Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527,** 3788–3805 (2024) https://doi.org/10.1093/mnras/stad3433
Advance Access publication 2023 November 2

The large-scale velocity field from the Cosmicflows-4 data

Yehuda Hoffman,[1★] Aurelien Valade,[2] Noam I. Libeskind,[2] Jenny G. Sorce ,[2,3,4] R. Brent Tully,[5] Simon Pfeifer,[2] Stefan Gottlöber[2] and Daniel Pomarède[6]

[1]*Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel*
[2]*Leibniz Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*
[3]*Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France*
[4]*Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, F-91405 Orsay, France*
[5]*Institute for Astronomy, University of Hawaii, Honolulu, HI 96822, USA*
[6]*Institut de Recherche sur les Lois Fondamentales de l'Univers, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*

Accepted 2023 October 30. Received 2023 October 26; in original form 2023 July 24

ABSTRACT

The reconstruction of the large-scale velocity field from the grouped Cosmicflows-4 (CF4) database is presented. The lognormal bias of the inferred distances and velocities data is corrected by the Bias Gaussianization correction scheme, and the linear density and velocity fields are reconstructed by means of the Wiener filter and constrained realizations (CRs) algorithm. These tools are tested against a suite of random and constrained Cosmicflows-3-like mock data. The CF4 data consist of three main subsamples – the 6dFGS and the SDSS data – and the 'others'. The individual contributions of the subsamples have been studied. The quantitative analysis of the velocity field is done mostly by the mean overdensity ($\Delta_L(R)$) and the bulk velocity ($V_{\rm bulk}(R)$) profiles of the velocity field out to 300 h^{-1} Mpc.

The $V_{\rm bulk}(R)$ and $\Delta_{\rm L}(R)$ profiles of the CF4 data without its 6dFGS component are consistent with the cosmic variance to within 1σ. The 6dFGS sample dominates the $V_{\rm bulk}$ ($\Delta_{\rm L}$) profile beyond $\sim 120\,h^{-1}$ Mpc, and drives it to roughly a 3.4σ (-1.9σ) excess (deficiency) relative to the cosmic variance at $R\sim 250\,(190)\ h^{-1}$ Mpc. The excess in the amplitude of $V_{\rm bulk}$ is dominated by its Supergalactic X component, roughly in the direction of the Shapley Concentration. The amplitude and alignment of the inferred velocity field from the CF4 data are at $\sim (2{-}3)\,\sigma$ discrepancy with respect to the Lambda cold dark matter model. Namely, it is somewhat atypical but yet there is no compelling tension with the model.

Key words: methods: data analysis – techniques: radial velocities – large-scale structure of Universe.

1 INTRODUCTION

In the standard model of cosmology, departures from uniform density and from a pure Hubble flow are strongly coupled – density irregularities induce peculiar velocities on top of the Hubble flow; peculiar velocities drive the matter away from uniform distribution. The equation of continuity tells it all (Peebles 1980; Weinberg 2008). This is why surveys of peculiar velocities of galaxies play such an important role in unveiling the underlying – luminous and dark – mass distribution in the nearby universe (Lilje, Yahil & Jones 1986; Lynden-Bell et al. 1988; Dekel, Bertschinger & Faber 1990; da Costa et al. 1996; Masters et al. 2006, is only a partial list). Peculiar velocity surveys have also been used to constrain cosmological parameters (e.g. Dekel et al. 1993; Zaroubi et al. 2001; Nusser & Davis 2011; Nusser 2017; Peery, Watkins & Feldman 2018). In fact, local surveys of peculiar velocities – extending out to a cosmological redshift of ~0.1 – are the only tracers that map the local (total) mass distribution directly. Yet, velocity surveys due to their large errors and sparse sampling are less effective in constraining the values of cosmological parameters compared with other probes – cosmic microwave background (CMB) anisotropies in particular (e.g. Planck Collaboration XVI 2014).

Of particular interest is the Cosmicflows project[1] of measuring and compiling distances and redshifts of galaxies, and thereby estimating their peculiar velocities. Four generations of data have been released so far: Cosmicflows-1 (Tully et al. 2008), Cosmicflows-2 (CF2; Tully et al. 2013), Cosmicflows-3 (CF3; Tully, Courtois & Sorce 2016), and Cosmicflow-4 (CF4; Tully et al. 2023). The Constrained Local UniversE Simulation's (CLUES) collaboration[2] primary focus is on the reconstruction of the present epoch density and velocity fields (e.g. Courtois et al. 2012, 2013; Tully et al. 2014; Hoffman et al. 2017; Pomarède et al. 2017, 2020) and on setting initial conditions for constrained simulations of the local universe (e.g. Hoffman et al. 2008, 2018; Forero-Romero et al. 2011; Sorce et al. 2014, 2016, 2021, 2023; Yepes, Gottlöber & Hoffman 2014; Sorce 2018; Libeskind et al. 2020; Ocvirk et al. 2020; Dolag et al. 2023; Pfeifer et al. 2023) from the Cosmicflows data.

★ E-mail: hoffman@huji.ac.il

[1] https://www.ip2i.in2p3.fr/projet/cosmicflows
[2] https://www.clues-project.org/cms/

Theorists do like peculiar velocities – their emergence in the standard cosmological model, the Lambda cold dark matter (ΛCDM) model, is well understood and in the linear regime the velocity and the density fields are related by a simple linear relation. This stands in sharp contrast to the difficulties arising in estimating velocities from observations. Velocity surveys are actually galaxy distance moduli and redshift surveys, from which the distances and peculiar radial velocities of the galaxies are derived.

The Cosmicflows data exemplify the complexities and intricacies of velocity surveys in general. The data is not homogeneously nor isotropically sampled. It is not sampled in a rigorous manner, namely it is not subjected to a given selection function. It is made of various subsamples of data, assembled by different observational groups, based on different methods and applying these to different type of galaxies (see Tully et al. 2023, for a detailed description of the composition of the CF4 data). Velocity surveys are sparse, very inhomogenous covering one part of the sky within a given redshift range and another part in another range, and in particular are very noisy. For the majority of Cosmicflows data, the typical distance error is of the order of 20 per cent. At a distance of, say, $100\,h^{-1}$ Mpc [where h is Hubble's constant (H_0) measured in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$], such an error in distance implies a $2000\ \mathrm{km\,s^{-1}}$ error in velocity. The standard model of cosmology – the ΛCDM model – predicts that the root mean square value of the radial velocities of galaxies is somewhat smaller than $400\ \mathrm{km\,s^{-1}}$. Namely, the signal-to-noise ratio of galaxies $100\,h^{-1}$ Mpc away is smaller than 20 per cent.

The analysis of such a database, and in particular the reconstruction of the entirety of the large-scale structure (LSS), i.e. the density and three-dimensional velocity fields on a regular grid, out to distances of hundreds of megaparsecs away is challenging. Yet, Bayesian algorithms, relying on the long-range correlations of velocities and assuming the ΛCDM model, have stood up to the challenge and have reconstructed the LSS remarkably well. This is the case of the linear reconstruction by means of the Wiener filter (WF) and constrained realizations (CRs) of Gaussian random fields (WF/CRs; Hoffman & Ribak 1991; Zaroubi et al. 1995; Zaroubi, Hoffman & Dekel 1999), and by Markov chain Monte Carlo (MCMC; Lavaux 2016; Graziani et al. 2019; Courtois et al. 2023) and by Hamiltonian Monte Carlo (Boruah, Lavaux & Hudson 2022; Valade et al. 2022, 2023; Prideaux-Ghee et al. 2023) methods. Hong et al. (2021) presented a new approach to the problem, applying a deep supervised machine-learning algorithm to the CF3 data to reconstruct the very local LSS. The method is still at its infancy stage but it will certainly play a key role in future studies.

Surveys of peculiar velocities like Cosmicflows pose further challenges to their analysis. Peculiar velocities are not directly observed but are rather inferred physical variables and as such inherit their errors from the distance and the (much smaller) redshift errors. The estimation of extragalactic distances suffers from a variety of Malmquist-like biases (see Strauss & Willick 1995, for a detailed discussion). In the Cosmicflows data, and other similar data bases, the distance errors are derived from the normally distributed errors on the distance moduli. It follows that the distance errors are lognormal distributed. In the context of the WF/CRs reconstruction, the bias in the distance-velocity distribution stems from that so-called lognormal bias (Hoffman et al. 2021).

The paper starts with a brief description of the CF4 grouped data (Section 2), followed by a review of the tools of reconstruction of the LSS from the Cosmicflows data (Section 3). A description of the tools of the analysis of the recovered LSS is presented in Section 4. A detailed description and analysis of the WF/CRs reconstructed nearby LSS is described in Section 5. A comparison with other recent reconstructions of the nearby LSS is given in Section 6 and a final summary and discussion are presented in Section 7. Appendices A and B describe the random and constrained mock CF3-like data, respectively, and their analysis. Appendix C presents a comparison of the present results with the WF/CRs application to the CF2 data.

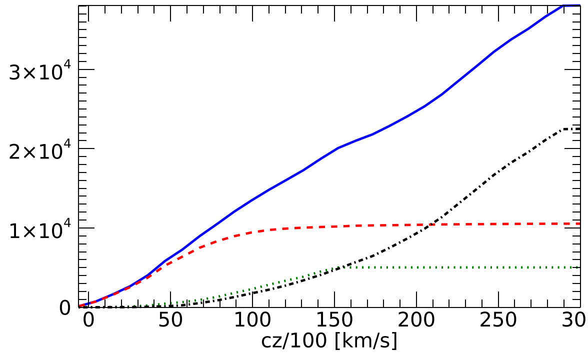

Figure 1. The cumulative number of data points within a given redshift is plotted against $cz/100$. The curves correspond to all the data points (blue, solid line), the 6dFGS data (green, dotted), the SDSS galaxies (black, dot–dashed line) and all the 'others' namely non-6dFGS and non-SDSS (red, dashed line).

2 COSMICFLOWS-4 DATA

A detailed description of the Cosmicflows database, and in particular the Cosmicflows-4 (CF4) data, is presented in Tully et al. (2023).[3] The CF4 data consist of distance moduli and redshifts of roughly 56 000 galaxies, gathered into $\sim$ 38 000 groups. Eight different distance measurement methodologies have been employed – the largest numbers coming from the correlation between the photometric and kinematic properties of spiral galaxies (Tully–Fisher; TF) and elliptical galaxies (fundamental plane). The CF4 data consist of three major sub-samples: the 6dFGS, the SDSS, and the 'others' samples (see Tully et al. 2023, for details). The redshift distributions of these samples are shown in Fig. 1 and their angular distribution in Fig. 2.

The reconstruction of the LSS of our local patch of the Universe from the CF4 data is challenging. The three main sub-samples have different radial and angular distributions, with a significant anisotropy that varies with depth. The 'others', 6dFGS and SDSS components range mostly within the redshift intervals of $cz/100 \sim$ [0–120], $\sim$ [60–160], and $\sim$ [60–300] $\mathrm{km\,s^{-1}}$, respectively. The 'others' are distributed roughly isotropically, outside the Zone of Avoidance, compared with the 6dFGS that is distributed mostly in the Southern Galactic hemisphere and the SDSS that lies entirely in the Northern Galactic hemisphere. Fig. 3 shows the redshift distribution of the velocity errors, normalized by $c\,z$ (where c is the speed of light and z is the redshift). For the vast majority of the data points, the typical fractional error is $\sim$ 20 per cent. It follows that the typical uncertainties in the CF4 inferred velocities amount to $\sim 2000\ \mathrm{km\,s^{-1}}$ at a redshift of $c\,z = 10\,000\ \mathrm{km\,s^{-1}}$ and that for the ΛCDM model the typical signal-to-noise ratio of inferred velocities is 0.15 at that redshift and about 0.05 at the edge of the data ($cz/100 = 300\ \mathrm{km\,s^{-1}}$).

---

[3] The CF4 data used here is the 2023 May 17 version.

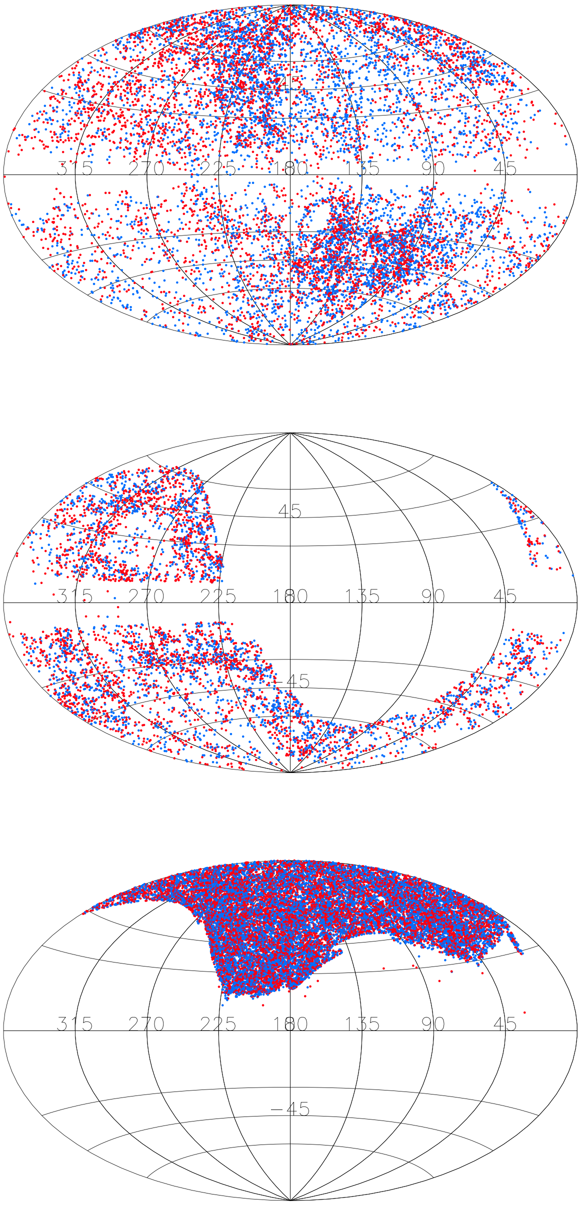

Figure 2. Aitoff projection in Galactic (l, b) coordinates of the distribution of the sub-samples of the CF4 data: SDSS (bottom panel), 6dFGS (middle panel), and all the others (upper panel). Positive peculiar velocities are marked in red and negative ones in blue.

