Bayesian reconstruction of dark matter distribution from peculiar velocities: accounting for inhomogeneous Malmquist bias

Supranta S. Boruah, Guilhem Lavaux, Michael J. Hudson

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **517,** 4529–4543 (2022)
Advance Access publication 2022 October 21

https://doi.org/10.1093/mnras/stac2985

Bayesian reconstruction of dark matter distribution from peculiar velocities: accounting for inhomogeneous Malmquist bias

Supranta S. Boruah [1]★ Guilhem Lavaux[2] and Michael J. Hudson [3,4,5]

[1] *Department of Astronomy and Steward Observatory, University of Arizona, 933 N Cherry Avenue, Tucson, AZ 85719, USA*
[2] *CNRS & Sorbonne Université, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[3] *Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[4] *Waterloo Centre for Astrophysics, University of Waterloo, 200, University Ave W, Waterloo, ON N2L 3G1, Canada*
[5] *Perimeter Institute for Theoretical Physics, 31 Caroline St N, Waterloo, ON N2L 2Y5, Canada*

Accepted 2022 October 10. Received 2022 October 5; in original form 2021 November 30

ABSTRACT

We present a Bayesian velocity field reconstruction algorithm that performs the reconstruction of the mass density field using only peculiar velocity data. Our method consistently accounts for the inhomogeneous Malmquist (IHM) bias using analytical integration along the line of sight. By testing our method on a simulation, we show that our method gives an unbiased reconstruction of the velocity field. We show that not accounting for the IHM bias can lead to significant biases in the Bayesian reconstructions. We applied our method to a peculiar velocity data set consisting of the SFI++ and 2MTF Tully–Fisher catalogues and the A2 supernovae compilation, thus obtaining a novel velocity reconstruction in the local Universe. Our velocity reconstructions have a cosmological power spectrum consistent with the theoretical expectation. Furthermore, we obtain a full description of the uncertainties on reconstruction through samples of the posterior distribution. We validate our velocity reconstruction of the local Universe by comparing it to an independent reconstruction using the 2M++ galaxy catalogue, obtaining good agreement between the two reconstructions. Using Bayesian model comparison, we find that our velocity model performs better than the adaptive kernel smoothed velocity with the same peculiar velocity data. However, our velocity model does not perform as well as the velocity reconstruction from the 2M++ galaxy catalogue, due to the sparse and noisy nature of the peculiar velocity tracer samples. The method presented here provides a way to include peculiar velocity data in initial condition reconstruction frameworks.

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

1 INTRODUCTION

The large-scale structure of the Universe sources the peculiar velocity of the galaxies. The study of the peculiar velocity of galaxies in the local Universe is important in cosmological applications for two reasons: (i) peculiar velocities are the only probe of the growth of large-scale structure in the low-redshift ($z \lesssim 0.1$) Universe, (ii) peculiar velocities are a nuisance parameter in the measurement of the Hubble parameter H_0 and thus need to be corrected for in such measurements (Boruah, Hudson & Lavaux 2021; Peterson et al. 2022). Correcting for the peculiar velocity contributions to the redshifts in the measurement of the expansion history relies on using a reconstruction of the peculiar velocity field in the local Universe. Given the current tension in the measurement of the Hubble constant, H_0 (Verde, Treu & Riess 2019), we need to account for all sources of uncertainties. In order to obtain unbiased estimates of the peculiar velocity corrections for the local H_0 measurement, we need accurate models of the local peculiar velocity field.

Therefore the reconstruction of the velocity field of the local Universe is important for these cosmological applications. Many methods of reconstruction of the peculiar velocity field have been proposed in the literature. We can categorize the different methods according to two criteria: (i) the data used for the reconstruction, i.e. whether galaxy catalogues or peculiar velocity tracers are used in the reconstruction and (ii) whether the method involves forward-modelling or a direct inversion from the data.

Most common reconstructions of the local velocity field use galaxy catalogues to reconstruct the galaxy density field (e.g. Carrick et al. 2015; Lilow & Nusser 2021). Following the reconstruction of the galaxy density field, the velocity field is calculated using linear perturbation theory. However, since galaxies are biased tracers of the matter field, we need to fit for the parameter $\beta = f/b$ to get the velocity field, where f is the logarithmic growth parameter and b is the linear galaxy bias. This is done by comparing the reconstructed velocity field to the peculiar velocity data. On the other hand, reconstruction methods relying directly on peculiar velocity data are not affected by galaxy bias, thus providing a complementary method for reconstructing the velocity field. On the flip side, the peculiar velocity catalogues are sparse and the velocity estimates obtained from peculiar velocity tracers are very noisy. Therefore reconstructing the density field is usually easier than reconstructing the velocity field. One well-known example of a reconstruction method relying on velocity data is the POTENT

★ E-mail: ssarmabo@email.arizona.edu

method (Bertschinger & Dekel 1989; Bertschinger et al. 1990) which uses peculiar velocity data directly to reconstruct the velocity field, assuming that the velocity field is proportional to the gradient of the gravitational potential field. Another such widely used method is the adaptive kernel-smoothed velocity reconstruction method (Springob et al. 2014, 2016) that smooths the peculiar velocity data to obtain the velocity reconstruction.

The other criteria that we can use to categorize the reconstruction methods is whether the reconstruction is performed using forward-modelling or uses a direct inversion from the data. Inverting non-linear problems from partial, noisy, observations is an ill-posed inverse problem, which makes forward-modelled Bayesian methods particularly suitable for the task of reconstruction of high-dimensional fields. Bayesian reconstruction methods have become increasingly popular in cosmology and have been applied in a range of different applications such as initial conditions reconstruction (Jasche & Wandelt 2013; Modi, Feng & Seljak 2018; Jasche & Lavaux 2019), weak lensing (Fiedorowicz et al. 2022; Porqueres et al. 2021, 2022; Boruah, Rozo & Fiedorowicz 2022), and CMB lensing (Millea et al. 2021; Millea, Anderes & Wandelt 2020). Such methods have also been used for the local velocity field reconstruction. The simplest of such methods uses a Wiener filtering technique (Zaroubi, Hoffman & Dekel 1999). This approach assumes that the density/velocity field is described as a Gaussian random field and the Wiener filtered reconstruction is the maximum-a-posteriori (MAP) solution for the problem. The Wiener filtering approach has been extended to account for uncertainties and biases in the reconstruction using a constrained realization approach (Hoffman & Ribak 1991; Hoffman, Courtois & Tully 2015; Hoffman et al. 2018; Lilow & Nusser 2021) An alternative way to account for the biases in the reconstruction in Wiener filtering is using the unbiased minimal variance approach (Zaroubi 2002). Another similar approach is the Bayesian hierarchical method, VIRBIUS (Lavaux 2016), which is based on the constrained realization approach but accounts for many different systematic effects in its analysis. This approach has been been applied to the *Cosmicflows-3* (Tully, Courtois & Sorce 2016) data set by Graziani et al. (2019). A similar reconstruction code, HAMLET, was introduced in Valade et al. (2022). However, these methods fail to account for the inhomogeneous Malmquist (IHM) bias which is an important source of systematic error in peculiar velocity analysis. The IHM bias arises from an incorrect assumption on the distribution of peculiar velocity tracers due to neglecting the line-of-sight inhomogeneities.

In this paper, we introduce a Bayesian reconstruction method that uses peculiar velocity data in its reconstruction while consistently accounting for the IHM bias in the analysis. As we show in this paper, not accounting for the IHM bias in the reconstruction may lead to substantial biases in the reconstruction. One approach to correct for the IHM bias is by marginalizing over an accurate model of galaxy distribution (e.g. Hudson 1994). We use such an approach to deal with the IHM bias within our peculiar velocity reconstruction method. Our method results in samples of reconstruction that are compatible with the data from a given peculiar velocity catalogue, and sampled from a posterior assuming a ΛCDM Gaussian prior. Furthermore, the full correlated uncertainties in the reconstruction can be estimated from the posterior samples. Our method presented here can be extended in a straightforward manner to include more complex gravity models such as initial condition reconstruction methods like BORG (Jasche & Wandelt 2013; Jasche & Lavaux 2019).

The paper is structured as follows: in Section 2, we describe the simulations and data used in this paper. In Section 3, we describe our reconstruction method in detail. Our method is validated with mock simulated catalogues in Section 4 followed by an application of our method to reconstruct the velocity field of the local Universe in Section 5. Finally, after a brief discussion in Section 6, we summarize our results in Section 7. In Appendix A, we investigate in further detail the various sources contributing to the reconstruction error.

2 SIMULATIONS AND DATA

In this section, we describe the simulations and the peculiar velocity surveys used in this work.

2.1 VELMASS simulation

In order to validate our method, we use an N-body simulation from the VELMASS suite of simulations.[1] The simulation was performed in a cubic box of size $(2\ h^{-1}\ \mathrm{Gpc})^3$ with 2048^3 particles with mass $9.387 \times 10^{10}\ h^{-1}\ \mathrm{M}_\odot$. The cosmological parameters used are: $\Omega_\mathrm{m} = 0.315$, $\Omega_\mathrm{b} = 0.049$, $H_0 = 68\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, $\sigma_8 = 0.81$, $n_s = 0.97$, and $Y_\mathrm{He} = 0.248$. The ROCKSTAR (Behroozi, Wechsler & Wu 2013) halo finding software was used to identify the dark matter haloes in the simulation. We only consider the haloes with mass $M > 3 \times 10^{12}\ h^{-1}\ \mathrm{M}_\odot$.

In order to create a mock peculiar velocity survey, we populate the haloes with mass $M > 3 \times 10^{12}\ h^{-1}\ \mathrm{M}_\odot$ with a standard candle with different levels of intrinsic scatter. That is, we assign it an absolute magnitude $\mathcal{M} = \mathcal{M}_0 + \epsilon$, where $\mathcal{M}_0$ is the standard candle absolute magnitude and ϵ is the intrinsic scatter drawn from a normal distribution, $\epsilon \sim \mathcal{N}(0, \sigma_\mathrm{int}^2)$. Here, σ_int is the standard deviation of the intrinsic scatter. We can calculate the apparent magnitude from this absolute magnitude and the distance to the haloes. In this paper, we report the results with three different intrinsic scatters, corresponding to distance uncertainties of 7, 15, and 20 per cent. Finally, we make an apparent magnitude cut to create a magnitude limited catalogue. From this magnitude limited sample, we randomly select 7000 haloes. The non-linear velocity of the haloes is calculated from the simulation output, which in turn is used to calculate the observed redshift, z_obs, of these haloes. While running the forward likelihood method described in Section 3.3, the method requires the estimated galaxy density field as an input.

2.2 Peculiar velocity surveys

In this work, we use peculiar velocity data from various surveys in our algorithm to reconstruct the velocity field of the local Universe. The three peculiar velocity data sets that we use in this work are the A2 supernovae compilation, SFI++ and 2MTF Tully–Fisher (TF) catalogues. Compared to our treatment of these data sets in Boruah, Hudson & Lavaux (2020), we do not fit the TF or the SNe light curve fitting parameters during the reconstruction. Instead we use the distances as provided in the catalogues. We however fit for a scaling factor that may arise due to miscalibration in the fitting of the zero-points of the TF relations as detailed in Section 3.3. We use a different scaling factor for each peculiar velocity survey. We use the galaxy density reconstruction of Carrick et al. (2015) to correct for the IHM bias. Since this reconstruction is limited to 125 h^{-1} Mpc in certain sky directions, we only consider peculiar velocity tracers within estimated distances $d < 100\,h^{-1}$ Mpc in this work, thus limiting the edge effects in our reconstruction. Note

[1]For more details on the simulation which we used, see Kodi Ramanah, Charnock & Lavaux (2019).

