

Prior-free cosmological parameter estimation of Cosmicflows-4

C. Duangchan^{1,2,*}, A. Valade³, N. I. Libeskind¹, and M. Steinmetz¹

¹ Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany

² Universität Potsdam, Institut für Physik und Astronomie, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

³ Aix Marseille Université, CNRS/IN2P3, CPPM, Marseille, France

Received 29 July 2025 / Accepted 17 February 2026

ABSTRACT

Context. As tracers of the underlying mass distributions, the peculiar velocities of galaxies are valuable probes of the Universe, allowing us to measure the Hubble constant or to map the large-scale structure and its dynamics. The catalogs of peculiar velocities, however, are noisy, scarce, and prone to various interpretation biases.

Aims. We measured the radial and bulk flow directly from the largest available sample of peculiar velocities and did not impose a cosmological prior on the velocity field. Furthermore, a minimum assumption on the shape of the radial flow at large distances enabled us to estimate the local Hubble constant.

Methods. To address these issues, we analyzed the Cosmicflows-4 catalog (CF4), the most extensive catalog of galaxy peculiar velocities, reaching a redshift $z = 0.1$. Specifically, we constructed a forward-modeling approach assuming only a flat Universe, which reconstructs the radial and bulk flows of the velocity field directly from measurements of peculiar velocities. Our method was tested on a series of 64 simulated catalogs that mimicked the complex selection function of CF4 in space and in magnitude. Based on our mock data, we propose a simulation-based correction method that we applied to the CF4 data.

Results. Our method recovers the radial flow and the direction and magnitude of the bulk flow throughout the covered volume without bias. The incompleteness of the data leads to a systematic amplification of the underlying flow, however, which in turn leads to a systematic overestimation of the amplitude of the bulk flow. By estimating the (cosmic) variance of the density field at large distances from the Lambda Cold Dark Matter (Λ CDM) model, we were able to extract a value of 75.9 ± 1 (stat) km/s/Mpc from the radial inflow. With regard to the bulk flow, a 3σ tension is found with Λ CDM on the supergalactic X direction and on the magnitude of the bulk flow around 140 Mpc/h and 240 Mpc/h. In summary, our work confirms the existing tension on the Hubble constant measured locally and a significant tension in the local bulk flow with Λ CDM predictions. More work is needed to handle all the possible sources of systematic errors inherent to the treating of composite catalogs such as CF4.

Key words. methods: data analysis – techniques: radial velocities – cosmological parameters

1. Introduction

The motion of galaxies inspired cosmologists for at least a century when Hubble noted that the redshifting of absorption lines was proportional to distance, thereby discovering the expansion of the Universe (Hubble 1929). The Hubble law, predicted by Friedman some years prior, is simple: $cz = H_0d$, where H_0 is (the Hubble) constant, z is the redshift, c is the speed of light, and d is the distance. The redshift Hubble used, however, is now known to be a combination of the velocity due to cosmic expansion and the line-of-sight component of a galaxy's peculiar velocity, namely the velocity caused by the net gravitational pull exerted by the full mass distribution of the Universe (Kaiser 1987; Dekel et al. 1993). At first order, the equation should read $cz = H_0d + v_r$, with (v_r) the radial component of the peculiar velocity. Thus, the peculiar radial velocity of a galaxy can be measured based on a joint measurement of redshift and distance,

$$v_r = cz - H_0d. \quad (1)$$

Catalogs combining redshifts and distances have demonstrated their relevance for cosmology. They enabled the discovery of the accelerated expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999), yielded a local value of H_0 (see

Di Valentino et al. 2025, for a review on the H_0 tension), permitted us to constrain cosmological parameters such as the growth factor $f\sigma_8$ (e.g., Carreres et al. 2023; Rosselli et al. 2025) or the matter density parameter Ω_M (Brout et al. 2022), tested non-standard models of (evolving) dark energy with cold dark matter (CDM) such as w_0w_a CDM (Brout et al. 2022), explored anisotropies in the expansion of the Universe (Giani et al. 2024; Boubel et al. 2025; Kalbouneh et al. 2025b,a; Stiskalek et al. 2026), or investigated the misalignment between local motion and the Cosmic Microwave Background (CMB) dipole (e.g., Sorrenti et al. 2023). Locally, peculiar velocities have been used to map the large-scale structure (e.g., Courtois et al. 2013; Valade et al. 2024) and the discovery of local dynamical structures such as the Great Attractor (Dressler et al. 1987), the Laniakea basin of attraction (Tully et al. 2014), or the dipole repeller (Hoffman et al. 2017). Last but not least, measurements of peculiar velocities can capture the so-called moments of the velocity field (Regos & Szalay 1989, namely the dipole, quadrupole, etc) which can be easily compared to theoretical predictions, providing a valuable test for cosmological models. This is the path we explore in this work.

While systematic redshift surveys with sufficient precision became available in the mid 1980s (de Lapparent et al. 1986), it remains difficult to measure the physical distances to galaxies. All distance measurements rely on essentially the same

* Corresponding author: cduangchan@aip.de

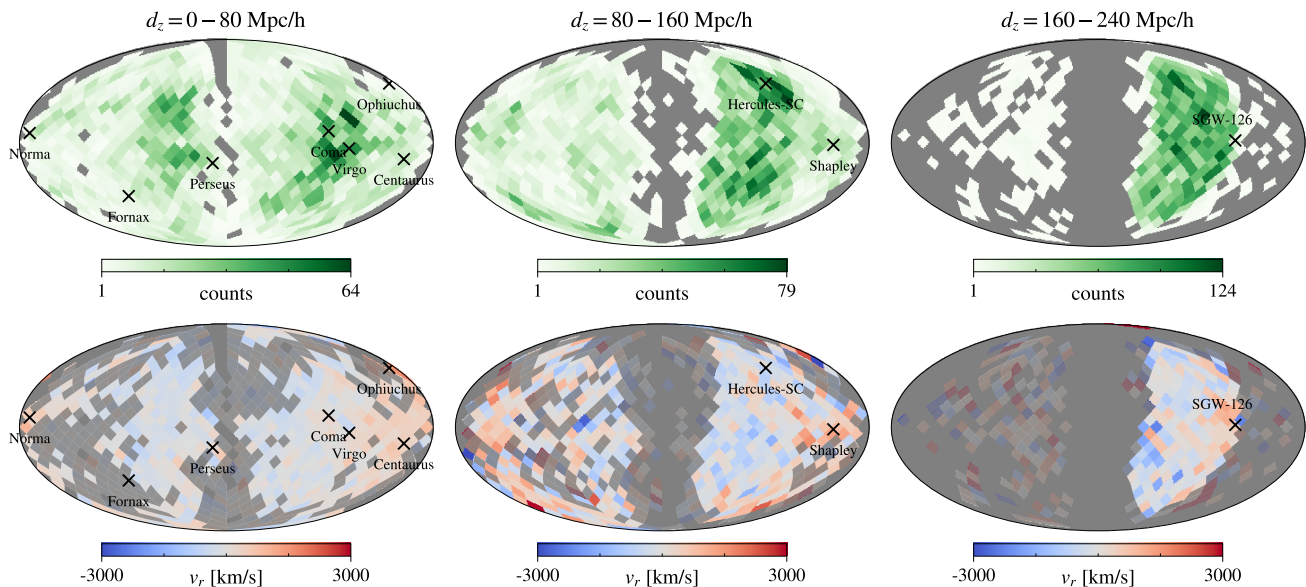


Fig. 1. Visualization of the CF4 data, split by redshift distance d_z . Some of the local clusters within the given range are marked with a black cross with their names. **Top:** Number of grouped galaxies within the HEALPix pixel. **Bottom:** Median of radial peculiar velocity for each direction. The pixels containing fewer than 10 data points are shaded in gray.

principle: the comparison between the observed magnitude or size with an assumed absolute magnitude or size. In the local Universe, where individual stars can be resolved, this is typically more accurate since (e.g.) the magnitude of stars at the tip of the red giant branch (Lee et al. 1993) is a well-established standard candle. Beyond this region, absolute magnitudes need to be derived from supernovae and scaling relations such as the Tully-Fisher (Tully & Fisher 1977) or the fundamental plane relations (Dressler et al. 1987). Supernovae, although observable to great distances, are rare, unfortunately.

For observational reasons, the vast majority of distance measurements are provided in distance moduli $\mu \propto \log(d)$. Issues arise when errors on μ (assumed to be normally distributed) are directly placed on the distance, since these then transform into log-normal distance errors, leading to biases (Strauss & Willick 1995). This breathing mode motivated the development of various distance estimators (e.g., Watkins & Feldman 2015; Sorce et al. 2016; Hoffman et al. 2021; Watkins et al. 2023; Sorce et al. 2024), which need to be applied to these data prior to any analysis, such as the moments of the velocity field we explore here. While carefully derived, these methods remain very spontaneous and result in the designing of heterogeneous analysis pipelines.

Methods designed to reconstruct the bulk flow from corrected velocity data (i.e., data from which this log-normal bias was removed) can be separated into three families. First, the maximum likelihood method proposed by Kaiser (1988) and applied to various datasets by, for example, Ma et al. (2011), Ma & Pan (2014), Qin et al. (2018), Qin et al. (2021), Howlett et al. (2022), and Whitford et al. (2023). Second, the minimum-variance method developed by Watkins et al. (2009), later applied by, for example, Feldman et al. (2010), Watkins & Feldman (2015), Watkins et al. (2023), Scrimgeour et al. (2016), Peery et al. (2018), and Whitford et al. (2023). Third, the Wiener filter framework, augmented by constrained realizations to sample uncertainties (Hoffman & Ribak 1992; Zaroubi et al. 1995), which has been applied in successive releases of the Cosmicflows catalogs (CF2, CF3, CF4) (Tully & Courtois 2012; Tully et al.

2016, 2023) and associated analyses (Hoffman et al. 2015, 2021, 2024). The ability to reconstruct the full 3D velocity field of this last approach, assuming a Λ CDM prior and the CMB-derived matter power spectrum, is among its strengths. Since it enforces a theoretical Λ CDM velocity-velocity correlation, this approach cannot truly probe the Λ CDM model, especially in regions beyond a few dozen megaparsec, where the noise dominates the signal and data are scarce.

An alternative approach, known as forward modeling, consists of deriving the full posterior probability of the model parameters of interest (e.g., radial flow, bulk flow, and H_0) given the data using a Bayesian framework. The posterior probability is then explored with a Monte Carlo algorithm. Although computationally expensive, forward modeling surpasses more traditional methods as it can describe the complexity of the distance observations with a fidelity higher than that of conventional distance estimators.

