



Observational biases in Lagrangian reconstructions of cosmic velocity fields

G. Lavaux, R. Mohayaee, S. Colombi, R. B. Tully, F. Bernardeau, J. Silk

► To cite this version:

G. Lavaux, R. Mohayaee, S. Colombi, R. B. Tully, F. Bernardeau, et al.. Observational biases in Lagrangian reconstructions of cosmic velocity fields. Monthly Notices of the Royal Astronomical Society, 2008, 383, pp.1292-1318. 10.1111/j.1365-2966.2007.12539.x . hal-03646501

HAL Id: hal-03646501

<https://hal.science/hal-03646501>

Submitted on 30 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Observational biases in Lagrangian reconstructions of cosmic velocity fields

G. Lavaux,¹^{*} R. Mohayaee,¹ S. Colombi,¹ R. B. Tully,² F. Bernardeau³ and J. Silk^{1,4}

¹*Institut d'Astrophysique de Paris – UMR 7095, 98bis bd Arago, CNRS/Université Pierre et Marie Curie, France*

²*Institute for Astronomy, University of Hawaii, Honolulu, USA*

³*Service de Physique Théorique, CEA/DSM/SPHT, Unité de recherche associée au CNRS, CEA/Saclay 91191 Gif-sur-Yvette cedex, France*

⁴*Department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH*

Accepted 2007 October 2. Received 2007 September 30; in original form 2007 July 24

ABSTRACT

Lagrangian reconstruction of large-scale peculiar velocity fields can be strongly affected by observational biases. We develop a thorough analysis of these systematic effects by relying on specially selected mock catalogues. For the purpose of this paper, we use the Monge–Ampère–Kantorovitch (MAK) reconstruction method, although any other Lagrangian reconstruction method should be sensitive to the same problems. We extensively study the uncertainty in the mass-to-light assignment due to incompleteness (missing luminous mass tracers), and the poorly determined relation between mass and luminosity. The impact of redshift distortion corrections is analysed in the context of MAK and we check the importance of edge and finite-volume effects on the reconstructed velocities. Using three mock catalogues with different average densities, we also study the effect of cosmic variance. In particular, one of them presents the same global features as found in observational catalogues that extend to 80 h^{-1} Mpc scales. We give recipes, checked using the aforementioned mock catalogues, to handle these particular observational effects, after having introduced them into the mock catalogues so as to quantitatively mimic the most densely sampled currently available galaxy catalogue of the nearby Universe. Once biases have been taken care of, the typical resulting error in reconstructed velocities is typically about a *quarter* of the overall velocity dispersion, and without significant bias. We finally model our reconstruction errors to propose an improved Bayesian approach to measure Ω_m in an unbiased way by comparing the reconstructed velocities to the measured ones in distance space, even though they may be plagued by large errors. We show that, in the context of observational data, it is possible to build a nearly unbiased estimator of Ω_m using MAK reconstruction.

Key words: methods: analytical and numerical – galaxies: distances and redshifts – dark matter – cosmological parameters.

1 INTRODUCTION

Galaxy redshift catalogues provide us with the radial velocities of the galaxies,

$$cz = H_0 r + v_r, \quad (1)$$

which are partly due to the global Hubble expansion ($H_0 r$ with H_0 the present value of the Hubble parameter) and partly due to the line-of-sight components of the peculiar velocities (v_r). Peculiar velocities are the deviations of galaxy velocities from the uniform Hubble expansion, due to the non-homogeneous distribution of matter in the Universe. The peculiar velocities are thus tracers of mass distribution in the Universe and can have far-reaching implications for

cosmology. As tracers of dark matter, peculiar velocities can be used to determine the local and global distribution of dark matter. From expression (1), it is evident that observations of galaxy redshifts (z) supplemented by measure of radial distances (r), would yield the peculiar velocities. However, measuring distances is a non-trivial exercise. The Tully–Fisher (TF) relation, surface brightness fluctuations, the Faber–Jackson relation for ellipticals (and their siblings, including the fundamental plane and the $D_n - \sigma$ methods, the Tip of the Red Giant Branch, Cepheids and SNIa) are the most usual methods for obtaining distances. The data gathered are, however, rather sparse: out of about a million galaxies whose redshifts are presently known with surveys such as 2dF and Sloan Digital Sky Survey (SDSS) only a few thousand have measured distances. Moreover, distances for most of these galaxies have too large peculiar velocity errors (due essentially to errors in distance measurements) to be useful in studying dynamics. For instance, distance

^{*}E-mail: lavaux@iap.fr

indicators such as the TF relation suffer from 20 per cent relative distance errors and thus produce quite noisy measurements at relatively moderate redshifts (i.e. $cz \gtrsim 3000 \text{ km s}^{-1}$). The data also suffer from selection biases (Strauss & Willick 1995; Tully & Pierce 2000). One way of reducing the error bars on distances is to average over many distance measurements for galaxies in clusters or groups and also by combining the results from different distance estimators. This treatment decreases the error bars on distances to about 8 per cent relative distance errors (Tully et al. 2007). Even though all these difficulties can be surmounted, one can finally hope to only have a sparse sample (as compared to redshift samples) of radial components of peculiar velocities. Fortunately, we now have Lagrangian velocity reconstruction schemes that are based solely on current redshift positions of mass tracers. The reconstructed velocities depend on cosmological parameters. Thus, comparing predictions obtained through Lagrangian reconstruction algorithms and the measured velocities may give estimations of these parameters.

This brings us to the *main point* that this paper tries to address: developing a *robust* and *unbiased* method of Lagrangian peculiar velocity reconstruction using redshift catalogues, in particular when observational effects distort most of the required data needed for the reconstruction of the dynamics. The reconstructed velocities are then compared to the measured ones using an *ad hoc* algorithm to yield a measurement of Ω_m , the mean matter density of the Universe.

Throughout the paper, we will try to mimic observational effects as they appear in the most densely sampled currently available galaxy catalogue of the nearby universe which has been compiled by one of the authors (R. B. Tully). This galaxy catalogue is built from different sources such as ZCAT (Zwicky Catalog; Huchra et al. 1992) and SSRS (Southern Sky Redshift Survey; da Costa et al. 1988). Only galaxies for which $cz \leq 8000 \text{ km s}^{-1}$ have been introduced in the catalogue. This catalogue is named NBG-8k, standing for NearBy Galaxy catalogue with a depth of 8000 km s^{-1} . Although selection criteria for this catalogue are not well defined, it will prove to be useful for the study of smaller galaxy catalogues such as NBG-3k (Tully et al. 2007).

For the purpose of this paper, we use a recently developed technique, called the Monge–Ampère–Kantorovitch reconstruction method (hereafter MAK), which is an approximation to the full non-linear dynamics to trace orbits back in time. This is a Lagrangian method, such as PIZA (Particle Interchange Zel’dovich Approximation; Croft & Gaztanaga 1997) or the Least-Action method (Peebles 1989), and not a Eulerian technique such as, e.g., POTENT (Bertschinger & Dekel 1989). One must note that the results of this paper are also valid for the other Lagrangian reconstruction methods as all the effects we are going to analyse are explainable in terms of gravitational dynamics. The MAK reconstruction has already been largely discussed when applied on numerical simulations (Brenier et al. 2003; Mohayaee et al. 2006). It is based on assuming that the dark matter displacement field is convex and potential, i.e. irrotational. In doing so, we exclude displacement fields which include multistreaming regions. The main result is that it is then possible to reconstruct accurately and uniquely the displacement field of dark matter particles between their original position and their current position. Practically, to solve the MAK problem, one must minimize a cost function for the assignment of a dark matter particle at the present comoving position \mathbf{x}_i and its initial comoving position \mathbf{q}_i :

$$S_\sigma = \sum_{i=1}^N [\mathbf{x}_i - \mathbf{q}_{\sigma(i)}]^2. \quad (2)$$

If the Universe is assumed to be initially homogeneous, which is a fair hypothesis supported by CMB data (e.g. Wilkinson Microwave Anisotropy Probe (WMAP) first year in Bennett et al. 2003),¹ then \mathbf{q}_i must be distributed on a uniform grid and the solution to the MAK problem is unique and given by the assignment σ which minimizes S_σ . The derived solution is then necessarily irrotational and derives from a convex potential. To solve this problem, we have implemented a parallel version of the so-called ‘auction’ algorithm proposed by Bertsekas (1979).² Of course, as we are using an approximation to the dynamics, the solution to the problem will be only valid above some scale (typically a few $h^{-1} \text{ Mpc}$). Once the solution is found, the immediate output of MAK reconstruction is the non-linear displacement field $\psi(\mathbf{q}) = \mathbf{x}(\mathbf{q}) - \mathbf{q}$, which can be used to find the peculiar velocity field \mathbf{v} using the first-order Zel’dovich approximation:

$$\mathbf{v}_i = \beta \psi_i, \quad (3)$$

where the subscript i indicates the comparison is achieved on the corresponding field averaged over the object i (i.e. in a Lagrangian way), and the linear growth factor $\beta \simeq \Omega_m^{5/9}$ (Bouchet et al. 1995). This best fit for β is valid as soon as $\Omega_m + \Omega_\Lambda = 1$, Ω_Λ being the present dark energy density. It appears then that a direct comparison of ψ_i against \mathbf{v}_i should in principle give us β and thus Ω_m . Though naive measurements (Mohayaee & Tully 2005) and preliminary studies (Branchini, Eldar & Nusser 2002; Phelps et al. 2006) on mock redshift catalogues have already been tried, the observational biases and systematic errors in the velocity–velocity comparison have never been studied thoroughly.

This paper is organized as follows. In Section 2, we describe the simulation and the basic mock catalogues that are used in the rest of this paper. Subsequent mock catalogues integrate more and more observational features but are still based on the same original basic mock catalogues presented in this section. Section 3 gives a model for the error distribution on MAK velocities and discuss the first problematic features of the comparison between MAK and measured velocities. This error distribution helps us in particular to establish the likelihood analysis in Section 7. We go then to the first main topic of this paper in Section 4 by studying the systematic errors introduced by arbitrary mass-to-light assignments in redshift catalogues. This section includes a study of missing mass correction (Section 4.1), unknown M/L function (Section 4.2) and incompleteness effects (Section 4.3; technical details are given in Appendix C). In Section 5, we discuss the problem of redshift distortions and the way to account for it during the MAK reconstruction. Section 6 is devoted to the handling of finite-volume and edge effects, i.e. issues related to the zone of avoidance (ZOA) (Section 6.1), the choice of the Lagrangian volume of the reconstruction (Section 6.2) and finally the so-called cosmic variance (Section 6.3). The last section (Section 7) of this paper investigates the effect of distance measurement errors on the comparison between reconstructed and

¹ Brenier et al. (2003) actually show the uniformity is even required to prevent singularities in the solution of the Euler–Poisson system of equations.

² We implemented a parallel version for shared-memory supercomputers and Message Passing Interface (MPI) clusters. On the Magique2 cluster, it needs 50 min on two processors to solve the assignment of 74 000 particles. The algorithm is already sparse, i.e. it only looks for candidates for assignment in a limited region of the catalogue. The MPI efficiency is here optimal using two processors. It must be noted that the time complexity depends highly on the catalogue that is being reconstructed. For a given catalogue, the time needed to solve the assignment problem increases as $N^{2.25}$ with N the number of particles.

measured velocities, and proposes a maximum-likelihood estimator (Section 7.2) to account for them in the measurement of Ω_m . Results given by this estimator are then discussed in Section 7.3.

2 MOCK CATALOGUES

To study various effects and systematic biases on the MAK reconstructed velocity field, we generated a number of mock catalogues extracted from a N -body simulation (Section 2.1). Although many recipes will be employed later to address various observational biases, we will always start from the same three³ ‘main’ halo catalogues as described in Section 2.2. The first catalogue aims to reproduce to some extent the main features of the local universe, in particular the presence of a large cluster at about $40 h^{-1}$ Mpc and a supercluster at about $70 h^{-1}$ Mpc. The second and the third catalogues have less salient features but represent locally overdense and underdense realizations in order to address the problem of cosmic variance.

2.1 The N -body sample

Our 128^3 particles N -body sample (Mohayaee et al. 2006) was generated with the public version of the N -body code HYDRA (Couchman, Thomas & Pearce 1995) to simulate collisionless structure formation in a standard Λ cold dark matter (Λ CDM) cosmology. The sample covers a comoving volume of $200^3 h^{-3}$ Mpc³. The mean matter density is $\Omega_m = 0.30$ and the cosmological constant $\Omega_\Lambda = 0.70$. The Hubble constant is $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The normalization of the density fluctuations in a sphere of radius $8 h^{-1}$ Mpc is $\sigma_8 = 0.99$. We note that this value of σ_8 is significantly larger than the value suggested by present WMAP data which set $\sigma_8 = 0.74$ (Spergel et al. 2007), but this should not significantly affect the results presented in this paper. In fact, a lower σ_8 compared to 0.99 would reduce both non-linearities and cosmic variance effects, hence improving the quality of the measurements.

As the velocity field presents significant fluctuations on a larger scale than for the density field, one may worry about the small size of the simulation volume. We have checked, using linear theory, that the velocity dispersion in $200^3 h^{-3}$ Mpc³, for our cosmology, is 40 km s^{-1} . This value has to be compared to the typical errors appearing while doing velocity reconstructions to ensure that cosmic variance effects are negligible for our purpose.

2.2 The basic mock catalogues

To build mock catalogues, we have selected haloes from the N -body experiment using a standard Friend-Of-Friend algorithm with a traditional value of the linking parameter given by $l = 0.2$ (Efsthathiou et al. 1988). Haloes with less than five particles, i.e. with mass smaller than $M_{\min} = 1.62 \times 10^{12} h^{-1} M_\odot$, were discarded. Fig. 1 shows the good agreement between the measured halo mass function and the Sheth & Tormen (2002) model for haloes with $M \gtrsim M_{\min}$. However, about 63 per cent of the mass is not clumped in these haloes and is distributed in the *background field*. In realistic galaxy samples such as the NBG-8k or the Two Micron All Sky Survey (2MASS) catalogue, the lower mass cut-off is of the order of $10^{11} M_\odot$, a value much smaller than our M_{\min} . To mimic galaxies with mass smaller than M_{\min} , as will be required in the following, we just use dark matter particles unassigned to any halo

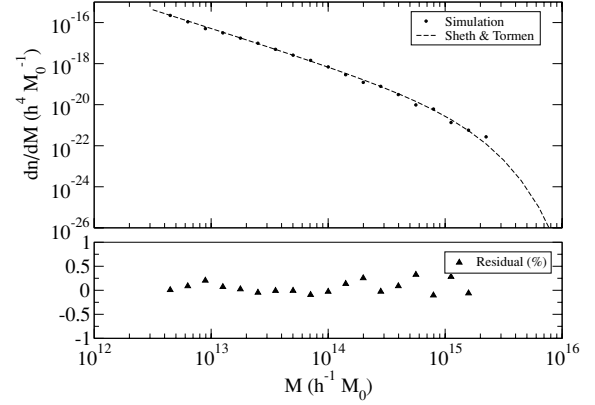


Figure 1. Sheth & Tormen mass function/diffuse mass. The top panel of this plot gives the number density of haloes in a mass bin as a function of the mass. The round points give the measurement of this function in the halo catalogue whereas the dashed line is obtained using the Sheth & Tormen (2002) theory. The residuals between the prediction and the measurement are given in the lower panel (relative differences). Most of the time, the points are within a few per cent of the theoretical prediction.

as tracers. The catalogue containing all the haloes and all the field particles will be called *FullMock*. One could here worry that the N -body sample that we are using has a too low resolution as the spatial distribution of small haloes is biased but not the particles of the background field. We have actually checked that using a 512^3 N -body sample with nearly the same cosmology [the simulation is described in Colombi, Chodorowski & Teyssier (2007)] does not change any measurements presented in Section 3.

Out of *FullMock*, we have extracted three spherical cuts of radius $40 h^{-1}$ Mpc (hereafter denoted by *4k-mockX*), where the velocity–velocity comparisons are conducted, and twice deeper counterparts (hereafter denoted by *8k-mockX*) are used to give better constraints (Section 6.2) on the reconstruction within the volume of analyses. Each of these catalogues is centred in a different place in the simulation such that

(i) *4k-mock6* is mildly overdense, with an effective mean matter density $\Omega_{\text{eff}} = 0.35$, and contains 495 haloes. It is designed in such a way that large voids and large concentrations of matter (clusters or superclusters) are present near its boundaries, similarly as found in real redshift catalogues of our local neighbourhood, such as the UZC (Upated Zwicky Catalog; Falco et al. 1999), the NBG-3k (Shaya, Peebles & Tully 1995; Tully et al. 2007) and the NBG-8k. This catalogue and its deeper counterpart, *8k-mock6*, are particularly suited to address edge effects on the NBG-3k (which terminates at Hydra and Centaurus clusters) and the NBG-8k (which stops at the Great Wall), respectively.

(ii) *4k-mock7* is highly overdense, with $\Omega_{\text{eff}} = 0.50$, and contains 656 haloes. Very little mass has come in and out of this volume: it behaves somewhat like an isolated universe, with small external tides.

(iii) *4k-mock12* is underdense, with $\Omega_{\text{eff}} = 0.19$, and contains 213 haloes. It presents as well a low level of density fluctuations along its boundary.

While there is no ambiguity in setting up a 128^3 MAK mesh when using all the haloes and the background particles (such as in *FullMock*), it is less trivial to consider lower resolution meshes that will be used in some of the subsequent analyses. Indeed, the number of mesh elements assigned to each tracer is not necessarily

³ The computationally high cost of the reconstruction considerably limits the number of possible realizations.

an integer anymore. Appendix A details the general procedure used to associate elements of the MAK mesh to each tracer.

3 ERRORS IN MAK VELOCITIES

Before going over observational issues, we address errors intrinsic to MAK reconstruction. First, there is scatter in the reconstruction of the displacement field itself which is expected to be rather small (Mohayaee et al. 2006). Second, there is scatter due to the Zel'dovich approximation one uses to convert a displacement field into a velocity field and to deal with redshift distortions. An accurate knowledge of the distribution of errors on the reconstructed velocities is eventually required for the likelihood analysis we want to introduce in Section 7.2. In this section, we measure such a distribution in real space while redshift space will be addressed in Section 5. In principle, the width of such a distribution is expected to increase when observational biases are taken into account while its shape should not change significantly.

We consider, in this section, reconstructions based on the catalogue *FullMock*, for which periodic boundary conditions are applied to avoid edge effect problems. We also assume that we know the mass of all of described catalogue objects (haloes and individual particles). Our subsequent reconstructions have a resolution within 64^3 and 128^3 mesh elements. We will thus present two reconstructions obtained on two different initial MAK mesh, 128^3 and 64^3 , obtained using the procedure presented in Appendix A. The results on the reconstructed displacement field are given in Fig. 2. These plots give the distribution of differences, P_{DE} , between the line of sight component of the reconstructed displacement field and the ‘exact’ one, given by the simulation.

The dot-dashed and dashed curves correspond to a least-square fit of the function P_{DE} corresponding to the 128^3 reconstruction, respectively, with a Gaussian fit, and a Lorentzian fit given by

$$P_{Lor}(x) = \frac{1}{\pi B} \frac{1}{1 + \frac{x^2}{B^2}}. \quad (4)$$

Examination of Fig. 2 supports the Lorentzian approximation with $B = 35 \text{ km s}^{-1}$, which reproduces better the long tails of P_{DE} than the Gaussian.

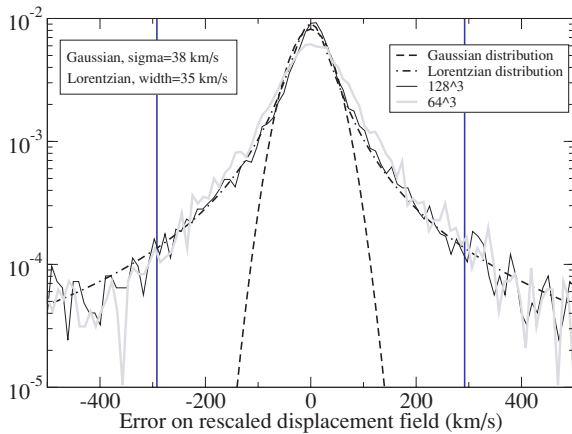


Figure 2. Error in reconstructed displacements. This plot displays the probability distribution of the quantity $\beta(\Psi_{r,rec} - \Psi_{r,sim})$ measured in *FullMock* (solid curve), where $\Psi_{r,rec}$ and $\Psi_{r,sim}$ are the line-of-sight component of the reconstructed and simulated displacement fields, respectively, after choosing an observer at the centre of the simulation box. The dashed and dot-dashed curves give the best fit of a Gaussian and a Lorentzian distribution, respectively.