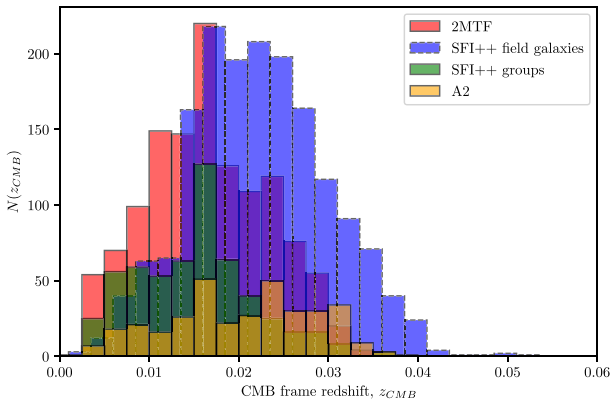

Figure 1. Redshift distribution of the tracers in the different peculiar velocity catalogues used in this work. We have imposed a distance cut of $d < 100\,h^{-1}$ Mpc on all catalogues.

that imposing such a distance cut does not introduce a selection bias in the forward peculiar velocity analysis methods (Strauss & Willick 1995). The redshift distribution of peculiar velocity tracers is shown in Fig. 1. In the following, we provide a brief description of the data used in this work. We refer readers to Boruah et al. (2020, 2021) for more details on the selection and the treatment of outliers, choosing only to highlight some main features of the data sets.

(i) *A2 Supernovae*: In Boruah et al. (2020), we compiled the second amendment (A2) data set of nearby supernovae from publicly available supernovae from the CfA supernovae sample (Hicken et al. 2009), Carnegie Supernovae Project-Data Release 3 (CSP-DR3; Krisciunas et al. 2017), the Lick observatory Supernova Survey (LOSS; Ganeshalingam, Li & Filippenko 2013), and the Foundation sample (Foley et al. 2018; Jones et al. 2019) of supernovae. We only consider the supernovae with estimated distance $d < 100\ h^{-1}$ Mpc. The median distance uncertainty, Δ_d, of the A2 supernovae is around 7 per cent, with the 16 and 84 percentile range of this uncertainty being 5 and 8 per cent, respectively.

(ii) *SFI++*: We use peculiar velocity data from the SFI++ (Masters et al. 2006; Springob et al. 2007) TF catalogue, which is an *I*-band TF survey. We follow the treatment of Boruah et al. (2020) to remove outliers. For galaxies in groups, we use the peculiar velocity data for the groups to suppress non-linear contributions to the peculiar velocity data. We only consider the galaxies/groups with estimated distance $d < 100\ h^{-1}$ Mpc. The median distance uncertainty, Δ_d, of the SFI++ galaxies is 20 per cent, with the 16 and 84 percentile range of this uncertainty being 18 and 24 per cent, respectively.

(iii) *2MTF*: The 2MTF survey (Masters, Springob & Huchra 2008; Hong et al. 2019) is an all sky TF galaxy catalogue in the near-infrared *J*, *H*, and *K* bands. The survey is limited to $cz < 10000$ km s^{-1}, approximately corresponding to $d < 100\ h^{-1}$ Mpc. The median distance uncertainty, Δ_d, of the 2MTF galaxies is 22 per cent, with the 16 and 84 percentile range of this uncertainty being 19 and 25 per cent, respectively.

With the given distance cuts, we select 345, 1682, 556, 1225 objects from the A2 catalogue, SFI++ field galaxy catalogue, SFI++ group catalogue, and 2MTF catalogue, respectively.

3 METHODOLOGY

We introduce a novel method for Bayesian velocity reconstruction which can consistently account for IHM bias in the reconstruction in this paper. In this section, we describe the theory and the methodology behind our method for reconstructing the velocity field with the peculiar velocity data.

3.1 Peculiar velocity theory

In the ΛCDM model, under linear perturbation theory, the present day peculiar velocity field can be expressed in terms of the density field as (Peebles 1980),

$$\boldsymbol{v_k} = \frac{i f H_0 \boldsymbol{k}}{k^2}\delta_{\boldsymbol{k}}, \tag{1}$$

where $f \approx \Omega_m^{0.55}$ is the logarithmic growth rate in the ΛCDM model, $i^2 = -1$. Since a factor of k appears in the denominator, the peculiar velocity field is sensitive to the large-scale density modes. Since these modes are well described with linear perturbation theory, linear theory predictions of peculiar velocity provide a fairly accurate result for the true velocity field. The peculiar velocity estimated using linear theory has been calibrated with simulations (Carrick et al. 2015; Hollinger & Hudson 2021) and the uncertainty in the linear theory velocity estimates due to the non-linearities was found to be $\sigma_{\rm NL} \approx 150$ km s^{-1}. In this work, we use linear perturbation theory to predict the velocity field. In the future, this can be extended to include non-linear differentiable velocity field models such as the Lagrangian perturbation theory (LPT; Bouchet et al. 1995) full particle mesh or COLA (Tassev, Zaldarriaga & Eisenstein 2013) simulations that have previously been used for Bayesian density reconstruction (e.g. Jasche & Lavaux 2019).

3.2 Inhomogeneous Malmquist bias

Peculiar velocity analyses are impacted by a number of different statistical biases, including the homogeneous and IHM biases. The homogeneous Malmquist bias arises due to the fact that the number of galaxies grows along the line of sight (LOS) according to the volume factor. In this paper, we mainly discuss the impact of IHM bias in Bayesian reconstruction with peculiar velocity data. Therefore, we discuss the basics of IHM bias in some detail in this section.

The IHM bias arises due to neglecting inhomogeneity in the radial distribution of the peculiar velocity tracers. Neglecting the overdensities along the LOS leads to a bias in the inferred distance, $d_{\rm est}$, compared to the true distance, $d_{\rm true}$. The magnitude of the bias is given as (Lynden-Bell et al. 1988; Strauss & Willick 1995),

$$\Delta_{\rm IHM} = \langle d_{\rm true}\rangle - \langle d_{\rm est}\rangle \approx d_{\rm est}\gamma\,\Delta_d^2, \tag{2}$$

where, Δ_d is the fractional uncertainty in the distance estimates and γ is the logarithmic slope in the galaxy overdensity along the LOS given by

$$\gamma = \frac{{\rm d}\ln n(r)}{{\rm d}\ln r}. \tag{3}$$

Since the slope changes sign in front of and behind an overdensity, the distance of the galaxies in front of the overdensity is underestimated and those at the back are overestimated. This may be interpreted as a spurious flow towards the overdensity if the IHM bias correction is neglected, and thus biases the magnitude of the velocity field to a higher value compared to the true value. Therefore, the inferred value of β is biased high when the IHM bias is not accounted for in velocity–velocity comparison methods (e.g. Boruah et al. 2020). As

we will show in Section 4, in Bayesian reconstruction from peculiar velocity data, this effect shows up as an increase in the inferred power spectrum of the reconstructed density field if the IHM bias is not corrected for.

3.3 Probability model

The method we present in this paper is a Bayesian method to reconstruct the velocity field using the data from a peculiar velocity survey. One such reconstruction method is the VIRBIUS algorithm (Lavaux 2016) which has previously been applied to the *Cosmicflows-3* (Tully et al. 2016) data set in Graziani et al. (2019). In their method, the true distance of the peculiar velocity tracers are used as latent parameters and the value of the distance is inferred in each Markov Chain Monte Carlo (MCMC) step. The homogeneous Malmquist bias is dealt with by using a skewed prior on these distances. However, the IHM bias is not explicitly accounted for in their analysis. In this work, we account for the IHM bias by using analytic integration over the LOS uncertainty. Marginalizing over the true distance also reduces the complexity since we do not have to infer their value in each MCMC step. Furthermore, the posterior distribution of the peculiar velocity tracer distances can potentially be highly multimodal due to the LOS inhomogeneities, which makes the analytical marginalization preferable to sampling based methods.

In Bayesian reconstructions, we want to infer the velocity field, $\{\boldsymbol{v}\}$, given the peculiar velocity data. The peculiar velocity data set consists of the observed redshift, $\{z_{\rm obs}\}$, and the estimate of its distance modulus, $\{\mu\}$, for a set of peculiar velocity tracers. The uncertainty associated with the distance modulus is denoted by σ_μ. In linear perturbation theory, the velocity field is calculated from the density field, $\{\delta\}$, using equation (1). Therefore, we can equivalently think about the process described here as a density reconstruction algorithm. Furthermore, we need an estimate of the galaxy density field to correct for the IHM bias. For this work, we provide the estimate of the galaxy density field, $\{\hat{\delta}_g\}$, as an input to our method.

In this work, we use a variant of the forward VELMOD method (Willick et al. 1997), which was introduced in Pike & Hudson (2005) to model the likelihood. For simplicity, we call this likelihood the *Forward Likelihood*. The forward likelihood method corrects for the IHM bias by accounting for the density dependent LOS distribution of peculiar velocity tracers. In the forward likelihood method, the observed redshift is predicted as a function of the distance estimates. That is, we model the likelihood using the conditional probability, $\mathcal{P}(\{cz_{\rm obs}\}|\{\mu\}, \{\boldsymbol{v}\}, \{\hat{\delta}_g\})$. We show the dependency of the various variables and the observables in a probabilistic graph in Fig. 2. Using Bayes theorem, the posterior for the density field δ can be written as

$$\mathcal{P}(\{\delta\}|\{cz_{\rm obs}\}, \{\mu\}, \{\hat{\delta}_g\}) \propto \mathcal{P}(\{cz_{\rm obs}\}|\{\mu\}, \{\delta\}, \{\hat{\delta}_g\})\mathcal{P}(\{\delta\}). \tag{4}$$

In our runs, we use a 128^3 cubic box where the density is inferred. That leads to a total of 128^3 density parameters. In order to sample from such high-dimensional parameter space, we used the Hamiltonian Monte Carlo (HMC; Neal 1993, 1996) algorithm. HMC can be used to sample from very high dimensional parameter space by using the gradient information of the log-posterior. In this paper, we implement our code by relying on the JAX package (Bradbury et al. 2018) which includes the gradients of the requisite function by using automatic differentiation. Sampling with HMC is highly sensitive to the choice of mass matrix. The optimal choice is to use the Hessian of the negative log-posterior as the mass matrix (Taylor, Ashdown & Hobson 2008). We create a number of Gaussian realizations of the density field and then calculate the second derivative of the log-posterior for 20 such fields. We use automatic differentiation to calculate the second derivative of the posterior. The mean of the Hessian calculated from these mocks is then used as our mass matrix.

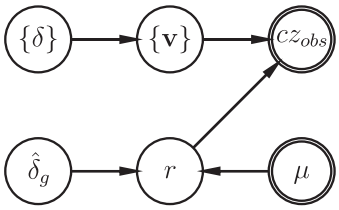

Figure 2. A probabilistic graph showing the dependency of the different observables and the underlying variables. The observables (redshift, $cz_{\rm obs}$, and the distance modulus, μ) are shown with a double circle. In the forward methods of peculiar velocity analysis (see Section 3.3), the redshift of the peculiar velocity tracers are predicted as a function of the distance estimates. This is reflected in the relationship between the observables. Note that, we use an external estimate of the galaxy density field, $\hat{\delta}_g$, in order to correct for the IHM bias.

The RHS of equation (4) consists of two terms, (i) the prior on the density field, $\mathcal{P}(\{\delta\})$ and (ii) the likelihood, $\mathcal{P}(\{cz_{\rm obs}\}|\{\mu\}, \{\delta\}, \{\hat{\delta}_g\})$. We will first discuss the prior we use in this work. In this work, we assume a Gaussian prior on the density field. In the Fourier space, we can write it as

$$\mathcal{P}(\{\delta\}) = \prod_{\boldsymbol{k}} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{|\delta_{\boldsymbol{k}}|^2}{2\sigma_k^2}\right], \tag{5}$$

where $\sigma_k^2 = V_s P(k)$ is the variance in $\delta_{\boldsymbol{k}}$. Here, $P(k)$ is the ΛCDM power spectrum and V_s is the volume of the simulation box. In this work, we fix our cosmological parameters to the fiducial cosmological parameters used in the VELMASS simulation. We note that the present day density field is significantly non-Gaussian. Thus it cannot be well-described using a Gaussian prior. In the future, we can incorporate the method presented here with the BORG framework which uses a non-linear, differentiable gravity model to forward model the present day velocity field from the density field at an initial time when the field is well-described as a Gaussian field. Nevertheless, due to the sensitivity of the velocity field to the large-scale density modes, the Gaussian prior still proves to be a useful approximation.