The application of forward modeling to reconstructing the velocity field was first proposed by Lavaux (2016) and was notably applied by Graziani et al. (2019) to the Cosmicflows-3 catalog. The method has been further improved, as described by Boruah et al. (2021) and Valade et al. (2022), and was applied to the Cosmicflows-4 data notably by Courtois et al. (2023) and Valade et al. (2024). While the forward-modeling approach has shown a greater ability to reconstruct small scales than the traditional Wiener filter approach (Valade et al. 2023), its ability to recover the tension with Λ CDM on the large scales is even more limited than that of the Wiener filter. In addition to assuming an underlying Λ CDM prior, the reconstructions are performed in periodic boxes (e.g., 1 Gpc/h in Valade et al. (2024)), which is known to further dampen the recovered large-scale motion due to the missing large-scale tidal modes. On the one hand, strong priors tend to hide the tension in the data with respect to the model. On the other hand, in the absence of priors, parameters of interest may be less well constrained and the scientific statement is more modest.

A forward-modeling reconstruction of the first spherical modes of the velocity field has been presented by Boubel et al.

(2024), who also provided an estimate of the Hubble constant H_0 . With a similar model, the anisotropy of the Hubble constant has been discussed by Boubel et al. (2025). The approach developed in these works, however, focused on distances obtained from the Tully-Fisher relation, thus limiting the application of their method to the Cosmicflows-4 Tully-Fisher sample (Kourkchi et al. 2022). Although it is affected by complex magnitude-selection effects, this dataset is relatively isotropic.

We present a forward modeling of the first two spherical modes of the velocity field (radial flow and bulk flow) from peculiar velocity data that makes minimum Λ CDM assumptions. As it relies on the Bayesian framework, the method can directly treat the observational data without relying on an intermediary distance estimator such as Watkins & Feldman (2015). Limiting the reconstruction to the first modes of the velocity enables us to depart from strong priors on the power spectrum of the velocity field, such as in Hoffman et al. (2024), minimizing the effect of the assumed cosmology on the resulting radial and bulk flows. Finally, compared to Boubel et al. (2024), our more generalist, but less realistic, treatment of the data allows us to apply our method to the entire Cosmicflows-4 catalog of peculiar velocities. This leads to results on the velocity flow at greater depth while raising issues of anisotropy coupled to the selection in magnitude (Andersen et al. 2016).

2. Observations

As mentioned in the preceding section, we analyzed the radial peculiar velocity field. For this purpose, we employed the Cosmicflows-4 (CF4) catalog (Tully et al. 2023), which constitutes the most comprehensive collection of galaxy peculiar velocities to date, containing about 56 000 galaxies gathered in about 38 000 groups of galaxies¹ so as to mitigate the effects for nonlinear motion within bound structures. This work is based on the grouped catalog, and all observed redshifts are considered in the CMB frame.

Rather than a single survey, the CF4 catalog is an intercalibrated compilation of several subcatalogs found in the literature: a catalog of catalogs. The fundamental-plane scaling relation (Djorgovski & Davis 1987) provides about three-quarters of the measured distances, with an uncertainty on the distance modulus of $\sigma_\mu \sim 0.4$. Many of these data are gathered from the Sloan Digital Sky Survey Peculiar Velocity (Howlett et al. 2022, SDSS-PV) catalog, which extends to $z = 0.1$ ($d \sim 300$ Mpc/h), but is restricted to the relatively narrow footprint; another significant component comes from the 6 degree field galaxy redshift survey (Magoulas et al. 2016, 6dFGRSv), which covers the entire southern celestial hemisphere up to $z = 0.055$ ($d \sim 160$ Mpc/h).

The Tully-Fisher scaling relation (Tully & Fisher 1977) is used to derive the majority of the remaining entries, with an uncertainty similar to that of the fundamental plane. These distances are mostly provided by the CF4-TF catalog (Kourkchi et al. 2022), which is an analysis of SDSS HI fluxes, line widths, and redshifts. CF4-TF covers roughly the entire sky, with a slight preference for the northern celestial hemisphere. The redshift distribution of the Tully-Fisher sample peaks at $z = 0.02$ ($d \sim 60$ Mpc/h) and slowly decreases until $z = 0.055$.

Cosmicflows-4 contains about 1 000 type 1a supernovae extracted from the Pantheon+ and SHOES catalogs (Riess et al. 2022; Scolnic et al. 2022, respectively), found mostly below $z = 0.03$ ($d \sim 120$ Mpc/h) and reaching $z \sim 0.1$. Finally, various

higher-precision distance estimators ($\sigma_\mu \leq 0.1$) are also present in the CF4 compilation, but are limited to $z < 0.02$.

The complex footprint of CF4 can be appreciated in Fig. 1, where we show the angular distribution of the catalog, colored by the number of groups in each pixel on the sky in three different radial shells. We considered three different regimes: dense and isotropic below 80 Mpc/h, relatively dense and relatively isotropic between 80–160 Mpc/h, and last, strongly anisotropic between 160–240 Mpc/h. The last shell was cropped to 240 Mpc/h so as to keep shells of constant thickness. The Milky Way shadows an opening angle of about 10° around the Galactic plane; a region of the sky dubbed the zone of avoidance, where it is relatively impossible to obtain distance modulus measurements due to foreground obscuration.

3. Method

In this section, we detail the model and the method we used to reconstruct the moments of the velocity field from measurements of peculiar velocities. The approach can be summarized as follows. The first two moments of the velocity field (i.e., a radial in or out flow plus a net flow in some direction) were chosen at random. Assuming that the errors in the redshift measurements are negligible, we reconstructed the distance to each galaxy from (a more accurate version of) Eq. (1) (see Eq. (4)). The inferred distances, transformed into distance moduli, were then compared with the Cosmicflows-4 observations. Finally, the moments of the velocity field were updated with a Monte Carlo step. The result is a large collection of probable values of radial and bulk flows compatible with the Cosmicflows-4 data. This approach, dubbed forward modeling, is detailed in the following section.

3.1. Forward modeling

Forward modeling heavily relies on Bayes’s theorem, which encodes the posterior probability of the model parameters π given that the data \mathcal{D} were obtained as

$$P(\pi | \mathcal{D}) \propto P(\mathcal{D} | \pi)P(\pi), \quad (2)$$

where the likelihood $P(\mathcal{D} | \pi)$ and the prior $P(\pi)$ encode the observational and the underlying physical models, respectively. As discussed in Section 6, we assumed no prior distribution on our parameters, that is, $P(\pi) = 1$.

We reconstructed the zeroth and first modes of the velocity field (Regos & Szalay 1989). These are often called the monopole and the dipole, respectively. However, since the term “dipole” has come to be understood in common parlance as the motion of an observer relative to background, we refer with it to the mean motion of galaxies flowing either radially (radial flow) or in some direction (bulk flow).

Throughout this paper, the radial flow is referred to as V_r . Although the radial flow has the dimensions of speed (km/s) and one of two directions (inward or outward), it is not a vector. The bulk flow is a 3D vector in the traditional sense and is referred to here as $\mathbf{V} = (V_x, V_y, V_z)$. Our parameter space thus consists of four parameters, namely $\pi = (V_r, V_x, V_y, V_z)$. As detailed below, these quantities fit by the forward model represent the velocity field in concentric shells, while the associated quantities in spheres were constructed as detailed in Section 3.3.

We considered $\mathcal{D} = \{\mu_i^{\text{obs}}\}_{i < n}$, where n is the number of observations (namely the number of entries in the CosmicFlows-4 catalog). The redshift does not feature in \mathcal{D} since the error on it was neglected and therefore did not need to be inferred.

¹ A group can also include a single galaxy.

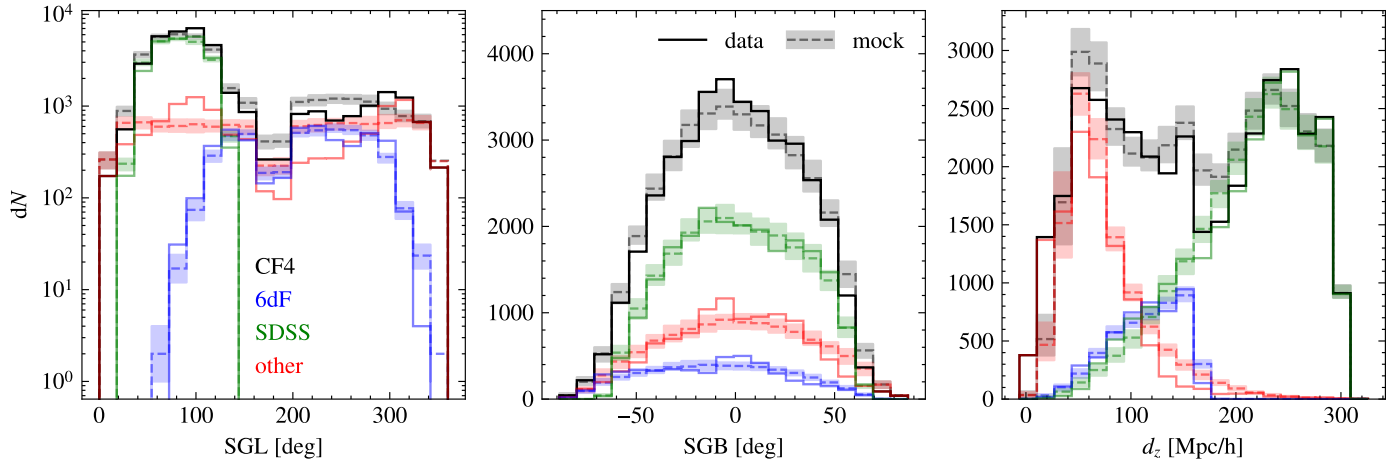


Fig. 2. 3D spatial distribution of the CF4 and mock catalog of each subtype shown as solid and dashed lines, respectively. The shade in a lighter color is the scatter among the mocks.

Moreover, distance moduli are not per se observables, but were derived from other observables through relations such as the Tully-Fisher relation or the fundamental plane.

