Testing the minimum variance method for estimating large-scale velocity moments

Shankar Agarwal,1⋆ Hume A. Feldman1⋆ and Richard Watkins2⋆

1Department of Physics & Astronomy, University of Kansas, Lawrence, KS 66045, USA
2Department of Physics, Willamette University, Salem, OR 97301, USA

Accepted 2012 May 18. Received 2012 May 15; in original form 2011 December 29

ABSTRACT
The estimation and analysis of large-scale bulk flow moments of peculiar velocity surveys is complicated by non-spherical survey geometry, the non-uniform sampling of the matter velocity field by the survey objects and the typically large measurement errors of the measured line-of-sight velocities. Previously, we have developed an optimal ‘minimum variance’ (MV) weighting scheme for using peculiar velocity data to estimate bulk flow moments for idealized, dense and isotropic surveys with Gaussian radial distributions, that avoids many of these complications. These moments are designed to be easy to interpret and are comparable between surveys. In this paper, we test the robustness of our MV estimators using numerical simulations. Using MV weights, we estimate the bulk flow moments for various mock catalogues extracted from the LasDamas and the Horizon Run numerical simulations and compare these estimates to the moments calculated directly from the simulation boxes. We show that the MV estimators are unbiased and negligibly affected by non-linear flows.

Key words: galaxies: kinematics and dynamics – galaxies: statistics – cosmology: observations – cosmology: theory – distance scale – large-scale structure of Universe.

1 INTRODUCTION

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

These difficulties have often been addressed by calculating statistics from peculiar velocity surveys that are designed to primarily reflect large-scale flows which are well described by linear theory. The most common statistic used is the bulk flow, which represents the average motion of the objects in a survey. The bulk flow statistic has been investigated extensively by many groups (Dressler & Faber 1990; Kaiser 1991; Feldman & Watkins 1994; Jaffe & Kaiser 1995; Strauss et al. 1995; Watkins & Feldman 1995; Hudson et al. 1999, 2004; da Costa et al. 2000a; Parnovsky & Tugay 2004; Sarkar, Feldman & Watkins 2007; Kashlinsky et al. 2008, 2010; Ma, Gordon & Feldman 2011; Macaulay et al. 2011; Nusser, Branchini & Davis 2011; Nusser & Davis 2011; Abate & Feldman 2012; Turnbull et al. 2012). However, bulk flow estimates can be difficult to interpret since how they sample the peculiar velocity field depends strongly on the characteristics of the particular survey being considered. In addition, results from bulk flow analyses have often been controversial, highlighting the importance of developing a robust bulk flow statistic that is easy to interpret and that can be compared between surveys with different geometries.

In Watkins, Feldman & Hudson (2009, hereafter Paper I) and Feldman, Watkins & Hudson (2010, hereafter Paper II), we developed the ‘minimum variance’ (MV) moments that were designed to estimate the bulk flow of a volume of a given scale rather than a particular peculiar velocity survey. We stress that the MV moments do not represent the bulk motion of the galaxies in a survey, rather they are estimates of the bulk motion of a given volume of space. The MV algorithm was designed to make a clean estimate of the large-scale bulk flow as a function of scale using the available peculiar velocity data. Essentially, each velocity datum in a real survey is weighted in a way that minimizes the variance of the difference

⋆E-mail: sagarwal@ku.edu (SA); feldman@ku.edu (HAF); rwatkins@willamette.edu (RW)
between the MV-weighted bulk flow of the real survey and an idealized survey bulk flow, on a characteristic scale $R$. The MV analysis suggested bulk flow velocities well in excess of expectations from a cold dark matter (LCDM) model with 7-year Wilkinson Microwave Anisotropy Probe (WMAP7; Larson et al. 2011) central parameters.

Indeed there are a few recent observations that suggest that the standard model may be incomplete: large-scale anomalies found in the maps of temperature anisotropies in the cosmic microwave background (CMB; Copi et al. 2010; Sarkar et al. 2011; Bennett et al. 2011); a recent estimate (Lee & Komatsu 2010) of the occurrence of high-velocity merging systems such as the bullet cluster is unlikely at a $\sim 6\sigma$ level; large excess of power in the statistical clustering of luminous red galaxies (LRG) in the photometric Sloan Digital Sky Survey (SDSS) galaxy sample (Thomas, Abdalla & Lahav 2011); Kovetz, Ben-David & Itzhaki (2010) find a unique direction in the CMB sky determined by anomalous mean temperature ring profiles, also centred about the direction of the flow detected above; larger than expected cross-correlation between samples of galaxies and lensing of the CMB (Hirata et al. 2008; Ho et al. 2008); Type Ia supernovae (SNla) seem to be brighter than expected at high redshift (Kowalski et al. 2008); small voids ($\sim 10$ Mpc) are observed to be much emptier than predicted (Gottlöber et al. 2003); observations indicate denser high concentration cluster haloes than the shallow low concentration and density profile predictions (de Blok 2005; Gentile et al. 2005).

