



HAL
open science

Extended fast action minimization method: application to SDSS-DR12 combined sample

E. Sarpa, A. Veropalumbo, C. Schimd, E. Branchini, S. Matarrese

► To cite this version:

E. Sarpa, A. Veropalumbo, C. Schimd, E. Branchini, S. Matarrese. Extended fast action minimization method: application to SDSS-DR12 combined sample. *Monthly Notices of the Royal Astronomical Society*, 2021, 503 (1), pp.540-556. 10.1093/mnras/stab378 . hal-02999512

HAL Id: hal-02999512

<https://hal.science/hal-02999512>

Submitted on 5 May 2023

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.

Extended fast action minimization method: application to SDSS-DR12 combined sample

E. Sarpa ^{1,2,3★} A. Veropalumbo ^{2,3} C. Schmid ¹ E. Branchini ^{2,3,4} and S. Matarrese ^{5,6,7,8}

¹Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France

²Dipartimento di Matematica e Fisica, Università degli studi Roma Tre, Via della Vasca Navale, 84, 00146 Roma, Italy

³INFN - Sezione di Roma Tre, via della Vasca Navale 84, I-00146 Roma, Italy

⁴INAF - Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monte Porzio Catone (RM), Italy

⁵Dipartimento di Fisica e Astronomia ‘Galileo Galilei’, Università degli studi di Padova, Via F. Marzolo, 8, I-35131 Padova, Italy

⁶INFN, Sezione di Padova, via F. Marzolo 8, I-35131, Padova, Italy

⁷INAF-Osservatorio Astronomico di Padova, Vicolo dell’Osservatorio 5, I-35122 Padova, Italy

⁸Gran Sasso Science Institute, Viale F. Crispi 7, I-67100 L’Aquila, Italy

Accepted 2021 February 5. Received 2021 February 5; in original form 2020 October 20

ABSTRACT

We present the first application of the extended Fast Action Minimization method (eFAM) to a real data set, the SDSS-DR12 Combined Sample, to reconstruct galaxies orbits back-in-time, their two-point correlation function (2PCF) in real-space, and enhance the baryon acoustic oscillation (BAO) peak. For this purpose, we introduce a new implementation of eFAM that accounts for selection effects, survey footprint, and galaxy bias. We use the reconstructed BAO peak to measure the angular diameter distance, $D_A(z)r_s^{\text{fid}}/r_s$, and the Hubble parameter, $H(z)r_s/r_s^{\text{fid}}$, normalized to the sound horizon scale for a fiducial cosmology r_s^{fid} , at the mean redshift of the sample $z = 0.38$, obtaining $D_A(z = 0.38)r_s^{\text{fid}}/r_s = 1090 \pm 29(\text{Mpc})^{-1}$, and $H(z = 0.38)r_s/r_s^{\text{fid}} = 83 \pm 3(\text{km s}^{-1} \text{Mpc}^{-1})$, in agreement with previous measurements on the same data set. The validation tests, performed using 400 publicly available SDSS-DR12 mock catalogues, reveal that eFAM performs well in reconstructing the 2PCF down to separations of $\sim 25h^{-1}\text{Mpc}$, i.e. well into the non-linear regime. Besides, eFAM successfully removes the anisotropies due to redshift-space distortion (RSD) at all redshifts including that of the survey, allowing us to decrease the number of free parameters in the model and fit the full-shape of the back-in-time reconstructed 2PCF well beyond the BAO peak. Recovering the real-space 2PCF, eFAM improves the precision on the estimates of the fitting parameters. When compared with the no-reconstruction case, eFAM reduces the uncertainty of the Alcock-Paczynski distortion parameters α_{\perp} and α_{\parallel} of about 40 per cent and that on the non-linear damping scale Σ_{\parallel} of about 70 per cent. These results show that eFAM can be successfully applied to existing redshift galaxy catalogues and should be considered as a reconstruction tool for next-generation surveys alternative to popular methods based on the Zel’dovich approximation.

Key words: methods: numerical – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

Baryon acoustic oscillations (BAO) are one of the main cosmological probes to investigate the nature of dark energy and search for deviations from General Relativity on cosmological scales. For this reason, they have been selected as primary science case of both current spectroscopic surveys, such as WiggleZ (Drinkwater, Blake & WiggleZ Team 2009), the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013), and its successor the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Blanton et al. 2017), and future surveys operated by the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016), the ESA Euclid mission (Laureijs et al. 2011), and the Roman Observatory (Green et al. 2012). Thanks to their wide sky coverage, all the new-generation surveys allow the measurement of the acoustic scale with unprecedented accuracy. However, to reach the ambitious goal of sub per cent precision and

extract the maximum information from the non-linear scales, one needs to pair the quality of observations with an accurate model of the small-scale clustering. To this end, the use of highly-optimized BAO reconstruction techniques able to minimize the non-linear effects that blur the acoustic feature and possibly trace the orbits of galaxies backwards-in-time is nowadays an essential step of any clustering analyses.

For the official SDSS-III/BOSS analysis, the Zel’dovich reconstruction technique (ZA; Eisenstein et al. 2007b; Padmanabhan et al. 2012) has been successfully adopted, yielding a remarkable improvement on the BAO measurements in both momentum (Beutler et al. 2017) and configuration space (Alam et al. 2017). However, the intrinsic linearity of the method prevents the accurate modelling of redshift-space distortions (RSD), usually overestimating their amplitude (Kazin et al. 2013), and limits the range of validity of the recovered linear two-point correlation to relatively large separations (Padmanabhan, White & Cohn 2009). Pursuing a more accurate description of the real-space linear correlation function, new non-linear BAO reconstruction techniques have been proposed.

