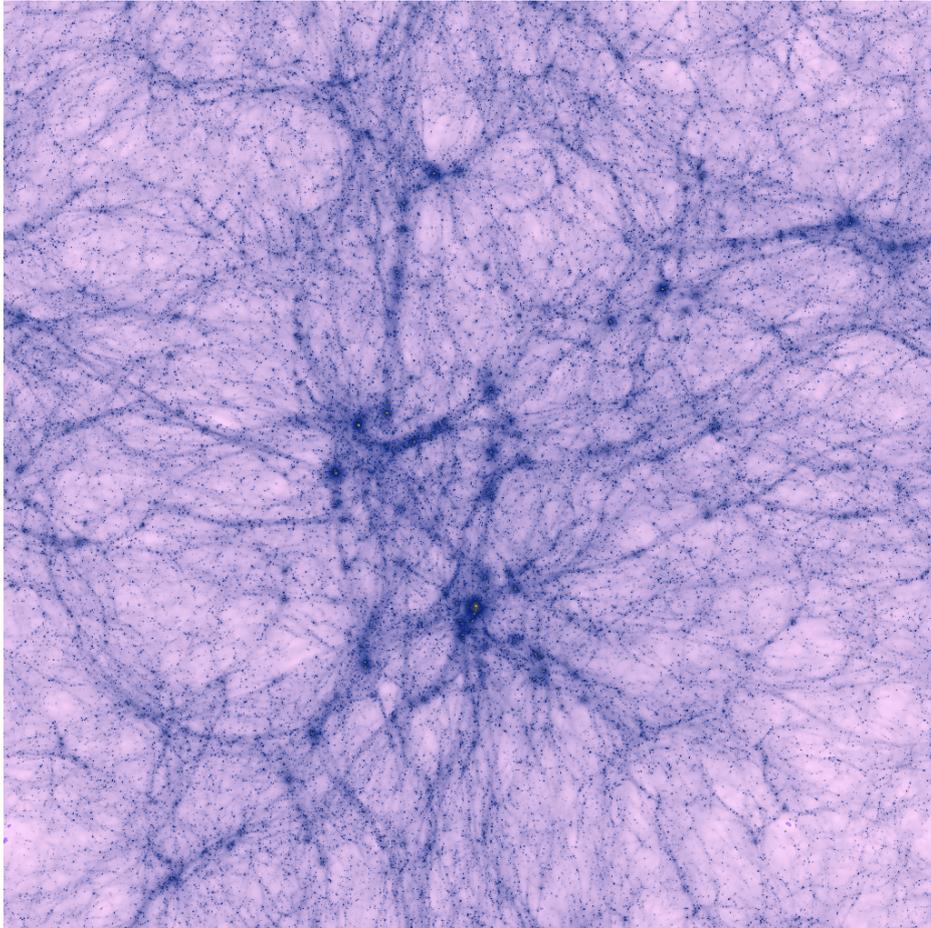


Doing Cosmology with Bulk Flow Magnitudes: Observational Effects



Credit: Project Horizon

Author: Per Andersen - blk555
Supervisor: Prof. Jens Hjorth

University of Copenhagen
August 3rd, 2015

Contents

1	Introduction	2
1.1	Brief Summary of Concordance Cosmology	2
1.2	Observables - What do we observe?	2
1.3	Scope of Thesis - Why do we observe?	3
2	Linear Theory	5
2.1	Comparison with Theory	7
3	Geometry Defined	8
4	Maximum Likelihood & Minimum Variance Bulk Flows	10
4.1	Maximum Likelihood Estimate (MLE)	10
4.2	Minimum Variance (MV)	12
5	Simulation: Horizon Run 2	15
6	Sampling, Completeness & Cosmic Variance	16
7	Mock Galaxy Surveys versus Dark Matter Halos	21
8	Dependence on survey geometry	23
9	Discussion	27
A	Estimating Peculiar Velocity Measurement Uncertainty	31
B	Computational Considerations	32
C	Bulk Flow Estimate Covariance matrix	33

Abstract

In this thesis we investigate the viability of comparing bulk flow magnitudes to predictions from linear theory, in order to test Λ CDM cosmology. This investigation is motivated by a current disagreement with Λ CDM by some bulk flow measurements in the literature, claiming observations of bulk flows larger than predicted in Λ CDM. We focus on observational effects that might bias our measurements of bulk flows, such as the effects of survey geometry, and the asymmetry of the bulk flow magnitude distribution. We find that all techniques to measure a value for the bulk flow overestimate its magnitude, and this is made worse both by sparse sampling and by incomplete sky coverage. Further work is needed before these effects are fully understood, and until that work has been carried out we conclude that the standard χ^2 analysis of the bulk flow vector components is still the most viable method to compare bulk flow measurements with theory. Finally, future work should perform the same analysis but calculate cosmological parameters rather than bulk flow velocity, and see if the same biases exist. Since our study shows observational effects will cause false detections of large bulk flows, it is critical to confirm that the same observational effects do not bias cosmological parameters.

Foreword

I would like to use this opportunity to thank my supervisor Prof. Jens Hjorth for his patience and help throughout the length of this project. Additionally I would like to express my deepest gratitude to Prof. Tamara Davis for her extensive guidance and feedback in all steps of this project.

1 Introduction

1.1 Brief Summary of Concordance Cosmology

The model that is currently the main contender for describing the evolution and structure formation of the universe is the Λ CDM model, where Λ is a dark energy component causing accelerated expansion of the universe and CDM is a Cold Dark Matter component dominating at the early stages of cosmological evolution. The Λ CDM model gained favour after Perlmutter et al. (1999) and Riess et al. (1998) observed a number of standardisable type Ia supernovae (SNe) and found that a Λ CDM cosmology with $\Omega_m \sim 0.3$ and $\Omega_\Lambda \sim 0.7$ was the best fit model to the data. Here Ω_m and Ω_Λ are the unitless matter and dark energy density components. The Λ CDM cosmology has been successful at explaining observations from the Cosmic Microwave Background (CMB), Baryon Acoustic Oscillations (BAO) and a number of standardisable candles, where type Ia SNe are currently the most prominent one. Where the Λ CDM model falls short is in explaining the actual physics behind the accelerated expansion of the universe, giving no physical justification for introducing the Λ parameter, which has gone in and out of favour in the astrophysical community since Einstein introduced it in 1917 (Einstein, 1917).

Additionally, tension currently exists in measurements of the cosmological bulk flow; the bulk flow of our local group with regards to the CMB. Some of these measurements are in apparent agreement with Λ CDM (Carrick et al., 2015; Colin et al., 2011; Dai et al., 2011; Hong et al., 2014; Lavaux et al., 2013; Ma & Pan, 2014; Ma & Scott, 2013; Nusser & Davis, 2011; Osborne et al., 2011; Planck Collaboration et al., 2014; Turnbull et al., 2012) where others are not (Abate & Feldman, 2012; Feix et al., 2014; Feldman et al., 2010; Kashlinsky et al., 2008; Watkins et al., 2009). Relieving this tension is important if we are to gain physical insight into whether dark energy actually exists, and if it does then what the nature of it is. It does not appear that cosmologists will be left with nothing productive to do in the near future, as this is not a simple task.

1.2 Observables - What do we observe?

An excellent way to probe cosmology is to measure galaxy peculiar velocities, and combine those measurements into a bulk flow of our local group of galaxies. The bulk flow refers to the total net velocity of a region of space with respect to the cosmic microwave background rest frame, and is represented by a vector with an element for each of the three spatial dimensions. Imprinted in the galaxy peculiar velocities is the fingerprint of large scale structure formation, which in turn is determined by the underlying cosmology. So by measuring the local bulk flow through peculiar velocities we gain constraining power on the underlying cosmology. A very popular probe of peculiar velocities is the standardisable candle type

Ia supernovae (SNe). In the following chapters the source of the peculiar velocity measurement is assumed to be one behaving like type Ia SNe, although the methods derived are equally valid for other distance indicators as long as they measure position on the sky, distance and peculiar velocity.

As mentioned above the pursued observable is the peculiar velocity of nearby galaxies. What we directly observe is the redshift of the host galaxy, z_{obs} , as well as the distance modulus, μ_{obs} , derived from fitting an empirical model to the light curve of the type Ia SN. To convert the distance modulus to a distance and recession redshift, z_r , we need to assume a cosmological model. See Fig. 1 for a schematic overview of the measurement method. This makes it more difficult to use type Ia SNe observations to distinguish between different models and determine which is the better model, but it has little effect if the objective is to perform best parameter estimation for a specific model. Finally we can combine the recession and observed redshift to derive the line-of-sight peculiar velocity of the galaxy

$$v = \mathbf{v} \cdot \hat{\mathbf{r}} = c \left(\frac{z_{obs} - z_r}{1 + z_r} \right) \quad (1)$$

where c is the speed of light in vacuum, z_{obs} is the observed redshift of the galaxy and z_r is the redshift due to the comoving expansion of space, also known as the Hubble flow. \mathbf{v} is the full velocity vector and $\hat{\mathbf{r}}$ is the unit vector pointing from the observer to the observed object. A small but important point is to not use the simple Hubble law $v = cz - H_0 d$, as is sometimes incorrectly done in the literature. The simple Hubble law is only appropriate for small redshifts, and will introduce a significant error on larger scales (Davis & Scrimgeour, 2014).

1.3 Scope of Thesis - Why do we observe?

The field of using large scale bulk flows to constrain cosmology has historically been plagued by systematic errors and poor results. As we begin using distance probes where the systematics are understood better, such as the type Ia SNe, we move into a regime where bulk flow measurements could become a highly valued addition to the effort of constraining the parameter space of models such as the Λ CDM. Traditionally a standard χ^2 analysis has been used to compare observations of the three individual bulk flow moments to the theoretical prediction of zero bulk flow that comes from assuming the cosmological principle to be true at all scales. When this analysis yields a larger than expected bulk flow it is important to be clear about what exactly that means. Does it mean that σ_8^1 is larger than expected? That bulk flows indicate that structure formation has proceeded faster than expected? Or does it mean that bulk flows are faster than expected

¹ σ_8 is the rms mass fluctuation amplitude in spheres of size 8 Mpc h^{-1} and measures the normalisation of the matter power spectrum.