In this paper, we use $N$-body simulations to investigate the robustness of our MV scheme for estimating the bulk flow moments of the velocity field, over a volume of a particular scale, $R$. First we extract a mock catalogue (described in Section 3) from $N$-body simulations. Given this mock catalogue, we use our MV algorithm (described in Section 2) to estimate the bulk flow moments $\{u_x, u_y, u_z\}$ of the velocity field over a volume of a particular scale. Then we position ourselves in the $N$-body simulation box at the location of the centre of the mock catalogue, and calculate the Gaussian-weighted moments $\{V_x, V_y, V_z\}$ by averaging the velocities of all the galaxies in the simulation box; each galaxy being weighted by a Gaussian radial distribution function $f(r) = e^{-r^2/2R^2}$. Note that a large number of particles in the simulation box are preferable to accurately calculate the Gaussian moments of the velocity field. Finally, we compare the MV-weighted moments $\{u_x, u_y, u_z\}$ with the Gaussian-weighted moments $\{V_x, V_y, V_z\}$ in Section 4. A close match between the two would indicate that the MV scheme accurately estimates the Gaussian bulk flow on scale $R$.

It is worth mentioning here the reason for our choice of a Gaussian profile $f(r)$ over, for example, a Tophat filter in developing the MV formalism. A Tophat filter gets contribution from small scales. As such, bulk flow calculated using a Tophat filter can be compared with theoretical expectations only if the observed velocity field is reasonably dense and uniform, so that the small-scale systematics average out. However, observations typically are sparse and non-uniform with large uncertainties. This leads to aliasing of small-scale power onto large scales, making comparison with theory difficult. A Gaussian filter, on the other hand, gets very little contribution from small scales and isolates the small-scale effects present in real surveys, thereby making comparison with theoretical predictions meaningful. Our MV method is specifically designed to convert the observed velocity field into a Gaussian field on a user-specified scale $R$.

In Section 2, we review the MV formalism. In Section 3, we describe the simulations we use and surveys we model to extract the mock catalogues. In Section 4, we compare the MV-weighted bulk flow moments with the Gaussian-weighted moments. We discuss our results and conclude in Section 5.

2 THE MINIMUM VARIANCE METHOD

Individual radial peculiar velocity measurements are plagued by large uncertainties and contributions from small-scale, non-linear processes that are difficult to model theoretically. Both of these problems can be greatly reduced if instead of considering individual velocities an average velocity over a sample, commonly called the bulk flow, is worked with. The three components of the bulk flow $u_i$ can be written as weighted averages of the measured radial peculiar velocities of a survey,

$$u_i = \sum_j w_{i,j} S_j,$$

where $S_j$ is the radial peculiar velocity of the $j$th galaxy and $w_{i,j}$ is the weight assigned to this velocity in the calculation of $u_i$. Throughout this paper, subscripts $i$, $j$, and $k$ run over the three components of the bulk flow, while subscripts $m$ and $n$ run over the galaxies. By far the most common weighting scheme used in studies of the bulk flow, which we will call the maximum likelihood estimate (MLE) method, is obtained from a maximum likelihood analysis introduced by Kaiser (1988). By modelling galaxy motions as being due to a uniform flow and assuming Gaussian-distributed measurement uncertainties, the likelihood function

$$L[u_i|\{S_n, \sigma_n\}] = \prod_n \frac{1}{\sqrt{\sigma_n^2 + \sigma_e^2}} \exp \left( -\frac{1}{2} \frac{(S_n - \hat{r}_n u_i)^2}{\sigma_n^2 + \sigma_e^2} \right)$$

is obtained, where $\hat{r}_n$ is the unit position vector of the $n$th galaxy, $\sigma_n$ is the measurement uncertainty of the $n$th galaxy and $\sigma_e$ is a 1D velocity dispersion accounting for smaller scale motions. Maximizing this likelihood gives a bulk flow estimate of the form of equation (1), with weights

$$w_{i,j} = \sum_{n=1}^3 A_{ij}^{-1} \frac{\hat{r}_n f_{n,j}}{\sigma_n^2 + \sigma_e^2}.$$

where

$$A_{ij} = \sum_n \frac{\hat{r}_n f_{n,j}}{\sigma_n^2 + \sigma_e^2}. $$

These weights play the dual roles of accounting for geometrical factors, e.g. picking out the $x$ component of velocities in the calculation of $u_x$, and down-weighting velocities with large uncertainties. However, the fact that velocity uncertainties are typically proportional to distance, together with the sparseness of velocity catalogues at their outer edges, means that nearby objects are greatly emphasized in calculations of the MLE bulk flow. Indeed, studies of the window functions of these moments (Paper I) have shown that MLE bulk flow moments of a survey are typically sensitive to flows on scales much smaller than the survey’s physical diameter, thus complicating their interpretation.

In Paper I, we introduced an alternative to the MLE weights that yield bulk flow moments that are much easier to interpret. First, we imagine an idealized survey containing radial velocities that are a large number of objects, all with zero measurement uncertainty. In Section 2, we review the MV formalism. In Section 3, we describe the simulations we use and surveys we model to extract the mock catalogues. In Section 4, we compare the MV-weighted bulk flow measurements with the Gaussian-weighted moments. We discuss our results and conclude in Section 5.
small-scale aliasing and which reflect the motion of a well-defined volume. Note that the difference between \( U_i \) and \( V_i \) (see Section 1 for the definition of \( V_i \)) is that \( U_i \) is calculated from an ideal (dense and isotropic) survey, while \( V_i \) is based on the galaxy distribution obtained from \( N \)-body simulations. In the limit that the simulations are dense enough, \( V_i \) will converge towards \( U_i \).