Assuming that the random components of the radial velocity for the various peculiar velocity tracers are independent, we can write

$$\mathcal{P}\left(\{cz_{\rm obs}\}|\{\mu\}, \{\delta\}, \{\hat{\delta}_g\}\right) = \prod_{i=1}^{N_{\rm PV}} \mathcal{P}\left(cz_{\rm obs}^{(i)}|\mu^{(i)}, \{\delta\}, \{\hat{\delta}_g\}\right). \tag{6}$$

Given the large uncertainties on the distance estimates, we marginalize over the true LOS distance, r, for the peculiar velocity tracer. The likelihood for the individual peculiar velocity tracers can be expressed as

$$\begin{aligned}&\mathcal{P}\left(cz_{\rm obs}^{(i)}|\mu^{(i)}, \{\delta\}, \{\hat{\delta}_g\}\right)\\ &= \int_0^\infty {\rm d}r \mathcal{P}\left(cz_{\rm obs}^{(i)}|r, \{\boldsymbol{v}\}\right)\mathcal{P}(r|\mu^{(i)}, \{\hat{\delta}_g\}\right).\end{aligned} \tag{7}$$

In the above, $\{\boldsymbol{v}\}$ is the linear velocity field calculated from the density field $\{\delta\}$ using equation (1). The first term inside the integral is modelled as a Gaussian with the uncertainty in the observed redshift given by $\sigma_{\rm NL}$, which is the uncertainty in the velocity induced by the

non-linearities in the density field. This term is given as

$$\mathcal{P}\left(cz_{\rm obs}^{(i)}|r,\{\boldsymbol{v}\}\right)=\frac{1}{\sqrt{2\pi\sigma_{\rm NL}^2}}\exp\left[-\frac{\left(cz_{\rm obs}^{(i)}-cz_{\rm pred}^{(i)}(r,\boldsymbol{v})\right)^2}{2\sigma_{\rm NL}^2}\right], \quad (8)$$

where

$$1+z_{\rm pred}(r,\boldsymbol{v})=[1+z_{\rm cos}(r)]\left(1+\frac{v_r(r)}{c}\right). \quad (9)$$

In the above equation, $z_{\rm cos}$ denotes the 'cosmological' recessional redshift. Also note that in equation (7) δ and $\hat{\delta}_g$ are assumed to be independent of each other (see end of this section for more details).

Choice of smoothing length: We use the velocity field calculated using linear theory, equation (1), to model the non-linear velocity of the haloes. Since the velocity field predicted by linear theory breaks down on small scales, we need to take care to make sure that the obtained velocity is unbiased. Therefore, we calculate a low-pass filtered velocity field, $\boldsymbol{v}^R_{\rm smooth}$, by smoothing the linear velocity field with a Gaussian filter of smoothing scale R. The relationship between this smoothed velocity field and the velocity of the haloes, $\boldsymbol{v}_{\rm halo}$, can be modelled as:

$$\boldsymbol{v}_{\rm halo}=m\cdot\boldsymbol{v}^R_{\rm smooth}+\boldsymbol{\epsilon}_v. \quad (10)$$

Here, $\boldsymbol{\epsilon}_v$ is the uncertainty in the velocity predicted using linear theory. $\boldsymbol{\epsilon}_v$ is modelled as a Gaussian with standard deviation, $\sigma_{\rm NL}$. It was shown in Berlind, Narayanan & Weinberg (2000) using N-body simulations that the slope, m, obtained by regressing the smoothed velocity with the 'true' halo velocity depends on the smoothing scale and is not always equal to 1, thus showing that the predicted velocity can be biased depending on the smoothing scale. In Carrick et al. (2015) and Hollinger & Hudson (2021), it was found that the velocity field smoothed with a Gaussian filter of scale $R_{\rm smooth}=4\ h^{-1}$ Mpc results in the slope of regression $m\approx 1$.[2] Since our forward model for the velocity field is the similar to that of Carrick et al. (2015) and Hollinger & Hudson (2021), we use the same smoothing scale of $R_{\rm smooth}=4\ h^{-1}$ Mpc for all subsequently reported velocity fields. At this smoothing scale, the noise in the velocity estimate is $\sigma_{\rm NL}\approx 150$ km s^{-1} (Carrick et al. 2015; Hollinger & Hudson 2021). We fix the value of $\sigma_{\rm NL}$ to this value in our inference. We note that the choice of smoothing scales can potentially be a source of systematic uncertainty. However, we defer a detailed study of this effect to a future work.

The 'forward' method for peculiar velocity analysis is susceptible to homogeneous and IHM biases (Hudson 1994; Strauss & Willick 1995). As mentioned earlier, the homogeneous Malmquist bias arises if we do not account for the fact that there are more galaxies at large radius. Moreover, galaxies are not uniformly distributed in the Universe, instead are clustered at regions of large overdensities. Neglecting these overdensities along the LOS leads to the IHM bias as discussed in Section 3.2. We use an estimate of the LOS galaxy density, $\hat{\delta}_g(r)$, to account for the IHM bias in the radial distribution. Using the galaxy density field, the expected number of peculiar velocity tracers along the LOS can be estimated as

$$n\left(r|\{\hat{\delta}_g\}\right)\propto r^2\left[1+\hat{\delta}_g(r)\right]. \quad (11)$$

Accounting for this dependence on the LOS density, we can write the distribution,

$$\mathcal{P}\left(r|\mu^{(i)},\{\hat{\delta}_g\}\right)\propto n\left(r|\{\hat{\delta}_g\}\right)\exp\left(-\frac{[\mu^{(i)}-\mu(r)]^2}{2\sigma_\mu^{(i)2}}\right), \quad (12)$$

where $\mu(r)=5\log_{10}(r/10\ {\rm pc})$ and $\sigma_\mu^{(i)}$ is the uncertainty on the estimate of the distance modulus. Note that we assume a Gaussian uncertainty on the distance modulus. One of the main sources of systematic uncertainty in estimating the distances to the peculiar velocity tracers is the calibration of the zero-point of the distance–luminosity relation. In order to take care of this uncertainty, we introduce a scale factor, A, so that all the distance estimates in a given catalogue are scaled by the same factor, $\mu\rightarrow\mu+5\log_{10}A$ (or equivalently $d\rightarrow Ad$). We then fit and marginalize over this factor during the reconstruction in a block sampling scheme. When using multiple catalogues, we fit a separate scaling factor for different catalogues.

In this work, the estimated galaxy density field, $\{\hat{\delta}_g\}$, is an additional input to our reconstruction code and is assumed to be independent of the density field that is inferred. This is not fully self-consistent and needs to be improved in the future. We note that we do not use any cross-correlation information between the estimated galaxy density field and the inferred velocity field. The estimated galaxy density field simply provides an estimate for the radial distance of tracer. For the test with the mock catalogues, we use the true underlying particle density to account for the IHM correction. For our run with the real data, we use an iteratively reconstructed density field from Carrick et al. (2015).

Table 1. Depth of the three mock peculiar velocity surveys used in this work. Δ_d is the fractional uncertainty of the peculiar velocity tracers. r_{95} is the radius of 95 per cent completeness for the mock survey and $r_{\rm peak}$ is the radius of peak number density of the peculiar velocity tracers.

	Δ_d	r_{95} (h^{-1} Mpc)	$r_{\rm peak}$ (h^{-1} Mpc)
Mock1	0.07	135	120
Mock2	0.15	143	117
Mock3	0.20	176	113

4 VALIDATION WITH SIMULATIONS

Having introduced our method in the previous section, we now present the results of running our code on a mock peculiar velocity survey. We validate our code by running the code on mock surveys described in Section 2.1 and comparing the density and velocity reconstructions with the ground truth.

In Table 1, we show the depth of the three mock surveys (corresponding to three different levels of distance uncertainty) we used in our reconstruction, showing the radius of 95 per cent completeness of the mock surveys and the peak of the tracer number density. As mentioned previously, our method requires an estimate of the galaxy density field to correct for the IHM bias. In our runs with the simulations, we use the halo density field of the simulation (smoothed with a Gaussian filter of scale 4 h^{-1} Mpc) as the estimate of the galaxy density field.

[2]We note that Lilow & Nusser (2021) found that a smoothing scale of $\sim 7\ h^{-1}$ Mpc results in unbiased velocities. However, the two approaches determine the unbiased smoothing scale in different ways. Lilow & Nusser (2021) reconstruct the linear velocity from mock Two-Micron All-Sky Redshift Survey (2MRS) galaxy catalogue Huchra et al. (2012) and compares it with mock *Cosmicflows-3* peculiar velocity catalogue to determine the value of the smoothing scale for unbiased reconstruction. On the other hand, in the approach used here (Carrick et al. 2015; Hollinger & Hudson 2021), halo velocities are compared against the smoothed N-body velocity field to determine the smoothing scale for unbiased velocity reconstruction. These differences makes the direct comparison between the two approaches difficult.

To compare the impact of the IHM bias correction, we also ran our method without any IHM correction by setting $\hat{\delta}_g = 0$. As we will see, neglecting the IHM bias correction impacts our reconstruction significantly. A visual comparison of the impact of the IHM correction in the velocity reconstruction is shown in Fig. 3, where we plot the reconstructed velocity field and the associated uncertainty with our algorithm in the $Z = 0$ plane. By comparing the reconstructed mean field to the true velocity field in the figure, we can see that our algorithm reconstructs the coherent large-scale velocities well. To estimate the mean and the variance of the reconstructions, we use a total of 250 reconstruction samples from our chain. The reconstructions samples from HMC-based reconstructions can be highly correlated. Therefore, we thin our MCMC samples by a factor of 30, roughly corresponding to the autocorrelation length of our MCMC chains. However, we see that without the IHM bias correction, the reconstructed velocity field has a higher amplitude compared to the true velocity field. This is due to the fact that the IHM bias leads to a spurious velocity in the reconstruction as discussed in Section 3.2. As expected from equation (2), this effect is most pronounced for a distance uncertainty of $\Delta_d = 0.20$. Finally, note that outside the data region, the mean velocity field is suppressed to a smaller value of the velocity. The impact of the IHM bias is most readily seen in the power spectrum of the reconstructed density field. We show the power spectrum of our reconstruction (with and without the IHM bias correction) in the top panel of Fig. 4. We can see that the reconstructed power spectrum without the IHM bias correction is biased high. This can be seen more clearly in the middle panel of Fig. 4 where we plot the ratio of the mean of the inferred power spectrum to the theoretical expectation. As we can see from the figure, the power spectrum of the reconstruction with IHM bias correction is consistent with the true power spectrum of the simulation. On the other hand, if we do not account for the IHM bias, we see excess power in the reconstructions. This is due to the fact that the IHM bias pushes the velocity to a larger value, which in turn increases the variance in the density field. In order to quantify the bias in the power spectrum, we calculate the quantity

$$\Delta_P = \sum_{i=1}^{N_{k-\mathrm{bins}}} \frac{\langle P_{\mathrm{sample}}(k_i)\rangle - P_{\mathrm{theory}}(k_i)}{\sigma[P(k_i)]}, \tag{13}$$

where $\sigma[P(k_i)]$ is the standard deviation in the power spectrum assuming a Gaussian covariance. For an unbiased reconstruction, we expect this quantity to have a value of zero. We compare the bias in the reconstruction by computing the bias in the power spectrum measured in 15 bins in k. For the reconstruction with the IHM bias correction, we obtain values of $\Delta_P = +0.46, -0.25, +0.55$ for the mock surveys with $\Delta_d = 0.07, 0.15, 0.20$, respectively. The same quantities when measured for the reconstruction without the IHM bias correction are $\Delta_P = +8.78, +15.24$, and $+16.99$, respectively, showing that the power spectrum inferred without IHM bias correction is biased high with high significance.