* E-mail: elena.sarpa@lam.fr

In particular, iterative reconstruction techniques (e.g. Schmittfull, Baldauf & Zaldarriaga 2017; Wang et al. 2017; Hada & Eisenstein 2018) have attracted special attention and have been successfully tested on N -body dark matter simulations. Mao et al. (2020) proposed an innovative approach which intends to recover the BAO signal using deep convolutional neural networks. The extended Fast Action Minimization method (eFAM; Sarpa et al. 2019), used in this work, is a variational reconstruction method designed to recover the geodesics of the mass tracers via minimization of the action of the system.

In this study, we present the first application of eFAM to spectroscopic data emphasizing its distinctive capability to simultaneously recover the real-space correlation function at both the observed and high redshift well into the mildly non-linear regime. To fully exploit the non-linearity of eFAM, we apply the reconstruction on the lower redshift bin $0.2 < z < 0.55$ extracted from the BOSS-DR12 Combined Sample, which is mostly affected by non-linear clustering. Finally, we measure the expansion rate $H(z)$ and the angular diameter distance $D_A(z)$ before and after reconstruction using the clustering wedges (Kazin, Sánchez & Blanton 2012).

The paper is organized as follows. In Section 2, we illustrate the new features of the eFAM algorithm accounting for masked regions, fibres collision, flux selection, and galaxy bias, as required for the application to wide-field, massive spectroscopic galaxy surveys. The BOSS-DR12 data and the mocks employed to test the method and build the covariance matrix are illustrated in Section 3. In Section 4, we describe the statistical tools used in the analysis, and motivate the choice of the fiducial fitting procedure in Section 5. We study the systematic behaviour of eFAM by reconstructing 400 realistic mocks in Section 6, and analyse SDSS data in Section 7. In Section 8, we discuss the main points of our study and present our conclusions.

2 EFAM FOR REAL GALAXY SURVEYS

The eFAM method, described in details in Sarpa et al. (2019), reconstructs the full-orbit of N point-like mass particles at their observed positions $\{\mathbf{x}_i(t)\}_{i=1,\dots,N}$, by minimizing the action of the self-gravitating system

$$S = \sum_{i=1}^N \int_0^{D_{\text{obs}}} dD \left[f E D a^2 \frac{1}{2} \left(\frac{d\mathbf{x}_i}{dD} \right)^2 + \frac{3\Omega_{m,0}}{8\pi f E D a} \phi_{\text{tot}}(\mathbf{x}_i) \right] + \sum_{i=1}^N \frac{1}{2} (f E D a)_{\text{obs}}^2 \left(\frac{d\mathbf{x}_{i,\text{obs}}}{dD} \cdot \mathbf{x}_{i,\text{obs}} \right)^2 \quad (1)$$

using mixed boundary conditions (Peebles 1989). Under the hypothesis that galaxies trace the underlying mass distribution, boundary conditions are set by the galaxies observed redshifts, $s_{i,\text{obs}}$, and their assumed vanishing peculiar velocities at early times. Here, a indicates the scale factor, D is the linear growth factor that we used as time variable, $f = d \ln D / d \ln a$ is the linear growth rate, and $E = H/H_0$ is the dimensionless Hubble parameter. Ω_m represents the mass density parameter and the subscripts ‘0’ and ‘obs’ indicate, respectively, quantities measured at the present epoch and at the redshift of the survey $z = z_{\text{obs}}$. Following Nusser & Branchini (2000), we model galaxies’ orbits as a linear combination of M time-dependent basis functions $\{q_n(D)\}_n$ viz.

$$\mathbf{x}_i(D) = \mathbf{x}_{i,\text{obs}} + \sum_{n=0}^M C_{i,n} q_n(D) \quad (2)$$

and minimize the action in equation (1) with respect to the expansion coefficients $C_{i,n}$.

In this work, we modify the original implementation of the eFAM algorithm (Sarpa et al. 2019), which has been already tested on simulated haloes catalogues, to optimize its application to real surveys. Specifically, we focus on the inclusion of unique selection effects, observational biases, and sample geometry in the computation of the total gravitational potential of the system, ϕ_{tot} .

2.1 Survey geometry

Real surveys probe the mass distribution within a large yet finite region of the Universe. The total gravitational potential at any point \mathbf{x} in this region, $\phi_{\text{tot}}(\mathbf{x})$, is given by the sum of two terms: the potential due to the matter distribution inside the volume of the survey, $\phi_{\text{int}}(\mathbf{x})$, and the potential due to the external mass distribution, $\phi_{\text{ext}}(\mathbf{x})$, which is unknown. Neglecting the external matter distribution in the estimation of ϕ_{tot} in equation (1) generates spurious tidal fields that affect the quality of the reconstruction. To mitigate their effect, a typical approach is to estimate $\phi_{\text{ext}}(\mathbf{x})$ assuming a homogeneous and isotropic distribution of matter outside the survey volume with the same mean density of the galaxy sample. This procedure removes the spurious tidal fields due to the geometry of the survey. However, since we do not model the clustering properties of the external mass distribution, residual tidal effects are expected to degrade the quality of the reconstruction as we approach the edge of the survey volume. For this reason, we define a buffer region at a distance d_{Buffer} from the edges that we use to perform the reconstruction but discard in the analysis of the reconstructed field. The optimal value for d_{Buffer} depends on the survey geometry and on the desired accuracy of the reconstruction and thus needs to be evaluated in each specific eFAM application.

A simple, brute force implementation of this strategy would be to embed the survey region in a much larger volume filled with an un-clustered distribution of synthetic objects with same angular and radial selection function and mean density as the real galaxies of the survey, and run eFAM over all objects, real and synthetic alike. However, the computational time required by this procedure is prohibitive since the number of synthetic objects needs to be very large to minimize shot-noise errors (with the Poisson solver implemented by eFAM, the CPU time required to compute the gravitational potential increases linearly with the number of objects).