Our goal is to construct estimators for the idealized survey bulk flow components \( U_i \), out of the measured radial peculiar velocities \( S_n \) and positions \( r_n \), contained in a real survey. We assume that \( S_n \) can be expressed as \( S_n = v_n + \delta_n \), where \( v_n \) is the radial component of the linear peculiar velocity field at the location of the object and \( \delta_n \) accounts for the measurement noise as well as any non-linear flow, e.g. infall into a cluster. In order to calculate the weights to use for the bulk flow estimators, we minimize the variance \( \langle (u_i - U_i)^2 \rangle \), where the average is over different realizations of a particular matter power spectrum. Expanding this expression using equation (1) for the bulk flow estimate, we obtain

\[
\langle (u_i - U_i)^2 \rangle = \sum_{m,n} w_{i,m} w_{i,n} \langle S_m S_n \rangle + \langle U_i^2 \rangle
\]

\[
-2 \sum_n w_{i,n} \langle U_i v_n \rangle,
\]

where we have used the fact that the measurement error included in \( S_n \) is uncorrelated with the bulk flow \( U_i \).

Before we minimize this expression with respect to the weights \( w_{i,n} \), we impose the following constraint introduced in Paper II. Suppose that the velocity field were a pure bulk flow, so that \( S_n = \sum U_i g_i(r_n) + \delta_n \), where \( U_i \) are the three bulk moments \( \{ U_z, U_r, U_z \} \); \( g_i(r_n) \) are the direction cosines of the \( i \)-th galaxy \( \{ \hat{r}_{nx}, \hat{r}_{ny}, \hat{r}_{nz} \} \) and \( \delta_n \) is the noise due to measurement error. We ask that the estimators \( u_i \) give the correct amplitude for the flow on average (over different realizations of the universe), namely that \( \langle u_i \rangle = U_i \). Plugging the expression for \( S_n \) into equation (1) gives the constraint that

\[
\sum_n w_{i,n} g_j(r_n) = \delta_{ij},
\]

\( \delta_{ij} \) being the Kronecker delta. This set of three constraints is implemented using Lagrange multipliers, so that we derive the desired weights by taking a derivative of the expression

\[
\sum_{m,n} w_{i,m} w_{i,n} \langle S_m S_n \rangle + \langle U_i^2 \rangle - 2 \sum_n w_{i,n} \langle U_i v_n \rangle
\]

\[+ \sum_{j=1}^{3} \lambda_{ij} \left( \sum_n w_{i,n} g_j(r_n) - \delta_{ij} \right) \]

with respect to \( w_{i,n} \) and setting the resulting expression equal to zero. Solving for the weights then gives

\[
w_{i,n} = \sum_m G_{mn}^{-1} \left[ \langle S_m U_i \rangle - \frac{1}{2} \sum_{j=1}^{3} \lambda_{ij} g_j(r_n) \right],
\]

where \( G \) is the covariance matrix of the individual measured velocities, \( G_{mn} = \langle S_m S_n \rangle \). The Lagrange multipliers can be found by plugging equation (8) into equation (6) and solving for \( \lambda_{ij} \),

\[
\lambda_{ij} = \sum_{k=1}^{3} \frac{1}{M_{ik}} \left( \sum_m G_{mn}^{-1} \langle S_m U_k \rangle g_j(r_n) - \delta_{jk} \right),
\]

where the matrix \( M \) is given by

\[
M_{ij} = \frac{1}{2} \sum_m G_{mn}^{-1} g_j(r_n) g_i(r_m).
\]

In linear theory, the correlation \( \langle S_m U_i \rangle \) and the covariance matrix \( G \) that appear in our expression for \( w_{i,n} \) can be calculated for a given matter power spectrum \( P(k) \) (for details see Paper II):

\[
\langle S_m U_i \rangle = \sum_{n=1}^{N'} w_{i,n}^m \langle S_m v_n \rangle
\]

\[= \sum_{n=1}^{N'} w_{i,n}^m H^2 \Omega^2 \frac{P(k)}{2\pi^2} \int dk \, P(k) f_m(k),
\]

where

\[
w_{i,n}^m = \frac{3}{N'} A_{ij}^{-1} \frac{F^m_j}{N'}
\]

are the weights of an ideal, isotropic survey consisting of \( N' \) exact radial velocities \( v_n \) measured at randomly selected positions \( r_n' \), with

\[
A_{ij} = \frac{\sum_{j=1}^{3} A_{ij}^{-1} \frac{F^m_j}{N'}}{N'}
\]

\[
G_{mn} = H^2 \Omega^2 \frac{P(k)}{2\pi^2} \int dk \, P(k) f_m(k) + \delta_{mn} \left( \sigma^2 + \sigma^2_n \right)
\]

\[= \langle \hat{r}_n \cdot \hat{v}(r_n) \rangle \hat{r}_m \cdot \langle \hat{v}(r_m) \rangle + \delta_{mn} \left( \sigma^2 + \sigma^2_n \right),
\]

where \( f_m(k) \) is the angle-averaged window function:

\[
f_m(k) = \frac{d^3 k}{4\pi} \left( \hat{r}_n \cdot \hat{k} \right) \left( \hat{r}_m \cdot \hat{k} \right)
\]

\[\times \exp(ik \cdot (r_n - r_m)).
\]