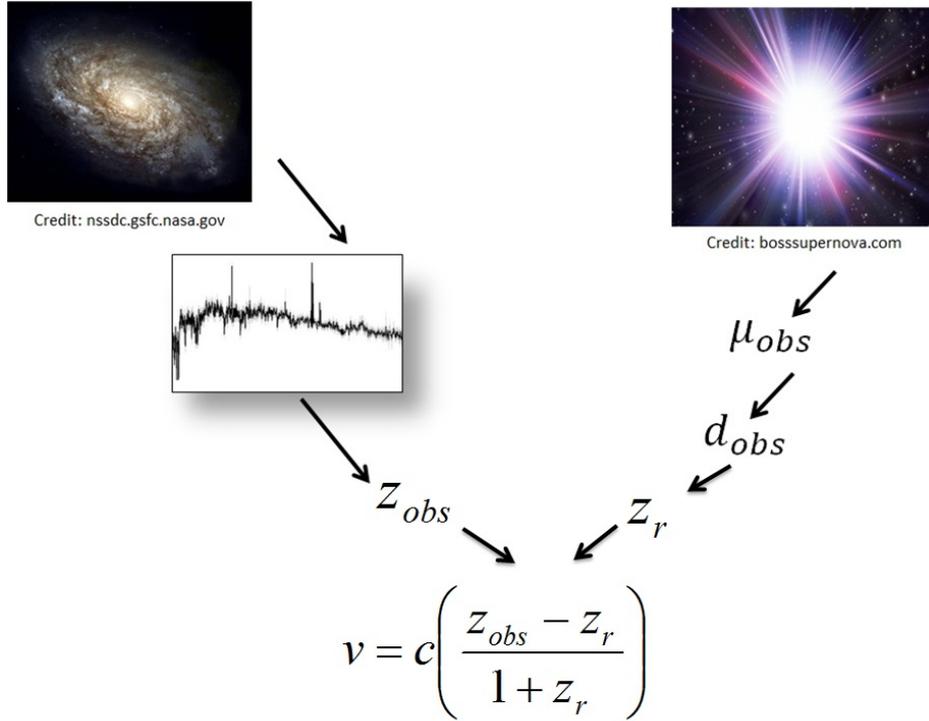


Figure 1: Schematic view of the measurements and method needed to detect peculiar velocities of nearby galaxies. By observing galaxies and type Ia SNe within those galaxies we can measure the peculiar velocities and use those observations to test Λ CDM.

given the structure we see? In one case the bulk flow has revealed a larger than expected structure, in the other case it indicates that the gravitational influence of structures is larger than we expected in standard Λ CDM cosmology. In this thesis we examine what it means to say that a measured bulk flow is larger than expected in Λ CDM. Through the use of a cosmological simulation we look at the predictions Λ CDM makes for the velocity distribution of galaxies in a given volume of space and note how this prediction should be dependent on the geometry of that volume, i.e. how the velocity distribution changes for equal volumes but different shapes.

As we shall see in chapter 2 we have an alternative to the standard χ^2 analysis of the bulk flow vector components from linear theory where the magnitude of the bulk flow measurement can be directly compared to the theoretical prediction of linear theory.

When working with bulk flow vector magnitudes, it is important to take into

consideration that the distribution of possible bulk flow magnitudes is not symmetric around zero, since only positive values are permitted. The main focus of this work is to investigate the effects of this asymmetry, as well as the effects of survey geometry, in the context of comparing measurements of the local bulk flow magnitude with the theoretical prediction made by linear theory. In chapter 2 we introduce linear theory as well as explain how we could practically compare an observation of bulk flow magnitude to the theoretical prediction. In chapter 3 we define exactly what is meant with geometry in this survey, and introduce the spherical cone geometry that we will work with in this thesis. Then in chapter 4 we will review the two methods of estimating a bulk flow from peculiar velocity observations that we apply in this thesis: the Maximum Likelihood Estimate (MLE) and the Minimum Variance (MV) method. These are the most common bulk flow estimators used in the literature. In chapter 5 we explain the details of the cosmological simulation used in this work, the Horizon Run 2 simulation. The method that we use to test the effects of survey geometry on our bulk flow magnitude measurements is explained in chapter 6, and in chapter 8 we apply that method to bulk flow magnitude measurements for various geometries using the MLE and MV method. Finally in chapter 9 we discuss and summarise the findings of this thesis.

2 Linear Theory

Before delving deeper into converting peculiar velocity observations to a bulk flow we will have a look at the theoretical background needed to compare our measured bulk flow magnitude with a theoretical prediction.

We start by applying the cosmological principle and assuming that the universe is statistically isotropic and homogeneous. This leads us to predict that the mean bulk flow velocity at any location is zero. The root mean square (rms) variance will however not be zero, and depends on the matter power spectrum, the scale at which it is measured as well as the window function for that measurement. The matter power spectrum is a way to measure structure in statistical fluctuations in the cosmological density field. It is measured in units of k which is given by $k = 2\pi/r$ (where r is the radial extent defined in chapter 3) and is thus inverse of length. This inverse relationship corresponds to looking at large scale fluctuations when we have small k values, and small scale fluctuations when k is large. The window function is a function of k that measures how sensitive we are to measuring the statistical fluctuations at a particular scale, given by the input value of k . If the window function is large for a particular k value it means that we are highly sensitive to measuring fluctuations at the scale k represents; the opposite is true as well with a smaller value of the window function for a particular k value

indicating less sensitivity to measuring fluctuations at the scale of k . Understanding the influence of the window function on our measurement is not trivial, but is nonetheless important when measuring bulk flow magnitudes, particularly for the Minimum Variance (MV) method.

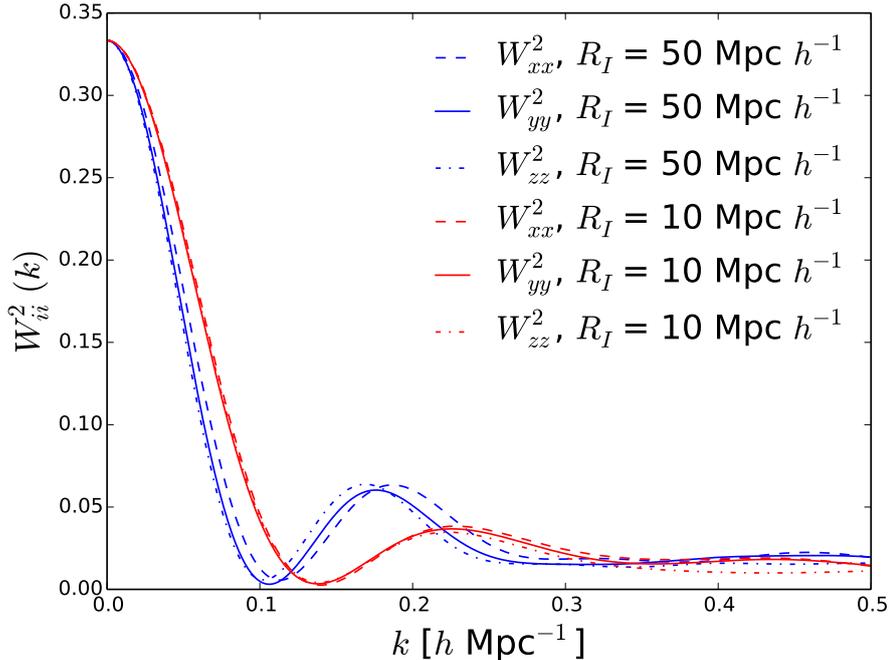


Figure 2: Window function for a spherical mock survey containing 100 galaxies. The window function has been calculated using Minimum Variance (MV) weights with an ideal radius of 50/10 $\text{Mpc } h^{-1}$ for the blue/red lines. The full, dashed and dash-dotted lines represent the three spatial components of the bulk flow vector.

If we assume that we are in the linear regime where the infall of mass is a simple process that includes no rotation or other non-linear terms², which is typically a good assumption on large enough scales, then linear theory tells us that the rms variance should be (Coles & Lucchin, 2002)

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

where the Hubble constant, H_0 , growth rate, f , and matter power spectrum $P(k)$ define our particular cosmology and $\widetilde{W}(k; R)$ is the Fourier transform of the win-

²The extreme opposite of the linear regime is the virial regime, which is typically only valid on small scales.

dow function, $W(R)$, of the specific bulk flow measurement. The window function will be dependent on the geometry of the measurements taken to derive the bulk flow, and is therefore unique for each particular survey. It is often approximated to be an all sky gaussian window function of the form

$$\widetilde{W}_G = \exp(-(kR)^2/2) \quad (3)$$

which is also what we will use in this work when using Eq. 2 to calculate the theoretical bulk flow variance for a given effective radius. The Fourier transform of a particular survey's window function will only have the same functional form if it is an all-sky survey, spherically symmetric and gaussian with an effective radius R (We will define R more carefully in chapter 3). How strongly the window function of a particular survey will deviate from this gaussian approximation will be determined by the geometry of the survey in question. To calculate the maximum likelihood bulk flow magnitude $V_{ML}(R)$ we assume that the peculiar velocity distribution is Maxwellian (Li et al., 2012) which gives us a probability distribution for the bulk flow amplitude of the form

$$p(V)dV = \sqrt{\frac{2}{\pi}} \left(\frac{3}{\sigma_V^2}\right)^{3/2} V^2 \exp\left(-\frac{3V^2}{2\sigma_V^2}\right) dV. \quad (4)$$

For this distribution the maximum probability is then given by the relation

$$V_{ML}(R) = \sqrt{2/3} \sigma_V(R). \quad (5)$$

When referring to the theoretical bulk flow magnitude throughout this thesis, it is this maximum likelihood value that we are referencing.

2.1 Comparison with Theory

In the above discussion of linear theory we derived an expression for the maximum likelihood bulk flow magnitude as a function of scale. We implicitly made the choice to consider only bulk flow magnitudes; that is we look at the magnitude of the full three dimensional vector and compare that to the theoretical prediction of equation 5. It is important to note that this theoretical model has an associated uncertainty, given by equation 2. Therefore, when using this method to compare with theory we need to compare our observed bulk flow magnitude with that from equation 5 and then take into account that both the observed and expected bulk flow magnitudes have uncertainties associated with them. One advantage of this approach is that we can fold non-linear behaviour into the powerspectrum, $P(k)$, term. This is especially useful on smaller scales where the assumption of being in the linear regime is not strictly held true. However, it requires that the scale at which we perform the measurement, R , is a well defined quantity. Later, in chapter

9, we will review the viability of comparing bulk flow magnitude measurements with linear theory in the fashion explained above, in the context of the main themes of this thesis; the asymmetry of the bulk flow magnitude distribution and survey geometry.

