Cosmic Flow Measurement and Mock Sampling Algorithm of Cosmicflows-4 Tully—Fisher Catalog

Fei Qin\textsuperscript{1,}, David Parkinson\textsuperscript{1,}, Cullan Howlett\textsuperscript{2,} and Khaled Said\textsuperscript{2,}$^{*}$

\textsuperscript{1} Korea Astronomy and Space Science Institute, Yuseong-gu, Daejeon-daero 776, Daejeon 34055, Republic of Korea; feiqin@kasi.re.kr
\textsuperscript{2} School of Mathematics and Physics, The University of Queensland, Brisbane, QLD 4072, Australia

Received 2021 June 10; revised 2021 August 7; accepted 2021 September 6; published 2021 November 19

Abstract

Measurements of cosmic flows enable us to test whether cosmological models can accurately describe the evolution of the density field in the nearby universe. In this paper, we measure the low-order kinematic moments of the cosmic flow field, namely bulk flow and shear moments, using the Cosmicflows-4 Tully—Fisher catalog (CF4TF). To make accurate cosmological inferences with the CF4TF sample, it is important to make realistic mock catalogs. We present the mock sampling algorithm of CF4TF. These mocks can accurately realize the survey geometry and luminosity selection function, enabling researchers to explore how these systematics affect the measurements. These mocks can also be further used to estimate the covariance matrix and errors of the power spectrum and two-point correlation function in future work. In this paper, we use the mocks to test the cosmic flow estimator and find that the measurements are unbiased. The measured bulk flow in the local universe is $376 \pm 23$ (error) $\pm 183$ (cosmic variance) km $s^{-1}$ at depth $d_{\text{MLE}} = 35$ Mpc $h^{-1}$, to the Galactic direction of $(l, b) = (298^\circ \pm 3^\circ, -6^\circ \pm 3^\circ)$. Both the measured bulk and shear moments are consistent with the concordance $\Lambda$ Cold Dark Matter cosmological model predictions.

Unified Astronomy Thesaurus concepts: Cosmology (343)

1. Introduction

Observations of galaxies indicate that there are fluctuations in the density field of local universe (Jarrett 2004 and references therein). The gravitational effects of the density perturbations exert additional velocity components to the galaxies’ Hubble recessional velocities, called “peculiar velocities.” These peculiar velocities are good probes of the density field, enabling us to constrain cosmological parameters and test cosmological models.

Due to the peculiar motions of galaxies, the apparent (or inferred) distance of a galaxy, $d$, is different from its true comoving distance, $d_c$. This difference is measurable and quantified by the “logarithmic distance ratio” for that galaxy, defined as

$$\eta = \log_{10} \frac{d_c}{d_b}.$$  (1)

The $\eta$ of late-type galaxies can be measured from the Tully—Fisher relation (Tully & Fisher 1977; Strauss & Willick 1995; Masters et al. 2008; Hong et al. 2014), while for early-type galaxies, $\eta$ can be measured from the Fundamental Plane (Djorgovski & Davis 1987; Dressler et al. 1987; Strauss & Willick 1995; Magoulas et al. 2012; Springob et al. 2014; Said et al. 2020). The peculiar velocity of a galaxy can be estimated from its log-distance ratio using velocity estimators (Davis & Scrimgeour 2014; Watkins & Feldman 2015; Adams & Blake 2017).

The cosmic flow field in the local universe arises from the peculiar velocities of galaxies. Measuring the low-order kinematic moments of the cosmic flow field, i.e., bulk flow and shear moments, from peculiar velocity surveys and comparing them to the $\Lambda$ Cold Dark Matter theory predictions enables us to test whether the theory accurately describes the motion of galaxies on large scales. In previous work, the bulk and shear moments are commonly measured using two different methods: maximum likelihood estimation (MLE, Kaiser 1988) and minimum variance estimation (Watkins et al. 2009; Feldman et al. 2010). The measurements generally agree so far with the $\Lambda$ Cold Dark Matter model ($\Lambda$CDM) prediction (Kaiser 1988; Staveley-Smith & Davies 1989; Jaffe & Kaiser 1995; Nusser & Davis 1995; Parnovsky et al. 2001; Nusser & Davis 2011; Turnbull et al. 2012; Ma & Scott 2013; Hong et al. 2014; Ma & Pan 2014; Scrimgeour et al. 2016; Qin et al. 2018, 2019b; Boruah et al. 2020; Qin 2021; Stahl et al. 2021).

In this paper we will measure the bulk flow and shear moments using the peculiar velocity catalog, Cosmicflows-4 Tully—Fisher catalog (CF4TF, Kourkchi et al. 2020b), and compare the measurements to the $\Lambda$CDM predictions to test the model. CF4TF is the currently largest full-sky catalog of Tully—Fisher galaxies, enabling us to more accurately measure the bulk and shear moments; the measurements are less affected by the anisotropic sky coverage, comparing to previous work.