Thus, given a peculiar velocity survey and a power spectrum model \( P(k) \) we can calculate the optimum weights \( w_{i,n} \) (see equation 8) for estimating the MV moments (see equation 1). We use the power spectrum model given by Eisenstein & Hu (1998) with WMAP7 (Larson et al. 2011) central parameters. Using the optimum weights \( w_{i,n} \) from equation (8), the angle-averaged tensor window function \( W_i^T(k) \) can be constructed (for details see Paper II) as

\[
W_i^T(k) = \sum_{m,n} w_{i,m} w_{j,n} f_{m,n}(k).
\]

The diagonal elements \( W_i^T(k) \) are the window functions of the bulk flow components \( u_i \). Given a velocity survey, \( W_i^T(k) \) estimated using the MV weights are the closest approximation to the ideal window functions. See Paper I for the MV-estimated window functions of the bulk flow components for a range of surveys.

### 3 Mock Catalogues

#### 3.1 N-body simulations

To check the robustness of our MV formalism, we calculated the bulk flow moments directly from numerical simulations. The N-body simulations we use in our analysis are (i) the Large Suite of Dark Matter Simulations (LasDamas; hereafter LD; McBride et al. 2009; McBride et al. in prep.) and (ii) the Horizon Run (hereafter HR; Kim et al. 2009). These are designed to model the SDSS observations. The LD (HR) simulation parameters are \( \Omega_m = 0.25 \) (0.26), \( \Omega_b = 0.04 \) (0.044), \( \Omega_c = 0.75 \) (0.74), \( h = 0.7 \) (0.72), \( \sigma_8 = 0.8 \) (0.794), \( n_s = 1.0 \) (0.96) and \( L_{\text{Box}} = (6.592) h^{-1} \) Gpc for

1 http://lss.phy.vanderbilt.edu/lasdamas/download.html
the matter, baryonic and cosmological constant normalized densities, the Hubble parameter, the amplitude of matter density fluctuations, the primordial scalar spectral index, the volume fraction of dark matter, and the gravitational force softening length.

**3.2 Catalogues**

We create mocks of three different peculiar velocity surveys from the simulations: (i) the 'DEEP' catalogue includes 103 SNIa (Tonry et al. 2003), 70 Spiral Galaxy Clusters (SC) Tully–Fisher (TF) clusters (Giovanelli et al. 1998; Dale et al. 1999a), 56 Streaming Motions of Abell Clusters (SMAC) Fundamental Plane (FP) clusters (Hudson et al. 1999, 2004), 50 Early-type Far Galaxies (EFAR) FP clusters (Colless et al. 2001) and 15 TF clusters (Willick 1999). The DEEP catalogue consists of 294 data points with a characteristic MLE depth of 50 km s$^{-1}$ Mpc, calculated using $\sum w_{n} \sigma_{n} / \sum w_{n}$ where the MLE weights are $w_{n} = 1/(\sigma_{n}^2 + \sigma_{z}^2)$. In this paper, we assume $\sigma_{z} = 150$ km s$^{-1}$.

**Table 1.** The cosmological parameters and the design specifications of the LD-Carmen and HR simulations.

| Parameter                     | LD-Carmen | HR   |
|-------------------------------|-----------|------|
| Cosmological parameters       |           |      |
| Matter density, $\Omega_m$    | 0.25      | 0.26 |
| Cosmological constant density, $\Omega_{\Lambda}$ | 0.75 | 0.74 |
| Baryon density, $\Omega_b$    | 0.04      | 0.044|
| Hubble parameter, $h$ (100 km s$^{-1}$ Mpc$^{-1}$) | 0.7       | 0.72 |
| Amplitude of matter density fluctuations, $\sigma_8$ | 0.8 | 0.794 |
| Primordial scalar spectral index, $n_s$ | 1.0      | 0.96 |

| Simulation design parameters  |           |      |
| Simulation box size on a side (h$^{-1}$ Mpc) | 1000     | 6592 |
| Number of CDM particles       | 1120$^3$ | 4120$^3$ |
| Initial redshift, $z$         | 49       | 23   |
| Particle mass, $m_p$ (10$^{10} h^{-1}$ M$_{\odot}$) | 4.938 | 29.6 |
| Gravitational force softening length, $f_s$ (h$^{-1}$ kpc) | 53 | 160 |

Figure 1. Top row: DEEP catalogue (left) and its radial distribution (right). Bottom row: DEEP mock catalogue (left) and its radial distribution (right).
estimates the bulk flow of this Gaussian-weighted box, by only using the mock catalogues of the kind shown in Figs 1 and 2 (bottom rows).