3.2. Likelihood

As mentioned above, the goal was to make a suggestion for each galaxy's (or galaxy group's) distance modulus, for the purpose of comparison with the measured value. Thus, the distance modulus for each entry was constructed as follows. First, the peculiar velocity of the constraint i was computed as

$$v_r^i = V_r + \mathbf{V} \cdot \hat{\mathbf{r}}^i, \quad (3)$$

where $\hat{\mathbf{r}}$ is a unit vector pointing toward the constraint. Assuming that the full redshift z is known with negligible uncertainty, we computed the proposed cosmological redshift (namely the redshift due to the Hubble expansion) from the proposed peculiar velocity,

$$z_{\text{cos}} = \frac{z - v_r/c}{1 + v_r/c}, \quad (4)$$

which is an accurate version of Eq. (1). Then, the proposed physical distance d was obtained by numerically integrating

$$d = \frac{c}{H_0} \int_0^{z_{\text{cos}}} \frac{dz}{\sqrt{(1+z)^3 \Omega_M + (1-\Omega_M)}}. \quad (5)$$

Last, the distance modulus was obtained,

$$\mu = 5 \log_{10} [(1 + z_{\text{cos}})d] + 25. \quad (6)$$

The Hubble parameter H_0 and the density parameter Ω_M were not explored by our forward model. Instead, their respective effects on the results are discussed in Section 4 and Appendix B. We note that Eq. (5) makes assumptions on the background cosmology, which are further discussed in Section 6.

After these calculations were carried out, the likelihood function was constructed, namely, the ensemble of proposed values of the distance moduli were compared with the observed values and their errors. The distance modulus errors were assumed to be normally distributed around the mean value and independent of each other, which is the standard approach (e.g., Hoffman et al. 2021; Boruah et al. 2021; Valade et al. 2022; Tully et al. 2023).

The likelihood over all the reconstructed distance moduli is expressed as the exponential product of the difference between the observed distance modulus and the proposed one, divided by the error on the distance modulus,

$$P(\{\mu_{i}^{\text{obs}}\} | \pi) = \prod_i P(\mu_{i}^{\text{obs}} | \pi) \quad (7)$$

$$\propto \prod_i \exp\left(-\frac{(\mu_{i}^{\text{obs}} - \mu_i(\pi))^2}{2\sigma_{\mu,i}^2}\right). \quad (8)$$

We did not attempt to model the volume or selection effects (i.e., $P(d) = \text{constant}$). This is loosely consistent with the distance distribution of the entire CF4, which is roughly flat, as shown in Fig. 2.

The posterior distribution was explored by means of a Hamiltonian Monte Carlo (HMC) technique (Neal 2011). An HMC is an efficient Monte Carlo technique that enables the exploration of the full permissible parameter space. Each Monte Carlo chain comprises 1000 states, preceded by a 100-step burn-in phase. Given the simplicity of our problem, any reasonable choice of hyperparameters led to an optimal convergence of the algorithm. The HMC implementation made use of TENSORFLOW-PROBABILITY (Dillon et al. 2017) with a JAX backend (Bradbury et al. 2018).

3.3. Inference of the velocity field in concentric shells

Since distances are measured via their observed distance moduli, the error on the peculiar velocity (σ_v) for each entry of the catalog is proportional to the measured error on the distance modulus (σ_μ) and its (true) distance. For the vast majority of data points, the redshift distance is more precise than the physical distance inferred from the distance ladder,

$$\sigma_v \approx \frac{\log(10)}{5} \sigma_\mu \times cz. \quad (9)$$

Owing to distance-dependent errors, most of the constraining power of the data is found near the observer. This implies that running the inference method on the entire catalog at once would lead to a bulk flow estimated at an effective radius

$$r_{\text{eff}} = \frac{c}{H_0} \frac{\sum_i z_i / \sigma_{v,i}}{\sum_i 1 / \sigma_{v,i}} \approx 100 \text{ Mpc} (\approx 75 \text{ Mpc}/h), \quad (10)$$

Table 1. Summary of the properties of each subcomponent of the CF4 mock catalogs.

Component	N	β	m_c	σ_μ
SDSS-PV	22 467	0.6	13.3	0.5
6dFGRSv	4768	1.5	11.7	0.5
other	10 822	0.45	10	$\mathcal{N}(0.4, 0.1)$

Notes. Here, N is the number of galaxies, β and m_c are the parameters fit to reproduce a magnitude-limited sample following the procedure given in Appendix A, and σ_μ is the typical error on the distance modulus. \mathcal{N} represents a normal distribution.

which is much shallower than that of the data (whose median distance is 155 Mpc/h). This is due to (1) the decrease in the number of constraints with distance and (2) the increase in individual error on the constraints. Additionally, it is to be noted that the CF4 catalog becomes very anisotropic beyond 165 Mpc/h ($z = 0.055$).

We designed an approach to overcome this limitation. Essentially, the Monte Carlo fits were first run independently in concentric shells, and then, the values in concentric spheres were reconstructed by averaging the contributions of the shells by their respective volume. More precisely, the catalog was split (according to each galaxy redshift distance $r_i \approx cz_i/H_{100}$ Mpc/h) into concentric shells of thickness $\Delta r = 20$ Mpc/h, in each of which the Monte Carlo algorithm was run separately. The value of the sphere of radius (R) was then computed as

$$\begin{aligned}
 r_j &= j \times \Delta r, \\
 \mathcal{V}_j &= \frac{4}{3}\pi(r_j^3 - r_{j-1}^3), \\
 Q^{(s)}(R) &= \frac{\sum_{j^{r_j < R}} Q_j^{(s)} \mathcal{V}_j}{\sum_{j^{r_j < R}} \mathcal{V}_j},
 \end{aligned} \tag{11}$$

where $Q_j^{(s)}$ is the s th Monte Carlo step of the quantity Q in the j th shell (i.e., Q_j). Equation (11) was applied independently to the radial flow (V_r) and the components of the bulk flow (V_X, V_Y, V_Z).

$Q^{(s)}(R)$ only constitute a pseudo-Monte Carlo chain and not an actual radius-dependent reconstruction. Namely, correlations between fitted values at different radii cannot be interpreted.

3.4. Mock of CF4

Before the forward-model method described above was applied to the CF4 data, we determined its precision and accuracy, namely how well it measures the two modes on the velocity field in which we were interested (the radial infall and the bulk flow). It was thus applied to mock observations of the CF4 grouped catalog. In this way, the estimated velocity field can be compared directly to the ground truth. Specifically, due to its complex footprint (anisotropy, sparse sampling, large errors, and luminosity limitation), the construction of reliable mocks of the CF catalog is nontrivial, and several methods have been developed (e.g., Doumler et al. 2013; Sorce et al. 2016; Qin et al. 2021; Valade et al. 2023; Whitford et al. 2023; Boubel et al. 2024). We note that applying the method to mock catalogs drawn from simulations necessitates choosing a cosmology. In this way, the accuracy of the method loses its prior-free nature. However, we recall that the assumption of a cosmology only affects the

interpretation of the results: the method itself remains largely cosmology-free.

The BIG MULTIDARK N -body simulation (Riebe et al. 2013), which covers a volume of $L_{\text{box}} = (2.5 \text{ Gpc}/h)^3$ with $N = 3840^3$ particles, was used for this purpose. The cosmology of the simulation is similar to that of Planck XIII ($h = 0.6777$, $\Omega_M = 0.307$; Planck Collaboration XIII 2016). Halos with more than 20 particles were considered, namely a lower-mass resolution of $4.7 \times 10^{11} M_\odot/h$.

The cosmological box was sliced into $4 \times 4 \times 4 = 64$ cubic subboxes of side length $(625 \text{ Mpc}/h)^3$, each of which was sufficiently large to enclose the full CF4 catalog. The observer was set at the center of each subbox. We acknowledge that this simplistic approach implies that the observers are statistically found in a cosmological underdense region, which affects results on the radial flow as discussed in Section 4.1. The box coordinates were (arbitrarily) associated with supergalactic coordinates (SGX, SGY, and SGZ).

To grasp the complexity of the CF4 dataset, we separately constructed three of its subcatalogs: SDSS-PV, 6dFGRSv, and ‘‘other’’. This enabled us to model the anisotropic footprints of the scaling-relation surveys contributing to CF4, since the catalog anisotropy is dominated by the two wide-angle sky surveys (SDSS-PV and 6dFGRSv).

Our method also focused on modeling a (simple) magnitude-limitation². The large number of instruments and distance-derivation methods on which the CF4 catalog is based makes a magnitude analysis very difficult. Not only are CF4 entries extracted from surveys with different magnitude limitations (among other method-dependent selection effects), but their magnitudes are observed in different bands. Furthermore, the simulation from which the halos are taken is dark-matter only, meaning that neither are halos populated by galaxies, nor are their luminosities simulated. We thus opted for an empirical approach detailed in Appendix A. A more accurate procedure is left to further works.

For each subbox, CF4-mock halos were extracted from the simulation according to the following recipe.

1. An angular mask was applied that mimicked the sky coverage of the surveys: the footprint of the SDSS-PV, the southern celestial hemisphere ($\text{dec} < 0^\circ$) for 6dFGRSv, and the full sky for other. We also included an additional cut of 10° around the Galactic plane so as to model the zone of avoidance. Halos were included or excluded based on whether they were inside or outside of the survey footprints and the zone of avoidance.
2. Each halo was given an observed redshift and distance modulus by applying Eqs. (4) to (6) to the radial peculiar velocity and distance of each halo.
3. Galaxies were painted onto halos by computing an absolute magnitude from the halo mass. For each subcatalog, the empirical luminosity-mass relation was fit alongside an upper (apparent) magnitude cut so as to reconstruct the radial distribution of the observations. The exact procedure is detailed in Appendix A.
4. Halos were then randomly selected out of the magnitude-limited sample.
5. An observational scatter depending on the subcatalog was added on the distance moduli following Table 1.

² This effect is also often referred to as a Malmquist bias. We refrain from using this term, as it has been used for a wider range of observational biases.

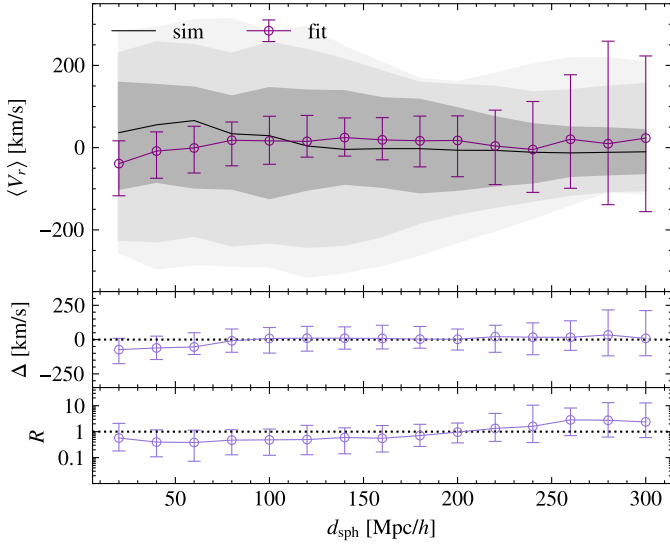


Fig. 3. Comparison between the average radial peculiar velocity from the simulation and the fit results. In the top panel, the black line represents the median of the true value across 64 mocks, with the shaded gray regions indicating 68%–95%–99.7% percentile intervals, equivalent to 1, 2, and 3σ . The purple error bars represent the median of the fit results, with 68% error bars. The middle and bottom rows show the difference Δ and ratio R between the estimated and simulation values, respectively. For each of these parameters, the median across the 64 mocks is shown by the central light purple line, with error bars indicating the percentile interval covering the central 68% of the distribution.