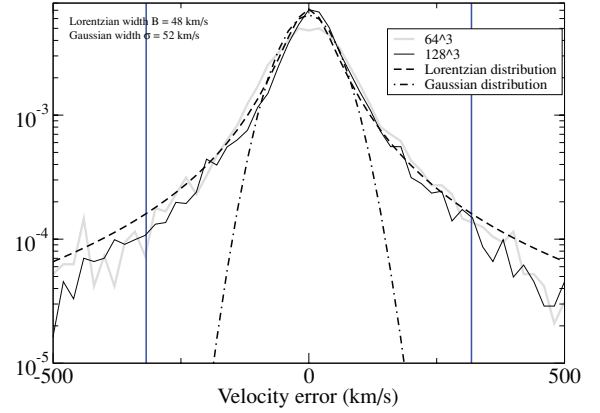


Figure 3. Error in reconstructed velocities. Same as in Fig. 2 but the solid curve corresponds to the probability distribution of the quantity $v_{r,rec} - v_{r,sim}$, where $v_{r,rec}$ and $v_{r,sim}$ are the line-of-sight reconstructed and simulated velocities, respectively.

The width, B , of P_{DE} is rather small compared to the line-of-sight dispersion, $\langle \beta^2 \Psi_r^2 \rangle^{1/2} \simeq 292 \text{ km s}^{-1}$, as expected. Naturally, the function P_{DE} is slightly flatter and larger for the 64^3 case than for the 128^3 one. However, the far end tails of P_{DE} are the same for 64^3 and 128^3 . In this regime, the measurements are not influenced by the resolution of the grid used to perform the reconstruction but rather by the inability of MAK to reproduce the internal dynamics of massive, relaxed objects (Mohayaee et al. 2006).

Fig. 3 is similar to Fig. 2, but considers line of sight reconstructed velocities versus ‘exact’ ones. Although Zel’dovich approximation introduces extra noise as shown by a wider width of the distribution, P_{DE} remains roughly Lorentzian with a small width $B = 48 \text{ km s}^{-1}$. This error variance is grossly 25 per cent higher than the expected velocity field variance on the simulation volume (Section 2.1). We are thus not affected by cosmic variance effects that could have been induced by modes larger than the box size of the simulation.

These results are fully supported by the examination of Fig. 4. However, the lower panels of this figure show that the joint distribution $P(v_{sim}, v_{rec})$ presents non-trivial tails above the diagonal line in the lower left quadrant and below the diagonal line in the upper right quadrant, respectively. These tails do not disappear even after smoothing of the velocity field with a $5 h^{-1} \text{ Mpc}$ Gaussian window. This is due to non-linear features in the dynamics not taken into account by our MAK+Zel’dovich prescription, which produces a slightly smoother velocity field than the real one. As a result, upper left-hand panel of Fig. 4, which corresponds to the reconstruction, is less contrasted than the upper right one, which corresponds to the simulation.

These non-linear tails give a propeller shape to $P(v_{sim}, v_{rec})$ which is susceptible to inducing a small bias on the final velocity–velocity comparison. For instance, one can estimate the slope of the lower left scatter plot of Fig. 4 using the ratio $s = \sigma_{v,rec}/\sigma_{v,sim}$, where $\sigma_{v,rec}^2$ and $\sigma_{v,sim}^2$ are the variances of the reconstructed and simulated velocity fields, respectively. In this case, the estimated β is biased to higher values by about 7 per cent. However, visually inspecting the scatter shows no measurement bias should occur if only the central part of the scatter is used for the computation. To achieve this, we have first applied an adaptive SPH filter on the scatter plot to produce a Probability Density Function (PDF), which is probed by the scatter in the points, on a regular mesh grid. We then compute the 1.5σ isocontour which encloses the region where the integrated PDF is equal to 68 per cent. This procedure has already been used in

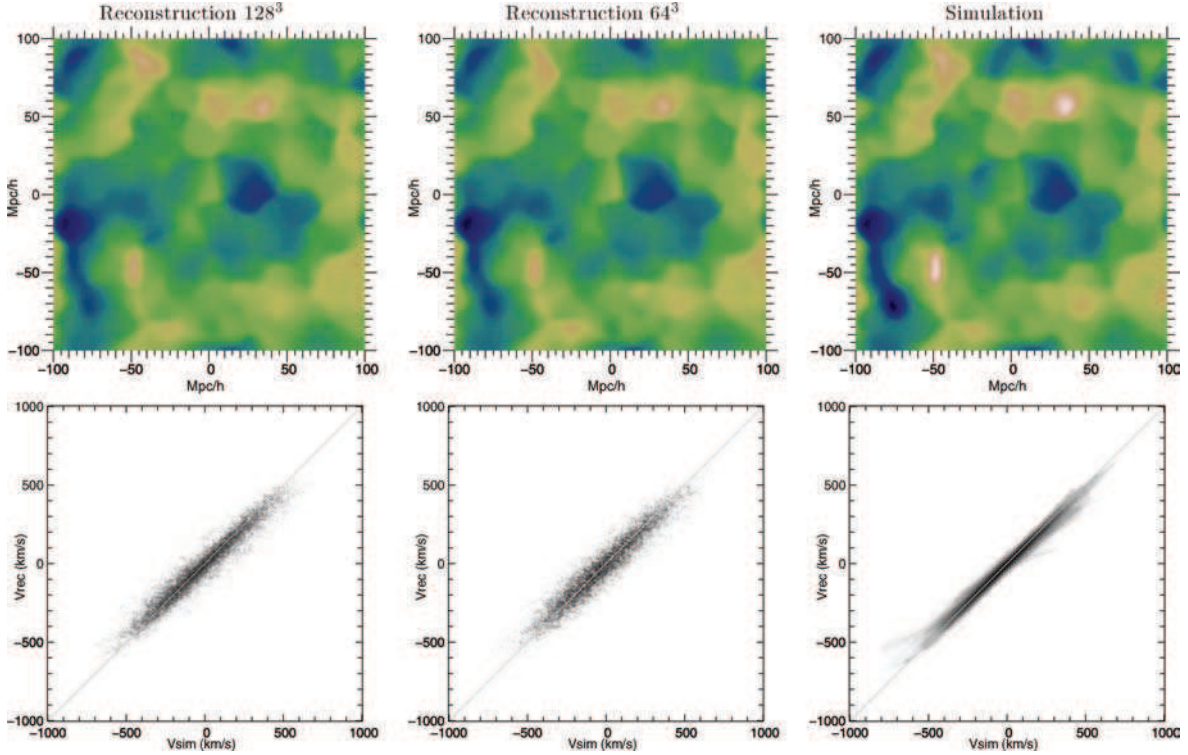


Figure 4. Velocity field reconstruction on FullMock. Top panels: a slice of the line-of-sight component of the simulated velocity field, $v_{r,\text{sim}}$, and the reconstructed one, $v_{r,\text{rec}}$, after smoothing with a $5 h^{-1}$ Mpc Gaussian window. The observer is at the centre of this slice. Bottom panels: scatter plots between $v_{r,\text{sim}}$ and $v_{r,\text{rec}}$ for individual haloes (left-hand panel) and after smoothing (right-hand panel).

Colombi et al. (2007) for the gravity–velocity comparison with total success. Only the points enclosed by the 1.5σ isocontour are used to compute the new $s_{\text{med},68}$ coefficient. The β parameter deduced from $s_{\text{med},68}$ is now statistically unbiased. Similarly, we define two other slope estimators $s_{\text{min},68}$ and $s_{\text{max},68}$ whose relevance is discussed in Appendix B. In this paper, until Section 7, we will only discuss the measurement of Ω_m obtained through the estimation of $s_{\text{med},68}$. The Ω_m obtained by this method is identified by a ‘ 1.5σ ’ to make a difference with the one obtained through the likelihood analysis that will be established in Section 7 and which is identified by a ‘ \mathcal{L} ’ in the tables and figures. A test of this method on a simulated scatter distribution, whose shape is built on analysis of reconstruction errors, is detailed in Appendix D.

4 MASS-TO-LIGHT ASSIGNMENT

Most reconstruction methods, including ours, infer the total matter distribution as a function of the visible matter distribution traced by galaxies. The fundamental assumption one usually makes is that the relation between these two distributions is highly deterministic. In other words, one assigns to each galaxy of a given luminosity L a dark matter concentration (a halo) of mass $M = f(L)$. However, there are several issues in this procedure:

(i) *Mass-to-light ratio.* The choice of a function $f(L)$ considerably influences the results and is expected to introduce significant bias on the measured β if performed unwisely. Now, the function $f(L)$ is coarsely determined (Marinoni & Hudson 2002, hereafter MH; Tully 2005) from direct measurements in observations. One way to infer this function is to rely on semi-analytic models of galaxy formation, but this represents a very strong prior on the measurements.

Furthermore, $f(L)$ remains a mean relation around which there can be some significant scatter. This dispersion can as well introduce some significant biases.

(ii) *Missing tracers/Magnitude limitation.* Even if function $f(L)$ is perfectly known, fainter galaxies are still missing in the catalogues due to the limitations of observational instruments. For instance, in magnitude-limited catalogues, the number density of detected galaxies decreases with distance from the observer. These missing tracers have unknown positions and correspond to a part of the dark matter distribution which is totally undefined. This missing mass has to be taken into account in some way.

In what follows, we will first address the second issue in a very simple way which assumes that the function $f(L)$ is well known (namely the masses of dark matter haloes themselves) but there is a fixed low-mass cut-off. The problem then consists in determining the unknown part of the dark matter distribution (namely the particles unassigned to any halo). Clearly, it is correlated with the detected mass tracers but less clustered. There are two following extreme ways to locate this missing mass:

- (a) associate it with the existing tracers as usually done with the analysis of real observations,
- (b) associate it with a uniform background.

Of course, the real solution is somewhat intermediate between (a) and (b) as will be shown in Section 4.1.

Then, we turn in Section 4.2 to the issue of the choice of $f(L)$. In this paper, we prefer to be as free as possible from strong priors so we deliberately do not use results from semi-analytic models of galaxy formation. Instead, we use determinations of $f(L)$ from observational data but, unfortunately, there are large uncertainties in

these measurements. The point here is to quantify, quite heuristically though, the effect of these uncertainties, random or systematic, on the measurement of β . Indeed, one is confronted with both a possibility of a wrong approximation of $f(L)$ and most probably a large scatter around this mean relation.

In sufficiently deep galaxy catalogues, the effect of the missing tracers is expected to be negligibly close to the observer and, in general, to increase with the distance from the observer. With appropriate weighting of the data, one can minimize the bias brought by the procedure used to infer the missing mass distribution far from the observer. In Section 4.3, we will illustrate this point by considering the case of a magnitude-limited catalogue where all the missing mass is associated with the existing tracers [Method (a) above].

4.1 Missing tracers

Fig. 5 shows the expected fraction of the total mass below a fixed threshold as a function of Ω_m , using the Sheth & Tormen (2002) model (see also Fig. 1). The solid line corresponds to the mass cut-off of haloes in *FullMock* and agrees, as expected, with the measurement in the simulation for $\Omega_m = 0.30$. Here, 63 per cent of the mass is outside of the haloes, which represent our ‘galaxies’ with known M/L ratio. The particles not linked to the haloes represent the missing mass. In Fig. 4, their exact location was used to perform the reconstruction. The only information available now is the distribution of ‘visible galaxies’. The missing mass needs to be redistributed using only these pieces of information. We propose two extreme ways to do so.

I. *All missing mass to background.* Prior to the reconstruction, the missing mass is divided into particles which are randomly put in the catalogue following a Poissonian distribution. In the example illus-

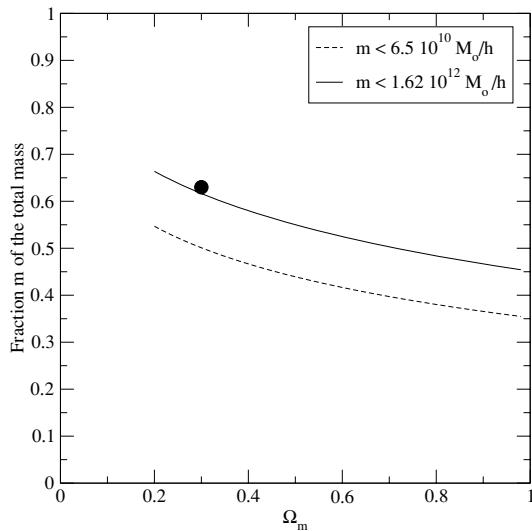


Figure 5. Diffuse mass. In this plot, we represent the fraction of the clustered mass below two mass resolutions for a standard Λ CDM type cosmology ($h = 0.65$, $\sigma_8 = 0.99$). We used a power spectrum as proposed by Bardeen et al. (1986). The curvature of the Universe is kept flat while Ω_m varies. This fraction is plotted for mass resolutions: $2.5 \times 10^{12} M_\odot$ (corresponding to the lower mass limit of haloes in our simulation) and $10^{11} M_\odot$ ($\approx 10^9 L_{\odot,B}$). The unclustered fraction in *FullMock* is given by the back filled circle. The fraction of mass below both of these limits is still considerable.

trated by the right-hand panels of Fig. 6, we choose for simplicity particles of the same mass as those in the simulation.

II. *All missing mass in haloes.* The missing mass is attributed to the existing haloes in proportion to their masses, as illustrated by left-hand panels of Fig. 6. This approach is equivalent, in real observations, to multiplying the M/L ratio of galaxies or group of galaxies by a constant $\alpha > 1$.

Obviously, in Case I, the screening effect due to the background is exaggerated, hence the reconstructed velocity is less contrasted and β is overestimated to compensate for this. In Case II, on the other hand, the potential wells are more contrasted than they should be, which leads to the opposite effect. At this point, it is extremely tempting to try to find a simple compromise between Cases I and II as illustrated by middle panels of Fig. 6 where 60 per cent of the missing mass was linked to the tracers and the remaining to a uniform background. With this particular choice of the redistribution, the match between the reconstructed and the simulated velocity fields is spectacular. This result is non-trivial given the simplicity of the handling of this 63 per cent missing mass all the more since the scatter on the middle lower panel of Fig. 6 is of the same order of that of the lower left panel of Fig. 4, where all the tracers contribute optimally.

Although the choice of the optimal redistribution remains a priori unknown in a real galaxy catalogue, one can at least infer error bars from Cases I and II. In that framework, Fig. 6 unfortunately provides quite a bad constraint on β , $0.36 \lesssim \beta \lesssim 0.85$. However, in real galaxy catalogues, such as the NBG-3k or the NBG-8k, the minimum luminosity is of the order of $10^9 L_\odot$. This corresponds to a less abrupt mass cut-off, $M_{\text{cut}} \sim 10^{11} M_\odot$, than in Fig. 6, where $M_{\text{cut}} = 2.5 \times 10^{12} M_\odot$. Therefore, one expects the problem of missing mass to be less salient in real observations, as illustrated by the dashed curve of Fig. 5. Furthermore, an appropriate use of mock catalogues can help at calibrating the redistribution of mass, as performed in middle panels of Fig. 6.

4.2 Mass-to-light ratio

To test how the choice of mass assignment to galaxies or group of galaxies affects the results, we consider the three following cases, as summarized in Fig. 7:

(i) *T-C case.* A galaxy catalogue is extracted from *FullMock* by associating a luminosity $L(M)$ to each dark matter halo or background particle using Tully’s latest best fit of the group mass–luminosity relation (Tully 2005, see fig. 8)

$$\frac{L_B}{L_\odot} = 2700 \left(\frac{M}{M_\odot} \right)^{0.59} e^{-6 \times 10^{11} M_\odot / M}, \quad (5)$$

which gives the luminosity in the B band for groups in the mass range $10^{11} - 10^{15} M_\odot$. Then, a new mass is given to each tracer assuming

$$M/L = \text{constant}, \quad (6)$$

as often used in the literature, and MAK reconstruction is performed on a resampling of this mass distribution.

(ii) *T-MH case.* A less extreme case than assuming $M/L = \text{constant}$ consists in separating the tracers in three broad classes: faint galaxies, luminous galaxies and group/clusters of galaxies, as performed by MH. To do this, they used a simple mapping between the Schechter luminosity function and the Press–Schechter mass

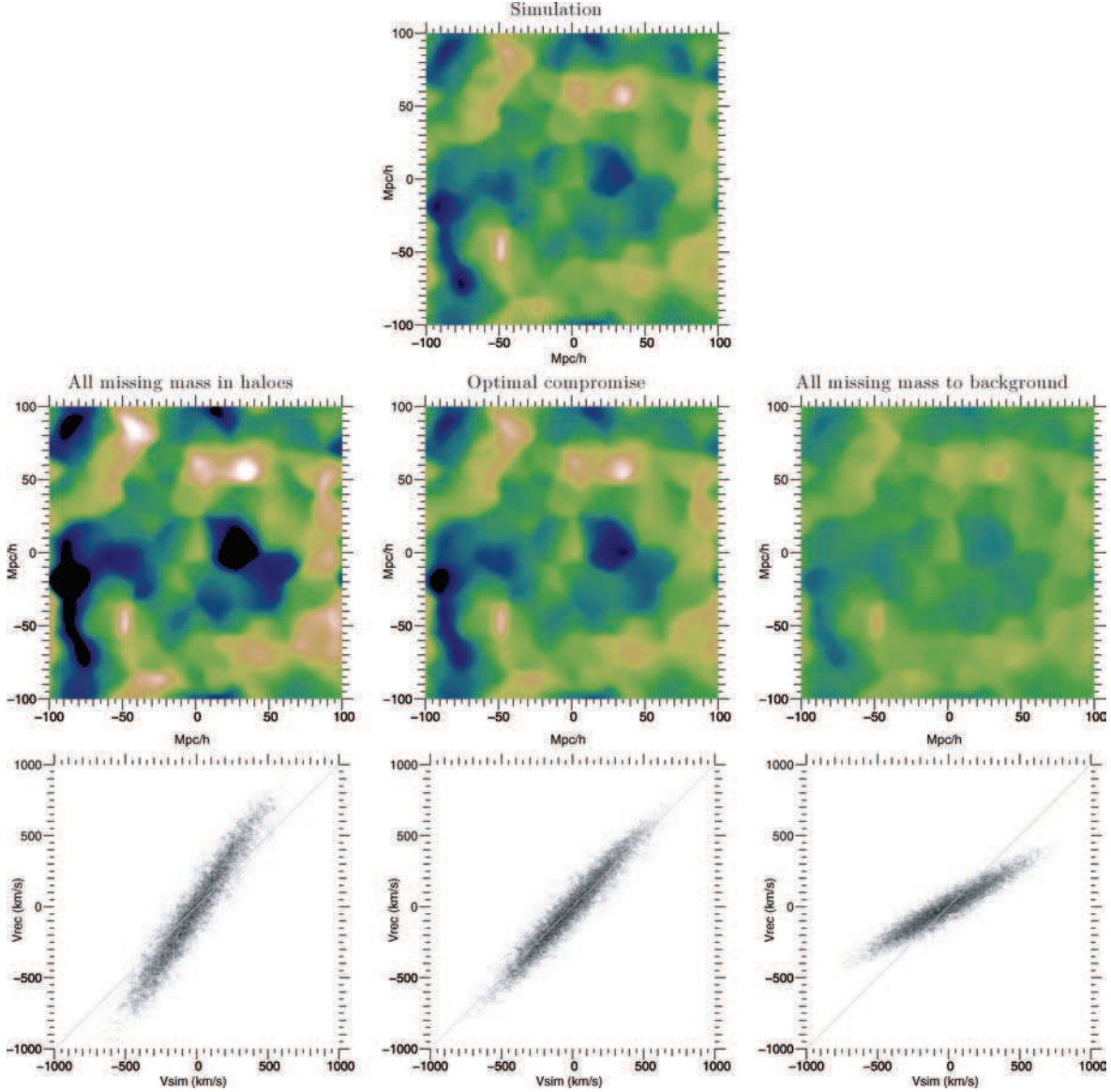


Figure 6. Diffuse mass correction. The top panel gives a slice of the line-of-sight component of the simulated velocity field, after smoothing with a $5 h^{-1}$ Mpc Gaussian window. The observer has been put at the centre of this slice. The second row of panels represents the line-of-sight component of the reconstructed velocity field, smoothed in the same way, for different corrections of the diffuse mass. The third row of panels gives the scatter distribution of individual reconstructed velocities of haloes *versus* simulated ones. The left-hand panels give the result of a reconstruction on a mock catalogue which only contains the haloes and not the background field but at the same time conserves the total mass of the catalogue by reassigning the missing mass to the haloes. The right-hand panels give the result for a reconstruction based on a mock catalogue for which the missing diffuse mass is represented by a background field composed of particles placed randomly in the catalogue. The centre panels give the result of a reconstruction on a mock catalogue which only contains the haloes and a random background field. The mass that has been initially removed from the mock catalogue (the background ‘galaxies’) is reassigned as follows: 60 per cent to haloes and 40 per cent to the background.

function that reads as follows:

$$\begin{aligned}
 M/L &= 1.15 \cdot 10^7 \left(\frac{L}{L_\odot} \right)^{-0.5} h \frac{M_\odot}{L_\odot} & \frac{L}{L_\odot} < 4 \times 10^{10} \\
 M/L &= 128 h \frac{M_\odot}{L_\odot} & 4 \times 10^{10} < \frac{L}{L_\odot} < 4 \times 10^{11} \\
 M/L &= 3.6 \cdot 10^{-4} \left(\frac{L}{L_\odot} \right)^{0.5} h \frac{M_\odot}{L_\odot} & \frac{L}{L_\odot} > 4 \times 10^{11}
 \end{aligned} \tag{7}$$

as shown in upper panel of Fig. 8. In this framework, we generated the same catalogue as in *T-C* case but it was analysed assuming the *M/L* function given by equation (7).

