70 h^{-1} \text{ Mpc}$, including Colin et al. (2011), Dai et al. (2011), Nusser & Davis (2011), Turnbull et al. (2012), and Feindt et al. (2013). Our direction is also consistent with the directions found by other studies.

We note again that the different surveys quoted in this table all have different window functions, so even those at the same effective distance may not be directly comparable.

3.9.7 Comparison with method of Magoulas et al., (2014)

We also show in Table 3.6 the result of Magoulas et al. (2014), who measure the bulk flow of 6dFGSv using a different method to us. They perform a Maximum Likelihood analysis using the measured logarithmic distance ratio $\eta = \log D_z/D_r$, without converting to linear velocities, and they constrain the bulk flow by including a model bulk flow in the fitting of the FP.

Their approach has the advantage of being entirely in logarithmic units, for which the uncertainties are Gaussian. Since we convert to linear velocities, which have lognormal uncertainties, this could potentially create a systematic in our method which they should avoid. By re-fitting the FP for each model bulk flow, they also retain the full information of the selection cuts. On the other hand, our use of the Minimum Variance estimator allows us to account for the survey window function,
Table 3.6: Summary of some recent bulk flow results in the literature, compared to the result in this work. For each measurement, we list the distance indicator used (DI), the number of peculiar velocities in the sample $N$, the radius of the measurement $R$, the measured bulk flow magnitude $|U|$, and the direction of the bulk flow in Galactic longitude $l$ and latitude $b$. A dash for the DI means a combination of datasets were used – these results all used the COMPOSITE sample. For measurements of the kSZ effect, $N$ shows the number of clusters used in combination with the CMB. A number of these results use the same, or overlapping, datasets, but apply different analyses, and the window functions differ for each survey.

| DI                      | N     | R (h$^{-1}$Mpc) | |U| (km s$^{-1}$) | l  | b   |
|-------------------------|-------|-----------------|-----------------|-----------------|-----|-----|
| 6dFGSv (this work)      | FP    | 8885            | 50              | 259 ± 54        | 317 ± 16 | 35 ± 11 |
|                         | FP    | 8885            | 70              | 254 ± 54        | 317 ± 16 | 35 ± 11 |
| 6dFGSv (M14)            | FP    | 8885            | ~ 70            | 366 ± 58        | 308 ± 9  | 20 ± 11 |
| Dressler et al. (1987a) | FP    | 423             | ≲ 60            | 599 ± 104       | 312 ± 11 | 6 ± 10  |
| Watkins et al. (2009)   | -     | 4481            | 50              | 407 ± 81        | 287 ± 9  | 8 ± 6   |
| Feldman et al. (2010)   | -     | 4536            | 50              | 416 ± 78        | 282 ± 11 | 6 ± 6   |
| Macaulay et al. (2012)  | -     | 4537            | 33              | 380$^{+99}_{-132}$ | 295 ± 18 | 14 ± 18 |
| Nusser & Davis (2011)   | TF    | 2859            | 40              | 333 ± 38        | 276 ± 3  | 14 ± 3  |
| Colin et al. (2011)     | SNe   | 142             | 160             | 260 ± 150       | 298$^{+62}_{-48}$ | 8$^{+34}_{-52}$ |
| Dai et al. (2011)       | SNe   | 132             | 150             | 188$^{+199}_{-103}$ | 290$^{+39}_{-31}$ | 20$^{+32}_{-32}$ |
| Turnbull et al. (2012)  | SNe   | 254             | 50              | 249 ± 76        | 319 ± 18 | 7 ± 14  |
| Feindt et al. (2013)    | SNe   | 128             | 74              | 243 ± 88        | 298 ± 25 | 15 ± 20 |
| Weyant2011              | SNe   | 30              | 112             | 446 ± 101       | 273 ± 11 | 46 ± 8  |
| Kashlinsky et al. 2008  | kSZ   | 782             | ~ 300-800       | ~ 600-1000      | 283 ± 14 | 12 ± 14 |
| Planck Collab. (2013)   | kSZ   | 1405            | 350             | < 390 (95% CL)  | 254 ± 14 | 12 ± 14 |

$^a$Magoulas et al. (2014)

$^b$The result for their lowest redshift shell
3.9. RESULTS AND DISCUSSION

and to specify the scale on which we measure the bulk flow, which should make it easier to directly compare our result to a theoretical model. As previously explained, the MV method is less sensitive to the positions and uncertainties of individual galaxies. We note that our MV results are weighted to a Gaussian radial profile, while the result of Magoulas et al. (2014) has a window function closer to a top-hat.

We also do not necessarily expect our MLE result to be the same as theirs, since we convert to linear velocities, and use a different method of estimating the velocity uncertainty contribution to the MLE weights, meaning our MLE bulk flows can potentially have slightly different window functions.

Despite these differences, we see in Figure 3.22 that our result is consistent in direction to theirs.

3.9.8 Implications for Cosmography

An important aim for bulk flow measurements has been to understand the motion of the Local Group (LG) with respect to the CMB, of \(627 \pm 22\) km s\(^{-1}\) towards \(l = 276 \pm 3^\circ\), \(b = 30 \pm 2^\circ\) (Kogut et al., 1993). From gravitational instability theory, this is expected to be caused by nearby structures, and to converge to the CMB dipole beyond them.

As we showed in Figure 3.22, the direction of our bulk flow is consistent with the direction of the Shapley Supercluster. We also saw in Figure 3.27 that the amplitude of the bulk flow remains fairly constant with distance, indicating that it is sourced by a distant rather than a nearby overdensity. This therefore seems to indicate that Shapley is the dominant source of the bulk flow motion we detect. Shapley is at a distance of \(152 h^{-1}\) Mpc, and is the largest supercluster in the local Universe out to \(200 h^{-1}\) Mpc (Lavaux & Hudson, 2011). Our result is consistent with many other bulk flow measurements that find directions close to Shapley (e.g. Feindt et al., 2013) and a source distance greater than \(\sim 50 - 80 h^{-1}\) Mpc as the origin of the flow (e.g. Hudson, 1994; Kocevski et al., 2004; Pike & Hudson, 2005; Watkins et al., 2009).

Lavaux & Hudson (2011) calculate, using linear theory applied to 6dFGS redshift data, that Shapley should be responsible for \(\sim 15\) per cent of the total velocity of the LG with respect to the CMB, or \(90 \pm 10\) km s\(^{-1}\), while the Horologium-Reticulum supercluster should generate \(\sim 60\) km s\(^{-1}\). However, it appears that our sample is
3. THE 6DF GALAXY SURVEY: THE MINIMUM VARIANCE BULK FLOW ON 50 – 70 $h^{-1}$ MPC SCALES

dominated mostly by Shapley. This makes it possible that its mass could be even larger than inferred from redshift data alone, which would agree with the finding of Feindt et al. (2013), who find that the bulk flow does not appear to reverse beyond Shapley, suggesting there could be more mass beyond it sourcing the bulk flow. They calculate that their bulk flow would be caused either if the mass of Shapley were twice as large as current estimates (from Muñoz & Loeb, 2008; Sheth & Diaferio, 2011), or if there were a more distant mass behind Shapley.

As we have previously noted however, the 6dFGSv sample extends beyond 100 $h^{-1}$ Mpc and partially samples the Shapley region, with no sampling at all of northern-sky structures, so this sampling could be partially responsible for Shapley dominating our results. More analysis would therefore be needed to confirm whether the bulk flow is truly closer to Shapley than any other structure.

3.9.9 Cosmological implications

Our 6dFGSv bulk flow is, as we have seen, larger than predicted by cosmic variance in ΛCDM, although still consistent within the 95.4% C.L. Our measured value of $\sigma_8 = 1.14^{+1.23}_{-0.51}$ and $\sigma_8 = 1.11^{+1.21}_{-0.50}$ at 50 and 70 $h^{-1}$ Mpc are also larger than the currently favoured value from Planck of $\sigma_8 = 0.83$, though consistent within 1σ. Our results do not appear to support the large bulk flows of WFH09 and Kashlinsky et al. (2008). Nonetheless, our results, and others which have found a bulk flow larger than the ΛCDM prediction, seem to indicate that we are in an unusual part of the Universe, with a high local bulk flow.

Our bulk flow appears to be a consequence of our proximity to the Shapley Supercluster, and the question may be reversed from the likelihood of the bulk flow, to the likelihood of being near a concentration the size of Shapley. Sheth & Diaferio (2011) use extreme value statistics to show that a Shapley-size cluster is not unexpected within any given 200 $h^{-1}$ Mpc radius, if the initial fluctuation field was Gaussian. This might suggest that such statistics should be applied to the bulk flow, and may further alleviate the apparent tension with ΛCDM that many peculiar velocity measurements have found.

There are also more exotic explanations of the bulk flow that go beyond ΛCDM. One of these is a ‘tilted Universe’ scenario (Turner 1991; Kashlinsky et al. 2008; Ma...
et al. (2011), in which pre-inflationary fluctuations cause an offset between the CMB frame and the Hubble expansion frame. This means that some part of the CMB dipole is caused by this offset, rather than local peculiar motions, and would result in an intrinsic bulk flow that does not converge in the CMB frame. Our measurement cannot constrain this directly, since it only probes scales up to $160 \, h^{-1} \text{Mpc}$, but the fact that we see a fairly constant MLE bulk flow in shells up to this distance would be consistent with the theory. However, the recent bulk flow measurement by the Planck team (Planck Collaboration et al. (2013b)), which measured the kSZ velocities of clusters at distances up to $2 \, h^{-1} \text{Gpc}$, did not find evidence of a bulk flow up to that scale, contradicting the result of Kashlinsky et al. (2008) and diminishing the possibility of a tilted universe.

Another suggestion by Wiltshire et al. (2013) is that the differential expansion of space due to the density gradients of nonlinear structure creates different frames for peculiar velocities in different parts of the sky. This would challenge the standard assumption of flat space with a defined CMB rest frame, in which peculiar velocities are interpreted. Further theoretical and observational work is required to determine whether this is a valid effect for the local velocity field.

### 3.10 Conclusion

The question of whether a large bulk flow exists in the local universe is still of much interest, and different surveys continue to give conflicting results. A good part of the disagreement can be assumed to be due to the noisy, sparse peculiar velocity samples to date, as well as possible unknown systematics including Malmquist bias and the combining of uncalibrated datasets. Our aim has been to shed light on this debate by making use of a large new peculiar velocity dataset, the 6 degree Field Galaxy Survey peculiar velocity sample, 6dFGSv. This sample is homogeneously selected, so avoids any bias from combining datasets, and the uncertainties and Malmquist biases have been carefully studied and accounted for (M12, S13). It is short the largest and most homogeneously selected peculiar velocity sample to date.

We have presented a new bulk flow analysis using this dataset. Using the ‘Minimum Variance’ bulk flow estimator, we find a bulk flow of magnitude $|U| = 259 \pm 54 \text{ km s}^{-1}$ in the direction $(l, b) = (317 \pm 16^\circ, 35 \pm 11^\circ)$ at a distance of
50 $h^{-1}$ Mpc, and $|U| = 254 \pm 54 \text{ km s}^{-1}$ in the same direction at a distance of 70 $h^{-1}$ Mpc. This is somewhat higher than the ΛCDM prediction on these scales, indicating a high value of $\sigma_8$, but consistent with Planck results within $2\sigma$. After marginalising over $\Omega_m$, from our two bulk flow measurements we find $\sigma_8 = 1.14_{-0.51}^{+1.23}$ and $\sigma_8 = 1.11_{-0.50}^{+1.21}$ respectively, consistent with the Planck value within 68.27% confidence.

Our result is in agreement with a number of recent measurements that also find a bulk flow consistent with ΛCDM, including Hong et al. (2014), Turnbull et al. (2012) and Feindt et al. (2013). Our result is also supported by the higher-redshift measurement of Planck Collaboration et al. (2013b), who used Planck CMB data combined with a large X-ray cluster catalogue, and found no evidence for a bulk flow from 350 $h^{-1}$ Mpc to 2 $h^{-1}$ Gpc scales.

A major challenge for the 6dFGSv analysis (and indeed, for any Fundamental Plane or Tully-Fisher sample) is accounting for the large, non-Gaussian uncertainties on the peculiar velocities. When combined with the fact that 6dFGSv only covers half the sky, these can result in a spurious polar bulk flow component if not properly accounted for. We have shown that it is important to propagate uncertainty from the Gaussian observable (in our case, the logarithmic distance ratio $\eta = \log_{10} D_z/D_r$) to the non-Gaussian velocity in a way that is independent of the $\eta \rightarrow v$ conversion itself, otherwise the bulk flow weights can correlate with the velocities. This would occur if the standard deviation $\sigma_{SDv}^n$ of the velocity probability distributions $P(v)$ was used as the velocity uncertainty. We instead make a linear fit to $\sigma_{SDv}^n$ with distance, and use this to estimate the velocity uncertainties. We recommend that such an approach be used by any other peculiar velocity dataset with large uncertainties, e.g. using the Tully-Fisher or Fundamental Plane relations.