For this reason, we adopt instead an equivalent, more effective approach based on the Newton’s shell theorem: (i) we assume a homogeneous isotropic distribution of mass throughout the whole Universe, inside and outside the survey volume, and set $\phi_{\text{tot}}(\mathbf{x}) = 0$ everywhere; (ii) we use a Poisson solver to numerically estimate the potential ϕ_{int} as generated by a uniform distribution of matter within the survey volume; (iii) we set $\phi_{\text{ext}} = -\phi_{\text{int}}$. As mentioned above, the CPU time required for the computation of the potential due to synthetic objects is larger than the one needed to compute it from galaxies. Though, here the calculation is performed only once per eFAM reconstruction and not at each time-step in equation (1). The increase in computational time is, therefore, negligible.

To assess the accuracy of this procedure, we apply eFAM to a set of dark matter haloes extracted from the DEUS simulation (Rasera et al. 2014) with same footprint and similar depth as the BOSS-DR12 survey. Fig. 1 illustrates the comparison between the ‘true’ velocity field in the N -body simulation (top-left panel) and the reconstructed one before (top-right panel) and after (bottom-left panel) including the effect of the model tidal field ϕ_{ext} . Our procedure effectively removes the spurious bulk-flow that would otherwise dominate the velocity field. The velocity residuals map (bottom-right panel) confirms that spurious flows are confined to

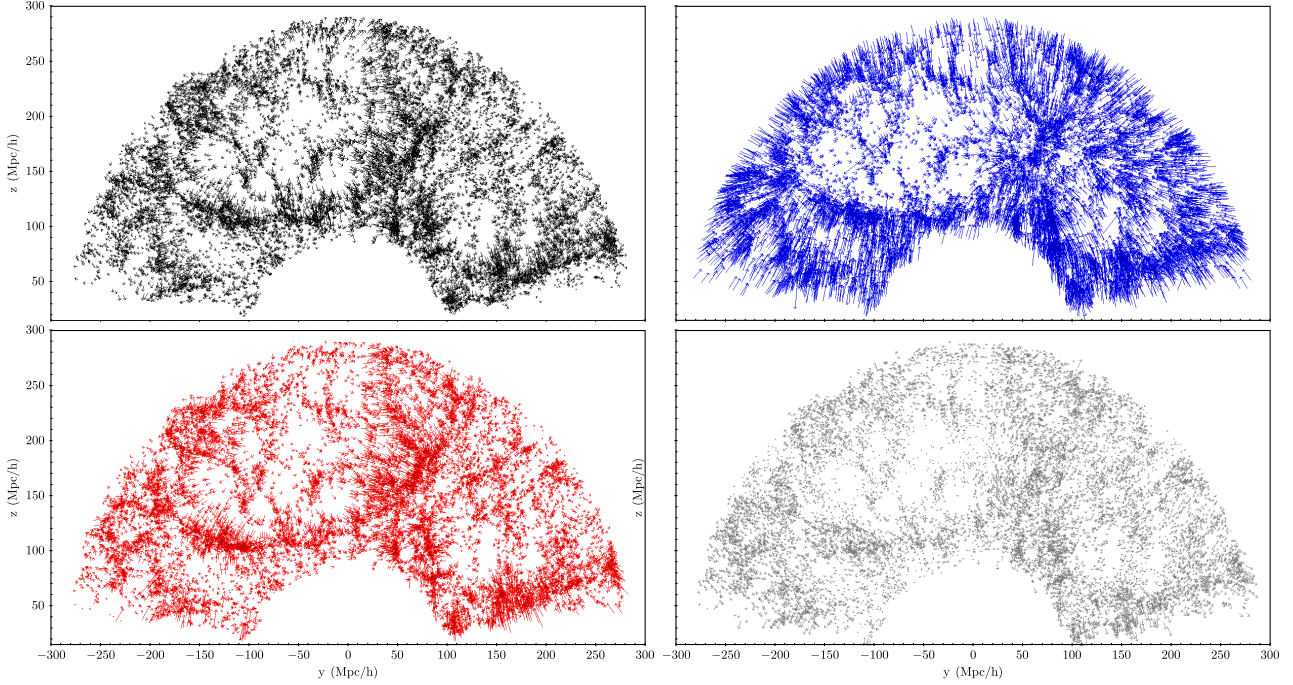


Figure 1. 2D velocity maps for the BOSS-like sub-sample carved from the DEUS dark matter haloes simulation. Top-left panel: N-body velocity field. Top-right panel: velocity field reconstructed without accounting for the external tidal field and assuming an empty universe outside the survey; the bulk flow induced by the geometry of the survey is dominant. Bottom left-hand panel: velocity field reconstructed including the external tidal field; the geometrical bulk flow is successfully removed and the N_{body} velocity field efficiently recovered. Bottom right-hand panel: residuals velocity field $V_{N_{\text{body}}} - V_{\text{eFAM}}$ after correcting for the external tidal field.

the regions near the edges justifying our choice of discarding those regions in the analysis. A quantitative assessment of the quality of the reconstruction and a test designed to evaluate systematic errors is illustrated in Appendix A.