Consequently, the mocks in Figs 1 and 2 have a relatively featureless radial distribution and the velocity errors of the survey objects. They must also have the same angular distribution as the real surveys of. We do not impose the additional constraint on the mocks that function about this point as the catalogue we are creating mocks.

3.3 Mock extraction procedure

Once we have identified a random point in the N-body simulation box, we extract a set of galaxies that has the same radial selection function about this point as the catalogue we are creating mocks of. We do not impose the additional constraint on the mocks that they must also have the same angular distribution as the real surveys for two reasons: (i) the N-body simulations are not dense enough to give us mocks that are exactly like the real surveys and (ii) the weights $w_i/N$ of the real surveys typically depend only on the radial distribution and the velocity errors of the survey objects. Consequently, the mocks in Figs 1 and 2 have a relatively featureless angular distribution. To make the mocks more realistic, we also impose a 10° latitude zone-of-avoidance cut.

From the simulations we find the angular position, the true line-of-sight peculiar velocity $v$, and the redshift $cz = d_i + v_i$, for each mock galaxy, where $d_i$ is the true radial distance of the mock galaxy from the random centre we selected, all in km s$^{-1}$. We then perturb the true radial distance $d_i$ of the mock galaxy with a velocity error drawn from a Gaussian distribution of width equal to the corresponding real galaxy’s velocity error, $\sigma_v$. Thus, $d_p = d_i + \delta d_i$, where $d_p$ is the perturbed radial distance of the mock galaxy (in km s$^{-1}$) and $\delta d_i$ is the velocity error. The mock galaxy’s measured line-of-sight peculiar velocity $v_p$ is then assigned to be $v_p = cz - d_p$, where $cz$ is the redshift we found above. The reason for this procedure is that the weight we assign to each galaxy in the mock catalogues will then be similar to the weights of the real catalogues, since these depend on the radial distribution errors of the survey objects.

This procedure of perturbing the distances $d_i$ and then assigning the velocities $v_p$ to the mock galaxies introduces a Malmquist bias. We have checked the effect of the bias by following a slightly different approach to generate the mocks. We used the exact distances $d_i$ and only perturbed the velocities as $v_p = v_i + \delta v$. We found the effect of Malmquist bias on our MV analyses to be negligible.

4 BULK FLOW MOMENTS

For each of the 4100 LD (5000 HR) mocks, we estimated the bulk flow moments $\{u, u, u\}$ using our MV weighting scheme (Section 2). We then compared the results to the Gaussian-weighted bulk moments $\{V_x, V_y, V_z\}$ calculated by going to the same central points for each of the 4100 LD (5000 HR) mock catalogues and averaging the velocities of all the galaxies in the simulation box, each galaxy being weighted by a Gaussian weight of width $R = 50 h^{-1}$ Mpc. Although the results that we show here are for a particular scale of $R = 50 h^{-1}$ Mpc, we have repeated our analysis for other values of $R$ with similar results. It is worth mentioning here that since the position and the velocity of every galaxy in the N-body simulations are known exactly, their respective uncertainties are zero. Here we present our results only from the LD simulations.

The HR simulation shows very similar results.

In Fig. 4, we show the probability distribution for the 4100 MV-weighted bulk flow moments $u_s$ (solid) and the Gaussian-weighted
moments $V_i$ (dashed) within a Gaussian window of radius $R = 50h^{-1}\,\text{Mpc}$ for the LD simulations. As shown in Fig. 4, the distributions for the MV-estimated bulk flow moments (solid histogram) and the Gaussian-weighted moments (dashed histogram) are both Gaussian distributed. This is as expected for large-scale moments and reflects the fact that non-linear motions, which can lead to non-Gaussian tails in the velocity distributions for individual galaxies, have been effectively averaged out. The widths of the distributions match well with the expectations from linear theory,

$$\sigma^2(R) = \frac{H_0^2 \Omega_m^{1/2}}{2\pi} \int dk \, P(k) W^2_k(kR), \quad (15)$$

where $\sigma(R)$ is the rms value of the peculiar velocity field smoothed with a suitable filter with a characteristic scale $R$; $W_k(k, R)$ is the window function (Fourier transform of the filter) and $P(k)$ is the matter power spectrum. A $\Lambda$CDM model with WMAP7 (Larson et al. 2010) central parameters, together with a Gaussian window function $W_k(k, R) = e^{-kR^2/2}$, predicts a 110 km s$^{-1}$ width for $R = 50h^{-1}\,\text{Mpc}$, virtually identical to the ones shown in Fig. 4. In Paper II we estimated that for a $\Lambda$CDM model with WMAP7 central parameters, the chance of getting a $\sim$400 km s$^{-1}$ bulk flow for a survey on scales of $50h^{-1}\,\text{Mpc}$ is $\sim$1 per cent. Examining Fig. 4 confirms that the probability will be similarly small. Indeed, the distribution widths obtained from the set of the mock catalogues are shown in Table 2, columns 7–9. Since each mock catalogue has in principle a slightly different expectation value, the probability distribution is also shown. The fact that the distributions are centred on zero demonstrates that the MV estimators are not biased.