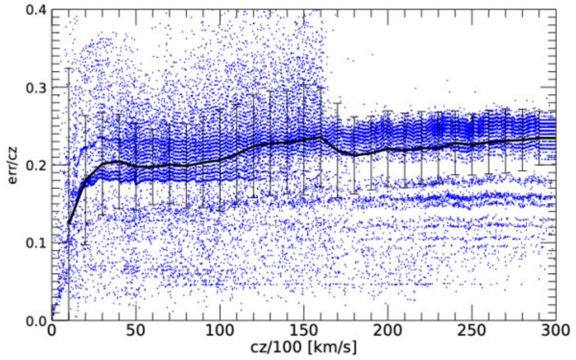

Figure 3. A scatter plot of the velocity errors, normalized by $c\,z$, against their redshift distance ($c\,z$ / 100). The solid line and the error bars show the mean and standard deviation of the normalized errors.

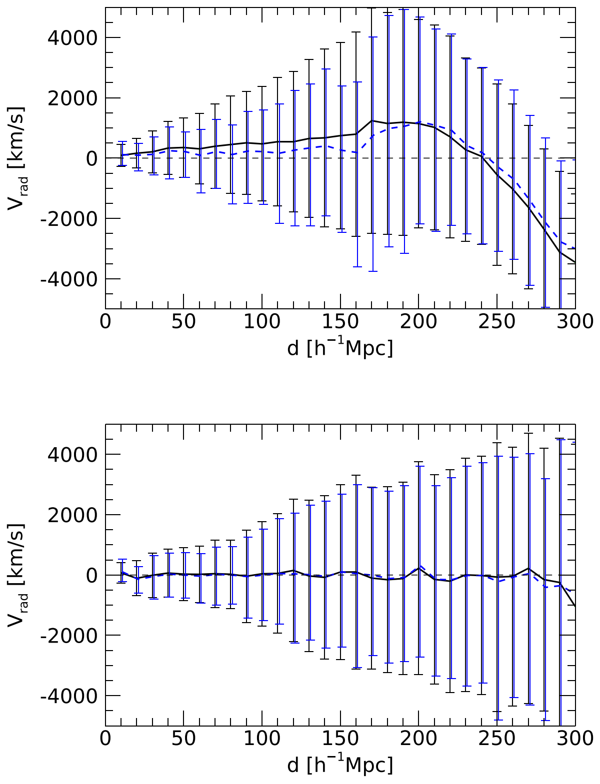

Figure 4. CF4 data: mean (black, solid line) and median (blue, dashed line) of the peculiar radial velocities versus their distances. Black error bars correspond to one standard deviation around the mean value and the blue ones to the first and third quartiles around the median, rescaled to correspond to the standard deviation of a normal distribution. The blue error bars are slightly shifted horizontally. The upper panel presents the uncorrected values of the radial velocities and distances, as derived from the measure distances moduli and redshifts. The lower panel presents the BGc corrected distances and velocities.

The lognormal bias is clearly manifested by the upper panel of Fig. 4, which presents the mean and the median of the distribution of the CF4 uncorrected velocities versus their uncorrected distances. It consists of an excess of positive peculiar velocities, namely an outflow, out to roughly $200\,h^{-1}$ Mpc that is followed by a strong inflow all the way to the edge of the data – in a strong disagreement with the ΛCDM model.

3 TOOLS OF RECONSTRUCTION

3.1 WF and CRs: reconstruction

In the framework of the standard model of cosmology, the LSS of the Universe has emerged out of primordial Gaussian random perturbations field. The ΛCDM model predicts that small scales have more power than the large ones hence they collapse and virialize first, while the long waves still remain in the linear regime. Hence, grouping the data act as a non-linear filter, where members of a collapsed group, for example, galaxies belonging to a given cluster, are replaced by one single data point whose dynamics are well approximated by the linear theory.

The Bayesian framework employed here is the linear theory of the ΛCDM standard model of cosmology. The algorithm of the WF and CRs is the optimal tool for dealing with data within that framework (Hoffman & Ribak 1991; Zaroubi et al. 1995). The essence of the algorithm is that in the case where the underlying field is assumed to be Gaussian the following linear estimators, the minimal variance solution, the conditional mean field given the data, the most probable field and the maximum entropy, are all identical. The WF provides an analytical expression for that estimator. The CRs constitute random realizations designed to obey the imposed constraints. In the case of 'strong' data, the WF estimators and the CRs follow closely the constraining data. Where the data is 'weak' the null field prevails and the CRs are essentially random realizations.

3.2 Lognormal bias correction

Peculiar velocities of surveys like the Cosmicflows are not directly observed but are rather inferred physical variables from surveys of galaxy distances. The Cosmicflows database, to be specific, is composed of the angular positions, redshifts, and distances of galaxies and their associated uncertainties. The major source of the uncertainties of the derived peculiar radial velocities (and from here on they are referred to as 'velocities') are the distance errors. Additionally, the estimation of extragalactic distances suffers from a variety of Malmquist-like biases (see Strauss & Willick 1995, for a detailed discussion). In the Cosmicflows data, and other similar data bases, the estimated luminosity distance ($d_{\rm L}$) is derived from the 'observed' distance modulus (μ) via

$$\mu = 5\log_{10}\left(\frac{d_{\rm L}}{10\,{\rm pc}}\right). \tag{1}$$

Consequently the normally distributed observational uncertainties on μ are transformed into lognormal errors on $d_{\rm L}$ and thereby also on proper distance and velocities. This so-called lognormal bias is analyzed in detail in Hoffman et al. (2021). In the context of the WF/CRs reconstruction, the lognormal bias leads to a nearby (faraway) spurious outflow (inflow) of the inferred velocities of the data points. It follows that the reconstruction of the LSS from velocity surveys cannot be done without accounting for that bias. Two approaches to the bias correction have been pursued. In the WF/CRs framework, the correction of the bias is done ahead of and independently of the application of the WF/CRs algorithm (Sorce 2015; Hoffman et al. 2021; Sorce, Stoica & Tempel 2023). In the MCMC and HMC approach, the undoing of the bias is done in conjunction with the estimation of the LSS, within a unified Bayesian approach.

The Bias Gaussianization correction (BGc) algorithm (Hoffman et al. 2021) is applied to the one-point probability distribution function of the inferred distances and velocities and is designed to transform these distributions from lognormal to normal ones. The BGc is applied here to the grouped CF4 data and this bias-corrected data are then used as an input for the WF/CRs reconstruction of the LSS of the nearby universe.

3.3 Random and constrained mock Cosmicflows-like data

Given the complex nature of the CF4 data and the approximate nature of the bias correction scheme and the WF/CRs reconstruction it is essential to examine these tools and test them against mock Cosmicflows-like data. The basic logic followed here is to use halos drawn from cosmological DM-only ΛCDM simulations for generating such mock catalogues. The simulations from which the mocks are drawn are the target against which the reconstructed LSS is to be compared and its merits are to be judged. Two kinds of mock data are constructed here – one is of random mock data where the mock data points are drawn from random ΛCDM simulations. Such mock data sets are 'random' in the sense that they are drawn from random realizations of the LSS that are consistent with the ΛCDM model and without imposing any constraints related to the particularities of our local realization of the Universe. The other kind is constrained mock data, where the data is drawn from constrained simulations. Random mock data have been recently used to test the BGc and HMC lognormal bias correction and reconstruction schemes (Hoffman et al. 2021; Valade et al. 2023) One of the main virtues of using random mock data is that it enables the probing of the cosmic variance, i.e. the variance introduced by moving the observer randomly in the universe, together with the errors and sampling variance.

The random mock CF3 data described by Hoffman et al. (2021) are used here. Ten different random observers are drawn from a random ΛCDM simulation, and 10 different errors realizations are constructed for each mock observer, resulting in an ensemble of 100 mock data sets. A detailed presentation and analysis of the monopole and dipole moments of the WF/CRs reconstruction from these mocks is given in Appendix A.

The reconstruction of the LSS from galaxy peculiar velocity surveys is very appealing – as the velocities constitute an unbiased probe of the underlying matter density field. Yet, the fact that galaxies, and their groups, are used as tracers of the velocity may lead to bias. Consider the case of voids in the galaxy distribution. The contribution of underdense regions to the velocity field is as important as the overdense ones (Courtois et al. 2017; Hoffman et al. 2017), and galaxies in voids partake in the flow field. Yet, galaxies are underabundant in voids hence also in galaxy velocities surveys. This can lead to a bias in the reconstruction from such surveys. A way to address such a sampling bias is to construct mock Cosmicflows-like data from constrained simulations of the local universe Doumler et al. (2013a, b, c). Here, we use the constrained CF3 mock data that has been drawn from the constrained DM-only simulation from the data (Sorce et al. 2023). That simulation was constrained by the CF2 data and it recovers all the prominent structures within the nearby $\sim 100\,h^{-1}\,$Mpc around the LG. Detailed testing of the WF/CRs reconstruction from the constrained mock CF3 data is presented in Appendix B.

A few words are due on why the mocks used here are CF3-like while the data is the CF4 catalogue. We wanted to test the algorithm based on random and constrained mocks. So far, only CF3-based constrained simulations are available and by construction they are not deep enough to construct realistic constrained mocks for CF4. Therefore, we decided to test the algorithm with CF3 mocks. Yet, Fig. 11 shows that the SDSS component of the CF4 data, the main component of CF4 which distinguishes it from CF3, hardly affects the monopole and dipole moments of the velocity fields. Hence, it seems very likely that the mock CF3 data are adequate for testing the reconstruction from the CF4 data by applying its methodology to the mock CF3 data.

The main conclusion one may draw from the comparison with the mock is that for the case studied here the WF/CRs estimated profiles faithfully reconstruct the bulk velocity and the mean over-density profiles of the target simulations for spherical volumes of $R \gtrsim 40\,h^{-1}\,$Mpc.

4 TOOLS OF ANALYSIS

4.1 Cosmic and constrained variance

The primordial density and velocity fields are random Gaussian fields. Furthermore, the standard cosmological model dictates that

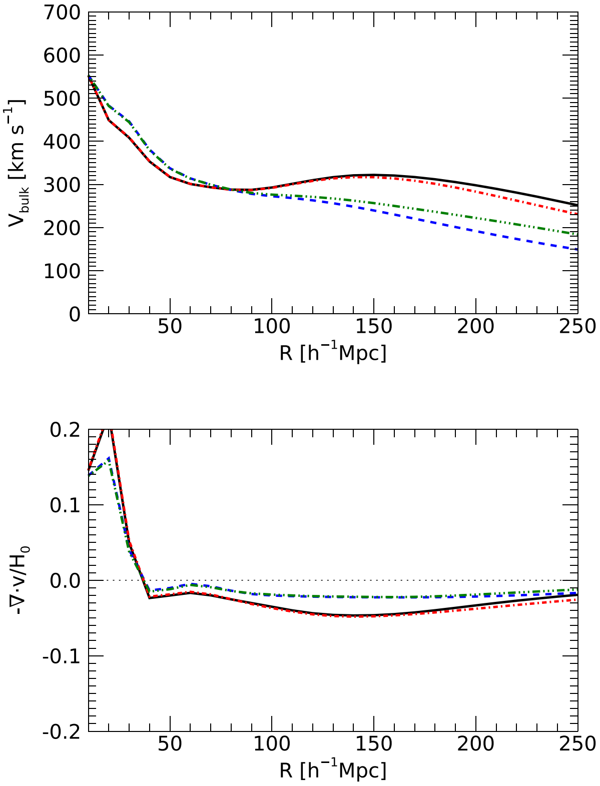

Figure 11. The amplitude of the WF reconstructed bulk velocity (upper panel) and of the monopole moment (lower panel) calculated from the entire CF4 data (black – solid line), the CF2-like data, (i.e. the CF4 without the SDSS and without the 6dFGS data; blue – dashed), the CF3-like data (i.e. CF4 without the SDSS data; red – dot–dash), and the CF4 without the 6dFGS data (green – dot–dot–dot–dash).

these fields are statistically homogeneous and isotropic. The variance of the fields, taken over large enough volume and at a given resolution, has therefore a universal value, also called the cosmic variance. The variance of the possible realizations of the δ and of the velocity fields constrained by the Cosmicflows data, say, is smaller than the cosmic variance, as it samples a sub-volume of all possible realizations of random realizations of the model. This so-call constrained variance depends on the strength and quality of the constraining data and on the nature of the assumed cosmological model. For the Cosmicflows data, one expects the constrained variance to converge to the cosmic variance away from the data zone – in configuration as well as in Fourier space. In the ΛCDM model, the high wavenumber (k) modes are dominated by non-linear dynamics and hence are very poorly constrained by the data, hence one expects that at the high k limit, namely at high resolution, the constrained variance converges to the cosmic one.

4.2 Ensembles of constrained and random realizations

The constrained and cosmological unconstrained variance are calculated here over an ensemble of 60 constrained and 3000 random realizations, respectively. Random realizations, namely unconstrained ones, are much ‘cheaper’ to construct, hence the imbalance in the number of realizations of the different kinds. Two sets of such realizations have been prepared, in boxes of side length of $L =$ 600 and $= 1000\,h^{-1}$ Mpc. In the presentation of the different results, the size of the box is clearly stated. The constrained and random realizations are calculated by means of an FFT algorithm, where periodic boundary conditions are assumed. The practice followed here for accounting for the missing power inherent to such an approach is to perform all the FFT calculations in computational boxes that are 4^3 times larger in volume and then crop the resulting fields to the desired box.

4.3 The linear density field

The WF/CRs methodology provides an estimation of the density and velocity fields within the framework of the linear regime, using the ΛCDM standard model of cosmology as the Bayesian prior model. Within the linear theory the fractional overdensity ($\delta = \rho/\bar{\rho} - 1$), where ρ is the local density and $\bar{\rho}$ is its mean cosmological value) and the velocity fields are related by

$$\delta_{\rm L} = -\frac{1}{H_0 f(\Omega)} \nabla \cdot \boldsymbol{v}, \tag{2}$$

where $f(\Omega)$ is the linear growth factor and the subscript L denotes that it is the linear δ. A comparison of the WF/CRs estimated density field from mock data should be made against $\delta_{\rm L}$ inferred from the velocity field of the target simulation.