Our measured bulk flow is very close to the direction of the Shapley Supercluster, consistent with many other measurements, and its amplitude appears to be fairly constant out to the distance of Shapley. This indicates that a large part of the bulk flow we measure is likely to sourced by Shapley.

We have also generated a set of ΛCDM mock catalogues of 6dFGSv, based on the GiggleZ $N$-body simulation, to be used for testing systematic biases in the dataset. These will be made available to other researchers for other analyses of the 6dFGSv sample.
4

Bulk Flow estimation for the SkyMapper Supernova and Transient Survey (SMT) and general survey optimisation

Scrimgeour, M. I. et al., in preparation

4.1 Abstract

The bulk flow, that is the dipole of the peculiar velocity field, is a sensitive probe of the amplitude of matter fluctuations on large scales, and allows for low-redshift tests of the standard cosmological model, ΛCDM. The bulk flow is currently of particular interest as some recent measurements have found it to be larger than predicted in ΛCDM. Type Ia Supernovae (SNe Ia) are currently the most accurate distance estimator that can be used to measure individual peculiar velocities. The SkyMapper Transient and Supernova Survey, running on the SkyMapper telescope at Siding Springs Observatory, aims to obtain accurate lightcurves for ∼ 420 low-redshift SNe Ia over five years. We have made simulations of this survey in order to predict the statistical accuracy with which a bulk flow measurement can be made. We also simulate more general supernova surveys, to provide guidelines for optimi-
sation of the bulk flow measurement. We have investigated the effect of survey field distribution, magnitude error and number of SNe Ia on the accuracy of the bulk flow measurement. We find that SkyMapper, in its current predicted configuration, should measure the bulk flow on a scale of $150 h^{-1}$ Mpc with a $\sim 90$ km s$^{-1}$ noise uncertainty, which would see an 18% improvement if the fields were distributed randomly over the whole southern sky, and a 26% improvement if the fields were distributed randomly over the entire sky. Alternatively, if the predicted distance modulus uncertainty $\Delta m$ for the SkyMapper SNe Ia was decreased from 0.14 to 0.10, the bulk flow noise uncertainty would decrease by 16%. Increasing the number of SNe Ia in the sample to 1000 would give a 20% improvement. Even in its most basic configuration, SkyMapper will provide a competitive constraint on the local bulk flow, and would be even more powerful in combination with a northern-hemisphere supernova survey.

4.2 Introduction

SkyMapper (Keller et al., 2007) is one of a new generation of dedicated, wide-field optical survey telescopes, and is located at Siding Spring Observatory near Coonabarabran, NSW. The primary role of SkyMapper is to conduct the Southern Sky Survey (S3), a multi-colour, multi-epoch photometric survey covering 20 000 deg$^2$ of the southern sky ($-90^\circ < \text{Dec} < 0^\circ$), to a depth comparable to the Sloan Digital Sky Survey. Part of the 25 per cent of time not dedicated to S3 (primarily the bad-seeing time) will focus on a rolling search for SNe Ia. This SkyMapper Transient and Supernova Survey (SMT) will provide continuous coverage of $\sim 1000$ deg$^2$ of sky, with a sampling rate of one observation every 3-4 days. It is expected to discover $\sim 100$ SNe Ia per year out to $z \lesssim 0.085$ within a total sky area of $\sim 5000$ deg$^2$.

Although the primary aims of the SMT are to obtain a low-redshift sample for investigating statistical properties of SNe Ia and to populate the low-redshift Hubble diagram, this sample can also be used to measure the peculiar velocities of the supernova host galaxies, when combined with a spectroscopic redshift for the host.

Peculiar velocities are the motions of galaxies caused by gravity, separate from the isotropic Hubble expansion. They are usually measured statistically via redshift-
4.2. INTRODUCTION

space distortions (Kaiser, 1987; Peacock et al., 2001; Tegmark et al., 2004; Guzzo et al., 2008) but can also be measured directly. The line-of-sight component of the peculiar velocity $v$ of a galaxy at position $\mathbf{r}$ is given by

$$v_p = c \left( \frac{z_{\text{obs}} - z_r}{1 + z_r} \right),$$  \hspace{1cm} (4.1)

where $c$ is the speed of light, $z_{\text{obs}}$ is the observed redshift, measured spectroscopically and corrected to the cosmic microwave background (CMB) rest frame, and $z_r$ is the redshift corresponding to the real comoving distance $r$.

Type Ia supernovae, as distance indicators, provide a redshift-independent measurement of $r$, allowing the peculiar velocity to be derived. Most distance indicators have large uncertainties, such as the Tully-Fisher relation (Tully & Fisher, 1977) with $\sim 20\%$ uncertainty and the Fundamental Plane relation (Djorgovski & Davis, 1987; Dressler et al., 1987b) with $\sim 30\%$ uncertainty. The uncertainty on $S$ is therefore almost always dominated by the uncertainty on $r$. With a distance accuracy of $5 - 6\%$, SNe Ia are the most accurate known distance indicator and hence the best available tool for measuring individual peculiar velocities\footnote{Surface brightness fluctuations provide a distance indicator of similar accuracy to SNe Ia; however, they are limited to very low redshift, $z \lesssim 0.03$ (Blakeslee et al., 1999), although they can be extended to larger redshifts if observed from space.}. However, SNe Ia are rare, and thus harder to obtain large samples than galaxies for the Tully-Fisher and Fundamental Plane. An important aim in making better peculiar velocity measurements must be to focus on obtaining larger SNe Ia samples.

Observations of the peculiar velocity field $\mathbf{v}(\mathbf{r})$ are useful, since in the linear regime $\mathbf{v}(\mathbf{r})$ is directly related to the density field $\delta(\mathbf{r})$, via (Peebles, 1980)

$$\mathbf{v}(\mathbf{r}) = \frac{H_0 f}{4\pi} \int d^3\mathbf{r}' \frac{\delta(\mathbf{r}') (\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3},$$  \hspace{1cm} (4.2)

where $f \equiv d \ln D / d \ln a$ is the present-day growth rate of cosmic structure, in terms of the linear growth factor $D$ and cosmic scale factor $a$, and $\delta(\mathbf{r}) = (\rho(\mathbf{r}) - \bar{\rho})/\bar{\rho}$ with $\bar{\rho}$ the average density of the Universe. Peculiar velocity measurements therefore allow us to trace the total matter distribution, including dark matter, without the complication of galaxy bias, and over a large range of scales.
4. BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION

There are different ways of using peculiar velocity information to test the cosmological model. A commonly used statistic is the dipole, or 'bulk flow' of the velocity field, which corresponds to the average velocity of galaxies within a given volume with respect to the CMB.

Bulk flow measurements have seen much interest in the literature, but have also produced conflicting results. In the past few years, several studies have claimed to find an unexpectedly large bulk flow in conflict with ΛCDM predictions, on ∼ 50 – 150 h⁻¹ Mpc scales in the direction l ∼ 300°, b ∼ 10° [Watkins et al., 2009; Feldman et al., 2010; Abate & Feldman, 2012; Lavaux et al., 2013]. Others find a smaller bulk flow amplitude consistent with ΛCDM [Colin et al., 2011; Nusser & Davis, 2011; Osborne et al., 2011; Dai et al., 2011; Turnbull et al., 2012; Ma & Scott, 2013; Planck Collaboration et al., 2013b]. Kashlinsky et al. (2008) claimed to find a large dipole in the WMAP kinetic Sunyaev-Zel’dovich (SZ) effect of 600-1000 km s⁻¹, although Keisler (2009) cast doubt on the significance of their result.

If real, the existence of a large bulk flow would be in conflict with most other cosmological probes, which support ΛCDM. In particular, it would imply a larger σ₈ than measured by the CMB [Komatsu et al., 2011; Planck Collaboration et al., 2013a]. Since it would imply the existence of large density fluctuations on scales larger than 100 h⁻¹ Mpc it would also contradict measurements of cosmic homogeneity above this scale (Hogg et al., 2005; Scrimgeour et al., 2012). One of the aims of a SkyMapper bulk flow measurement is to provide a new test of the local bulk flow with a larger, more homogeneous supernova sample than has been previously used and test the apparent conflict.

A particularly interesting result is that of Watkins et al. (2009) (hereafter WFH09), who used the largest peculiar velocity sample to date (COMPOSITE, a compilation of several different surveys with a total of 4481 peculiar velocities), along with an optimised weighting method. They found a bulk flow of 407 ± 81 km s⁻¹ in the direction l = 287° ± 9°, b = 8° ± 6°, on a scale of 50 h⁻¹ Mpc, which they claim is inconsistent with the ΛCDM linear theory prediction at the 98% confidence level. Their result was disputed by Ma & Scott (2013), who questioned their method of combining the different surveys, as well as their assumed value of small-scale velocity dispersion. However, for our analysis, we use this measurement as a case study, and find the constraints that SkyMapper would measure if the WFH09 bulk flow

143
measurement were true.

The window function of the supernova sample is crucial in determining how well the peculiar velocities will do at measuring the bulk flow. Since the bulk flow is the dipole of the velocity field, to measure it requires covering as much of the sky as possible, with close to uniform sampling and a large number of objects. Such samples are difficult to obtain with SN Ia surveys, so we are interested to find out firstly how well a given planned SN Ia survey that is not optimised for a bulk flow measurement will do, and secondly how such a survey can be optimised to improve the bulk flow measurement. Previous work suggests that significant constraints can be made by peculiar velocity surveys even without full sky coverage (Feldman & Watkins, 1998).

In this paper, we make an accurate simulation of the SMT, to find the number of SNe Ia it will observe over its 5-year span, given the expected field distribution, magnitude errors and magnitude/redshift limit. We also simulate random surveys with different sky distributions and numbers of SNe Ia. We then apply an analytic formula for calculating the uncertainty on the bulk flow measurement (from both shot noise and cosmic variance) to each of these simulated surveys, and see how this depends on sky distribution, magnitude error and number of SNe Ia.

Throughout this work, we assume a standard flat ΛCDM cosmology with Planck parameters $(\Omega_m, \Omega_\Lambda, \Omega_b) = (0.32, 0.68, 0.049)$ and $H_0 = 100h \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$ with $h = 0.67$.

4.3 SkyMapper Supernova Survey Fields

The SMT fields were selected from an initial tiling of the whole Southern Hemisphere, via a realistic simulation of the year-round observing conditions using the AAO weather logs over a 5-year period (Scalzo et al., in preparation). Each field is $2.373^\circ \times 2.395^\circ$ degrees. Only fields were selected for which the Galactic extinction computed at their centre, using the maps of Schlegel et al. (1998), is less than 0.05 mag in the $g$ band, and which are observable from Siding Spring at an airmass less than 1.5. This left 1650 fields usable for the search, covering two distinct regions:

---

Planck Collaboration et al. (2013a). We use the Planck-only best fit values.

lat: 31S, long: 149E, alt: 1200m
A large southern patch, located between RA = 18h and RA = 6h, and a smaller equatorial patch, between RA = 8h and RA = 16h.

A 5-year survey was then simulated, using an algorithm for selecting fields that optimised the use of good seeing time, and which maintained a number of 200 fields (1000 deg$^2$) being observed at any given time (Scalzo et al., in preparation). The algorithm took into account (1) how long ago each field was observed, ensuring each field was observed at least once every five days, (2) the seeing on a given night, taken from weather logs recorded by the AAO between 2000 and 2005, (3) the galactic absorption affecting each field.

This resulted in a simulated survey with 900 observed fields over 5 years. Such a survey was found to be able to generate about 150 SNe Ia per year, up to $z = 0.12$, with an excellent lightcurve sampling (i.e. at least one photometry point in the $v$, $g$, $r$ and $i$ bands seven days before max, one point 30 days after max, and a minimum of nine 4-band photometric observations in total). This would require about 150-200 additional hours of good-seeing observing time per year, in addition to the bad-seeing time. These fields are shown in Figure 4.1; we assume these fields in this work.

The currently envisioned SMT will have a magnitude cutoff of $g = 19$ mag, corresponding to a maximum redshift of $z_{\text{max}} \sim 0.09$, and a magnitude uncertainty of $\Delta m = 0.14$, which corresponds to a distance uncertainty of $\sim 6.5\%$.

4.4 Survey Simulation

We first perform a realistic simulation of the SMT based on a $\Lambda$CDM $N$-body simulation, in order to predict the number of SNe that it will observe over 5 years given the current survey parameters. Using an $N$-body simulation provides a way of simulating the rate and distribution of SNe Ia detections we would expect to observe in a $\Lambda$CDM universe. We use the Millennium Run simulation (MR, Springel et al. 2005), a dark matter simulation within a box size of 500 $h^{-1}$Mpc$^3$, with a $\Lambda$CDM cosmology with WMAP 1-year parameters.