2.2 Galaxy bias

The original implementation of FAM (Nusser & Branchini 2000) assumes that the average mass density of the survey volume matches the cosmological mean and that all the mass is associated with individual, point-like visible objects. However, real galaxies are known to be biased tracers of the mass density field. To account for galaxy bias, we assume that bias is a local, linear and deterministic phenomenon and that the total mass distribution can be split in two components: a distribution of point-like discrete masses accounting for both the luminous and dark matter associated to each observed galaxy, $\rho'_g(\mathbf{x})$, surrounded by a uniform, smooth distribution of dark matter, $\bar{\rho}_{\text{DM}}$, filling the survey volume, i.e.

$$\rho_{\text{tot}}(\mathbf{x}) \equiv \rho'_g(\mathbf{x}) + \bar{\rho}_{\text{DM}} \sum_{\mathbf{g}} [1 - \delta_{\text{D}}(\mathbf{x} - \mathbf{x}_{\mathbf{g}})], \quad (3)$$

where the sum runs over galaxy positions $\mathbf{x}_{\mathbf{g}}$. The relation between the observed galaxy density field, ρ_g , and ρ'_g is set by the linear biasing assumption $\delta_{\text{tot}} = \delta_g/b$ to

$$\rho'_g(\mathbf{x}) = \left(\frac{\delta_g(\mathbf{x})}{b} + 1 \right) \Omega_{\text{m},0} \frac{3H^2}{8\pi G} \sum_{\mathbf{g}} \delta_{\text{D}}(\mathbf{x} - \mathbf{x}_{\mathbf{g}}), \quad (4)$$

where $\Omega_{\text{m},0} 3H^2 / (8\pi G) = \bar{\rho}_{\text{tot}}$, $\delta_{\text{tot}} = 1 + \rho_{\text{tot}}/\bar{\rho}_{\text{tot}}$, and $\delta_g = 1 + \rho_g/\bar{\rho}_g$.

Equation (4) implies that the net effect of the bias is to down-weight the mass of individual objects, hence lowering their contribution to

the mass density below the cosmological mean. Since eFAM solely accounts for the mass component associated with discrete objects, the down-weighting of the masses will cause the survey volume to be treated as an under-dense region and a net outflow will be predicted. To restore the density balance, we thus need to account for the gravitational pull of the smooth dark component. Our strategy is to model the gravitational potential generated by the smooth dark matter distribution, ϕ_{smooth} , as an extra unknown external field to be added to the one exerted by the homogeneous external distribution described in Section 2.1. Similarly to the procedure adopted for the external tidal field, we infer the new external gravitational potential $\phi'_{\text{ext}} \equiv \phi_{\text{ext}} + \phi_{\text{smooth}}$ setting $\phi'_{\text{ext}} = -\phi'_{\text{int}}$, where ϕ'_{int} is now estimated having fixed the mean density of the smooth matter distribution within the survey to $\bar{\rho}'_{\text{R}} = \bar{\rho}_{\text{tot}} - \bar{\rho}_{\text{DM}}$.

In our fiducial biasing scheme, we treat all mass tracers as point-like masses, ignoring the physical extension of both the observed galaxies and their hosting haloes. Although this approximation does not constitute a problem as long as the inter-object separation is significantly larger than the halo radius (Sarpa et al. 2019), it becomes relevant when attempting to trace the orbits of galaxies in high-density environments, like groups, or in a $\Omega_m = 1$ cosmology, where haloes are very extended objects (see e.g. Fig. 5 in Branchini & Carlberg (1994)). In this work, we reconstruct the dynamic of a relatively low-density galaxy sample, $\bar{n}_g = 0.003$, evolving in a Λ CDM universe, i.e. with a mean galaxy separation larger than their typical halo size. We can, therefore, safely ignore the physical extension of the galaxies.

The second peculiarity of our bias model is the conservation of the number of tracers through time. As a result, eFAM fails to recover the merging history of both observed galaxies and their hosting haloes effectively tracing the trajectories of the centre-of-mass of

the clustered matter distribution associated with visible objects. This approximation should not spoil the quality of reconstruction as long as the mismatch between the position of the centre-of-mass and that of the parent haloes is small compared to the scale of the structures we wish to trace. Since most of the displacement along the trajectory occurs at recent epochs, where major mergers are less common, we expect this assumption to produce negligible effects on our clustering measurements at high redshifts. The goodness of this assumption has been verified by the success of eFAM at recovering the linear two-point correlation functions at relatively high redshifts when applied to simulated halo catalogues (Sarpa et al. 2019)).

3 DATA AND SIMULATIONS

3.1 SDSS-DR12 Combined Sample

We apply eFAM reconstruction algorithm to the galaxies in the low-redshift bin $0.2 < z < 0.55$ of the ‘BOSS Combined Sample’, corresponding to the first redshift bin considered in the clustering analyses of the BOSS collaboration (Kitaura et al. 2016; Ross et al. 2017; Vargas-Magaña et al. 2018). The ‘Combined Sample’ spans the redshift range $0.2 < z < 0.75$ and includes both the Constant Stellar Mass sample (CMASS), designed to be approximately stellar-mass limited above $z = 0.45$, and the low-redshift (LOWZ) sample, targeting galaxies in the redshift range $0.15 < z < 0.43$. Having considered the survey footprint, we decide to focus on the northern galactic cap to reduce both the computational cost of the reconstruction and the impact of the external tidal field, which would be large in the thinner southern strip. Following (Ross et al. 2017, hereafter Ross17), we assign a statistical weight to each galaxy to account for fibre collision, veto flag, and the need to minimize the variance of the clustering estimator. The latter aspect is dealt with using the so-called FKP weight (Feldman, Kaiser & Peacock 1994) $\omega_{\text{FKP}} = 1/[1 + \bar{n}(z_i)P(k_0)]$ where \bar{n} is the redshift distribution at the galaxy redshift (z_i), and $P(k_0) = 40\,000(h^{-1}\text{Mpc})^3$ is the amplitude of the power spectrum evaluated approximately at the BAO scale for SDSS Luminous Red Galaxies (LRG). In the eFAM implementation, we use these weights to assign an effective mass to each individual object.