4.4 Visualization of the velocity field by streamlines

Streamlines are a graphical visualization of a velocity field. The equation of ‘motion’ of the line element of a given streamline s(l), where l is the line parameter, is:

$$\frac{{\rm d}\boldsymbol{s}}{{\rm d}\boldsymbol{l}} = \boldsymbol{v}\,(\boldsymbol{s}\,(\boldsymbol{l})) \tag{3}$$

The choice of the seeding points of the streamlines is a matter of convention, randomly or uniformly distributed on a grid. Particles move along streamlines at a given moment, yet they do not follow the entirety of given streamlines. The flow field is represented here by colored streamlines whose colour reflects the local amplitude of the velocity and the tangent of the line is in the direction of the velocity vector.

4.5 Multipole expansion

A common presentation of the velocity field is by means of the monopole and dipole moments of the velocity field. Given a velocity field evaluated over a regular grid, these are defined as the volume average over a sphere of radius R of the monopole moment either as

$$\nabla \cdot \boldsymbol{v}(R) = \frac{1}{4\pi R^3/3} \int_{<R} \nabla \cdot \boldsymbol{v}\, {\rm d}^3 r, \tag{4}$$

or as $\Delta_{\rm L}(R) = -\nabla \cdot \boldsymbol{v}(R)/(H_0\, f\,(\Omega))$. For the ΛCDM standard model, $f(\Omega) \sim 0.52$. We use the term ‘monopole’ in a somewhat loose sense – it stands here for the isotropic component of the linear expansion of the velocity field, namely the ‘breathing’ mode of the velocity field. The dipole moment, which is also the bulk velocity of the sphere, is given by

$$\boldsymbol{V}_{\rm bulk}(R) = \frac{1}{4\pi R^3/3} \int_{<R} \boldsymbol{v}\, {\rm d}^3 r. \tag{5}$$

4.6 χ^2 statistics

The Cartesian components of the bulk velocity vector of given spheres of radius Rconstitute a set of correlated Gaussian variables, and as such are best suited for a χ^2 statistics to measure their likelihood given the ΛCDM model. Following Watkins et al. (2023), we consider here the bulk velocity of a given sphere and vary the radius of that sphere. The analysis consists of calculating the $\chi^2(R)$ of the three Cartesian components of the mean of the ensemble of the bulk velocity of an ensemble of CRs, $\boldsymbol{V}^{\rm CRs}_{\rm bulk}(R)$,

$$\chi^2(R) = V^{\rm CRs}_{{\rm bulk},\alpha} < V_{{\rm bulk},\,\alpha} V_{{\rm bulk},\,\beta} >^{-1} V^{\rm CRs}_{{\rm bulk},\beta}, \tag{6}$$

where $< V_{{\rm bulk},\,\alpha} V_{{\rm bulk},\,\beta} >$ is the covariance matrix of the bulk velocity and the angular brackets denote an ensemble average. In principle, this covariance can be calculated analytically, yet given the fact that it corresponds to the velocity field calculated over a finite grid we have chosen to evaluate it numerically by calculating it as an average evaluated over an ensemble of the bulk velocity of 3000 random realizations of the ΛCDM model.

4.7 Tidal decomposition

An inherent feature of the present implementation of the WF/CRs algorithm is the recovery, in principle, of the full (linear) velocity field, i.e. the velocity field induced by all the matter in the whole Universe. Practical considerations limit the evaluation of the velocity field to a given finite computational box, that often encompasses the entire data zone. That recovered velocity field can be decomposed into two components – the one introduced by the matter distribution within the box, or a sub-volume of that box and the one induced by the matter outside that volume. These are the ‘local’ (divergent) and the tidal components, respectively. By construction, the tidal field within that volume is divergence-less. The technique employed here for the tidal decomposition uses the WF reconstructed density and velocity fields within a box (see Zaroubi, Hoffman & Dekel 1999; Hoffman et al. 2001). Fourier/FFT decomposition is applied to the density field within the box, and employing the assumption of the linear theory, the divergent velocity field and the density are related by

$$\boldsymbol{v}^{div}_k = -\frac{1}{H_0 f(\Omega)}\frac{\mathbf{k}}{k^2}\,\delta_k, \tag{7}$$

where $\boldsymbol{v}^{div}_k$ and δ_k are the Fourier-transformed divergent velocity and density fields, respectively. The tidal component is the residual of the full velocity field from the local one. The volume considered here for the tidal decomposition is a sphere of radius $R = 300\,h^{-1}$ Mpc.

5 RECONSTRUCTION FROM CF4 DATA

5.1 Large-scale structure

The main focus of the paper is the quantitative analysis of the LSS reconstructed from the CF4 data, within the framework of the ΛCDM serving as the Bayesian prior. Yet, we start here with a qualitative visual overview of the recovered density and velocity fields (Fig. 5). An insight into the issue of the validation of the reconstructed LSS is provided by Fig. 6, which presents the large-scale WF reconstructed overdensity field, $\delta = (\rho - \bar{\rho})/\bar{\rho}$ (where $\bar{\rho}$ is the mean density of the Universe). The distribution of the LEDA galaxies (Makarov et al. 2014) is overlaid on the overdensity colour/contour map, for the sake of orientation and visual inspection of the quality of the reconstruction. The large-scale velocity field is shown by means of flow lines (Fig. 5).

A detailed analysis of the reconstruction of the LSS of our local patch of the Universe is beyond the scope of the current paper. Yet, we note here that the LSS that emerges here is in very good agreement with the one inferred from the CF2 WF reconstruction (see Appendix C). A series of papers on the linear WF/CRs out of the CF2 data has been published and the interested readers are referred to these (e.g. Tully et al. 2014; Hoffman, Courtois & Tully 2015; Hoffman et al. 2017)

5.2 Cosmic versus constrained variance

Fig. 7 depicts the behaviour of the constrained variance manifested by an ensemble of CRs constrained by the CF4 data. The constrained variance is calculated as the variance of the residual of the CRs from their mean value, namely the WF field. Two different resolutions are used – the case of no smoothing on a grid of $N = 128^3$ and a box of side $L = 600\,h^{-1}$ Mpc (left panel) and Gaussian smoothing with a kernel of $Rs = 10.0\,h^{-1}$ Mpc (right panel). The contour/colour maps present the cosmic variance normalized by the (resolution dependent) cosmic variance. The no smoothing case is dominated by the virtually unconstrained short waves, hence the constrained variance is virtually identical to the cosmic one. For the $Rs = 10.0\,h^{-1}$ Mpc smoothed CRs, one finds an inner region of a radius of ~ 25 (70) h^{-1} Mpc with the constrained variance smaller than 20 (50) per cent of the cosmic variance. It follows that the Virgo cluster (at a distance of $\sim 10\,h^{-1}$ Mpc) is strongly constrained by the data and the Coma cluster is less so.

5.3 Monopole and dipole moments

The monopole and dipole moments, namely $\Delta_{\rm L}(R)$ and $V_{\rm bulk}(R)$, have been calculated for the ensembles of the CRs and of the random realizations of box sized of $L = 1000\,h^{-1}$ Mpc, for which the mean profiles and the scatter (namely standard deviation) around the mean have been calculated. Fig. 8 presents the mean and variance of the constrained and random realizations of the norm of the bulk velocity (upper panel) and its three Supergalactic Cartesian components. Fig. 9 shows the mean and scatter of the ensemble of the constrained and random realizations of the monopole moment. Table 1 shows the mean and scatter of the ensemble of CRs of the alignment of $\mathrm{V}_{\rm bulk}(R)$ with CMB dipole velocity ($\boldsymbol{V}_{CMB} = [-410, 353, -324]$ km s^{-1}), $\mu_{CMB}(R) = \hat{\boldsymbol{V}}_{bulk}(\mathrm{R}) \cdot \hat{\boldsymbol{V}}_{\rm CMB}$. The statistical significance of the estimated alignment, $\mu_{\rm CMB}(R)$ is tested against the alignment calculated for an ensemble of 3000 random realizations. As these realizations do not have a counterpart to $V_{\rm CMB}$, the alignment is calculated between the bulk velocity at a radius R and its value at a minimal radius, taken to represent the zero lag, $\mu_{self}(R) = \hat{\boldsymbol{V}}_{bulk}(R_{\rm min}) \cdot \hat{\boldsymbol{V}}_{bulk}(\mathrm{R})$. Here, $R_{\rm min} = 10\,h^{-1}$ Mpc is chosen. In practice, this is done by constructing 3000 RAN, evaluating the bulk velocity profile for each realization and calculating the resulting $\mu_{\rm self}(R)$ profiles. This is done also for 60 CRs of the actual CF4 data (Fig. 10). Given the distribution of the $\mu_{\rm self}(R)$ profiles of the random realizations the likelihood of a given $\mu_{\rm self}(R)$ profile of a given CR at a given radius R is gauged as follows. The mean and scatter, over the ensemble of CRs, of the fraction of the RAN for which $\mu_{\rm self}(R)$ is smaller than the corresponding value of a given CR is 0.8 ± 0.05, with little variation as a function of R. Namely, the alignment of the bulk velocity with itself at zero lag is consistent with ΛCDM to within roughly $\sim 1.5\sigma$.

The cosmological mean profile of $V_{\rm bulk}(R)$ is a monotonically decreasing function of depth (R). The constrained mean radial profile of $V_{\rm bulk}$ starts very close to CMB dipole velocity, in magnitude and direction, at the first radial bin considered here ($R = 10\,h^{-1}$ Mpc), and then decays out to $R \sim 70\,h^{-1}$ Mpc, follows by a hump that peaks at $R \sim 160\,h^{-1}$ Mpc. Fig. 11 helps one to trace the contribution of the three major sub-samples of the CF4 data to monopole (lower panel) and dipole (upper panel) moments. It shows the moments of the WF

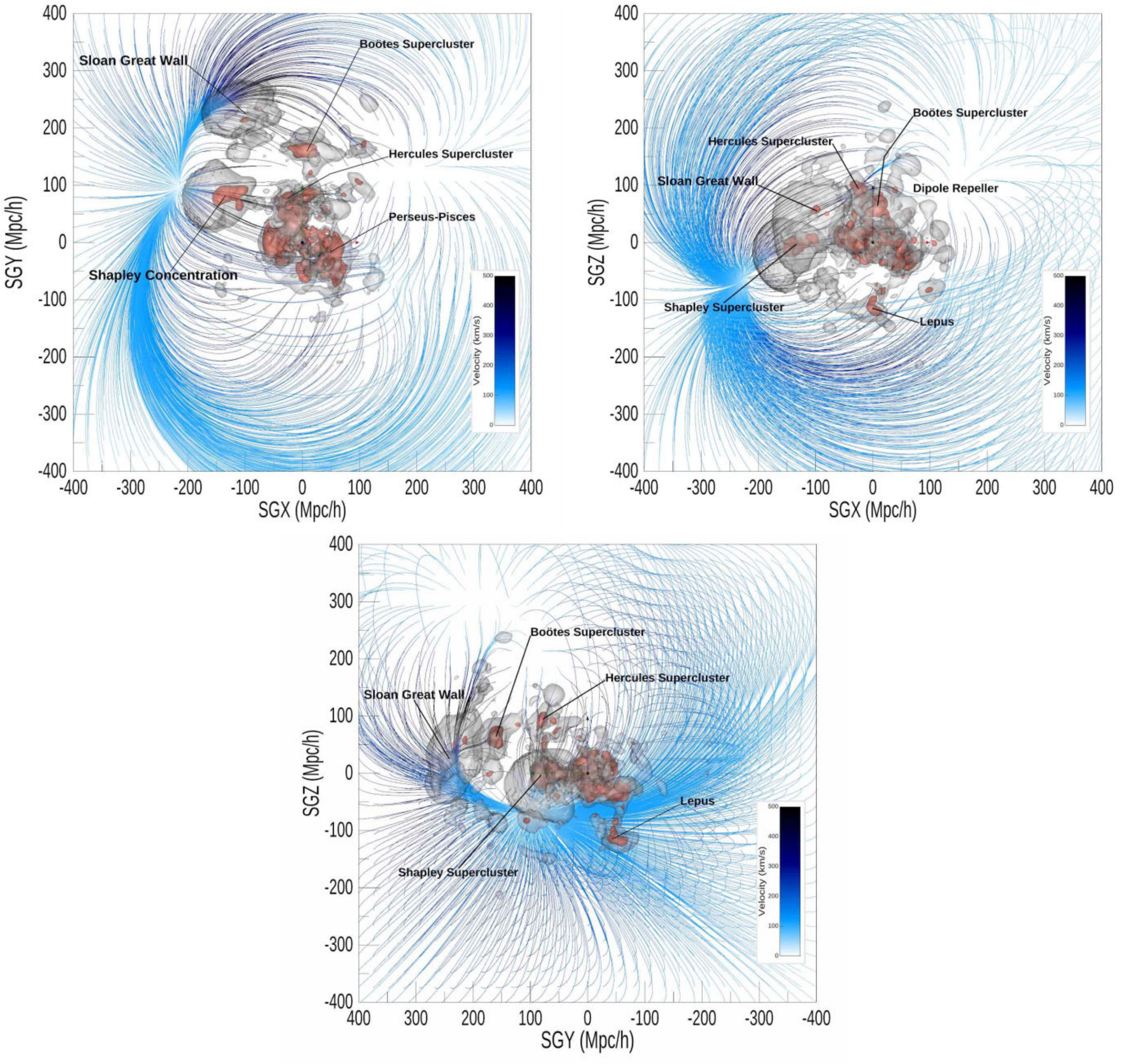

Figure 5. The WF reconstruction of the velocity field from the BGc corrected CF4 data, assuming $H_0 = 74.6\,\mathrm{km\,s^{-1}Mpc^{-1}}$. The field is represented by streamlines, with colours representing the amplitude of velocity field within the plane. Density isosurfaces correspond to $\delta_{\mathrm{L}} = 0.2$ (grey) and 0.5 (red). Labels denote the prominent nearby structures.

reconstructed velocity field from the full CF4 data, the CF4 without the 6dFGS and SDSS components (dubbed here as the 'others'), of the CF4 without the 6dFGS component and the combined 'other' and SDSS data. One should note here that the 'others' data constitutes an updated version of the CF2 data, and the 'others' plus the 6dFGS data are the current equivalent of the CF3 data. The lower panel of Fig. 8 shows that the SGY and SGZ components of V_{bulk} lie within or close to the 1σ of the cosmic variance as opposed to the SGX component – the one that deviates from the 2σ cosmic variance outside of $R \sim 100\,h^{-1}$ Mpc. Fig. 11 depicts the radial profiles of the WF calculated bulk velocity (upper panel) and the monopole moment (lower panel). The plots further decompose the contributions to these moments by the main components of the CF4 data. Fig. 11 clearly shows that the hump in the V_{bulk} profile is contributed by the 6dFGS sub-sample. The CF4 data without the 6dFGS do not show that rise in the V_{bulk} profile. Fig. 8 shows that it is the SGX component of the V_{bulk} that is a responsible for the excess power.