3 Geometry Defined

To determine the effects of survey geometry on measurements of the bulk flow, it is fruitful to first properly define what is meant by geometry. By geometry we refer to the spatial distribution of the galaxies in the survey used to measure the bulk flow. More specifically, to define the geometry of a survey, the redshifts, right ascensions and declinations (or equivalent) need to be known, but not the measured peculiar velocities. A typical survey would observe a patch of the sky and all the galaxies in such a survey would lie within a spherical cone, illustrated in Fig. 3, with a radius given by

$$r = \left(\frac{3V}{2\pi(1 - \cos(\theta/2))} \right)^{1/3}. \quad (6)$$

Here V is the volume and θ the opening angle of the spherical cone. By increasing the angle to $\theta = \pi$ the spherical cone becomes a hemisphere survey, and for $\theta = 2\pi$ it becomes an all-sky survey. Most real surveys can be approximated as having a spherical cone geometry, being either a narrow pencil beam like survey, a hemisphere survey or an all-sky survey. Therefore the geometries investigated in this thesis will be spherical cones like the ones illustrated in Fig. 3 with radius given by Eq. 6.

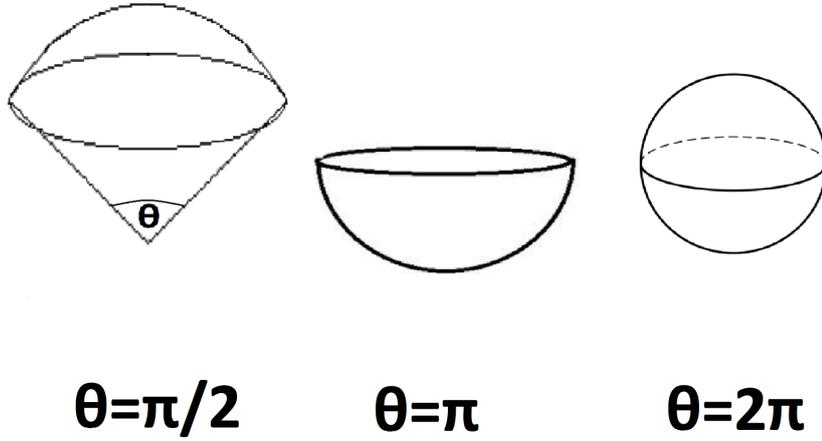


Figure 3: Illustration of the spherical cone geometry. When $\theta = \pi$ the spherical cone becomes a hemisphere, when $\theta = 2\pi$ the geometry is a sphere. Since the volume of the geometry is kept constant, the radius of the spherical cone geometry decreases as θ increases towards the maximum 2π . Likewise, the radius diverges towards positive infinity as θ approaches 0. This puts a practical limit on how small θ can be set.

In the following analysis there are three radii we reference frequently. They are:

1. Radial extent - r
2. Effective radius - R
3. Ideal radius - R_I

The first is the radial extent of the data, $r(z)$. The radial extent is the radius of the largest redshift in the redshift distribution, assuming a certain cosmology to be true; or put in other words, it is simply the distance to the most distant galaxy in the survey. The second radius is the effective radius, R , which defines the radius at which we are aiming to calculate the bulk flow. For a specific survey with surveyed volume V the effective radius is $R = (3V/4\pi)^{1/3}$. This corresponds to converting the surveyed volume into a sphere with volume V and calculating the radius associated with that sphere. The effective radius is often used when comparing results of different surveys with linear theory through Eq. 2. In Fig. 3 we show spherical cones of constant volume but varying opening angle. Since the volume of all these spherical cones is kept constant, the effective radius R will also be constant. The final radius introduced is the ideal radius R_I . The ideal radius is used in the construction of an ideal survey, which is given as an input in the Minimum Variance (MV) method of estimating the bulk flow. The

constructed ideal radius has a radial distribution that follows the functional form $n(r) \propto r^2 \exp(-r^2/2R_I^2)$. In chapter 4 we will go into more detail with the role of the ideal radius, R_I , and its effect on the MV bulk flow method.

The ideal radius R_I and effective radius R are often set approximately equal. It can however often be quite difficult to get a meaningful estimate for the effective radius R for a specific survey, especially if the survey is far from being spherical.

4 Maximum Likelihood & Minimum Variance Bulk Flows

To compare the measured bulk flow with the theoretical prediction, it is necessary to have a method to turn the individually observed peculiar velocities into a bulk flow. In this thesis we focus on two methods, the maximum likelihood (MLE) and the minimum variance (MV) methods.

First let us review the assumptions that are made for both methods. The MLE and the MV methods have in common that they assume that the observational errors on the peculiar velocities are exactly gaussian in nature. Furthermore linear theory is assumed to hold so that the line-of-sight peculiar velocity is much smaller than the cosmological Hubble flow; it will be of order 10-20% of the Hubble flow for redshifts up to $z \sim 0.01$. Finally it is assumed that the non-linear flows on small scales can be taken into account by a constant cosmic variance term, σ_* , which should be smaller than the average measured peculiar velocities; typically it is assumed that $\sigma_* \sim 250 \text{ km s}^{-1}$ which is comparable to the expected bulk flow amplitude. Despite the fact that the assumptions do not always hold true in the strictest sense, measurements of the local bulk flow are nonetheless a valuable probe of cosmology, as long as we look in the limit where the assumptions are true. Furthermore, modern surveys that have just recently been carried out allow us to move into a regime where the assumptions made above will be adhered to. This will make bulk flow measurements even stronger at differentiating between cosmological models that predict the same expansion history than they have been previously.

4.1 Maximum Likelihood Estimate (MLE)

Of the two methods to estimate a bulk flow, the MLE method is by far the easiest to implement and is computationally much cheaper than the MV method. It does however have several disadvantages. Since it is more likely to observe a galaxy in a dense region than in an underdense one, the MLE method will be density

weighted rather than volume weighted. Galaxies that are nearer to us typically have smaller errors on their distance and peculiar velocity measurements, which means that the MLE method will weigh the nearby galaxies higher than the distant ones and probe the local bulk flow on smaller scales. From linear theory we expect a larger bulk flow on average on these smaller scales. This will bias the MLE method towards measuring bulk flows that are larger than the actual underlying bulk flow of the observed volume. Finally it will have a window function that is a complex function of survey geometry and measurement uncertainties, which will be unique for each particular survey. This makes it very difficult to compare MLE estimates of the bulk flow from different surveys. Nevertheless, its ease of use and fairly straightforward interpretation have meant that it is the most common bulk flow estimator used in the literature.

When calculating the Maximum Likelihood Estimate (MLE) of the bulk flow, the final result is a vector containing the velocity components corresponding to each of the three spatial dimensions. Each of the three components is given by a sum over the individual peculiar velocity components multiplied by some weight. The sum has the form

$$u_i = \sum_n w_{i,n} S_n \quad (7)$$

where i is the placeholder for either the x , y or z index and the sum goes over all n peculiar velocities. S_n is the n 'th measured peculiar velocity, $w_{i,n}$ is the associated weight for that peculiar velocity and u_i is the calculated bulk flow where again $i = (x, y, z)$. This equation holds true for both the MLE and the MV methods. Where they differ is how they go about calculating the $w_{i,n}$ weights.

For the MLE the weights are given by

$$w_{i,n} = \sum_j \frac{\hat{x}_j \cdot \hat{r}_n}{(\sigma_n^2 + \sigma_\star^2)} A_{ij}^{-1}. \quad (8)$$

The sum is over the $j = (x, y, z)$ components, and $\hat{x}_j \cdot \hat{r}_n$ is the projection of the unit vector \hat{r} pointing from the observer to the galaxy in question. σ_n is the uncertainty on the velocity of the n 'th measurement, and σ_\star is a constant of order 250 km s^{-1} meant to account for the non-linear flows on smaller scales. Finally A_{ij}^{-1} is the inverse of matrix A_{ij} given by

$$A_{ij} = \sum_n \frac{(\hat{x}_i \cdot \hat{r}_n)(\hat{x}_j \cdot \hat{r}_n)}{\sigma_n^2 + \sigma_\star^2}. \quad (9)$$

In practise when calculating the MLE weights the first step is to calculate the A_{ij} matrix, taking advantage of the symmetry $A_{ij} = A_{ji}$. The inverted matrix

A_{ij}^{-1} is then computed, and the weights $w_{i,n}$ are calculated. This is a fairly simple process, and is cheap in computation time needed. As described earlier the MLE method does have a number of downsides that makes comparisons between different surveys difficult. For those situations the MV method might be a stronger, but computationally more expensive, alternative that alleviates some of the concerns that one might have when using the MLE method.

4.2 Minimum Variance (MV)

The MV method (Watkins et al., 2009) builds upon the MLE method, but constructs weights that are volume weighted, instead of density weighted. It is computationally much more expensive than the MLE method. It incorporates a constructed ideal survey and minimises the variance between the bulk flow measured by the survey sample and the bulk flow that would be measured by this ideal survey. The effect of this is that the MV method has a much simpler window function that is not strongly a function of the survey geometry, and can probe the bulk flow at larger scales where non-linear effects from small scale flows are less significant. The constructed ideal survey is typically an all-sky spherically symmetric gaussian survey with an ideal radius, R_I , that is typically set to 50 Mpc h^{-1} (Feldman et al., 2010; Ma & Scott, 2013; Watkins et al., 2009), but can be set to any value dependent on which scale one wishes to probe the bulk flow at.

The ideal survey is constructed by generating x,y,z coordinates uniformly randomly in the range $[-4R_I, 4R_I]$ and then drawing points according to the distribution $n(r) \propto r^2 \exp(-r^2/2R_I^2)$. This constructed ideal survey is spherically symmetric and isotropic. It is constructed such that the window function of the MV method is sensitive in the range where we wish to probe the bulk flow, namely on scales of R_I . In order to stay consistent R_I will be set to 50 Mpc h^{-1} in this work, unless otherwise stated. The total number N of points in the constructed ideal survey is set to 1200 throughout this work. It was found that increasing the number of points in the ideal survey beyond 1200 did not contribute to the stability of the MV method but only served to increase the already considerable computation time.