3.2 MultiDark-Patchy mock catalogues

To test the performances of eFAM and estimate the covariant errors employed in the likelihood analysis, we make use of the publicly available MultiDark-Patchy (hereafter MD-Patchy) mocks (Kitaura et al. 2016; Rodríguez-Torres et al. 2016). MD-Patchy are extracted from the Big-MultiDark Planck simulation (Klypin et al. 2016) which assumes a spatially flat Λ CDM *Planck* cosmology with $\Omega_m = 0.307115$, $\Omega_\Lambda = 0.692885$, $\Omega_b = 0.048$, $\sigma_8 = 0.8288$, $n_s = 0.9611$, $h = 0.6777$. Furthermore, a non-linear, deterministic, stochastic bias recipe has been applied to the galaxy catalogue to match the clustering properties of the BOSS LRG sample.

In the following analysis, we use the V6-C version of the ‘Combined’ sample mocks, built to estimate the covariance matrix of the 1, 2, and 3-points clustering statistic of SDSS-DR12/BOSS. Since we are interested in computing the 2PCF of the mock galaxies, we also use the associated mock random catalogue which contains 50 times as many objects as the SDSS-DR12 mocks and has the same footprint, selection function, and statistical weights. To reduce the computational cost of our analysis, we decide to consider only 400 mocks out to the 1000 employed by Ross17. The impact of this choice is discussed in Section 8.

4 STATISTICAL TOOLS

4.1 Two-point correlation function estimator

To gauge the performance of the eFAM method, we perform a similar clustering analysis to that already carried out by the BOSS collaboration aimed at estimating the Hubble parameter, H , and the diameter angular distance, D_A , from the analysis of the BAO peak in the anisotropic 2PCF, $\xi(s, \mu)$. Here, s represents the separation of the galaxy pair and μ is the cosine of the angle between \mathbf{s} and the bisector of the angle to the pair. To estimate the $\xi(s, \mu)$ of each sample, we use the minimum variance, unbiased Landy & Szalay (1993) estimator

$$\xi(s, \mu) = \frac{DD(s, \mu) - 2DR(s, \mu) + RR(s, \mu)}{RR(s, \mu)}, \quad (5)$$

where the data-data, DD , data-random, DR , and random-random, RR , counts are evaluated in bins of $\Delta s = 5h^{-1}\text{Mpc}$ and $\Delta\mu = 0.05$, within the ranges $30 < s < 190$ and $0 < \mu < 1$, respectively. We measure the 2PCF of three types of data sets: (i) the observed, redshift-space (hereafter labelled ‘Obs’), in which SDSS galaxies are placed at their comoving coordinates using measured redshift as distance proxy and assuming the cosmology of the mocks as fiducial model, (ii) the reconstructed real-space galaxies at the epoch of observation (labelled ‘RecZ’), obtained by placing galaxies at their reconstructed real-space positions, and (iii) the back-in-time reconstructed real-space positions (‘RecL’) in which galaxies are displaced back-in-time to the real-space positions they occupied at an earlier epoch. Unlike the standard ZA reconstruction, eFAM does not require any spatial smoothing of the observed density field to obtain the reconstructed one. We can thus use the estimator in equation (5) rather than the modified version of Padmanabhan et al. (2012), which additionally requires to reconstruct the back-in-time orbits of the random objects, with a consequent important computational load.

To minimize the impact of the external tidal field, we estimate the 2PCF of the sample after having discarded all the objects, real and random alike, that reside in the buffer region (see Section 2.1). Besides, since reconstruction displaces the objects from their original positions, we trim the reconstructed catalogs to match the survey footprint and depth. Finally, the 2PCF is estimated by assigning to each object its statistical weight accounting for FKP, fibre collision, and veto flag.

4.2 Clustering wedges

We analyse the anisotropic 2PCF focusing on the wedge clustering statistics introduced by Kazin et al. (2012) and applied by Kazin et al. (2013) to the study of the SDSS-DR11 galaxy clustering. We shall consider the two wedges that model clustering along the parallel (||) and transverse (\perp) direction to the line-of-sight:

$$\xi_{||(\perp)}(s) = \frac{1}{\Delta\mu} \int_{0.5(0)}^{1(0.5)} d\mu \xi(s, \mu). \quad (6)$$

Following Kazin et al. (2013), we focus on the BAO peak whose scale depends on a combination of H , D_A and the sound horizon scale during baryon drag epoch, r_s . Since the radial clustering wedge $\xi_{||}$ is mostly sensitive to H while the transverse one, ξ_{\perp} , is sensitive to D_A , one can measure $\xi_{||}$ and ξ_{\perp} to break this degeneracy. This is done by exploiting the Alcock-Paczynski effect (Alcock & Paczynski 1979) (hereafter AP), which quantifies the dependence of the measured separations along and across the line of sight, $s_{||}^{\text{fid}}$ and s_{\perp}^{fid} , on a