(iii) *TS-T case.* Assuming that we have an unbiased estimator of the *M/L* function, there can still be a scatter around this mean value that can increase the errors and also introduce systematic bias. We test this by multiplying the mass of each halo of *FullMock* by a random number x such that $\log_{10} x$ is uniformly distributed in $[-1, 1]$, prior to MAK reconstruction, which is performed on a resampling of the halo catalogue following the procedure explained in Appendix A. Note that the mass of background particles remains unchanged during the process, which corresponds to 63 per cent of the matter distribution being unaffected by the scattering. However, applying the scatter to small mass haloes only introduces a local additional noise which should not have any significant consequences

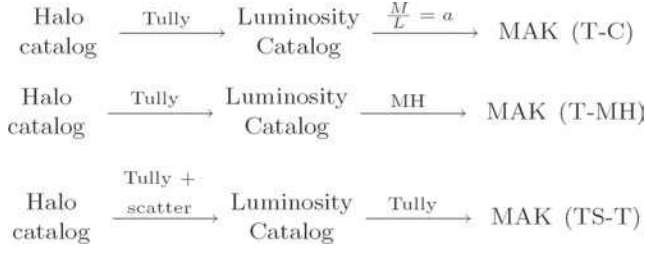


Figure 7. M/L assignment. Sketch of the procedures used to test the influence of a choice of a M/L assignment, as explained in the main text.

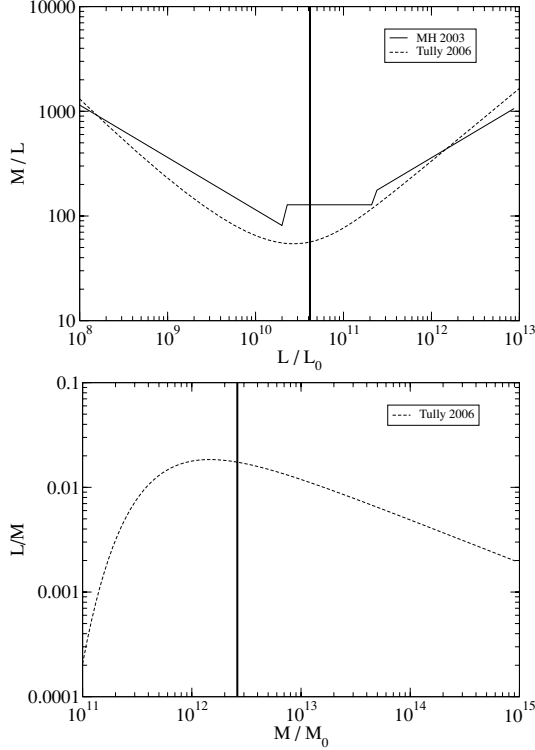


Figure 8. M/L function. The above two plots give the forward and inverse mass-to-light functions for both Tully (2005) and Marinoni & Hudson (2002) fits. The top panel gives the M/L as a function of the luminosity L , the bottom panel gives L/M as a function of the mass M .

on the reconstruction accuracy for which deeper potential wells are in fact more critical.

We want to highlight the fact that each of these transformations, actually corresponding to transforming the mass of an object of *FullMock* through a $M \rightarrow L \rightarrow M$ operation, does not correspond to an identity. One actually gets a new set of masses attached to each tracer which is different from the original one. Moreover, the output mass distribution $P_{\text{mass,out}}(M)$ may be fundamentally different from the input one $P_{\text{mass,in}}(M)$. Indeed, computing $P_{\text{mass,out}}(M)$ is equivalent to performing a weighted average of $P_{\text{mass,in}}(M)$. This procedure induces a global reshaping of the distribution. Consequently, the statistical properties of the corresponding mass density field may be affected.

More technically, during the procedure used to construct all the catalogues above, total mass conservation is enforced. Note that the total mass depends on $\Omega_m h^2$, but this normalization does not affect MAK displacements, which are sensitive to density contrasts only. Parameters Ω_m and h in fact intervene while performing velocity–

velocity comparison and while converting distances to velocities (Section 7), respectively.

Results are summarized in Fig. 9 and Table 1.

As expected, random uncertainty on the mass determination does not introduce any bias, it only increases the scatter in the measurements as illustrated by the lower left panel of Fig. 9. A more important issue is the global knowledge of the M/L relation. Indeed, it seems that the slope of this relation greatly influences the results, as illustrated by the middle and right-hand panels of Fig. 9. Clearly, if the galaxies follow the Tully formula (5), it is definitely wrong to assume constant M/L and even the MH fit introduces a significant bias, although it is well within the observational errors compared to equation (5). It must be noted that this bias can be turned into an advantage if one does not want to measure Ω_m but the M/L relation. Indeed, WMAP experiment (Bennett et al. 2003; Spergel et al. 2007) coupled with an analysis of the power spectrum of large-scale density of galaxies (Tegmark et al. 2006) gives good constraints on the real Ω_m now. Our method, on the other hand, is able to measure the discrepancy between the measured β and the expected growth factor $\beta_{\text{expected}} = \Omega_m^{5/9}$ (i.e. the bias). This measurement may give an idea of how wrong is the assumed M/L relation prior to the reconstruction and may push us to try different plausible M/L functions. Thus, our method is able to measure the way that the matter is distributed in the Universe once it is given its average density Ω_m . On the other hand, if the above bias is well understood, this method helps at reducing the degeneracy in the determination of cosmological parameters. Indeed, our posterior probability on (Ω_m, h) gives a constraint orthogonal (for example, see the results in Mohayaee & Tully 2005) to the one obtained from the WMAP experiment and from the galaxy statistics of the SDSS.

4.3 Magnitude limitation

Magnitude-limited sampling of mass tracers introduces a new type of problem: flux limitation decreases the mass resolution towards the outer edges of the catalogue contrary to the homogeneous case studied in Section 4.1. Usually, the incompleteness is handled by uniformly boosting the luminosities of galaxies at a given distance from the observer (Branchini et al. 2002), prior to conversion of luminosities into masses. This is a fair approach if $M/L = \text{constant}$, within the limits of the issues discussed in Section 4.1. However, this method is in general questionable for non-trivial M/L relations as in equation (5) or if different M/L s are assigned to galaxies with different types. In these last two cases, the missing mass correction should be applied to the mass distribution itself instead of the luminosity one, to avoid systematic errors on mass assignments, hence on reconstructed velocities. This unfortunately requires a prior assumption on the value of Ω_m , but only slightly complicates the analyses.

In the observational data, galaxies are separated into two populations: groups⁴ of galaxies (Tully 1987) and field galaxies. These two populations should be treated separately, keeping in mind that the groups are the most critical because their gravitational influence is much larger than individual field galaxies and they have better peculiar velocity measurements.

The full procedure consisting of creating a magnitude-limited mock catalogue and recovering the mass distribution is detailed in Appendix C. Let us recall that, in our mock catalogues, groups

⁴ Groups are defined here as compact sets of five galaxies or more.

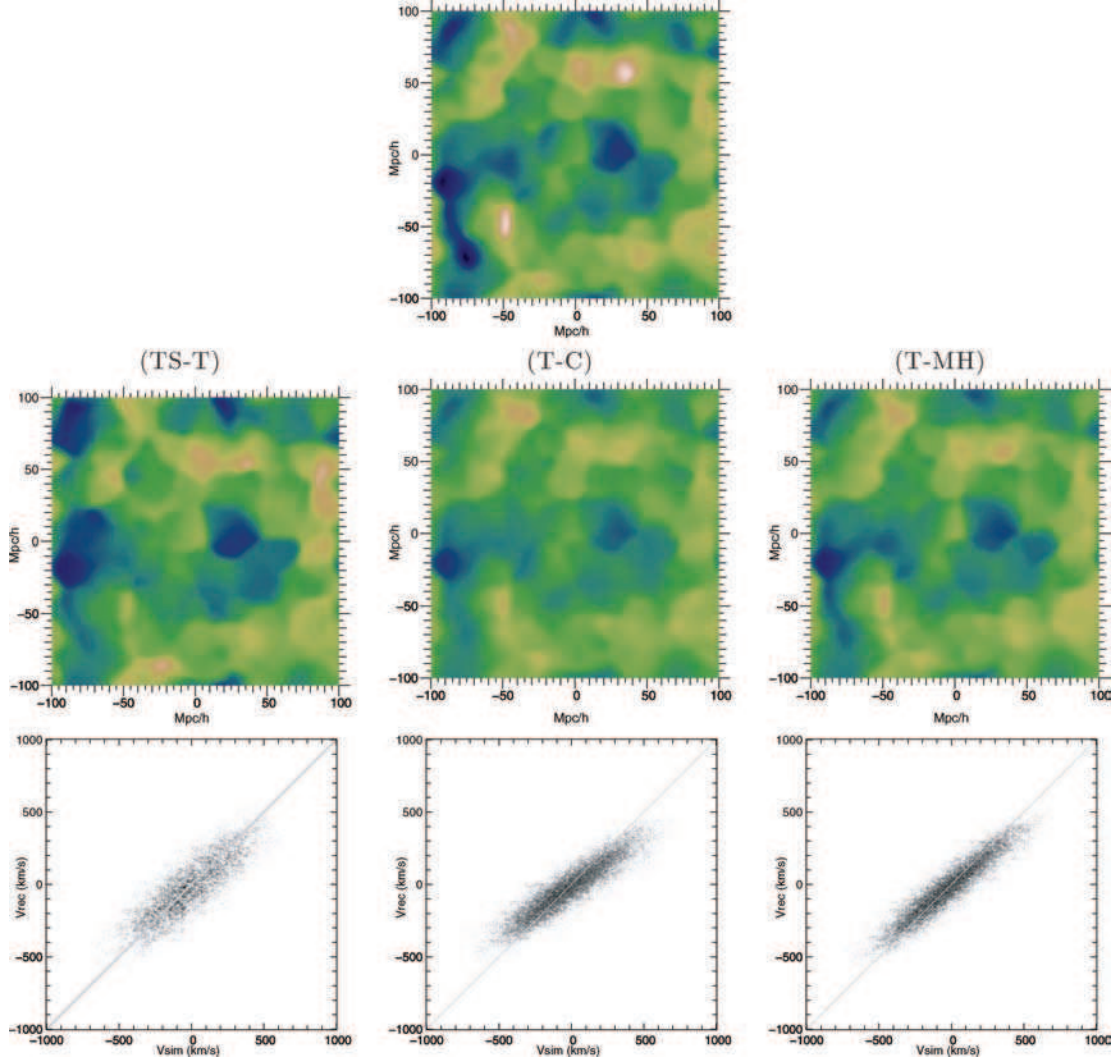


Figure 9. M/L bias. The top panel gives the expected line-of-sight component v_r of the velocity field, smoothed with a $5 h^{-1}$ Mpc Gaussian filter, as given by the simulation in a thin slice of the simulation containing the observer. The middle panels give the reconstructed v_r field, with the same smoothing, after having applied each of the transformations specified in Fig. 7 to *FullMock*. The lower panels give the scatter between the reconstructed and simulated peculiar velocities for each of the transformations.

Table 1. M/L bias effect. This table gives the results obtained using different statistical tools. We also measured Ω_m using six different methods: the label s means we used the slope estimated by using all objects, the label \mathcal{L} is used when Ω_m has been determined using the likelihood analysis and the label 1.5σ is used when the slope is estimated using only the objects within the 1.5σ isocontour of the PDF between reconstructed velocities and simulated velocities (method described in Section 3).

| Transf. | Velocity | | | $\Omega_m (s)$ | Ω_M | | | Ω_M | |
|---------|----------|------|----------|----------------|------------------------|------------------------|-------------------------------|-------------------------|-------------------------|
| | s | r | σ | | (\mathcal{L}_{\min}) | (\mathcal{L}_{\max}) | $(1.5\sigma, s_{\text{med}})$ | $(1.5\sigma, s_{\min})$ | $(1.5\sigma, s_{\max})$ |
| None | 0.88 | 0.89 | 0.58 | 0.38 | 0.30 | 0.31 | 0.30 | 0.28 | 0.31 |
| TS-T | 0.90 | 0.78 | 0.64 | 0.36 | 0.26 | 0.30 | 0.28 | 0.24 | 0.33 |
| T-MH | 0.80 | 0.80 | 0.60 | 0.45 | 0.33 | 0.38 | 0.36 | 0.32 | 0.40 |
| T-C | 0.71 | 0.78 | 0.63 | 0.55 | 0.40 | 0.48 | 0.44 | 0.37 | 0.51 |

of galaxies are simulated dark matter haloes with more than five particles while background galaxies are identified with dark matter particles unassigned to any halo. We list here the key steps we used to correct for incompleteness:

I. The total apparent luminosities of groups of galaxies is obtained assuming a global or a local Schechter luminosity distribu-

tion for the considered groups. The intrinsic luminosity is computed trivially from the total apparent luminosity and the redshift of the group.

II. The intrinsic luminosity of the remaining unbound galaxies (thus field galaxies) is also determined, straightforwardly.

III. Then, masses are estimated by assigning appropriate M/L to each object of Steps I and II.

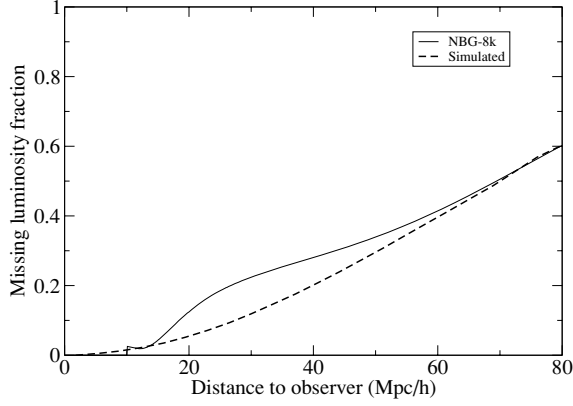


Figure 10. Magnitude limitation. Solid line: NBG-8k predicted luminosity incompleteness at the given distance from the observer. Dashed line: simulated luminosity incompleteness in *8k-mock6*. The incompleteness is expressed in terms of missing luminosity fraction at the specified distance.

IV. The local missing mass from undetected background galaxies is inferred from the detected mass distribution. This requires a prior on Ω_m .

V. This missing mass may either be reassigned locally to detected field galaxies of II (our choice) or be introduced by the mean of new randomly positioned tracers, as discussed in Section 4.1.

To examine the effects of systematics in the correction for incompleteness, we use *8k-mock6* and choose a flux limit such that the resulting mock catalogue has an incompleteness similar to NBG-8k, as shown in Fig. 10. Results are summarized in Fig. 11 and in Table 2.

The reconstructed radial peculiar velocities $v_{r,\text{rec}}$ are behaving extremely well. On average, the comparison between simulated and reconstructed velocity fields is surprisingly good in a volume of radius $80 h^{-1}$ Mpc, even though the edge misses locally 98 per cent of the field galaxies which represents 60 per cent of the total mass in our mock catalogue. It means that, though we keep only 2 per cent of the field galaxies, they suffice, in addition to the groups, for

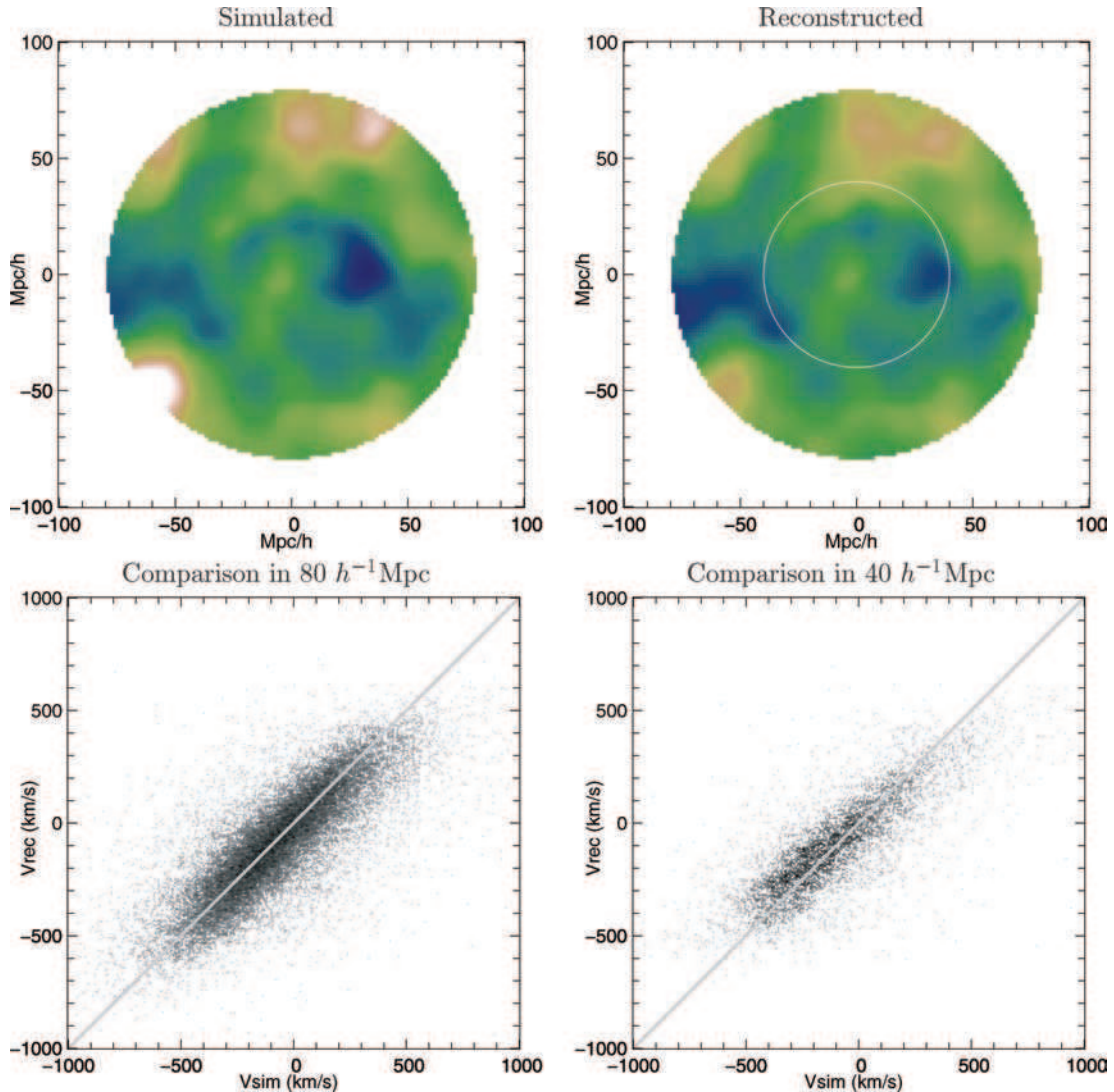


Figure 11. Incompleteness: magnitude limitation. Top panels: a slice of the line-of-sight component of the simulated velocity field in *8k-mock6* and the reconstructed one, after smoothing with a $5 h^{-1}$ Mpc Gaussian window. The displayed slice is chosen to include the observer in (0, 0). The white circle in the right-hand panel gives the size of the $40 h^{-1}$ Mpc sphere embedded in the $80 h^{-1}$ Mpc one. Bottom panels: the scatter plots compare the reconstructed and simulated velocities of objects in the $80 h^{-1}$ Mpc region (left-hand panel) and in the $40 h^{-1}$ Mpc volume (right-hand panel).

Table 2. Incompleteness: magnitude limitation. Column description is given in the caption of Table 1.

| Volume | Velocity field | | | Ω_m | Ω_m | Ω_m | Ω_m | Ω_m | Ω_m |
|--------|----------------|------|----------|------------|---------------------|---------------------|---------------------------------|---------------------------|---------------------------|
| | s | r | σ | (s) | (Σ_{\min}) | (Σ_{\max}) | ($1.5\sigma, s_{\text{med}}$) | ($1.5\sigma, s_{\min}$) | ($1.5\sigma, s_{\max}$) |
| 8k | 0.86 | 0.77 | 0.64 | 0.39 | 0.26 | 0.31 | 0.29 | 0.25 | 0.34 |
| 4k | 0.77 | 0.75 | 0.66 | 0.48 | 0.37 | 0.45 | 0.38 | 0.30 | 0.47 |

a reasonably fair recovery of the large-scale peculiar velocity field. Note the small bias in the scatter of the lower right panel of Fig. 11, resulting in a slightly larger $\Omega_m = 0.38$ than the expected value of 0.30, but in good agreement with the effective value of 0.35 expected in the corresponding volume (see Section 6.3 on cosmic variance effects). This bias might be the consequence of our treatment of the missing mass coming from undetected tracers as discussed in detail in Appendix C (point B).