While the power spectrum gives a useful consistency check, the power spectrum does not provide an assessment about the phases of the Fourier modes. In order to test that the phases of the density modes are reconstructed correctly, we compute the cross-correlation rate of the reconstructed density field with the true underlying density field, which is defined as follows:

$$\rho_c(k) = \frac{\langle \delta_{\mathrm{rec}}(\boldsymbol{k})\delta^*_{\mathrm{sim}}(\boldsymbol{k})\rangle_{\boldsymbol{k}\in k\ \mathrm{bin}}}{V_s\sqrt{P_{\mathrm{rec}}(k)P_{\mathrm{sim}}(k)}}, \tag{14}$$

where P_{rec} is the power spectrum of the reconstruction and P_{sim} is the true power spectrum of the simulation. V_s is the volume of the survey. We show the cross-correlation of our reconstructed density samples with the true underlying density field of the simulation in the bottom panels of Fig. 4. As can be seen, the reconstructed large-scale density modes show a cross-correlation of $\sim$0.5 with the true density field, showing that the large-scale density modes are reconstructed well with our method. The cross-correlation approaches zero for the small-scale ($k \gtrsim 0.04\ h\ \mathrm{Mpc}^{-1}$) modes. We can see the impact of the IHM bias on the cross-correlation from this plot. For the reconstruction without IHM bias correction on the mock survey with larger distance uncertainty, cross-correlations of the large-scale density modes are reduced compared to the reconstruction with IHM bias correction. This effect is most readily seen in the reconstruction with $\Delta_d = 0.2$

We however note that, while the cross-correlation in Fourier space is an interpretable metric, it may not be the best metric to compare the two fields. This is because we only have the peculiar velocity data in a small volume of our simulation box. Therefore, we also compare the radial velocity in our reconstruction to the true radial velocity in shells of radial distances. In each shell, we compute the slope, m, and the scatter, σ_v, between the velocity estimates by regressing the reconstructed velocity on the true velocity field. The slope of regression is given as:

$$m = \frac{\left\langle V^r_{\mathrm{rec}} V^r_{\mathrm{true}}\right\rangle}{\left\langle V^{r\ 2}_{\mathrm{true}}\right\rangle} = \rho_c \frac{\sigma_v^{\mathrm{rec}}}{\sigma_v^{\mathrm{true}}}, \tag{15}$$

where V^r_{rec} is the reconstructed velocity and V^r_{true} is the true velocity field. ρ_c is the cross-correlation between the two velocity fields and $\sigma_v^{\mathrm{rec/true}}$ is the standard deviation of the reconstructed and true velocity fields. In case the two velocity fields have the same standard deviation, as is the case for unbiased reconstructions, the slope is equal to the cross-correlation between the two fields in real space. We show the slope from this comparison in Fig. 5. As it can be seen from the figure, the IHM bias-corrected reconstruction has a slope of $m \gtrsim 0.75$ in the data region ($R \lesssim 120\ h^{-1}$ Mpc), showing good cross-correlation with the truth. However, the quality of the reconstruction reduces at high R, reflected in the fact that the value of $m \sim 0$ at these scales. On the other hand, the IHM bias leads the reconstructed velocity to be biased high, due to which $\sigma_v^{\mathrm{rec}} > \sigma_v^{\mathrm{true}}$, resulting in a slope, m, higher than 1, as seen in Fig. 5. Furthermore, we compare the reconstruction error for the different runs of our method. The reconstruction error is defined as the root mean squared (RMS) error between the reconstruction and the true velocity, $(\Delta V_r)^2 = |V_r^{\mathrm{samples}} - V_r^{\mathrm{true}}|^2$. We see that the IHM bias-corrected reconstruction leads to a smaller error in the reconstructed velocity compared to the reconstruction without IHM bias correction. At large radius, the velocity scatter between the reconstructed velocity and the true velocity approaches the scatter of the haloes in the VELMASS simulation. We discuss in more detail the different sources contributing to this reconstruction error in Appendix A.

The above tests show that we get an unbiased reconstruction of the velocity field with our Bayesian reconstruction method after correcting for IHM bias. On the other hand, not correcting for the IHM bias can substantially bias the velocity field reconstruction.

5 VELOCITY RECONSTRUCTION IN THE LOCAL UNIVERSE

In this section, we apply our method on real peculiar velocity data to reconstruct the velocity field of the local Universe. The description of the peculiar velocity surveys used in this work is provided in Section 2.2. By sampling density fields from the posterior (4), we get samples of reconstruction that are consistent with the observed

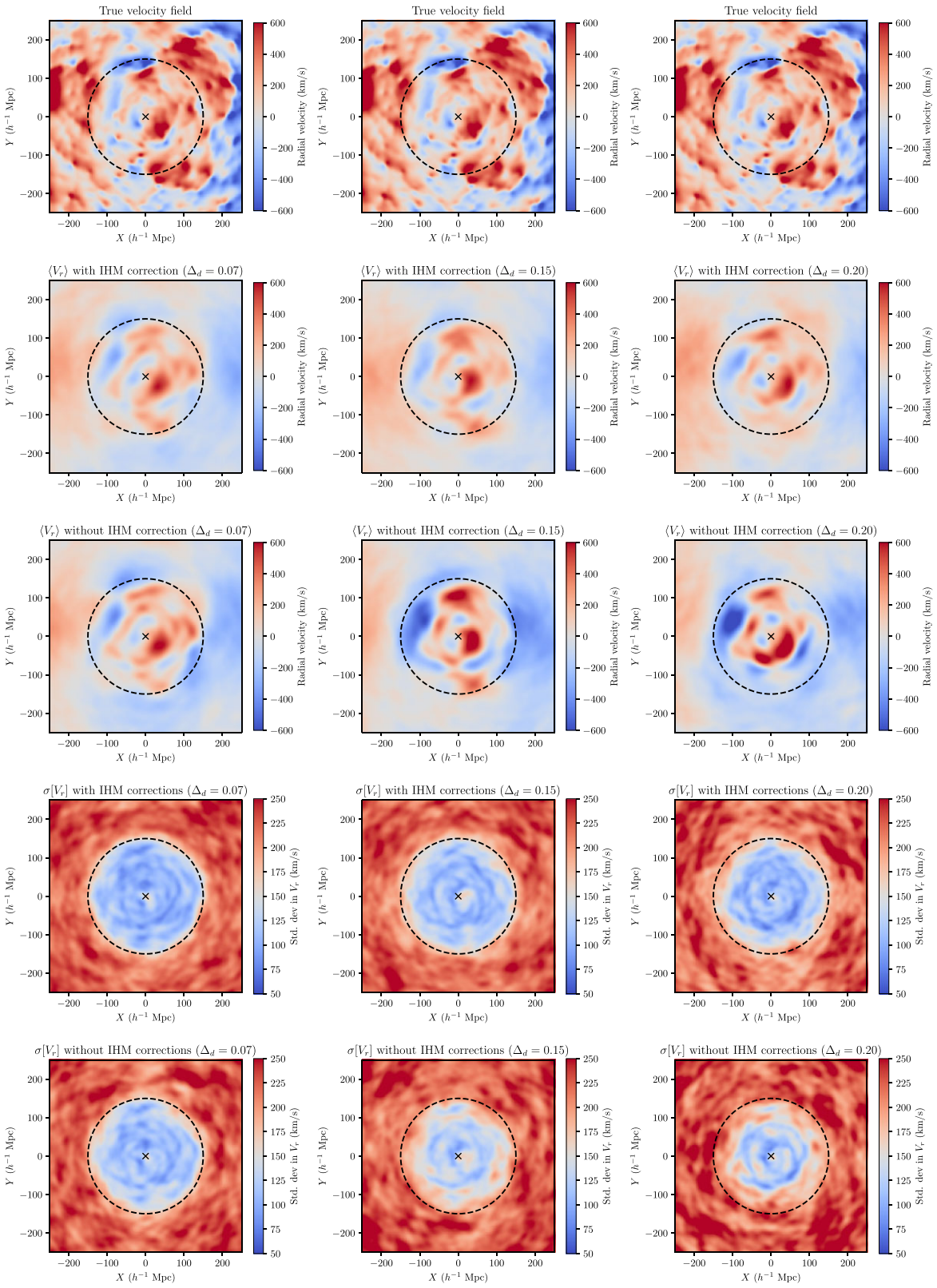

Figure 3. A visual comparison of the reconstructed radial velocity field and the associated uncertainties with the true velocity field in the $Z = 0$ plane in our mock simulation. We show the reconstruction for three different mock surveys with a distance error of $\Delta_d = 0.07$ (left), $\Delta_d = 0.15$ (centre), and $\Delta_d = 0.20$ (right). The top row is the true velocity field in the simulation in the $Z = 0$ plane. The black dashed circle is at a radius of 150 h^{-1} Mpc, which is close to the 95 per cent completeness radius for each survey scenario. The second and the third rows show the mean velocity for the samples of our reconstructions with and without IHM bias corrections, respectively. As can be seen, if we do not correct for the IHM bias, the velocity field in the data region is boosted up compared to the true velocity field. The bottom two rows show the standard deviation in the reconstruction samples with and without the IHM bias correction.