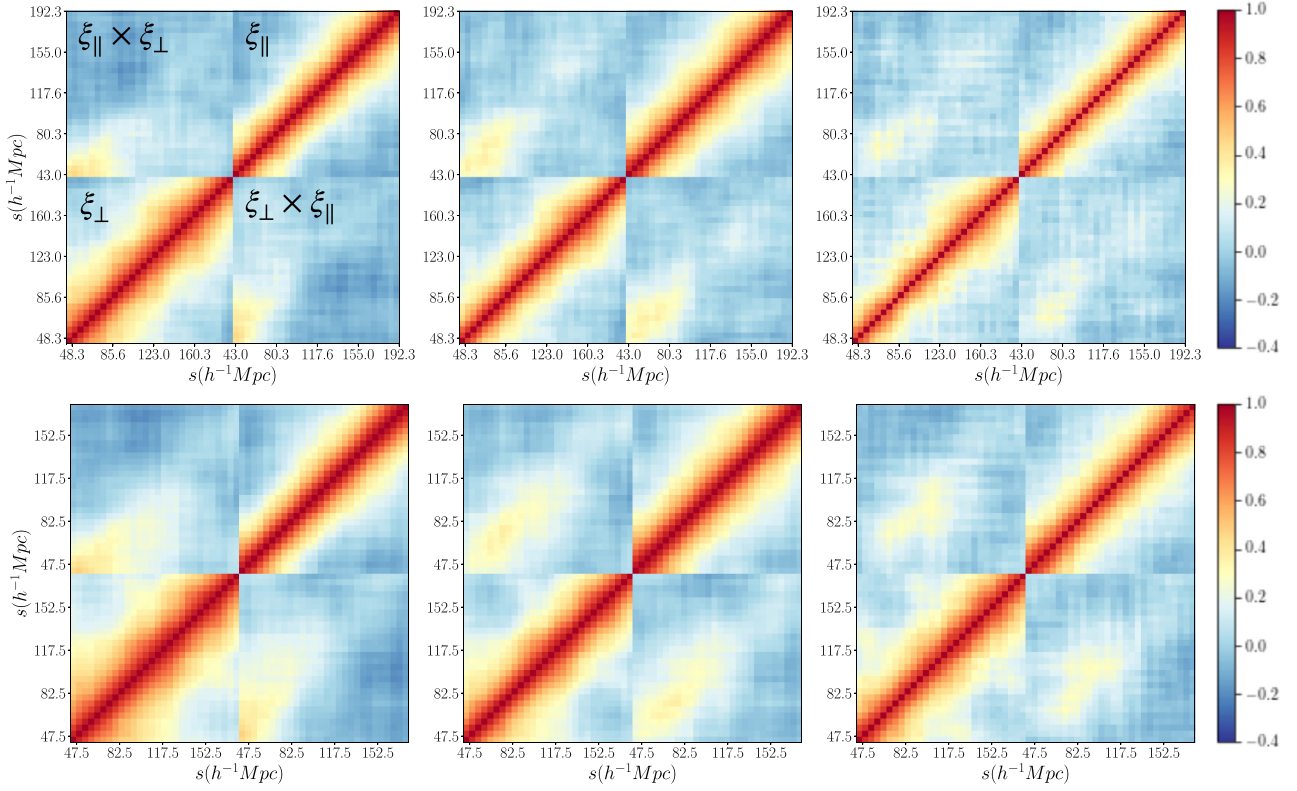


Figure 2. Correlation matrices of the clustering wedges estimated from 400 MD-Patchy mocks. Left panels: Obs sample. Middle panels: RecZ samples. Right panels RecL samples. Top panels: catalogs extracted by excluding all objects laying within $d_{\text{Buffer}} = 200h^{-1}\text{Mpc}$ from the survey edges. Bottom panels: $d_{\text{Buffer}} = 125h^{-1}\text{Mpc}$. The colour code is set accordingly to the normalized covariance amplitude.

fiducial cosmological model, i.e.

$$s_{\perp}^t = \alpha_{\perp} s_{\perp}^{\text{fid}} \quad \text{with} \quad \alpha_{\perp} \equiv \frac{D_A^t r_s^{\text{fid}}}{D_A^{\text{fid}} r_s^t}, \quad (7)$$

$$s_{\parallel}^t = \alpha_{\parallel} s_{\parallel}^{\text{fid}} \quad \text{with} \quad \alpha_{\parallel} \equiv \frac{H^{\text{fid}} r_s^{\text{fid}}}{H^t r_s^t}, \quad (8)$$

to infer the ‘true’ values H^t and D^t from the modelling of the anisotropy in the measured 2PCF. The discrepancy between the fiducial and true separation, and thus between the fiducial and true cosmology, is quantified by the parallel and perpendicular dilation parameters, α_{\parallel} and α_{\perp} .

In redshift-space, anisotropies in the 2PCF are generated by both RSD and AP-distortions. Hence, one needs to disentangle the two effects to accurately estimate H^t and D^t . In this work, we use eFAM to remove RSD. However, since the reconstruction algorithm demands to set a value for the linear growth rate, f , we shall not attempt to estimate its value in the likelihood analysis of reconstructed samples. Instead, we will check the adequacy of the fiducial f by verifying that the RSD have been consistently removed in both reconstructed data sets, RecZ and RecL.

In addition to clustering wedges, we study the anisotropy in the 2PCF considering its multipoles (Padmanabhan & White 2008)

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^1 d\mu \xi(s, \mu) \mathcal{L}_l(\mu), \quad (9)$$

where $\mathcal{L}_l(\mu)$ are the Legendre polynomials, which provide an equivalent analysis of the clustering properties to that of the wedges but ease the estimation of anisotropies as a non-vanishing quadrupole.

4.3 Covariance matrix

Given the strong correlation between ξ_{\perp} and ξ_{\parallel} , the covariance matrix of the clustering wedges C_s is built by combining the auto-covariance and the cross-covariance of the binned parallel and perpendicular wedges measured from the 400 mocks. As shown by Hartlap, Simon & Schneider (2007), Percival et al. (2014), the precision matrix, Ψ , estimated by inverting the measured covariance matrix, C_s , is a biased estimate of the true one, with the amplitude of this bias depending on both the number of mocks n_s used to measure C_s and its dimension $n_b \times n_b$. To correct for this bias we set

$$\Psi = \left(1 - \frac{n_b + 1}{n_s - 1}\right) C_s^{-1}. \quad (10)$$

The core of our analysis focuses on the properties of the 2PCF in the separation range $[50, 150]h^{-1}\text{Mpc}$. For this application, we sample the clustering wedges in 20 bins of $5h^{-1}\text{Mpc}$ covering the whole range of interest.

Fig. 2 illustrates the auto and cross-correlation matrices of the clustering wedges measured for the Obs (left panels), RecZ (middle) and RecL (right) mock samples. The difference between the top and bottom panels is due to the different extent of the buffer region that contains objects excluded from the analysis. A simple visual inspection reveals that the application of eFAM significantly reduces the amplitude of the off-diagonal terms in both RecZ and RecL samples.