5 REDSHIFT DISTORTION

The input of MAK reconstruction is the position of objects in real space as needed by equation (2). However, redshift catalogues give us galaxy positions in redshift space, namely $s_r = Hd + v_r$, where s_r is the redshift distance, d is the luminosity distance between the observer and the object and v_r is the line-of-sight peculiar velocity. To account for redshift distortions, we must correct for two major effects:

- (i) ‘Fingers-of-god’ correspond to an elongation of dense structures along the line of sight, such as clusters of galaxies, due to random motions of galaxies within these structures.
- (ii) Kaiser effect (Kaiser 1987) is a large-scale effect coming from the coherent part of the cosmic flows, which, for instance, increase the overall density contrast.

Fingers-of-god effects can be easily removed by simply collapsing groups or clusters to a single point, as usually performed in the literature. However, such a procedure is generally carried out in a rather ad hoc way and is certainly not free of biases.

The Kaiser effect can be accounted for by modifying the cost function (2) using the Zel’dovich approximation to infer line-of-sight peculiar velocities as functions of the sought displacement field (Valentine, Saunders & Taylor 2000; Mohayaee & Tully 2005). If $s(q)$ is the redshift coordinate of a particle originally at q , then the total cost (2) of the association σ becomes

$$I_\sigma = \sum_{i=1}^N \left[(s_i - q_{\sigma(i)})^2 - \frac{\beta(2+\beta)}{(1+\beta)^2} \frac{((s_i - q_{\sigma(i)}) \cdot s_i)^2}{||s_i||^2} \right], \quad (8)$$

where β is the linear growth factor. Once the redshift displacement $\psi^s = s - q$ has been computed, the reconstructed radial peculiar velocity of the object i can be obtained by

$$v_{r,\text{rec}}^s = \frac{\beta}{1+\beta} \frac{s \cdot \psi^s}{||s||}. \quad (9)$$

The cost function I_σ leads to the exact result in the case of a Zel’dovich displacement field without shell crossing after redshift distortion. However, in general, the second term (accounting for redshift distortion) of equation (8) becomes of the same order as the first term (the real space cost term) near the origin. In this case, the reconstruction becomes ill defined because of the loss of convexity of functional I_σ . We expect thus the central part of all catalogues to be, in general, poorly reconstructed. The size of

such a region is roughly determined by the magnitude v_{obs} of the large-scale flow nearby the observer with respect to the Cosmic Microwave Background. The velocity v_{obs} determines the relative contribution of the first term with respect to the second term of equation (8). In practice, v_{obs} is of the order of a few hundred km s^{-1} (for instance, the Local Group velocity is 630 km s^{-1} ; Erdoğan et al. 2006) which gives us a region of ‘exclusion’ of radius of about a few $h^{-1} \text{ Mpc}$.⁵

Again, MAK reconstruction fails in regions where shell crossings occur. Projection in redshift space generates such shell crossings along the line-of-sight. These shell crossings are dramatic because of their anisotropic nature. In particular, filaments can cross each other while passing from real to redshift space, implying the reconstruction will fail in a large region of the catalogue encompassing the gravitational influence of these filaments. In this area, most of the reconstructed radial velocities will have the opposite sign compared to the true velocity. Of course, shell crossings in redshift space can have more complex consequences but this simple example suggests that MAK reconstruction should not work as well in redshift space as in real space.⁶

Another problem of this method is that one must assume β prior to the reconstruction. As for Section 4.3, where we had to guess the undetected mass, we choose a value $\Omega_{m,\text{in}}$, thus an assumed β_{in} , then we make a redshift reconstruction and measure a $\Omega_{m,\text{out}}$. In practice, the ‘true’ Ω_m of the catalogue was chosen to be the one for which $\Omega_{m,\text{in}} = \Omega_{m,\text{out}}$, which corresponds to having self-consistent orbits modelling when doing MAK reconstruction and when one makes a comparison with measured velocities.

Fig. 12 shows both reconstructed and simulated velocity fields and the scatter between $v_{r,\text{rec}}^s$ and $v_{r,\text{sim}}$. The first impression when comparing the two top panels of Fig. 12 is that the redshift reconstruction behaves really well. However, some potentially worrying localized features are as present:

- (i) Some important structures have their velocities badly reconstructed. Two important examples are the green-yellowish finger just above the centre of the upper right panel of Fig. 12 and the big velocity peak at the top of this same panel. In the left-hand panel, these two structures are not so prominent. The difference can be understood by studying the impact of the Kaiser effect on the reconstructed velocity field. Basically, two nearby filaments can merge in redshift space and give birth to a filament with a higher apparent density. The reconstruction is not able to separate these two filaments, which leads to an area with higher reconstructed velocities than the true ones. Thus, we expect in observational data to meet problems in the neighbourhood of the Great Wall, which is a supercluster of filaments compressed by redshift distortion.

- (ii) The velocity field in the immediate ($5\text{--}10 h^{-1} \text{ Mpc}$) neighbourhood of the mock observer has lost its spatial structure and

⁵ See e.g. Colombi et al. (2007) for a similar discussion.

⁶ This is also true for the Least-Action method for which multiple solutions quickly arise.

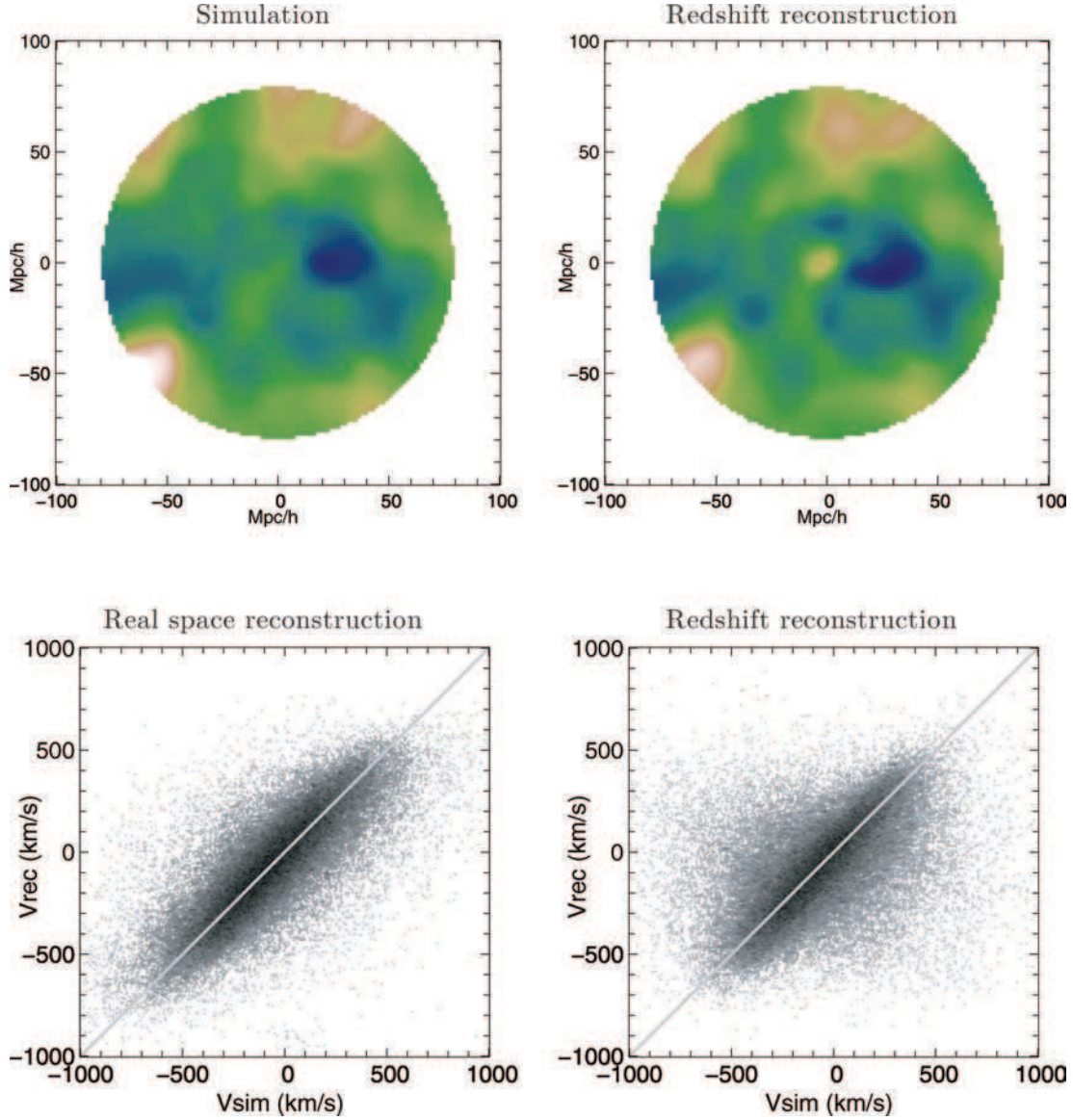


Figure 12. Redshift distortion correction. Top panels: a slice of the smoothed velocity field $v_{r,\text{sim}}$ and $v_{r,\text{rec}}$ is shown in the left- and right-hand panels, respectively. The two fields have been smoothed with a $5 h^{-1}$ Mpc Gaussian window, with objects put at their real (simulated and reconstructed) comoving coordinates. Bottom panels: scatter plots between v_r and $v_{r,\text{rec}}$ for individual mass tracers. The left-hand panel (right-hand panel) was produced using a real space reconstruction (redshift space reconstruction, respectively). In both cases, only objects within a sphere of 8000 km s^{-1} are shown.

even presents a spurious peak. This is, most unfortunately, an expected problem that is linked to the above discussion on the problems of I_σ near the observer. Indeed, in the neighbourhood of the observer, I_σ becomes singular and the reconstruction misses, most likely, the right orbits. Analysing the smoothed velocity field seems to show that this effect looks in practice much like the one just above: the reconstructed velocity field may be boosted by the merging of different structures in the neighbourhood of the observer.

(iii) The lower right panel presents two additional off-diagonal tails compared to lower left panel. As discussed earlier, these tails are due to shell-crossings occurring along the lines-of-sight when passing from real to redshift space. These extra shell-crossings result in some reconstructed velocities acquiring a sign opposite to the true velocities.

Similarly as in Section 3, we have computed in Fig. 13 the distribution of differences P_{VE}^s between $v_{r,\text{rec}}^s$ and $v_{r,\text{sim}}^s$ for a redshift

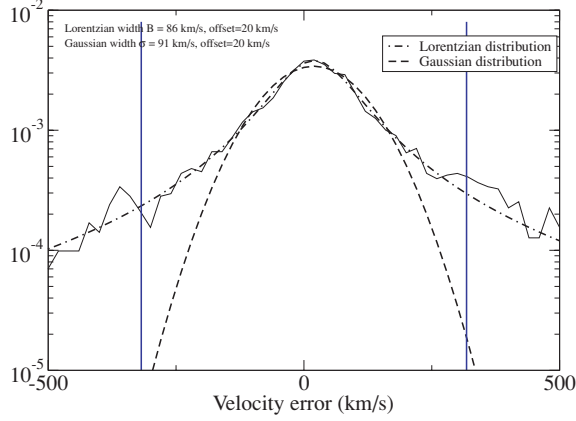
reconstruction applied on 8k-mock6 based on a 64^3 mesh.⁷ Though the distribution is of course wider than in Fig. 3, the previously drawn conclusions are still valid. P_{VE}^s is better fitted by a Lorentzian distribution with $B = 86 \text{ km s}^{-1}$ than by a Gaussian of width $\sigma = 91 \text{ km s}^{-1}$, particularly in the tails.

To check the effects of redshift distortion on the quality of the reconstruction, one can compare Table 3 to the first row of Table 1. As usual, the s parameter is slightly biased below unity due to non-linear effects discussed in Section 3, which seem, not surprisingly, to be slightly enhanced by redshift distortions. The appearance of the off-diagonal tails in the lower right panel of Fig. 12 increases the level of scattering, hence the correlation coefficient r decreases and the signal-to-noise ratio increases. Reducing the analysis to

⁷ The handling of the finiteness of the catalogue volume is handled in Section 6.2.

Table 3. Redshift reconstruction. Column description is given in the caption of Table 1.

| s | r | σ | $\Omega_m(s)$ | Ω_m (\mathcal{L}_{\min}) | Ω_m (\mathcal{L}_{\max}) | Ω_m ($1.5\sigma, s_{\text{med}}$) | Ω_m ($1.5\sigma, s_{\min}$) | Ω_m ($1.5\sigma, s_{\max}$) |
|------|------|----------|---------------|--|--|---|---|---|
| 0.83 | 0.46 | 0.95 | 0.50 | 0.22 | 0.29 | 0.27 | 0.22 | 0.33 |

**Figure 13.** Error distribution of the reconstructed velocity field, redshift space. Same as in Fig. 3 but the solid curve corresponds to the probability distribution of the quantity $v_{\text{r,rec}}^s - v_{\text{r,sim}}^s$, where $v_{\text{r,rec}}^s$ and $v_{\text{r,sim}}^s$ are the line-of-sight reconstructed and simulated velocities, respectively.

the region inside 1.5σ isocontour greatly improves the results, as expected, but still leads to a value of Ω_m slightly biased to lower values, $\Omega_m = 0.27$.

6 EFFECTS OF CATALOGUE GEOMETRY

In practice, real galaxy catalogues are not spatially periodic as is our simulation. They represent a region of finite volume with non-trivial geometry. In particular, two kinds of problems arise:

(i) *Edge effects.* Reconstruction of the galaxy trajectories without any piece of information on what may affect them dynamically from the outer parts of the catalogue is likely to introduce significant sources of errors, possibly systematic. We separate here edge effects into two subclasses: the effects of the obscuration by our galaxy, which defines a ZOA and the effects of finite depth of the catalogue. These two effects need a separate treatment detailed in Sections 6.1 and 6.2.

(ii) *Cosmic variance.* The finite volume of the accessible part of the Universe might be a potentially unfair realization of the random process underlying the properties of the large-scale matter distribution. We must investigate whether our method, including handling of edge effects, is robust to the recovering of the statistical properties of the whole Universe from observations of only a fraction of it.

6.1 Zone of avoidance

Dust present in the Milky Way’s galactic plane highly attenuates the light, thus galaxy catalogues generally do not provide any data in this direction (approximately the region within $|b| < 5^\circ$, where b is the galactic latitude) of the ZOA. This strong attenuation introduces a boundary effect, which has the unpleasant feature of being present at any distance from the observer and may thus severely affect the measurements. As this area is none the less relatively small, particularly at low redshift, a simple correction should be able to greatly remove the boundary effect in the inner region of the catalogue.

Simulating the effect is made easy by putting an observer at the centre of the simulation volume and by removing all mass tracers in the neighbourhood of the galactic plane $z = 0$, i.e. which have $|b| < \alpha$. This gives us *FullMockZOA*.⁸

Though more advanced ways of filling the ZOA exists (e.g. Lahav et al. 1994; Fontanot, Monaco & Borgani 2003), this latter is here sufficiently small to be dealt with by the following simple algorithm. Since the statistical properties of the galaxies should not change across the boundaries of the ZOA, the objects in its neighbourhood can be used to fill the zone. We build new mass tracers to fill the obscured area by applying a locally planar symmetry transformation to the galaxies and groups with $-3\alpha < b < -\alpha$ according to the ‘plane’ $-\alpha$. We execute the same operation on objects with $+\alpha < b < +3\alpha$ but according to the ‘plane’ $+\alpha$. In the end, the masses of the copied haloes in the ZOA are divided by two and we only take half of the field galaxies. This method has been used previously to fill the ZOA in NBG-3k (Shaya et al. 1995) and NBG-8k. This folding procedure has been applied to *FullMockZOA*, slightly moving some of the newly created objects to enforce the periodicity of the simulation box to avoid mixing the effect of the ZOA with other boundary effects. The results are presented in Fig. 14. As expected, the ZOA has a clear impact on errors of the reconstructed velocities.

The typical errors on the reconstructed velocities, represented in the left-hand panel of this figure, rise substantially in the vicinity of the obscured area. Fortunately, they remain well below the natural velocity dispersion of the simulation (dashed line). As we are comparing velocity fields filtered with a $5 h^{-1}$ Mpc Gaussian window, we expect the reconstructed velocity field to be nearly error free for all points nearer than about $60 h^{-1}$ Mpc.⁹ It is also fortunate we have not introduced an extra bias using the filling algorithm, as shown by both comparing Table 4 to the first row of Table 1 and looking at the scatter plot in the right-hand panel of the Fig. 14. We none the less highlight that the edge effect is not at all localized near the ZOA but extends quite far away and becomes negligible only for $|b| > 20^\circ$. Table 4 shows that the above extra noise does not have any impact on the measured Ω_m .

6.2 Lagrangian domain

The inputs to MAK reconstruction are the present coordinates of the objects, i.e. \mathbf{x} in equation (2) or s in equation (8), and the knowledge of the Lagrangian domain, i.e. \mathbf{q} in equation (2) or (8). Redshift catalogues give the present ‘positions’ of the objects, i.e. s in equation (8); however we have no observations that would give us the corresponding Lagrangian domain \mathbf{q} . We are thus limited to make guesses, though in the end, for huge catalogues, the details of the guess does not matter as gravitational forces are screened on large scales by the nearly homogeneous distribution of matter in the universe. Consequently, what happens at the boundaries should not

⁸ $\alpha = 5^\circ$ in our case.

⁹ This corresponds to taking a 5° wide ZOA and computing at what distance the window is smaller than the ZOA.

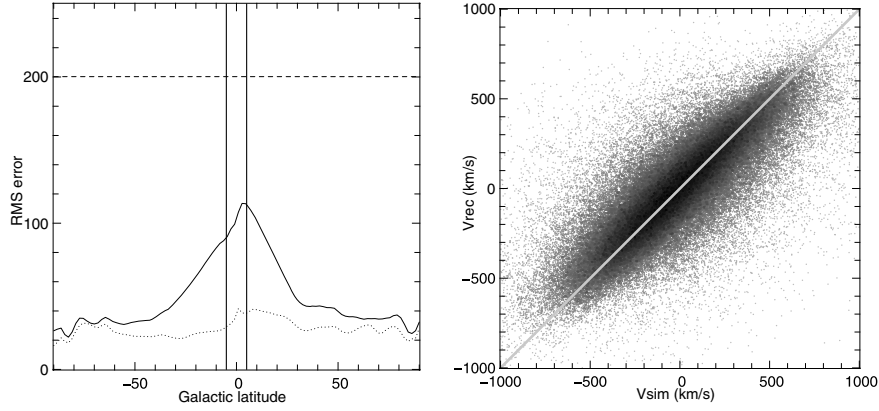


Figure 14. ZOA/velocity field. Left-hand panel: binned average rms error on the smoothed velocity field. As usual, the velocity field window has been smoothed with a $5 \text{ Mpc } h^{-1}$ Gaussian window. Each point is computed by averaging the square of the deviation of the velocity field along the line-of-sight and for all line-of-sights belonging to the same $\sin(b)$ bin, where b is the ‘galactic’ latitude. The solid line gives the rms error in the presence of a ZOA at $b = 0$. The dotted line gives the rms error for a reconstruction on a catalogue without ZOA. The dashed line gives the rms of the smoothed velocity field itself. Right-hand panel: scatter plots between $v_{r,\text{rec}}$ and $v_{r,\text{sim}}$ for individual mass tracers.

Table 4. ZOA. Noise and biasing summary. Column description is given in the caption of Table 1.

| s | r | σ | $\Omega_m(s)$ | $\Omega_m(\mathcal{L}_{\min})$ | $\Omega_m(\mathcal{L}_{\max})$ | $\Omega_m(1.5\sigma, s_{\text{med}})$ | $\Omega_m(1.5\sigma, s_{\min})$ | $\Omega_m(1.5\sigma, s_{\max})$ |
|------|------|----------|---------------|--------------------------------|--------------------------------|---------------------------------------|---------------------------------|---------------------------------|
| 0.89 | 0.79 | 0.61 | 0.37 | 0.30 | 0.35 | 0.32 | 0.285 | 0.36 |

strongly affect the central part of the catalogue though some guesses may be better at confining the edge effects on the boundaries. The naive solution is to assume that the Lagrangian domain is not so different from the volume of the catalogue itself. This assumption only begins to be a good approximation for volume enclosed in a sphere for which radius is big enough. For our $80 h^{-1} \text{ Mpc}$ sample, the mass going in and out of the volume (from initial to present time) already represents about 16 per cent of the total mass. For a $40 h^{-1} \text{ Mpc}$ sphere, the mass flow is even greater: it may vary between 30 and 63 per cent of the total mass depending on the 8k-mock catalogue considered. Though tidal field and cosmic variance effects become negligible on a $80 h^{-1} \text{ Mpc}$ scale, they still affect the boundaries of the Lagrangian domain of a given catalogue in a non-trivial way. As we will show, these problems are further enhanced by redshift distortion.