Like for the MLE method the MV bulk flow vector components u_i are given by

$$u_i = \sum_n w_{i,n} S_n. \quad (10)$$

The tricky bit is then again to calculate the weights $w_{i,n}$. For $i = (x, y, z)$. For readability we use matrix notation so that $w_{i,n}$ becomes column matrix \mathbf{w}_i of n

elements. We compute \mathbf{w}_i with

$$\mathbf{w}_i = (\mathbf{G} + \lambda\mathbf{P})^{-1}\mathbf{Q}_i. \quad (11)$$

First let's discuss the matrices \mathbf{G} and \mathbf{P} . \mathbf{G} is a symmetric square n by m matrix where n and m correspond to the n 'th and m 'th measurement. The matrix \mathbf{G} is the covariance matrix for the individual velocities S_n and S_m . In linear theory we can write the matrix elements G_{nm} as a sum of two terms

$$G_{nm} = \langle S_n S_m \rangle \quad (12)$$

$$= \langle v_n v_m \rangle + \delta_{nm}(\sigma_*^2 + \sigma_n^2). \quad (13)$$

The second term is known as the noise term and is the Kronecker delta function; 0 for $n \neq m$ but $\sigma_*^2 + \sigma_n^2$ when $n = m$. The first term is the geometry term which is given by

$$\langle v_n v_m \rangle = \frac{\Omega_m^{1.1} H_0^2}{2\pi^2} \int dk P(k) f_{mn}(k) \quad (14)$$

where H_0 is the Hubble constant³ in units of $h \text{ km s}^{-1} \text{ Mpc}^{-1}$, and $\Omega_m^{1.1}$ is the growth of structure parameter $f^2 \approx \Omega_m^{1.1}$. $P(k)$ is the matter power spectrum, which in this work is calculated using CAMB⁴ (Lewis et al., 2000). The function $f_{mn}(k)$ is the angle averaged window function which is explicitly given as

$$f_{mn}(k) = \int \frac{d^2 \hat{k}}{4\pi} (\hat{\mathbf{r}}_n \cdot \hat{\mathbf{k}})(\hat{\mathbf{r}}_m \cdot \hat{\mathbf{k}}) \times \exp[ik\hat{\mathbf{k}} \cdot (\hat{\mathbf{r}}_n - \hat{\mathbf{r}}_m)]. \quad (15)$$

Although Eq. 15 is often quoted in the literature as the function used to calculate $f_{mn}(k)$ it is far from being a practical expression and in reality the expression used is from Ma et al. (2011) who showed that we can express the angle averaged window function as

$$f_{mn}(k) = \frac{1}{3} \cos(\alpha(j_0(kA) - 2j_2(kA))) + \frac{1}{A^2} j_2(kA) r_n r_m \sin^2(\alpha) \quad (16)$$

where

$$A = (r_n^2 + r_m^2 - 2r_n r_m \cos(\alpha))^{0.5} \quad (17)$$

and α is the angle between the n 'th and m 'th galaxy given by

$$\alpha = \arccos(\hat{\mathbf{r}}_n \cdot \hat{\mathbf{r}}_m). \quad (18)$$

³Which is always 100, per definition of $h = (H_0/100) \text{ km s}^{-1} \text{ Mpc}^{-1}$.

⁴<http://camb.info/readme.html>

The $j_0(x)$ and $j_2(x)$ functions are spherical Bessel functions given by

$$j_0(x) = \frac{\sin(x)}{x}, \quad j_2(x) = \left(\frac{3}{x^2} - 1\right) \frac{\sin(x)}{x} - \frac{3 \cos(x)}{x^2}. \quad (19)$$

Putting all this together gives us the G_{nm} elements. Finding the P_{nm} elements of \mathbf{P} is then fairly simple as it is simply the $k = 0$ limit of f_{nm} which is

$$P_{nm} = \frac{1}{3} \cos(\alpha). \quad (20)$$

The principal idea of the MV method is to minimise the variance between the bulk flow measured by the galaxy survey and the bulk flow that would be measured by an ideal survey. The \mathbf{G} and \mathbf{P} matrices that we have explained how to calculate above are the components of the weight calculating that take as input the measured data. The last component, the \mathbf{Q} matrix, takes as input the position and peculiar velocities from the galaxies of the constructed ideal survey. It is calculated in much the same way as the G_{nm} elements with the $Q_{i,n}$ elements being given by

$$Q_{i,n} = \sum_{n'=1}^{N'} w'_{i,n'} \langle v_{n'} v_n \rangle \quad (21)$$

and

$$\langle v_{n'} v_n \rangle = \frac{\Omega_m^{1.1} H_0^2}{2\pi^2} \int dk P(k) f_{n'n}(k) \quad (22)$$

where $f_{n'n}(k)$ is analogous to Eq. 16 but with the difference that n' and n run over the galaxies in the constructed ideal survey, in contrast to n and m that run over the galaxies from the actual observed galaxies of our survey. The ideal weights $w'_{i,n'}$ will be given by

$$w'_{i,n'} = 3 \frac{\hat{\mathbf{x}}_i \cdot \hat{\mathbf{r}}_n}{N} \quad (23)$$

where N is the total number of galaxies in the constructed ideal survey.

Now the final piece of the puzzle is to solve for the value for λ . λ is a Lagrange multiplier inherent from the minimisation process. It enforces the normalisation constraint

$$\sum_m \sum_n w_{i,n} w_{i,m} P_{nm} = \frac{1}{3}. \quad (24)$$

A simple method to solve for λ is to vary λ and calculate the above sum for each of them, until a value for λ that makes the above equality true is found.

Calculating the MV bulk flow vector is a rather involved process and is orders of magnitude more expensive computationally than the MLE method. In this

work we will perform our analysis using both the MLE and the MV method, as both have merit in certain scenarios. The MLE method is favourable when the input data is close to being spherically symmetric, the effects of associated uncertainties are well understood and computation time is a concern. If the data is not close to being spherically symmetric the MV method is a stronger choice, at the cost of computational time. Optimally, both methods should be carried out if possible.

5 Simulation: Horizon Run 2

The cosmological simulation used throughout this thesis is the Horizon Run 2 (HR2) (Kim et al., 2011) containing 216 billion particles spanning a $(7.2 \text{ Gpc } h^{-1})^3$ volume. It is a dark matter (DM) halo simulation, tracing the motions of dark matter particles through time. The mass resolution goes down to $1.25 \cdot 10^{11} M_{\odot} h^{-1}$. This mass resolution allows for resolution of galaxy-size halos with mean particle separation of $1.2 \text{ Mpc } h^{-1}$. The power spectrum, correlation function, mass function and basic halo properties match those predicted by Λ CDM and linear theory to percent level accuracy.

From the full HR2 dataset we draw spherical subsets with radius $1 \text{ Gpc } h^{-1}$. The origin of each subset is chosen randomly, so that some will be chosen in higher than average density regions and some in lower than average density regions. Knowledge of our local galactic surroundings could have been folded into the selection of origins, so that the subsets chosen would more closely represent the local environment that we find ourselves in. We have not done this, which means that the results of this work are the zero-knowledge results with no assumptions made about our position in the cosmological density field. It would be enlightening to investigate what the effects of assuming the observer to be in an overdense or underdense region would be on the measured bulk flow for a particular geometry, but our current knowledge of the local density field is too uncertain to be able to do this with great confidence.

The HR2 subsets consist of approximately $3.1 \cdot 10^6$ dark matter haloes, each with six dimensional phase space information. Unfortunately a mock galaxy survey that fills the entire volume of the simulation doesn't exist, so in our analysis we assume that each DM halo corresponds to one galaxy. The smallest of the DM haloes are of a mass comparable to that of a galaxy, but the largest DM haloes of the HR2 simulation have a mass that would be equivalent to hundreds of galaxies. Effectively we are grouping galaxies in massive clusters into just one datapoint with the same probability of being subsampled as any other galaxy.

It is hard to predict exactly how this simplification effects our calculated cosmic variance, but fortunately a limited number of mock SDSSIII galaxy catalogues have been produced for the HR2 simulation. In chapter 7 we perform an analysis of the bulk flow magnitude distribution of galaxies from one such mock catalogue, and compare the distributions derived from DM halo velocities. Fortunately that analysis show that the distributions are similar, so that our use of the HR2 simulation in this thesis is justified.

6 Sampling, Completeness & Cosmic Variance

In this chapter the method used to test the effects of a particular survey geometry on the distribution of bulk flow magnitudes will be broken into three parts, where each part adds a layer of complexity to the process.

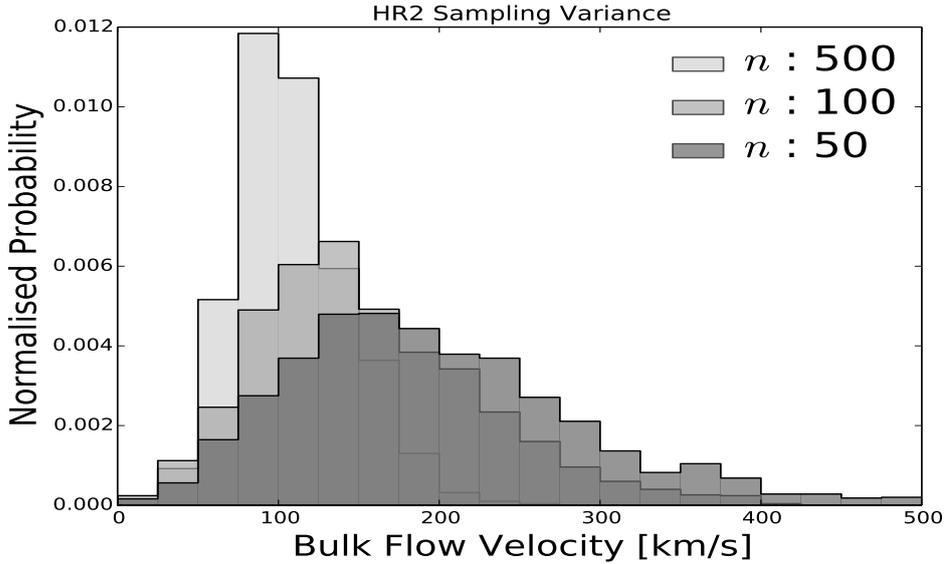


Figure 4: MLE Sampling variance for a 90 degree opening angle spherical cone, varying the number n of galaxies used per bulk flow calculation. The asymmetry in the distribution of bulk flow magnitudes causes the the variance as well as the mean of the distribution to be a function of n .