5.4 Dipole velocity: χ^2 statistics

The top panel of Fig. 12 shows $\chi^2(R)$ plotted out to $R = 250\,h^{-1}$ Mpc. The lower panel shows the probability of obtaining a χ^2 equal or larger than the one calculated here, recalling a χ^2 distribution with 3 degrees of freedom (dof). Table 1 presents these numbers for a few cases of R. In particular, we find here at $R = 150\,(200)\,h^{-1}$ Mpc a $\chi^2 = 9.99\,(12.48)$ and $P(> \chi^2) = 1.9 \times 10^{-2}\,(5.9 \times 10^{-3})$. These results are further discussed in Section 6.2.

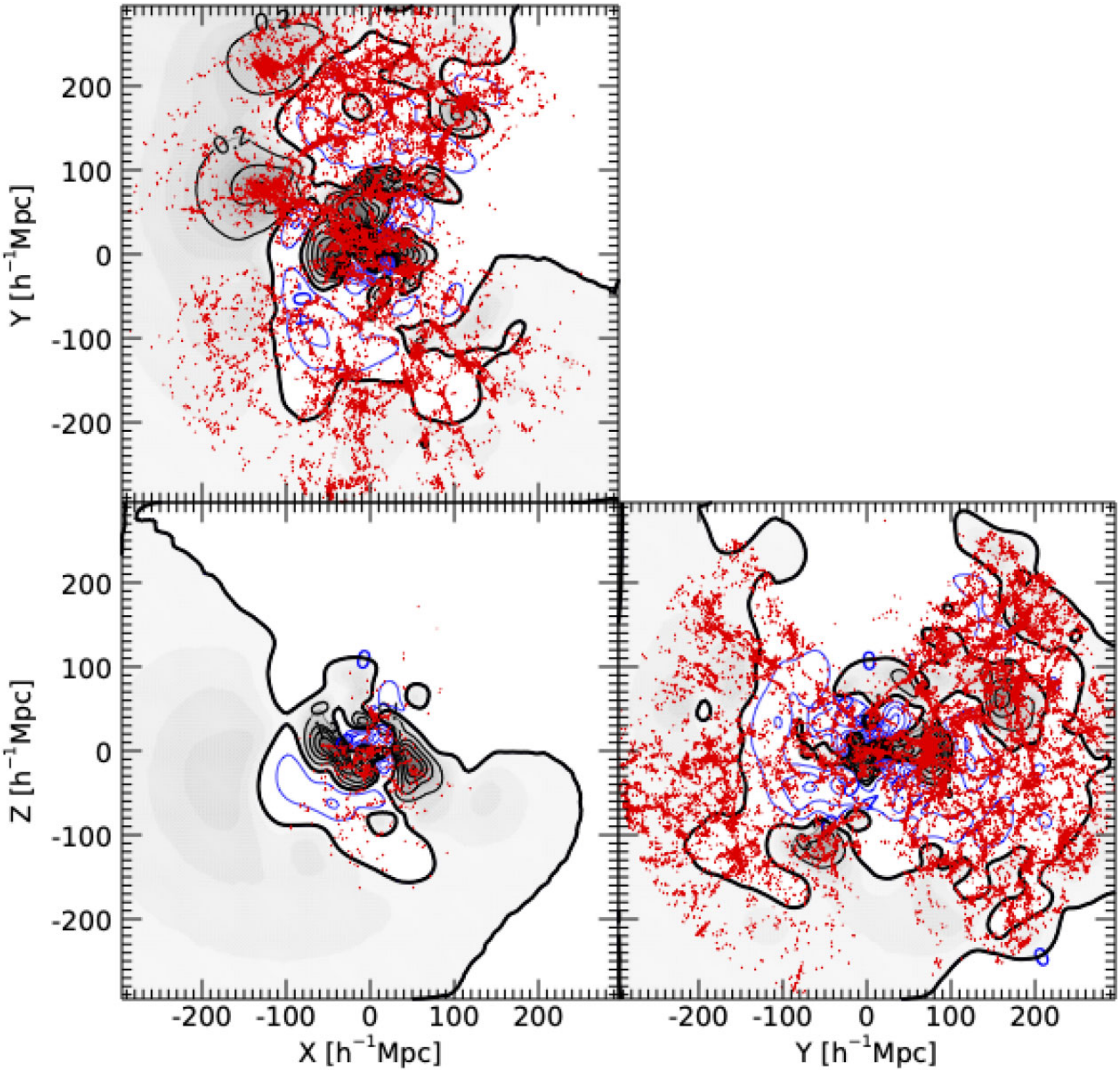

Figure 6. The WF reconstruction of the linear density field ($\delta_{\rm L}$) from the BGc-corrected CF4 data, assuming $H_0 = 74.6\,{\rm km\,s^{-1}Mpc^{-1}}$. The field is represented by means of contour/colour maps on the three principal planes of the Supergalactic Cartesian coordinates. The LEDA galaxies brighter than $M = -20$, within slabs of $\pm\, 5\,h^{-1}$ Mpc on to the principal planes are presented as a gauge of the quality of the reconstruction.

5.5 Dipole velocity: tidal decomposition

Fig. 13 shows the tidal decomposition of the WF reconstructed velocity field from the CF4 data, with respect to a sphere of a radius $R = 300\,h^{-1}$ Mpc. The main feature is that the tidal component of the bulk velocity is essentially constant at $V_{\rm bulk}^{\rm tidal} \sim 129$ km s^{-1}. This means that ~ 20 per cent of the CMB dipole velocity, i.e. the motion of the LG in the CMB frame of reference, is induce by structures in the universe beyond the edge of the CF4 data.

There are two reasons for dwelling here on the issue of the tidal decomposition. One is the issue of the convergence of the bulk flow. Namely, the bulk velocity converges to zero as the radius R increases. For the CF4 data, with its effective depth of $\sim 300\,h^{-1}$ Mpc, and within the ΛCDM model, the tidal residual flow, induced by structures beyond that depth, is roughly 63 km s^{-1}. The other motivation for the performance of such decomposition is for the sake of comparison with other reconstruction methods of the large-scale velocity field.

Commonly, reconstructions of the local flow from galaxy redshift surveys are based on the estimation of the local density, from which the velocity field is recovered by using the linear theory (equation 7). That reconstructed 'local' velocity field needs to be augmented by an estimated tidal component, which is often done by requiring that the sum of the divergent and tidal fields equals the CMB dipole velocity field at the Local Group. Our evaluation of the divergent component enables a direct comparison of the WF reconstructed velocity field with that calculated from the galaxy distribution, and thereby bypassing the need to evaluate a tidal external field. A similar problem is encountered by Bayesian methods applied to peculiar velocities where the target of the reconstruction is the density field in its Fourier representation – such as in the MCMC (Lavaux 2016; Graziani et al. 2019) and the HMC (Boruah, Lavaux & Hudson 2022; Valade et al. 2022, 2023). The periodic boundary conditions employed by the FFT suppresses the tidal component of the velocity field. The comparison of our work with those FFT-

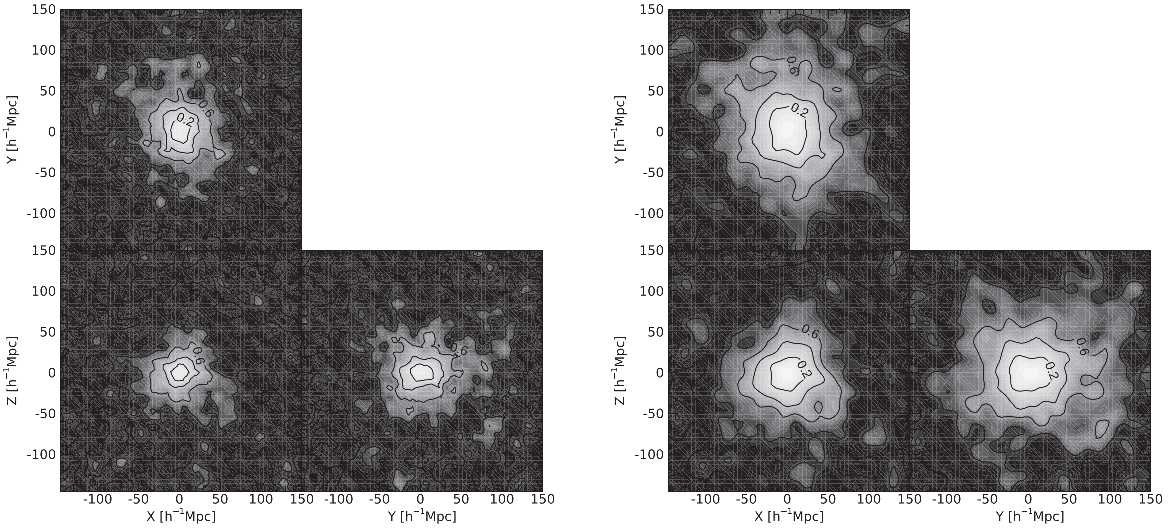

Figure 7. Contour map of the constrained variance, normalized by the cosmic variance, taken over an ensemble of 50 CRs for the un-smoothed density field (left panel) and with Gaussian smoothing with $R_s = 10.0\,h^{-1}$ Mpc (right panel). The CRs are evaluated on a cubic grid of $N = 128^3$ and a box of side length of $L = 600\,h^{-1}$ Mpc.

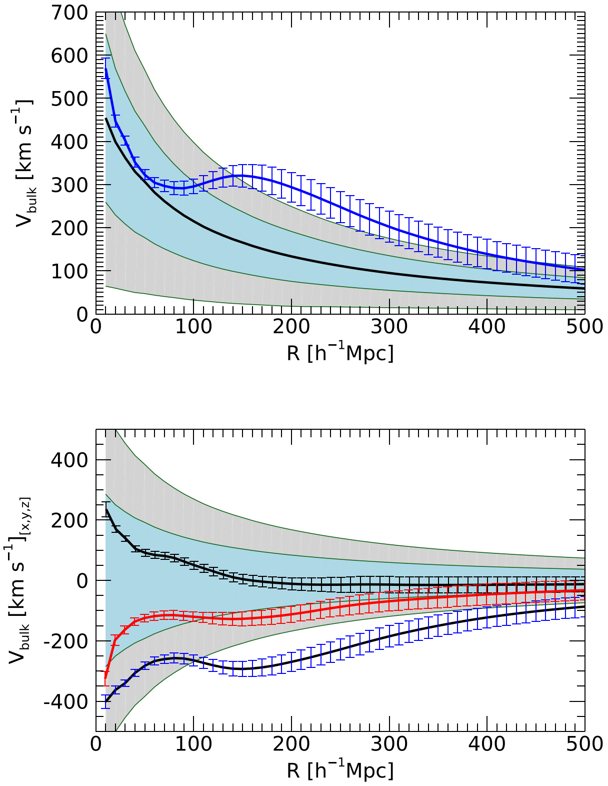

Figure 8. The amplitude of the bulk velocity (upper panel): the mean and scatter taken over the ensembles of the constrained (blue solid line, error bars) and of the random realizations (black solid lines, 1σ uncertainty (light blue) and 2σ (light grey) shaded regions). The bulk velocity is calculated within spheres of radius R. The three Supergalactic Cartesian components of the bulk velocity (lower panel): The mean and scatter of SGX (blue), SGY (black), and SGZ (red) components of the bulk velocity. The shaded regions (1σ, light blue, and 2σ, light grey) show the cosmic variance of one Cartesian component of the bulk velocity. The cosmic mean value of each of the individual components is zero and therefore is not shown.

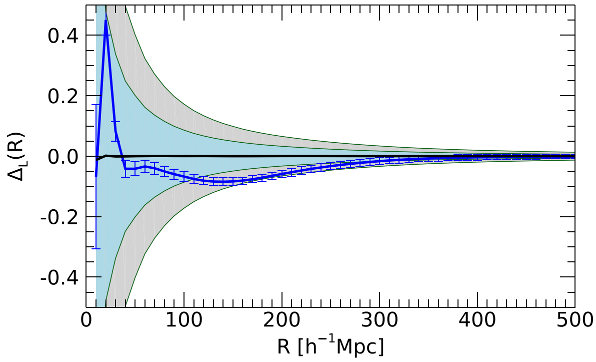

Figure 9. The estimated cumulative linear density field in spheres of radius R ($\Delta_{\rm L}(R)$); (conventions of the different lines are identical to those employed in the upper panel of Fig. 8).

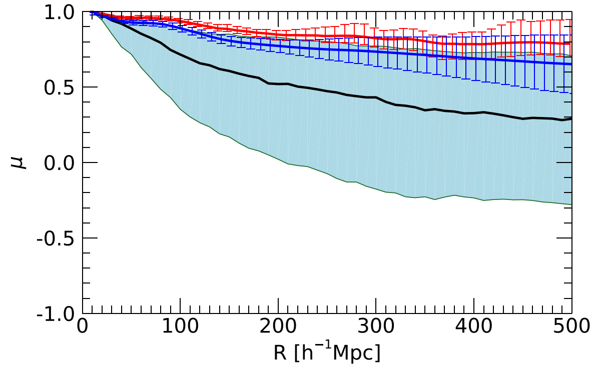

Figure 10. The alignment of the bulk velocity of a sphere of radius R with the CMB dipole velocity ($\mu_{\rm CMB}$) is presented by mean and variance taken over an ensemble 60 CRs (blue). To enable a comparison with the random realizations, the alignment with the CMB dipole is closely approximated by plotting the alignment of the bulk velocity at R with itself at zero lag, $\mu_{\rm self}(R) = \hat{V}_{bulk}(R_{\rm min}) \cdot \hat{V}_{bulk}$(R). The plot shows the median and the first and third quartiles of the distribution of the ensembles of 60 CRs (red) and 3000 random realizations (median in black solid line and the uncertainty in shaded light blue). Here $R_{\rm min} = 10\,h^{-1}$ Mpc.