To achieve a meaningful comparison, we have run a reconstruction on 8k-mock6 using the Lagrangian domain given by the simulation; this reconstruction is called *TrueDom*. Now, we confront the results of *TrueDom* for two different reconstruction setups that try to recover the Lagrangian domain:

(i) *NaiveDom* reconstruction is obtained by assuming a naive spherical Lagrangian domain for 8k-mock6. In that case, all the mass that is presently in the 8k-mock6 catalogue was uniformly in a sphere of radius $80 h^{-1} \text{ Mpc}$. Equivalently, it means no significant mass flow must have gone through the comoving boundaries in the past.

(ii) *PaddedDom* reconstruction is obtained by padding homogeneously the 8k-mock6 catalogue. The padding is chosen such that the final MAK mesh that will be reconstructed is an inhomogeneous cube (as in the right-hand panel of the second row of Figs 15 and 16). The cube must be sufficiently big to absorb density fluctuations present at the boundary of the catalogue (typically a $20 h^{-1} \text{ Mpc}$ buffer zone is needed). With real data, we are bound to assume that

the catalogue is totally representative of the whole universe, i.e. its effective mean matter density is equal to Ω_m .

Fig. 15 shows the result of a *TrueDom*, *NaiveDom* and *PaddedDom* reconstruction applied to 8k-mock6 in the absence of redshift distortion. Fig. 16 gives the same reconstructions when applied to a redshift catalogue. Table 5 summarizes the value of the moments of $P(v_{r,\text{sim}}, v_{r,\text{rec}})$ for different cases. We will now first confront the results of real space reconstructions, and second redshift space reconstructions.

TrueDom reconstruction does not yield any significant bias at $80 h^{-1} \text{ Mpc}$. However, at $40 h^{-1} \text{ Mpc}$, cosmic variance effects introduce a notable systematic error in the direction of higher Ω_m that will be discussed in Section 5.3. Compared to *TrueDom*, *NaiveDom* gives good overall results though the central blue region of *TrueDom* turns to dark blue in *NaiveDom*, which would suggest the velocity field is biased. This analysis is confirmed by looking at the bottom scatter plot. The Ω_m measurement (Table 5) is underestimated by about 26 per cent even in the central region of the catalogue which is normally less affected by boundary effects. *PaddedDom*, on the other hand, does not yield such a sharp discrepancy in the middle of 8k-mock6, namely in the 4k-mock6 region. Both the bottom scatter plot and the Ω_m measurement confirm that the reconstructed velocities are nearly bias-free in the central region. As expected, the velocities in the neighbourhood of the boundaries are completely wrong for the two methods.

Now, the catalogues are cut in redshift space. Redshift distortion biases the velocity distribution of objects on the catalogue boundary: the catalogue receives more infalling objects than outfalling ones. In some cases, one may even find objects seemingly artificially separated from the main volume of the catalogue (they look ‘disconnected’). In those cases, the hypothesis of convexity is definitely lost for those objects. This problem will enhance boundary problems. The case of *TrueDom* reconstruction has been discussed in Section 5. As previously, the peculiar velocities in *NaiveDom* and

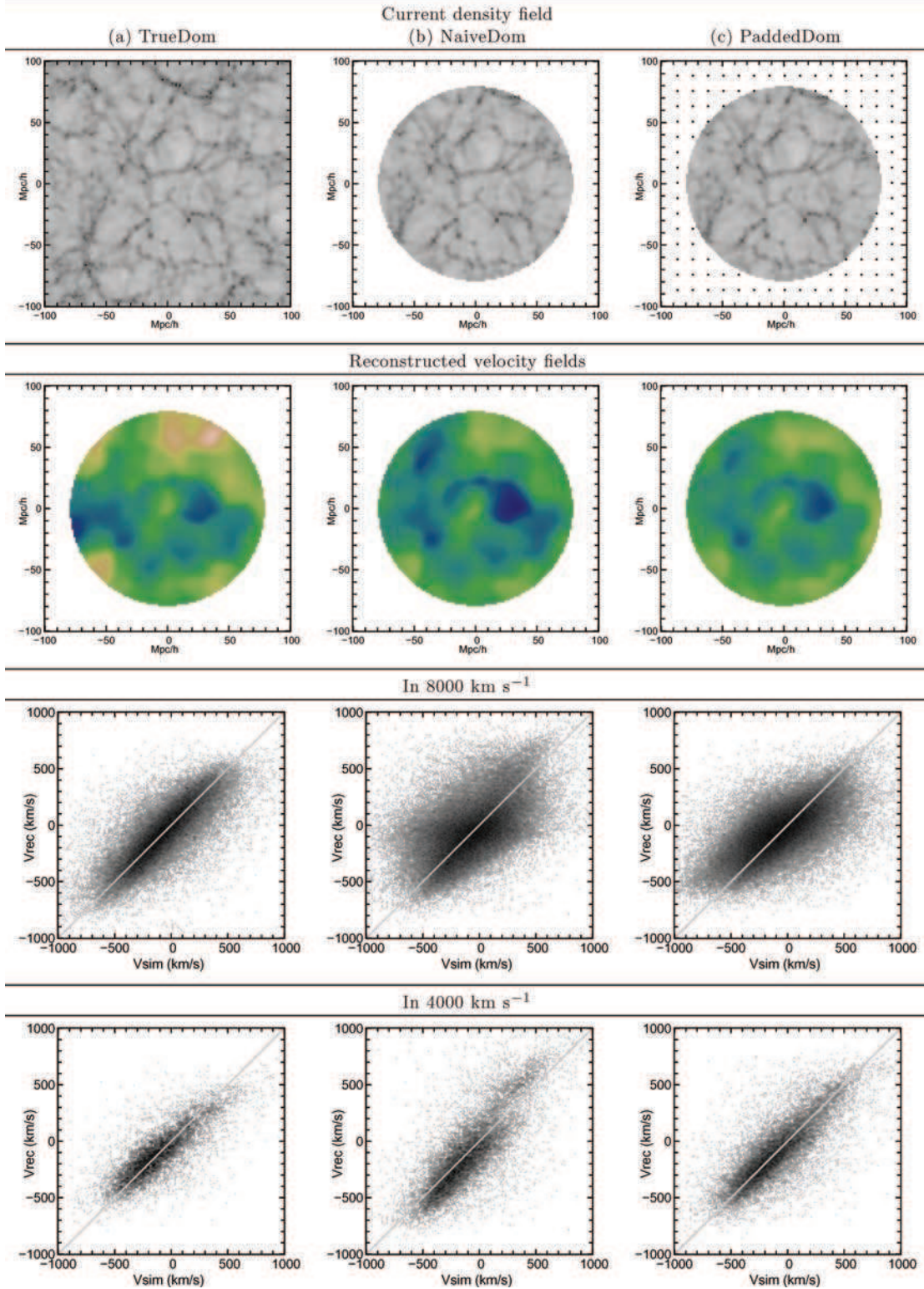


Figure 15. Lagrangian domain/without redshift distortions. This figure summarizes the results obtained on reconstructions that have limited information on the Lagrangian domain. The left-hand column illustrates the *TrueDom* reconstruction, the middle column the *NaiveDom* one and the right-hand column the *PaddedDom* one. For space occupation reasons, the original velocity field given by the simulation is not remembered but can be found in Fig. 11. The *top row* illustrates the three schemes for handling boundary effects on the density field: in the left-hand column, one retains information of large-scale tidal fields, in the middle column one cuts the catalogue spherically and does a reconstruction on it, in the right-hand column one pads the spherically cut catalogue with particles homogeneously distributed on a grid. The *second row* gives the reconstructed velocity field in each case, smoothed with a $5 h^{-1}$ Mpc Gaussian window as usual. The colour coding is the same as for the other figures, i.e. dark blue is -1000 km s^{-1} and white is $+1000 \text{ km s}^{-1}$. The *third row* compares the individual (not smoothed) reconstructed and simulated velocities of objects in the 8k-mock6 catalogue. The *fourth row* does the same comparison but objects lying only in the 4000 km s^{-1} region of the 8k-mock6 catalogue.

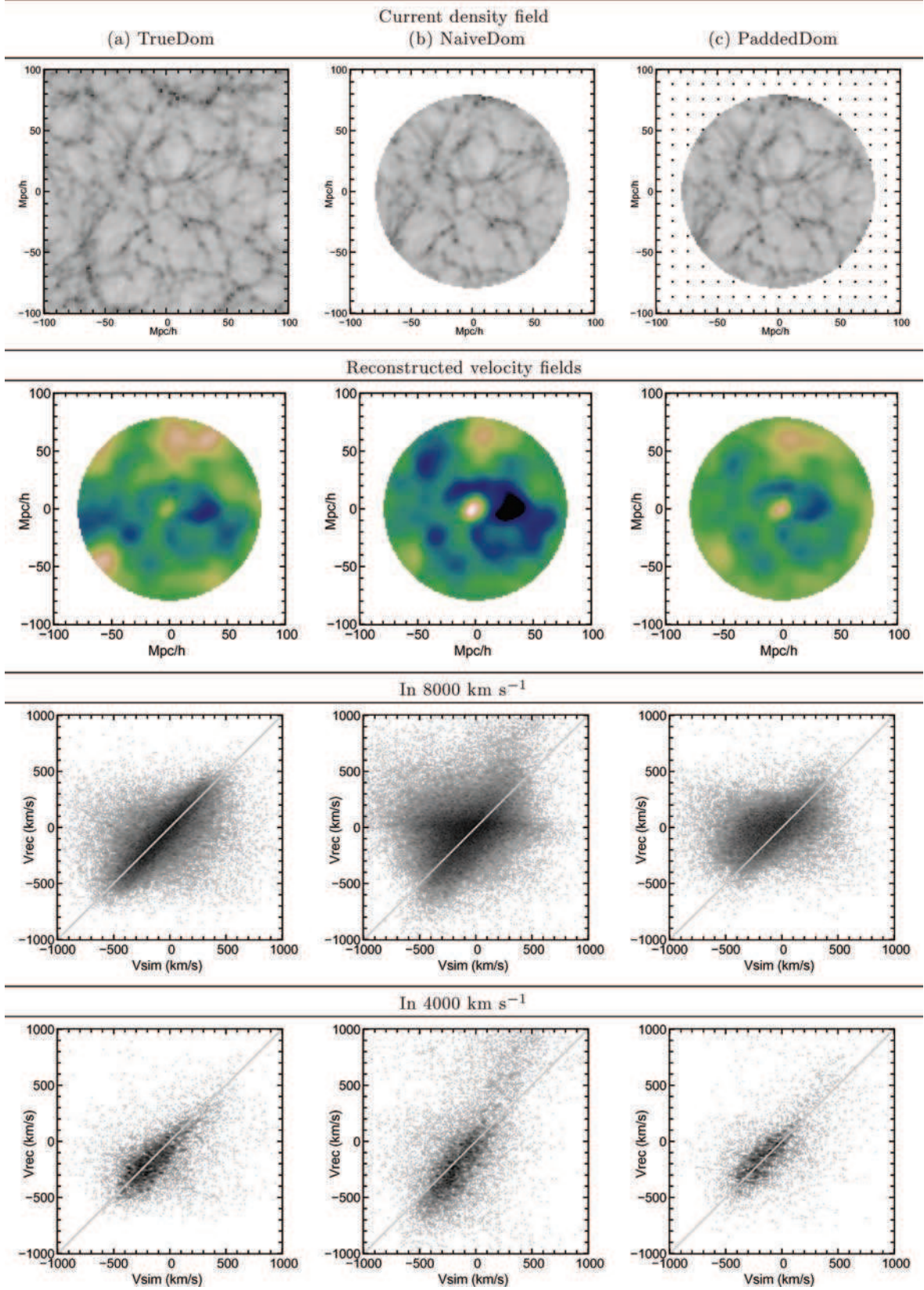


Figure 16. Lagrangian domain/with redshift distortions. Same as Fig. 15, but for mock catalogues including redshift distortion.

Table 5. Lagrangian volume. Residual error after the correction. Description for some columns is given in the caption of Table 1. ‘Radius’ gives the spatial size of the sphere on which the velocity–velocity comparison is conducted. ‘Reconstruction type’ indicates the type of Lagrangian domain reconstruction and whether it is mixed with redshift distortion effect. Details on the meaning of each name are given in Section 6.2.

| Reconstruction type | Radius (h^{-1} Mpc) | Velocities | | | Ω_m | Ω_m | Ω_m | Ω_m | Ω_m | Ω_m |
|--------------------------|------------------------|------------|------|----------|------------|--------------------------|--------------------------|---------------------------------|---------------------------|---------------------------|
| | | s | r | σ | (s) | (\mathcal{L}_{\min}) | (\mathcal{L}_{\max}) | ($1.5\sigma, s_{\text{med}}$) | ($1.5\sigma, s_{\min}$) | ($1.5\sigma, s_{\max}$) |
| TrueDom/Real space | 80 | 0.91 | 0.77 | 0.66 | 0.35 | 0.28 | 0.31 | 0.27 | 0.233 | 0.32 |
| | 40 | 0.80 | 0.76 | 0.65 | 0.45 | 0.28 | 0.38 | 0.35 | 0.28 | 0.43 |
| NaiveDom/Real space | 80 | 0.87 | 0.52 | 0.92 | 0.38 | 0.20 | 0.28 | 0.42 | 0.20 | 0.87 |
| | 40 | 1.11 | 0.77 | 0.73 | 0.25 | 0.20 | 0.24 | 0.244 | 0.19 | 0.31 |
| PaddedDom/Real space | 80 | 0.73 | 0.65 | 0.77 | 0.53 | 0.36 | 0.48 | 0.45 | 0.27 | 0.75 |
| | 40 | 0.91 | 0.77 | 0.64 | 0.35 | 0.28 | 0.34 | 0.32 | 0.26 | 0.38 |
| NaiveDom/Redshift space | 40 | 1.49 | 0.51 | 1.31 | 0.11 | 0.15 | 0.26 | 0.20 | 0.12 | 0.37 |
| PaddedDom/Redshift space | 40 | 0.93 | 0.53 | 0.94 | 0.36 | 0.18 | 0.34 | 0.38 | 0.20 | 0.79 |

in *PaddedDom* are largely uncorrelated in the full 8k-mock6 volume (Fig. 16). However, peculiar velocities reconstructed by *NaiveDom* are more strongly overestimated than by using *PaddedDom*’s, as shown in Table 5. For *NaiveDom*, the scatter is plagued by a horizontal alignment in Fig. 16, mid-lower panels, which is a signature of a strong edge effect. This spurious alignment was already present, though much less apparent, in the real space case. On the other hand, *PaddedDom* does not present this feature but only a large scatter. We have verified that objects belonging the horizontal alignment are essentially near the $80 h^{-1}$ Mpc boundary, contrarily to velocities reconstructed using *PaddedDom* which are more or less uniformly distributed and essentially uncorrelated to simulated velocities.¹⁰ This means that *PaddedDom* is at least better at screening edge effects than *NaiveDom* in the sense the errors are more evenly distributed and less systematic. Though impressively low in the last two rows of Table 5, the correlation coefficient r is actually spoiled by the long tails of the PDF shown in the scatter plots in Fig. 16. Concerning Ω_m , *NaiveDom* seems less robust to produce an unbiased estimation than *PaddedDom*. Indeed, looking at Table 5, one may note that the interval delimited by s_{med} , s_{\min} and s_{\max} nearly does not contain $\Omega_m = 0.30$ for *NaiveDom*/Real space/ $40 h^{-1}$ Mpc, and does not contain it at all for *NaiveDom*/Redshift space. On the contrary, $\Omega_m = 0.30$ is always selected by the three s parameters using *PaddedDom* reconstruction. In the rest of this paper, whenever it is needed, we will thus use the *PaddedDom* reconstruction.

6.3 Cosmic variance

We generally assume that galaxy catalogues give a fair representation of the whole universe, but of course we have no guarantee that this assumption is correct. Thus, the result of a MAK reconstruction may be affected by inhomogeneities above the catalogue scale. For instance, our galaxy may reside in a particularly extreme region (overdense or underdense), which would produce unusual

peculiar velocities. This effect, known as *cosmic variance*, can be investigated by our three original basic mock catalogues: 4k-mock6, 4k-mock7 and 4k-mock12 (Section 1). The cosmic variance effect is here further enhanced by the finiteness of the sampled volume. The volume is sufficiently small here to have a non-zero average line-of-sight velocity. On a $40 h^{-1}$ Mpc scale, this effect can substantially modify the Ω_m measurement (put $\Omega_{m,\text{mes}}$ in this section) by cutting the $P(v_{\text{rec}}^s, v_{\text{sim}})$ distribution at an inadequate place.

The results of the reconstruction on these three mock catalogues are given in Fig. 17. In Table 6, we give, for each mock catalogue, the best achievable result (thus highlighting purely the effect of choosing this mock catalogue) and the results one would obtain through observation of this piece of the universe. Unknown Lagrangian domain, redshift distortion and incompleteness effects are added to the considered mock catalogue. The problems of mass-to-light assignment and the ZOA are left apart for the sake of clarity. Their imprint on the velocities should most likely remain the same as we have shown in the corresponding previous sections, i.e. biasing for the first and increase of the scatter for the second. Only the cases with the forementioned observational effects are represented in Fig. 17.

Visual inspection of lower scatter plots in Fig. 17 shows that volume finiteness is likely making the $\Omega_{m,\text{mes}}$ measurement sensitive to the ‘local’ Ω_m (Ω_{eff} in the table). This assertion is supported by the estimation of s and Ω_m for *TrueDom* reconstructions given in Table 6. Moreover, experiments conducted with the spherical collapse model show that $\Omega_{m,\text{mes}}$ is indeed a weighted average between Ω_{eff} and Ω_m .

More specifically, reconstructed velocities in *4k-mock7* (including observational effects) are apparently giving the Ω_m of the simulation but they present a large scatter rendering the slope estimation dubious. Indeed, doing the same reconstruction but without observational effects gives a measured $\Omega_{m,\text{mes}} = 0.40$, which is the exact average between the simulation $\Omega_{m,\text{simu}} = 0.30$ and $\Omega_{\text{eff}} = 0.50$.¹¹ The aforementioned scatter is expected for this mock catalogue: the velocity field is badly reconstructed near the observer in that case (middle panels) because the local cosmic flow is higher than usual ($\sim 1000 \text{ km s}^{-1}$) and the non-linearities are stronger. Thus, the convexity of the problem is lost on an extended region around the observer when the reconstruction is conducted in redshift space (see Section 5). A particularly salient misreconstruction is given by the outflowing ‘bubble’ at the centre which disappears in the reconstructed velocity field. The size of the affected region is about $20 h^{-1}$ Mpc around the observer in 4k-mock7 and thus limits the

¹⁰ This behaviour is expected from an algorithmic point of view. The objects nearby the boundary cannot acquire any displacement using MAK because of the ‘pressure’/competition of objects inside the sphere. This problem is further enhanced in redshift space because generally these objects come from outside the sphere and are selected because their infall velocity is high. In *NaiveDom*, they cannot escape from the assumed spherical Lagrangian domain which thus leads to zeroing their velocity. On the other hand, *PaddedDom* is much less strict on the boundary, which leaves the freedom for MAK reconstruction to have a non-zero velocity even for objects on the boundary of the catalogue.

¹¹ Spherical collapse rather predicts $\Omega_{m,\text{mes}} = 0.35$ for the same setup.