The mock CF4 catalog was constructed by merging the mock SDSS-PV, 6dFGS_z, and the other data.

In Fig. 2 we show the angular and redshift distributions of the mock observations and the CF4 catalog. The spatial distribution of the CF4 catalog is reproduced with great fidelity, demonstrating the ability of our mock procedure to reproduce the complexity of the CF4 compilation well.

While the distributions of SDSS-PV and 6dFGS_v agree well in Fig. 2, some disparities can be seen in the other component. Although the distribution in d_z is well captured, the angular distributions (SGB and SGL) have minor differences. This is due to the crude modeling of the other component, which in reality contains a number of different anisotropic surveys.

4. Results: Mock data

4.1. Radial flow

The reconstruction of the radial flow of the 64 mocks in spherical volumes out to 300 Mpc/h is presented in the top panel of Fig. 3. The recovered radial flow closely follows the ground truth of the simulations, demonstrating that our method reliably reconstructs the radial flow within the expected statistical uncertainty. The error in the method due to the sampling increases with distance, with a remarkable change in behavior after 160 Mpc/h. As expected, this effect stems from the observational limitations modeled in our mock data, namely scarcity of the data, large observational uncertainties, and luminosity limitations.

In the bottom panels of Fig. 3, the under- or overestimate of the radial flow as a function of distance is shown in two ways. We focus on the misestimate in km/s represented by $\Delta = V_r - V_r^{\text{sim}}$, namely the estimated infall velocity at a given distance minus the corresponding value from the simulation. We also examined the ratio $R = |V_r|/|V_r^{\text{sim}}|$. The method underestimates the slight

outflow found below 50 Mpc/h seen in the simulations. The rising of the median $\langle V_r \rangle$ within this region reflects the observer’s underdense position. The underestimation of this outflow is due to the (poor) simulation resolution (in the exact region where the observations can see the faintest objects) and implies a disparity between the CF4 catalog and our mocks in the innermost region (also seen in Fig. 2). In the absence of constraints, the mean tends to zero while the scatter increases, but due to the absence of a Λ CDM prior, this convergence is expected to be weaker than it is for field-level inference methods such as Hoffman et al. (2024).

On larger scales, beyond ~ 200 Mpc/h, the method slightly underestimates the true infall velocity by about 20 km/s, which is small with respect to the expected Λ CDM cosmic variance (more than 100 km/s at 200 Mpc/h). More prominently, observational limitations amplify the signal starting 160 Mpc/h, reaching a factor of two at the 200 Mpc/h. The reconstructed velocity and true velocity begin to diverge at a distance of about 160 Mpc/h. This result is expected, as CF4 becomes strongly anisotropic beyond this distance, as it is then limited to the SDSS-PV sample, which covers only a small fraction of the sky.

4.2. Bulk flow

How well the forward model proposed here works is shown in the bottom two rows of Fig. 4. Out to ~ 140 Mpc/h, where the CF4 catalog is relatively isotropic, the method is able to recover each component with a variance comparable to that of Λ CDM. No significant over- or underestimation is found in any of the components. Beyond 160 Mpc/h, where only the SDSS-PV dominates, the method amplifies the underlying signal, leading to errors that exceed cosmic variance and reach a factor of 5 at the edge of the catalog. We note that the R factor increases faster on the Y component than on X and Z, which we ascribe to the geometry of the survey: the SDSS-PV is concentrated around the Y-axis.

The magnitude of the bulk flow (namely the square root of the sum of the squares of each component) is presented in the right-most panel of Fig. 4. We show the direct application of Eq. (11) to the Monte Carlo computed magnitude $Q^{(s)} = |V|^{(s)}$ (in purple). The estimated magnitude of the bulk flow is overestimated at all distances: the greater the distance, the greater the overestimation. We distinguish three regimes: (1) up to 60 Mpc/h, the reconstructed amplitude is accurate, (2) between 60 Mpc/h and 160 Mpc/h, the bulk flow decreases more slowly than the Λ CDM predictions; it is characterized by a slow, distance-increasing amplification, and (3) beyond 160 Mpc/h, the amplification becomes stronger and the reconstructed bulk flow begins to increase artificially, which is at odds with the Λ CDM prediction.

The salient points to be gleaned from Fig. 4 are thus that

- The components and the amplitude of the bulk flow are well estimated within ~ 60 Mpc/h;
- A modest amplification is found on all components, mostly Y, between 60 Mpc/h and 160 Mpc/h, resulting in a noticeable overestimation of the amplitude of the bulk flow in this distance range;
- A strong amplification of each of the components and thus of the amplitude of the bulk flow beyond 160 Mpc/h appears to be symptomatic of the strong anisotropy of CF4 in this region.

This analysis shows that the CF4 observational limits affect the reconstructed bulk flow and lead to an artificial amplification of the estimated bulk velocity. Thus, a naive interpretation of the

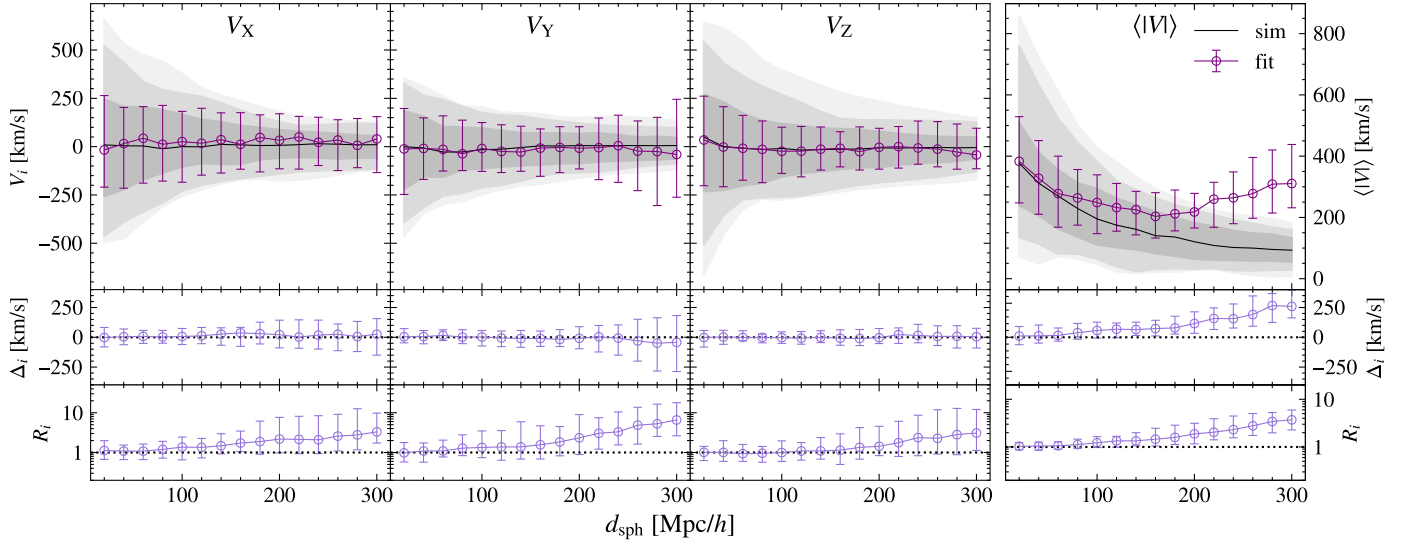


Fig. 4. Same as Fig. 3 but for the three Cartesian components of the bulk flow ($V_{X/Y/Z}$, first three columns) and its magnitude (right-most panel).

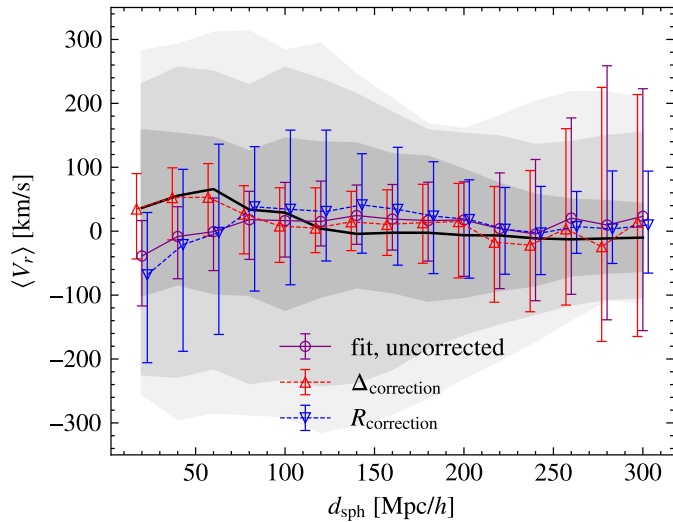


Fig. 5. Application of the corrections to the radial flow fit of the mocks. The median of the radial flow in the simulation, expected to be minimal, among the mocks is shown as a black line, with gray shades indicating the 68%–95%–99.7% intervals demonstrating the cosmic variance. Corrections based on differences and ratios are shown as dashed red and blue lines, respectively, with error bars indicating the 68% interval across the 64 corrected radial flows.

reconstructed bulk flow would result in an important but artificial tension with the cosmological model.

4.3. Simulation-based correction

The tests on mock catalogs conducted in the preceding section not only give us confidence in the regime where we expect the method to perform well, but also indicate how we can interpret (and thus correct for) the regions in which the method overestimates the radial or bulk flow. In each shell of each component of the bulk flow and separately for the radial flow, we can use the value of R and Δ to correct the measured value and thus recover the true velocity field moments. This was done by

1. Removing the mean over- or underestimation, namely subtracting the value of Δ , (the Δ correction);
 2. Dividing by a factor of the amplification R (R correction).
- The correction we applied was based on a set of mocks, and there is an obvious variance across these mocks and their corrections. In other words, in some of our mocks, the method might perform well, while in others, it might perform worse. In general, we took the median correction, but we also included this variance in the error bar of our corrected results.