The first step is to calculate the sampling variance of a subset of the HR2 simulation. The sampling variance is a measure of how well the observed volume is being sampled. If all galaxies in a volume have their peculiar velocity measured - which would be full sampling - the sampling variance would be zero. From the N total galaxies in the subset, n galaxies are drawn randomly. For these n galaxies both

the MLE and MV bulk flow is calculated. Another n galaxies are then drawn, and the MLE and MV bulk flows are calculated again. This process of drawing new subsamples of n galaxies and calculating the MLE and MV bulk flows is done until the sampling variance converges. The sampling variance is the variance⁵ in the distribution of these calculated bulk flows, for the MLE and MV methods respectively. This is illustrated in the leftmost part of Fig. 7 where the purple gaussian represents the scatter in the calculated bulk flow values for the particular subset of the HR2 simulation.

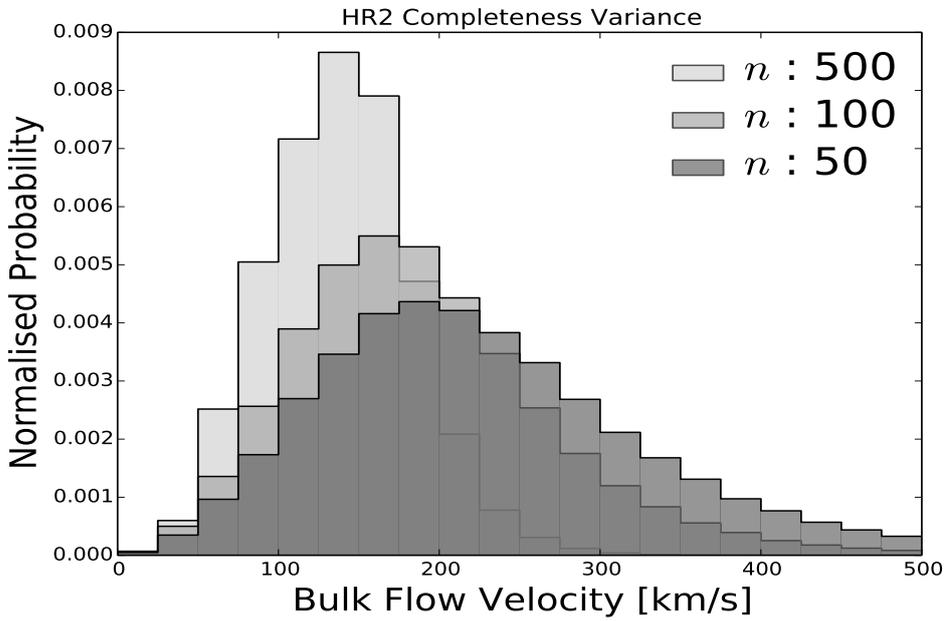


Figure 5: MLE Completeness variance for a 90 degree opening angle spherical cone, varying the number n of galaxies used per bulk flow calculation. The asymmetry in the distribution of bulk flow magnitudes causes the the variance as well as the mean of the distribution to be a function of n .

In Fig. 4 the MLE sampling variance is computed for a spherical cone geometry with an opening angle of 90 degrees. As explained above the sampling variance is simply the variance in the distribution of bulk flows calculated from subsamples of n galaxies, all taken from the same sample of N galaxies. We would expect that the sampling variance would decrease when we increase n , with the extreme being the fully sampled situation where $n = N$ and the sampling variance would be zero. In Fig. 4 we do indeed see that the sampling variance decreases as a function of

⁵Variance is defined as $\text{Var}(x) = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$ where μ is the mean of the distribution.

n , but furthermore we see that the mean of the bulk flow distribution is also a function of n , decreasing when n is increased. This is a result of the possible values of bulk flow magnitudes being restricted to only positive values. The distribution of possible bulk flows is not a symmetric distribution centred on zero, but instead an asymmetrical distribution restricted to positive values. An example that illustrates this well is an observer embedded in a completely isotropic and spherically symmetric bulk flow constructed so that the sum of all the peculiar velocities is exactly zero. If we incompletely sampled this bulk flow, the peculiar velocities in our sample wouldn't add up to zero, and we would estimate some non-zero bulk flow magnitude. This would be true no matter how many times we would perform this incomplete sampling - each time we would come up with some non-zero bulk flow magnitude so that in the end we would converge towards a distribution with some non-zero mean and variance. The less complete our sampling would be, the larger this variance would be, so that we more often than not would find ourselves estimating a fairly large bulk flow magnitude. This in turn effects the mean of the distribution; the variance and the mean of the bulk flow magnitude distribution are related in much the same way that the variance and mean of the Poisson distribution are.

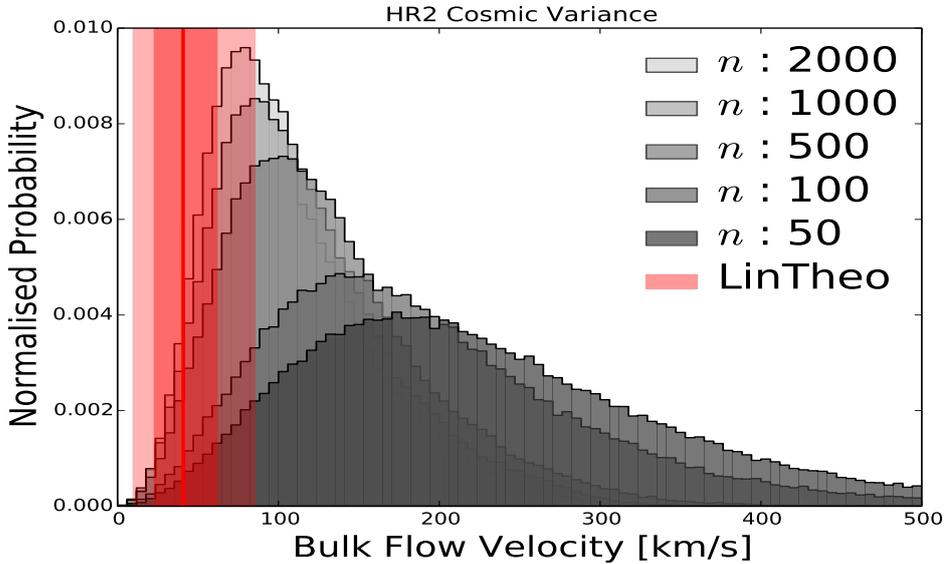


Figure 6: MLE Cosmic variance for a 90 degree opening angle spherical cone, varying the number n of galaxies used per bulk flow calculation. The asymmetry in the distribution of bulk flow magnitudes causes the the variance as well as the mean of the distribution to be a function of n . The red vertical line and red shaded areas are the maximum likelihood bulk flow $V_{ML}(R)$ of Eq. 5 with associated 1 and 2 σ uncertainties.

The second step is to calculate the completeness variance. The completeness variance builds upon the method of sampling variance, but adds rotation. For a particular geometry a subset of n galaxies is drawn, and the MLE and MV bulk flows are calculated. The geometry is then rotated about origin, creating a new subset of HR2. From this new subset we again draw n galaxies and calculate the MLE and MV bulk flow. This process of rotating the geometry, drawing galaxies and calculating the associated MLE and MV bulk flows is continued until the completeness variance converges. The completeness variance is the variance on the total distribution of calculated bulk flows for all rotations of the geometry. This process is illustrated in the centre part of Fig. 7. For fully spherically symmetric geometries the completeness variance is equal to the sampling variance, since rotation has no effect on the geometry. Just like the sampling variance the completeness variance is also a function of the number of galaxies, n , used to calculate the bulk flow magnitude. This is illustrated in Fig. 5 where the completeness variance has been computed for the same sample of galaxies as was used when calculating the sampling variance in Fig. 4. From the two figures it is obvious that the completeness variance is larger than the sample variance, which will typically - although not always - be the case.

Both the sampling and the completeness variance are specific for the chosen origin. The sampling variance is specific both to the chosen origin, and the chosen orientation of the survey geometry, where the completeness variance is specific only to the chosen origin as the orientation dependence is eliminated through rotation of the geometry around the origin. The final step in the process is to eliminate the dependence on the chosen origin; this is done when computing the cosmic variance.

The cosmic variance builds upon the completeness variance and adds translation. When each subset of n galaxies is drawn, the geometry is then rotated and translated randomly to a new origin in the full HR2 simulation. This is illustrated in the right part of Fig. 7. The cosmic variance is a measurement of the variance in the cosmological velocity field. It will depend on the geometry and scale chosen for the measurement, decreasing with larger scales where the cosmological principle becomes a better approximation. Since when deriving the cosmic variance we are restricted to calculating bulk flow magnitudes that are exclusively positive, the cosmic variance is also a function of the number of galaxies in each subsample, n , just like the sampling and completeness variance. This is illustrated in Fig. 6. In the figure we also have the linear theory prediction, calculated with Eq. 2, shown as the red vertical line, with the red shading being the one and two sigma uncertainties. The theoretical prediction does not exactly agree with the peak of the bulk flow magnitude distribution. Increasing the number of galaxies used to

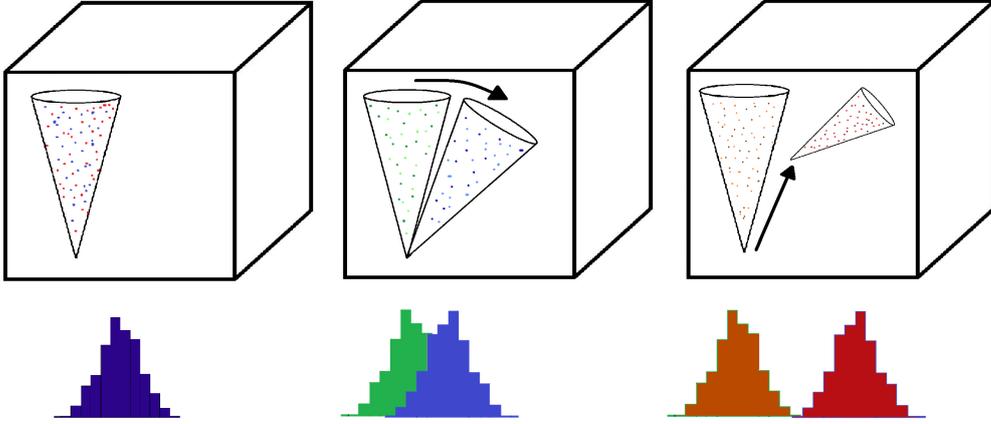


Figure 7: Left/centre/right parts illustrates the method used to derive sampling/completeness/cosmic variance. To derive sampling variance a number of different subsets are drawn from the same distribution, each subsample is then used to calculate a bulk flow and the standard deviation of the calculated bulk flows give the sampling variance. The completeness variance is similar to sampling variance in that it draws subsamples from a distribution, and calculates a bulk flow for each of these subsamples. Additionally it then adds rotation so that a number of subsamples are drawn with the observer position kept constant, but the orientation of distribution to draw samples from rotates around the observer position. The completeness variance is then derived by calculating the bulk flow for all of the drawn subsamples and calculating the standard deviation of these bulk flows. The cosmic variance is similar to the completeness variance in that it draws subsamples from different distributions through rotation, but it differs in that it also adds translation and does not keep the observer position constant.

calculate each bulk flow, n , above the maximum plotted $n = 2000$ was attempted, but did not significantly change this, so the discrepancy between the theoretical prediction and the prediction from the simulation must come from some other source. There are two immediate suspects in explaining this discrepancy. The first is that the theoretical prediction is based on a window function that assumes a perfectly spherical and isotropic survey. What we used was a spherical cone with a 90 degree opening angle which has a window function that would functionally be different from the assumed ideal window function. The second source is that the figures throughout this chapter have been produced by calculating the MLE bulk flow a great many times to derive the underlying distributions. The MLE has a bias towards measuring larger bulk flows than the underlying actual bulk flow, which could also be part of the explanation of this discrepancy. Please note that this bias of the MLE is dependent on the uncertainties on nearby peculiar velocities being smaller than on more distant measurements. In Appendix A we explain how the peculiar velocity uncertainties were derived throughout this thesis.