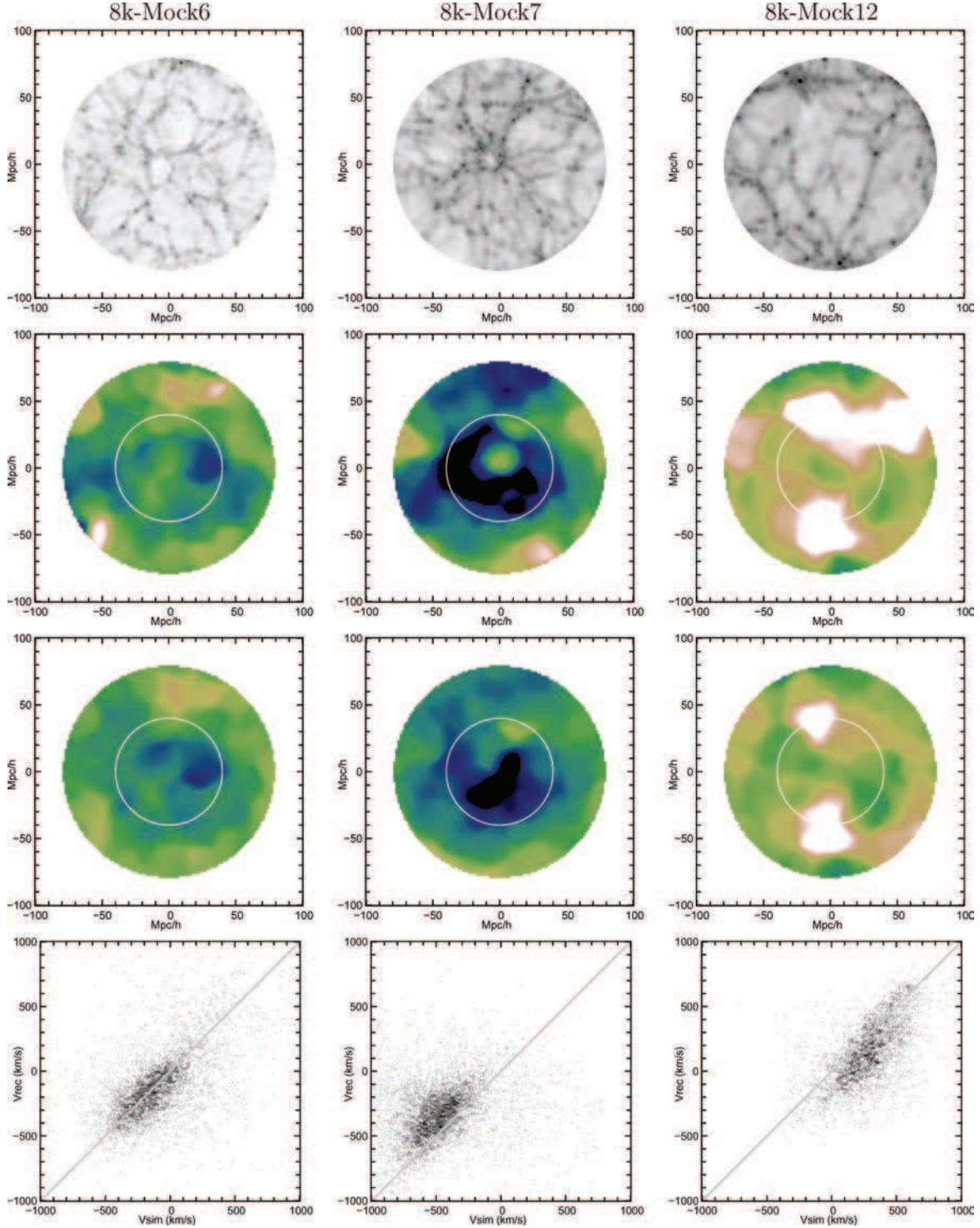


Figure 17. Cosmic variance. This figure gives a visual comparison of the three mock catalogues used to study cosmic variance effects. Top panels: adaptively smoothed density fields of the considered mock catalogues. In each case, we have represented the central thin slice that contains the observer. Second row: simulated velocity field, after smoothing with a $5 h^{-1}$ Mpc Gaussian window. The white circle gives the limit of the $40 h^{-1}$ Mpc volume. Third row: same as second row, but for the reconstructed velocity field. Fourth row: comparison between reconstructed and simulated peculiar velocities.

number of objects having both good reconstructed and observable peculiar velocities.

In an opposite way, velocities in *4k-mock12* are reconstructed with a better correlation, as shown by Table 6, but Ω_m measurement is strongly weighted toward Ω_{eff} . These two ‘features’ are largely due

to the huge central void. First, MAK reconstruction and Zel’dovich approximation are known to work better in low-density regions and being centred on a void results in inhibiting blueshift distortion as galaxies are principally going away from the observer, rendering the reconstruction problem convex in equation (8). Second, the

Table 6. Cosmic variance. Summary of measurements conducted on the three mock catalogues. The reconstruction is conducted either on the basic catalogue without any observational effect besides cosmic variance (labelled *Original*) or on the same catalogue but affected by redshift distortion, incompleteness and for which the Lagrangian domain is determined using *PaddedDom* (reconstruction labelled *Full*). The description of the other columns is given in Table 1.

| Catalogue | Reconstruction type | Velocity field | | | $\Omega_m(s)$ | Ω_m | Ω_m | Ω_m | Ω_m | Ω_m |
|--|---------------------|----------------|------|----------|--------------------------|--------------------------|---------------------------------|---------------------------|---------------------------|------------|
| | | s | r | σ | (\mathcal{L}_{\min}) | (\mathcal{L}_{\max}) | ($1.5\sigma, s_{\text{med}}$) | ($1.5\sigma, s_{\min}$) | ($1.5\sigma, s_{\max}$) | |
| 4k-mock6 ($\Omega_{\text{eff}} = 0.35$) | Original | 0.80 | 0.76 | 0.65 | 0.313 | 0.28 | 0.38 | 0.35 | 0.28 | 0.43 |
| | Full | 0.94 | 0.50 | 0.96 | 0.35 | 0.13 | 0.31 | 0.31 | 0.16 | 0.70 |
| 4k-mock7 ($\Omega_{\text{eff}} = 0.5$) | Original | 0.70 | 0.67 | 0.76 | 0.57 | 0.39 | 0.47 | 0.40 | 0.33 | 0.48 |
| | Full | 0.88 | 0.11 | 1.33 | 0.43 | 0.41 | 1.62 | 0.30 | 0.09 | 1.29 |
| 4k-mock12 ($\Omega_{\text{eff}} = 0.19$) | Original | 1.12 | 0.81 | 0.66 | 0.24 | 0.235 | 0.27 | 0.24 | 0.22 | 0.26 |
| | Full | 1.08 | 0.58 | 1.11 | 0.24 | 0.29 | 0.62 | 0.15 | 0.08 | 0.31 |

low-density region largely affects the statistical velocity distribution, which in this case leads to a measured $\Omega_{m,\text{mes}}$ weighted more strongly towards the Ω_{eff} of 4k-mock12.¹² This leads us to a $\Omega_{m,\text{mes}}$ that is nearer $\Omega_{m,\text{simu}}$ in 4k-mock12 than the mean matter density of the whole simulation. The volume finiteness also produces an apparent offset between reconstructed velocities and measured ones. This is expected as doing a statistical analysis on a finite-volume catalogue must introduce a selection bias effect. We have indeed checked that the point set $\{(v_{r,i}, \psi_{r,i}^s)\}$, obtained through a MAK reconstruction applied on 4k-mock12, is a subset of the corresponding set built from a reconstruction on 8k-mock12. Looking at our ‘standard’ 4k-mock6, one can note that the simulated velocity distribution is generally more symmetric according to the null velocity than for the two other mock catalogues, with no visual bias while comparing reconstructed velocity to simulated velocity. This supports the initial assertion linking $\Omega_{m,\text{mes}}$ to ($\Omega_{m,\text{simu}}, \Omega_{\text{eff}}$) and the asymmetric distribution of velocities. Potentially, one could recover the true Ω_m of the Universe (or here the simulation) from the measured velocities of any catalogues by predicting how the velocity distribution asymmetry is linked to local density contrast. However, the simplest, and more robust solution, would still be to extend the depth of current catalogues to reach a volume where velocities are normally distributed.

From a prediction point of view, comparing visually the velocity fields inside the white circles shows that, if we know Ω_m , we reconstruct plausible velocity fields for the three mock catalogues. Outside the white circles, the reconstructed velocity field is nearly completely uncorrelated compared to the simulated one as we have discussed in the previous section. It must be noted that the velocity field goes smoothly to zero (green colour) on the edge of all mock catalogues: this is an expected side effect of the homogeneous padding which tends to smooth out any fluctuation on the edge (velocity and density field).

7 VELOCITY MEASUREMENT ERRORS

7.1 The need for a likelihood analysis?

All the effects already described in this paper are present in a redshift catalogue. Though we expect most of the observational biases should be independent, some of them may correlate and give worse systematic errors. We present in Fig. 18 the progressive deterioration of the velocity–velocity comparison for 4k-mock6 based on

a reconstruction conducted on 8k-mock6. The effects are piled up from left- to right-hand side. The Ω_m measurements for the 1.5σ method are indicated below each panel. The obvious conclusion is that the measurements are progressively affected but that no extra correlated error seems to happen when mixing the effects. Another fortunate event is that bias seems to counterbalance themselves to give in the end a nearly unbiased result (last but one panel). Going from *TrueDom*/Real to Redshift tends to decrease Ω_m as has been seen previously. On the contrary, injecting incompleteness pushes the measurement to higher Ω_m as we have noted in Section 4.3. The 1.5σ method seems to give the right Ω_m value in all cases, which means that we should be able to use it on galaxy catalogues provided we have sufficient precision on velocity measurements. However, looking at the last panel (bottom right) of Fig. 18 shows that injecting random velocity measurement errors (here we introduced an optimistic error of 8 per cent of the distance to the object, corresponding to an error on distance modula of $\sigma_\mu = 0.17$), renders slope estimation much more difficult. In that case, the measured Ω_m is severely biased. This is expected as the 1.5σ method relies mostly on the central part of the scatter, which in turn is the one that is the most affected by random errors. This leads to a circularization of the 1.5σ isocontour and thus a completely wrong estimation of the slope. On the other hand, looking at the global structure of the scatter shows that the right-hand slope is still hidden in the data, but one should then take into account the tails of the distribution. This last test shows the limit of a direct velocity–velocity comparison in real cases. It might be possible to recover the original distribution of the scatter by deconvolving from the noise. However, it seems to be a difficult operation and we prefer to first try a maximum-likelihood approach. Its main advantage would be to work using distances, thus rendering the error in measurements more tractable.

7.2 Maximum-likelihood analysis

Observations of galaxies first give us access to their distances and not their peculiar velocities. A method based on distances to make a comparison between a model and observations is potentially less sensitive to distance measurement errors. Indeed, by comparing directly distances, one has a small relative error on each measurement instead of a huge one when peculiar velocities are considered. Below, we discuss galaxy selection bias and zero-point calibration errors in distance measurements while keeping the notation of Strauss & Willick (1995).

Presentation of the Bayesian chain. For the TF relation, one makes an estimate of the absolute magnitude of a galaxy as a function of its linewidth: the slope between the two quantities can be biased because the sample is limited in magnitude (Strauss & Willick 1995).

¹² The spherical collapse model would predict a measured $\Omega_{m,\text{mes}} = 0.26$ and this is in good agreement with the value measured when no observational effects are injected in the mock catalogue.

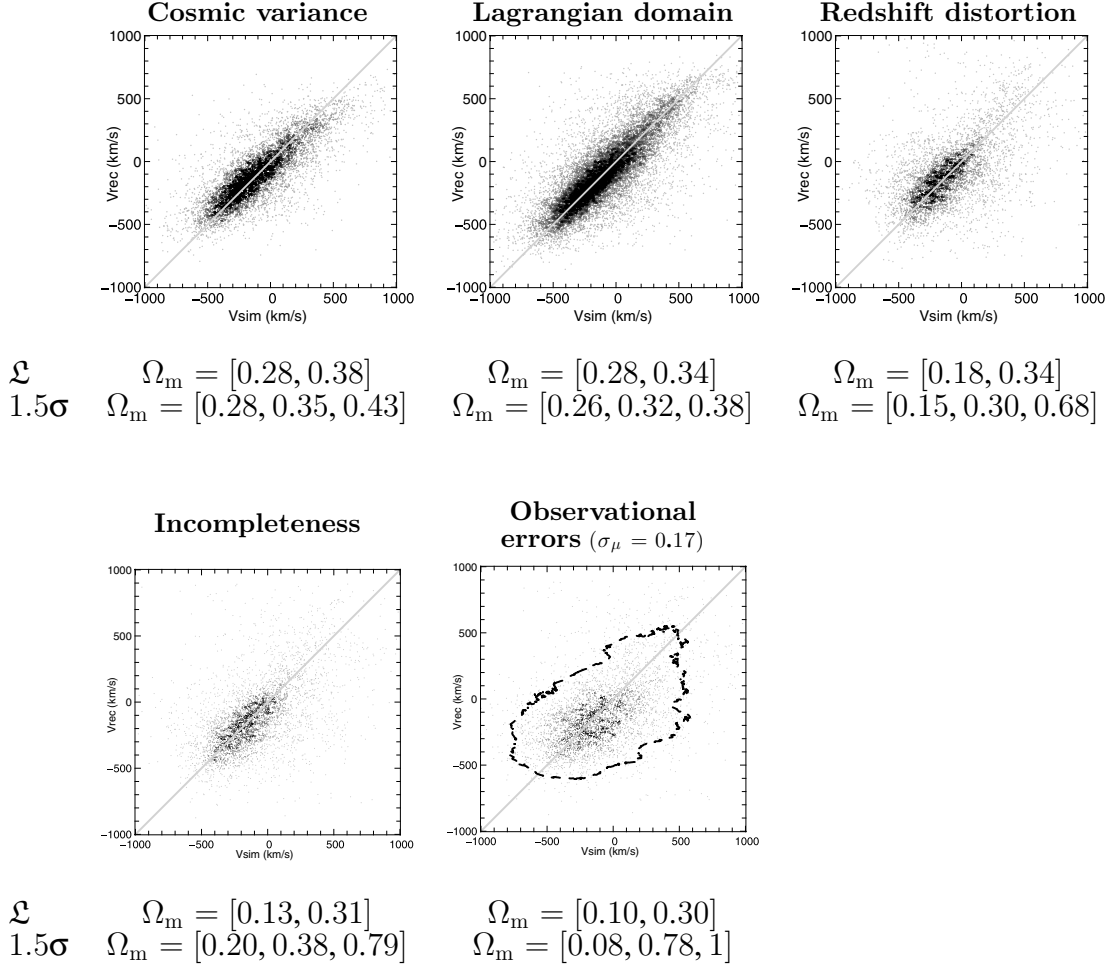


Figure 18. This figure gives the evolution of the scatter distribution and of the measurement of Ω_m using it while more and more observational effects are added to 8k-mock6 catalogue. All measurements of Ω_m are given between brackets and are sorted as follows. For measurements obtained through the likelihood analysis, labelled by \mathcal{L} , the first number corresponds to \mathcal{L}_{\min} and the second to \mathcal{L}_{\max} . For measurements obtained using the 1.5σ method, the first number corresponds to s_{\min} , then s_{med} and finally s_{\max} . The last (lower right) panel uses the full likelihood function of equation (31). All others use a restricted likelihood analysis with $\sigma_0/e = 0$, which is nearly equivalent to using equation (30) for each $(v_{r,i}, \psi_{r,i})$ pair. The 1.5σ isocontour has been plotted with a thick dashed line in the last panel.

This effect which is known as *selection bias* is purely statistical and if not correctly taken into account can lead to large systematic errors. Using these absolute magnitudes, occasionally combined to form groups of galaxies, and the apparent magnitudes of the same group, one builds the distance modulus

$$\mu(r) = m(r) - M = 5 \log_{10} \left(\frac{r}{10 \text{ pc}} \right) \quad (10)$$

with r the distance of the considered object (group of galaxies or galaxy). In addition to the forementioned statistical bias, peculiar velocity obtained from redshift positions through a Lagrangian reconstruction, here MAK, is sometimes very noisy, as shown in Fig. 18. Another more subtle effect is introduced by the Gaussian distribution of our velocity sample that we are going to analyse. We need to take care of this ‘selection bias’ to avoid being spoiled by eventual large reconstruction errors present for objects with a high velocity. Thus, we need a Bayesian approach to account for all these statistical effects.

In principle, the likelihood function gives a probability for the data, i.e. here redshift positions $\mathcal{Z} = \{z_i\}$, with i running from 1 to N , and distance moduli $\mathcal{M} = \{\mu_i\}$, assuming some model described

by the vector parameter p . Additionally, we assume that we have an estimation of measurement errors on \mathcal{M} through the set \mathcal{S} . The exact description of \mathcal{S} will be given in the next paragraph. Typically, errors on redshift measurements are of the order of $50\text{--}60 \text{ km s}^{-1}$. This means that we can consider them as negligible if we consider objects farther than $R_z = 6\text{--}10 h^{-1} \text{ Mpc}$. The volume enclosed by the sphere of radius R_z is, in any case, also poorly reconstructed because of the singularity introduced by redshift distortions near the observer (Section 5). In the following analysis, we will consider redshift measurements as negligible by avoiding the objects located at less than $10 h^{-1} \text{ Mpc}$ from the observer, thus we have¹³

$$P(\mathcal{M}, \mathcal{Z}|p, \mathcal{S}) \propto P(\mathcal{M}|\mathcal{Z}, p, \mathcal{S}) = \mathcal{L}(p). \quad (11)$$

The end of this section is devoted to computing the right-hand part of this equation. To achieve this, we will decompose the probability

¹³ Though it is in theory possible to avoid this hypothesis, it is in practice highly difficult for computational reason as one would need to run several MAK reconstructions to evaluate the extra integral that would be needed in equation (11).

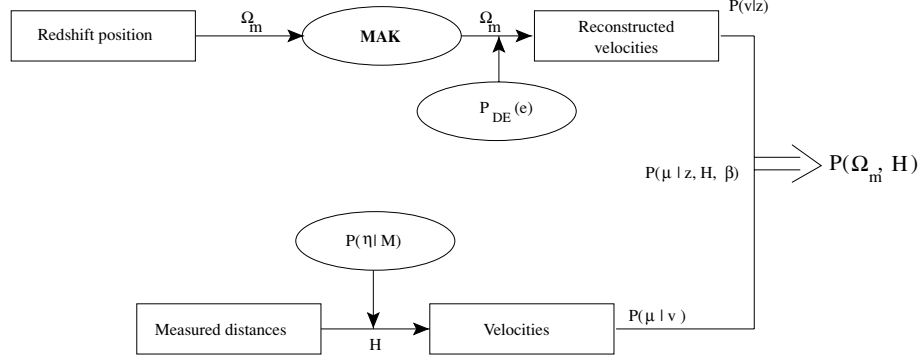


Figure 19. Maximum-likelihood analysis. This sketch illustrates the Bayesian chain used to establish the likelihood function. The input data are located on the left-hand side and the output posterior distribution $P(\beta, H)$ on the right-hand side.

into small pieces:

$$P(\mathcal{M}|\mathcal{Z}, \mathcal{S}, p) = \int \int \int_{\mathcal{M}_r, \mathcal{V}, D} P(D) P(\mathcal{M}|\mathcal{M}_r, \mathcal{S}, D, p) \times P(\mathcal{M}_r|\mathcal{V}, \mathcal{Z}, p) P(\mathcal{V}|\mathcal{Z}, p) d\mathcal{M}_r d\mathcal{V} dD \quad (12)$$

with $\mathcal{M}_r = \{\mu_{1,r}, \dots, \mu_{N,r}\}$ representing the ‘true’ distance moduli, with $\mu_{i,r} \in [-\infty, +\infty]$ and $\mathcal{V} = \{v_1, \dots, v_n\}$ the ‘true’ object peculiar velocities. $P(\mathcal{M}|\mathcal{M}_r, \mathcal{S}, p)$ is the probability of measuring the set of distance moduli \mathcal{M} given that the real set of distance moduli is \mathcal{M}_r and the expected error on the measurement is given by \mathcal{S} . $P(\mathcal{M}_r|\mathcal{V}, p)$ is the probability of obtaining the set of distance moduli \mathcal{M}_r given the reconstructed velocities \mathcal{V} . $P(\mathcal{V}|\mathcal{Z}, p)$ is the probability the velocities are well reconstructed from the redshift data \mathcal{Z} . The probability $P(D)$ is going to be introduced in the last paragraph to account for uncertainty in the calibration of the Tully-Fisher relation. All those probabilities are computed assuming the model parameters p . We will establish the likelihood function $\mathcal{L}(p)$ in three steps.

(i) First, the error distributions linked to observations are considered to get an unbiased distance estimator for groups. This analysis yields the probability $P(\mu_i|\mu, \sigma_{\mu,i}, p)$.

(ii) Second, the errors on reconstructed velocities are considered to compute $P(v|\mathcal{Z}, p)$.

(iix) Last, the two analyses are merged as given above to produce the likelihood function which gives the posterior distribution of β and the Hubble constant H .

A picture of the above Bayesian chain is given in Fig. 19.

Distance modulus error distribution. To establish the likelihood function comparing the measured distance to the reconstructed velocity field, we assume the distance catalogues are obtained using the inverse TF relation (Shaya et al. 1995),

$$\eta^0(M) = -e(M + D), \quad (13)$$

where M is the absolute magnitude of the considered galaxy, $\eta_0(M)$ is its predicted linewidth, e is the slope and D is the zero-point calibration (the latter two are assumed to be known exactly). It is known that inverse TF relation is less sensitive to the selection bias as compared to forward TF relation (Strauss & Willick 1995). Observational data show that the differences between the predicted linewidth $\eta^0(M)$ and the measured linewidth η for an object of absolute magnitude M are Gaussian distributed¹⁴ (Tully & Pierce 2000; Pizagno et al. 2007).