To account for the variance stemming from the correction, we write

$$P_{\Delta}(Q_{\text{corr}} | D) = \int P(Q_{\text{fit}} = Q_{\text{corr}} - \Delta | D)P(\Delta)d\Delta, \quad (12)$$

$$P_R(Q_{\text{corr}} | D) = \int P(Q_{\text{fit}} = Q_{\text{corr}} \times R | D)P(R)dR, \quad (13)$$

where $P(Q_{\text{fit}} | D)$ is the result of the forward model described above, and $P(\Delta)$ and $P(R)$ are the probability distributions of the correction factors. In practice, the uncorrected output of the forward model was redressed separately by each of the 64 R or Δ corrections derived from the 64 mock universes, so as to sample the uncertainty in the correction arising from cosmic variance. The final results consist thus of 64 corrected profiles, on which the median and quantiles were computed. We note that this method might overestimate the uncertainty.

As shown in Figs. 5 and 6, the R and Δ corrections behave differently depending on whether the bulk flow or the radial flow is examined: the Δ correction is better suited to radial measurements, while the R correction works ideally for the bulk flow. We highlight that the correction on the bulk flow was performed on the components, not on the amplitude.

5. Results: Cosmicflows-4 data

5.1. Hubble parameter and radial flow

Catalogs of calibrated distances, such as CF4, are the best way to obtain a local estimate of the Hubble parameter H_0 . A simple regression of the $cz = H_0d$ scatter plot might be run, but this is prone to several issues. First, the fact that most of the information is concentrated near the observer (where the observations are

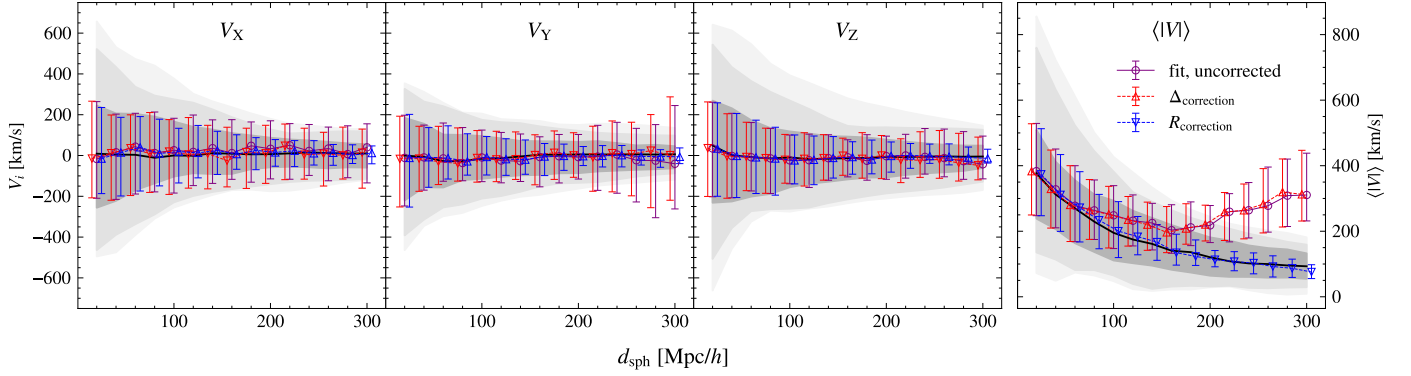


Fig. 6. Same as Fig. 5 but for the bulk flow components and magnitude. Note that corrections were applied to the components, not to the magnitude of the vector.

dense and uncertainties are small) leads to a very small effective radius of the fitting. Second, this approach overlooks the large-scale radial modes of the velocity. Furthermore, in the presence of an anisotropic sampling of the sky, such as CF4, the bulk flow might pollute the estimate of H_0 and so might higher modes of the velocity field, although observations tend to show that they have significantly less power (Kalbouneh et al. 2025a).

We introduce a more elaborate approach. For several values of H_0 , the radial component of the velocity flow was fit to the data. The corresponding profiles are presented in Fig. 7, where the uncorrected (top) and Δ corrected (bottom) mean radial flow is presented. Two regimes appear. Up to 200 Mpc/h, the velocity profiles are straight, with a slope depending on the value of H_0 . This is the expected behavior in the presence of an offset in the value of H_0 . Beyond 200 Mpc/h, all the profiles plummet. This can be explained by the combination of (1) the limitation in the sky of the coverage in the SDSS-PV volume, and (2) the overwhelming effect of the Sloan Great Wall (Gott et al. 2005; Valade et al. 2024) in this region, whose distance coincides with the elbow in the radial flow (i.e., about 240 Mpc/h). In essence, all entries in the CF4 catalog beyond 200 Mpc/h fall onto this massive object, creating a local positive radial flow between 200 Mpc/h and 240 Mpc/h and a negative radial flow beyond this distance. A correction of the radial flow changes the local inflow, as indicated in Fig. 3, and the method underestimates the mean radial flow below 100 Mpc/h.

In the absence of prior information on the radial flow, all the values of H_0 are equally likely. In other words, without knowing on which scale no radial flow is expected, it is impossible to further constrain the expansion rate. This scale is a prediction of any cosmological model (e.g., a consequence of the power spectrum of initial perturbations). In the Λ CDM model, $\langle V_r \rangle$ is driven to zero with increasing distance according to the Copernican principle. Symptomatic of the so-called cosmic variance, the progressive convergence to homogeneity can be described by a radius-dependent probability law on $\langle V_r \rangle$. This allowed us to constrain H_0 at each distance probabilistically. Our method essentially reconstructs the probability of H_0 at a certain distance by marginalizing the fitted radial flows over the values predicted by cosmic variance.

The expected Λ CDM $\langle V_r \rangle$ is a Gaussian distribution, namely $P(\langle V_r \rangle) = \mathcal{N}(0, \sigma_{\langle V_r \rangle})$. That this is centered on zero reflects the cosmological principle of homogeneity, while $\sigma_{\langle V_r \rangle}$ is the scatter in the radial flow due to cosmic variance. $\sigma_{\langle V_r \rangle}$ can be analytically derived assuming a power spectrum or empirically sampled from a Λ CDM simulation. In the scope of the

linear theory of structure formation, $\langle V_r \rangle$ and $\sigma_{\langle V_r \rangle}$ both decrease as one over distance³.

The probability of a given H_0 given the measured $\langle V_r \rangle$ and the Λ CDM prior assumption of the scatter $\sigma_{\langle V_r \rangle}$ is

$$P(H_0 | \text{CF4}, \sigma_{\langle V_r \rangle}) = \int P(H_0 | \langle V_r \rangle, \text{CF4}) P(\langle V_r \rangle | \sigma_{\langle V_r \rangle}) d\langle V_r \rangle \quad (14)$$

$$= \int P(\langle V_r \rangle | H_0, \text{CF4}) \mathcal{N}(0, \sigma_{\langle V_r \rangle}) d\langle V_r \rangle. \quad (15)$$

We considered that $P(H_0 | \langle V_r \rangle, \text{CF4})$ was obtained from the forward model of Section 3.2 and does not depend on the cosmological model⁴. The equality $P(\langle V_r \rangle | H_0, \text{CF4}) = P(H_0 | \langle V_r \rangle, \text{CF4})$ is the application of the Bayes theorem in the absence of a prior on either H_0 or $\langle V_r \rangle$, which is consistent with the absence of a prior in the forward model. The quantity $P(\langle V_r \rangle | H_0, \text{CF4})$ is effectively the posterior of the forward model described above, whose results are shown in Fig. 7 at various distances. The integral Eq. (15) was computed numerically.

Figure 8 shows the resulting distribution of H_0 for the corrected and uncorrected radial flow estimator at several distances. The distributions appear to be close to normal, and the interpretation is therefore straightforward. The corrected and uncorrected values agree except at the smallest distance (60 Mpc/h). As they are computed in concentric spheres, the values of H_0 at different distances are not independent. To summarize our results, we chose 160 Mpc/h as our fiducial distance. This is the largest distance we trusted – it lies at the edge of the 6dF data, outside the volume dominated by the Sloan Great Wall, and appears insensitive to the correction (see Fig. 5). The final result thus reads $H_0 = 75.9 \pm 1$ km/s/Mpc. We highlight that the uncertainty might increase strongly with a proper treatment of further systematics errors in the data and in the approach⁵.

³ In a uniform density field superimposed with a power spectrum of fluctuations, the radial velocity as a function of distance is $v_r(r) = -\frac{1}{3}H_0 f r \bar{\delta}(r)$, where f is the linear growth factor, and $\bar{\delta}$ is the density contrast. The scatter depends directly on the scatter in the density field:

$$\sigma_{\langle V_r \rangle}^2(r) = \left(\frac{H_0 f r}{3}\right)^2 \sigma_{\bar{\delta}}^2(r).$$

⁴ Which means here that the forward model does not depend on $\sigma_{\langle V_r \rangle}$, while it is true that the cosmological model enters the redshift-distance relation Eq. (5); see the discussion in Section 6.

⁵ Boubel et al. (2024) reported a systematic error of 3 km/s/Mpc for the CF4 Tully-Fisher sample.

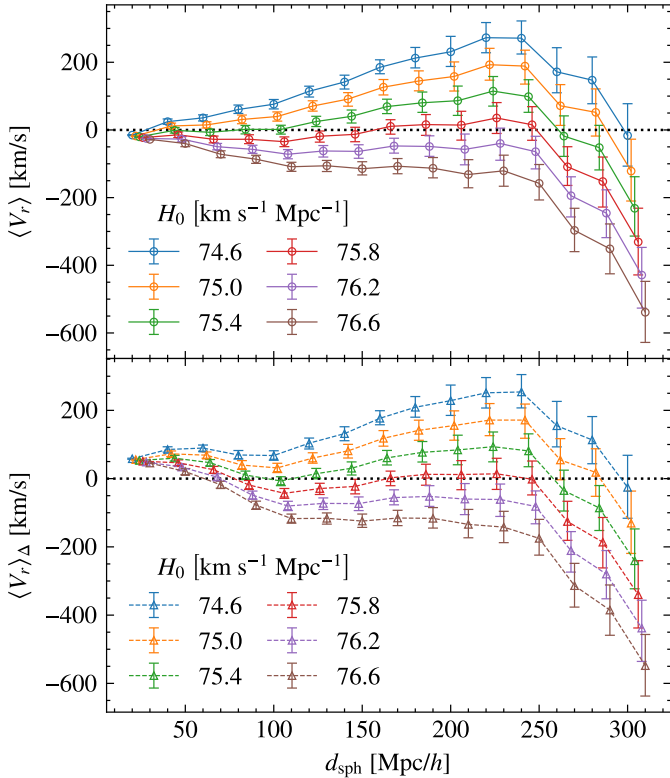


Fig. 7. Radial flow fit of the CF4 data. The top panel shows the result directly from the forward modeling as solid lines, and the bottom panel shows the Δ -corrected result as dashed lines. Each line, corresponding to a different value of H_0 assumed in the model, is distinguished by color. The error bars are the 68% interval of the Monte Carlo steps.