	SDSSIII Mock	DM Halo
Sampling variance	$n : 50 - (209.7 \pm 98.0)\text{km s}^{-1}$	$n : 50 - (198.2 \pm 91.1)\text{km s}^{-1}$
	$n : 100 - (175.6 \pm 78.3)\text{km s}^{-1}$	$n : 100 - (169.8 \pm 70.7)\text{km s}^{-1}$
	$n : 500 - (131.0 \pm 41.5)\text{km s}^{-1}$	$n : 500 - (137.8 \pm 36.6)\text{km s}^{-1}$
Completeness variance	$n : 50 - (209.1 \pm 96.3)\text{km s}^{-1}$	$n : 50 - (202.6 \pm 92.7)\text{km s}^{-1}$
	$n : 100 - (169.0 \pm 75.5)\text{km s}^{-1}$	$n : 100 - (166.9 \pm 73.4)\text{km s}^{-1}$
	$n : 500 - (114.1 \pm 46.5)\text{km s}^{-1}$	$n : 500 - (123.8 \pm 49.3)\text{km s}^{-1}$

Table 1: Sampling and completeness variance for SDSSIII mock survey galaxy catalogue and DM halo slice of the full HR2 simulation, for varying number of galaxies per bulk flow calculation, n . The numbers should be compared across horizontally. All the numbers are within 0.1σ of each other, which shows that using DM Halos gives comparable results to using a mock galaxy catalogue.

7 Mock Galaxy Surveys versus Dark Matter Halos

As explained in chapter 5 a number of mock SDSSIII galaxy catalogues have been produced from the HR2 cosmological DM halo simulation. One of these mock SDSSIII catalogues is a sphere with radius $1 \text{ Gpc } h^{-1}$ and origin at $(x, y, z) = (1.8, 1.8, 1.8) \text{ Gpc } h^{-1}$. From the full HR2 DM halo simulation we slice a sphere that also has radius $1 \text{ Gpc } h^{-1}$ and origin at $(x, y, z) = (1.8, 1.8, 1.8) \text{ Gpc } h^{-1}$. The sampling and completeness variance are then calculated for both the SDSSIII mock catalogue and the sliced sphere of DM halos. The sampling variances are shown in Figure 8 and the completeness variances in Figure 9. What we see is that, when we use the same number of galaxies per bulk flow, n , the distributions look very similar. The means and variances of the distributions are shown in Table 1. From Figure 8 and 9, and Table 1 we can see that the distributions of bulk flow magnitudes, as well as their mean values and variances, are in good agreement. This shows that it is indeed possible to use the DM halos of the full HR2 simulation to perform our analysis, including investigating the effects of survey geometry on the measurements of bulk flow magnitudes.

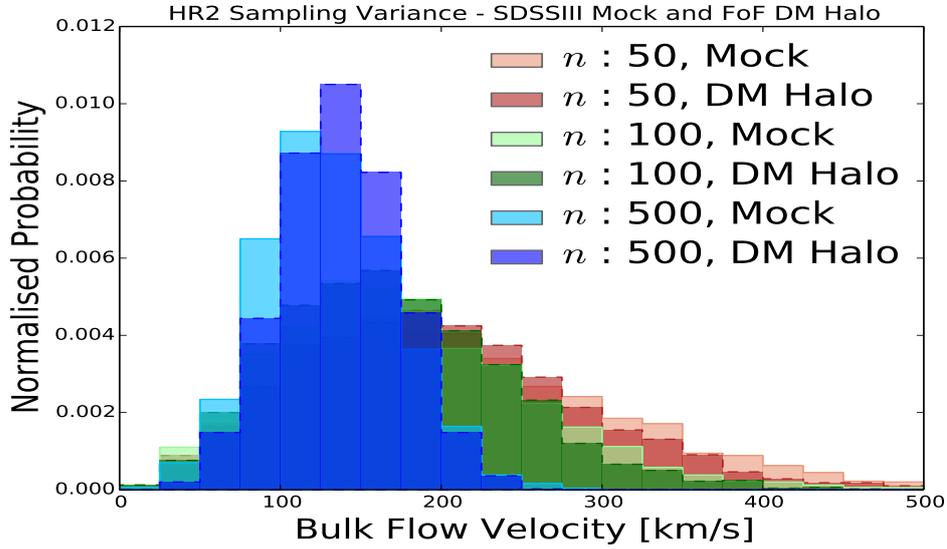


Figure 8: MLE Sampling variances for SDSSIII mock galaxy catalogue subsamples and DM halo subsamples, both taken from the same position in the full HR2 simulation. The bulk flow magnitude variances for the DM halo subsamples are labelled ‘DM Halo’, with the variances for the SDSSIII mock catalogue samples labelled ‘Mock’. The individual pairs of bulk flow magnitude distributions (e.g. $n = 500$, $n = 100$ and $n = 50$) all show similar behaviour in their velocity distributions.

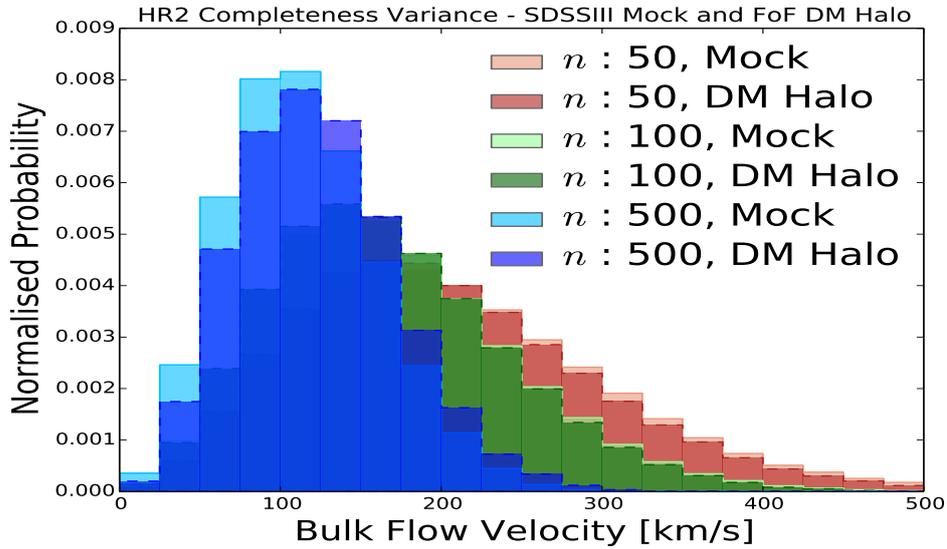


Figure 9: Same as Fig. 8 but for completeness variance.

8 Dependence on survey geometry

With the above discussion of how we derive the cosmic variance for a specific geometry, we are now ready to probe the cosmic variance as a function of survey geometry. The method used to estimate the cosmic variance is the one explained above in chapter 6, and the simulation we draw subsamples from is the Horizon Run 2 simulation, as discussed in chapter 3. Finally, we apply both the Maximum Likelihood (MLE) and Minimum Variance (MV) methods of estimating the bulk flow.

In summary, the process carried out to estimate the cosmic variance for a specific geometry is

1. Choose random point (x, y, z) from HR2 simulation.
2. Place geometry at chosen point (x, y, z) .
3. Rotate geometry to random orientation.
4. Sum the velocities of all the N peculiar velocities that now lie within the geometry to derive actual underlying bulk flow magnitude. Save this value⁶.
5. From the total N peculiar velocities that lie within the geometry draw n peculiar velocities randomly.
6. Calculate MLE/MV bulk flows for the chosen n data points.
7. Rotate the geometry to new random orientation.
8. Compute completeness variance. If converged, store calculated MLE/MV bulk flows. If not converged, return to and continue from step 3.
9. Compute cosmic variance. If converged, end run. If not converged, return to and continue from step 1.

Above step 4 performs a simple sum over all the peculiar velocities that lie within the geometry, and then computes the actual underlying bulk flow magnitude. The mean and variance of the actual underlying bulk flow is therefore also computed as a function of geometry, but should remain fairly constant as we are keeping the volume constant as we vary the opening angle of our geometry. We then use this mean and variance of the actual underlying bulk flow as a benchmark that the MLE and MV bulk flow magnitude estimates should strive to recreate.

The first scenario we will look at is a spherical cone geometry with a constant

⁶Used for the ‘VelocitySum’ grey band in Fig. 10 and 11

volume. The constant volume is set such that when the opening angle is the maximum 360 degrees possible the radial extent is $210 \text{ Mpc } h^{-1}$ ($z \sim 0.07$). The opening angle is then decreased to a minimum of 22.5 degrees where, to keep the volume constant, the radial extent is increased to $1 \text{ Gpc } h^{-1}$ ($z \sim 0.25$). The constructed ideal surveys used in the MV bulk flow estimate have a constant ideal radius, R_I , of $50 \text{ Mpc } h^{-1}$ for all geometries. In Figure 10 we see the results of this analysis. As we expect the mean and variance summed velocities, estimating the actual underlying bulk flow mean and variance, stay approximately constant as we vary the opening angle. The summed velocity mean is shown as the dashed grey line with the grey shaded band being the 1σ variance. The mean of the summed velocities is of order 100 km s^{-1} , which is in decent agreement with the theoretical prediction from equation 5 of $60_{-45}^{+65} \text{ km s}^{-1}$, using an effective radius R of $210 \text{ Mpc } h^{-1}$.