Table 1. The mean and standard deviation taken over and ensemble of 60 CRs of the velocity field constrained by the CF4 data. The bulk velocity of an ensemble of 60 random realizations, with the same phases as the corresponding CRs, is shown for reference. The different columns present the mean and scatter of the amplitude of the bulk velocity taken over an ensemble of constrained (V_{bulk}^{CRs}) and random ($V_{\mathrm{bulk}}^{\mathrm{RANs}}$) realizations, of the Supergalactic Cartesian components of the bulk velocity ($V_{\mathrm{bulk,\,x}}$, $V_{\mathrm{bulk,\,y}}$, $V_{\mathrm{bulk,\,z}}$), the spherical mean linear overdensity ($\Delta_{\mathrm{L}}(R)$), the alignment of the bulk velocity with the CMB dipole (μ_{CMB}), the χ^2 value of V_{bulk} and the probability of having such a χ^2 of 3 degrees of freedom larger than the estimated value ($P(>\chi^2)$).

R (h^{-1} Mpc)	V_{bulk}^{CRs}	V_{bulk}^{RANs}	$V_{bulk,\,x}$	$V_{bulk,\,y}$	$V_{bulk,\,z}$ [km s^{-1}]	$\Delta_{\mathrm{L}}(R)$	μ_{CMB}	χ^2	$P(>\chi^2)$
10	557 ± 11	441 ± 187	-404 ± 8	245 ± 10	-293 ± 13	0.33 ± 0.11	0.99 ± 0.002	4.29	2.3e-01
40	354 ± 10	320 ± 133	-308 ± 10	103 ± 9	-137 ± 14	-0.04 ± 0.03	0.93 ± 0.009	3.51	3.2e-01
50	320 ± 13	295 ± 123	-280 ± 13	90 ± 10	-125 ± 13	-0.04 ± 0.02	0.93 ± 0.014	3.29	3.5e-01
100	295 ± 18	211 ± 84	-264 ± 19	46 ± 11	-121 ± 16	-0.07 ± 0.01	0.88 ± 0.021	5.34	1.5e-01
150	318 ± 27	162 ± 65	-290 ± 26	4 ± 15	-127 ± 21	-0.06 ± 0.01	0.81 ± 0.031	10.03	1.8e-02
200	292 ± 30	131 ± 51	-267 ± 31	-8 ± 19	-111 ± 27	-0.06 ± 0.01	0.78 ± 0.042	12.14	6.9e-03
250	247 ± 123	109 ± 43	-227 ± 32	-14 ± 23	-87 ± 29	-0.03 ± 0.01	0.75 ± 0.059	11.43	9.6e-03
300	210 ± 36	94 ± 40	-185 ± 32	-13 ± 26	-72 ± 31	-0.02 ± 0.01	0.73 ± 0.097	11.14	1.1e-02
500	102 ± 32	59 ± 24	-86 ± 34	-12 ± 24	-32 ± 32	-0.005 ± 0.01	0.65 ± 0.195	6.36	9.5e-02

Table 2. Comparison with other studies (see the text for details). The radius given here is the effective radius whose definition depends on the weighting scheme. The ΛCDM column provides the cosmic variance of the bulk velocity. The numbers that appear in the parentheses in the V_{bulk} column for this work correspond to the bulk velocity of the local flow induced within a sphere of $R = 300\,h^{-1}$ Mpc (see Section 5.5).

Source	Data	Weighting	Radius (h^{-1} Mpc)	V_{bulk} [km s^{-1}]	ΛCDM (km s^{-1})	l (°)	b (°)	χ^2
Carrick et al. (2015)	2M++ (redshift)	Gaussian	50	230 ± 30		293 ± 8	14 ± 10	
Hoffman, Courtois & Tully (2015)	CF2	Volume	50	258 ± 21	292 ± 147	280	18	
Scrimgeour et al. (2016)	6dFGS	Minimum variance	50	248 ± 58		318 ± 20	40 ± 13	
Qin et al. (2019)	2MTF	Maximum likelihood	37	259 ± 15	231^{+118}_{-101}	300 ± 4	23 ± 3	2.76
Lilow & Nusser (2021)	2MRS (redshift)	Gaussian	50	274 ± 50		287 ± 9	11 ± 10	
Boruah, Lavaux & Hudson (2022)	SFI++ & 2MTF	Gaussian	40	220 ± 21		295 ± 6	21 ± 5	
Watkins et al. (2023)	CF4	Minimum variance	150	387 ± 28	139	297 ± 4	-6 ± 3	19.34
	CF4	Minimum variance	200	419 ± 36	120	298 ± 5	-8 ± 4	29.13
Courtois et al. (2023)	CF4	Volume	50	255 ± 49		296	15	
	CF4	Volume	150	272 ± 105		295	14	
	CF4	Volume	200	205 ± 123		279	60	
Whitford, Howlett & Davis (2023)	CF4	Maximum likelihood	49	408 ± 165	196 ± 82	301	-18	6.2
	CF4	Minimum variance	173	420 ± 108	194 ± 86	297	5	16.0
This work	CF4	Gaussian	40	$303(196) \pm 14$	229 ± 129	291 ± 2	9 ± 1	4.20
	CF4	Gaussian	50	$303(191) \pm 16$	200 ± 94	292 ± 2	5 ± 2	5.06
	CF4	Volume	50	$320(217) \pm 13$	291 ± 914	291 ± 2	13 ± 1	3.12
	CF4	Volume	150	$318(200) \pm 26$	158 ± 65	293 ± 3	-1 ± 2	10.41
	CF4	Volume	200	$292(174) \pm 30$	127 ± 53	294 ± 5	-4 ± 3	13.17
	CF4	Volume	300	$202(78) \pm 36$	94 ± 40	297 ± 9	-5 ± 7	11.14

based reconstructions needs to be done with respect to the divergent velocity field only.

5.6 Dependence on Hubble's constant

At its core the Cosmicflows is a database of the observed galaxy redshifts and distance moduli, and an H_0 needs to be assumed in order to recast it as a distances and radial peculiar velocities catalogue. Obviously, the most appropriate value of H_0 to use is the one obtained by fitting to the data. Random mock data catalogues have been used to assess the uncertainties in the determination of H_0 in the ΛCDM standard model and the Cosmicflows data in the BGc framework (Hoffman et al. 2021). The mock data sampled the cosmic variance, namely the variation with the choice of the random observers, and the errors variance, i.e. the variation due the observational errors. The BGc analysis of the CF3 data yielded $H_0 = 75.8 \pm 0.1 \pm 1.0 \pm 0.4$ km s^{-1} Mpc^{-1}, where first, second and third uncertainties are due to fitting formal error, the cosmic variance and the errors variance, respectively (Hoffman et al. 2021). Repeating that analysis and assuming its assessed uncertainties, we find here $H_0 = 74.6 \pm 1.1$ km s^{-1} Mpc^{-1}, where the uncertainty includes the cosmic variance. Without the cosmic variance, the uncertainty drops to 0.4 km s^{-1} Mpc^{-1}. This stands in excellent agreement with the $H_0 = 74.6 \pm 0.8$ km s^{-1} Mpc^{-1} of Tully et al. (2023). The latter

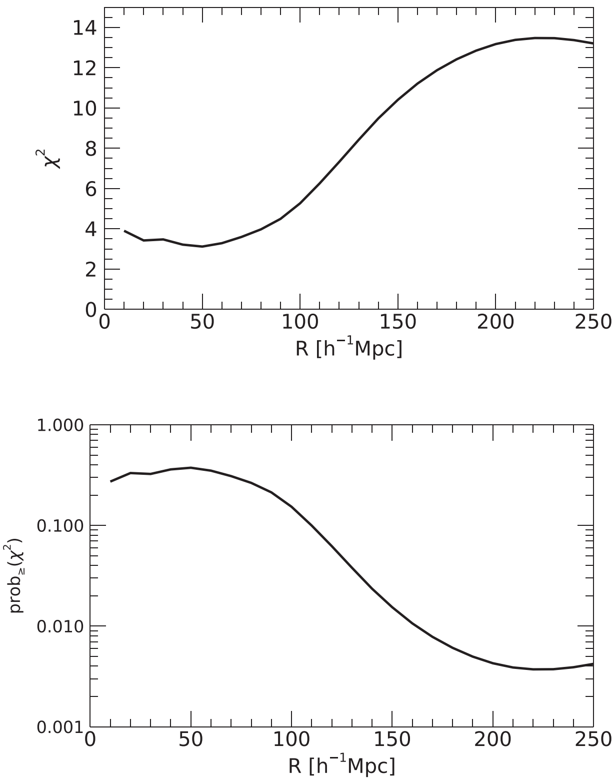

Figure 12. The χ^2 (of 3 dof) statistics of the amplitude of the bulk velocity: the upper panel shows the χ^2 profile as a function of R – the radius within which $V_{\rm bulk}(R)$ is calculated. The lower panel presents the probability of obtaining a χ^2 equal or larger than the one of the estimated $V_{\rm bulk}(R)$.

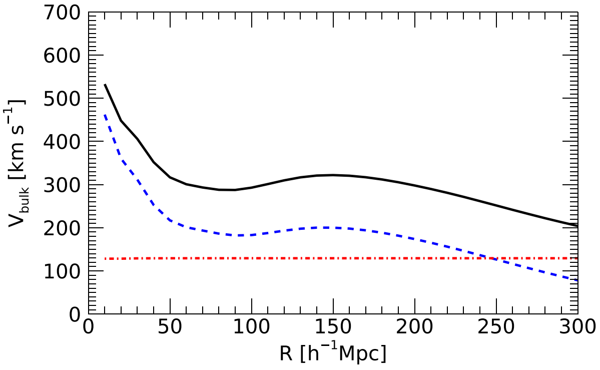

Figure 13. Decomposition of the velocity field into its tidal and local components with respect to sphere of radius $R = 300\,h^{-1}$ Mpc (see the text in Section 5.5): The bulk velocity of the full velocity field (solid, black), of the local component (dashed, blue) and of the tidal component (dot–dashed, red). The error bars of the case of the full flow (shown in Fig. 8) serve as upper limits on the bulk velocities of the two components.

result refers to the actual CF4 data, observed by humans on earth, hence it does not account for the cosmic variance. It should be stressed here that we do not account here to the possibility of systematic errors – the CF4 data are taken here at face value. In particular, the zero-point calibration of the sub-samples of the CF4 data is taken here as is from the Cosmicflows analysis.

Given the H_0 controversy, the dependence of the estimated monopole and dipole moments on H_0 is of interest and is examined here. The monopole moment corresponds to the local deviation from the global value of H_0 – hence a strong dependence on the assumed H_0 is expected, as indeed depicted by Fig. 14. Considering the bulk velocity then such a dependence is possible only in the case of an anisotropic distribution of the data points. For perfectly isotropically distributed data, the monopole and dipole moments are mutually orthogonal, hence the bulk flow cannot be affected by the choice of H_0. The CF4 data are extremely anisotropic and inhomogeneous (Fig. 2) hence a dependence of $V_{\rm bulk}$ on H_0 is expected. Fig. 14 indeed confirms that expectation, yet $V_{\rm bulk}$ hardly changes for H_0 in the range of $(40 - 60)\,h^{-1}$ Mpc.

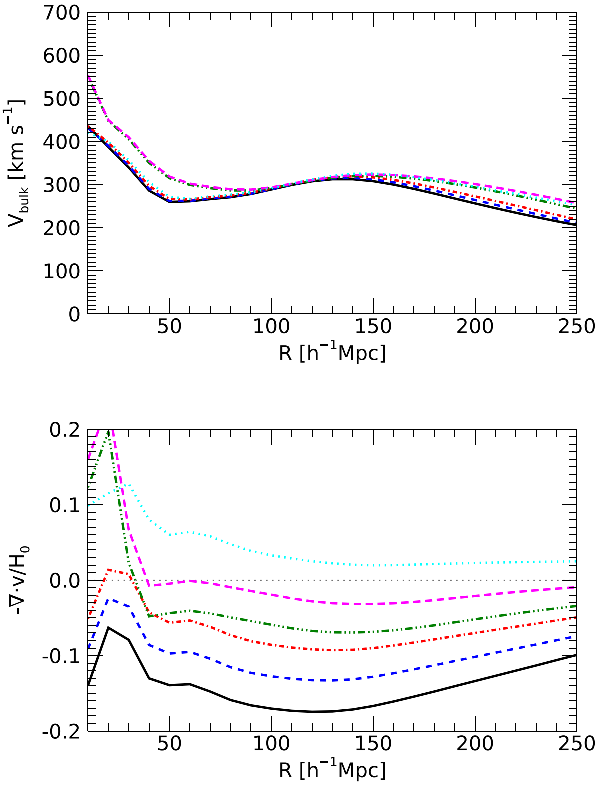

Figure 14. The dependence on the monopole (lower panel) and dipole (upper panel) moments moments on Hubble's constant: $H_0 = 71.0$, (black, solid), 72.0 (blue, dashed), 73.0 (red, dot–dashed), 74.0 (green, dot–dot–dashed), 75.0 (magenta, dashed), and $76.0\,h^{-1}$ Mpc (cyan, dotted).

6 INTERPRETATION AND COMPARISON

6.1 Validation and assessment

The quality of the combined BGc and WF/CRs algorithm is tested here against ensembles of random mock CF3 data (Appendix A) and of constrained mock CF3 data (Appendix B). The details of mock data sets, their constructions and their analysis are given in the two appendices.

We use here the monopole and dipole moments, namely the mean of the linear overdensity (up to the scaling by the growth factor) and the bulk velocity in a sphere of radius R, as tracers of the large-scale velocity field. Detailed analysis of reconstruction of these from random and constrained mock data is presented in Appendices A and B. A brief summary is given here. The main conclusion that follows from that analysis is that the residual of the WF reconstructed monopole and dipole moments from the ones of the target velocity field of the simulation is always smaller, and often much smaller, than the cosmic variance of these moments. Over the majority of ranges of R considered here the WF reconstruction is within 1σ to 2σ of the target simulation, where sigma is the constrained variance of the ensemble of CRs.

The conclusion we draw here is that the WF reconstruction provides a good estimation of the underlying velocity to within one to two times the constrained variance that is materialized by the scatter of the CRs.

6.2 Comparison with other studies: bulk velocity

The bulk velocity is arguably the most basic and common characteristic of the large-scale velocity field and therefore we compare here our estimation of it with some of its recent estimations from galaxy peculiar velocities and of galaxy redshift surveys (Table 2). We start by a brief description of these studies:

(i) Carrick et al. (2015): The velocity field is calculated by means of the linear theory from the 2M++ galaxy redshift compilation and the external tidal contribution has been fitted so as to minimize the residual from the observed redshifts. The bulk velocity was estimated by Gaussian weighting.

(ii) Hoffman, Courtois & Tully (2015): The LSS was reconstructed from the CF2 data by means of the Bayesian WF/CRs methodology, same as the one employed here, but with a different bias correction scheme. The bulk velocity is defined by the volume weighting in a sphere.