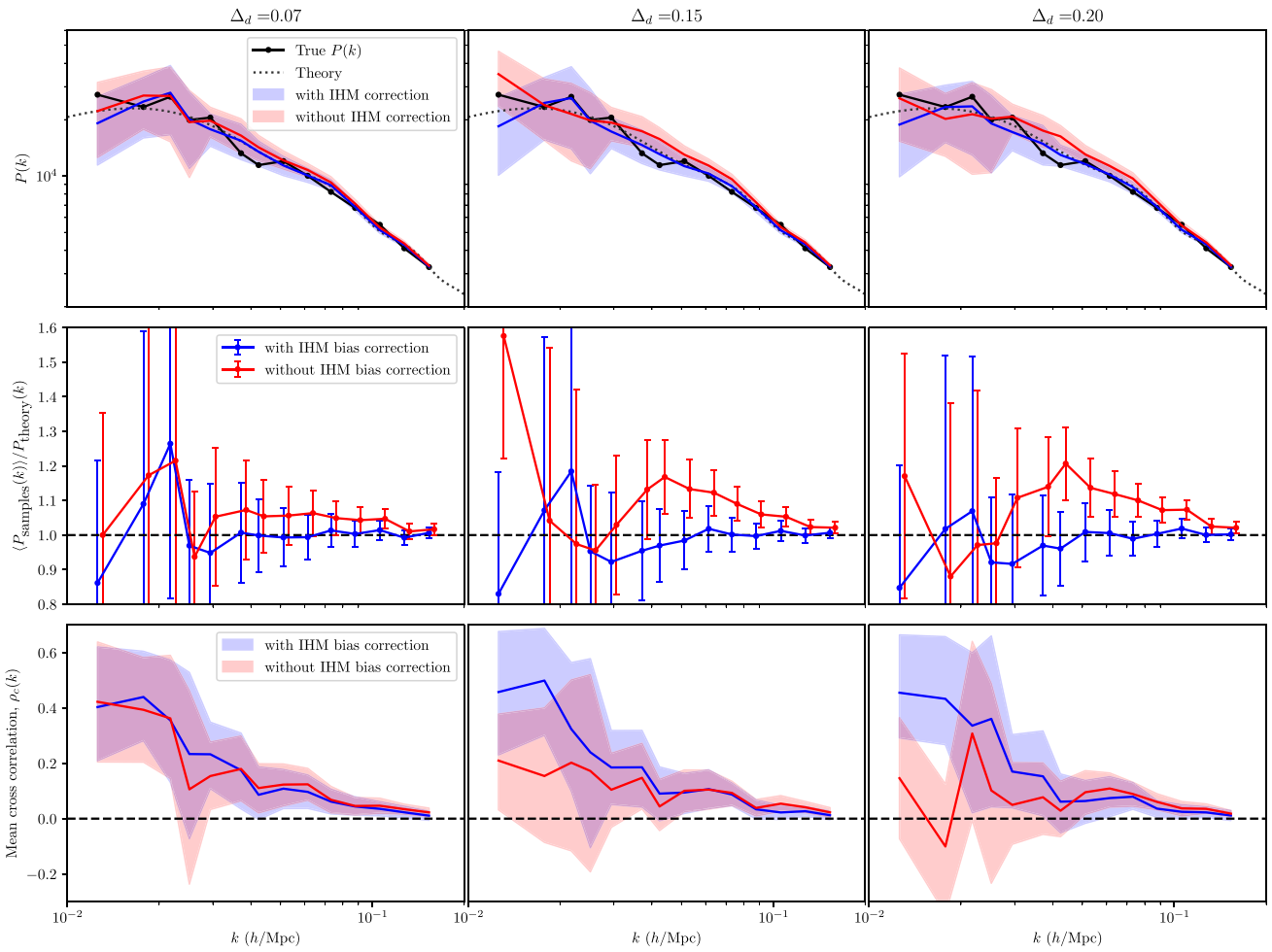

Figure 4. The power spectrum (top), ratio of the power spectrum to the theoretical expectation (middle) and the cross-correlation of the reconstructed density field with the true density field (bottom) for the three mock surveys with distance uncertainties, $\Delta_d = 0.07$ (left), $\Delta_d = 0.15$ (centre), and $\Delta_d = 0.20$ (right). In the top panels, we plot the power spectrum inferred from our reconstruction and compare it with the true power spectrum of the simulation (shown with black line) and the theoretical expectation (shown as a black dotted line). The red lines show the power spectrum of the reconstruction without IHM bias correction and the blue lines are the corresponding curves with IHM bias correction. The shaded region show the 68 per cent confidence interval as calculated from the reconstruction samples. In the middle panels, we plot the ratio of the mean of the power spectrum of our reconstructions to the theoretical expectation. The error bars are calculated assuming a Gaussian covariance for the power spectrum. Note that this is different from the error bars in the top panel. The two quantities will be the same only if the reconstruction samples are drawn from the prior distribution. As we can see the power spectrum of the reconstruction without IHM bias correction is biased high. The magnitude of the bias is higher for the reconstruction with larger distance uncertainty. Finally, in the bottom panel, we plot the cross-correlation of the reconstructed density field with the true density field of the simulation. Similar to the top panel, the shaded region shows the 68 per cent confidence interval as calculated from the reconstruction samples. The cross-correlation is the largest at large scales (small k) and it decreases to 0 at small scales (high k). However, we see that if the IHM bias is not corrected, the large-scale cross-correlation goes down, suggesting that these density modes are not reconstructed well.

data. One useful side product of sampling from the posterior is that we get an estimate of the correlated uncertainties in the reconstructed velocity field. For the reconstruction of the local Universe, we use a cubic sampling box of side length, $L_{\rm box} = 500\ h^{-1}$ Mpc with 128^3 grids. In Fig. 6, we plot the power spectrum of the samples of our reconstruction. In the Bayesian reconstruction approach, we expect the reconstructed field to have a similar distribution as the prior, unless the data contradicts the assumed prior. Here we assume a Gaussian prior on the density field, with the power spectrum following a cosmological power spectrum consistent with Planck CMB results (Planck Collaboration VI 2020). Therefore, we expect the reconstruction to have a cosmological power spectrum. However, as we saw in Section 4, IHM bias may artificially enhance the inferred power spectrum. Thus, consistency between the theoretical power spectrum and the inferred power spectrum shows that the IHM bias is corrected in our reconstruction. As can be seen from the figure, this is indeed true and the power spectrum computed from the samples of our posteriors agree very well with the theoretical prediction from the ΛCDM model.

5.1 Comparison with 2M++ reconstruction

Unlike simulations, for the real data we do not have access to the true velocity field. None the less, we can compare our reconstruction to velocity fields reconstructed using other methods. Therefore, we compare our reconstructed velocities to the reconstruction of Carrick et al. (2015) who used an iterative reconstruction procedure,

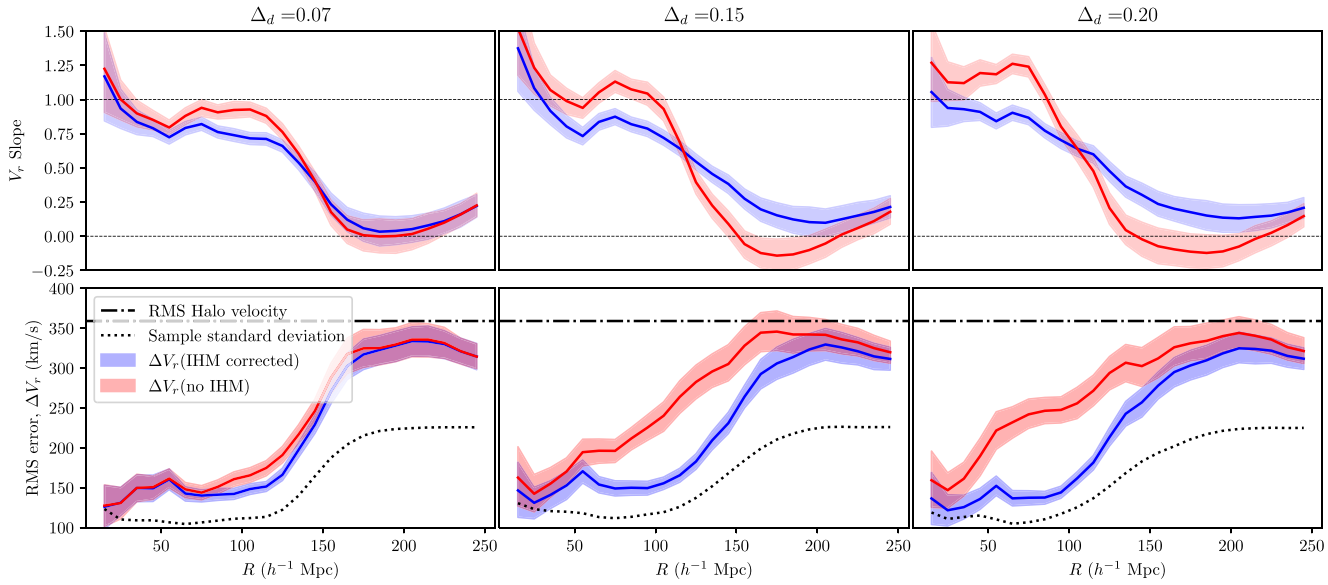

Figure 5. (Top) Slope of regression of the reconstructed V_r on the true V_r values from the simulation, binned in radial shells and (bottom) the RMS error between the reconstructed velocity and the true velocity. In order to regress the velocity fields, we compare the point estimates of the radial velocity calculated in voxels of the reconstruction grid within a given radial shell. The three horizontal panels show the results for mock surveys with three distance uncertainties: $\Delta_d = 0.07$ (left), $\Delta_d = 0.15$ (centre), and $\Delta_d = 0.20$ (right). In the top panel, we plot the slopes of regression in radial shells. The red lines show the results without the IHM bias correction and the blue lines are the results after correcting for the IHM bias. We see that the slope of regression $m \sim 1$ deep inside the data region and then drops to $m \sim 0$ at large radius. Without the IHM bias correction, the slope can be higher than 1 since IHM bias introduces spurious velocities. In the bottom panel, we plot the reconstruction error in radial bins. The blue and the red lines show the RMS of the difference of reconstructed and true values of the radial velocity, with and without the IHM bias correction as a function of radius. The dotted line shows the standard deviation in our reconstruction samples. The dashed–dotted line shows the velocity scatter of the haloes in the VELMASS simulation. At large radius, the velocity error in our reconstruction approaches the velocity scatter of the haloes.

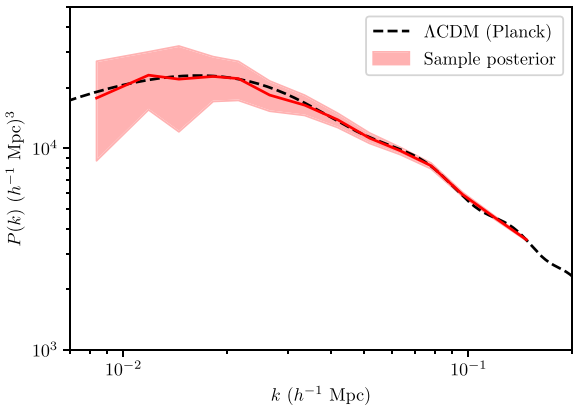

Figure 6. Power spectrum of the samples of reconstruction of the local Universe. The reconstruction was performed using the 2MTF and the SFI++ TF catalogues and the A2 supernovae compilation. The density reconstruction from Carrick et al. (2015) was used for the IHM bias correction.

where the galaxy density is reconstructed from the galaxies in the 2M++ galaxy compilation (Lavaux & Hudson 2011). The velocity is predicted from the galaxy density field using linear perturbation theory. However, since galaxies are a biased tracer of the underlying density field, the velocity estimate from the galaxy field needs to be scaled by a factor of $\beta = f/b$. The value of β is then fitted by comparing the predicted velocities to the the peculiar velocity data from a peculiar velocity survey. In this work, we use the best fit β and $\boldsymbol{V}_{\rm ext}$ value fitted in Carrick et al. (2015). We show a visual comparison of our velocity reconstruction with the reconstruction of Carrick et al. (2015; hereafter C15) in the galactic X, Y, Z planes in Fig. 7. The top panels of Fig. 7 show the radial velocity field in the reconstruction of C15. The middle and the bottom panels show the mean and the standard deviation of the velocity field samples reconstructed using our algorithm. Similar to Section 4, we use a total of 250 independent samples from the chain for estimating the mean and the standard deviation from the chains. By comparing our velocity reconstruction to the 2M++ reconstruction, we see the similarity between some of the same large-scale features in both reconstruction. Note that in the region without data, the mean field approaches very small values. This is similar to the Wiener filtered fields, where in the region without data, the field is strongly suppressed such that in the limit of no data, the velocity field approaches zero. In the bottom panel, we plot the uncertainty in the velocity estimates in these planes. Since our method produces samples of reconstruction from the posterior with the Gaussian prior, we get the correlated uncertainties in velocity estimates of our reconstruction. From the figure, we see that the uncertainties in the velocity estimates are substantially lower in the inner region than the outer region which does not have peculiar velocity data. This is because the inner region contains almost all the peculiar velocity tracers. We note that the uncertainty obtained from our reconstructions does not account for the error associated with our modelling choices, e.g. the assumption of linear theory, choice of smoothing scale, etc. Quantifying these sources of 'systematic' uncertainties is beyond the scope of this work. Therefore the uncertainty obtained from our reconstructions should be treated as the statistical error. We also plot the velocity field