Since we require galaxy properties for our SN simulation, we use the $z = 0$ sem-
4.4. SURVEY SIMULATION

Figure 4.1: Fields for the 5-year SkyMapper Supernova Survey assumed in this work, in Equatorial coordinates and in Mollweide projection. The CMB dipole (stars) and WFH09 bulk flow dipole (diamonds) are also shown; the blue-red colours indicate the direction, with blue moving away from us and red moving towards us. The grey band shows galactic latitude $|b| < 10^\circ$.

analytic catalogue of [Croton et al. (2006)], which contains over 9 million galaxies and provides for each its position, stellar mass and star formation rate.

4.4.1 Survey simulation algorithm

Our method for generating a simulated SkyMapper Supernova Survey is as follows:

1. **Field selection.** We first place the observer at the centre of the box, and then overlay the survey selection function, i.e. selecting all the galaxies within the SMT fields.

2. **Implement Supernova rate.** We generate supernovae by applying the Type Ia supernova rate ($\text{SNR}_{\text{Ia}}$) from [Sullivan et al. (2006)], as a function of galaxy stellar mass ($M_\text{st}$) and mean star formation rate ($\dot{M}_\text{new}$, measured in solar masses per year):

$$\text{SNR}_{\text{Ia}}(t) = AM_\text{st}(t) + B\dot{M}_\text{new}(t),$$  

(4.3)

with $A = 5.3 \pm 1.1 \times 10^{-14}(H_0/70)^2$ SNe yr$^{-1}$ $M_\odot^{-1}$ and $B = 3.9 \pm 0.7 \times 10^{-4}(H_0/70)^2$ SNe yr$^{-1}$ $(M_\odot$ yr$^{-1})^{-1}$. We multiply this rate by the survey
duration to generate a random sample of SNe Ia that occur within the survey volume during the 5 years of observation.

To test consistency, we compute the SN volume rate that this produces, and compare to literature values. The volume rate produced by our method is

$$R = 2.67 \times 10^{-5} \text{ SNe yr}^{-1} \text{Mpc}^{-3},$$

which is consistent with empirical rates, such as the recent rate found in SDSS at $z = 0.1$ by Graur & Maoz (2013), of $R = 2.47^{+0.45}_{-0.57} \times 10^{-5} \text{ SNe yr}^{-1} \text{Mpc}^{-3}$, and that of Blanc et al. (2004) at $z \sim 0.13$, of $2.26^{+1.32}_{-1.15} \times 10^{-5} \text{ SNe yr}^{-1} \text{Mpc}^{-3}$.

3. **Lightcurve sampling.** To find the number of these supernovae with good photometric sampling, we assign each supernova a random time for light curve maximum during the survey duration and use the simulated observing logs of Scalzo et al. (in preparation) to see when the light curves are observed. For “good photometric sampling” we require an observation at least 7 days before maximum, and at least 3 observations in at least 3 different bands within 30 days after maximum, with at least 9 observations in total.

We also require that the supernovae are “spectroscopically identified,” meaning that the $g$-band magnitude $m_g$ goes below a certain threshold (19 mag) during the supernova life, where

$$m_g = 5 \log_{10}(d_L/\text{Mpc}) + 25 + M_g + \delta_m$$

with the absolute $g$-band magnitude $M_g$ assumed to be -19 mag, $d_L$ is the luminosity distance, and $\delta_m$ is a random observational error drawn from a Gaussian of width $\Delta m$.

4. **Creating observables ($z$, $v_{\text{rad}}$) with uncertainties**

- **Redshift $z$:** We calculate the real-space redshift $z_r$ for each SN from the inverse of

$$d_c(z_r) = \frac{c}{H_0} \int_0^{z_r} \frac{dz}{E(z)},$$

...
4.4. SURVEY SIMULATION

Here, $d_c$ is the comoving distance of the SN Ia in the simulation, and $E(z) = \sqrt{\Omega_m (1 + z)^3 + \Omega_\Lambda}$. We assume redshift uncertainties are negligible. The observed redshift $z_{\text{obs}}$ is obtained by adding the radial peculiar velocity via $(1 + z_{\text{obs}}) = (1 + z_r)(1 + v_p/c)$.

- **Uncertainty on radial peculiar velocity $S$:**

We calculate the uncertainty on the distance measurement of each SN due to a magnitude error of $\Delta m = 0.14$. This corresponds to a peculiar velocity error $\sigma_S$ of order 6.5% of the Hubble flow,

$$\frac{\sigma_S}{H_0d_c} = \frac{\ln(10)}{5} \Delta m \approx 6.5\%.$$  \hspace{1cm} (4.7)

We make 100 realisations of the SMT using the above method, to find the mean number of SNe we would expect it to observe. The histogram of the final number of SNe for the 100 realisations is shown in Figure 4.2. We find that SkyMapper should obtain $\sim 420 \pm 30$ SNe that are both spectroscopically identified and well observed, over the 5 year duration.

Figure 4.2: Histogram of the number of SNe obtained in 100 realisations of the SkyMapper Supernova Survey.
4. BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION

Figure 4.3: Redshift distribution of our simulated 5-year SkyMapper SNe Ia survey. Black line histogram: all SNe occurring within the survey fields. Magenta line histogram: well-sampled SNe. Blue solid histogram: spectroscopically identified SNe. Green solid histogram: the final sample of well-sampled and spectroscopically identified SNe.

For the results in this paper, we therefore use a simulation with 420 SNe. We show the redshift distribution of well-sampled and spectroscopically identified SNe for this simulation in Figure 4.3. A large number of SNe will occur within the survey fields during the 5 years, but only a small fraction of these are both well-sampled and spectroscopically identified. Their sky distribution is shown in Figure 4.4 and the redshift distribution of the final sample is shown in Figure 4.5.

4.4.2 Effective depth

In order to determine the effective depth of our mock survey, we calculate the characteristic ‘Maximum Likelihood Estimate (MLE) depth’ of SkyMapper, via

\[ d_{\text{MLE}} = \frac{\sum d_{c,n}w_n}{\sum w_n}, \]  

(4.8)
4.4. SURVEY SIMULATION

Figure 4.4: Simulated 5-year SkyMapper SN Ia survey, with 420 well-sampled SNe Ia, plotted in Equatorial coordinates and in Mollweide projection. The grey shaded region shows galactic latitude $|b| < 10^\circ$.

Figure 4.5: Redshift distribution of our simulated 5-year SkyMapper SNe Ia survey, with 420 SNe.
where the MLE weights are \( w_n = 1/(\sigma_{S,n}^2 + \sigma_*^2) \) and \( \sigma_* \) is a velocity dispersion noise term, which we assume to be 250 km s\(^{-1}\).

For our SMT simulation we find the MLE depth to be 152 \( h^{-1}\)Mpc.

### 4.5 Different survey geometries

An important aim of our predictions is to determine how the accuracy of a bulk flow measurement from a supernova survey depends on parameters of the survey – such as depth, sky coverage and number of SNe Ia. This will enable optimisation of SkyMapper, as well as provide general guidelines for optimising any supernova survey for performing a bulk flow measurement.

Varying the depth of the survey simply means varying the scale on which the bulk flow is measured. For any particular scale of bulk flow, the determining factor in how well we can measure it is the sky coverage of the survey. In this paper, we therefore consider only changes to the distribution of the survey fields, while keeping the redshift distribution constant.

To do this we generate a variety of mock surveys, all with the same number of fields (900), and the same redshift distribution as our SMT mock made using the Millennium Run, but with different arrangements of the fields on the sky. The number of SNe is set to either 420 (the predicted number for SkyMapper) or 1000. Our aim is to calculate the uncertainty that each of these surveys would measure for the bulk flow.

The method of generating these surveys is similar to the method used to simulate the SMT. However, since we are now interested only in the effect of survey geometry, we do not need to generate \( \Lambda CDM \) realisations of these surveys. Therefore, at the step of selecting Millennium Run galaxies within the survey fields (Step 2 in Section 3), we randomise the positions of the galaxies, while keeping their stellar mass and star formation rates. This maintains the correct overall volume rate of SNe Ia, but without including large scale structure from a particular \( \Lambda CDM \) realisation. The same could be achieved by simply using a known SN Ia volume rate. Also, since we do not have simulated observation logs for these surveys as we do for the SMT, we obtain the desired number of SNe by adjusting the total observation time.

The survey geometries we consider are (all shown in Figure 4.6):

151
4.5. DIFFERENT SURVEY GEOMETRIES

1. **All-sky isotropic.** We start by tiling the entire celestial sphere with SkyMapper-sized fields, then we randomly subsample these to obtain the desired number.

2. **Southern-sky isotropic.** We randomly subsample a tiling of fields of the whole southern hemisphere only.

3. **All-sky isotropic survey, excluding ZOA.** We randomly select fields from an all-sky tiling, but avoiding the Zone of Avoidance (ZOA) ($|b| > 10^\circ$ or $|b| > 20^\circ$).

4. **Southern-sky isotropic survey, excluding ZOA.** We select fields randomly from a southern sky tiling, avoiding the ZOA ($|b| > 10^\circ$ or $|b| > 20^\circ$).

5. **‘Ideal SkyMapper’: Dec < 10° isotropic survey, excluding ZOA.** We select fields randomly from an all-sky tiling with Dec < 10° (corresponding to the maximum declination of SkyMapper), avoiding the ZOA to a degree similar to the planned SMT ($|b| > 20^\circ$).

6. **All-sky glass survey.** [Haugbølle et al. (2007)] argue that in order to measure higher-order moments of the velocity field, it is necessary to minimise the size of the largest holes in the sky coverage. [Hannestad et al. (2008)] suggest measuring a glass-like distribution of SNe as a way of doing this. To see if this is also useful for the bulk flow, we place a glass-like distribution of fields over the whole sky, following the method used by [Hannestad et al. (2008)]:

   (a) Start with the full set of SkyMapper fields, covering the whole Southern sky. For an all-sky survey, mirror these in the northern hemisphere.

   (b) Select a random subset of these, containing $N_{\text{start}}$ fields.

   (c) The field outside this subset, with the maximal minimum angular distance from the already selected fields, is added to the subset.

   (d) The last point is iterated until 900 fields are obtained.
Figure 4.6: Various distributions of fields tested in this report, as labelled above each field, plotted in Equatorial coordinates in Mollweide projection. They all have 900 fields (black squares), but differ in their sky distribution. SNe discovered in those fields are shown by blue dots (either 500 or 1000 per survey as labelled). Some distributions exclude the zone of avoidance, either by 10 degrees in galactic latitude ($|b| > 10^\circ$) or 20 degrees ($|b| > 20^\circ$). The bottom right panel shows the glass distribution of supernovae without the fields, for clarity. Having a glass distribution of fields does not produce a glass distribution of SNe, since they occur randomly.
4.6 Bulk Flow estimation

The bulk flow $\mathbf{U} = (u_x, u_y, u_z)$ is essentially a weighted sum of the peculiar velocities within a volume,

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

(4.9)

where $S_n$ is the observed peculiar velocity of the $n^{th}$ galaxy, and $w_{i,n}$ is its weight in the $i^{th}$ direction. It has the advantage that it is sensitive to large scales ($\gtrsim 100 h^{-1} \text{Mpc}$), and hence is in the linear regime.

However, although $\mathbf{U}$ is in the linear regime, the individual velocities are affected by small-scale, nonlinear motions. This is a problem since they are in general sparsely sampled, and this can lead to an aliasing of small-scale power to larger scales in a moments analysis. This has historically made it difficult to compare bulk flow measurements between different sparse surveys, and with theoretical models [Kaiser, 1988; Watkins & Feldman, 1995; Juszkiewicz et al., 2000; Feldman et al., 2003; Hudson, 2003; Sarkar et al., 2007; Watkins & Feldman, 2007; Feldman & Watkins, 2008].

Recently, Watkins et al. (2009); Feldman et al. (2010) proposed a method for determining the bulk flow using an optimal weighting scheme that minimises the influence of small scales due to survey geometry, while minimising the uncertainties due to measurement noise. They name this the “minimum variance” (MV) method - the aim of the MV weights is to minimise the variance between the measured bulk flow, and the bulk flow that would be measured for an ‘ideal’ survey with a specified window function. We have used this method to estimate the accuracy with which SkyMapper and our generalised supernova surveys will measure the bulk flow. It provides an analytic way to estimate the shot noise and cosmic variance of the bulk flow measurement, dependent only on the weights and an input matter power spectrum.

An alternative method for determining the uncertainty would be to generate many mock peculiar velocity surveys, using multiple $N$-body simulation realisations, and calculate the variance in the measured bulk flow, incorporating measurement errors. However, the analytic estimate is simpler to perform and would give the same result.
4. BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION

4.6.1 Minimum Variance (MV) method

In this subsection we provide a brief overview of the MV method; see Watkins et al. (2009); Feldman et al. (2010) for a full description.