In addition, as CF4TF is one of the key components of the future full Cosmicflows-4 catalog, we present the mock sampling algorithm for CF4TF in this paper. Combining these CF4TF mocks with the 6dFGSv mocks (Qin et al. 2018, 2019a), we can obtain mock catalogs for the two largest subsets of the final full Cosmicflows-4 catalog. Our mocks can model the luminosity selection and survey geometry of the real surveys, enabling researchers to explore how these systematics affect the measurements (Qin et al. 2018, 2019b). In future work, these mocks can be used to study the power spectrum and two-point correlation of velocities of the Cosmicflows-4 catalog. They are the key to estimate the covariance matrix and errors of these measurements. They can also be used to test the methods of these measurements to identify any possible biases. The mock catalogs underlying this article will be shared upon a reasonable request to the corresponding author. In this paper, these mocks are used to test the cosmic flow (bulk and shear moments) estimator to explore how well the estimator recovers the true moments.
The paper is structured as follows. In Section 2, we introduce the CF4TF data. In Section 3, we introduce the L-PICOLA simulation (Howlett et al. 2015a, 2015b) and present the mock sampling algorithm. In Section 4, we introduce the cosmic flow estimator and test the estimator using mocks. We present the final results and discussion in Section 5. A conclusion is presented in Section 6.

We adopt a spatially flat ΛCDM cosmology as the fiducial model. The cosmological parameters are: \( \Omega_m = 0.3121 \), \( \Omega_{\Lambda} = 0.6825 \), and \( H_0 = 100 \, h \, \text{km s}^{-1} \, \text{Mpc}^{-1} \) with \( h = 0.6751 \).

### 2. Data

The CF4TF (Kourkchi et al. 2020b) is a full-sky catalog of 9790 galaxies. The sky coverage of the CF4TF galaxies in Galactic coordinates is shown in Figure 1. The redshift of the CF4TF galaxies reaches 20,000 km s\(^{-1}\); the redshift distribution is shown in Figure 2. The distances of galaxies are measured using the Tully–Fisher relation (Tully & Fisher 1977; Kourkchi et al. 2020b), which is a linear relation between the H\(\text{I} \) rotation widths and photometric magnitudes.

The H\(\text{I} \) data are taken from the following four catalogs. The primary catalog is the All Digital H\(\text{I} \) catalog (ADHI, Courtois et al. 2009); the galaxies with H\(\text{I} \) line width uncertainties \( \lesssim 20 \, \text{km s}^{-1} \) are selected. In the ADHI, the galaxies below decl. \( \delta = -45^\circ \) are observed by the Parkes Telescope (Courtois et al. 2011); the number density of galaxies is lower in this region. Most of the remainder are from the Arecibo Legacy Fast ALFA Survey (ALFALFA, Haynes et al. 2018, 2011), mainly distributed in the decl. range \([0^\circ, 38^\circ]\), and the galaxies with an H\(\text{I} \) spectrum signal-to-noise ratio \( S/N > 10 \) are selected. The rest are from the Springob/Cornell catalog (Springob et al. 2005) and the Pre-Digital H\(\text{I} \) catalog (Fisher & Tully 1981; Huchmeyer & Richter 1989).

The \( u-, g-, r-, i-, \) and \( z- \)band photometry are taken from the Sloan Digital Sky Survey (SDSS) Data Release 12 (DR12, York et al. 2000). The \( w_1 \) and \( w_2 \) band data are taken from the Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010). The \( i- \)band absolute magnitude cut \( M_i \lesssim -17 \, \text{Mag} \) is applied to the data (Kourkchi et al. 2020b, 2020a).

![Figure 1](image_url). The sky coverage of the CF4TF galaxies in Galactic coordinates. The colors of the dots represent the redshift of the galaxies according to the color bar.

### 3. Mocks

#### 3.1. L-PICOLA Simulation

Producing realistic mock surveys that model nonlinear gravitational interactions requires \( N \)-body simulations. However, such full \( N \)-body simulations, which model the gravitational forces and motions down to a fine resolution, can require an extensive HPC infrastructure and a large amount of CPU time. To make a large ensemble of mock surveys quickly and efficiently, we use the COMoving Lagrangian Acceleration (COLA) approach (Tassev et al. 2013) implemented in the L-PICOLA code (Howlett et al. 2015a, 2015b).

We use L-PICOLA to generate 250 dark matter simulations. Though approximate, this method and code have been demonstrated to reproduce the clustering of dark matter particles and halos well on all scales of interest for galaxy and peculiar velocity clustering (Howlett et al. 2015a, 2015b; Blot et al. 2019), at a \( \gtrsim 95\% \) accuracy for both the power spectrum and bispectrum at \( k = 0.3h \, \text{Mpc}^{-1} \). Halos are identified in these simulations using the 3D friends-of-friends algorithm in the VELOCIraptor code (Elahi 2009). This code has also been demonstrated to recover realistic halos, subhalos, and halo merger trees (although we only use the first of these here), in agreement with a variety of other halo finders (Onions et al. 2012; Knebe et al. 2013).

The fiducial cosmology is given by \( \Omega_m = 0.3121 \), \( \Omega_{\Lambda} = 0.0488 \), \( \sigma_8 = 0.815 \), and \( h = 0.6751 \). The redshift of the simulations is at \( z = 0 \). The box size is 1800 Mpc \( \text{h}^{-1} \). The number of particles in each simulation is 2560\(^3\). The minimum halo mass is \( \sim 5 \times 10^{11} M_{\odot} \, \text{h}^{-1} \) (20 particles per halo).