(iii) Lilow & Nusser (2021): The velocity field is estimated from the 2MRS galaxy redshift survey by means of linear CRs, assuming the ΛCDM standard model as a prior. The external tidal component is estimating by fitting to the CF3 data. A Gaussian window function is used to calculate the bulk flow.

(iv) Scrimgeour et al. (2016): Minimal variance estimation of the bulk flow from the 6dFGS velocities survey.

(v) Qin et al. (2019): Bulk flow is estimated by means of a maximum likelihood analysis from the 2MTF velocity survey.

(vi) Boruah, Lavaux & Hudson (2022): The large-scale velocity field is reconstructed by a linear Bayesian estimator – the Hamiltonian Monte Carlo algorithm – from the combined SFI++ and 2MTF velocity surveys. The bulk flow is calculated by means of a Gaussian filter.

(vii) Watkins et al. (2023): A minimal variance estimation of the bulk velocity of a spherical volume from the CF4 data. The authors use a novel bias correction algorithm of the observed radial peculiar velocities.

(viii) Courtois et al. (2023): A Bayesian reconstruction of the LSS from the grouped CF4 data by means of the Hamiltonian Monte Carlo algorithm. The bulk velocity is the volume-weighted average within spheres of radius R.

(ix) Whitford, Howlett & Davis (2023): An estimation of the bulk flow from the CF4 data by means of the maximum likelihood and by the minimum variance estimators.

(x) Current work: This work estimations of the bulk velocity for an effective radii of 40 and 50 h^{-1} Mpc are expressed in terms of the Gaussian weighting, so as to enable a direct comparison with the other studies. The Gaussian weighting calculation is based on a small ensemble of 60 random realizations, hence the uncertainty of the calculated χ^2 is $\sim$0.2.

One should recall that all the HMC-based reconstructions (Boruah, Lavaux & Hudson 2022; Courtois et al. 2023; Valade et al., in preparation) are done in Fourier space employing the periodic boundary conditions, hence recovering the local component of the velocity field.

The fact that all the estimations of the direction of the bulk velocity quoted here agree to within their formal errors is gratifying. The angular positions of the data points are very precisely determined, hence in spite of the sparseness of the data and the very incomplete sky coverage of some of the databases, all the methods and databases considered here provide consistent results about the direction of the bulk velocity. This is a sanity check that all studies have passed successfully. This is not the case with the amplitude of the bulk velocity. We focus here mostly on the comparison of our results with the other ones quoted here. Our best agreement is found with the V_{bulk} profile of fig. 11 of Lilow & Nusser (2021). We are also in good agreement with Qin et al. (2019) both in terms of the bulk velocity itself, at a single effective distance, and also with respect to its χ^2 significance with respect to the ΛCDM model.

All the Bayesian reconstructions considered here – the present WF/CRs and the HMC ones – reconstruct the full LSS, densities, and velocities, out of which the bulk velocity is calculated by means of a volume-weighted average within spherical volumes. This is the case with the bulk flow presented in fig. 10 of Boruah, Lavaux & Hudson (2022), given that it is the local component of the velocity field that is calculated by the HMC algorithm, it needs to be compared with the dashed-blue curve of Fig. 13. Visual inspection of the two figures finds an agreement out to $R \approx 50\,h^{-1}$ Mpc, beyond which the results diverge. The bulk velocity of Boruah, Lavaux & Hudson (2022) decreases monotonically with R over the entire plotted range, compared with our results that show an upturn in the V_{bulk} profile with a ‘hump’ that peaks at $R \approx 150\,h^{-1}$ Mpc. A similar upturn is exhibited by the V_{bulk} profile of Lilow & Nusser (2021). Yet, Fig. 11, which shows the V_{bulk} profiles for the three components of the CF4 data, provides the explanation of the discrepancy. The V_{bulk} profile of the ‘other’ component, after correcting for its local component, is similar in shape and consistent in amplitude with the profile of Boruah, Lavaux & Hudson (2022). We note here that the effective depths of the ‘others’ component (Fig. 1) and of the combined SFI++ and 2MTF (used by Boruah, Lavaux & Hudson 2022) are similar, $cz \sim 10^4$ km s^{-1} and both have quite a uniform angular distribution outside the ZOA. We conclude here that for similar data, the Boruah, Lavaux & Hudson (2022) and our results are in good agreement.

The HMC analysis of the CF4 grouped data by Courtois et al. (2023) stands in formal agreement with the present results. Namely, given the much larger uncertainty of the estimated V_{bulk} of these authors the present results are within the 1σ uncertainty of that work. Yet, an inspection of fig. 3 of Courtois et al. (2023) finds a radial profile of the bulk velocity that is very different than the one found here (Fig. 8). This work does not find any hint of the second hump, which peaks at $\sim 275\,h^{-1}$ Mpc, of Courtois et al. (2023). A closer look at Table 1 of that work finds a marked difference not only in the amplitude of V_{bulk} but also in its direction. The entry for the case of $R = 300\,h^{-1}$ Mpc (for the grouped CF4 data) shows $V_{\mathrm{bulk}} = 230 \pm 136$ km s^{-1} ($\boldsymbol{V}_{bulk} = (1, 229, -29)$ km s^{-1}) (no error bars on the Supergalactic Cartesian components are given), compared with our $V_{\mathrm{bulk}} = 202 \pm 36$ km s^{-1} ($\boldsymbol{V}_{bulk} = (-185 \pm 35, -13 \pm 26, -69 \pm 32)$ km s^{-1}). The directions of the two estimated bulk flows are very different. Comparable disagreements are found for other values of R.

Our work stands in disagreement with Watkins et al. (2023). That study uses the same CF4 data that is used here, yet the bias correction scheme is very different from the BGc used here. Also, a different weighting scheme for $V_{\rm bulk}$ is used there. On the face of it one cannot compared the minimal variance and the WF/CRs estimations at their quoted depth, R, yet the fact that cosmic variance calculated by the two methods are essentially the same argues that such a comparison is meaningful. There is a strong discrepancy, of a few sigma of one method from the other. Moreover, the statistical significance of the results, as gauged by the χ^2 statistics with respect to the standard ΛCDM model, varies dramatically between to the two works. The probability found by Watkins et al. (2023) of finding a χ^2 as large or larger than the estimated one is 2.3×10^{-4} and 2.1×10^{-6} for $R = 150$ and $200 h^{-1}$ Mpc, respectively. Here we find 6.9×10^{-3} and 9.6×10^{-3} for the same radii (Table 1). Watkins et al. (2023) result constitute a ~3.5 and ~4.5 sigma deviation from the standard ΛCDM model, compared with the ~2.6 of this work. Given that the two studies use the same data, it seems that the differences between the bulk velocities stem from the different bias correction schemes of the two studies.

The minimal variance analysis of the 6dFGS survey (Scrimgeour et al. 2016) finds at a depth of $R = 50\,h^{-1}$ Mpc an amplitude of $V_{\rm bulk}$ in close agreement with the present finding, yet the direction of the vector deviates from the one calculated here by ~40°, which constitutes a more than 1σ discrepancy. One might argue that such a discrepancy is not surprising given the very anisotropic distribution over the sky of the 6dFGS data point (see Fig. 2).

There is also tension between this work and the minimum variance estimation of Whitford, Howlett & Davis (2023). These authors estimate $V_{\rm bulk}$ at an effective depth of $R = 173\,h^{-1}$ Mpc to be 428 ± 108 km s^{-1} ($\boldsymbol{V}_{\rm bulk} = (-391 \pm 104, -119 \pm 93, -126 \pm 122)$ km s^{-1}). We find here $V_{\rm bulk} = 311 \pm 28$ km s^{-1} ($\boldsymbol{V}_{bulk} = (-285 \pm 28, -1 \pm 17, -122 \pm 24)$ km s^{-1}) at $R = 170\,h^{-1}$ Mpc. The ΛCDM cosmic variance at that depth is 194 ± 86 for the minimum variance estimator and 144 ± 60 km s^{-1} for the current WF/CRs estimator. One can fit an effective depth, of the WF/CRs algorithm, that corresponds to the effective depth of the minimum variance estimation by equating their corresponding ΛCDM predictions. This yields $R \sim 110\,h^{-1}$ Mpc for the $173\,h^{-1}$ Mpc of Whitford, Howlett & Davis (2023). For that depth, we find $V_{\rm bulk} = 303 \pm 20$ km s^{-1} ($\boldsymbol{V}_{\rm bulk} = (-273 \pm 21, 35 \pm 12, -124 \pm 16)$ km s^{-1}) – still some tension exists between the current results and Whitford, Howlett & Davis (2023), for the amplitude and the direction of the bulk velocity.

6.3 Comparison with other studies: density

The WF estimated linear density field provides a good proxy to the actual density field on scales, i.e resolution, of roughly $20\,h^{-1}$ Mpc or larger. It follows that the WF predicted cumulative overdensity profile ($\Delta_{\rm L}(R)$) constitutes a good approximation to the real density field for $R \gtrsim 20\,h^{-1}$ Mpc.

The prediction of the linear density field profile, $\Delta_{\rm L}(R)$, with the density inferred from surveys of galaxies or galaxy clusters is hampered by issues of selections and bias. We are tempted here to compare the WF predicted $\Delta_{\rm L}(R)$ with the corresponding density profile of the compilation of CLASSIX galaxy clusters (Böhringer, Chon & Collins 2020). Fig. 11 of these authors shows the cumulative profile of the clusters number density and using a simple bias model also that of the total matter density. Detailed analysis is to be presented elsewhere and only a qualitative comparison shows that both the measured (from clusters) and the predicted (from the CF4 data) local underdensity extends over the range of $30 \lesssim R \lesssim 250\,h^{-1}$ Mpc. Note that range of the local underdensity quite strongly depends on the value of H_0 (cf. Fig. 14). The agreement quoted here is obtained from $H_0 = 74.6$ km s^{-1} Mpc^{-1}.

The maximal statistical departure of $\Delta_{\rm L}(R)$ occurs at $R \sim 190\,h^{-1}$ Mpc, where it attains a value of roughly -1.9 times the ΛCDM cosmic variance.

7 SUMMARY AND DISCUSSION

The LSS of the Universe out to a distance of $\sim 300\,h^{-1}$ Mpc is reconstructed here from the CF4 grouped data of redshifts and distance moduli. An unbiased transformation of the input data to distances and peculiar velocities is performed by the BGc algorithm. The reconstruction is performed by means of the WF and CRs assuming the linear theory of the ΛCDM standard cosmological model as the Bayesian prior. The combined BGc and WF/CRs algorithm has been tested against a constrained and a set of random mock CF3-like data sets.

The main conclusion that follows from the testing of the BGc and WF/CRs algorithm against the mock data is that the WF reconstruction of the velocity field recovers the actual underlying density field to within one to two times the scatter exhibited by the ensemble of the CRs, namely the constrained variance. Furthermore, for the CF4 data and within the ΛCDM prior the constrained variance is significantly smaller than the cosmic one.

The WF/CRs predicted underdensity over the range of $30 \lesssim R \lesssim 250\,h^{-1}$ Mpc stands in qualitative agreement with the density inferred from the compilation of the CLASSIX galaxy clusters (Böhringer, Chon & Collins 2020). That range depends quite strongly on H_0, hence the agreement with the distribution of clusters provides an extra support to the claim that local probes give rise to higher values of H_0 and thereby intensifies the so-called Hubble constant tension.

The WF/CRs approach is a conservative one. Namely, where the data are ‘weak’ the Bayesian prior dominates the reconstruction and nothing new is learnt from the reconstruction. In the case of the standard cosmological model, in regions in configuration or Fourier spaces dominated by ‘weak’ data the WF reconstructed density and velocity fields converges towards the null fields and the CRs become random realizations. Yet in spite of this reservation, it is of interest to check in what sense and to what degree our local ‘patch’ of the Universe is typical, or maybe atypical. Our analysis shows that on distances exceeding $\sim 150\,h^{-1}$ Mpc and all the way to the edge of the data the local Universe is somewhat an outlier, on the 2σ to 3σ level. This is manifested by the amplitude of the $V_{\rm bulk}$ and the monopole moment profiles. This is also indicated by the χ^2 analysis of the bulk velocity vector. Also, the coherent nature of the direction of the bulk velocity vector is somewhat atypical. Our local patch of the Universe is a ‘bit’ unusual, at the $\sim 2\sigma$ level, in the sense of the coherence of the direction of the bulk velocity vector. This is also the maximal level of discrepancy of the density profile, $\Delta_{\rm L}(R)$.

How meaningful is such discrepancy? It is clear that being at the roughly 2σ edge of the cosmic variance is not enough to declare a possible tension with the ΛCDM model. Yet, it is enough to recognize our local cosmological neighborhood is somewhat atypical. It is interesting to see whether such a roughly 2σ discrepancy can be corroborated by other reconstructions of the local Universe, in particular, from redshift surveys of galaxies.

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of Nick Kaiser, whose seminal paper (Kaiser 1988) set the stage and provided the framework for subsequent studies of cosmological velocity surveys. This work has been done within the framework of the CLUES simulations. YH has been partially supported by the Israel Science Foundation grant ISF 1358/18. NIL and AV acknowledge financial support of the Project IDEXLYON at the University of Lyon under the Investments for the Future Program (ANR-16-IDEX-0005). JS acknowledges support from the ANR LOCALIZATION project, grant ANR-21-CE31-0019 of the French Agence Nationale de la Recherche. SP acknowledges financial support from the Deutsche Forschungs Gemeinschaft joint Polish–German research project LI 2015/7-1 (LUSTRE)

DATA AVAILABILITY

The estimated density and velocity fields are available upon reasonable request to the authors.