5.2. Bulk flow

The V_Y and V_Z components of the bulk flow are generally well behaved and fully consistent with the Λ CDM predictions, with and without the R correction. This is seen since the magenta and blue curves are within the 3σ Λ CDM expectation for essentially all distances, with a tension below 1σ before 160 Mpc/h. Beyond 160 Mpc/h, the R -correction brings the bulk flow components in line with Λ CDM expectations and alleviates the moderate tension found in the direct results of the forward model.

This is markedly not the case for V_X , which exits the 3σ range of Λ CDM at a distance of ~ 140 Mpc/h, reaching a maximal tension with the cosmological paradigm at more 4σ level at 240 Mpc/h for the corrected and about 8σ at 240 Mpc/h for the uncorrected curves. Although the uncorrected case looks fully inconsistent with the Λ CDM model in the region 140–300 Mpc/h, the corrected V_X is marginally consistent (i.e., at the 3σ level) with Λ CDM between 140 – 220 Mpc/h.

We now analyze the amplitude of the bulk flow, which we show in Fig. 9. The direct application of the method and the corrected bulk flow decrease with increasing radius for the innermost ~ 60 Mpc/h. This behavior is typical of Λ CDM. Beyond this distance, however, the CF4 data display a rise which, in the case of the corrected flow, peaks at 350 km/s at ~ 140 Mpc/h. This value is well within the 3σ scatter of the Λ CDM model. We assume that this behavior is due to the discrepant V_X component. The (uncorrected) bulk flow, on the other hand, exploded and reached a peak of more than 600 km/s at 240 Mpc/h.

We summarize the results of the bulk flow by noting the following: while a rise in one component (V_X) and in the ampli-

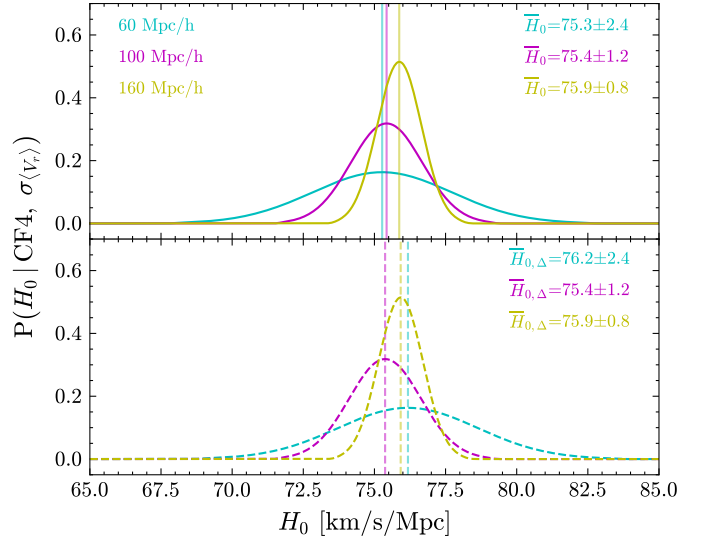


Fig. 8. Probability distribution of the Hubble constant given the CF4 data and Λ CDM radial flow variance at different sphere radii shown in colors, with $\mu \pm 1\sigma$ of each curve. The vertical lines emphasize the curve mean. The marginalized likelihood function is proportional to the (un-) corrected radial flow distribution, whose resulting distributions are shown in the (top) bottom panel.

tude of the bulk flow is rare in Λ CDM, the result presented here is consistent with all other published analyses of CF4, and indeed with a few other studies that were based on other methods and data. Figure 9 presents the values of bulk flow found in the literature at various distances, using various methods. Our approach yields results that are consistent with the literature and the Λ CDM model for the regime out to 60 Mpc/h, where the direct reconstruction of the bulk flow appears to be reliable, and beyond 60 Mpc/h, where the simulation-based correction is effective. The rise in the amplitude of the bulk flow is unusual but appears to be real, or is at least a feature of the currently available data. Some methods even measured even higher bulk flows than presented here, specifically when compared to the corrected bulk flow.

6. Discussion

We have explored a forward-modeling approach to jointly reconstruct the first two modes of the velocity field from measurements of peculiar velocities: the radial inflow, and the bulk flow. The forward model was fit independently on the data, which were split into concentric shells before they were reassembled in spherical volumes, each shell being weighed by its volume. This enabled us to reconstruct the radial and bulk flow as a function of the distance. The method was applied to the Cosmicflows-4 (CF4) dataset after being tested on a simulated mock data. From the analysis of the radial flow in a sphere of 160 Mpc/h, augmented by a simple assumption on the shape of the velocity field on large scales, we provided an estimate of the Hubble parameter H_0 .

The application of our method to a series of 64 realistic mock observations of the CF4 catalog yielded two results. First, we quantified the selection effects in space and magnitude on the radial and bulk flow. Although no systematic shift appears, the radial flow and the Cartesian components of the bulk flow are amplified with respect to the underlying signal, as previously discussed in Andersen et al. (2016). This results in an overestimated

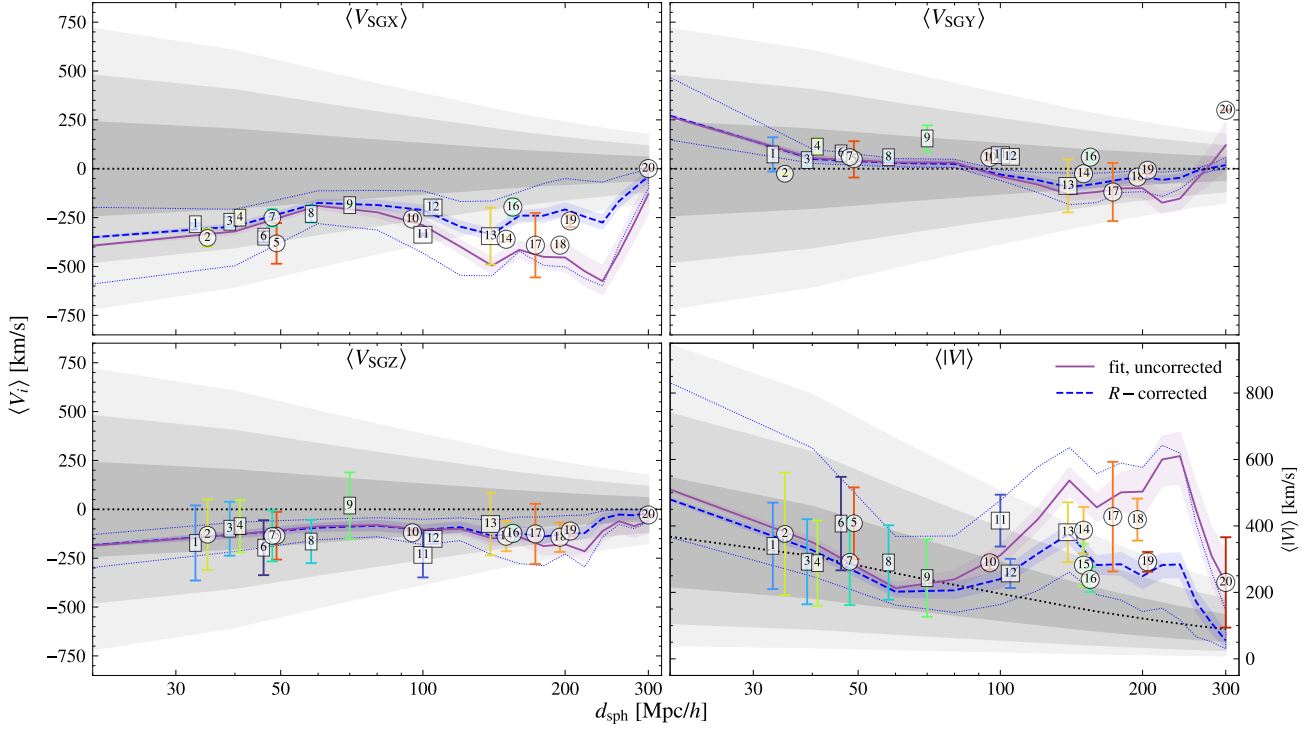


Fig. 9. Bulk flow of the CF4 catalog, given $H_0 = 75.9$ km/s/Mpc. The purple solid line shows the direct application of our method with its 1σ uncertainty band, and the dashed blue line shows the R -corrected results presented in Section 4.3. The color shaded region corresponds to the uncertainty stemming from the forward model while the dotted blue lines delimit the corridor of the uncertainty coming from the R correction described in Eq. (13). Other reported values of the bulk flow are denoted by numbers referring to Table 2, with values derived from the Cosmicflows series shown as circles and those from other surveys shown as squares. The radius for some literature values is approximate for illustrative purposes; the exact values are provided in Table 2.

amplitude of the bulk flow, with a qualitative degradation of the reconstructed amplitude beyond 140 Mpc/h. Second, a simple simulation-based correction was extracted from the application of our method to the 64 mock catalogs. This approach enabled a corrected estimation of the radial and bulk flow of the CF4 data at all distances up to the edge of the data.

The uncorrected bulk flow differs significantly from Λ CDM beyond 140 Mpc/h in the $-SGX$ direction and on the magnitude of the bulk flow. The tension is extreme: the bulk flow measured naively rules out Λ CDM. The increase in the bulk flow around 100–200 Mpc/h has been consistently described in the literature, with various amplitudes. Beyond this distance, the amplitude predicted by the uncorrected approach stagnates, while the one predicted by the corrected estimator decreases slowly. We note that our uncorrected estimation of the bulk flow of CF4 is consistent with that of Whitford et al. (2023) and Watkins et al. (2023), which might indicate that their methods are also sensitive to spurious amplification issues. However, our studies of the effect of the CF4 survey footprint (more specifically, the angular and radial selection function) have demonstrated that the method overestimates the true velocity field. When this is taken into account, the results are more consistent with Λ CDM. This does not need to be the case: our Λ CDM based correction does not force agreement between the method and the model. It simply indicates that if we lived in a Λ CDM model, the bulk flow estimated by our forward model would overestimate the true one by a quantifiable amount.

The radial flow in spheres is degenerate with the value of H_0 chosen when running the fitting procedure. Instead of exploring the tension with Λ CDM using the H_0 provided by Tully et al.