Given a mock catalogue, the theoretical expectation value for the width of the distribution, i.e. $(\langle u_i - U_i \rangle)^{1/2}$, can be calculated in linear theory using equations (5), (11), (12) and (13). To check the robustness of our MV method, this can then be compared with the distribution width $(\langle u_i - V_i \rangle)^{1/2}$ calculated directly from the simulations [i.e. $(u_i - V_i)$] distribution is shown in Fig. 6] using the same cosmological model. The theoretical widths $(\langle u_i - U_i \rangle)^{1/2}$ for the 4100 LD mocks are shown in Table 2, columns 7–9. Since each mock catalogue has a slightly different expectation value, we quote the average and standard deviation of the widths obtained from the set of the mock catalogues. The widths $(\langle u_i - V_i \rangle)^{1/2}$ found in the simulations are shown in Table 2, columns 4–6.

Comparing linear theory predictions [$(\langle u_i - U_i \rangle)^{1/2}$ in Table 2, columns 7–9] with the widths found in the numerical simulations [$(\langle u_i - V_i \rangle)^{1/2}$, columns 4–6], we see that the distribution widths
The much improved performance of MV formalism over the widely used MLE scheme is also evident in Fig. 8, where we show the window functions $W^2_{ii}$ of the bulk flow components, calculated using MV (thick) and MLE (thin) methods. These window functions correspond to the DEEP (left-hand column) and COMPOSITE (right-hand column) real catalogues, for $R = 20 \, h^{-1} \, \text{Mpc}$ (top row) and $R = 50 \, h^{-1} \, \text{Mpc}$ (bottom row). For both DEEP and COMPOSITE mock catalogues, the MV window functions are a reasonable match to the ideal ones. The MLE window functions are not only contaminated by small-scale power, but they are also very different for the $x$-, $y$- and $z$-directions – making it difficult to interpret the MLE bulk flow moments. On the other hand, by directly controlling the survey window functions the MV formalism effectively suppresses the small-scale contribution to the bulk flow. Since it is the small scales that are predominantly plagued by non-linear effects, the MV scheme is able to make a clean estimate (compared to MLE) of the bulk flow components, while keeping the non-linear contamination to a minimum.

In Table 2, columns 1–3, we also show the values of the theoretical widths $\langle (u_i - U_i)^2 \rangle^{1/2}$ from the real catalogues on which the mocks are based. We see that the theoretical widths for the real catalogues (columns 1–3) are somewhat larger than the theoretical widths for the mocks (columns 7–9). This is due primarily to the fact that the objects in the simulated catalogues are less clumped than in the real catalogues, even though they have similar radial distribution functions. This is evident in Figs 1 and 2, where the mock catalogues can be seen as having a relatively featureless spatial distribution. Less clumping and fewer close by galaxies in the simulations lower the MV-weighted bulk flow moments $u_i$, resulting in somewhat lower widths $\langle (u_i - U_i)^2 \rangle^{1/2}$ than the real catalogue widths. The creation of mock catalogues with widths that more closely matched the real catalogue widths would require simulations with higher resolution.

We also found that the sparser the mock catalogue is (e.g. DEEP), the higher the chances of getting large velocities (see the extended tails in the velocity distributions for the DEEP mocks in Fig. 4), but in a way that is consistent with the larger uncertainties found in the simulations are somewhat different than the widths predicted by linear theory.

The differences between the MV- and Gaussian-weighted moments for the $x$- and $z$-directions are shown in Table 2, columns 4–6. As we do not show the $y$-direction since it is statistically identical to the $x$-direction. The SFI++ catalogue shows very similar trends and so was not displayed.

Figure 6. Histograms showing the normalized probability distribution for the differences between the MV- and Gaussian-weighted moments for the $x$- and $z$-directions in the top and bottom rows, respectively. The solid histograms show the quantities $\langle (u_i - V_i) \rangle$ for the 4100 mock catalogues extracted from the 41 LD simulation boxes for $R = 50 \, h^{-1} \, \text{Mpc}$: DEEP (left-hand column) and COMPOSITE (right-hand column). Superimposed on the histograms are Gaussians centred at zero and with the same width, $\langle (u_i - V_i) \rangle^{1/2}$, as the corresponding histogram. The fact that the distributions are centred on zero demonstrates that the MV estimators are not biased. We do not show the $y$-direction since it is statistically identical to the $x$-direction. The SFI++ catalogue shows very similar trends and so was not displayed.