REFERENCES

Böhringer H., Chon G., Collins C. A., 2020, A&A, 633, A19

Boruah S. S., Lavaux G., Hudson M. J., 2022, MNRAS, 517, 4529

Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, MNRAS, 450, 317

Courtois H. M., Dupuy A., Guinet D., Baulieu G., Ruppin F., Brenas P., 2023, A&A, 670, L15

Courtois H. M., Hoffman Y., Tully R. B., Gottlöber S., 2012, ApJ, 744, 43

Courtois H. M., Pomarède D., Tully R. B., Hoffman Y., Courtois D., 2013, AJ, 146, 69

Courtois H. M., Tully R. B., Hoffman Y., Pomarède D., Graziani R., Dupuy A., 2017, ApJ, 847, L6

da Costa L. N., Freudling W., Wegner G., Giovanelli R., Haynes M. P., Salzer J. J., 1996, ApJ, 468, L5

Dekel A., Bertschinger E., Faber S. M., 1990, ApJ, 364, 349

Dekel A., Bertschinger E., Yahil A., Strauss M. A., Davis M., Huchra J. P., 1993, ApJ, 412, 1

Dolag K., Sorce J. G., Pilipenko S., Hernández-Martínez E., Valentini M., Gottlöber S., Aghanim N., Khabibullin I., 2023, A&A, 677, 16

Doumler T., Courtois H., Gottlöber S., Hoffman Y., 2013b, MNRAS, 430, 902

Doumler T., Gottlöber S., Hoffman Y., Courtois H., 2013a, MNRAS, 430, 912

Doumler T., Hoffman Y., Courtois H., Gottlöber S., 2013c, MNRAS, 430, 888

Forero-Romero J. E., Hoffman Y., Yepes G., Gottlöber S., Piontek R., Klypin A., Steinmetz M., 2011, MNRAS, 417, 1434

Graziani R., Courtois H. M., Lavaux G., Hoffman Y., Tully R. B., Copin Y., Pomarède D., 2019, MNRAS, 488, 5438

Hoffman Y. et al., 2018, Nat. Astron., 2, 680

Hoffman Y., Courtois H. M., Tully R. B., 2015, MNRAS, 449, 4494

Hoffman Y., Eldar A., Zaroubi S., Dekel A., 2001, preprint (astro-ph/0102190)

Hoffman Y., Martinez-Vaquero L. A., Yepes G., Gottlöber S., 2008, MNRAS, 386, 390

Hoffman Y., Nusser A., Valade A., Libeskind N. I., Tully R. B., 2021, MNRAS, 505, 3380

Hoffman Y., Pomarède D., Tully R. B., Courtois H. M., 2017, Nat. Astron., 1, 0036

Hoffman Y., Ribak E., 1991, ApJ, 380, L5

Hong S. E., Jeong D., Hwang H. S., Kim J., 2021, ApJ, 913, 76

Huchra J. P. et al., 2012, ApJS, 199, 26

Kaiser N., 1988, MNRAS, 231, 149

Klypin A., 2016, MNRAS, 457, 4340

Klypin A., Hoffman Y., Kravtsov A. V., Gottlöber S., 2003, ApJ, 596, 19

Lavaux G., 2016, MNRAS, 457, 172

Libeskind N. I. et al., 2020, MNRAS, 498, 2968

Lilje P. B., Yahil A., Jones B. J. T., 1986, ApJ, 307, 91

Lilow R., Nusser A., 2021, MNRAS, 507, 1557

Lynden-Bell D., Faber S. M., Burstein D., Davies R. L., Dressler A., Terlevich R. J., Wegner G., 1988, ApJ, 326, 19

Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, A&A, 570, A13

Masters K. L., Springob C. M., Haynes M. P., Giovanelli R., 2006, ApJ, 653, 861

Nusser A., 2017, MNRAS, 470, 445

Nusser A., Davis M., 2011, ApJ, 736, 93

Ocvirk P. et al., 2020, MNRAS, 496, 4087

Peebles P. J. E., 1980, The Large-Scale Structure of the Universe,Princeton University Press, Princeton, New Jersey

Peery S., Watkins R., Feldman H. A., 2018, MNRAS, 481, 1368

Pfeifer S., Valade A., Gottlöber S., Hoffman Y., Libeskind N. I., Hellwing W. A., 2023, MNRAS, 523, 5985

Planck Collaboration XVI, 2014, A&A, 571, A16

Pomarède D., Hoffman Y., Courtois H. M., Tully R. B., 2017, ApJ, 845, 55

Pomarède D., Tully R. B., Graziani R., Courtois H. M., Hoffman Y., Lezmy J., 2020, ApJ, 897, 133

Prideaux-Ghee J., Leclercq F., Lavaux G., Heavens A., Jasche J., 2023, MNRAS, 518, 4191

Qin F., Howlett C., Staveley-Smith L., Hong T., 2019, MNRAS, 482, 1920

Scrimgeour M. I. et al., 2016, MNRAS, 455, 386

Sorce J. G. et al., 2016, MNRAS, 455, 2078

Sorce J. G., 2015, MNRAS, 450, 2644

Sorce J. G., 2018, MNRAS, 478, 5199

Sorce J. G., Courtois H. M., Gottlöber S., Hoffman Y., Tully R. B., 2014, MNRAS, 437, 3586

Sorce J. G., Dubois Y., Blaizot J., McGee S. L., Yepes G., Knebe A., 2021, MNRAS, 504, 2998

Sorce J. G., Mohayaee R., Aghanim N., Dolag K., Malavasi N., 2023, A&A, 679, 1

Sorce J. G., Stoica R. S., Tempel E., 2023, A&A, 679, 25

Strauss M. A., Willick J. A., 1995, Phys. Rep., 261, 271

Tully R. B. et al., 2013, AJ, 146, 86

Tully R. B. et al., 2023, ApJ, 944, 94

Tully R. B., Courtois H. M., Sorce J. G., 2016, AJ, 152, 50

Tully R. B., Courtois H., Hoffman Y., Pomarède D., 2014, Nature, 513, 71

Tully R. B., Shaya E. J., Karachentsev I. D., Courtois H. M., Kocevski D. D., Rizzi L., Peel A., 2008, ApJ, 676, 184

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

Watkins R. et al., 2023, MNRAS, 524, 1885

Weinberg S., 2008, Cosmology, Oxford University Press, Oxford

Whitford A. M., Howlett C., Davis T. M., 2023, MNRAS, 526, 3051

Yepes G., Gottlöber S., Hoffman Y., 2014, New Astron. Rev., 58, 1

Zaroubi S., Bernardi M., da Costa L. N., Hoffman Y., Alonso M. V., Wegner G., Willmer C. N. A., Pellegrini P. S., 2001, MNRAS, 326, 375

Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446

APPENDIX A: WF/BGC RECONSTRUCTION: RANDOM MOCK CF3-LIKE DATA

Hoffman et al. (2021) present a detailed description of the mock catalogues used here. A brief description of these follows. Mock catalogues of the grouped CF3 data were constructed so as to test the BGc bias correction scheme. The mock catalogues are drawn from the publicly available Multi-Dark-2 simulation[4] (Klypin et al. 2016). This is a DM-only *N*-body simulation with $N = 3840^3$ in a periodic box of side length $1.0h^{-1}$ Gpc particles, assuming a Planck cosmology ($H_0 = 67.7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.69$, $\Omega_b = 0.04$, $\Omega_m = 0.31$, and $\sigma_8 = 0.82$) (Planck Collaboration XVI 2014).

---

[4] see https://www.cosmosim.org

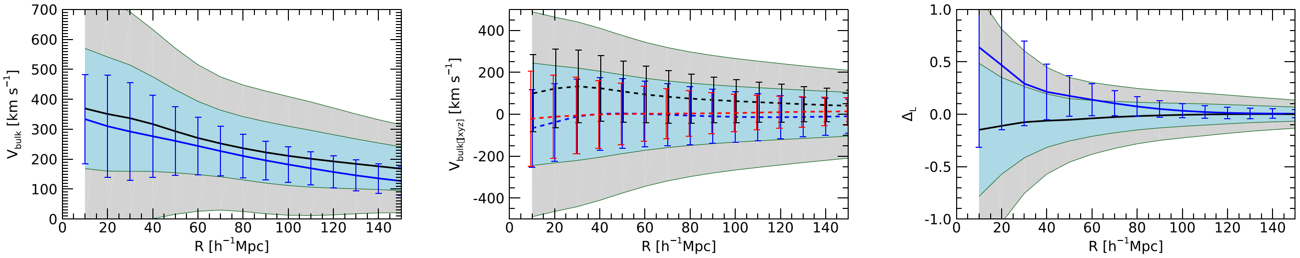

Figure A1. The mean and variance of an ensemble of the WF reconstruction of 100 mock data realizations, of the 10 random observers and of the 10 errors realizations for each observer. The shaded uncertainties correspond to an ensemble of 200 random realizations and are given for reference. Colour and line style conventions are the ones used in Figs 8 and 9.

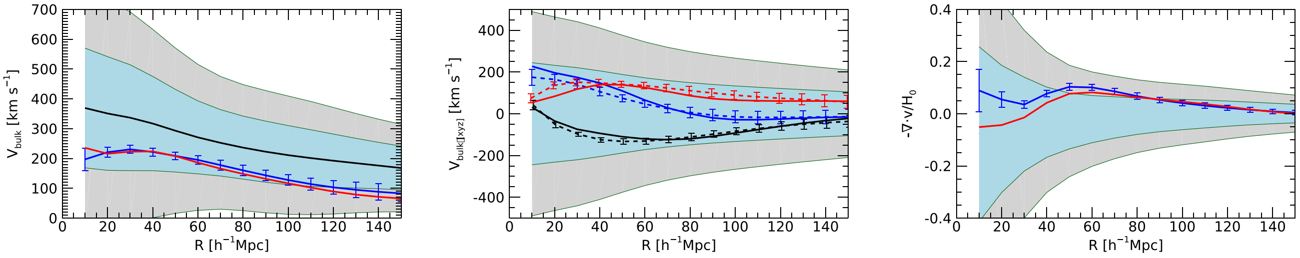

Figure A2. The mean and variance of an ensemble of 30 CRs of a single errors realization of one of the mock observers. Colour and line style conventions are the ones used in Figs 8 and 9. In addition, the red solid line represents the corresponding profiles of that particular target simulation.

An FOF algorithm is run on the $z = 0$ particle distribution and all groups larger than 20 particles are retained. Ten different mock observers have been selected at random within the computational box, subject to a single constraint – that they ‘reside’ within $\approx 10^{12}\,\mathrm{M}_\odot$ halos. The distribution of the mock data points was designed to reproduce the spatial distribution of the CF3 data and the measurement uncertainties are inherited from the actual CF3 data (for details Hoffman et al. 2021). Ten different random realizations of the observational errors were constructed for each mock observer, resulting in a total of 100 mock catalogues – representing the cosmic variance (10 mock observers) and the error variance (10 errors realizations).

The BGc algorithm has been applied to the mock catalogues, so as to correct the lognormal bias. The LSS has been reconstructed, within the linear regime, by means of the WF/CRs algorithm. Fig. A1 presents the monopole and dipole moments of the WF reconstructed from the 100 mock catalogues. The plots of the figures present the mean and scatter of the ensemble of the 10 errors realizations of the mock data of each observer. The WF reconstruction is done on a cubic grid of $N = 128^3$ spanning a box of side length $L = 500\,h^{-1}$ Mpc. The resolution of the first two bins of the profiles ($R = 10$ and $20\,h^{-1}$ Mpc) is rather poor but we opted to show them nevertheless.

The mean and variance of the amplitude and the Supergalactic coordinates of the bulk velocity and of the cumulative linear density profiles of the ensemble of WF reconstructions from an ensemble of 100 mock data sets are presented in Fig. A1. The mean and variance of an ensemble of random realizations is presented in shaded confidence intervals for reference. The bulk velocity plots show that the WF reconstruction recovers the ‘true’ underlying profiles to better than one standard deviation of the distribution of the WF profiles. The figure further shows the expected underestimation of the WF reconstruction. The mean spherical density profiles deviate from the expected random realizations. This is a reflection of the fact that the random observers are selected to coincides with $\approx 10^{12} h^{-1}\,\mathrm{M}_\odot$ halos. Such halos tend to reside in overdense environments, and hence the bias.

Next, the more relevant analysis of an ensemble of 30 CRs using one errors realization of one particular observer, namely one of the 100 cosmic and errors realization, is considered here. Such an analysis is actually employed in the case of the actual CF4 data, drawn from our actual Universe. Fig. A2 presents the mean and cosmic variance of the amplitude and the values of the three Cartesian components of the bulk velocity and of the monopole moment. The plots show the constrained mean and variance, calculated over an ensemble of 30 CRs, and the corresponding ‘true’ profiles of the underling velocity field of the target simulation. The mean and variance of an ensemble of 30 random realizations are shown as well as a manifestation of the cosmic variance. The plots show that the constrained variance is much smaller that the cosmic one and that the residual of the mean of the CRs, which is effectively equal to the WF estimated profiles, is always much smaller than the cosmic variance. It should be recalled here that the scatter of residual of the CRs from the actual field depends only on the measurement uncertainties and not the actual values of the underlying field (Hoffman & Ribak 1991), hence all statements made here about the residual from the ‘truth’ apply to all 10 different mock observers.

APPENDIX B: WF/BGC RECONSTRUCTION: CONSTRAINED CF3-LIKE MOCK DATA

Doumler et al. (2013a, c) presented a new and alternative way for constructing mock galaxy velocity catalogues. This was done by drawing the mock data from a constrained simulation, one that was constrained by actual observed data (Klypin et al. 2003; Sorce et al. 2023). The main merit of such constrained mock data is that the choices of the mock data points are correlated with the

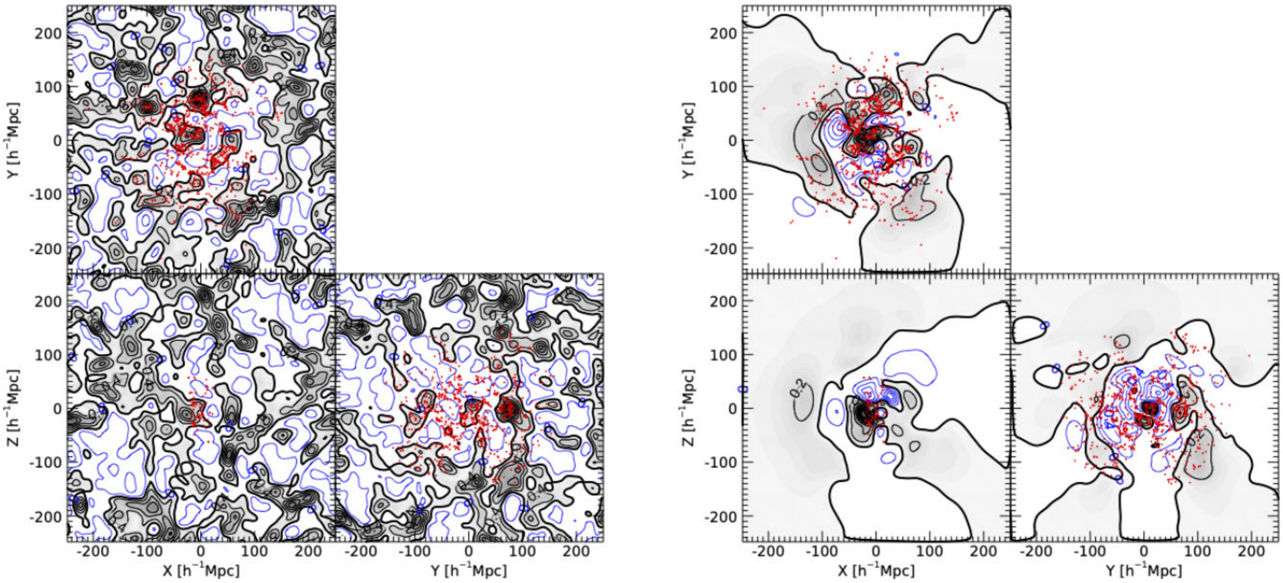

Figure B1. A comparison of the target CF2-constrained simulation with the WF reconstruction: The CIC interpolated linear density field ($\delta_{\rm L}$) of the target simulation (left panel) is compared with the WF reconstructed linear density field (right panel). Both fields are Gaussian smoothed with a kernel of $R_s = 5\,h^{-1}$ Mpc. The constrained CF3-like mock data points (within a slab of $\pm 5\,h^{-1}$ Mpc) are superimposed on both images (red dots).