The MV method provides a way of determining the bulk flow \( \mathbf{U} \) (Equation 4.9) by calculating weights for each galaxy in order to make the survey match an ‘ideal’ window function. Here, we choose a spherical window function with a Gaussian fall-off of radius \( R_I = 150h^{-1}\) Mpc, which we find to be a good fit to the SkyMapper radial distribution. It has a Gaussian density profile,

\[
\rho(r) \propto \exp\left(-\frac{r^2}{2R_I^2}\right)
\]

which is shown in Figure 4.7 and its radial number distribution is

\[
n(r) \propto r^2 \exp\left(-\frac{r^2}{2R_I^2}\right).
\]

which is shown in Figure 4.8.

Since the magnitude limit means that the SMT has a redshift cutoff at \( z \sim 0.09 \), or \( \sim 275h^{-1}\) Mpc, we also apply this cutoff to the Gaussian density profile of our ideal window function.

We then calculate MV weights for SkyMapper, which weight each SN to match our ideal window function. In short, the weights are calculated by

\[
\mathbf{w}_i = (\mathbf{G} + \lambda\mathbf{P})^{-1}\mathbf{Q}_i,
\]

where \( \mathbf{w}_i \) is the vector of weights for each SN in the \( i^{th} \) direction. \( \mathbf{G} \) is the covariance matrix of the individual velocities, which includes both a noise term and a cosmic variance term determined by an input model power spectrum. \( \mathbf{P} \) is the \( k = 0 \) limit of the angle-averaged window function of the SNe, and \( \mathbf{Q}_i \) incorporates the input ideal window function. The \( \lambda \) term is a Lagrange multiplier. In this way, the MV weights depend on the actual window function of the SNe, the ‘ideal’ input window function, and the velocity covariance predicted by a given model power spectrum (although the weights are not sensitive to the exact parameters of the input model).

For a given survey, we can then calculate the uncertainties on the three bulk flow
moments (in the $x, y, z$ directions) from

$$\sigma_i^2 = \sum_n w_{i,n}^2 (\sigma_{S,n}^2 + \sigma_\star^2) + \frac{f^2 H_0^2}{2\pi^2} \int dk P(k) W^2_{ii}(k),$$

(4.13)

where $w_{i,n}$ are the MV weights, $\sigma_{S,n}$ is the measurement uncertainty of the $n$th peculiar velocity, $\sigma_\star$ is a velocity dispersion noise term, $\Omega_m$ is the cosmic matter density, $P(k)$ is the matter power spectrum, and $W^2_{ii}(k)$ is the angle-averaged tensor window function for the $i$th moment. (See WFH09 for further details of how $w_{i,n}$ and $W^2_{ii}(k)$ are calculated.)

The first term in the above equation is the shot noise component of the uncertainty, while the second term is the cosmic variance component. The shot noise is the limiting factor of any measurement of a bulk flow, since we only measure one, local bulk flow — i.e. we only have one realisation of the local Universe, and the shot noise determines how well we measure it. Cosmic variance, on the other hand, would be applied when we compare a bulk flow measurement to the prediction of a particular model such as $\Lambda$CDM, and hence it indicates the constraining power of a particular bulk flow measurement.

Note that the peculiar velocities themselves, $S_n$, do not enter into the equation for the bulk flow uncertainty — only the SN positions and velocity uncertainties $\sigma_{S,n}$. Our results are therefore not dependent on the particular realisation of peculiar velocities in our mocks.

### 4.7 Results

For each survey geometry and number of SNe, we calculate the uncertainty on the bulk flow measurement in the Equatorial $x, y, z$ directions (with $z$ pointing towards the North Pole), using the analytic formula in Equation 4.13. We then use these to calculate the uncertainty on the bulk flow amplitude, assuming a bulk flow in the direction of the WFH09 bulk flow. The results are shown in Tables 4.1 and 4.2.

Table 4.1 shows the uncertainties on the bulk flow components, $\delta x, \delta y, \delta z$, while Table 4.2 shows the uncertainties on the bulk flow amplitude, $\delta |U|$, assuming the direction of the bulk flow is that in WFH09. This provides the easiest way to compare the difference between the surveys. The uncertainties shown are the shot noise (first
4. **BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION**

![Figure 4.7: Radial density profile $\rho(r)$ of the Gaussian filter with $R_I = 150 h^{-1}$ Mpc. The filter is cut off at a radius corresponding to the maximum redshift of SkyMapper.](image)

Figure 4.7: Radial density profile $\rho(r)$ of the Gaussian filter with $R_I = 150 h^{-1}$ Mpc. The filter is cut off at a radius corresponding to the maximum redshift of SkyMapper.

...term of Equation 4.13), and the cosmic variance (second term of Equation 4.13). All results assume a magnitude error of $\Delta m = 0.14$, although in Table 4.2 we also show the results for $\Delta m = 0.10$ for certain surveys.

The last row of the tables shows the prediction for SkyMapper combined with the Palomar Transient Factory (PTF) [Law et al. 2009], a northern-sky survey of 1250 SNe Ia. The two cases are for 420 SkyMapper SNe (1670 total) and 1000 SkyMapper SNe (2250 total).

Table 4.2 allows a quick, easy comparison of the relative uncertainty on the bulk flow that the different surveys would measure. For a survey of 420 SNe, SkyMapper in its current form should measure the local bulk flow with an uncertainty of 92 km s$^{-1}$, assuming the direction of the flow is that measured by WFH09. This uncertainty decreases to 75 km s$^{-1}$, an 18% improvement, for an all-southern-sky survey, and 68 km s$^{-1}$, a 26% improvement, for an all-sky isotropic survey. If the number of SNe is increased to 1000, we find an uncertainty of 73 km s$^{-1}$, a 20% improvement. For certain surveys we show the results obtained when the magnitude...
4.7. RESULTS

Figure 4.8: Radial number distribution $N(r)$ of the Gaussian filter with $R_I = 150h^{-1}\ Mpc$ (black curve), compared to the scaled radial distribution of SkyMapper (red histogram). The filter is cut off at the maximum distance of SkyMapper.

The linear-theory $\Lambda$CDM prediction of the most likely bulk flow on this scale is $\sim 85\ km\ s^{-1}$ (see Figure 4.9). So assuming $\Lambda$CDM, SkyMapper would not be expected to measure the bulk flow at $150h^{-1}\ Mpc$; however, if the bulk flow is much larger than the $\Lambda$CDM prediction, as has been found by other surveys, then SkyMapper may be able to detect it if it extends to this scale.

The most ideal survey is an all-sky, isotropic random or glass survey (with $\sim 70\ km\ s^{-1}$ noise uncertainty for 420 SNe). However, if SkyMapper were combined with PTF, even better uncertainty could be achieved for SkyMapper in its current form due to the large number of SNe (which would give 55 km s$^{-1}$ noise uncertainty).

The decrease in uncertainty between SkyMapper ($92\ km\ s^{-1}$) and, say, an all-sky isotropic survey ($70\ km\ s^{-1}$), is not necessarily a factor of $\sqrt{2}$ as one might expect. This is because the uncertainty depends on the direction of the bulk flow. To increase the accuracy of the measurement requires increasing coverage of the
Figure 4.9: The predicted uncertainty of the SkyMapper Supernova and Transient Survey (SMT) bulk flow measurement (red triangle and error bar), compared to other measurements in the literature (blue). These are: Lavaux et al. (2013) (L13), WFH09 (W09), Turnbull et al. (2012) (T12), Colin et al. (2011) (C11), Planck Collaboration et al. (2013b) (P13), and Dai et al. (2011) (D11). All error bars are 1σ, while the two Planck arrows are the 95% upper limits. The solid line shows the linear-theory ΛCDM predicted bulk flow amplitude, assuming an all-sky Gaussian survey window function, with the 1σ and 2σ levels shown by the dark grey and light grey shaded regions. (Note the measurements cannot be directly compared with this prediction since they generally have different window functions; this is mainly for illustration).
4.7. RESULTS

Table 4.1: Bulk flow errors \((\delta x, \delta y, \delta z)\) due to shot noise (first two columns) and cosmic variance (second two columns), in \(\text{km s}^{-1}\) on a scale of \(150 \ h^{-1} \text{Mpc}\), for surveys with 900 fields, \(z_{\text{max}} = 0.09\), and different field distributions. For SkyMapper + PTF, this \(z_{\text{max}}\) and number of fields only apply to SkyMapper. PTF itself has 1250 SNe Ia, in addition to the number of SkyMapper SNe Ia.

<table>
<thead>
<tr>
<th>Geometry</th>
<th>Number of SNe</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>420</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shot noise errors</td>
<td>Cosmic variance errors</td>
</tr>
<tr>
<td>SkyMapper</td>
<td>(69,80,104)</td>
<td>(53,58,84)</td>
<td>(43,37,57)</td>
</tr>
<tr>
<td>Dec &lt; 10°, (</td>
<td>b</td>
<td>&gt; 20°)</td>
<td>(63,75,86)</td>
</tr>
<tr>
<td>Southern-sky isotropic, (</td>
<td>b</td>
<td>&gt; 20°)</td>
<td>(61,80,84)</td>
</tr>
<tr>
<td>Southern-sky isotropic, (</td>
<td>b</td>
<td>&gt; 10°)</td>
<td>(66,73,77)</td>
</tr>
<tr>
<td>Southern-sky isotropic</td>
<td>(70,70,78)</td>
<td>(49,49,62)</td>
<td>(33,34,53)</td>
</tr>
<tr>
<td>Southern-sky glass</td>
<td>(63,68,81)</td>
<td>(46,49,65)</td>
<td>(35,33,52)</td>
</tr>
<tr>
<td>All-sky isotropic, (</td>
<td>b</td>
<td>&gt; 20°)</td>
<td>(64,75,70)</td>
</tr>
<tr>
<td>All-sky isotropic, (</td>
<td>b</td>
<td>&gt; 10°)</td>
<td>(68,75,69)</td>
</tr>
<tr>
<td>All-sky isotropic</td>
<td>(67,69,70)</td>
<td>(48,49,49)</td>
<td>(30,30,29)</td>
</tr>
<tr>
<td>All-sky glass</td>
<td>(66,71,73)</td>
<td>(48,49,49)</td>
<td>(31,30,30)</td>
</tr>
<tr>
<td>SkyMapper + PTF (+1250)</td>
<td>(48,60,60)</td>
<td>(40,49,53)</td>
<td>(36,33,33)</td>
</tr>
</tbody>
</table>

The dipole cosine on the sky. As can be seen in Figure 4.1, the WFH09 bulk flow is already partially measured by SkyMapper, so we only see a 24% improvement when moving to an all-sky survey.

This is illustrated more clearly in Figure 4.10, where we show how well a given SN survey constrains the bulk flow cosine on the sky, assuming the peculiar velocity of each SN Ia is due purely to the bulk flow, and assuming that the bulk flow is in the direction found by WFH09. Specifically, we plot \((\cos \theta_n)\) where \(\theta_n\) is the angle between the supernova and the WFH09 bulk flow, perturbed by a Gaussian error with standard deviation \(\sigma = (\sigma_{S,n}/U_{\text{rad},n}) \cos \theta_n\). Here, \(U_{\text{rad},n} = U_{\text{WFH09}} \cdot \hat{r}_n\) is the radial component of the WFH09 bulk flow at the position of the \(n^{th}\) supernova and \(\sigma_{S,n}\) is the peculiar velocity measurement uncertainty. SkyMapper in its current form already covers part of the cosine, although better coverage is achieved by an all-southern sky survey or an all-sky survey.

If we instead consider a bulk flow along the north-south direction, which is the
4. BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION

Table 4.2: Same as Table 4.1 except showing the uncertainty on the bulk flow magnitude (in km s$^{-1}$) on a scale of 150 $h^{-1}$ Mpc, assuming a bulk flow in the WFH09 direction. The uncertainties are shown for shot noise, with cosmic variance in parentheses. All uncertainties are for a magnitude error of $\Delta m = 0.14$; however, for certain surveys we also show the result with $\Delta m = 0.10$.