#### 3.2. Mock Sampling Algorithm for CF4TF

We have 250 simulation boxes in total; each simulation provides us with the masses \( m_{\text{halo}} \) of halos and the Cartesian positions \([x, y, z]\) and velocities \([v_x, v_y, v_z]\) of halos. We divide each simulation box into eight identically sized cubes. In each cube, the observer is placed at a galaxy close to the center of that cube. Therefore, we have generated 2000 mock CF4TF catalogs in total.
The halos in the simulations are parent halos; it is required to generate subhalos and galaxies from these parent halos. The starting point is the halo concentration, defined as

\[ c_v = \frac{r_v}{r_s} \]  

(2)

which characterizes the mass distribution in a halo. The term \( r_v \) is the virial radius of a halo, \( r_s \) is the break radius between an inner and outer density profile in a halo, and \( c_v \) is computed purely as a function of parent halo mass \( m_p \) using the fitted relationship from Prada et al. (2012), calibrated on high-resolution N-body simulations. The virial radius is also computed using the parent halo mass and assuming the halo is both spherical and has a density 200× the critical density of the universe. Given \( c_v \) and \( r_v \), one can generate the subhalos and galaxies using the Navarro–Frenk–White (NFW) profile (Navarro et al. 1997).

The subhalo generation algorithm is based on Vale & Ostriker (2004), Conroy et al. (2006), Howlett et al. (2015c, 2017) and Qin et al. (2019a) and is detailed below. First, we assume a power-law distribution for the mass ratio between a subhalo and its parent halo. From this power-law distribution, we can compute the expected number of subhalos as (Springel et al. 2008; Giocoli et al. 2008; Elahi et al. 2018)

\[ \lambda_{\text{POI}} = \int_{f_{\text{min}}}^{1} A f_M^{-c_\text{q}} df_M \]  

(3)

where \( f_M \) is the mass ratio between the subhalo and its parent. Ratio \( f_{\text{min}} \) is the minimum mass ratio we consider, which we set, based on the minimum halo mass of the simulation (see Section 3.1), to 20 times the dark matter particle mass divided by the parent halo mass (i.e., a fixed minimum mass, such that the minimum mass ratio \( f_{\text{min}} \) varies based on the parent halo). The free parameters \( A \) and \( c_\text{q} \) will be fitted by matching the density power spectrum of mocks to real data.

To add realism to the sampling process, the actual number of subhalos in each parent halo \( N_{\text{sub}} \) is generated assuming a Poisson distribution with mean \( \lambda_{\text{POI}} \). The mass ratios of the \( N_{\text{sub}} \) subhalos are generated by drawing uniform random numbers in the interval \([0, 1]\), and inverting the quantile distribution for Equation (3). This results in mass ratios in the interval \([f_{\text{min}}, 1]\). The subhalo masses are then obtained by multiplying the set of \( f_M \) by the mass of the parent halo. As the largest value of \( f_M \) generated for the \( N_{\text{sub}} \) subhalos is 1, the masses of the subhalos will not be larger than their parent halos.

The positions of subhalos are generated from the NFW profile following the arguments in Robotham & Howlett (2018), and rely on the concentration and virial radius for each parent halo computed as described previously. We generate \( N_{\text{sub}} \) random numbers in the interval \( R \in [0, 1] \), then calculate the position parameter (Robotham & Howlett 2018)

\[ p = R \left( \ln(1 + c_v) - \frac{c_v}{1 + c_v} \right) \]  

(4)

and the radius

\[ r = -\frac{r_v}{c_v} \left( 1 + \frac{1}{W_0(-e^{-p-1})} \right) \]  

(5)

which gives the radial position of the subhalos relative to the center of their parent halo. Here, \( W_0 \) is the Lambert Function.\(^3\) \( N_{\text{sub}} \) random numbers in the intervals \([-\pi, \pi]\) and \([0, 2\pi]\) are generated as the polar coordinates of the subhalos, respectively. Then we convert \( r \) into Cartesian coordinates using the polar coordinates. Finally we add them to the Cartesian positions of their parent halo and subtract the observer’s position to obtain the final positions of subhalos.

The velocities of subhalos are calculated from (Navarro et al. 1997; Howlett et al. 2015c)

\[ v = \sqrt{\frac{Gm_v s}{r_v}} \]  

(6)

where \( G \) is the Newtonian Gravitational Constant and \( s \) is the ratio between the velocities of subhalos and the circular velocity of their parent halo, given by (Navarro et al. 1997, Howlett et al. 2015c)

\[ s = \frac{1}{q} \ln\left( 1 + c_v q \right) - \frac{c_v q}{1 + c_v q} \]  

(7)

Then we randomly draw the Cartesian components of the velocities \( v_x, v_y, \) and \( v_z \) for the subhalos from the Gaussian function with the mean equal to the velocity of their parent halo and the standard deviation (std) equal to \( v/\sqrt{3} \) (Howlett et al. 2015c).

The \( i \)-band absolute magnitude is generated using the luminosity function (see Appendix A for more discussion about the \( i \)-band luminosity function). We re-order these magnitudes in descending order, then assign them to the halos (both parent and subhalos) in descending order of mass. We add additional scatter in this one-to-one assignment via a standard deviation parameter \( \sigma_{\text{log} M} \) to account for the expected scatter between halo/subhalo mass and galaxy luminosity. This is used to draw a “proxy” luminosity for the pairwise-matching of halo mass and luminosity using a Gaussian centered on the actual simulated luminosity. Note that this “proxy” is used only for the matching; the simulated luminosity is the one actually stored and used for later applications of the mocks. Then, we obtain mock galaxies. The normalization factor of the luminosity function will be fitted by matching the density power spectrum of mocks to real data.