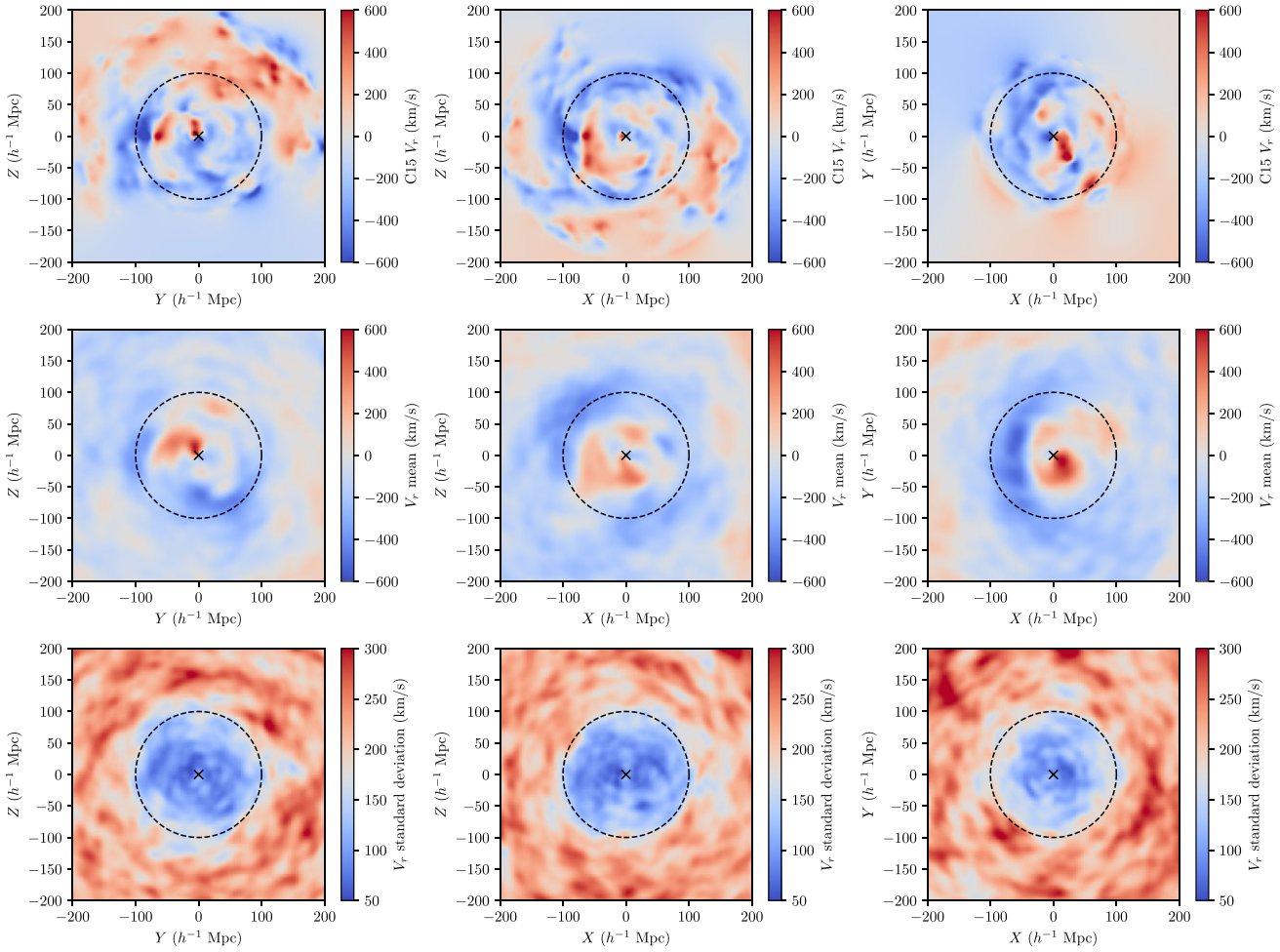

Figure 7. A comparison of our reconstruction with the reconstruction of Carrick et al. (2015) in the galactic X (left), Y (centre), and Z (right) planes. The top panels show the reconstruction of Carrick et al. (2015), the middle and the bottom panels show the mean and the standard deviation of the velocity reconstruction from our method. We can see similar large-scale features in the velocity field in both our reconstruction and the C15 reconstruction. We show the radial distance of 100 h^{-1} Mpc with a black dashed circle. This is the distance cut we use for our peculiar velocity tracer sample. Note that the uncertainty in the velocity field reconstruction increases drastically beyond this boundary.

and the density field in the supergalactic plane, comparing also the position of some prominent galaxy clusters in Fig. 8. From the mean field, we see that while the overdensities around the Perseus–Pisces (PP), Virgo and Coma clusters are well-reconstructed, we do not see prominent overdensities around the Norma and the Shapley clusters. Since the Shapley cluster lies outside the 100 h^{-1} Mpc, the threshold of our reconstruction, and the Norma cluster lies along the galactic plane, there is lack of peculiar velocity data around these clusters. The lack of peculiar velocity data around the cluster leads to poor reconstruction quality around these regions.

Finally, we compare our reconstruction to the C15 reconstruction by comparing the velocity fields in radial shells. For this comparison, we compared the radial velocity from C15 and our reconstruction in radial shells of width 20 h^{-1} Mpc. In each shell, we compare the radial velocities from our reconstruction samples and the reconstruction from C15. We fit the slope and the standard deviation between the two velocity estimates by regressing the velocity estimates of our reconstruction on the C15 velocity estimates. The fitted slope can be thought of as a proxy for the cross-correlation between the two velocity estimates. The slope and the scatter between the two velocity fields are shown in Fig. 9. As we can see from the figure, the two velocity fields have a large degree of correlation in the nearby Universe ($R \lesssim 120\ h^{-1}$ Mpc). Since the peculiar velocity data is restricted to $d < 100\,h^{-1}$ Mpc, we do not expect the reconstructed velocity field to be a good estimate beyond this boundary. This is reflected in the fact that the slope rapidly approaches zero beyond the peculiar velocity data boundary. We also see that the scatter between the two increases with increasing radius. At very low radius ($R \lesssim 50\ h^{-1}$ Mpc), the scatter between the two velocity fields is less than or comparable to the uncertainty due to the non-linearities, $\sigma_{\rm NL}$ $\approx$ 150 km s^{-1}. At large radius, the scatter approaches the sample standard deviation ($\sim$250 km s^{-1}).

5.2 Bulk flow

Due to the sensitivity to large-scale density modes, the velocity field is correlated on very large scales. The large-scale velocity field is often expanded in terms of its kinematic moments, such as the dipole and the quadrupole of the velocity field (Jaffe & Kaiser 1995),

$$\boldsymbol{V}(d) = \boldsymbol{U} + d\mathbf{Q}\cdot\hat{\boldsymbol{r}} + ..., \tag{16}$$

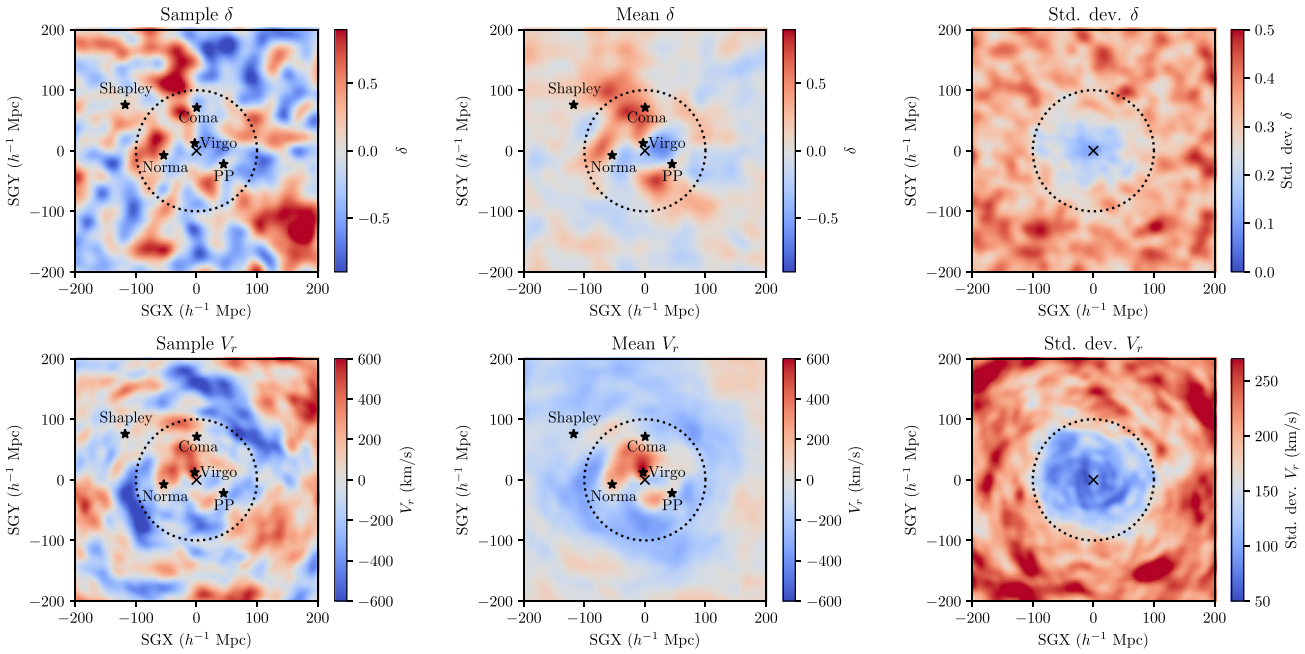

Figure 8. The density and the velocity field reconstruction in the supergalactic plane. The top panels shows the density field smoothed using a Gaussian filter of smoothing scale 10 h^{-1} Mpc and the bottom panels shows the velocity field reconstruction. In the left-hand panels, we show the density and the velocity field of a randomly selected sample in our reconstruction. The middle panels show the mean density and velocity field. The rightmost panels show the standard deviation in our density and velocity field reconstruction samples. We show the location of several prominent clusters in the supergalactic plane, namely Virgo, Norma, PP, Coma, and Shapley, with a black star. The black dotted line shows the radius at 100 h^{-1} Mpc, which is the selection limit imposed on the sample.

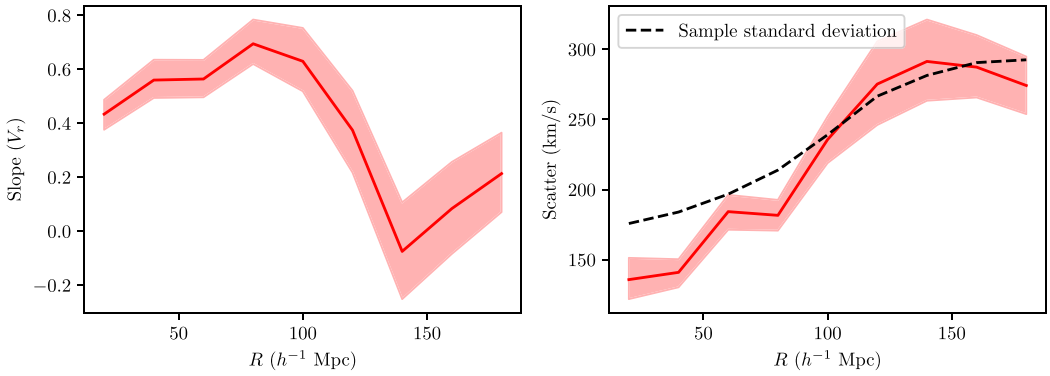

Figure 9. Comparison of (left) the slope and (right) the RMS difference between the velocity estimates of our reconstruction and the C15 reconstruction for voxels binned in radial shells. The slope can be interpreted as the cross-correlation between the two velocity fields. We see a large degree of cross-correlation in the region with the peculiar velocity data ($R \lesssim 100\ h^{-1}$ Mpc), while it rapidly goes to zero beyond this region. The right-hand panel shows the RMS error between the two velocity reconstructions. At very small radius, $R \lesssim 50\ h^{-1}$ Mpc, the scatter approaches the value $\sigma_{\rm NL} \approx 150$ km s^{-1}. The black dashed line show the standard deviation in our reconstruction sample with $\sigma_{\rm NL}$ added in quadrature.

where $\boldsymbol{U}$ is the dipolar bulk flow, and $\mathbf{Q}$ is the trace-free, symmetric quadrupolar shear moment and d is the radial distance. The bulk flow is sensitive to the large-scale power of the density field and therefore can also be used to constrain cosmological parameters due to its sensitivity to the power spectrum shape parameter (Feldman & Watkins 2008).

We can measure the bulk flow and the shear moments directly from our reconstruction. In order to measure the bulk flow in our reconstructions at a given scale, we smooth the velocity field with a window function of characteristic scale, $R_{\rm bulk}$. We use a Gaussian filter in order to measure the bulk flow in our reconstruction. The bulk flow at a scale of $R_{\rm bulk} = 40\ h^{-1}$ Mpc for our reconstruction is $V_{\rm bulk} = 220^{+20}_{-21}$ km s^{-1} in the direction $l = 295^\circ \pm 6^\circ$, $b = 21^\circ \pm 5^\circ$ deg. As noted in the previous section, the error bar obtained from the reconstruction should be treated as the statistical error bar and does not account for systematic uncertainties. The bulk flow as measured in our reconstruction is compared to a number of different results in the literature in Table 2. We also show the total magnitude and the direction of the bulk flow as measured from our reconstruction samples in Fig. 10. As we can see from the figure, both the magnitude and the direction of the bulk flow are consistent with other results in the literature. Of the compared results, Scrimgeour et al. (2016) and

Table 2. Comparison of the bulk flow results with other studies in the literature. The results are also shown in Fig. 10. We see that our bulk flow result is in excellent agreement with the other results in literature.