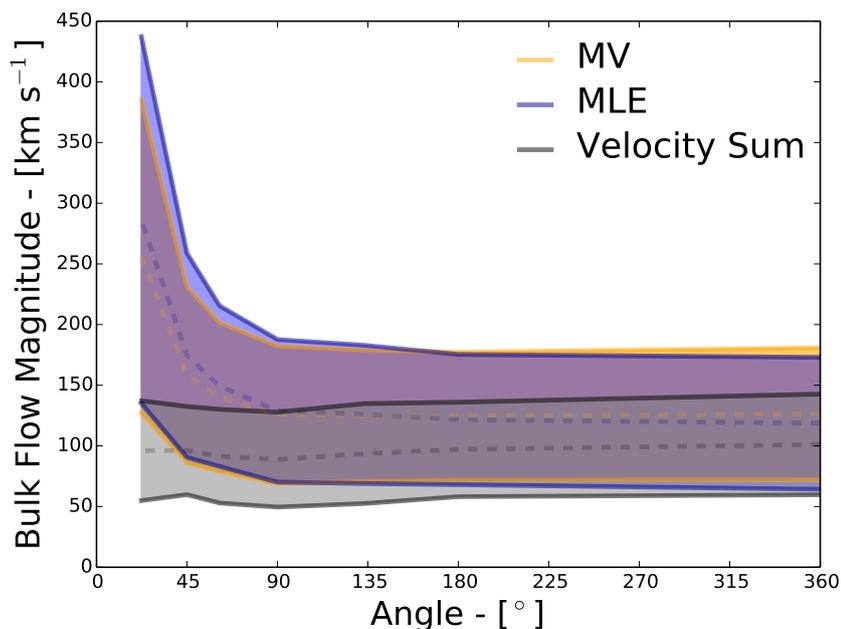


Figure 10: Cosmic variance for spherical cone geometries. The angle is varied from a full sphere of 360 degrees to 22.5 degrees. For more narrow geometries the radial extent of the spherical cone diverges quickly towards infinity, since the volume is kept constant. The Minimum Variance bulk flow is calculated with an ideal radius of $50 \text{ Mpc } h^{-1}$.

The mean value of the MLE is shown as the blue dashed line, with the 1σ variance shown as the blue shaded band. Similarly the mean value of the MV estimate is

shown as the dashed orange line, with the 1σ variance shown as the orange shaded area. For geometries close to the fully spherical geometry with an opening angle of 360 degrees both the MLE and MV estimate agree fairly well with the summed velocity, but they do both on average overestimate the bulk flow magnitude by approximately 20 km s^{-1} . Here it is important to remember from chapter 6 that the mean of the bulk flow magnitude distribution is a function of the number of galaxies used in the bulk flow estimate, n . In this analysis we used $n = 300$, which maybe the cause of this overestimation. Unfortunately due to computing restrictions it was not possible to confirm if this is indeed the case.

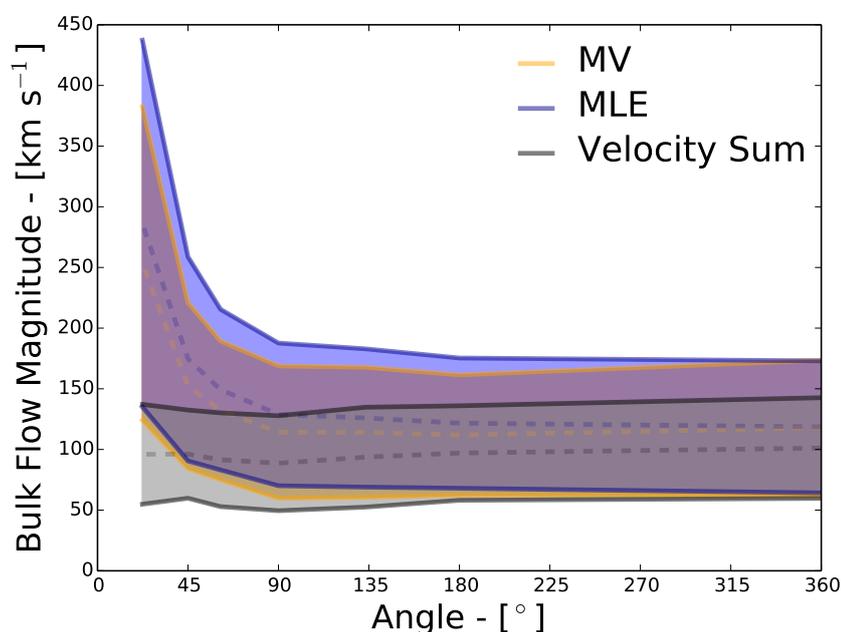


Figure 11: Cosmic variance for spherical cone geometries. The angle is varied from a full sphere of 360 degrees to 22.5 degrees. For more narrow geometries the radius of the spherical cone diverges quickly towards infinity, since the volume is kept constant. The Minimum Variance bulk flow is calculated with an ideal radius equal to half of the radial extent, increasing as opening angle decreases.

This slight overestimation of the bulk flow is a constant trend until the opening angle goes below approximately 180 degrees. As we move into the regime where the opening angle is below 180 degrees both the MLE and MV methods increasingly overestimate the mean bulk flow magnitude. This trend continues all the way to the minimum opening angle of 22.5 degrees. If the opening angle were to be decreased below the minimum 22.5 degrees of this analysis, the radius would

diverge rapidly, and become too large to fit into the HR2 simulation cube with side length $7.2 \text{ Gpc } h^{-1}$. Therefore we can only speculate about what would happen if the opening angle was further decreased below 22.5 degrees, but the trend indicates that more narrow opening angles leads to a larger overestimates on average of the underlying bulk flow. For the MLE method we could have suspected this, since we know from chapter 4 that the MLE method weights the peculiar velocity measurements by their measurement uncertainty, which increases with distance. This means that for very narrow geometries, where most of the peculiar velocity measurements are at large distances, those few nearby points will be up-weighted heavily. From chapter 2 on linear theory we additionally know that on smaller scales we expect a larger bulk flow measurement, so these few upweighted nearby points will bias the bulk flow measurements to larger than expected values.

Surprisingly, the MV method doesn't fare much better than the MLE method; the results of MLE and MV methods are virtually indistinguishable in this analysis. The result of the MV method is dependent on the input constructed ideal survey, which in this case had an ideal radius, R_I , of $50 \text{ Mpc } h^{-1}$. It could be that the ideal radius of the constructed ideal survey is simple too small and the MV method therefore suffers from the same upweighting of nearby datapoints that then dominate the bulk flow estimate, and biases it towards higher than expected average bulk flow measurements. To determine whether or not this is the case we repeat the analysis that produced Figure 10, except this time the ideal radius, R_I , of the constructed ideal survey is not kept constant. Rather, it is set to be equal to half of the radial extent, r , of the geometry in question. The precise values used as a function of opening angle can be seen in Table 2.

In Figure 11 we see the results of this analysis where the ideal radius, R_I , is increased with the radial extent, r . The MLE and summed velocities results are included for comparison purposes, and are exactly equal to the results of Figure 10. The only difference between Figure 10 and 11 is a slight shift of the mean MV bulk flow estimate in Figure 11 to lower values, closer to the actual underlying mean and variance. So even if it seems that increasing the ideal radius, R_I , seems to improve things and move the MV estimates closer to the expected values, the shift is so small that the improvement is far from being conclusive. From previous studies (Watkins et al., 2009) we know that the choice of ideal radius may have a significant impact on the output value of the MV bulk flow estimate, but in our analysis neither the mean nor the cosmic variance are shifted significantly when using a different ideal radius.

	Opening Angle - θ	Radial extent - r	Ideal Radius - R_I
1:	360°	210 Mpc h^{-1}	100 Mpc h^{-1}
2:	270°	224 Mpc h^{-1}	110 Mpc h^{-1}
3:	180°	267 Mpc h^{-1}	130 Mpc h^{-1}
4:	135°	316 Mpc h^{-1}	160 Mpc h^{-1}
5:	90°	402 Mpc h^{-1}	200 Mpc h^{-1}
6:	60°	516 Mpc h^{-1}	260 Mpc h^{-1}
7:	45°	631 Mpc h^{-1}	320 Mpc h^{-1}
8:	22.5°	1003 Mpc h^{-1}	500 Mpc h^{-1}

Table 2: The ideal radii, R_I , values used to construct ideal surveys as a function of opening angle θ in the production of Figure 11.

9 Discussion

In the above chapters we have introduced the necessary background to use the Horizon Run 2 (HR2) simulation to test the effects of survey geometry on the mean and cosmic variance of the MLE and MV bulk flow estimates, when applied to points drawn from the spherical cone geometry with varied opening angle but constant volume. We found that when we restrict ourselves to look at bulk flow magnitudes both the variance and the mean are functions of the number of galaxies used in the bulk flow calculation, n . On top of this we found that when we calculate the cosmic variance as a function of opening angle for a constant volume spherical cone geometry both the MLE and MV bulk flow estimates have an increasing bias towards larger than expected bulk flows as the opening angle is decreased; in other words, the bias towards measuring larger than expected bulk flow increases when we move away from fully spherical survey geometries and into more narrow pencil beam-like surveys. These effects make it complicated to use measurements of the bulk flow magnitude to compare with linear theory as described in chapter 2, or with results of other surveys that may not have the same number of observed peculiar velocities or survey geometry. If one is to perform such a comparison, great care has to be taken to make sure that the asymmetry of the bulk flow velocity magnitude distribution as well as the survey geometry are accounted for.