<table>
<thead>
<tr>
<th>Geometry</th>
<th>Number of SNe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>420</td>
</tr>
<tr>
<td>SkyMapper</td>
<td>92</td>
</tr>
<tr>
<td>SkyMapper, $\Delta m = 0.10$</td>
<td>77</td>
</tr>
<tr>
<td>Dec $&lt; 10^\circ$, $</td>
<td>b</td>
</tr>
<tr>
<td>Southern-sky isotropic, $</td>
<td>b</td>
</tr>
<tr>
<td>Southern-sky isotropic, $</td>
<td>b</td>
</tr>
<tr>
<td>Southern-sky glass</td>
<td>75</td>
</tr>
<tr>
<td>All-sky isotropic, $</td>
<td>b</td>
</tr>
<tr>
<td>All-sky isotropic, $</td>
<td>b</td>
</tr>
<tr>
<td>All-sky isotropic</td>
<td>68</td>
</tr>
<tr>
<td>All-sky glass</td>
<td>70</td>
</tr>
<tr>
<td>All-sky glass, $\Delta m = 0.10$</td>
<td>53</td>
</tr>
<tr>
<td>SkyMapper + PTF (+1250)</td>
<td>55</td>
</tr>
<tr>
<td>SkyMapper + PTF (+1250), $\Delta m = 0.10$</td>
<td>50</td>
</tr>
</tbody>
</table>

direction in which SkyMapper has the least optimal sky coverage, the bulk flow uncertainty is given purely by the uncertainty in the $z$ direction: 104 km s$^{-1}$ noise uncertainty for SkyMapper. Since the $z$-direction uncertainty reduces to 73 km s$^{-1}$ for an all-sky survey, this represents a $\sim 30\%$ improvement.

One interesting point is that when the magnitude error is decreased (and so the uncertainty on the measured peculiar velocities decreases), shot noise also decreases, but cosmic variance increases. One way to explain this is that having larger velocity errors means that large-scale structure is effectively ‘smoothed’, and so cosmic variance becomes less significant. The reverse then holds when the magnitude error decreases.
4.7. RESULTS

Figure 4.10: Illustration of how surveys of different sky coverage constrain the bulk flow dipole on the sky, assuming the bulk flow is in the direction found by WFH09. Top panel: SkyMapper in its current configuration, with 420 SNe. Middle panel: a southern-sky isotropic survey, excluding the ZOA with $|b| > 10^\circ$ and 420 SNe. Bottom panel: an all-sky, isotropic survey with 420 SNe. The bulk flow cosine is shown by the blue solid line. Each black star shows the position of a SN with respect to the bulk flow direction. Errorbars are shown in grey.
4. BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION

4.8 Discussion

The accuracy of the SkyMapper bulk flow measurement that we predict depends on its scale, which is given by the effective depth of SkyMapper (which we have shown is $\sim 150 h^{-1}\text{Mpc}$). We can compare this to the $\Lambda\text{CDM}$ prediction for the rms velocity dispersion $\sigma_V(R)$, which is given in linear theory by

$$\sigma^2_V(R) = \frac{H_0^2 f^2}{2\pi^2} \int_{k=0}^\infty P_{\delta\delta}^{(L)}(k) \tilde{W}(k; R)^2 dk$$

(4.14)

where $\tilde{W}(k; R)$ is the survey window function in Fourier space with effective radius $R$, and $P_{\delta\delta}^{(L)}(k)$ is the linear matter power spectrum.

If the density field is close to Gaussian, then on large, linear scales the peculiar velocity field is also close to Gaussian, and the velocity probability distribution can be modelled as Maxwellian (Kashlinsky et al., 2008; Li et al., 2012). The most likely value of the bulk flow amplitude $U_{ML}$ can then be calculated from $U_{ML} = \sqrt{2/3} \sigma_V$, with a $1\sigma$ range given by $\Delta U_1 = 0.389 \sigma_V$. We show this prediction for our fiducial cosmology, and an all-sky Gaussian window $\tilde{W}_G = \exp (-k^2 R^2/2)$ in Figure 4.9 along with our SkyMapper prediction.

Our prediction for the accuracy of the SkyMapper bulk flow measurement indicates that it should be competitive with existing bulk flow measurements. We list below some of the most competitive recent bulk flow measurements in the literature, and we also plot these in Figure 4.9. Note that different papers use different methods for determining the bulk flow, and for quoting the scale on which their bulk flow is determined. They also generally have different window functions, and are on different scales to SkyMapper, and so they cannot be directly compared to the SkyMapper prediction. All uncertainties quoted are total uncertainty (noise + cosmic variance), except where indicated otherwise.

• Watkins et al. (2009) measured a bulk flow in a $50 h^{-1}\text{Mpc}$ Gaussian window (i.e. a distance 3× smaller than the SkyMapper window), using the all-sky COMPOSITE sample with $\sim 4500$ galaxies, and the Minimum Variance method. Their result was $407 \pm 81 \text{ km s}^{-1}$ (with $65 \text{ km s}^{-1}$ noise) In comparison, the $\Lambda\text{CDM}$ predicted 1D rms velocity for a $50 h^{-1}\text{Mpc}$ Gaussian is $\sim 110 \text{ km s}^{-1}$, or $\sim 190 \text{ km s}^{-1}$ for the bulk flow amplitude.
• Colin et al. (2011) measured a bulk flow within $18\,000\,\text{km\,s}^{-1}$ ($\sim 260\,h^{-1}\text{Mpc}$, about $1.5\times$ the scale of SkyMapper), using the Union2 catalogue of 557 SNe Ia, and a Maximum Likelihood (ML) method, and found $250^{+190}_{-160}\,\text{km\,s}^{-1}$.

• Dai et al. (2011) measured the bulk flow at $z < 0.05$ ($148\,h^{-1}\text{Mpc}$), using the Union2 dataset. They found a bulk flow of $188^{+119}_{-103}\,\text{km\,s}^{-1}$ for SNe with $z < 0.05$, and no significant flow for SNe with $z > 0.05$.

• Turnbull et al. (2012) made a measurement at $50\,h^{-1}\text{Mpc}$, using the First Amendment (A1) compilation of 245 SNe (also all-sky). They use both a Maximum Likelihood (ML) and Minimum Variance (in $50\,h^{-1}\text{Mpc}$ Gaussian window) method, and find $197 \pm 56\,\text{km\,s}^{-1}$ for the ML (uncertainty not including cosmic variance) and $249 \pm 76\,\text{km\,s}^{-1}$ for the MV method.

• Lavaux et al. (2013) made a measurement within $50\,h^{-1}\text{Mpc}$, using the kinetic Sunyaev-Zel’dovich (kSZ) signal from WMAP 7-year data, and found $533 \pm 263\,\text{km\,s}^{-1}$.

• Planck Collaboration et al. (2013b) measured the bulk flow at $z = 0.18$ using the kSZ signal from Planck. They found a limit of $390\,\text{km\,s}^{-1}$ within spheres of $350\,h^{-1}\text{Mpc}$ at 95% C.L., and on larger scales, a bulk flow below $254\,\text{km\,s}^{-1}$ at 95% C.L. for a radius of $2\,h^{-1}\text{Gpc}$. (Note that Planck and WMAP cannot put constraints on the local bulk flow, at $r < 350h^{-1}\text{Mpc}$).

SkyMapper is therefore competitive with many of the existing datasets, including the Union2 and First Amendment supernova datasets. By combining SkyMapper with other available peculiar velocity datasets, an even more precise measurement of the local bulk flow should be achievable.

We note that in our above analysis, we have not considered systematic effects, only statistical effects. Perhaps the most important systematic for the SkyMapper bulk flow measurement is the fact that it only surveys the southern hemisphere, not the full sky. We leave investigation of such systematics to future work.

We also note that the Minimum Variance method that we use assumes that the velocity uncertainties are Gaussian. Work on other peculiar velocity surveys with different distance indicators, such as the 6 degree Field Galaxy Survey (6dFGS)
which uses the Fundamental Plane, shows that having non-Gaussian uncertainties can bias the bulk flow measurement (Scrimgeour et al. 2014, in preparation). For SNe Ia, it is generally assumed that the magnitude errors are Gaussian, which would mean the uncertainty on distance should be lognormal. However, since the size of the uncertainties for SNe Ia are small, we assume they can be approximated as Gaussian.

This is justified in Figure 4.11, where for our SMT simulation we compare the ‘true’ errors resulting from a Gaussian uncertainty on magnitude, and the 1σ error we would expect if we had Gaussian uncertainties on the peculiar velocities corresponding to 6.5% of the Hubble Flow. The distribution of true errors appears reasonably consistent with the Gaussian assumption, and ~70% lie within the 1σ shaded region, although there is clearly more scatter towards negative values. Also, selection bias causes an absence of SNe in the final sample (black points in Figure 4.11) with low values of δv at large comoving distance. Both of these effects would need to be corrected for in an actual bulk flow measurement; we ignore them here in estimating the bulk flow uncertainty. Since biases from non-Gaussian errors will increase as numbers of SNe increase in future samples, the assumption of Gaussian uncertainties should be revisited in future work.

### 4.9 Conclusion

We have made a realistic prediction of the SkyMapper Supernova Survey, using the simulated fields and observing logs of Scalzo et al. (in preparation) along with updated predictions of the magnitude limit and uncertainty Δm. Using this we predict that SkyMapper will produce about 80 spectroscopically identified SNe per year, assuming the observation conditions and additional hours of good-seeing time suggested by Scalzo et al (in preparation). The updated magnitude error of Δm = 0.14 leads to a predicted peculiar velocity uncertainty for each SN Ia of ~6.5% of the Hubble flow.

We have predicted the accuracy with which the SMT will measure the local bulk

---

5Near the magnitude limit of the survey, SNe that have their apparent magnitudes perturbed to larger values (making them appear more distant, and hence having a more negative peculiar velocity) will fall off the apparent magnitude threshold, and will not be spectroscopically identified.
4.9. CONCLUSION

Figure 4.11: Comparison of the ‘true’ peculiar velocity errors $\delta_v$ of our simulated SNe Ia, to the assumed Gaussian errors, as a function of comoving distance. Black points: the ‘true’ error $\delta_v$ in the derived peculiar velocity of each SN, resulting from a Gaussian distance modulus uncertainty. Grey stars: the same, but for the well-sampled SNe that did not meet the mag $< 19$ requirement and so did not make it into the final sample. Shaded region: the assumed 6.5% Hubble Flow Gaussian error, i.e. $6.5\% \times H_0 d_c$. The lack of black points with small/negative $\delta_v$ at large distances is due to selection bias.

flow, and find it will have $\sim 90$ km s$^{-1}$ noise uncertainty, assuming the WFH09 bulk flow direction. We investigate the effect of survey geometry, number of SNe Ia and magnitude error on the bulk flow uncertainty, by generating mock surveys with different field distributions. To determine the uncertainty (from both noise and cosmic variance), we use the analytic formula from WFH09. We find that distributing the SkyMapper fields evenly over the whole southern sky would lead to an 18% improvement in noise uncertainty, while an isotropic all-sky distribution would have a 26% improvement. Increasing the number of SNe Ia to 1000 would give a 20% improvement. Decreasing the magnitude uncertainty from $\Delta m = 0.14$ to $\Delta m = 0.10$ would give a 16% improvement.

In conclusion, SkyMapper in its current form will be competitive with many
4. **BULK FLOW ESTIMATION FOR THE SKYMAPPER SUPERNOVA AND TRANSIENT SURVEY (SMT) AND GENERAL SURVEY OPTIMISATION**

current constraints of the bulk flow. It could be further improved by increasing the sky coverage of the fields, or extending the observing time to obtain more SNe Ia. Even without further modifications, SkyMapper will provide an important southern-sky sample of SNe that will add significant constraint when combined with existing northern-sky SN surveys.
5

Summary and Future Work

5.1 Thesis Summary

In this thesis, we have presented two different measurements of large-scale structure that provide a test of the standard model of cosmology, ΛCDM.

In Chapter 2 we investigated the transition to homogeneity in the clustering of galaxies in the WiggleZ Dark Energy Survey. This is an important test of the ΛCDM model, since ΛCDM is based on the assumption that the large-scale Universe is homogeneous – the ‘Cosmological Principle’ – which allows the use of the Friedmann-Robertson-Walker metric in which all calculations within the model are made. Our measurement is the largest-volume, and most precise, measurement of the scale of homogeneity to date. By measuring both the mean counts-in-spheres $N(< r)$ and fractal correlation dimension $D_2(r)$ of the dataset, and seeing how these change with scale, we find a clear transition from small-scale clustering, to an absence of clustering on large scales. Due to the depth of WiggleZ ($z < 1$) we were also able to make the first measurement to date of how the transition to homogeneity varies with cosmic epoch.