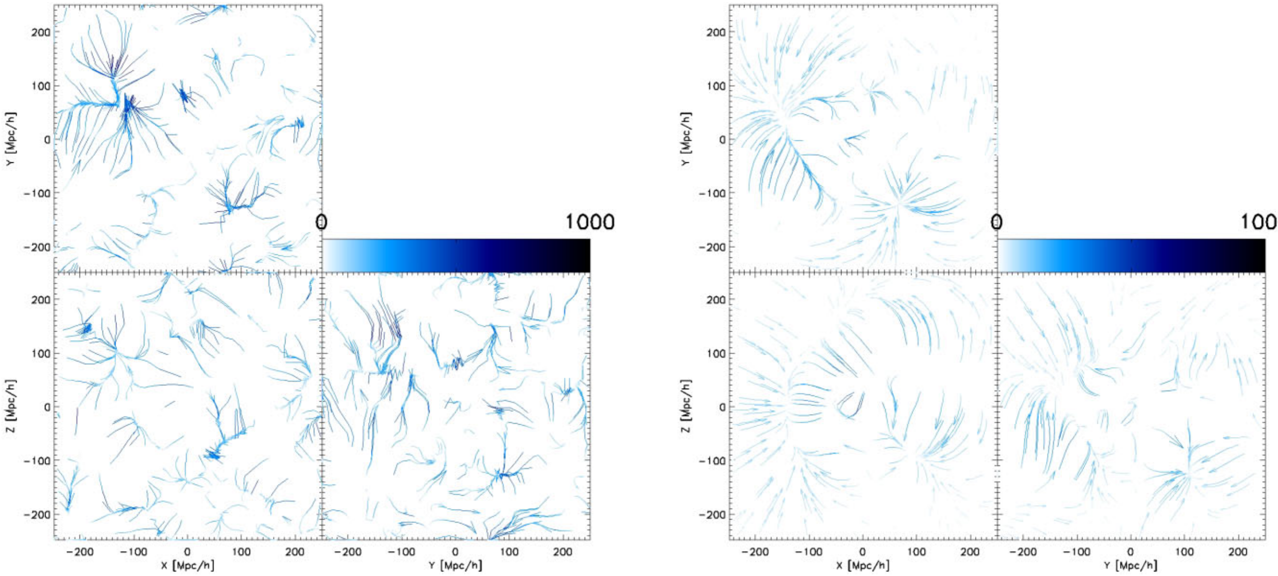

Figure B2. A comparison of the target CF2-constrained simulation with the WF reconstruction: The CIC interpolated velocity field of the target simulation (left panel) is compared with the WF reconstructed velocity field (right panel). Both fields are Gaussian smoothed with a kernel of $R_s = 5\,h^{-1}$ Mpc and the field is presented by streamlines.

underlying density field of the actual Universe. Namely, the constrained simulation reproduce, to a certain degree, the neighborhood of prominent structures such as the Virgo cluster, say. The cluster itself is represented by one data point only, in the grouped data catalogue, of a reduced error, and is surrounded by many nearby data points. Such a data distribution enhances the constraining power of the data and makes the Virgo cluster and its neighborhood more constrained than other nearby neighborhoods. The constrained mock data captures that feature of the actual data.

A CF3-like constrained mock grouped data, drawn from a ΛCDM simulation constrained by the CF2 data have been recently constructed (Sorce et al. 2023). The BGc bias correction scheme has been applied to that data and the WF/CRs algorithm has been applied to it. The quality of the BGc and the WF/CRs algorithms is gauged by the comparison with the target-constrained simulation. Fig. B1 presents a comparison of the $\delta_{\rm L}$ field (equation 2; Gaussian smoothed with a kernel of $5\,h^{-1}$ Mpc) with the WF reconstructed δ field. The constrained CF3-like mock data points (within a slab of $\pm 5\,h^{-1}$ Mpc) are superimposed for the sake of orientation and

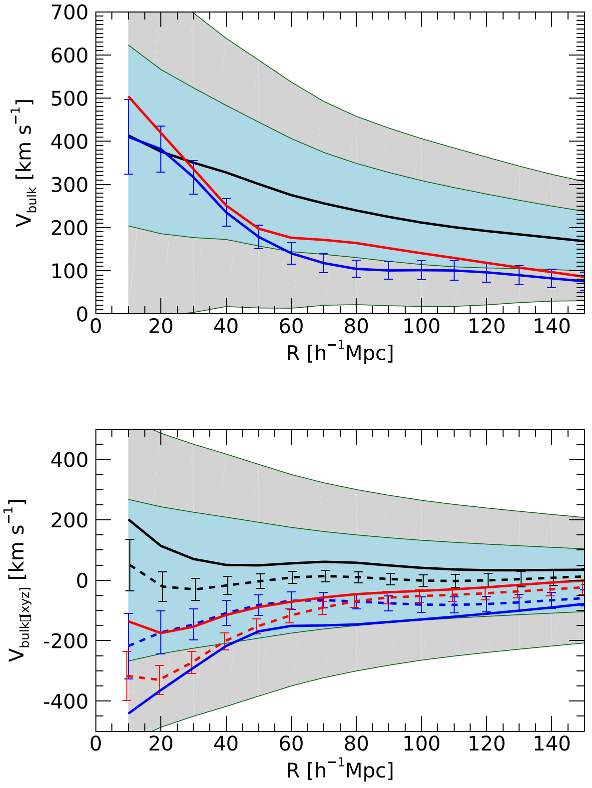

Figure B3. A comparison of the target CF2-constrained simulation with the WF/CRs reconstruction from a CF3-like mock data drawn from that simulation. The mean and scatter are calculated for ensembles of 60 constrained and 60 random realizations. The amplitude of the bulk velocity (upper panel): the mean and scatter taken over the ensembles of the constrained (blue solid line, error bars) and of the random realizations [black solid lines, 1σ uncertainty (light blue) and 2σ (light grey) shaded regions]. The bulk velocity is calculated within spheres of radius R. The red solid line presents the target bulk velocity profile. The three Supergalactic Cartesian components of the bulk velocity (lower panel): the mean and scatter of SGX (blue), SGY (black), and SGZ (red) components of the bulk velocity. The dashed lines with error bars correspond to the WF/Crs reconstructions and the solid lines to the target profiles. The shaded regions (1σ, light blue, and 2σ, light grey) show the cosmic variance of one Cartesian component of the bulk velocity. The cosmic mean value of each of the individual components is zero and therefore is not shown.

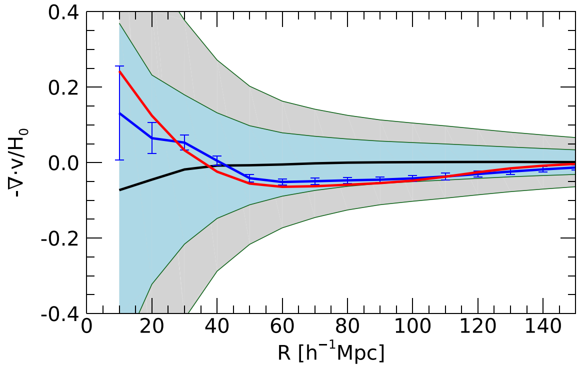

Figure B4. The monopole moment of the velocity field ($-\nabla \cdot v/H_0$), calculated in spheres of radius R (conventions of the different lines are identical to those employed in the upper panel of Fig. 8. The minus sign and the H_0 scaling are introduced so as to make it proportional to the linear density.

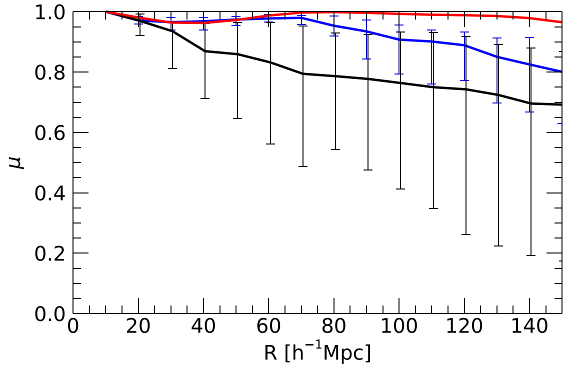

Figure B5. The alignment of the bulk velocity of a sphere of radius R with itself at zero lag, $\mu_{\rm self}(R) = \hat{V}_{\rm bulk}(R_{\rm min}) \cdot \hat{V}_{\rm bulk}(R)$. The zero lag is defined by $R_{\rm min} = 10\,h^{-1}$ Mpc. The plot shows the median and the lower and upper quartiles of the distribution the ensembles of 50 CRs (blue) and the corresponding 50 random realizations (black). The alignment of the bulk velocity of the target simulation with itself at zero lag is show as well (red).

comparison. The effective depth of the CF3 data is $\sim 160\,h^{-1}$ Mpc and beyond it the quality of the reconstruction degrades to the null field. As expected, the reconstructed velocity field recovers the LSS of the target simulation better than the density field (Fig. B2).

Fig. B3 presents the amplitude and the three Cartesian coordinates of the bulk velocity out to a distance of $R = 150\,h^{-1}$ Mpc. Over most of that range, the $V_{\rm bulk}$ of the target simulation lies within the scatter around the mean of the ensemble of CRs. The discrepancy is larger than that, within the 2σ level, within the range of $60 < R < 100\,h^{-1}$ Mpc. As expected, the discrepancy of the individual Cartesian components is somewhat larger, yet it is considerably smaller than the cosmic variance. The monopole moment of the target simulation is very well recovered by WF/CRs reconstruction (Fig. B4). The constrained variance around the mean profile of the CRs, hence of the WF, is much smaller than the cosmic variance. The slight deviation from zero of the mean of the CRs is due to the very finite sampling at small radii and the relatively small number of CRs. The alignment of $V_{\rm bulk}$ with itself (at zero lag) is considerably larger than that of RAN, with the constrained variance of the alignment being much smaller than the cosmic variance (Fig. B5).

APPENDIX C: WF RECONSTRUCTION: CF VERSUS CF2

A comparison of the large-scale density field reconstructed from the CF4 data with the one from the CF2 is depicted in Fig. C1. The 2MRS galaxies (Huchra et al. 2012) are superimposed on the density maps as a means of validation of the reconstruction and for the sake of orientation.

The CF2 grouped data consist of 4814 data points within an effective depth of $\sim 100\,h^{-1}$ Mpc, compared with the 38 060 data points within $\sim 300\,h^{-1}$ Mpc of the grouped CF4 catalogue. Also, the lognormal bias correction of the CF4 data is much better than the one used for the CF2 data. Yet, the density field of the CF4 reconstruction and of the CF2 data are in good agreement within the CF2 data zone.

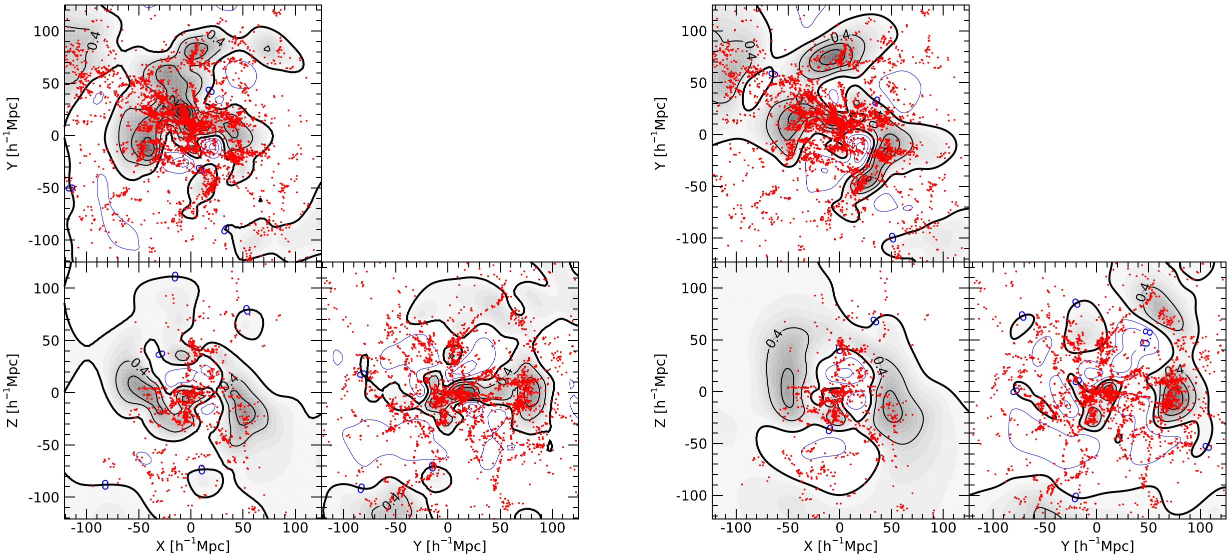

Figure C1. The WF reconstructed overdensity field of the CF4 data (left panel is compared with that of the CF2 one of the right panel). The density fields are smoothed with a $5.0\,h^{-1}$ Mpc Gaussian kernel. The 2MRS galaxies, within ±5.0, h^{-1} Mpc of the principal Supergalactic planes, are superimposed as a mean for validation of the reconstruction and for the sake of orientation.

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