\(^3\) We use PYTHON function scipy.special.lambertw.
We calculate the sky completeness of the CF4TF survey by splitting into five distinct patches (see Appendix B) and comparing to the 2M++ redshift survey (Lavaux et al. 2010). To do this, the CF4TF and 2M++ are gridded using HEALPix (Górski et al. 2005; Zonca et al. 2019). For each HEALPix pixel, we then calculate the ratio of galaxies in CF4TF and 2M++, treating this as the completeness. However there are two caveats to this. First, 2M++ is not quite uniform; it is 1 K-band magnitude deeper in the Sloan Digital Sky Survey (SDSS) and 6dF regions, which we correct by normalizing the sum of the values in the HEALPix pixels in these regions to be the same. Second, we would not expect every galaxy with a redshift in 2M++ to be capable of providing a Tully–Fisher (TF) distance. However, as we are only interested in measuring the completeness of the CF4TF measurements in one region of the sky relative to other regions, the completeness ratio is then normalized such that the mean completeness of the two patches containing ALFALFA data is one. If the completeness of a pixel is greater than one, it is set to be one. We note that this does not give an absolute measure of completeness (i.e., given all of the galaxies that could have TF measurements in a region of the real universe, how many actually have measurements in CF4TF), but it does provide a relative measure, which is all that is needed for the mocks given the following steps. Importantly, this procedure removes any signatures of large-scale structure from the completeness mask, as these structures should be present in both CF4TF and 2M++. For each mock, the simulated galaxies are subsampled to match the smoothed redshift distribution of CF4TF data in each of the five distinct sky patches separately. Given the relative sky completeness above, this downsampling ensures the mocks have comparable numbers of objects to those we actually have TF measurements for, and that the completeness in both the angular and radial coordinates is representative of the real universe.

In a given redshift bin of the real CF4TF data, we use the Gaussian kernel distribution function (KDE) to smooth the distribution of $\epsilon$ (which denotes the measurement error of log-distance ratio $\eta$) of that redshift bin. Then we calculate the spline cumulative distribution function (CDF) from the Gaussian KDE for $\epsilon$ of that redshift bin. The spline CDF is used to generate measurements errors of log-distance ratios for mock galaxies of that redshift bin. We repeat this step for all of the other redshift bins to obtain the measurements errors of log-distance ratio $\epsilon_{\text{mock}}$ for all mock galaxies. The measured log-distance ratio for each mock galaxy is generated using a Gaussian function centered on the true log-distance ratio $\eta_t$ with standard deviation $\epsilon_{\text{mock}}$. As the measured log-distance ratios are centered on the true values, our mocks do not reproduce possible systematics such as Malmquist bias that may affect the real data. To do so would require instead simulating the observed quantities of the Tully–Fisher relationship, and running these through the same pipeline/fitting procedure as the real data. However, such a procedure was already performed when obtaining the CF4TF data in

$^4$ We use the PYTHON package scipy.stats.Gaussian-kde.

$^5$ The true log-distance ratio of a mock galaxy is given by $\eta = \log_{10} D_{\text{mock}} D_{\text{true}}$, where $D_{\text{mock}}$ is calculated from the true position of the mock galaxy. $D_{\text{true}}$ is calculated from its “observed” redshift $z_{\text{obs}}$ given by $z_{\text{obs}} = (1 + z)(1 + v/c) - 1$, where $z$ is the true redshift calculated from the true position. The term $v/c$ is the true line-of-sight velocity of the mock galaxy calculated from its true position and velocity.

| $A$ | $\alpha_h$ | $\sigma_{\log M}$ | $\phi_*$ |
|-----|------------|-------------------|--------|
| 2.882 | 0.102 | 2.121 | 0.005 |

Kourkchi et al. (2020b), so here we assume the Malmquist bias in the data has been corrected for appropriately, and the mocks only need to reproduce the remaining effects of cosmic variance and measurement errors.

Finally, we can measure the redshift space density power spectra of the 2000 mocks using the method in Howlett (2019) and Qin et al. (2019a). This is shown in Figure 3, which compares the average of these measured power spectra (blue curve) to the density power spectrum of the real CF4TF data (black filled squares) to find the best values for the parameters $A$, $\alpha_h$, $\sigma_{\log M}$ and $\phi_*$, as presented in Table 1. The $\chi^2$/d.o.f. = 39/(38 − 4) indicates that the mocks are in excellent agreement with the data.

Figure 4 shows the comparison between the real CF4TF data and the mocks. The top left-hand-side panel shows the i-band absolute magnitude distribution. The black curve is the average of 2000 mocks, and the shaded areas indicate the 1σ, 2σ, and 3σ regions. Here, $\sigma$ denotes the standard deviation of the magnitude distributions of the 2000 mocks. The dashed blue curve is for the real CF4TF data. The top right-hand-side panel shows the distribution of the log-distance ratio for the data and mock average. The bottom left-hand-side panel shows the distribution of the error of log-distance ratio for the data and mock average. The bottom right-hand-side panel shows the redshift distribution of the data and mock average. In Figure 5, the top panel shows the sky coverage of an example mock in equatorial coordinates. For comparison, the bottom panel shows the sky coverage of the real CF4TF data in equatorial coordinates.

$^6$ In this method, the galaxies are gridded in using their observed redshifts, and then Equation (2.1.3) of Feldman et al. (1994) and Equation (6) of Bianchi et al. (2015), as well as Equation (11) of Yamamoto et al. (2006), are used to estimate the density power spectrum. This means that we are measuring the redshift space rather than real space power spectrum, which would need to be accounted for if we were performing any theoretical modeling, but avoids the impact of the Malmquist bias or other biases that would arise if one were using the true distances to place the galaxies on the grid.
coordinates. The colors of the dots indicate the redshifts based on the color bars.

The mocks are in good agreement with the data. They will be continuously updated based on the updates of the real CF4TF catalog and L-PICOLA simulations.