Thus, the probability of measuring the linewidth η , given that the object has an absolute magnitude M , and assuming that the TF relation $\eta^0(M)$ is known, is

$$P(\eta|M, e, D) = \frac{1}{\sqrt{2\pi}\sigma_\eta(M)} e^{-\frac{(\eta - \eta^0(M))^2}{\sigma_\eta^2(M)}} \quad (14)$$

with $\sigma_\eta(M)$ the linewidth estimation error for the absolute magnitude M . Distance catalogues are composed of estimated distance moduli μ_e from the inverse TF relation. These estimated distance moduli are built from the statistics on a single group. Therefore, the joint probability of having a galaxy in a group with both a linewidth η and an absolute magnitude M , assuming the TF relation $\eta^0(M)$, is

$$P(\eta, M|e, D) = F(M)P(\eta|M, e, D), \quad (15)$$

where $F(M)$ is the normalized absolute luminosity function of the group.¹⁵

The estimator for the distance modulus is given by

$$\mu_e = m - M_e(\eta) = M + \mu_0(r) + D' + \frac{\eta}{e'}, \quad (16)$$

where D' and e' are the estimated inverse TF parameters of equation (13) and $\mu_0(r) = 5 \log(\frac{r}{10 \text{ pc}})$ the true distance modulus of the considered group. The conditional probability that the estimated distance modulus for the group is μ , assuming that the estimated Tully-Fisher parameters are e' and D' and that the real parameters for this group are e and D , can be written as

$$\begin{aligned} P(\mu|\mu_0(r), e, e', D, D') &= \langle \delta_D (\mu - \mu_e) \rangle_{\text{group}} \\ &= \left\langle \delta_D \left(\mu - \mu_0(r) + M + D' + \frac{\eta}{e'} \right) \right\rangle_{\text{group}} \\ &= \int_M \frac{e' F(M)}{\sqrt{2\pi}\sigma_\eta(M)} e^{-\frac{(e'(\mu_0(r) - \mu) + e' D' - e D + (e' - e)M)^2}{2\sigma_\eta^2(M)}} dM. \end{aligned} \quad (17)$$

While working with the inverse TF relation, one can assume that the slope e' is completely determined and $e' = e$. Since the observed $\sigma_\eta(M)$ varies little with M , it is chosen to be equal to a constant σ_0 .

measurement of line width, which may reach 10 per cent because of the uncertainty in galaxy inclination correction, and the intrinsic modelling errors of the TF relation itself.

¹⁵ Note that the selection function is assumed to be independent of η and is hence absorbed in $F(M)$. $F(M)$ corresponds to $\Phi(M) S(M, \eta)$ in Strauss & Willick (1995) notation (e.g. equation 188).

¹⁴ In fact, in writing equation (14), two effects are mixed: the error on the

The previous probability reduces to

$$P(\mu|\mu_0(r), e, D, D', \sigma_0) = \frac{e}{\sqrt{2\pi}\sigma_0} e^{-\frac{(a-\mu_0(r)+D'-D)^2}{2\sigma_0^2}}. \quad (18)$$

Though the slope e' is well determined, the zero-point calibration D may still be affected by non-negligible errors.¹⁶ The set describing errors on distance moduli is thus $\mathfrak{S} = \{\sigma_{0,1}/e, \dots, \sigma_{0,N}/e\} = \{\sigma_{\mu,1}, \dots, \sigma_{\mu,N}\}$. The error on this calibration will affect the distances globally. As a first approximation we model the error on the zero-point by a Gaussian centred on D with a standard deviation of σ_d .

Linking distance modulus to velocity. The second probability function in equation (12) is $P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p)$, which is actually a distribution linking the velocities and redshifts to distance modulus. This principally corresponds to a change of variable and we give directly the expression of it, which is inspired by equation (1):

$$P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) = \prod_{i=1}^N H 10^{\mu_{r,i}/5} \delta_D(z_i - v_i - 10 \text{ pc} \times H 10^{\mu_{r,i}/5}). \quad (19)$$

Reconstructed velocity distribution. We are now going to establish the expression of $P(v|\mathfrak{Z}, p)$ with the vector of parameters of our chosen model $p = (H, \beta, B_v, \sigma_v, \gamma_*, e) - \sigma_v$ and γ_* are going to be introduced in the next immediate paragraphs. One may decompose $P(v|\mathfrak{Z}, p)$ that way

$$P(v|\mathfrak{Z}, p) = \int_{\mathfrak{P}} P(v|\mathfrak{P}, p) P(\mathfrak{P}|\mathfrak{Z}, p) d\mathfrak{P}, \quad (20)$$

with $\mathfrak{P} = \{\psi_{r,i}\}$ the reconstructed displacements. As MAK reconstruction is deterministic once β has been assumed (Section 4), the second probability distribution is simply given in our case by

$$P(\mathfrak{P}|\mathfrak{Z}, p) = \prod_{i=1}^n \delta_D(\psi_{r,i} - \psi_i(\mathfrak{Z}, \beta)) \quad (21)$$

with ψ_i representing the MAK reconstructed displacement of the i th object, being a function of all redshift coordinates and β . Thus, studying $P(\mathcal{V}|\mathfrak{Z}, p)$ reduces to examine $P(\mathcal{V}|\mathfrak{P}(p_0), p')$, with $p' = (H, \beta', B_v, \sigma_v, \gamma_*)$, $p_0 = (H, \beta_0, B_v, \sigma_v, \gamma_*)$, β_0 being the assumed growth factor to compute the set $\mathfrak{P}(p_0)$ using the redshift reconstruction. $P(\mathcal{V}|\mathfrak{Z}, p)$ and $P(\mathcal{V}|\mathfrak{P}(p_0), p')$ equalize only if $p = p' = p_0$. Thus, one needs a several redshift reconstructions to build the probability function $P(\mathcal{V}|\mathfrak{Z}, p)$. Working with the intermediary set \mathfrak{P} is easier than with \mathfrak{Z} , we thus put the reduced likelihood function:

$$\mathcal{L}'_{\beta_0}(p') = \int \int_{\mathfrak{M}_r, \mathcal{V}} P(\mathfrak{M}|\mathfrak{M}_r, \mathfrak{S}) P(\mathfrak{M}_r|\mathcal{V}, \mathfrak{Z}, p) P(\mathcal{V}|\mathfrak{P}(p_0), p') d\mathfrak{M}_r d\mathcal{V} \quad (22)$$

and we are going to establish the expression of the elementary probability function $P(v_r|\psi_r, p)$ which will yield

$$P(\mathcal{V}|\mathfrak{P}(p_0), p) = \prod_{i=1}^N P(v_{r,i}|\psi_{r,i}, p) \quad (23)$$

assuming statistical independence of all $\{v_{r,i}, \psi_{r,i}\}$ duets, and that $\mathfrak{P}(p_0)$ is obtained using a redshift reconstruction for which $\beta = \beta_0$.

\mathcal{L}' may be written in a factorized way:

$$\mathcal{L}'_{\beta_0}(p') = \prod_{i=1}^N \iint_{\mu_r, v_r} P(\mu_i|\mu_r, \sigma_{0,i}/e) P(\mu_r|v_r, z_i, p) P(v_r|\psi_{r,i}, p) d\mu_r dv_r. \quad (24)$$

The computation of \mathcal{L}' is clearly helped using this factorized form. We may now concentrate on the third probability function of the above equation.

As has been established in Section 2, the distribution of errors on the reconstructed velocity field is the Lorentzian

$$P_{DE}(e_{\psi v}) \propto \frac{1}{1 + \left(\frac{e_{\psi v}}{B_v}\right)^2}, \quad (25)$$

where $B_v = 86 \text{ km s}^{-1}$ (redshift reconstruction), with $e_{\psi v}$ the distance between the reconstructed velocity $\beta\psi_r$ and the true velocity v_r . This formulation is different from saying that the reconstructed velocity is affected by error when compared to the true velocity, and permits some errors in the MAK reconstructed displacement field. As has been seen in Section 6.3, the reconstructed velocities may also contain an extra offset that needs to be removed while measuring β . The error distance $e_{\psi v}$ is thus

$$e_{\psi v} = \alpha_* v_r - \beta_* \psi_r + \gamma_* \quad (26)$$

with

$$\alpha_*^2 + \beta_*^2 = 1 \text{ and } \beta = \beta_*/\alpha_* \quad (27)$$

and γ_* to account for a potential spurious offset in reconstructed velocities. From linear theory (Peebles 1980), we know that the line-of-sight component of the velocity field must be distributed like a Gaussian function. We now assume that the absolute probability for an object to have a velocity v is given by a Gaussian distribution:

$$P_{\text{vel}}(v|p) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{v^2}{2\sigma_v^2}}. \quad (28)$$

It must be noted that it is likely that the observational data do not encompass a sufficiently large volume so that measured velocities follow this law. Moreover, this prior is of some importance when we have to deal with highly scattered data. The shortcomings of such an approach will be discussed in the next section. One can recover the standard uniform prior on velocities by taking the limit $\sigma_v \rightarrow +\infty$ in the next equations. Assuming $e_{\psi v}$, as a random variable, is independent of v_r and these two quantities are themselves statistically independent from β and γ_* , we may now write the joint probability of reconstructing ψ_r , having a true velocity v_r :

$$\begin{aligned} P(v_r, \psi_r, \beta, \gamma_*|B_v, \sigma_v) &= \beta_* P_{DE}(e_{\psi v}|B_v, \sigma_v) \\ &\times P(\beta, \gamma_*|B_v, \sigma_v) P(v_r|B_v, \sigma_v) P(\psi_r|B_v, \sigma_v) \\ &= \beta_* C(B_v, \sigma_v) P(\beta, \gamma_*|B_v, \sigma_v) \frac{P(\psi_r|B_v, \sigma_v) e^{-\frac{v_r^2}{2\sigma_v^2}}}{1 + \left(\frac{\beta_* \psi_r - \alpha_* v_r + \gamma_*}{B_v}\right)^2}, \end{aligned} \quad (29)$$

where C is a function eventually depending on B_v and σ_v . The conditional probability that the true velocity is v_r given the reconstructed

¹⁶ The latest calibration is given in Tully et al. (2007).

displacement ψ_r is now exactly

$$P(v_r | \psi_r, p) = \frac{e^{-\frac{v_r^2}{2\sigma_v^2}} \left[1 + \left(\frac{\beta_* \psi_r - \alpha_* v_r + \gamma_*}{B_v} \right)^2 \right]^{-1}}{\int_{v=-\infty}^{+\infty} e^{-\frac{v^2}{2\sigma_v^2}} \left[1 + \left(\frac{\beta_* \psi_r - \alpha_* v + \gamma_*}{B_v} \right)^2 \right]^{-1} dv}. \quad (30)$$

The denominator of the right-hand part of this equation must be computed numerically.¹⁷ It can be shown that, in the limit $\sigma_v \rightarrow +\infty$, $P(v_r | \psi_r, \beta_*, \gamma_*)$ reverts to a pure Lorentzian form.

Merging the probability distributions. We may now establish the ‘elementary’ conditional probability for an object i to get a measured distance μ_i given that its reconstructed displacement is $\psi_{r,i}$, its redshift is z_i , the error on the linewidth measurement is $\sigma_{0,i}$ and the model parameters are p' in the notation of this section:

$$\begin{aligned} P(\mu_i | \psi_{r,i}(p_0), z_i, \sigma_{0,i}, D', p') \\ = \int \int_{v, \mu_r} P(\mu_i | \mu_r, \sigma_{0,i}, D', p') P(\mu_r | v, z_i, p') \\ \times P(v | \psi_{r,i}(p_0), p') d\mu_r dv \\ \propto \frac{eH}{\sigma_{0,i}} \int_{\mu_r} 10^{\mu_r/5} e^{-\frac{e^2}{2\sigma_{0,i}^2}(\mu_r - \mu_i + D - D')^2} \\ \times P[v = (z_i - 10 \text{ pc} \times H 10^{\mu_r/5}) | \psi_{r,i}(p_0), p'] d\mu_r, \end{aligned} \quad (31)$$

with $p' = (H, \beta', B_v, \sigma_v, \gamma_*)$, $\psi_{r,i}(p_0)$ being computed assuming the parameters p_0 . Looking closely at this probability, one may note that changing $D \rightarrow D' = D + \Delta$ is equivalent to changing $H \rightarrow H' = H \exp(\Delta/5)$. Thus, the uncertainty in the zero-point calibration translates only in an uncertainty on H and not on the parameters of the model.

We may now write the full formal expression of $\mathcal{L}'(p)$, as already sketched in equation (22). As specified in the discussion we take $P(D')$ to be a Gaussian distribution centred on D and with a standard deviation σ_D . Now, we may replace and get

$$\mathcal{L}'_{\beta_0}(p') \propto \int_{D'=-\infty}^{\infty} e^{-\frac{(D-D')^2}{2\sigma_D^2}} \prod_i P(\mu_i | \psi_{r,i}(p_0), z_i, \sigma_{0,i}, e, p') dD, \quad (32)$$

with i running on objects of the catalogue. Assuming a uniform prior on β, H and γ_* and taking care of the relation between \mathcal{L}'_{β} and \mathcal{L} as mentioned above, the Bayes theorem permits us to write

$$P(H, \beta, \gamma_* | \mathfrak{M}, \mathfrak{S}, \mathfrak{Z}, e, D', B_v, \sigma_v) \propto \mathcal{L}(p) = \mathcal{L}'_{\beta}(p). \quad (33)$$

We now have access to the posterior distribution of (H, β, γ_*) .

7.3 Results

The results of measuring Ω_m using the maximum-likelihood estimator are presented in the tables using the label \mathcal{L} .

Except in the case where we consider observational errors, we use a simplified version of \mathcal{L} by taking $\sigma_{0,i} = 0$. While it would have been natural to find the maximum of the likelihood for *all* parameters (including B_v, σ_v, γ_*), we quickly noted that it was leading to unacceptably biased measurements and to an unnecessary increase of the parameter space. Moreover, the results quite strongly depend on σ_v and γ_* , especially when the reconstruction noise becomes high

as in redshift reconstructions (see Appendix D). We thus propose to discuss the values obtained by setting $\gamma_* = 0$, $B_v = 90 \text{ km s}^{-1}$ and choosing two values for σ_v . First, we use linear theory to predict the average velocity dispersion of haloes in the universe, this leads to take $\sigma_v = 326 \text{ km s}^{-1}$ (the Ω_m measured that way is labelled \mathcal{L}_{max}). Second, $\sigma_v = +\infty$ is used to check the influence of recovering a uniform prior on the velocity distribution (labelled \mathcal{L}_{min} , respectively).

By looking at all tables of this paper, we noted that the difference between the two measured Ω_m is mostly following the interval defined by s_{min} and s_{max} . We were expecting such a behaviour (σ_v is more or less controlling the statistical bias of the likelihood function) but not that it would so clearly follow the other method. The more the scatter is important, the more the measurement becomes imprecise as expected. It must, however, be noted that on average the measure \mathcal{L}_{max} suffers less systematic bias than \mathcal{L}_{min} . This behaviour is supported by the tests conducted in Appendix D.

The seemingly well-estimated Ω_m in the lower right-hand panel of Fig. 18 has been computed using the full likelihood analysis. Actually, compared to the 1.5σ method for which the measured slope is undefined, \mathcal{L}_{min} and \mathcal{L}_{max} are basically *the same* as when *no observational errors* are introduced.

The correction based on a Gaussian velocity distribution assumption cannot be entirely trusted for 4k-mock7 and 4k-mock12. As one may note in Fig. 17, the velocity distribution is highly non-Gaussian in these cases. This renders incorrect our distribution modelling in Section 7.2. Looking at Table 6, we note that though the measurements on ‘Original’ reconstruction is not strongly affected, we cannot say the same thing using data obtained from ‘Full’ reconstruction. In the first case, the noise is sufficiently low so that the prior does not have much importance whereas in the second case the wrong modelling of the velocity distribution leads to a strong error on the measured Ω_m . Fortunately, the scatter distribution presents different types of properties that led to compatible measurements in Table 6 between the maximum likelihood ($\sigma_v = +\infty$ to remove the Gaussian prior) and the 1.5σ method. For 4k-mock7 and 4k-mock12, the slope estimate is helped by probing velocities with high magnitudes, leading to less possibility of systematic error on the slope.

One is thus led to use a sufficiently deep distance catalogue to ensure the velocity distribution is more or less Gaussian to be able to apply the correction to the likelihood analysis. In this case, one may rely on the value given by \mathcal{L}_{max} . If, on the contrary, the velocity distribution is highly non-Gaussian, one must use \mathcal{L}_{min} . If possible, a visual inspection of the velocity–velocity scatter plot must be conducted to give a check on the amount of statistical biasing.

8 CONCLUSION

The MAK method has been applied with success to reconstruct the velocity field and the density field of simulations (Mohayaee et al. 2006), providing an interesting tool to apply to galaxy catalogues in order to recover the dynamics of our local universe. This method presents the interesting advantage of finding the exact solution of an approximated dynamical problem written in Lagrangian coordinates. The Lagrangian description presents two major advantages. First, it gives a real estimation of peculiar velocities for each galaxies or groups of galaxies, as opposed to a field description which would give an average value at a given spatial position (which is also possible to build using the Lagrangian description). Second, it permits us to use the Zel’dovich approximation, which gives better peculiar velocity prediction than linear Eulerian theory applied to the same

¹⁷ This function is known as a Voigt profile.

dark matter density field. It means that we expect this method to give better results and more spatially resolved than, e.g. the POTENT method (Bertschinger & Dekel 1989) or velocity field reconstruction through spherical harmonics (Regos & Szalay 1989). Now, most previous analyses of Lagrangian peculiar velocity reconstruction have been run mostly on particle catalogues coming from simulations. However, galaxy catalogues are not as simple, and the main problems are as follows.

- (i) Catalogues mostly provide redshift positions of galaxies and for a few objects their physical distances from us.
- (ii) The luminosity is the only known ‘dynamical’ quantity for most objects in catalogues and so we need to extrapolate the M/L relation for known objects to the ones that we do not know.
- (iii) Incompleteness effects have to be taken into account: either because of magnitude limitation or due to extinction of objects by the galactic plane.
- (iv) The MAK reconstruction also needs the Lagrangian domain of the galaxy catalogue.

All these biases and unknown quantities render the reconstruction problem much more difficult than in simulations. We propose here both to test the feasibility of such a reconstruction on galaxy catalogues and the methods to overcome the problems that we have just mentioned. We have tried to address the following problems:

- (i) Reducing the introduced systematic errors due to unknown bias between mass and luminosity tracers. The dark mass can be either put uniformly into the catalogue or put in the detected haloes (Section 4.1). It appears that there exists an optimum way to distribute the mass, as can be seen in Fig. 6, which gives *unbiased* and *noiseless* reconstructed velocities, even though the exact location of 63 per cent of the mass in the universe remains unknown. In addition to the previous, global, problem, the relative mass distribution between objects in the catalogue is also uncertain as we do not know their true M/L . The induced systematic errors have been studied in Section 4.2 and we show that the naive approach corresponding to using $M/L = \text{constant}$ inevitably gives a large bias on reconstructed velocities. Even a reasonable guess, for instance, the one proposed by MH, is still significantly biased. This suggests some more work must be done on the M/L relation, particularly on the high-mass end. However, on the positive side, large random errors on M/L do not yield any systematic effect and only increase the scatter in the velocity–velocity comparison.

- (ii) We proposed a slightly improved way to correct for incompleteness effects in galaxy catalogues and its effect on reconstruction. Though it has given good results, we do not expect this method to be completely bias-free as it presents the same deficiencies as the previous item. However, by enforcing the correction on the *mass* distribution, we managed to preserve the dynamics in the observational data in a better way than would be the case if we had enforced it on the *luminosity* distribution.

- (iii) We investigated the eventual systematic errors in redshift reconstructions as proposed previously by Mohayaee & Tully (2005) and which corresponds to the inverse redshift operator studied by Valentine et al. (2000). It appears that, though the bias is small, Ω_m tends to be always underestimated.

- (iv) Two solutions to overcome the Lagrangian volume uncertainty for the case of finite-volume catalogues have been investigated. The reconstruction method which gives better result seems to be *PaddedDom*. The other alternative, *NaiveDom*, appears to bias the reconstructed velocities, especially in the case of a redshift reconstruction.

- (v) The efficiency of the correction for the ZOA as proposed by Shaya et al. (1995) has been checked (Section 4.1). It appears that the correction is bias-free and only introduces a small, but notable, additional noise for objects in the direction of the ZOA.

- (vi) We checked that the resulting errors of each effect are uncorrelated so they only pile up without producing a strong additional bias. It is fortunate that some observational effects produce complementary biases: incompleteness effect tends to overestimate Ω_m whereas redshift distortion underestimates Ω_m . The resulting bias is thus not so important.

- (vii) We finally tried two estimators to measure Ω_m from both reconstructed displacement and distance measurement (Section 7): the 1.5σ and the maximum-likelihood estimator. However, the first one is not able to work with noisy measured velocities, and the second one is badly affected by large distribution tails in redshift reconstruction. Adding a prior on the distribution of velocities in the catalogue helps to reduce the bias at the cost of having a good measurement of the width of this distribution. A good estimate of Ω_m is thus rendered more problematic though we have shown that it should be feasible in principle.

We intend to continue this work in the following directions.

- (i) This method can be applied to make a measurement of Ω_m in NBG-8k/NBG-3k catalogues and in the upcoming 6dFGS redshift and distance catalogues.

- (ii) A better comparison to the acoustic peaks of the CMB can potentially be obtained using the reconstructed displacement field (Eisenstein et al. 2006).

- (iii) We can apply MAK reconstruction on SDSS and 2MASS catalogues to obtain the initial Lagrangian positions and velocities of objects in our local universe. This would render the possibility of a re-simulation of our local universe for the first time and check the MAK prediction and correction schemes on real observations.