Using this method, we calculate the homogeneity scale $R_H$ for WiggleZ to be $71 \pm 8 \, h^{-1}\text{Mpc}$ at $z \sim 0.2$, $70 \pm 5 \, h^{-1}\text{Mpc}$ at $z \sim 0.4$, $81 \pm 5 \, h^{-1}\text{Mpc}$ at $z \sim 0.6$, and $75 \pm 4 \, h^{-1}\text{Mpc}$ at $z \sim 0.8$. Our results are consistent with other measurements in the literature which find homogeneity at around $70 - 150 \, h^{-1}\text{Mpc}$, such as Guzzo [1997].
Martínez et al. (1998), Amendola & Palladino (1999), Pan & Coles (2000), Hogg et al. (2005), Yadav et al. (2005) and Sarkar et al. (2009). We make the small caveat that the exact homogeneity scale will depend on the bias of the galaxy sample, and so a direct comparison between datasets cannot be made without accounting for this. Also, the bias of WiggleZ changes with redshift, and so we do not see the monotonic increase of $R_H$ with epoch as we might expect. Our results are in disagreement with a number of studies that have claimed to find evidence for fractal structure extending up to large scales ($\gtrsim 100\, h^{-1}\, \text{Mpc}$), such as Coleman & Pietronero (1992), Joyce et al. (1999), Sylos Labini et al. (2009) and Sylos Labini (2011). We find no evidence of large-scale fractal behaviour in WiggleZ. Furthermore we have made a thorough test of the robustness of our measurement against systematic effects from the survey window function, using a suite of fractal mock catalogues, so we can be confident that our result is not being biased to a false detection of homogeneity.

Our measurement provides a strong consistency test of the FRW-based ΛCDM model. It is not entirely independent of the assumption of the FRW metric, since we must use this in order to convert galaxy redshifts to distances. However, it would be extremely unlikely for our measurements of $N(<r)$ and $D_2(r)$ to so closely agree with the FRW-based ΛCDM model if the distribution were actually inhomogeneous, and the FRW metric had been incorrectly assumed.

Subsequent measurements have provided further support for our conclusions, including Marinoni et al. (2012) who measure the scale of isotropy to be $\sim 150\, h^{-1}\, \text{Mpc}$ for the SDSS Data Release 7, Hoyle et al. (2013) who use the star formation history of the SDSS LRG sample to find no evidence for inhomogeneity up to redshift $z < 0.5$, and Nadathur (2013) who shows that recent measurements of large, Gigaparsec-size structures in SDSS (by Clowes et al., 2013) are not in contradiction with the homogeneity scale we measure.

In Chapter 3, we presented a bulk flow analysis of the 6dFGSv survey, which is the largest and most homogeneously selected peculiar velocity sample to date. The bulk flow, or the dipole of the velocity field, is a measure of the overall motion of the local Universe, and its variance is predicted by ΛCDM. Recent measurements in the literature have given inconsistent results, but a number of them have suggested the local bulk flow is larger than predicted by ΛCDM (Kashlinsky et al., 2008; Watkins
et al., 2009; Abate & Feldman, 2012; Lavaux et al., 2013). Applying the ‘Minimum Variance’ bulk flow estimator introduced by Watkins et al. (2009) and Feldman et al. (2010), we measure the bulk flow amplitude to be $|U| = 259 \pm 54 \text{ km s}^{-1}$ in the direction $(l, b) = (317 \pm 16^\circ, 35 \pm 11^\circ)$ on a scale of $50 \text{ h}^{-1} \text{ Mpc}$, and $|U| = 254 \pm 54 \text{ km s}^{-1}$ in the direction $(l, b) = (317 \pm 16^\circ, 35 \pm 11^\circ)$ at a distance of $70 \text{ h}^{-1} \text{ Mpc}$. Our result is one of the most precise bulk flow measurements to date, thanks to the large number of galaxies (8885) in the 6dFSGv sample. Our result is larger than the predicted variance in ΛCDM, but is not in significant disagreement; our bulk flow constraints on $\Omega_m$ and $\sigma_8$ are within 2σ of the values measured by the Planck first data release. From the measured bulk flow, we determine the constraint on $\sigma_8$ to be $\sigma_8 = 1.14^{+1.23}_{-0.51}$ for the measurement at $50 \text{ h}^{-1} \text{ Mpc}$, and $\sigma_8 = 1.11^{+1.21}_{-0.50}$ for the measurement at $70 \text{ h}^{-1} \text{ Mpc}$ (both 68.27% C.L.). Both of these values are consistent with the Planck best-fit value of 0.83 within 68.27% confidence. Our result is consistent with a number of recent bulk flow measurements that also do not find an amplitude significantly larger than the ΛCDM prediction, including Colin et al. (2011), Dai et al. (2011), Nusser & Davis (2011), Turnbull et al. (2012), and Feindt et al. (2013).

Nonetheless, the fact that we measure a relatively large, non-zero bulk flow indicates that there must be a large nearby overdensity sourcing it. The direction of our measured bulk flow is close to the Shapley Supercluster, consistent with the finding of Magoulas (2012) who also analysed 6dFGSv, as well as other measurements, indicating that this is responsible for a good part of the local motion. This goes part of the way to answering the question of the origin of the Local Group motion (Kogut et al., 1993; Kocevski & Ebeling, 2006; Erdoğan et al., 2006a,b). However, Shapley is at the edge of the 6dFGSv sample, and some measurements have found evidence that there may be further mass sourcing the bulk flow beyond this – e.g. Feindt et al. (2013), and Magoulas (2012) who compare the 6dFGSv velocity field with the reconstructed 2MRS density and velocity fields, and find a large residual bulk flow of $|U_{\text{res}}| = 319 \pm 41 \text{ km s}^{-1}$, suggesting there may be contributions to the bulk flow from structures beyond the sample. Higher-redshift peculiar measurements are needed to resolve this.

It is perhaps surprising that there appears to be a relatively large local bulk flow - or on the other hand, that there is such a large nearby overdensity as Shapley
- in light of the strong evidence for homogeneity on $\sim 70 \, h^{-1} \, \text{Mpc}$ scales that we find in Chapter 2. One explanation is that the homogeneity measurement we make is a volume-averaged measurement, and so does not necessarily rule out fluctuations above the homogeneity scale we measure. Secondly, the existence of a nearby Shapley-size structure is not necessarily surprising - Sheth & Diaferio (2011) show, using extreme value statistics, that such a structure is not unexpected within any $200 \, h^{-1} \, \text{Mpc}$ radius, if the initial fluctuation field was Gaussian. Nonetheless, more measurements are needed, of both the velocity and density field, to distances beyond Shapley to fully understand the origin of the bulk flow.

In Chapter 4 we presented a realistic simulation of the SkyMapper Supernova and Transient Survey (SMT), which is currently running on the SkyMapper telescope at Siding Spring Observatory. This survey will observe $\sim 420$ SNe Ia at $z < 0.1$ over $\sim 1000 \, \text{deg}^2$, for which peculiar velocities can be derived. Using the Minimum Variance method, we find that SMT should measure the bulk flow on a scale of $150 \, h^{-1} \, \text{Mpc}$ with an uncertainty of $\sim 90 \, \text{km s}^{-1}$.

We also investigated how properties of a general SN Ia survey, such as the distribution of fields and the magnitude uncertainty, can be varied to optimise the bulk flow constraints. We find that distributing the SMT fields isotropically over the southern sky leads to a 18% improvement in bulk flow accuracy, while distributing them isotropically over the whole sky leads to a 26% improvement. Decreasing the magnitude uncertainty of SMT from 0.14 to 0.10 mag would give a 16% improvement, while increasing the number of SNe to 1000 would give a 20% improvement. These findings will be of use in guiding the survey design of low-redshift SN Ia surveys that aim to measure the local bulk flow.

### 5.2 Future Work

This is a very exciting time for cosmology, with huge recent advantages in knowledge provided by the CMB, SNe, galaxy redshift surveys and weak lensing. A large effort is going into developing new telescopes and surveys, which will provide a wealth of new data. At the same time, there is ongoing work on improving theoretical
models, in particular to understand structure growth on nonlinear scales, which will improve the information that can be extracted from the data. This will help to advance our current measurements, and hopefully provide new ways of searching for physics beyond ΛCDM. We give below a summary of some upcoming peculiar velocity and redshift surveys and how they can be used to extend the work in this thesis.

5.2.1 Further Work with 6dFGSv and SkyMapper

First, there are a number of ways in which our work with the 6dFGSv and SMT surveys could be extended. Our bulk flow analysis of 6dFGSv could be easily extended to investigate the higher-order moments of the velocity field, such as the quadrupole and octupole, as described in Feldman et al. (2010), and a related method, the expansion of the velocity field into multipoles, as described by Haugbølle et al. (2007) and Hannestad et al. (2008). Our bulk flow constraint can also be used, along with the shear and octupole moments, to constrain the amplitude of the matter power spectrum, as in Macaulay et al. (2011, 2012), and provide improved constraints on Ω_m and σ_8.

We have presented a prediction for the accuracy with which SMT will measure the bulk flow. This survey is currently running on the SkyMapper telescope, and once the data is available we will be able to use it for a bulk flow analysis. This dataset can be combined with current and upcoming northern-sky SN Ia surveys, which we describe in the following subsection, for a more precise measurement.

5.2.2 Current and Future Peculiar Velocity Surveys

The potential of peculiar velocities as an independent cosmological probe, as well as intriguing questions raised about the properties of the local velocity field from the apparently large bulk flow, have motivated a number of different peculiar velocity surveys, which we summarise below.

A new peculiar velocity survey, the 2MASS Tully-Fisher (2MTF) survey (Masters et al., 2008; Hong et al., 2013) is almost complete, and comprises the largest and most uniform Tully-Fisher sample to date. It combines near-infrared and 21-cm radio observations to measure the infrared Tully-Fisher relation, which has less intrinsic
scatter than the optical analogue (Aaronson et al. [1982]). 2MTF was selected from the Two Micron All-Sky Extended Source Catalog (2MASS XSC; Jarrett et al., 2000) and will contain about 3000 HI widths from the Parkes, Green Bank and Arecibo radio telescopes. These cover the whole sky, with galactic latitude $|b| > 5^\circ$, out to a redshift of $cz < 10000 \text{ km s}^{-1}$. This will provide an excellent new sample for peculiar velocity analysis, with smaller distance uncertainties than 6dFGSv ($\sim 20\%$) due to both the decreased scatter of the infrared Tully-Fisher relation compared to the FP, and the lower redshift of the sample. It should also be less susceptible to systematic biases due to sky coverage.

Peculiar velocity surveys up to now have been hampered by sparse sampling, large errors and inhomogeneous selection. 6dFGSv, the largest and most homogeneous sample to date, is limited by only covering half the sky. 2MTF will be a major step forward, by providing a large, uniformly selected all-sky TF sample. This would make it an excellent dataset for an improved bulk flow analysis, since, as we showed in Chapter 4, good coverage of the dipole cosine on the sky is crucial. It would also allow a more accurate measurement of the local velocity field to be made (as in e.g. Erdogdu et al. [2006b] Lavaux et al. [2010]), which when combined with the density field provided by 2MASS will allow a constraint on the parameter combination $f\sigma_8$, and the growth index $\gamma$ (since $f \sim \Omega_m^\gamma$), which would allow a test of gravity (as in e.g. Hudson & Turnbull [2012]). The growth rate measurement provided by a survey like 2MTF will provide an important low-redshift complement to measurements of the growth rate from redshift-space distortions using galaxy surveys at higher redshift (e.g. Guzzo et al. [2008] Blake et al. [2011a] Beutler et al. [2012]). Having a low-redshift measurement is important, since the density of dark energy is highest at low redshift.

The TAIPAN survey, an all-southern-sky optical galaxy redshift survey with the UK Schmidt Telescope (UKST), is an upcoming survey using an upgraded version of the 6dF spectrograph that will extend the depth of 6dFGS. This will cover a similar area of sky to 6dFGS, but will extend it in redshift to $z \simeq 0.08$. TAIPAN is currently in the early planning stage, and will have one of two possible configurations: a deeper survey of 406 000 galaxies to a magnitude limit of $r = 17$, or a shallower survey of

---

1TAIPAN: Transforming Astronomical Imaging surveys through Polychromatic Analysis of Nebulae
221,000 galaxies to $r = 16.5$. TAIPAN will provide important information about the peculiar velocity field beyond Shapley (which is just at the edge of the 6dFGS volume), and may help to resolve questions about the origin of the bulk flow.

We have seen the benefit of low-redshift SNe Ia for making peculiar velocity measurements, and there are a number of new current and future nearby SN Ia surveys that will extend the available samples. These include the Palomar Transient Factory (PTF, Law et al., 2009), a northern-sky survey of 1250 SNe Ia, which we included in our bulk flow prediction in Chapter 4 and the Nearby Supernova Factory (SNfactory, Aldering et al., 2002) of 117 SNe in the redshift range $0.03 < z < 0.08$, which has already been used to measure the bulk flow (Feindt et al., 2013). The precision of SN Ia distance measurements makes them an important complement to larger, but more noisy, FP and TF samples, and would also allow for a cross-correlation to explore systematic biases in peculiar velocity measurements. Additionally, a number of analysis techniques have been suggested to make use of the precise velocity measurements from SNe Ia. These include measuring the covariance of luminosity distance fluctuations caused by peculiar velocities, as a probe of $\Omega_m$ and the growth index $\gamma$ (Hui & Greene, 2006; Gordon et al., 2007; Abate & Lahav, 2008), and an angular expansion of the velocity field as measured by SNe Ia (Haugbølle et al., 2007; Hannestad et al., 2008). (With the growing number of peculiar velocities from less precise distance indicators like the FP and TF relations, these methods will also be extendable to these datasets).