Table 2. The theoretical distribution width $\langle (u_i - U_i)^2 \rangle^{1/2}$ for the real catalogues in the first (x), second (y) and third (z) columns, calculated in linear theory using equations (5), (11), (12) and (13). In the fourth (x), fifth (y) and sixth (z) columns, we show the widths $\langle (u_i - V_i) \rangle^{1/2}$ of the $u_i - V_i$ histograms for the LD mocks (see Fig. 6), this should be compared to the first three columns. The theoretical widths for the LD mocks are shown in the seventh (x), eighth (y) and ninth (z) columns. For the LD mocks, we quote the mean and standard deviation values of $\langle (u_i - U_i)^2 \rangle^{1/2}$, for the 4100 mock catalogues. These values are based on WMAP7 (Larson et al. 2010) central power spectrum parameters. In the last column, we show the width of the distribution of the moments $u_i$ over the 4100 mock catalogues (see Fig. 4). Since the widths $u_x$, $u_y$ and $u_z$ were all found to be very similar, we only quote a single value for $u_i$ in the last column. All values are in km s$^{-1}$.
Figure 7. Left: the distribution of galaxies around the location of the centre of a typical mock catalogue. Each galaxy is weighted with a Gaussian radial distribution function $f(r) = e^{-r^2/2R^2}$ (here $R = 50h^{-1}$ Mpc). Right: the window functions $W_{ii}$ (see equation 14) of the bulk flow components $u_i$ for $R = 50h^{-1}$ Mpc. The x, y and z components are dash–dotted, short-dashed and long-dashed lines, respectively, and correspond to the distribution in the left-hand panel. The solid line is the ideal window function (since the ideal survey is isotropic, all components are the same).

Figure 8. The window functions $W_{ii}$ of the bulk flow components calculated using MV (thick) weights (see equation 8) and MLE (thin) weights (see equation 3) for $R = 20h^{-1}$ Mpc (top row) and $R = 50h^{-1}$ Mpc (bottom row) for the DEEP (left-hand column) and COMPOSITE (right-hand column) real catalogues. The x, y and z components are dash–dotted, short-dashed and long-dashed lines, respectively. The solid line is the ideal window function.

associated with the estimators derived from these mock catalogues.

This can be seen by comparing the predicted distribution widths $\langle (u_i - U_i)^2 \rangle^{1/2}$ for the DEEP and COMPOSITE mock catalogues in Table 2, columns 7–9. The DEEP mocks, being sparser compared to the COMPOSITE mocks, have larger widths. Comparing the widths of $(u_i - V_i)$ histograms (Table 2, columns 4–6) found in the simulations (Fig. 6), we again see that the DEEP mocks have marginally larger uncertainties in the bulk estimators, as expected.

5 DISCUSSION AND CONCLUSIONS

In previous papers (Papers I and II), we developed a weighting scheme for analysing peculiar velocity surveys that gives estimators of idealized bulk flow moments that reflect the flow of a volume of a particular scale centred on our location rather than the characteristics of a particular survey. Given a peculiar velocity survey, the MV method is capable of ‘redesigning’ the survey window function in a way that minimizes the aliasing of small-scale power on to large scales, thereby making comparisons with linear theory as well as among independent surveys possible. The direct control over a survey window function makes the MV formalism an extremely useful tool when comparing bulk flow results across independent surveys with varying characteristics.

Using mock catalogues drawn from numerical simulations, we have demonstrated that the MV formalism, within errors, recovers the bulk flow moments of the underlying matter distribution and that the MV moments are unbiased estimators of the bulk flow of a volume of a given scale, regardless of the geometry of a particular survey. The MV moments are unbiased, in that on average they give the correct values for the idealized bulk flow components.

We calculated the variance of the bulk estimator using (i) linear theory $\langle (u_i - U_i)^2 \rangle^{1/2}$ and (ii) numerical simulations $\langle (u_i - V_i)^2 \rangle^{1/2}$. Although the variance calculated using the simulations was found to be somewhat different from the linear theory predictions, we argued that this is due to the simulations being underdense and thus not having enough galaxies. For numerical simulations with higher resolution (more galaxies), we expect the Gaussian-weighted moments $V_i$ to approach the ideal moments $U_i$ and give a much closer match. We found the variance estimates using simulations and linear theory to be significantly closer to each other for the DEEP catalogue, which has fewer close by galaxies and thus performed much better than the SFI++ and COMPOSITE catalogues when testing the MV formalism. These results validate our use of linear theory in the development of the MV method and confirms the fact that non-linear, small-scale motions do not significantly affect the MV estimators.
We tested many facets of the MV formalism and found agreement in all the tests we performed using the LD and HR simulations. We found that the chance of getting large flows ($\sim$400 km s$^{-1}$) in a ΛCDM universe is of the order of $\sim$1 per cent. The bulk moments $u_i$ estimated using our MV formalism are, within errors, the same as the moments $V_i$ of the volume as traced by all the galaxies in the simulation box and linear theory correctly predicts the variance of the estimators. Further, since the formalism allows for exploration of all scales where there are data, we can reliably explore flows on many scales and track the dynamics of volumes of different scales (parametrized by a radius of a Gaussian sphere $R$).

ACKNOWLEDGMENTS

We would like to thank Mike Hudson for many thoughtful and useful comments. We are also grateful to Róman Scoccimarro and the LasDamas collaboration and Changbom Park and the Horizon Run collaboration for providing us with the simulations. This work was supported by the National Science Foundation through TeraGrid resources provided by the NCSA.