Figure 4. Comparing the mock averages (black curves) to the real CF4TF data (blue dashed curves). The shaded areas indicate the 1σ, 2σ, and 3σ regions. The top panels are for the distribution of $i$-band absolute magnitudes and log-distance ratios, respectively. The bottom panels are for the distribution of errors of log-distance ratios and the redshifts, respectively.

Figure 5. The top panel shows the sky coverage of an example mock in equatorial coordinates. The bottom panel shows the sky coverage of the CF4TF data in equatorial coordinates. The colors of the dots represent the redshifts of galaxies based on the color bars.

4. Bulk and Shear Moments Measurements

4.1. Bulk and Shear Moments

Using the Taylor series expansion, the line-of-sight peculiar velocity field, $v(d_h)$, can be expanded to the first order (Jaffe & Kaiser 1995; Parnovsky et al. 2001; Feldman & Watkins 2008; Feldman et al. 2010; Qin et al. 2019b)

$$v(d_h) = \sum_{p=1}^{9} U_p g_p(d_h),$$

where $U_p$ are the nine moment components

$$U_p = \{B_x, B_y, B_z, Q_{xx}, Q_{yy}, Q_{zz}, Q_{xy}, Q_{xz}, Q_{yz}\},$$

and where the zeroth-order vector $[B_x, B_y, B_z]$ is the so-called bulk flow velocity. The first-order symmetric tensor $Q_{ij}$, $(i, j = x, y, z)$ is the so-called shear moment. The mode functions are given by

$$g_p(d_h) = \{\hat{r}_x, \hat{r}_y, \hat{r}_z, d_h\hat{r}_x^2, d_h\hat{r}_y^2, d_h\hat{r}_z^2, 2d_h\hat{r}_x\hat{r}_y, 2d_h\hat{r}_x\hat{r}_z, 2d_h\hat{r}_y\hat{r}_z\},$$

where the true comoving distance $d_h$ of a galaxy is given by

$$d_h(z_h) = \frac{c}{H_0} \int_0^{z_h} \frac{dz'}{\sqrt{\Omega_m (1 + z_h)^3 + \Omega_\Lambda}},$$

and where $z_h$ is the Hubble recessional redshift of that galaxy. $H_0$ is the Hubble constant of the present-day universe. The terms $\Omega_m$ and $\Omega_\Lambda$ are the matter and dark energy densities of the present-day universe, respectively.
4.2. Bulk and Shear Moments Estimator

Measuring cosmic flows can provide us with an intuitive understanding of the amplitude and direction of the measured (not the modeled or reconstructed) velocity field. In addition, comparing the measured $U_p$ to the $\Lambda$CDM prediction enables us to test whether $\Lambda$CDM accurately describe the motions of galaxies in the nearby universe. To avoid the non-Gaussianity of the peculiar velocities, we use the so-called $\eta$MLE (Nusser & Davis 1995, 2011; Qin et al. 2018, 2019b) to estimate the nine moment components $U_p$ from the full-sky galaxy catalog CF4TF.

Assuming the log-distance ratio of the $n$th galaxy, $\eta_n$, has Gaussian errors $\epsilon_n$, then the Gaussian likelihood function of a set of $N$ galaxies can be written as (Nusser & Davis 1995, 2011; Qin et al. 2018, 2019b)

$$P(U_p, \epsilon_s) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi(\epsilon_n^2 + \epsilon_{s,n}^2)}} \exp \left( -\frac{1}{2} \frac{\left( \eta_n - \tilde{\eta}_n(U_p) \right)^2}{\epsilon_n^2 + \epsilon_{s,n}^2} \right),$$

where the intrinsic scatter of the log-distance ratio $\epsilon_s$ arises from the nonlinear motions of the galaxies. We set it as a free parameter to fit.

Following the steps of the second paragraph in Section 4.1 of Qin et al. (2019b), the modeled log-distance ratio $\tilde{\eta}_n(U_p)$ of the $n$th galaxy is converted by first substituting Equation (8) into the usual peculiar velocity estimator (Colless et al. 2001; Hui & Greene 2006; Davis & Scrimgeour 2014; Scrimgeour et al. 2016; Qin et al. 2019b).
The Astrophysical Journal, 922:59 (13pp), 2021 November 20

Qin et al. 2018, 2019b)

\[ \chi^2 = \frac{1}{9} \times \frac{1}{1000} - \frac{1}{2} (U_m - U_i)C^{-1}(U_m - U_i)^T, \]

where the measured moments \( U_m \) and the true moments \( U_i \) contain 9000 elements (1000 mocks times 9 moments, respectively). \( C \) is the covariance matrix. Equation (8) gives the nine moments \( U_p \) as the components of the cosmic flow field. Since these moments are all being measured from a single observation location, and with incomplete sky coverage, uncertainty of the measurement of one of these moments will be correlated with those of another, and this is true for all moments. \( C \) is a 9000 × 9000 matrix with 1000 9 × 9 diagonal blocks and zeros elsewhere. The 1000 diagonal blocks are calculated from the MCMC chains. The \( \chi^2_{\text{red}} = 1.806 \). This slightly larger value of \( \chi^2_{\text{red}} \) is due to the intrinsic scatter between the measured moments and true moments, which arises from the intrinsic scatter of the peculiar velocities \( \alpha_n \) (or \( \epsilon_t \), log-distance ratio). Parameter \( \alpha_n \) (or \( \epsilon_t \)) is to encapsulate the nonlinear peculiar motions of galaxies. The peculiar velocity estimator Equation (13)