Table 1. Fitting parameters in the likelihood analysis and their prior range. Top: parameters used to fit the correlation function averaged over the mocks (ξ) (Section 6.1). Bottom: parameters used to fit the 2PCF of the data (Section 7 and that of a single mock (Section 6.2).

sample	α_{\perp}	α_{\parallel}	$\Sigma_{\perp} (h^{-1}\text{Mpc})$	$\Sigma_{\parallel} (h^{-1}\text{Mpc})$	b	f	$\Sigma_s (h^{-1}\text{Mpc})$	$\{A_{p,i}\}_{i=0,1,2}$
All	Flat: [0.7,1.3]	Flat [0.7,1.3]	Flat: [0,15]	(ξ) (Section 6.1) Flat: [0,15]	Flat: [0,10]	Flat: [0,5]	Flat: [0,10]	Flat: [-10, 10]
Obs	Flat: [0.7,1.3]	Flat: [0.7,1.3]	Fixed to (ξ)	Fixed to (ξ)	Fixed to (ξ)	Fixed to (ξ)	Fixed to (ξ)	Flat: [-10, 10]
RecL	Flat: [0.7,1.3]	Flat: [0.7,1.3]	Fixed to (ξ)	Fixed to (ξ)	Fixed to (ξ)	Fixed to 0	Fixed to 0	Flat: [-10, 10]

5 MODELLING THE 2PCF

5.1 Two-point correlation function model

The goal of this study is to model the anisotropic 2PCF over a large range of scales including the BAO peak. Hence, we need to take into account RSD, AP-distortions and non-linear clustering evolution. The first step is to model the non-linear mass power spectrum in real-space, $P^R(k, \mu)$, using the Eisenstein & Hu (1998) model

$$P^R(k, \mu) = [P_l(k) - P_{\text{nw}}(k)] e^{-k^2 [(1-\mu^2)\Sigma_{\perp}^2/2 + \mu^2\Sigma_{\parallel}^2/2]} + P_{\text{nw}}(k),$$

where $P_l(k)$ is the linear power spectrum, P_{nw} is the no-wiggle power spectrum, and Σ_{\parallel} and Σ_{\perp} quantify the non-linear damping along and across the line-of-sight (Crocco & Scoccimarro 2006; Eisenstein, Seo & White 2007a; Crocco & Scoccimarro 2008). P_l is estimated using CAMB (Lewis, Challinor & Lasenby 2000) having assumed the same cosmological model as the mock catalogs while P_{nw} is computed analytically.

To model RSD, we combine linear theory to model large-scale coherent motion (Kaiser 1987), with the streaming model (Peacock & Dodds 1994) to account for small-scale incoherent velocities obtaining

$$P^S(k, \mu) = \frac{(1 + \mu^2 f/b)^2}{1 + (k^2 \mu^2 \Sigma_s^2)^2} P^R(k, \mu), \quad (11)$$

where the superscript ‘S’ indicates quantities measured in redshift-space, μ is the cosine angle between the wave vector and the line-of-sight direction, b is the linear bias parameter, and Σ_s is the velocity dispersion parameter. The redshift-space template is only used to fit the 2PCF of the Obs sample, i.e. before performing eFAM reconstructions; it is not included to model the reconstructed 2PCF of the RecZ and RecL sample since we find that the reconstructed 2PCF of the RecZ and RecL samples is well fit by a model with $f = \Sigma_s = 0$ (see Section 6.1).

Finally, we model the measured clustering wedges, $\{\bar{\xi}_p^{R(S)}\}_{p=\perp,\parallel}$, as follows: i) we estimate the multipoles of the anisotropic power spectrum, $P^{R(S)}(k, \mu)$, defined as

$$P_l^{R(S)}(s) = \frac{2l+1}{2} \int_{-1}^1 d\mu P^{R(S)}(k, \mu) \mathcal{L}_l(\mu); \quad (12)$$

ii) we compute the corresponding 2PCF multipoles ξ_l performing a Fourier transform of each multipole P_l ; (iii) we combine the 2PCF multipoles to obtain the clustering wedges, $\xi_p^{R(S)}(s)$ as described in Kazin et al. (2012), and (iv) we obtain the fiducial template by adding the full-shape term $A_p(s)$, viz.

$$\bar{\xi}_p^{R(S)}(s; \alpha_{\perp}, \alpha_{\parallel}) = \xi_p^{R(S)}(s; \alpha_{\perp}, \alpha_{\parallel}) + A_p(s). \quad (13)$$

Here, the two dilation parameters are free parameters accounting for the mismatch between the fiducial and the true cosmology while the full-shape term $A_p(s) = A_{p,0} + A_{p,1/s} + A_{p,2/s^2}$ is introduced to

account for possible systematic errors in reconstructing the broadband shape of the 2PCF. Differently from the fiducial BOSS template (Ross17), we omit the shape parameter B that would be highly degenerate with those controlling the RSD.

5.2 Likelihood analysis

5.2.1 Maximization of the likelihood

To fit the model to the measured 2PCF, we search for the maximum of the likelihood in the space of the free parameters using the modelling routine in the CosmoBolognaLib library (Marulli, Veropalumbo & Moresco 2016). Maxima are searched for using a two-step procedure. First, the Nelder-Med method (Nelder & Mead 1965) is used to approach the best-fitting value. Here, the N_p -dimensional parameters space is probed by evaluating the likelihood at the vertex of a running $(N_p + 1)$ -dimensional simplex which progressively approaches the nearest maximum. Second, once convergence is attained, we refine the search by running a Markov Chain around the best-fitting value. In the following, we perform different likelihood analyses varying the number of free parameters. Table 1 summarizes our choices for the analysis of the mean correlation function averaged over 400 mocks, (ξ), and for the study of individual mocks.