(2023), we chose to estimate H_0 in the light of the CF4 data in a probabilistic way, assigning each value of H_0 a probability based on the likelihood of that corresponding mean inflow in Λ CDM. This analysis resulted in a value of $H_0 = 75.9 \pm 1$ km/s/Mpc, which is at the higher end of the range of local values of H_0 found by other authors. We note that systematic errors on the value should be treated more carefully to obtain a more realistic uncertainty range. We also highlight that this estimate differs from a crude regression of the $cz = H_0 d$ on the data, as in Tully et al. (2023). In the presence of an anisotropic dataset, the radial and bulk flow are correlated, and the bulk flow thus biases the estimation of H_0 . Our procedure extracts the radial component from the signal, resulting in a value of H_0 that is not polluted by the directional bulk flow.

Throughout this work, we tried to control the effect of the assumed cosmology on the reported results.

- To reconstruct the radial and bulk flows, an isotropic H_0 was assumed, as well as the existence of a velocity field. No prior on the power spectrum of this field was assumed, however, as opposed to previous works (e.g., Courtois et al. 2023; Valade et al. 2024; Hoffman et al. 2024).
- To derive the distance-redshift relation in Eq. (5), a flat Λ CDM Universe was considered. Any cosmology assuming an FLRW metric (e.g., Robertson 1933) yields a similar relation, however.
- Our series of synthetic observations of the CF4 catalog and the resulting correction were based on Λ CDM simulations. The departure of the reconstructed velocity profile with respect to the ground truth is essentially due to the selection function of the data and is thus a priori unrelated to the

Table 2. Literature values of bulk flow from different methods and data, grouped for every 50 Mpc/h interval by horizontal lines.

No.	Authors	d_{sph} [Mpc/h]	B_{SGX} [km/s]	B_{SGY} [km/s]	B_{SGZ} [km/s]	$ B $ [km/s]
1	Ma et al. (2011)	33	-283.9	73.2	-172.2	340
2	Qin et al. (2021)	35	-353.7	-25.7	-129.9	376
3	Hong et al. (2014)	40	-271.4	44.3	-99.2	292.3
4	Qin et al. (2018)	40	-250.4	113	-86	288
5	Whitford et al. (2023)	49	-382	48	-135	408
6	Watkins et al. (2009)	50	-347.7	78.7	-196.3	407
7	Watkins & Feldman (2015)	50	-251	60	-136.3	292
8	Ma & Pan (2014)	58	-231.2	58.9	-164.8	290
9	Scrimgeour et al. (2016)	70	-186.1	155	19.2	243
10	Courtois et al. (2025)	100	-257	58	-120	289
11	Feldman et al. (2010)	100	-337.4	69.6	-233.2	416
12	Nusser & Davis (2011)	100	-198	61	-151	257
13	Howlett et al. (2022)	139	-345	-88	-75	381
14	Watkins et al. (2023)	150	-360.8	-25.5	-137.6	387
15	Peery et al. (2018)	150				282
16	Hoffman et al. (2015)	150	-196	59	-123	239
17	Whitford et al. (2023)	173	-391	-119	-126	428
18	Watkins et al. (2023)	200	-391	-43	-143	419
19	Hoffman et al. (2024)	200	-267	-8	-111	292
20	Courtois et al. (2023)	300	1	299	-29	230

Notes. Entry numbers refers to Fig. 9.

simulation cosmology. A series of tests that we do not present here demonstrated, however, that the correction scheme depends on cosmological parameters such as $f\sigma_8$. We would like to highlight that assuming Λ CDM in the correction model is consistent with our aim to test Λ CDM as a cosmological model. Furthermore, the correction works in favor of Λ CDM and might only dampen the tension in the data.

- In order to give an estimate of H_0 , we assumed the cosmological principle, i.e., that the radial velocities of spheres of increasing radii asymptotically converge to zero. Furthermore, a single-parameter scatter on the radial flow of a sphere of 160 Mpc/h was introduced to model the local cosmic variance. The numerical value of this scatter was extracted from our Λ CDM simulations. We highlight, however, that (1) Λ CDM agrees with linear theory on these scales, making our chosen value more general than Λ CDM; (2) that this scatter only affects the uncertainty on H_0 and not its value, and (3) that this result can easily be extended to any other cosmology, given that the cosmic variance on the radial velocity of sphere of 160 Mpc/h can be derived.
- It appears that the bulk flow reconstructed by our algorithm is mildly sensitive to the value of H_0 . Considering, for instance, Planck's $H_0 = 67.77$ km/s/Mpc value slightly decreases the magnitude of the bulk flow, but simultaneously induces extreme outflows (on the order of thousands of km/s) at the edge of the data. This only marginally mitigates the tension on the amplitude of the bulk flow while creating a major (unphysical) radial outflow of matter.

This method differs from traditional bulk flow estimators, such as those presented in Whitford et al. (2023), from several perspectives. First, the forward-modeling approach enables the writing of complex, nonlinear, but more realistic observational models (e.g., Boubel et al. 2024), thus bypassing external Gaussianization methods (e.g., Watkins & Feldman 2015; Qin et al. 2021; Hoffman et al. 2021). The fitting of the forward model in

shells that are combined in the postprocessing into spheres constitutes a new weighting technique. While a simple cut in the data at the high-redshift end causes all the profiles to plateau to their values at small distances because the information is concentrated near the observer, this approach enables a better estimation of the evolution of the fitted values with distance. Compared to field-level inference approaches (e.g., Courtois et al. 2023; Hoffman et al. 2024), our method does not rely on the linear theory of fields, nor does it assume a correlation function of the velocity field, which is known to dampen the tension between the data and the underlying model. Last but not least, our approach is similar to that of Boubel et al. (2024, 2025). However, we treated the entire CF4 catalog with a more general, but less realistic, model than did Boubel et al. (2024), who focused on the Tully-Fisher relation and were thus limited to about 10 000 constraints out of the 38 000 of the entire CF4 grouped catalog. While interpreting the dipolar anisotropic expansion as bulk flow, Kalbouneh et al. (2025a) reported similar bulk flow results using the ungrouped CF4 data, complementing the findings of Stiskalek et al. (2026) using the redshift-limited ($z \lesssim 0.05$) CF4 subsamples, where no local H_0 anisotropy was detected, and the results remain consistent with Λ CDM at these shallower scales.

In future work, we suggest improving the model to include possible systematics in the observational data, notably issues of selection function or intercalibration, which have been proven to lead to significant spurious flows (Whitford et al. 2023, Fig. 9). The correction framework could be improved by considering more elaborated methods. We also look forward to more peculiar velocity data, specifically, to data covering the heavens more spherically.

Acknowledgements. CD appreciates the DPST scholarship. AV thanks A. Whitford for sharing the literature data used in Fig. 9. AV has been supported by the Agence Nationale de la Recherche of the French government through the program ANR-21-CE31-0016-03. NIL acknowledges funding from the European Union Horizon Europe research and innovation program (EXCOSM, grant No.101159513)/The MultiDark Database used in this paper and the web

application providing online access to it were constructed as part of the activities of the German Astrophysical Virtual Observatory as a result of a collaboration between the Leibniz-Institute for Astrophysics Potsdam (AIP) and the Spanish MultiDark Consolider Project CSD2009-00064. The Bolshoi and MultiDark simulations were run on the NASA's Pleiades supercomputer at the NASA Ames Research Center. The MultiDark-Planck (MDPL) and the BigMD simulation suite have been performed in the Supermuc supercomputer at LRZ using time granted by PRACE. Plots are made with Python libraries: *Healpy* and *Matplotlib*. HMC module is made available by *TensorFlow Probability*, coupled with *NumPy*, *Scipy*, and *Astropy* for handy mathematical constants and expressions.