\[ \chi^2 = \frac{1}{9} \times \frac{1}{1000} - \frac{1}{2} (U_m - U_i)C^{-1}(U_m - U_i)^T, \]

where the measured moments \( U_m \) and the true moments \( U_i \) contain 9000 elements (1000 mocks times 9 moments, respectively). \( C \) is the covariance matrix. Equation (8) gives the nine moments \( U_p \) as the components of the cosmic flow field. Since these moments are all being measured from a single observation location, and with incomplete sky coverage, uncertainty of the measurement of one of these moments will be correlated with those of another, and this is true for all moments. \( C \) is a 9000 × 9000 matrix with 1000 9 × 9 diagonal blocks and zeros elsewhere. The 1000 diagonal blocks are calculated from the MCMC chains. The \( \chi^2_{\text{red}} = 1.806 \). This slightly larger value of \( \chi^2_{\text{red}} \) is due to the intrinsic scatter between the measured moments and true moments, which arises from the intrinsic scatter of the peculiar velocities \( \alpha_n \) (or \( \epsilon_t \), log-distance ratio). Parameter \( \alpha_n \) (or \( \epsilon_t \)) is to encapsulate the nonlinear peculiar motions of galaxies. The peculiar velocity estimator Equation (13)
5. Results and Discussion

5.1. Results and Comparing to $\Lambda$CDM Prediction

Applying $\eta$MLE to the real CF4TF data, we obtain the measured bulk and shear moments in the local universe, as presented in Table 2. The moments are measured in Galactic coordinates. Figure 7 shows the 2D contours and histograms of the MCMC samples of the moments. The shaded areas of the histograms indicate the 1σ errors of the moments. To compare to theory, the characteristic scale of cosmic flow measurement is defined as (Scrimgeour et al. 2016)

$$d_{\text{MLE}} = \frac{\sum |d_{h,\text{MLE}}| W_n}{\sum W_n},$$

where the weight factors $W_n = 1/(\alpha_n^2 + \alpha_n^2)$.

The $\Lambda$CDM predicted moments should have a zero mean and “cosmic root mean square” (CRMS) variation (Feldman et al. 2010). The CRMS is calculated from the following equation (Feldman et al. 2010; Ma et al. 2011; Johnson et al. 2014)
Figure 8. The window functions for CF4TF in Galactic coordinates. The top panels are for the bulk flow components. The middle panels are for the diagonal elements of the shear moments. The bottom panels are for the off-diagonal elements of the shear moments.

Where the matter density power spectrum $P(k)$ is generated using the CAMB package (Lewis et al. 2000). The window function is given by (Feldman et al. 2010; Ma et al. 2011; Johnson et al. 2014):

$$W_{pq}^2(k) = \sum_{m,n} w_p w_q f_{mn}(k),$$

where $f_{mn}(k)$ is given by Ma et al. (2011) and Johnson et al. (2014),

$$f_{mn}(k) = \frac{1}{3} \left( j_0(kD_{mn}) - 2j_2(kD_{mn}) \right) r_n \cdot \hat{r}_m,$$

$$\times \frac{1}{D_{mn}^2} j_2(kD_{mn}) r_n \sin^2(\alpha_{mn}),$$

and where $D_{mn} \equiv |r_m - r_n|$ and $\alpha_{mn} = \cos^{-1}(\hat{r}_m \cdot \hat{r}_n)$. Variable $r_n$ is the position vector of the $n$th galaxy, and $j_0(x)$ and $j_2(x)$ are the zeroth- and second-order spherical Bessel functions of the first kind. The weight factors are given by Equation (16).

Figure 8 shows the $p=q$ components of the window function $W_{pq}^2$ of the CF4TF data in Galactic coordinates. In Table 2, we list the CRMS predicted by $\Lambda$CDM. As shown in the top panel of Figure 8, the amplitudes of window functions for the three bulk flow components are similar; therefore, the co-responding CRMS of the three bulk flow components are similar in Table 2. As shown in the middle and bottom panels of Figure 8, the window functions for the $y$ direction related components $[Q_{yy}, Q_{xy}, Q_{yx}]$ have larger amplitudes, indicating a larger cosmic variance in this direction. This is due to the anisotropic sky coverage caused by the denser region contributed by ALFALFA galaxies. Therefore, in Table 2, the CRMS for the $y$ direction related components, especially $Q_{yy}$, is larger than other shear moment components.

The $\Lambda$CDM predicted moments should have a zero mean. Therefore, the $\chi^2$ between the nine measured moments and the $\Lambda$CDM prediction is given by:

$$\chi^2 = U_p(C_{pq} + R_{pq}^2)^{-1}U_p^T,$$

where $C_{pq}$ is the covariance of the measurement errors and is calculated from the MCMC samples. The $\chi^2 = 10.4$, and the corresponding $p$-value is 0.319, indicating the measurements are consistent with the $\Lambda$CDM prediction.

5.2. Comparing to Other Literature

To visualize the comparison of different bulk flow measurements from other surveys on a single figure, we need to standardize the window function due to the differing geometries and depths of the surveys. In this work, we use the spherical top-hat window function:

$$W(k) = 3(\sin kr - kr \cos kr)/(kr)^3.$$
smallest error bar; they measure the bulk flow by comparing the reconstructed velocity field of 2M++ (Carrick et al. 2015) to the measured velocity field of 465 supernovae. However, as presented in Section 5.1 of Carrick et al. (2015), the reconstructed field has many potential systematic effects that are hard to test. Q19 measures the bulk flow using the Cosmicflows-3 catalog (Tully et al. 2016), which contains 391 Type Ia supernovae and, therefore, has a smaller error bar too. However, the major part of the published Cosmicflows-3 catalog is the 6dFGSv, which does not have a perfect...
Malmquist bias correction (Qin et al. 2018); therefore, this measurement has potential systematic errors.