As an alternative to looking at bulk flow magnitudes, we might instead choose to look at the full covariance matrix that includes the individual bulk flow moments, v_x , v_y and v_z . These bulk flow moments are all symmetric distributions around 0, which means that only the variance, and emphatically *not* the mean, is a function of the number of peculiar velocities, n , used in the bulk flow estimate. If we wanted to use the full covariance matrix to compare with theory we could perform a χ^2 analysis, as is also common in the literature (Feix et al., 2014;

Feldman et al., 2010; Watkins et al., 2009). The χ^2 statistic is here defined as

$$\chi^2 = \sum_{i,j} u_i R_{ij}^{-1} u_j \quad (25)$$

where u_i and u_j are components of the bulk flow vector, with i and j being either the x , y or z component respectively. R_{ij} is the full covariance for the bulk flow moments for a specific set of cosmological parameters. This means that the sum contains nine terms, one for each element of the covariance matrix R_{ij} . The important thing to note here is that χ^2 contains no term from the theoretical prediction, since the theoretical prediction for the bulk flow components is zero, independent of scale. The theoretical prediction is however in the form of the covariance matrix. In most analyses the data are compared to a model and the purpose of the covariance matrix is to weight the sum by the uncertainties on each data point. In contrast, here the covariance matrix is constructed from the model, and represents the model prediction. While the mean bulk flow is predicted to be zero, the mean absolute value of the bulk flow will not be zero. In other words, the rms of the bulk flow is what we would expect to see if we find ourselves at a random position in the universe. The covariance matrix encapsulates this prediction and varies with the cosmological model that is being tested.

To use the χ^2 statistic defined in eq. 25 we then need to calculate the bulk flow vector components u_x , u_y and u_z as well as the full covariance matrix R_{ij} for a range of parameters, e.g. Ω_m and H_0 , and compare their χ^2 values. The parameter set that results in the smallest χ^2 value is then the best fit parameter set. For a description of how we can calculate the full covariance matrix R_{ij} please see Appendix C.

Using predictions of linear theory to directly compare bulk flow magnitude measurements with theory, as an alternative to performing a standard χ^2 analysis, is an enticing idea, but as our analysis shows one has to be very careful to take into account the asymmetry of the bulk flow velocity magnitude distribution, as well as survey geometry when performing such an analysis. Before such effects are better understood, it is preferable to rely on the χ^2 analysis of the bulk flow vector components, when comparing bulk flow velocity measurements with theory.

Additionally, future work should redo the analysis of this thesis, but calculate cosmological parameters rather than bulk flow velocity, and see if the same biases caused by geometry and asymmetry exist. Since our study shows observational effects will cause false detections of large bulk flows, it is critical to confirm that the same observational effects do not bias cosmological parameters.

References

- Abate, A., & Feldman, H. A. 2012, MNRAS, 419, 3482
- Carrick, J., Turnbull, S. J., Lavaux, G., & Hudson, M. J. 2015, MNRAS, 450, 317
- Coles, P., & Lucchin, F. 2002, *Cosmology: The Origin and Evolution of Cosmic Structure*, Second Edition
- Colin, J., Mohayaee, R., Sarkar, S., & Shafieloo, A. 2011, MNRAS, 414, 264
- Dai, D.-C., Kinney, W. H., & Stojkovic, D. 2011, JCAP, 4, 15
- Davis, T. M., & Scrimgeour, M. I. 2014, MNRAS, 442, 1117
- Davis, T. M., et al. 2011, ApJ, 741, 67
- Einstein, A. 1917, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, Seite 142-152., 142
- Feix, M., Nusser, A., & Branchini, E. 2014, JCAP, 9, 19
- Feldman, H. A., Watkins, R., & Hudson, M. J. 2010, MNRAS, 407, 2328
- Hong, T., et al. 2014, MNRAS, 445, 402
- Kashlinsky, A., Atrio-Barandela, F., Kocevski, D., & Ebeling, H. 2008, ApJ, 686, L49
- Kim, J., Park, C., Rossi, G., Lee, S. M., & Gott, III, J. R. 2011, *Journal of Korean Astronomical Society*, 44, 217
- Lavaux, G., Afshordi, N., & Hudson, M. J. 2013, MNRAS, 430, 1617
- Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
- Li, M., et al. 2012, ApJ, 761, 151
- Ma, Y.-Z., Gordon, C., & Feldman, H. A. 2011, *Phys. Rev. D*, 83, 103002
- Ma, Y.-Z., & Pan, J. 2014, MNRAS, 437, 1996
- Ma, Y.-Z., & Scott, D. 2013, MNRAS, 428, 2017
- Nusser, A., & Davis, M. 2011, ApJ, 736, 93
- Osborne, S. J., Mak, D. S. Y., Church, S. E., & Pierpaoli, E. 2011, ApJ, 737, 98
- Perlmutter, S., et al. 1999, ApJ, 517, 565

Planck Collaboration et al. 2014, A&A, 561, A97

Riess, A. G., et al. 1998, AJ, 116, 1009

Turnbull, S. J., Hudson, M. J., Feldman, H. A., Hicken, M., Kirshner, R. P., &
Watkins, R. 2012, MNRAS, 420, 447

Watkins, R., Feldman, H. A., & Hudson, M. J. 2009, MNRAS, 392, 743

A Estimating Peculiar Velocity Measurement Uncertainty

To estimate the peculiar velocity measurement uncertainty, σ_v , as a function of redshift we follow the approach of Davis et al. (2011). Using the terminology of Davis et al. (2011) the measurement uncertainty is

$$\sigma_v = c \cdot \sigma_z = c \cdot \sigma_\mu \cdot \frac{\ln(10)}{5} \frac{\bar{z}(1 + \bar{z}/2)}{1 + \bar{z}} \quad (26)$$

where c is the speed of light in vacuum, \bar{z} is the recession redshift and σ_μ is the uncertainty on the distance modulus measurement. To obtain an estimate for the peculiar velocity measurement uncertainty one has to assume a value for σ_μ , we have chosen to set $\sigma_\mu = 0.1$ throughout this thesis, as it is the value of σ_μ that modern type Ia SNe surveys can achieve, although it is a bit lower than what was possible for legacy surveys where a value of $\sigma_\mu = 0.15$ would be more appropriate. Below is a plot of σ_v as a function of recession redshift \bar{z} , assuming different values of σ_μ .

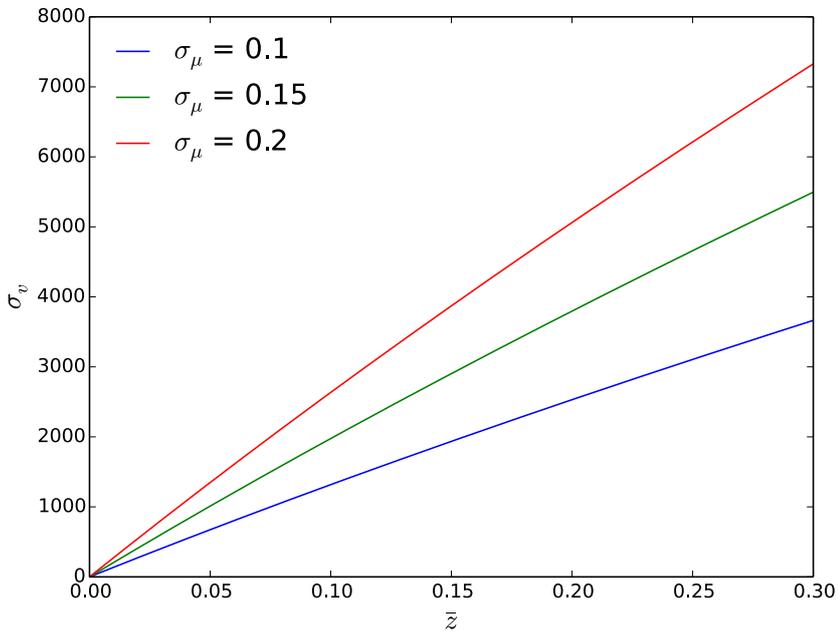


Figure 12: Peculiar velocity measurement uncertainty σ_v as a function of recession redshift \bar{z} for various values of σ_μ , following the approach of Davis et al. (2011).

B Computational Considerations

To perform the analysis of this thesis many thousands of lines of code have been written. All of the code as well as instructions on how to use the code is available for public download at <https://github.com/Arkioss/MV-MLE-BulkFlow>. In this appendix we will cover some of the broad term considerations that have been made in this analysis with concerns to the languages used, and the structure of the programs.

We wanted a code that was easy to understand and modify, but still flexible so that we could modify it in case we came across something interesting that we wanted to explore that wasn't part of the original plan. Therefore we chose to program as much as possible in Python, since it is a high level language with lots of useful functions and modules. All of the plotting has been performed in Python using the matplotlib package, and all the programs that deal with the initial data handling and sorting have also been written in Python.

When efficiency was a concern we instead used C++, as it offers excellent management of typing and memory handling which is either impossible or just very complicated to achieve in Python. Since we would have to calculate both MLE and MV bulk flow estimates many thousand times, dedicated code was written in C++ for both the MLE and MV bulk flow estimates respectively. The code used for the MV bulk flow estimates is based off of Dr. Morag Scrimgeour's code which is available for download at <https://github.com/mscrim/MVBulkFlow>.

Whenever possible we apply both the MV and MLE methods, but since MV estimates take orders of magnitude longer to calculate than MLE estimates we need to prioritise where to spend the available CPU time. Since the MV method attempts to address some of the geometry dependence issues that exist with the MLE, when computing the cosmic variance as a function of opening angle for the spherical cone geometry we did that using both the MLE and the MV method. However, when we look at the distribution of bulk flow magnitudes for the sampling, completeness and cosmic variance we applied only the MLE method. This is because the main point, that the distribution of bulk flow magnitudes is a function of the number of galaxies used to estimate the bulk flow, is a valid point for both the MLE and MV method, and although it would have been preferable to have the results for both MLE and MV methods it was simply not possible with the available computing power.

C Bulk Flow Estimate Covariance matrix

This appendix explains how to calculate the full covariance matrix R_{ij} for the bulk flow estimate. The discussion here is based on the equations of chapter 2 and 4, and assumes that the reader is familiar with the terminology introduced in those chapters. The full covariance matrix R_{ij} is a sum of two terms, the noise term $R_{ij}^{(\epsilon)}$ as well as the geometry term $R_{ij}^{(\nu)}$ such that we have

$$R_{ij} = R_{ij}^{(\epsilon)} + R_{ij}^{(\nu)}. \quad (27)$$

The noise term is given by

$$R_{ij}^{(\epsilon)} = \sum_n w_{i,n} w_{j,n} (\sigma_n^2 + \sigma_\star^2) \quad (28)$$

and the geometry term by

$$R_{ij}^{(\nu)} = \frac{\Omega^{1.1} H_0^2}{2\pi^2} \int_0^\infty dk W_{ij}^2(k) P(k). \quad (29)$$

If we from the full covariance matrix R_{ij} then wish to extract the uncertainties on the individual x , y and z bulk flow moments we simply get the square root of the diagonal terms so that if $i = (x, y, z)$ then

$$\sigma_i = \sqrt{R_{ii}}. \quad (30)$$