References

- Andersen, P., Davis, T. M., & Howlett, C. 2016, *MNRAS*, **463**, 4083
- Boruah, S. S., Hudson, M. J., & Lavaux, G. 2021, *MNRAS*, **507**, 2697
- Boubel, P., Colless, M., Said, K., & Staveley-Smith, L. 2024, *MNRAS*, **533**, 1550
- Boubel, P., Colless, M., Said, K., & Staveley-Smith, L. 2025, *JCAP*, **2025**, 066
- Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, <https://github.com/jax-ml/jax>
- Brout, D., Scolnic, D., Popovic, B., et al. 2022, *ApJ*, **938**, 110
- Carreres, B., Bautista, J. E., Feinstein, F., et al. 2023, *A&A*, **674**, A197
- Courtois, H. M., Pomarède, D., Tully, R. B., Hoffman, Y., & Courtois, D. 2013, *AJ*, **146**, 69
- Courtois, H. M., Dupuy, A., Guinet, D., et al. 2023, *A&A*, **670**, L15
- Courtois, H. M., Mould, J., Hollinger, A. M., Dupuy, A., & Zhang, C. P. 2025, *A&A*, **701**, A187
- de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, *ApJ*, **302**, L1
- Dekel, A., Bertschinger, E., Yahil, A., et al. 1993, *ApJ*, **412**, 1
- Di Valentino, E., Levi Said, J., Riess, A., et al. 2025, *Phys. Dark Univ.*, **49**, 101965
- Dillon, J. V., Langmore, I., Tran, D., et al. 2017, arxiv e-print [arXiv:1711.10604]
- Djorgovski, S., & Davis, M. 1987, *ApJ*, **313**, 59
- Doumler, T., Gottlöber, S., Hoffman, Y., & Courtois, H. 2013, *MNRAS*, **430**, 912
- Dressler, A., Lynden-Bell, D., Burstein, D., et al. 1987, *ApJ*, **313**, 42
- Feldman, H. A., Watkins, R., & Hudson, M. J. 2010, *MNRAS*, **407**, 2328
- Giani, L., Howlett, C., Said, K., Davis, T., & Vagnozzi, S. 2024, *JCAP*, **2024**, 071
- Gott, J. R., III., Jurić, M., Schlegel, D., et al. 2005, *ApJ*, **624**, 463
- Graziani, R., Courtois, H. M., Lavaux, G., et al. 2019, *MNRAS*, **488**, 5438
- Hoffman, Y., & Ribak, E. 1992, *ApJ*, **384**, 448
- Hoffman, Y., Courtois, H. M., & Tully, R. B. 2015, *MNRAS*, **449**, 4494
- Hoffman, Y., Pomarède, D., Tully, R. B., & Courtois, H. M. 2017, *Nat. Astron.*, **1**
- Hoffman, Y., Nusser, A., Valade, A., Libeskind, N. I., & Tully, R. B. 2021, *MNRAS*, **505**, 3380
- Hoffman, Y., Valade, A., Libeskind, N. I., et al. 2024, *MNRAS*, **527**, 3788
- Hong, T., Springob, C. M., Staveley-Smith, L., et al. 2014, *MNRAS*, **445**, 402
- Howlett, C., Said, K., Lucey, J. R., et al. 2022, *MNRAS*, **515**, 953
- Hubble, E. 1929, *Proc. Nat. Acad. Sci.*, **15**, 168
- Kaiser, N. 1987, *MNRAS*, **227**, 1
- Kaiser, N. 1988, *MNRAS*, **231**, 149
- Kalbouneh, B., Marinoni, C., Maartens, R., et al. 2025a, arXiv e-prints [arXiv:2510.02510]
- Kalbouneh, B., Santiago, J., Marinoni, C., et al. 2025b, *JCAP*, **2025**, 076
- Kourkchi, E., Tully, R. B., Courtois, H. M., Dupuy, A., & Guinet, D. 2022, *MNRAS*, **511**, 6160
- Lavaux, G. 2016, *MNRAS*, **457**, 172
- Lee, M. G., Freedman, W. L., & Madore, B. F. 1993, *ApJ*, **417**, 553
- Ma, Y.-Z., & Pan, J. 2014, *MNRAS*, **437**, 1996
- Ma, Y.-Z., Gordon, C., & Feldman, H. A. 2011, *Phys. Rev. D*, **83**, 103002
- Magoulas, C., Springob, C., Colless, M., et al. 2016, in *The Zeldovich Universe: Genesis and Growth of the Cosmic Web*, eds. R. van de Weygaert, S. Shandarin, E. Saar, & J. Einasto, 308, 336
- Neal, R. 2011, *Handbook of Markov Chain Monte Carlo* (Chapman and Hall/CRC), 113
- Nusser, A., & Davis, M. 2011, *ApJ*, **736**, 93
- Peery, S., Watkins, R., & Feldman, H. A. 2018, *MNRAS*, **481**, 1368
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, **517**, 565
- Planck Collaboration XIII. 2016, *A&A*, **594**, A13
- Qin, F., Howlett, C., Staveley-Smith, L., & Hong, T. 2018, *MNRAS*, **477**, 5150
- Qin, F., Parkinson, D., Howlett, C., & Said, K. 2021, *ApJ*, **922**, 59
- Regos, E., & Szalay, A. S. 1989, *ApJ*, **345**, 627
- Riebe, K., Partl, A. M., Enke, H., et al. 2013, *Astron. Nachr.*, **334**, 691
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, **116**, 1009
- Riess, A. G., Yuan, W., Macri, L. M., et al. 2022, *ApJ*, **934**, L7
- Robertson, H. P. 1933, *Rev. Mod. Phys.*, **5**, 62
- Rosselli, D., Carreres, B., Ravoux, C., et al. 2025, *A&A*, **701**, A119
- Scolnic, D., Brout, D., Carr, A., et al. 2022, *ApJ*, **938**, 113
- Scrimgeour, M. I., Davis, T. M., Blake, C., et al. 2016, *MNRAS*, **455**, 386
- Sorce, J. G., Gottlöber, S., Hoffman, Y., & Yepes, G. 2016, *MNRAS*, **460**, 2015
- Sorce, J. G., Mohayaee, R., Aghanim, N., Dolag, K., & Malavasi, N. 2024, *A&A*, **687**, A85
- Sorrenti, F., Durrer, R., & Kunz, M. 2023, *JCAP*, **2023**, 054
- Stiskalek, R., Desmond, H., & Lavaux, G. 2026, *MNRAS*, **546**, staf2048
- Strauss, M. A., & Willick, J. A. 1995, *Phys. Rep.*, **261**, 271
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, **688**, 709
- Tully, R. B., & Courtois, H. M. 2012, *ApJ*, **749**, 78
- Tully, R. B., & Fisher, J. R. 1977, *A&A*, **500**, 105
- Tully, R. B., Courtois, H., Hoffman, Y., & Pomarède, D. 2014, *Nature*, **513**, 71
- Tully, R. B., Courtois, H. M., & Sorce, J. G. 2016, *AJ*, **152**, 50
- Tully, R. B., Kourkchi, E., Courtois, H. M., et al. 2023, *ApJ*, **944**, 94
- Valade, A., Hoffman, Y., Libeskind, N. I., & Graziani, R. 2022, *MNRAS*, **513**, 5148
- Valade, A., Libeskind, N. I., Hoffman, Y., & Pfeifer, S. 2023, *MNRAS*, **519**, 2981
- Valade, A., Libeskind, N. I., Pomarède, D., et al. 2024, *Nat. Astron.*, **8**, 1610
- Watkins, R., & Feldman, H. A. 2015, *MNRAS*, **450**, 1868
- Watkins, R., Feldman, H. A., & Hudson, M. J. 2009, *MNRAS*, **392**, 743
- Watkins, R., Allen, T., Bradford, C. J., et al. 2023, *MNRAS*, **524**, 1885
- Whitford, A. M., Howlett, C., & Davis, T. M. 2023, *MNRAS*, **526**, 3051
- Zaroubi, S., Hoffman, Y., Fisher, K. B., & Lahav, O. 1995, *ApJ*, **449**, 446

Appendix A: Modeling the magnitude-limited mocks

The mocks presented in this work are based on halos issued from a dark matter-only simulation. In order to properly model the magnitude cut, the halos have to be affected a luminosity. This is done by introducing a simplistic mass-luminosity relationship

$$\frac{L}{L_0} = \alpha \left(\frac{M}{M_0} \right)^\beta \quad (\text{A.1})$$

where L is the constructed luminosity, M is the tabulated mass of the halos (using the 200c convention) and $L_0 = 10^{10}$ and $M_0 = 10^{12}$ are normalizing constants. Given the halo mass function $P(\log(M))$ predicted by Tinker et al. (2008), the expected number of halos as function of distance in the presence of a magnitude cut at m_{cut} is given by

$$P^{\text{mod}}(r) = r^2 \times \delta^K(r < r_{\text{cut}}) \times \int_{-\infty}^{m_{\text{cut}}} P(\log(M) = \beta^{-1}(m - 5 \log(r) - 25)) dm \quad (\text{A.2})$$

where δ^K is the Kronecker delta function and $\delta^K(r < r_{\text{cut}})$ is a crude modeling of the cut in redshift space present in the 6dF and SDSS catalogs. Note that we have set $\alpha = 1$ as this parameter is perfectly degenerated with the parameter m_{cut} .

The parameters β and m_{cut} are fitted to each one of the major subcomponents of the CF4 catalog. The optimized quantity is

$$\chi^2 = \sum_i (P^{\text{obs}}(r_i) - P^{\text{mod}}(r_i))^2 \quad (\text{A.3})$$

where $r_i = i \times \Delta r$ with $\Delta r = 5 \text{ Mpc}/h$ and $P^{\text{obs}}(r)$ is the histogram of the CF4 redshift-derived distances.

The results of the fitting procedure are presented in table 1 and in fig. 2 (right panel). The mass-distance relationship, displaying the simulated magnitude cut, is presented in fig. A.1 in the appendix. A visual evaluation of the distance histogram shows a good convergence of the approach. The variations in the CF4 observations appear in agreement with the cosmic variance over our 64 mocks, represented by the shaded corridor around the mean histogram. Even though the simplicity of the model prevents a physical interpretation for the fitted values of m_{cut} and β , it is noted that all the subcatalogs first display a volume-limited regime before being limited in magnitude, at 60 Mpc/h for the other component, 90 Mpc/h for 6dF and 230 Mpc/h for SDSS.

Appendix B: Effect of Ω_M

Equation (5) introduces a dependence of our model on H_0 and Ω_M . As demonstrated in fig. B.1, the effect of Ω_M is of second order compared to that of H_0 . A strong variation of Ω_M ($\pm 3\sigma$ the uncertainty of Brout et al. (2022)) results in a variation of less than 100 km/s in the radial flow at 240 Mpc/h, comparable to a variation of H_0 of about $\approx 0.4 \text{ km/s/Mpc}$. In future works, this should be added in the systematic errors budget on H_0 .

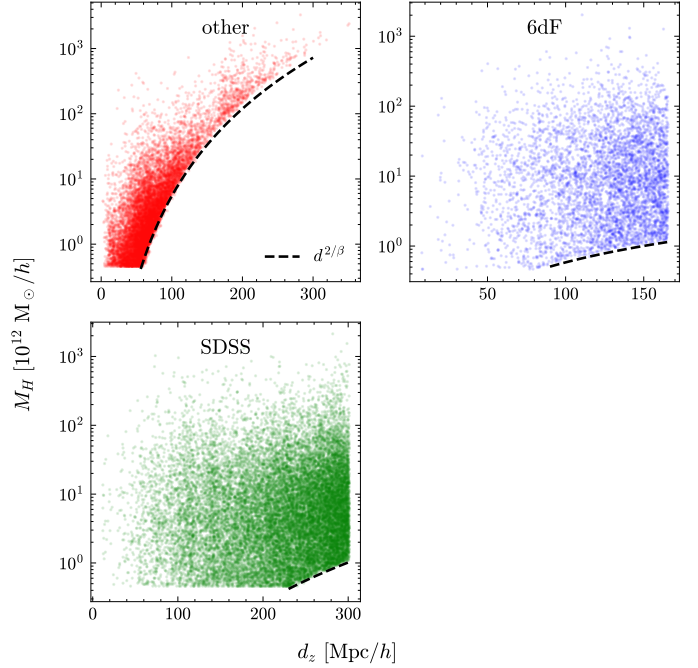


Fig. A.1. An example of the halo mass as a function of redshift distance of the selected halo in the mock, given the criteria in Table 1. The minimum halo mass cut (equivalent to the minimum magnitude cut) is in a black dashed line.

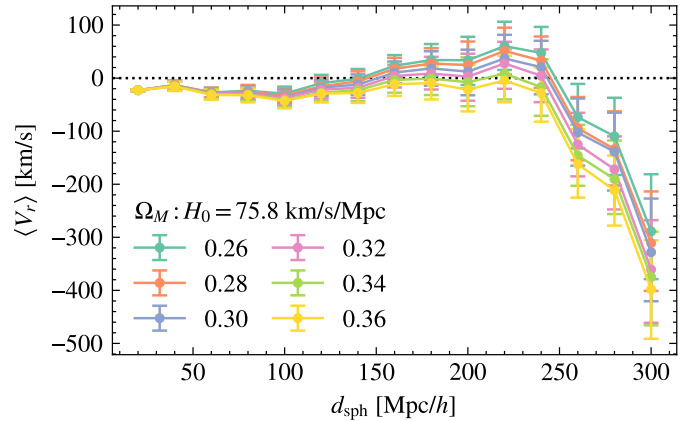


Fig. B.1. Radial flow fit of CF4 data with different matter density parameter, given the same H_0 .