Reference	Abbreviation	Effective radius	$\lvert \boldsymbol{V}_{\rm bulk} \rvert$ (km s^{-1})	l (deg)	b (deg)
Scrimgeour et al. (2016)	S16	$50\,h^{-1}$ Mpc	248 ± 58	318 ± 20	40 ± 13
Qin et al. (2019)	Q17	$37\,h^{-1}$ Mpc	259 ± 15	300 ± 4	23 ± 3
Boruah et al. (2020)	B20	$40\,h^{-1}$ Mpc	252 ± 11	293 ± 5	14 ± 5
Lilow & Nusser (2021)	L21	$50\,h^{-1}$ Mpc	274 ± 50	287 ± 9	11 ± 10
This work	–	$\mathbf{40\,h^{-1}}$ **Mpc**	$\mathbf{220 \pm 21}$	$\mathbf{295 \pm 6}$	$\mathbf{21 \pm 5}$

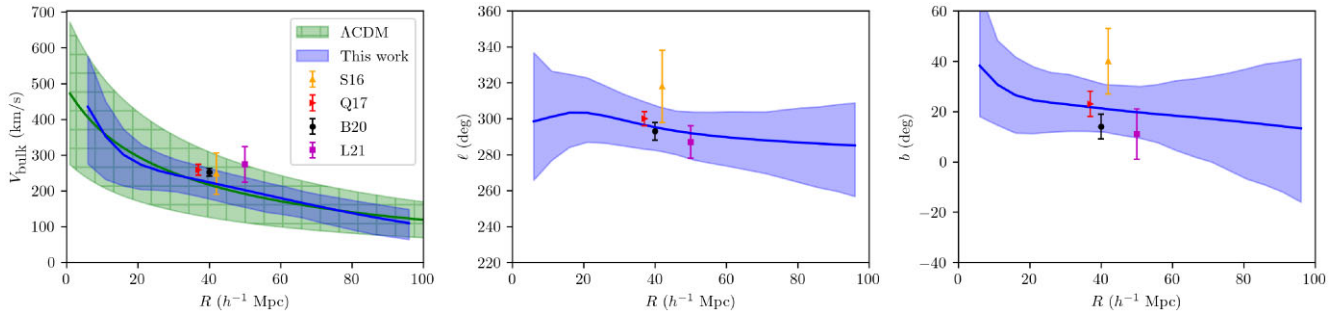

Figure 10. Comparison of the bulk flow measured in our reconstruction and compared to the ΛCDM expectation as well as the other results in the literature. The left-hand panel shows the magnitude of the bulk flow, while the middle and the right-hand panels show the direction of the bulk flow in terms of the galactic longitude and latitude, respectively. As we can see, the magnitude of the bulk flow in our reconstruction (shown in blue) is consistent with both the ΛCDM expectation (shown in green hatch) as well as other results in the literature. The direction of the bulk flow as inferred from our reconstruction is also consistent with other results in the literature. The results in the literature compared in this results are: Scrimgeour et al. (2016) (S16), Qin et al. (2019) (Q17), Boruah et al. (2020) (B20), and Lilow & Nusser (2021) (L21).

Qin et al. (2019) use an estimator to estimate the bulk flow directly from the peculiar velocity data. On the other hand, Boruah et al. (2020) and Lilow & Nusser (2021) explicitly fit an external dipole in their flow model to account for velocity contribution from outside the survey volume. Note that the approach presented here is different from these approaches. We do not explicitly model an additional dipole. The density modes we fit for in our reconstruction not only account for the bulk flow but also other higher kinematic moments of the large-scale velocity field.

5.3 Velocity field comparison

We introduced a Bayesian model comparison framework in Boruah et al. (2021) to compare the performance of different velocity reconstruction models. In this model comparison framework, we look at the Bayes factor between two models, $\mathcal{M}_1$ and $\mathcal{M}_2$,

$$\text{Bayes factor} = \frac{\mathcal{P}(\mathcal{D}|\mathcal{M}_1)}{\mathcal{P}(\mathcal{D}|\mathcal{M}_2)}. \tag{17}$$

If the Bayes factor is greater than 1, the model $\mathcal{M}_1$ is preferred over the model $\mathcal{M}_2$ and vice versa. We use this model comparison framework to assess the quality of the velocity reconstruction method introduced in this paper. We compare our velocity model to two different velocity reconstructions of the local Universe – (i) reconstruction of C15, (ii) an adaptive kernel-smoothed velocity reconstruction, where the peculiar velocity data is directly smoothed (Springob et al. 2014, 2016). We use an adaptive-kernel smoothed velocity field obtained from a combined TF catalogue consisting of the SFI++ and the 2MTF velocity field. The Bayesian model comparison method relies on our capacity to predict new data points from the model. This is called a posterior predictive test, and thus requires what is sometimes called a 'test set' of peculiar velocities. Therefore for this comparison, we perform a reconstruction using only the TF catalogues for the reconstruction and use the A2 sample as a test set for the posterior predictive analysis. We note that this is the same peculiar velocity data set used for the adaptive kernel-smoothed reconstruction.

Table 3. Comparison of the logarithm of Bayes factor for various redshift selection. A positive value of the logarithm of Bayes factor implies that our model is favored over the compared model, while a negative value of the logarithm of the Bayes factor implies the compared model is preferred over our model. From the table, we can see that our model performs better than the adaptive kernel-smoothed velocity field. On the other hand, C15 velocity field performs better than our Bayesian velocity reconstruction.

Test set	Selection	$\ln\left(\frac{\mathcal{P}_{\text{fwd-PV}}}{\mathcal{P}^{\text{TF}}_{\text{adaptive}}}\right)$	$\ln\left(\frac{\mathcal{P}_{\text{fwd-PV}}}{\mathcal{P}_{\text{2M++}}}\right)$	N_{tracers}
A2	$z < 0.01$	7.44	−2.12	49
	$z < 0.015$	10.88	−5.92	92
	$z < 0.02$	21.73	−20.69	168
	$z < 0.03$	41.16	−27.65	310

The results of the model comparison are shown in Table 3. Using our model comparison, we find that the Bayesian velocity reconstruction performs better than the kernel-smoothed velocity field, even though both reconstructions use the same data sets. Our velocity field however does not perform as well at predicting new velocity data compared to the reconstruction of Carrick et al. (2015), which uses the 2M++ galaxies to map the density field, and linear perturbation theory to predict peculiar velocities. This is not a failure of our method, but rather a limitation of the peculiar velocity data used for the reconstruction. Since the peculiar velocity data used in this work are very noisy and sparse, the quality of reconstruction is not comparable to the 2M++ reconstruction. In the future, with denser peculiar velocity samples, we expect the quality of reconstruction to be significantly better. There is

also the possibility that the approximation of linear perturbation theory may fail which would require more complex models based, e.g. on BORG. Furthermore, we see that both the adaptive-kernel smoothed and the Bayesian velocity field perform worse at higher redshifts compared to the reconstruction of Carrick et al. (2015). This degradation is explained by two factors that affect the PV tracer based reconstructions: (i) The PV tracers are sparser at higher redshift. (ii) The absolute uncertainty of distances (and therefore velocity) is higher at higher redshifts.

6 DISCUSSION

Our work provides a way towards including peculiar velocity data into the initial condition reconstruction framework BORG. A density reconstruction by combining both galaxy and peculiar velocity data in BORG potentially has multiple advantages. First, using BORG, we can access the fully non-linear velocity field, which is accurate to much smaller scales as compared to the velocity field from linear perturbation theory. For example, BORG particle mesh runs have been shown to perform better than linear velocity fields in *N*-body simulations (Mukherjee et al. 2021). Furthermore, the vorticity of the velocity field, which is a uniquely non-linear phenomenon was already reconstructed in Jasche & Lavaux (2019). In Prideaux-Ghee et al. (2022), the first steps were taken towards the inclusion of the peculiar velocity field in BORG. Second, adding the velocity data to the Bayesian reconstructions can potentially make the reconstruction robust to the details of the galaxy bias model. As was shown in Nguyen et al. (2021), the amplitude of the inferred power spectrum in BORG reconstructions with galaxy/halo catalogues alone may be substantially biased. This arises because the higher order bias terms are not sufficiently informative to break the degeneracy between linear galaxy bias and the amplitude of matter clustering, σ_8. Since peculiar velocity data is sensitive to the total matter density, the addition of peculiar velocity data to galaxy data would make the BORG reconstruction robust to the details of the galaxy bias. Adding velocity information to the Bayesian reconstruction methods can potentially lead to improved reconstruction of the full phase space structure (density + velocity) of dark matter (Leclercq et al. 2017). Such phase space reconstruction can then potentially guide observation efforts for new discoveries (Kostić et al. 2022).

In the future, we can also extend our method to simultaneously constrain cosmological parameters from peculiar velocity data. As we highlight in this paper, not accounting for the IHM bias can lead to the amplitude of the peculiar velocities and so the inferred power spectrum of the density/velocity field to be biased high. This is crucial for the inference of the cosmological parameters, since this bias will likely translate to the biased inference of cosmological parameters. Thus, this paper provides the foundation of inferring cosmological parameters with peculiar velocity data in a Bayesian framework. Also note that our method is sensitive to the large-scale flows, which in turn are a sensitive probe of the break in the matter power spectrum (e.g. Feldman & Watkins 2008). Thus, the large-scale power spectrum can constrain the shape parameter, $\Gamma \approx \Omega_{\rm m} h$. Combining these constraints from large-scale flows with traditional velocity–velocity comparison methods can potentially provide additional cosmological information. The use of the non-linear velocity field using BORG can also potentially break the $f\sigma_8$ degeneracy through the non-linearities. This feature was already noted in the context of weak lensing by using a forward model based on Lagrangian Perturbation theory within the BORG framework for mock weak lensing data (Porqueres et al. 2022).

Another possible application of our Bayesian velocity reconstructions is in using the kinetic Sunyaev–Zeldovich (kSZ) data to reconstruct the large-scale velocity modes. kSZ-based velocity reconstruction has been shown to improve the constraints on primordial non-Gaussianities through multitracer analyses (Giri & Smith 2022).

7 SUMMARY

In this paper, we introduced a forward-modelled velocity reconstruction method which uses peculiar velocity data to reconstruct the velocity field. Using an external estimate of the LOS density, we consistently correct for the IHM bias. Using mocks created from an *N*-body simulation, we validated that our method leads to unbiased velocity field reconstruction after accounting for the IHM bias. However, as we show using mock simulations, not accounting for the IHM bias in forward-modelled methods can lead to significantly biased reconstruction. Since the IHM bias may induce spurious flows, the inferred power spectrum of the reconstruction is biased high if the IHM bias is not corrected for. We then applied our method to the 2MTF, SFI++ TF catalogues and the A2 supernovae compilation, resulting in a novel forward-modelled velocity field reconstruction of the local Universe. Since we sample our velocity field reconstruction samples from the field-level posterior, we get an estimate of the full correlated uncertainties in the peculiar velocity estimates. Furthermore, we found that the bulk flow calculated from our reconstruction is consistent with other results in the literature. Using a Bayesian model comparison framework, we showed that the reconstructed forward modelled velocity fields perform better than the widely used adaptive kernel-smoothed velocity fields constructed from the same data. However, the reconstruction presented here does not perform as well as the velocity field reconstructed from the 2M++ galaxy catalogue. This is likely because peculiar velocity data sets are sparser and noisier compared to galaxy catalogues. Our method provides a way to extend forward-modelled initial condition reconstruction algorithms such as BORG by including peculiar velocity data sets.