Figure 10 shows directions of bulk flows measured from different surveys in Galactic coordinates. The S16 measurements is significantly biased from other surveys. The revised result is shown as “Q18-6dFGSv.”

6. Conclusion

In this paper, we present the mock sampling algorithm of CF4TF data. These mocks can realize the luminosity selection, survey geometry, and clustering of the real CF4TF data. A combination of these CF4TF mocks and 6dFGSv mocks (Qin et al. 2018, 2019a) provide us with the mock catalogs for the two largest subsets of the final full Cosmicflows-4 catalog. The mocks can be used to further study cosmology in future work, for example, estimating the covariance matrix and errors of the power spectrum and the two-point correlation of velocities. The mock catalogs underlying this article will be shared upon a reasonable request to the corresponding author.

We use mocks to test the cosmic flow estimator $\eta_{\text{MLE}}$, and we find that the estimator works well to recover the true moments. We measure the bulk and shear moments of the local universe using the CF4TF data. We find that the bulk flow in the local universe is $376 \pm 23$ (error) $\pm 183$ (cosmic variance) km s$^{-1}$ at depth $35$ Mpc $h^{-1}$, to the Galactic direction of $(l, b) = (298^\circ \pm 3^\circ, -6^\circ \pm 3^\circ)$. Both the measured bulk and shear moments are consistent with the $\Lambda$ Cold Dark Matter model prediction.

F.Q. and D.P. are supported by the project “Understanding Dark Universe Using Large Scale Structure of the Universe”, funded by the Ministry of Science. C.H. and K.S. are supported by the Australian Government through the Australian Research Councils Laureate Fellowship funding scheme (project FL180100168).

Facility: The L-PICOLA simulation and CF4TF mock sampling algorithm were performed on the OzSTAR national facility at Swinburne University of Technology. The Cosmicflows-4 Tully–Fisher catalog is downloaded from the Extragalactic Distance Database (EDD) http://edd.ifa.hawaii.edu/.

Software: EMCEE (Foreman-Mackey et al. 2013), CHAIN-CONSUMER (Hinton 2016), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), HEALPIX (Zonca et al. 2019; Górski et al. 2005).

Appendix A

I-band Luminosity Function

In the CF4TF catalog, we have the fully corrected $i$, $g$, $r$, $z$, $w_1$, and $w_2$ apparent magnitudes. However, none of these bands are for all galaxies. To fit the luminosity function of all CF4TF galaxies, we need to work out a single band of magnitudes for all galaxies.

There are 6776 galaxies that have $i$-band apparent magnitudes $m_i$, while 5037 galaxies have $w_1$-band apparent magnitudes $m_{w_1}$. There are 2033 galaxies that have both $w_1$-band and $i$-band apparent magnitudes in the CF4TF catalog. Following the arguments in Kourkchi et al. (2020b), we can first calculate the differences between the magnitudes of the two bands for these common galaxies, $\Delta_{i, w_1} = m_i - m_{w_1}$.

Then, we use the equation shown in Figure 15 of Kourkchi et al. (2019) to convert the $w_2$-band main principal components $P_{1, w_2}$ to the $w_1$-band main principal components $P_{1, w_1}$. Finally, following the arguments in Appendix B4 of Kourkchi et al. (2020b), we use the Random Forest Regressor from the Python package SCIKIT-LEARN to obtain the $w_1$-band apparent magnitudes of all CF4TF galaxies.

Since the $i$-band magnitude cut is the commonly quoted one in the literature (Kourkchi et al. 2019, 2020a, 2020b), we then need to convert $w_1$-band magnitudes to $i$-band magnitudes. The blue dots in Figure 11 shows the relation between the $i$-band apparent magnitudes and $w_1$-band apparent magnitudes of the
6776 galaxies. Then, we use the Random Forest Regressor to obtain the $i$-band apparent magnitudes of the rest of the galaxies, as shown as the yellow dots in Figure 11. Thus, we obtain the $i$-band apparent magnitudes for all CF4TF galaxies and convert the $i$-band apparent magnitudes to absolute magnitudes using the distance modulus data in the CF4TF catalog. The distribution of $i$-band absolute magnitudes $M_i$ is shown in the bottom panel of Figure 12.

To generate mock catalogs for CF4TF, we need to fit the $i$-band luminosity function for the data. We use the Schechter function (Schechter 1976)

$$\phi(M) = 0.4 \ln(10) \phi_* 10^{-0.4(M-M_*)/(a+1)} \exp[-10^{-0.4(M-M_*)}],$$

(A1)

The best-fit values of parameters $M_*$ and $\alpha$ are given in the Table 6 of Kourkchi et al. (2020a). However, the shape of the luminosity function given by their parameters cannot match the measurements; see the green curve in the top panel of Figure 12. In addition, Kourkchi et al. (2020a) do not provide the normalization parameter $\phi_*$ of the luminosity function. Therefore, instead of using the values in Table 6 of Kourkchi et al. (2020a), we fit our own luminosity function. The best-fit values are $M_* = -21.528^{+0.030}_{-0.029}$ and $\alpha = -0.280^{+0.020}_{-0.018}$. The fit result is shown as the blue curve in the top panel of Figure 12.