Looking further ahead, a major upcoming wide-area survey is the Widefield ASKAP L-band Legacy All-sky Blind surveY (WALLABY), a blind HI survey with the Australian SKA Pathfinder telescope (ASKAP). WALLABY will observe $\sim 500,000$ to $600,000$ galaxies with a mean redshift of $z \simeq 0.04$ (Duffy et al., 2012). A complementary northern-sky survey will also be made, the Westerbork Northern HI Sky Survey (WNSHS), with the Apertif radio telescope. Together, WALLABY and WNSHS will provide $\sim 30,000$ TF distances up to $z = 0.05$, the largest peculiar velocity survey ever made (3-4 times larger than existing surveys).

A promising technique for peculiar velocity analysis is to combine the velocity and density field information to cancel the cosmic variance, similar to the method used in the PTF. This technique can be extended to the larger datasets provided by WALLABY and WNSHS.
of combining tracers of the density field proposed by McDonald & Seljak (2009). Recently, Koda et al. (2013) have investigated the potential of this method, and find that TAIPAN and WALLABY/WNSHS will be able to measure the growth rate $f\sigma_8$ to a precision of 3 per cent at $z \sim 0.025$, a 40 per cent improvement on the constraint from redshifts alone from the same surveys. This method would also allow a measurement of the growth rate as a function of scale, and the above surveys should measure this to a precision of 15-30 per cent in bins of wavenumber with width $\Delta k = 0.01 h \text{Mpc}^{-1}$, also a large improvement over galaxy redshifts alone. Such a scale-dependent measurement of the growth rate would allow for tests of modified gravity models, such as $f(R)$ models, which predict a scale-dependent strength of gravity.

With the wealth of upcoming data, the future looks very promising for peculiar velocities, which will provide independent constraints to existing cosmological probes, and a unique test of the nature of gravity at low redshifts where the effect of dark energy, or modified gravity, is strongest.

5.2.3 Current and Future Redshift Surveys

Galaxy redshift surveys have been extremely successful in recent years in providing competitive constraints on cosmological models, via several different techniques including the galaxy power spectrum, redshift-space distortions and baryon acoustic oscillations. As a result a large effort is focusing on new galaxy redshift surveys to provide new tests of $\Lambda$CDM, constraints on the nature of dark energy via measurements of its equation of state $w$, and tests of modified gravity and non-Gaussianity. We summarise some of these surveys relevant for our analysis below.

An important survey that is just finishing is the Baryon Oscillation Spectroscopic Survey (BOSS, Schlegel et al. [2009]), part of SDSS III and running to 2014. BOSS will consist of a redshift survey of $\sim 1.5$ million LRGs at $0.2 < z < 0.8$ over 10,000 deg$^2$, and a quasar survey of 160,000 objects at $2.3 < z < 2.8$ over 8000 deg$^2$. BOSS was designed to make percent-level measurements of the BAO scale. In the future, e-BOSS will extend BOSS beyond $z = 0.6$, measuring emission line galaxies (ELGs) at $0.6 < z < 1$ and quasars at $1 < z < 2$. It will make competitive new BAO constraints at $0.6 < z < 2$. And further ahead, the proposed BigBOSS survey
(Schlegel et al., 2011) aims to measure BAO at $0.2 < z < 3.5$, as well as measure 15.3 million ELGs, 3.4 million luminous red galaxies (LRGs) and 630,000 quasars.

BOSS is the largest-volume galaxy survey ever made and would allow the most precise measurement to date of the transition to homogeneity and large-scale modes of the power spectrum. In addition, there are prospects for measuring the FP relation with SDSS and BOSS galaxies, to derive distances for measuring peculiar velocities. If successful, this would allow for high-redshift peculiar velocity measurements, which would greatly extend their use as a cosmological probe. It would allow tests of growth of structure at high redshift, which so far is only possible with redshift-space distortions. Since peculiar velocities directly trace the density field, while redshift-space distortions are complicated by galaxy bias, this would be very useful. The uncertainties on such peculiar velocity measurements would be very large at such high redshift, but is possible this could be mitigated by the large number of objects. Simulations or predictions are needed to determine whether the signal would be significant from such measurements.

Further in the future, the Square Kilometre Array (SKA) telescope is a huge radio project that will combine telescopes in both Australia and South Africa using interferometry to detect the distribution of hydrogen in the Universe up to the epoch of reionisation (Blake et al., 2004; Schilizzi et al., 2008; Rawlings, 2011). Following on from WALLABY, it will provide both information on the galaxy distribution, and a huge new sample of $\text{H} I$ linewidths that should allow us to map out the velocity field of the local and higher-redshift Universe in incredibly precise detail.

These surveys will greatly expand our knowledge about the Universe, and allow for even more precise tests of $\Lambda$CDM and the forces driving cosmic acceleration.
Appendices
A Gaussian vs true probability distributions $p_D[D_2(r)]$

In our likelihood analysis for the homogeneity scale (Section 2.6.3) we assume that the distribution of the 100 lognormal realisations in each bin, $p_D[D_2(r)]$, is Gaussian. This allows us to interpolate between the data points and errors, so create more finely-spaced $p_D[D_2(r)]$ distributions, in order to determine a smoother PDF for the homogeneity scale, $P(R_H \leq r)$. However, it is not obvious that these distributions should be Gaussian. We therefore repeat the analysis, but use the true distributions given by the lognormal realisations.

This gives the $p_D[D_2(r)]$ distributions shown in Figure 5.1. They are not smooth Gaussians, although they are close to Gaussian. Their resolution is limited by the number of lognormal realisations, so they could be improved by using more lognormal realisations, although we do not do this here.

We then use these to calculate the PDF for the homogeneity scale, $P(R_H \leq r)$, in the same way as we did in Section 2.6.3 for the Gaussian distributions. This gives the PDFs shown in Figure 5.2 for each redshift slice. The mean values and errors are shown in Table 5.1 along with those from our original analysis. The values are very similar, though the errors are larger. This is because we cannot interpolate between data points as easily, and there are a finite number of lognormal realisations contributing to the distribution for each datapoint. This means the distribution is effectively smoothed, giving larger uncertainties.
Figure 5.1: The probability distributions $p_D[D_2(r)]$ for each of the $r$ bins (blue-to-red gradient indicates small to large radius) in the $0.5 < z < 0.7$ redshift slice. Unlike in Figure 2.10, we do not assume these are Gaussian distributions; rather, we plot the distributions given by our 100 lognormal realisations.

Figure 5.2: Probability distributions for the homogeneity scale, $p(R_H)$, for WiggleZ galaxies in each of the four redshift slices. These are calculated from the probability distributions $p_D[D_2(r)]$ of the 100 lognormal realisations, rather than assuming Gaussians.
**B. DERIVATION OF PECULIAR VELOCITY PROBABILITY DISTRIBUTION** $P(v)$ **FROM THE DISTANCE PROBABILITY DISTRIBUTION** $P(\eta)$

Table 5.1: Comparison of the most probable $R_H$ values from a likelihood analysis using the true $p_D[D_2(r)]$ distributions from lognormal realisations, and assuming Gaussian distributions.

<table>
<thead>
<tr>
<th>Redshift</th>
<th>$R_H$ for true $p_D[D_2(r)]$ distributions [h$^{-1}$Mpc]</th>
<th>$R_H$ assuming Gaussians [h$^{-1}$Mpc]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1 &lt; z &lt; 0.3$</td>
<td>$79 \pm 19$</td>
<td>$71 \pm 8$</td>
</tr>
<tr>
<td>$0.3 &lt; z &lt; 0.5$</td>
<td>$71 \pm 13$</td>
<td>$70 \pm 5$</td>
</tr>
<tr>
<td>$0.5 &lt; z &lt; 0.7$</td>
<td>$91 \pm 16$</td>
<td>$81 \pm 5$</td>
</tr>
<tr>
<td>$0.7 &lt; z &lt; 0.9$</td>
<td>$84 \pm 16$</td>
<td>$75 \pm 4$</td>
</tr>
</tbody>
</table>

B Derivation of peculiar velocity probability distribution $P(v)$ from the distance probability distribution $P(\eta)$

The output of the Fundamental Plane fitting ([Springob et al., 2014](#)) was a log distance probability distribution for each galaxy, $P(\eta)$, where

$$\eta = \log_{10}(D_z/D_r).$$  \hfill (5.1)

Here, $D_z$ is the redshift-space distance calculated from the observed redshift $z$, and $D_r$ is the co-moving distance corresponding to the angular diameter distance inferred from the Fundamental Plane.

We want to convert $P(\eta)$ into a probability distribution of peculiar velocity, $P(v)$. We start from the relation

$$1 + z = (1 + z_r)(1 + z_p),$$  \hfill (5.2)

where $z$ is the total, observed redshift, $z_r$ is the redshift corresponding to $D_r$ in the
assumed cosmology, and the peculiar velocity redshift $z_p$ is defined by $v_p = cz_p$. We
can then write the peculiar velocity in terms of these redshifts as:

$$v_p/c = (z - z_r)/(1 + z_r).$$  \hspace{1cm} (5.3)

For each galaxy, $z$ is known. So we can use this equation to determine the
relation for $v$ in terms of $z_r$, hence $D_r$, and hence $\eta$ (since $D_z$ is known).

We can now relate the different probability densities as follows. We need to
calculate

$$P(v) = P(\eta) \frac{d\eta}{dv_p} = P(\eta) \frac{d\eta}{dD_r} \frac{dD_r}{dv_p} = P(\eta) \frac{d\eta}{dD_r} \frac{dD_r}{dz_r} \frac{dz_r}{dv_p}.$$  \hspace{1cm} (5.4)

We can express $P(v)$ in terms of the probability density of $z_r$, as

$$P(v) = P(z_r) \frac{dz_r}{dv_p} = P(z_r) \frac{-(1 + z_r)^2}{c(1 + z)}.$$  \hspace{1cm} (5.5)

Then we substitute in

$$P(z_r) = P(\eta) \frac{d\eta}{dD_r} \frac{dD_r}{dz_r}.$$  \hspace{1cm} (5.6)

We have

$$\frac{d\eta}{dD_r} = \frac{-1}{D_r \ln(10)},$$  \hspace{1cm} (5.7)

since $D_z$ is a constant for each galaxy, and $dD_r/dz_r$ can be calculated using numerical
differentiation.
C. MINIMUM VARIANCE BULK FLOW METHOD FROM WATKINS ET AL. (2009)

So substituting Equations 5.5-5.7 into Equation 5.4 gives

\[ P(v) = P(\eta) \frac{1}{[D_r \ln(10)]} \frac{dD_r}{dz_r} \frac{(1 + z_r)^2}{c(1 + z)^2} \]  \hspace{1cm} (5.8)

where for each \( v \), each term is evaluated at the corresponding values of \( z_r, D_r \) and \( \eta \) derived from Equation 5.3.

The reverse conversion is

\[ P(\eta) = P(v)D_r \ln(10) \frac{dz_r}{dD_r} \frac{c(1 + z)}{(1 + z_r)^2}. \]  \hspace{1cm} (5.9)

C Minimum Variance Bulk Flow Method from Watkins et al. (2009)

The Minimum Variance (MV) method (Watkins et al., 2009; Feldman et al., 2010) provides a way of estimating the bulk flow for a given peculiar velocity sample, by taking into account both the measurement errors of the sample, and the survey geometry. It accounts for the survey geometry by weighting the sample to match an ‘idealised’ survey, with a spherically symmetric distribution of objects with a given radial distribution function and exact line-of-sight velocities. It therefore provides a way of measuring the bulk flow on a specified scale from a given sample, and to compare this measurement between samples of different geometry, as well as with a theoretical prediction.

The bulk flow is a 3D vector defined by

\[ \mathbf{U} = (U_1 \hat{x}_1, U_2 \hat{x}_2, U_3 \hat{x}_3). \]  \hspace{1cm} (5.10)
In practice, we have a data set consisting of positions \( r_n = (x_n, y_n, z_n) \) and measured velocities \( S_n \). The measured line-of-sight velocity is assumed to have the form \( S_n = v_n + \delta_n \), where \( v_n \) is the true velocity and \( \delta_n \) is drawn from a Gaussian distribution with variance \( \sigma_n^2 + \sigma^2_\ast \). Here \( \sigma_n \) is the measurement error of the \( n^{th} \) galaxy, and \( \sigma_\ast \) is the velocity noise, which accounts for smaller scale flows not included in the model.