ACKNOWLEDGEMENTS

We thank Eleni Tsaprazi, Florent Leclercq, and James Prideaux-Ghee for useful feedback on the manuscript. This work has been done as part of the activities of the Domaine d'Intérêt Majeur (DIM) 'Astrophysique et Conditions d'Apparition de la Vie' (ACAV), and received financial support from Région Ile-de-France. GL acknowledges financial support from the ANR BIG4, under reference ANR-16-CE23-0002. GL acknowledges support by the Simons Collaboration on 'Learning the Universe'. MH acknowledges the support of an NSERC Discovery Grant. This work is done within the Aquila Consortium.[3] This work has made use of the Infinity Cluster hosted by Institut d'Astrophysique de Paris. We thank Stephane Rouberol for running this cluster smoothly for us.

DATA AVAILABILITY

The Second Amendment (A2) supernovae compilation is available with the supplementary data of https://doi.org/10.1093/mnras/staa2485. The 2MTF and the SFI++ catalogues are publicly available with their respective publications that are cited in Section 2. The 2M++ reconstruction used in this work is publicly available at

[3]https://www.aquila-consortium.org/

https://cosmicflows.iap.fr/. The data products generated in this work will be shared on reasonable request to the authors.

REFERENCES

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Berlind A. A., Narayanan V. K., Weinberg D. H., 2000, ApJ, 537, 537
Bertschinger E., Dekel A., 1989, ApJ, 336, L5
Bertschinger E., Dekel A., Faber S. M., Dressler A., Burstein D., 1990, ApJ, 364, 370
Boruah S. S., Hudson M. J., Lavaux G., 2020, MNRAS, 498, 2703
Boruah S. S., Hudson M. J., Lavaux G., 2021, MNRAS, 507, 2697
Boruah S. S., Rozo E., Fiedorowicz P., 2022, MNRAS, 516, 4111
Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575
Bradbury J., Frostig R., Hawkins P., Johnson M. J., Leary C., Maclaurin D., Wanderman-Milne S., 2018, JAX: Composable Transformations of Python + NumPy Programs. Available at: http://github.com/google/jax
Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, MNRAS, 450, 317
Feldman H. A., Watkins R., 2008, MNRAS, 387, 825
Fiedorowicz P., Rozo E., Boruah S. S., Chang C., Gatti M., 2022, MNRAS, 512, 73
Foley R. J. et al., 2018, MNRAS, 475, 193
Ganeshalingam M., Li W., Filippenko A. V., 2013, MNRAS, 433, 2240
Giri U., Smith K. M., 2022, J. Cosmol. Astropart. Phys., 2022, 028
Graziani R., Courtois H. M., Lavaux G., Hoffman Y., Tully R. B., Copin Y., Pomarède D., 2019, MNRAS, 488, 5438
Hicken M., Wood-Vasey W. M., Blondin S., Challis P., Jha S., Kelly P. L., Rest A., Kirshner R. P., 2009, ApJ, 700, 1097
Hoffman Y., Ribak E., 1991, ApJ, 380, L5
Hoffman Y., Courtois H. M., Tully R. B., 2015, MNRAS, 449, 4494
Hoffman Y. et al., 2018, Nat. Astron., 2, 680
Hollinger A. M., Hudson M. J., 2021, MNRAS, 502, 3723
Hong T. et al., 2019, MNRAS, 487, 2061
Huchra J. P. et al., 2012, ApJS, 199, 26
Hudson M. J., 1994, MNRAS, 266, 468
Jaffe A. H., Kaiser N., 1995, ApJ, 455, 26
Jasche J., Lavaux G., 2019, A&A, 625, A64
Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894
Jones D. O. et al., 2019, ApJ, 881, 19
Kodi Ramanah D., Charnock T., Lavaux G., 2019, Phys. Rev. D, 100, 043515
Kostić A., Jasche J., Kodi Ramanah D., Lavaux G., 2022, A&A, 657, L17
Krisciunas K. et al., 2017, AJ, 154, 211
Lavaux G., 2016, MNRAS, 457, 172
Lavaux G., Hudson M. J., 2011, MNRAS, 416, 2840
Leclercq F., Jasche J., Lavaux G., Wandelt B., Percival W., 2017, J. Cosmol. Astropart. Phys., 2017, 049
Lilow R., Nusser A., 2021, MNRAS, 507, 1557
Lynden-Bell D., Faber S. M., Burstein D., Davies R. L., Dressler A., Terlevich R. J., Wegner G., 1988, ApJ, 326, 19
Masters K. L., Springob C. M., Haynes M. P., Giovanelli R., 2006, ApJ, 653, 861
Masters K. L., Springob C. M., Huchra J. P., 2008, AJ, 135, 1738
Millea M. et al., 2021, ApJ, 922, 259
Millea M., Anderes E., Wandelt B. D., 2020, Phys. Rev. D, 102, 123542
Modi C., Feng Y., Seljak U., 2018, J. Cosmol. Astropart. Phys., 2018, 028
Mukherjee S., Lavaux G., Bouchet F. R., Jasche J., Wandelt B. D., Nissanke S., Leclercq F., Hotokezaka K., 2021, A&A, 646, A65
Neal R., 1993, Technical Report CRG-TR-93-1. Dept. of Computer Science. University of Toronto, Toronto
Neal R., 1996, Lecture Notes in Statistics, No. 118. Springer-Verlag, Berlin
Nguyen N.-M., Schmidt F., Lavaux G., Jasche J., 2021, J. Cosmol. Astropart. Phys., 2021, 058
Peebles P. J. E., 1980, The Large-scale Structure of the Universe. Princeton Univ. Press, Princeton
Peterson E. R. et al., 2022, ApJ, 938, 112
Pike R. W., Hudson M. J., 2005, ApJ, 635, 11
Planck Collaboration VI, 2020, A&A, 641, A6
Porqueres N., Heavens A., Mortlock D., Lavaux G., 2021, MNRAS, 502, 3035
Porqueres N., Heavens A., Mortlock D., Lavaux G., 2022, MNRAS, 509, 3194
Prideaux-Ghee J., Leclercq F., Lavaux G., Heavens A., Jasche J., 2022, preprint (arXiv:2204.00023)
Qin F., Howlett C., Staveley-Smith L., Hong T., 2019, MNRAS, 482, 1920
Scrimgeour M. I. et al., 2016, MNRAS, 455, 386
Springob C. M., Masters K. L., Haynes M. P., Giovanelli R., Marinoni C., 2007, ApJS, 172, 599
Springob C. M. et al., 2014, MNRAS, 445, 2677
Springob C. M. et al., 2016, MNRAS, 456, 1886
Strauss M. A., Willick J. A., 1995, Phys. Rep., 261, 271
Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 2013, 036
Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, MNRAS, 389, 1284
Tully R. B., Courtois H. M., Sorce J. G., 2016, AJ, 152, 50
Valade A., Hoffman Y., Libeskind N. I., Graziani R., 2022, MNRAS, 513, 5148
Verde L., Treu T., Riess A. G., 2019, Nat. Astron., 3, 891
Willick J. A., Strauss M. A., Dekel A., Kolatt T., 1997, ApJ, 486, 629
Zaroubi S., 2002, MNRAS, 331, 901
Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

APPENDIX: PHYSICAL ORIGIN OF THE RECONSTRUCTION ERROR

In this appendix, we try to isolate different physical origins of the uncertainties in our velocity reconstruction. As shown in Section 4, without the use of the correct line-of-sight (LOS) distribution for the peculiar velocity tracers, the reconstructed velocity field is biased high. This bias also leads to increased uncertainty in the reconstructed velocity. However, another factor contributes to the increased error in the velocity reconstruction. In equation (12), there are two contributions to the distribution of peculiar velocity tracers along the LOS – (i) The uncertainty in the distance, (ii) The LOS density field. Due to the latter contribution, peculiar velocity tracers are clustered around regions of high density. Therefore, when we take the LOS density into account, the radial distribution may be more peaked than naively expected from the measured distance uncertainty. This is shown in Fig. A1, where we plot an example from

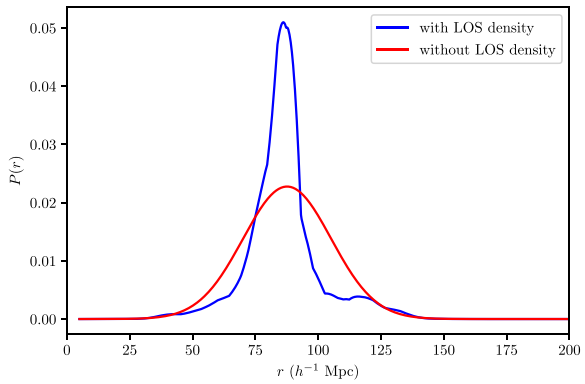

Figure A1. The impact of the LOS density field on the radial distribution of the peculiar velocity tracer. Here, we show the radial distribution of a peculiar velocity tracer in our mock survey. The red curve shows the expected radial distribution ignoring the LOS inhomogeneities and the blue curve shows the distribution after accounting for the LOS inhomogeneities. As we can see, radial distribution after accounting for the LOS inhomogeneities is more peaked than the naive distribution expected from the measured distance error.

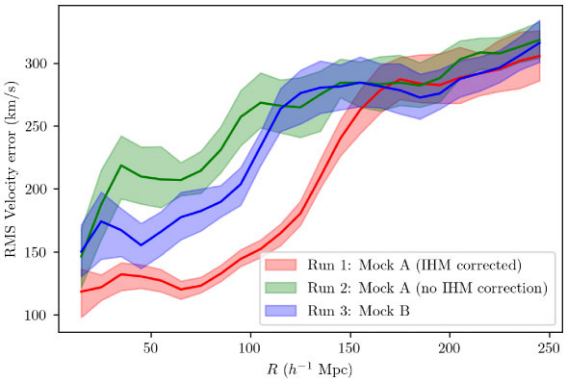

Figure A2. The RMS error in the velocity reconstructions in our three runs. The peculiar velocity tracers in Mock A are sampled by accounting for the LOS inhomogeneities and therefore the reconstruction is susceptible to IHM bias. On the other hand, in Mock B, the peculiar velocity tracers are sampled homogeneously and therefore does not suffer from IHM bias. We notice that the RMS error in Run 1 is much lower than Run 2 and Run 3. On the other hand, the RMS error in Run 3 is closer to the error in Run 2 (compared to Run 1). This suggests that the error in the IHM corrected reconstruction has the additional advantage of reducing the reconstruction error by better localization of the peculiar velocity tracers.

our mocks used in Section 4. In this figure, the radial distribution of the tracer including the LOS density is much sharply peaked compared to the distribution is we ignored the LOS.

Therefore, when we account for the LOS density, we not only correct for the IHM bias, but we also sharpen the distance estimate and therefore decrease the uncertainty in the reconstruction. In order to investigate the impact of this factor, we use our method on two different mock surveys. In mock survey A, we populate the tracers according to equation (12), accounting for the LOS density field. In mock B, we populate tracers homogeneously, i.e. we do not account for the LOS inhomogeneity. Therefore, by design, inference with mock B does not suffer from IHM bias. Since the velocity tracers are not clustered around density peaks, we are not able to determine the position of the tracer better than the distance uncertainty. We then make three runs of our code on the two mocks:

(i) Run 1: Mock A, corrected for IHM bias.
(ii) Run 2: Mock A, not corrected for IHM bias.
(iii) Run 3: Mock B, not corrected for IHM bias (by design, Mock B is unaffected by IHM bias).

The results of the three runs are shown in Fig. A2, where we plot the root mean squared (RMS) error in the velocity reconstruction binned in radial bins. We see that the RMS error for the Run 1 is substantially lower than Run2 and Run3. Furthermore the RMS error on Run 2 is higher than Run 3. Note that only Run 2 is impacted by the IHM bias. We see that the RMS error in Run 3 is closer to Run 2 than to Run 1, thus suggesting that extra error of the IHM uncorrected run over the IHM corrected run is driven by the fact that the distance estimates are better localized when we account for the LOS density field, rather than the IHM bias. Thus, by accounting for the LOS density, we not only correct for the IHM bias, but also reduce the reconstruction error by better localization of the peculiar velocity tracers.

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