REFERENCES

Abate A., Feldman H. A., 2012, MNRAS, 419, 3482
Bennett C. L. et al., 2011, ApJS, 192, 17
Bernardi M., Alonso M. V., da Costa L. N., Willmer C. N. A., Wegner G., Pellegrini P. S., Rite C., Maia M. A. G., 2002, AJ, 123, 2990
Colless M., Saglia R. P., Burstein D., Davies R. L., McMahon R. K., Wegner G., 2001, MNRAS, 321, 277
Copi C. J., Huterer D., Schwartz D. J., Starkman G. D., 2010, Adv. Astron., 2010, 78
da Costa L. N., Bernardi M., Alonso M. V., Wegner G., Willmer C. N. A., Pellegrini P. S., Maia M. A. G., Zaronbli S., 2000a, ApJL, 537, L81
da Costa L. N., Bernardi M., Alonso M. V., Wegner G., Willmer C. N. A., Pellegrini P. S., Rite C., Maia M. A. G., 2000b, AJ, 120, 95
Dale D. A., Giovanelli R., Haynes M. P., Campuzano L. E., Hardy E., 1999, AJ, 118, 1489
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
de Blok W. J. G., 2005, ApJ, 634, 227
Dressler A., Faber S. M., 1990, ApJ, 354, 13
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Feldman H. A., Watkins R., 1994, ApJ, 430, L17
Feldman H. A., Watkins R., Hudson M. J., 2010, MNRAS, 407, 2328 (Paper II)
Gardner J. P., Connolly A., McBride C., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 69
Gentile G., Burkert A., Salucci P., Klein U., Walter F., 2005, ApJ, 634, L145
Giovanelli R., Haynes M. P., Salzer J. J., Wegner G., da Costa L. N., Freudling W., 1998, AJ, 116, 2632
Gottlöber S., Łokas E. L., Klypin A., Hoffman Y., 2003, MNRAS, 344, 715

Minimum variance velocity moments

Hirata C. M., Ho S., Padmanabhan N., Seljak U., Bahcall N. A., 2008, Phys. Rev. D., 78, 043520
Ho S., Hirata C., Padmanabhan N., Seljak U., Bahcall N., 2008, Phys. Rev. D., 78, 043519
Hudson M. J., Smith R. J., Lucey J. R., Schlegel D. J., Davies R. L., 1999, ApJ, 512, L79
Hudson M. J., Smith R. J., Lucey J. R., Branchini E., 2004, MNRAS, 352, 61
Jaffe A. H., Kaiser N., 1995, ApJ, 455, 26
Kaiser N., 1988, MNRAS, 231, 149
Kaiser N., 1991, ApJ, 366, 388
Kashlinsky A., Atrio-Barandela F., Kocevski D., Ebeling H., 2008, ApJ, 686, L49
Kashlinsky A., Atrio-Barandela F., Ebeling H., Edge A., Kocevski D., 2010, ApJ, 712, L81
Kim J., Park C., Choi Y.-Y., 2008, ApJ, 683, 123
Kim J., Park C., Gott J. R., III, Dubinski J., 2009, ApJ, 701, 1547
Kovetz E. D., Ben-David A., Izhaki N., 2010, ApJ, 724, 374
Kowalski M. et al., 2008, ApJ, 686, 749
Larson D. et al., 2011, ApJS, 192, 16
Lee J., Komatsu E., 2010, ApJ, 718, 60
Ma Y.-Z., Gordon C., Feldman H. A., 2011, Phys. Rev. D., 83, 103002
Macaulay E., Feldman H., Ferreira P. G., Hudson M. J., Watkins R., 2011, MNRAS, 414, 621
McBride C., Berlind A., Scoccimarro R., Wechsler R., Busha M., Gardner J., van den Bosch F., 2009, BAAS, 41, 425.06
Masters K. L., Springob C. M., Haynes M. P., Giovanelli R., 2006, ApJ, 653, 861
Nusser A., Davis M., 2011, ApJ, 736, 93
Nusser A., Branchini E., Davis M., 2011, ApJ, 735, 77
Parnovsky S. L., Tugay A. V., 2004, Astron. Lett., 30, 357
Sarkar D., Feldman H. A., Watkins R., 2007, MNRAS, 375, 691
Sarkar D., Huterer D., Copi C. J., Starkman G. D., Schwarz D. J., 2011, Astropart. Phys., 34, 591
Springob C. M., Masters K. L., Haynes M. P., Giovanelli R., Marinoni C., 2007, ApJS, 172, 599
Springob C. M., Masters K. L., Haynes M. P., Giovanelli R., Marinoni C., 2009, ApJS, 182, 474
Strauss M. A., Cen R., Ostriker J. P., Lauer T. R., Postman M., 1995, ApJ, 444, 507
Thomas S. A., Abdalla F. B., Lahav O., 2011, Phys. Rev. Lett., 106, 241301
Tonry J. L., Dressler A., Blakeslee J. P., Ajhar E. A., Fletcher A. B., Luppino G. A., Metzger M. R., Moore C. B., 2001, ApJ, 546, 681
Tonry J. L. et al., 2003, ApJ, 594, 1
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., 1995, ApJ, 453, L73
Watkins R., Feldman H. A., Hudson M. J., 2009, MNRAS, 392, 743 (Paper I)
Wegner G. et al., 2003, AJ, 126, 2268
Willick J. A., 1999, ApJ, 522, 647

This paper has been typeset from a TeX/LATEX file prepared by the author.