- (iv) We want also to improve the reconstruction itself and propose a new algorithm to include further gravitational effect during orbit reconstructions. This will never give us the internal structure of objects but potentially will give better reconstructed velocities while keeping the power of the MAK reconstruction.

ACKNOWLEDGMENTS

We are grateful to S. Prunet and C. Pichon for useful discussions and comments on maximum-likelihood methods. We would also like to thank D. Weinberg and M. Chodorowski for useful discussions. GL thanks the support and the hospitality of the Institute for Astronomy (University of Hawaii). This work has been supported by the HORIZON project (<http://www.projet-horizon.fr>).

REFERENCES

- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
- Bennett C. L. et al., 2003, *ApJS*, 148, 1
- Bertschinger E., Dekel A., 1989, *ApJ*, 336, L5
- Bertsekas D. P., 1979, *A Distributed Algorithm for the Assignment Problem*. MIT Press, Cambridge, MA
- Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, *A&A*, 296, 575
- Branchini E., Eldar A., Nusser A., 2002, *MNRAS*, 335, 53
- Brenier Y., Frisch U., Hénon M., Loeper G., Matarrese S., Mohayaee R., Sobolevskii A., 2003, *MNRAS*, 346, 501
- Colombi S., Chodorowski M. J., Teyssier R., 2007, *MNRAS*, 375, 348
- Couchman H. M. P., Thomas P. A., Pearce F. R., 1995, *ApJ*, 452, 797
- Croft R. A. C., Gaztanaga E., 1997, *MNRAS*, 285, 793
- da Costa L. N. et al., 1988, *ApJ*, 327, 544

- Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
- Eisenstein D. J., Seo H.-j., Sirko E., Spergel D., 2006, ApJ, 664, 675
- Erdoğan P. et al., 2006, MNRAS, 373, 45
- Falco E. E. et al., 1999, PASP, 111, 438
- Fontanot F., Monaco P., Borgani S., 2003, MNRAS, 341, 692
- Lahav O., Fisher K. B., Hoffman Y., Scharf C. A., Zaroubi S., 1994, ApJ, 423, L93
- Marinoni C., Hudson M. J., 2002, ApJ, 569, 101 (MH)
- Mohayaee R., Tully R. B., 2005, ApJ, 635, L113
- Mohayaee R., Mathis H., Colombi S., Silk J., 2006, MNRAS, 365, 939
- Huchra J. P., Geller M. J., Clemens C. M., Tokarz S. P., Michel A., 1992, Bull. Inf. Cent. de Données Stellaires, 41, 31
- Kaiser N., 1987, MNRAS, 227, 1
- Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Research Supported by the National Science Foundation. Princeton University Press, Princeton, NJ, p. 435
- Peebles P. J. E., 1989, ApJ, 344, L53
- Phelps S. D., Desjacques V., Nusser A., Shaya E. J., 2006, MNRAS, 370, 1361
- Pizagno J. et al., 2007, ApJ, 134, 945
- Regos E., Szalay A. S., 1989, ApJ, 345, 627
- Shaya E. J., Peebles P. J. E., Tully R. B., 1995, ApJ, 454, 15
- Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
- Spergel D. N. et al., 2007, ApJS, 170, 377
- Strauss M. A., Willick J. A., 1995, Phys. Rept., 261, 271
- Tegmark M. et al., 2006, Phys. Rev. D, 74, 123507
- Tully R. B., 1987, ApJ, 321, 280
- Tully R. B., 2005, ApJ, 618, 214
- Tully R. B., Pierce M. J., 2000, ApJ, 533, 744
- Tully R. B., Shaya E. J., Karachentsev I. D., Courtois H., Kocevski D. D., Rizzi L., Peel A., 2007, AJ, submitted (arXiv:0705.4139)
- Valentine H., Saunders W., Taylor A., 2000, MNRAS, 319, L13
- Zandivarez A., Martínez H. J., Merchán M. E., 2006, ApJ, 650, 137

APPENDIX A: CONSTRUCTION OF A MAK MESH

MAK reconstructions require a sampling of the matter distribution with ‘particles’ of equal mass corresponding to nodes of a homogeneous mesh. When considering the simulation, one uses a full periodic cubic mesh. However, in real galaxy catalogues, the relevant Lagrangian volume is a non-periodic compact subset inscribed in a larger rectangular mesh. In that case, the assignment is performed only for ‘particles’ belonging to this initial volume. Note that the determination of this initial volume is by itself a great challenge and a poor guess can have dramatic consequences.

Given a number of ‘galaxies’, or tracers, for which the individual masses M_i are known and a choice of the mass resolution of the MAK grid, m_R , the problem is now to determine how many ‘particles’ have to be assigned to tracer i . This number should be $n_i = M_i/m_R$ which is rarely an integer. To address this issue, we construct an integer function \tilde{n}_i such that the quantity

$$\chi^2 \equiv \sum_i \left(\frac{\tilde{n}_i m_R - M_i}{M_i} \right)^2 \quad (\text{A1})$$

is minimized given the constrain

$$\sum_i \tilde{n}_i = N_{\text{MAK}}, \quad (\text{A2})$$

where N is the total number of nodes on the MAK grid, such that $N_{\text{MAK}} \times m_R$ is as close as possible to the total mass, $\sum M_i$. The minimization of χ^2 is performed iteratively until convergence. Note that the solution of such a minimization is, in general, not unique due

to the possible permutations between objects of the same mass. Due to this degeneracy, it is needed to shuffle randomly the tracers prior to the minimization in order to avoid possible systematic effects.

Note finally that one must make sure that there is at least a few particles per tracer, $n_i \geq \alpha$ with $\alpha > 1$. This brings constraints on m_R and therefore on the size of the MAK mesh. Unfortunately, it is not always possible to have $\alpha > 1$ due to the prohibitive CPU cost it would imply for the MAK reconstruction in the present paper. To address this problem, we separate the catalogue into groups of galaxies and field galaxies. For the groups, the χ^2 minimization is performed, as explained above, with a possible loss of the lightest ones since n_i can still be smaller than unity. For the field galaxies, we use a simpler procedure as follows. Given the mass M_i of a galaxy i , a MAK tracer is randomly assigned to it with occurrence probability M_i/m_R .

APPENDIX B: TOOLS FOR ERROR ANALYSIS

To check the accuracy of the reconstructions, we compute the moment of the joint probability distribution of the reconstructed velocities $\mathbf{v}_{\text{rec},i}$ of object i and the simulated velocities of those objects $\mathbf{v}_{\text{sim},i}$. We write $\langle A \rangle$ the average of the quantity A

$$\langle A \rangle = \frac{1}{N} \sum_{i=1}^N A_i. \quad (\text{B1})$$

We define three second moments (after subtraction of the average):

$$\sigma_r^2 = \langle v_{\text{rec}}^2 \rangle, \sigma_s^2 = \langle v_{\text{sim}}^2 \rangle, \sigma_{r,s} = \langle v_{\text{rec}} v_{\text{sim}} \rangle. \quad (\text{B2})$$

From these moments, we can build the correlation coefficient:

$$r = \frac{\sigma_{r,s}}{\sigma_r \sigma_s} \quad (\text{B3})$$

and the ratio between the width of the reconstructed field PDF (density or velocity) and the width of simulated – mock – field PDF

$$s = \frac{\sigma_r}{\sigma_s}. \quad (\text{B4})$$

For these two quantities, the optimum value is 1. Alternatively, two other ‘slope’ estimator of the reconstructed velocities versus the simulated ones can be built from the above momenta

$$s_{\min} = \frac{\sigma_{r,s}}{\sigma_s^2} = sr \text{ and } s_{\max} = \frac{\sigma_r^2}{\sigma_{r,s}} = s/r. \quad (\text{B5})$$

These two slopes are interesting when one makes an estimation of Ω_m through s and needs an evaluation of the uncertainty. The two extra slopes determined using this way should, ideally, be equal to s but due to the lack of perfect correlation ($r < 1$), they are actually different from it in realistic cases. In fact, we have $s_{\min} < s_{\text{med}} < s_{\max}$.

Please note that we can define the relative dispersion

$$\sigma^2 = \frac{\langle (v_{\text{rec}} - v_{\text{sim}})^2 \rangle}{\sigma_s^2} = 1 + s^2 - 2sr, \quad (\text{B6})$$

which is a measure of the noise-to-signal ratio: high σ corresponds to low signal. Ideally, one wants $\sigma = 0$.

APPENDIX C: SIMULATING MAGNITUDE-LIMITED CATALOGUES

Having only a halo catalogue, we must generate a ‘galaxy catalogue’ including incompleteness effects. The main difficulty in that

construction is that the distribution of galaxies in the universe is a non-trivial, non-linear functional of the total matter density field. For instance, bright galaxies tend to concentrate in massive structures (Zandivarez, Martínez & Merchán 2006). It means that, though most of the field galaxies are missed, the major groups can still be easily seen due to the bright galaxies they contain. Thus, the galaxy distribution should mostly trace large haloes at large distances, potentially introducing a bias in the reconstructed peculiar velocities if incompleteness corrections are performed unwisely. In what follows, we generate mock galaxy catalogues like NBG-8k/3k. To take properly into account the effects discussed above, we separate groups of galaxies from field galaxies. Groups are populated with galaxies following the universal Schechter form for simplicity, but with a different normalization to account for their non-trivial M/L .

Statistically, NBG-8k/3k catalogues are composed of galaxies measured in the B band and distributed according to the Schechter form

$$n(L) dL \simeq n_0 L^{-1} e^{-L/L_*} dL, \quad (C1)$$

with $L_* \simeq 5.7 \times 10^{10} L_\odot$ and $n_0 \simeq 0.03 h^3 \text{Mpc}^{-3}$. Moreover, the NBG-8k catalogue is complete above $3 \times 10^9 - 4 \times 10^9 L_\odot$ inside a sphere of radius $d_{\text{comp}} = 12 h^{-1} \text{Mpc}$. As the mean ‘galaxy’ (particle) density in the simulation is $n_{\text{sim}} = 0.26 h^3 \text{Mpc}^{-3}$ and about $n_{\text{cat}} = 0.08 h^3 \text{Mpc}^{-3} \simeq 0.30 n_{\text{sim}}$ in NBG-8k, we must dilute the simulation to get a mock catalogue similar to NBG-8k. The luminosity L_G of a detected galaxy at a distance d from the observer must satisfy the constraint

$$L_G > 4\pi d_{\text{cut}}^2 \quad (C2)$$

with l_{cut} the minimum flux detectable by the observer. The fraction of galaxies detected at the distance d in the galaxy mock catalogue is thus

$$f_{\text{field}}(d) = \begin{cases} 0.30 & \text{if } d < d_{\text{comp}} \\ \frac{\int_{4\pi d_{\text{cut}}^2}^{\infty} n(L) dL}{\int_{L_{\min}}^{\infty} n(L) dL} & \text{otherwise} \end{cases} \quad (C3)$$

with l_{cut} the minimum flux detectable by the observer. The fraction is saturated at 0.30 to follow the dilution constraint expressed above. We enforce the continuity of $f_{\text{field}}(d)$ by choosing L_{\min} such that $f_{\text{field}}(d_{\text{comp}}) = 0.30$.

The mock galaxy and group of galaxies catalogue is now built:

I. We take a halo A from *FullMock* and assume it is a group of galaxies. We thus deduce the intrinsic luminosity L_A from the mass M_A of this object using equation (5).

II. The observed luminosity L'_A of A is computed assuming that its galaxy population follows (C1) but with a different normalization to achieve the intrinsic luminosity L_A . If d_A is the distance between the observer and the halo A , then the galaxies detected in this halo verify (C2) for $d = d_A$. The total observable luminosity for A is thus

$$L'_A = L_A f_L(d_A) \quad (C4)$$

with, assuming $L_{\min} \ll L_*$,

$$f_L(d) = \begin{cases} \sim 1 & d < d_{\text{comp}} \\ \frac{\int_{4\pi d_{\text{cut}}^2}^{\infty} L n(L) dL}{\int_{L_{\min}}^{\infty} L n(L) dL} \simeq e^{-\frac{4\pi d^2 l_{\text{cut}}}{L_*}} & d \geq d_{\text{comp}} \end{cases} \quad (C5)$$

III. If $L'_A < 4\pi d_A^2 l_{\text{cut}}$ then A is removed from the catalogue, otherwise it is kept.

IV. This gives us the group component of our magnitude-limited catalogue.

V. The case of the ‘field galaxies’ is treated separately. Galaxies are identified with dark matter particles and their luminosity is assigned following (C1). More specifically, we choose a shell S_d put at a distance d from the observer. The probability of keeping a ‘galaxy’ G in S_d is given by (C3). Inside the shell S_d , the selected ‘galaxies’ share now a luminosity

$$L_f(d) = \int_{4\pi d_{\text{cut}}^2}^{\infty} L n(L) dL \quad (C6)$$

which is distributed evenly among them. Strictly speaking, such a repartition should be performed randomly according to (C1). That would add a small additional noise on the reconstructed velocities. This noise should be of insignificant consequence as supported by the discussion of the $TS-T$ case in Section 4.2.

We have now a realistic mock catalogue and we try to account for its incompleteness as we would for NBG-8k:

A. The missing luminosity in groups is corrected. In order to do this, we compute, in a thin shell S_d at some distance d , the ratio between the expected total luminosity and the observed luminosity

$$b(d) = \frac{\int_0^{\infty} L n(L) dL}{\int_{4\pi d_{\text{cut}}^2}^{\infty} L n(L) dL} = e^{4\pi d^2 l_{\text{cut}}/L_*}. \quad (C7)$$

The intrinsic luminosity L_A of a group A in S_d is recovered with

$$L_A = L_{\text{obs},A} b(d). \quad (C8)$$

The mass M_A of A can then be obtained using the *non-linear* relation (5).

B. The remaining missing mass in S_d can be written as

$$M_{\text{missed},d} = \Upsilon b(d) (L_{\text{field,obs},d} + L_{\text{group,obs},d}) - M_{\text{field,obs},d} - M_{\text{group,obs},d}, \quad (C9)$$

with $\Upsilon = 93 \frac{M_\odot}{L_\odot}$ the average M/L ,¹⁸ $L_{\text{group,obs},d}$ the observed luminosity of groups, $M_{\text{group,obs},d}$ the masses of groups obtained after the above correction and $L_{\text{field,obs},d}$ the luminosity of field galaxies. The quantity $M_{\text{missed},d}$ comes from both missing galaxies and missing group of galaxies. If $M_{\text{missed},d} > 0$ and without any further information, the missing mass may either be assigned evenly to field galaxies of S_d (our choice, as usually performed in the literature), or distributed uniformly in S_d using new random tracers. If $M_{\text{missed},d} \leq 0$, the mass distribution in S_d is untouched.

This procedure is certainly not free from biases. For instance, the contrasts between shells are partly smoothed out, as illustrated by Fig. C1. This is equivalent to reducing the overall magnitude of fluctuations in the density field. As a result, a small bias towards larger Ω_m might occur, as in the lower right-hand panel of Fig. 6. On the opposite, if the missing mass is assigned to detected background galaxies, the estimated Ω_m is expected to underestimate the true value as discussed in Section 4.1.

C. Note that the mass of the ‘field galaxies’ is not the mass of a single particle anymore. Procedure explained in Appendix A is facilitated as follows, for the sake of algorithmic simplicity. With $v \in (0; 1]$, a uniform random variable, a galaxy G of mass m_G is splitted into n_G subcomponents of mass m_{particle} such that

$$n_G = \begin{cases} r_G & \text{if } \left(\frac{m_G}{m_{\text{particle}}} - r_G \right) < v \\ r_G + 1 & \text{otherwise} \end{cases}, \quad (C10)$$

¹⁸ Note that a prior assumption on the value of Ω_m is obviously needed to estimate Υ .

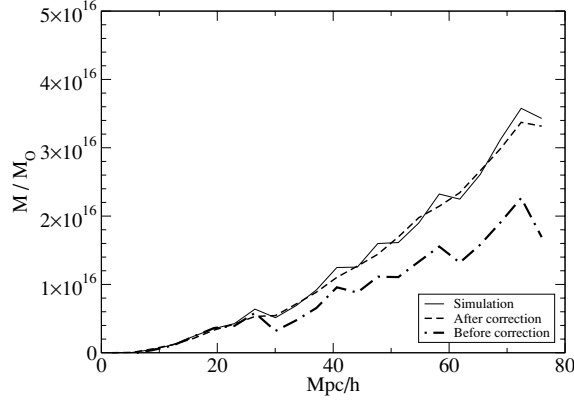


Figure C1. Magnitude limitation/filling missing mass. This plot gives the measured amount of mass in a thin shell at different distances from the observer. The solid line gives the original mass distribution in the simulation, the dot-dashed line the mass distribution after mimicking incompleteness and the dashed line the recovered mass distribution after correction for incompleteness as described in Appendix C.

with $r_G = \lfloor \frac{m_G}{m_{\text{particle}}} \rfloor$, $\lfloor x \rfloor$ being the integer part of x . Each of the subcomponent is now considered as a ‘field galaxy’ in the procedure explained in Appendix A.

APPENDIX D: STATISTICAL BIAS IN THE SLOPE ESTIMATION

The two methods that we used for slope estimation are known to be biased. A more precise treatment of this bias is beyond the scope of this paper. However, we propose here to check the order of magnitude of the systematic effect of the statistical analysis itself. To achieve this, we produced a set of randomly generated ‘velocities’ v and their ‘reconstructed velocities’ v_r counterpart. The probability for a point (v, v_r) to have a velocity v is given by

$$P_v(v) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-v^2/(2\sigma_v^2)}, \quad (\text{D1})$$

with $\sigma_v = 300 \text{ km s}^{-1}$ typically. The probability for it to have a reconstructed velocity v_r is given by the *same* probability law. We now compute the error e between v_r and v , which must be distributed according to the Lorentzian form

$$P_{\text{DE}}(e) = \frac{1}{\pi B} \frac{1}{1 + \left(\frac{e}{B}\right)^2}, \quad (\text{D2})$$

with $B = 86 \text{ km s}^{-1}$. The error e is related to v and v_r by

$$e = \alpha_* v - \beta_* v_r. \quad (\text{D3})$$

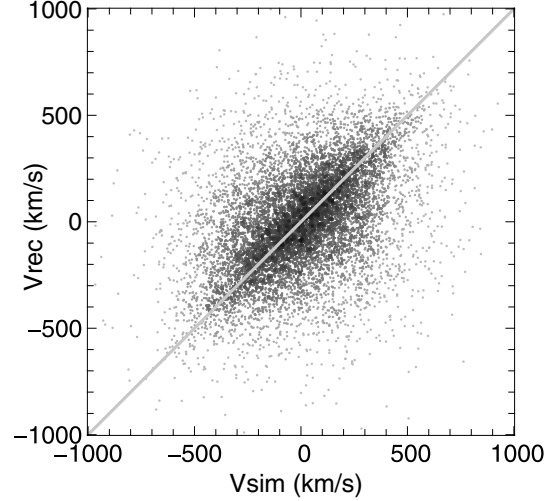


Figure D1. Statistical bias. Scatter plot of 10 000 randomly generated points following the approximated probability laws found between reconstructed velocities and simulated velocities.

For the rest of the Appendix, we take $\alpha_* = \beta_* = 1/\sqrt{2}$. The probability of keeping a point (v, v_r) with an error e is given by integrating $P_{\text{DE}}(e')$ between $-e$ and $+e$,

$$P_{\text{keep}}(e) = \frac{2}{\pi} \tan^{-1} \left(\frac{e}{B} \right). \quad (\text{D4})$$

We represented in Fig. D1 a scatter plot of 10 000 points generated using this procedure. As one can see, it does look like a real scatter plot of a redshift reconstruction.

Conducting a 1.5σ analysis on this set of points, we find a slope $\beta_*/\alpha_* = 1.0 \pm 0.20$. In our case, this would give $\Omega_m = 0.30 \pm 0.10$. Estimating the slope using the maximum-likelihood approach gives, with $\sigma_v = +\infty$, $\beta_*/\alpha_* = 0.81 \pm 0.01$ ($\Omega_m = 0.20 \pm 0.02$) and with $\sigma_v = 300 \text{ km s}^{-1}$, $\beta_*/\alpha_* = 1.074 \pm 0.012$ ($\Omega_m = 0.34 \pm 0.02$). Putting $B = 40 \text{ km s}^{-1}$, both for generated data and likelihood function, as for real space reconstructions, reduces the error and gives $\Omega_m = 0.31 \pm 0.02$, thus highlighting the importance of the reconstruction noise for a good estimation of Ω_m .

Consequently, though one must rely on the likelihood analysis, it may be strongly biased by the structure of reconstruction errors mixed with the non-uniform distribution of observables. We tried to make a good approximate model of the errors, though it seems to quite depend on the value of σ_v . Whenever possible, of course, one must crosscheck the result of the likelihood by a visual inspection of the scatter plot.

This paper has been typeset from a \LaTeX file prepared by the author.