The bulk flow estimate is given by

\[
    u_i = \sum_{n=1}^{N} w_{i,n} S_n
\]

where \( w_{i,n} \) are the MV weights for each of the three bulk flow components \( i \). Given an idealised survey with bulk flow moments \( U_i \), the MV method aims to determine weights \( w_{i,n} \) such that \( u_i \) gives the best possible estimate for \( U_i \); in other words, it aims to minimise the average variance \( \langle (U_i - u_i)^2 \rangle \).

To calculate the weights, the authors apply constraints to ensure that the estimator gives the correct average amplitudes for the velocity moments, i.e. \( \langle u_i \rangle = U_i \), of the form

\[
    \sum_n w_{i,n} g_j(r_n) = \delta_{ij}.
\]

Here, \( g_j(r) \) are the mode functions corresponding to given moments of the velocity field; for the three bulk flow moments, they are

\[
    g_j(r) = \{ \hat{r}_x, \hat{r}_y, \hat{r}_z \}.
\]

The authors implement the set of constraints in Equation 5.12 using Lagrange mul-
C. MINIMUM VARIANCE BULK FLOW METHOD FROM WATKINS ET AL. (2009)

tipliers, and so the quantity to be minimised is

\[ \langle (U_i - u_i)^2 \rangle + \sum_j \lambda_{ij} \left[ \sum_n w_{i,n} g_j(r_n) - \delta_{ij} \right]. \]  

(5.14)

Doing this, the \( N \)-dimensional vector of weights specifying the \( i \)th moment \( u_i \) is calculated to be:

\[ w_i = (G + \lambda P)^{-1} Q_i. \]  

(5.15)

The terms in this equation are described below.

C.1 G, P and \( f_{mn}(k) \)

The covariance matrix for the individual velocities, \( G_{nm} = \langle S_n S_m \rangle \) can be calculated for a given power spectrum. In linear theory it can be written in terms of the velocity field \( v(r) \) as

\[
G_{nm} = \langle S_n S_m \rangle = \langle v_n v_m \rangle + \delta_{nm} (\sigma_n^2 + \sigma_m^2). \]

(5.16)

The first, 'geometrical' term can be expressed as an integral over the density power spectrum \( P(k) \):

\[
\langle v_n v_m \rangle = \frac{f(\Omega_m)^2 H_0^2}{2\pi^2} \int dk \ P(k) f_{mn}(k), \]

(5.17)

where \( H_0 \) is the Hubble constant in units of \((h \text{ km s}^{-1}\text{Mpc}^{-1})\), and the function \( f_{mn}(k) \) is the angle averaged window function,

\[
f_{mn}(k) = \int \frac{d^2 \hat{k}}{4\pi} (\hat{r}_n \cdot \hat{k})(\hat{r}_m \cdot \hat{k}) \times \exp[i k \cdot (r_n - r_m)]. \]

(5.18)
$P_{nm}$ is the $k = 0$ limit of $f_{mn}$:

$$P_{nm} = \int \frac{d^2 k}{4\pi} (\hat{r}_n \cdot \hat{k})(\hat{r}_m \cdot \hat{k}).$$ \hspace{1cm} (5.19)$$

Equations 5.18 and 5.19 can be evaluated analytically as follows. We have that $\hat{r}_n = (\sin \theta_n \cos \phi_n, \sin \theta_n \sin \phi_n, \cos \theta_n)$ and $\hat{k} = (\sin \theta_k \cos \phi_k, \sin \theta_k \sin \phi_k, \cos \theta_k)$. So

$$ (\hat{r}_n \cdot \hat{k})(\hat{r}_m \cdot \hat{k}) = (\sin \theta_n \cos \phi_n \sin \theta_k \cos \phi_k + \sin \theta_n \sin \phi_n \sin \theta_k \sin \phi_k + \cos \theta_n \cos \theta_k) \times$$

$$\times (\sin \theta_m \cos \phi_m \sin \theta_k \cos \phi_k + \sin \theta_m \sin \phi_m \sin \theta_k \sin \phi_k + \cos \theta_m \cos \theta_k).$$ \hspace{1cm} (5.20)$$

If, for each pair of galaxies, we rotate to the frame in which both galaxies have azimuthal angle $\phi = 0$, then this simplifies to:

$$ (\hat{r}_n \cdot \hat{k})(\hat{r}_m \cdot \hat{k}) = (\sin \theta_n \sin \theta_k \cos \phi_k + \cos \theta_n \cos \theta_k) \times$$

$$\times (\sin \theta_m \sin \theta_k \cos \phi_k + \cos \theta_m \cos \theta_k).$$ \hspace{1cm} (5.21)$$

Ma et al. (2011) show (in their appendix) that:

$$ (\hat{r}_n \cdot \hat{k})(\hat{r}_m \cdot \hat{k}) = \cos \theta (\sin \alpha \sin \theta \sin \phi + \cos \alpha \cos \theta)$$ \hspace{1cm} (5.22)$$

185
and
\[
f_{nm} = \frac{1}{3} \cos \alpha (j_0(kA) - 2j_2(kA)) + \frac{1}{A^2} j_2(kA) r_n r_m \sin^2 \alpha \tag{5.23}
\]
where
\[
A = [r_1^2 + r_2^2 - 2r_1 r_2 \cos \alpha]^{1/2} \tag{5.24}
\]
and \(\alpha\) is the angle between the two galaxies,
\[
\alpha = \arccos (\hat{r}_n \cdot \hat{r}_m). \tag{5.25}
\]
The matrix \(P_{nm}\) is just the \(k = 0\) limit of \(f_{mn}\):
\[
P_{nm} = \frac{1}{3} \cos \alpha \tag{5.26}
\]

since the zeroth and second order Bessel functions evaluated at zero are \(j_0(0) = 1\) and \(j_2(0) = 0\), respectively.

### C. 2 Q

The correlation matrix \(Q_{i,n}\) is calculated in a similar way, but incorporates the window function of the input ‘ideal’ survey. It is evaluated by generating an ideal survey with \(N'\) random positions \(r'_{n'}\) with the desired radial distribution function. \(Q_{i,n}\) is then given by
\[
Q_{i,n} = \langle U_i v_n \rangle = \sum_{n'=1}^{N'} w_{i,n'}' \langle v_{n'} v_n \rangle. \tag{5.27}
\]
The weights \( w'_{i,n} \) for the ideal survey simply give the bulk flow as the average of the projections of the radial velocities on the three coordinate axis directions,

\[
w_{i,n} = \frac{3 \hat{x}_i \cdot \hat{r}_n}{N}.
\] (5.28)

(Note in WFH09 the factor of 3 has been omitted from this equation). Following Watkins et al. (2009) we create an ‘ideal’ survey with \( N' = 10^4 \) and a Gaussian radial density \( n(r) \propto \exp(-r^2/2R_I^2) \), where \( R_I \) is the effective radius of the Gaussian.

Then, we evaluate \( \langle v_{n'}v_n \rangle \) by

\[
\langle v_{n'}v_n \rangle = \frac{f(\Omega_m)^2 H_0^2}{2 \pi^2} \int dk P(k) f_{n'n}(k).
\] (5.29)

### C.3 \( w \) and \( \lambda \)

Feldman et al. (2010) show that the weights can be evaluated as:

\[
w_{i,n} = \sum_m G^{-1}_{nm} \left( Q_{im} - \frac{1}{2} \sum_j \lambda_{ij} g_j(r_m) \right),
\] (5.30)

with the \( \lambda_{ij} \) given by

\[
\lambda_{ij} = \sum_{l=1}^3 \left[ M^{-1}_{il} \left( \sum_{m,n} G^{-1}_{nm} Q_{lm} g_j(r_n) - \delta_{lj} \right) \right]
\] (5.31)

where

\[
M_{ij} = \frac{1}{2} \sum_{n,m} G^{-1}_{nm} g_i(r_n) g_j(r_m).
\] (5.32)
C. MINIMUM VARIANCE BULK FLOW METHOD FROM WATKINS ET AL. (2009)

For the bulk flow, with \( g_i(r) = \hat{r}_i \), this equation becomes

\[
M_{ij} = \frac{1}{2} \sum_{n,m} G_{nm}^{-1} \hat{r}_i(n) \hat{r}_j(m). \tag{5.33}
\]

C.4 Window Functions

The angle-averaged tensor window function is

\[
W_{ij}^2(k) = \sum_{n,m} w_{i,n} w_{j,m} \int \frac{d^2 \hat{k}}{4\pi} (\hat{r}_n \cdot \hat{k})(\hat{r}_m \cdot \hat{k}) \times 
\exp[i k \hat{k} \cdot (r_n - r_m)]
= \sum_{n,m} w_{i,n} w_{j,m} f_{mn}(k). \tag{5.34}
\]

For the case \( i = j \), this gives the angle-averaged window function for the moment \( u_i \),

\[
W_{ii}^2(k) = \sum_{n,m} w_{i,n} w_{i,m} f_{mn}(k). \tag{5.35}
\]

C.5 Bulk flow uncertainties

The covariance matrix of the bulk flow moments, \( R_{ij} \), can be written as:

\[
R_{ij} = \langle u_i u_j \rangle = \sum_{mn} w_{im} w_{jn} G_{mn} = R_{ij}^{(e)} + R_{ij}^{(v)}, \tag{5.36}
\]

where \( R_{ij}^{(e)} \) represents the noise contribution,

\[
R_{ij}^{(e)} = \sum_n w_{i,n} w_{j,n} (\sigma_n^2 + \sigma_v^2), \tag{5.37}
\]
and $R_{ij}^{(v)}$ represents the cosmic variance contribution,

$$R_{ij}^{(v)} = \frac{f(\Omega_m, z) H_0^2}{2\pi^2} \int_0^\infty dk W_{ij}(k) P(k). \quad (5.38)$$

The errors on the bulk flow moments, $\sigma_i$, are then $\sigma_i = \sqrt{R_{ii}}$, and the error on the bulk flow magnitude is $\sigma_U^2 = JR_{ij}J^T$, where $J$ is the Jacobian of the bulk flow magnitude $U$,

$$J = \left( \frac{\partial U}{\partial u_x}, \frac{\partial U}{\partial u_y}, \frac{\partial U}{\partial u_z} \right) = \frac{1}{U} (u_x, u_y, u_z). \quad (5.39)$$
Bibliography


Alpher, R. A., & Herman, R. C. 1948, Physical Review, 74, 1737


Baumann, D. 2009, ArXiv e-prints, 0907.5424

Behrend, J., Brown, I. A., & Robbers, G. 2008, JCAP, 1, 13, 0710.4964


BIBLIOGRAPHY


——. 2008, General Relativity and Gravitation, 40, 467, 0707.2153


192


Dodelson, S. 2003, Modern cosmology


Einstein, A. 1916, Annalen der Physik, 354, 769


Gabrielli, A. 2005, Statistical physics for cosmic structures, Lecture Notes in Physics (Springer)


195
BIBLIOGRAPHY


Hannestad, S., Haugbølle, T., & Thomsen, B. 2008, JCAP, 2, 22, 0705.0979


Hart, L., & Davies, R. D. 1982, Nat, 297, 191


Heavens, A. F., Jimenez, R., & Maartens, R. 2011, JCAP, 9, 35, 1107.5910


Hong, T. et al. 2014, submitted


Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168


197


Maartens, R. 2011, Royal Society of London Philosophical Transactions Series A, 369, 5115, 1104.1300


Magoulas, C. 2012, PhD thesis, The University of Melbourne


——. 2014, in preparation

Marinoni, C., Bel, J., & Buzzi, A. 2012, JCAP, 10, 36, 1205.3309


200


McDonald, P., & Seljak, U. 2009, JCAP, 10, 7, 0810.0323


Peacock, J. A. 1999, Cosmological Physics, ed. Peacock, J. A.


Peebles, P. J. E. 1980, The large-scale structure of the universe


Planck Collaboration et al. 2013, ArXiv e-prints, 1303.5062

Planck collaboration et al. 2013, ArXiv e-prints, 1303.5075

Planck Collaboration et al. 2013a, ArXiv e-prints, 1303.5076
——. 2013b, ArXiv e-prints, 1303.5090


Pskovskii, I. P. 1977, SOVAST, 21, 675

——. 2011, Classical and Quantum Gravity, 28, 164008, 1102.0408

Rawlings, S. 2011, ArXiv e-prints, 1105.6333


Ruiz-Lapuente, P. 2010, Dark Energy: Observational and Theoretical Approaches


Song, Y.-S., Sabiu, C. G., Kayo, I., & Nichol, R. C. 2011, JCAP, 5, 20, 1006.4630


Springob, C. M. et al. 2014, submitted (S14)


Sylos Labini, F. 2011, EPL (Europhysics Letters), 96, 59001, 1110.4041


——. 2007b, Physical Review Letters, 99, 251101, 0709.0732

——. 2009, Phys. Rev. D, 80, 123512, 0909.0749


Zwicky, F. 1933, Helvetica Physica Acta, 6, 110