To emphasize, in this stage we only want to know the shape parameters of the luminosity function; i.e., we only want to know the values of $M_*$ and $\alpha$. The normalization parameter $\phi_*$ is fit in Section 3, as it is difficult to obtain the survey volume of CF4TF due to the complex survey geometry.

### Appendix B

#### The Five CF4TF Patches

In building our mock catalogs, we treat the inhomogeneous CF4TF data as consisting of five distinct patches, each of which has its own unique redshift distribution. Figure 13 shows the five patches of the CF4TF sky coverage we use, which are defined as and correspond to:

1. The blue region: $0^\circ < \text{decl.} < 38^\circ$ and $111^\circ < \text{R.A.} < 250^\circ$. It covers the ALFALFA data in the northern Galactic region.
2. The green region: $0^\circ < \text{decl.} < 38^\circ$ and $0^\circ < \text{R.A.} < 360^\circ$. It covers the ALFALFA data in the southern Galactic region.
3. The red region: $-45^\circ < \text{decl.} < -45^\circ$. Here are relatively uniform galaxies, mainly from ADHI with additions from the Springob/Cornell and Pre-Digital HI catalogs.
4. The orange region: $149^\circ < l < 255^\circ$ and $-5^\circ < b < 5^\circ$. This has sparse data in the Galactic plane from ADHI, the Springob/Cornell, and Pre-Digital HI catalogs, where Galactic extinction reduces the number density of observed TF objects.
5. The black region: $0^\circ < \text{decl.} < 38^\circ$. This includes in the main galaxies from the ADHI but observed using the Parkes Telescope, which results in a lower number density of these galaxies relative to the green region.

**ORCID iDs**

Fei Qin [orcid.org/0000-0001-7950-7864]
David Parkinson [orcid.org/0000-0002-7464-2351]
Khaled Said [orcid.org/0000-0002-1809-6325]

**References**

Adams, C., & Blake, C. 2017, MNRAS, 471, 839
Bianchi, D., Gil-Marín, H., Ruggeri, R., & Percival, W. J. 2015, MNRAS, 453, L11
Blot, I., Crocce, M., Sefusatti, E., et al. 2019, MNRAS, 485, 2806
Boruah, S. S., Hudson, M. J., & Lavaux, G. 2020, MNRAS, 498, 2703
Carrick, J., Turnbull, S. J., Lavaux, G., & Hudson, M. J. 2015, MNRAS, 450, 317
Colin, J., Mohayaee, R., Sarkar, S., & Shahfieloo, A. 2011, MNRAS, 414, 264
Colless, M., Saglia, R. P., Burstein, D., et al. 2001, MNRAS, 321, 277
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Courtois, H. M., Tully, R. B., Fisher, J. R., et al. 2009, AJ, 138, 1938
Courtois, H. M., Tully, R. B., Makarov, D. O., et al. 2011, MNRAS, 414, 2005
Dai, D.-C., Kinney, W. H., & Stojkovic, D. 2011, JCAP, 2011, 015
Davis, M.,® , & Scrimgeour, M. I. 2014, MNRAS, 442, 1117
Djorgovski, S., & Davis, M. 1987, ApJ, 313, 59
Dressler, A., Lynden-Bell, D., Burstein, D., et al. 1987, ApJ, 313, 42
Elahi, P. J. 2009, PhD thesis, Queen’s Univ. Canada
Elahi, P. J., Welker, C., Power, C., et al. 2018, MNRAS, 475, 5338
Feldman, F. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23
Feldman, H. A., & Watkins, R. 2008, MNRAS, 387, 825
Feldman, H. A., Watkins, R., & Hudson, M. J. 2010, MNRAS, 407, 2328
Fisher, J. R., & Tully, R. B. 1981, ApJS, 47, 139
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gioioli, C., Tormen, G., & van den Bosch, F. C. 2008, MNRAS, 386, 2135
Görski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, ApJ, 861, 49
Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, AJ, 142, 170
Hinton, S. R. 2016, JOSS, 1, 00045
Hong, T., Springob, C. M., Staveley-Smith, L., et al. 2014, MNRAS, 445, 402
Howlett, C. 2019, MNRAS, 487, 5209
Howlett, C., Manera, M., & Percival, W. J. 2015a, A&A, 12, 109
Howlett, C., Manera, M., & Percival, W. J. 2015b, L-PICOLA: Fast dark matter simulation code, Astrophysics Source Code Library, ascl:1507.004
Howlett, C., Ross, A. J., Samushia, L., Percival, W. J., & Manera, M. 2015c, MNRAS, 449, 848
Howlett, C., Staveley-Smith, L., Elahi, P. J., et al. 2017, MNRAS, 471, 3135
Huchmeier, W. K., & Richter, O. G. 1989, A General Catalog of HI Observations of Galaxies. The Reference Catalog (Berlin: Springer), 350
Hui, L., & Greene, P. B. 2006, PhilD, 73, 123526
Hunter, J. D. 2007, CSE, 9, 90
Jaffe, A. H., & Kaiser, N. 1995, ApJ, 455, 26
Jarrett, T. 2004, PASA, 21, 396
Johnson, A., Blake, C., Koda, J., et al. 2014, MNRAS, 444, 3926
Kaiser, N. 1988, MNRAS, 231, 149
Knebe, A., Pearce, F. R., Luc, H., et al. 2013, MNRAS, 435, 1618
Kourkchi, E., Tully, R. B., Anand, G. S., et al. 2020a, ApJ, 896, 3