Similarly to the majorities of the multidimensional optimization algorithms, the Nelder-Med method approximates a local rather than the global maximum. To ensure convergence to the global maximum, it is crucial to set the initial size of the simplex so as to cover a sufficiently large portion of the N_p -dimensional domain. We do so by starting the maximization of the likelihood from an equilateral simplex and re-scaling the fitting parameters so that to have similar magnitudes. The parameters $\alpha_{\perp,\parallel}$, $\Sigma_{\perp,\parallel}$, Σ_s , f , and b represents physical quantities thus their amplitude can be guessed from perturbation theory (Crocco & Scoccimarro 2008). Ross17 estimate them to be in the range [1,10]. On the contrary, the values of the shape parameters $\{A_{p,i}\}_i$ is *a priori* unknown and their values can significantly vary. To avoid handling quantities with very different magnitudes we normalize the values of $A_{p,i}$ by the factor $3\sigma_{\xi_p(r_{\text{ref}})} r_{\text{ref}}^i$, where $\sigma_{\xi_p(r_{\text{ref}})}$ is the standard deviation of $\xi_p(r_{\text{ref}})$ among the mocks evaluated at an arbitrary scale r_{ref} . With this parametrization, and assuming $A_{p,i} = 0 \forall i$ for the mean wedges averaged over all the mocks ($\langle \xi_p \rangle$), we are able to model the 2PCF at r_{ref} for the almost totality of the mocks using $|A_{p,i}| \lesssim 1$. Aiming at modelling the distortion of the 2PCF that can not be described solely by physical parameters (e.g. a negative value of the correlation function at small scales), we set $r_{\text{ref}} \sim 80h^{-1}\text{Mpc}$, corresponding to the deep of the correlation function at small separations.

5.2.2 Prior and posterior

To minimize the risk of biased results, we assign flat, non-informative, priors to the parameters of the model. This approach

differs from the one of Ross17, Anderson et al. (2014), Xu et al. (2012), which assigned a Gaussian prior to the shape parameter B . The downside of this choice is the risk of hitting a local maxima of the likelihood, $\mathcal{P}_{\mathcal{L}}(\mathbf{p})$, hence of increasing the uncertainty of the parameters estimate. To minimize this effect, we infer the median and the variance of the i -th parameter p_i by performing a robust-sigma estimation (Section 3.2 in Longobardi et al. 2015) on the corresponding marginalized one-dimensional posterior

$$\mathcal{P}_{\mathcal{L}}(p_i) = \int_D dp_1 \dots dp_{i-1} dp_{i+1} \dots dp_{N_p} \mathcal{P}_{\mathcal{L}}(\mathbf{p}), \quad (14)$$

where D is the interval over which the prior of all j parameters ($j \neq i$) are defined. We do so by defining the 2σ -clipped posterior, $\mathcal{P}_{2\sigma}(p_i) = \mathcal{P}_{\mathcal{L}}(p_i) \Theta(p_i - \langle p_i \rangle_{\mathcal{L}} - \sigma_{\mathcal{L},i}) \Theta(\langle p_i \rangle_{\mathcal{L}} - \sigma_{\mathcal{L},i} - p_i)$, where $\langle p_i \rangle_{\mathcal{L}}$ and $\sigma_{\mathcal{L},i}$ are the mean and variance of $\mathcal{P}_{\mathcal{L}}(p_i)$ prior to sigma-clipping, and Θ is the Heaviside step function. The variance of $\mathcal{P}_{2\sigma}(p_i)$ is assigned to p_i , scaled by a numerical factor determined from Monte Carlo simulations to complete the 2σ -clipped distribution to a complete Gaussian.

6 EFAM RECONSTRUCTION IN THE MOCK CATALOGS. VALIDATION TESTS

In this section, we validate and investigate the performance of eFAM reconstruction focusing on two key aspects: its ability to remove RSD, and that of improving the signal-to-noise ratio of the BAO feature. For the following tests, we use the 400 MD-Patchy mocks mimicking the SDSS catalog described in Section 3.2.

6.1 From redshift to real-space

To assess the quality of the redshift-to-real space reconstruction, we apply eFAM to each of the 400 mock catalogues and measure the 2PCF monopole and quadrupole moments as well as the clustering wedges of the observed (Obs) and reconstructed (RecZ and RecL) samples defined in Section 4.1. To minimize the effect of tidal fields, we discard all objects within $d_{\text{Buffer}} = 200 h^{-1} \text{Mpc}$ from the edges of the survey. The results are illustrated in Fig. 3. Top panels illustrates the results obtained from the Obs sample. Middle and Bottom panel refer to the RecZ and RecL cases, respectively. Panels on the left show the 2PCF parallel (red) and perpendicular (blue) clustering wedges. Panels on the right illustrate the 2PCF monopole and the quadrupole moments. All quantities are multiplied by the square of the spatial separation s . Thin curves show the measurement for each of the 400 mocks, thick curves show their average values, and shaded areas indicate the standard deviation among the mocks, corresponding to the square root of the diagonal elements of the covariance matrix. Finally, yellow thick curves on the left panels represent the best-fitting model for the clustering wedges.

In redshift-space (Obs, top-left), the parallel and perpendicular wedges are significantly different from each other. This discrepancy is expected. Indeed, RSD simultaneously reduce the amplitude of the clustering signal along the line-of-sight and boost up the amplitude of the perpendicular wedge. Moreover, RSD generate a quadrupole moment that is clearly visible in the top-right panel. A visual inspection of the mid panels (RecZ) shows the success of eFAM at removing RSD; the two wedges are now in agreement, almost superimposed to each other (left-hand panel), and the quadrupole moment is consistent with zero (marked by the black, dotted horizontal line in the right-hand panels). Remarkably enough, the statistical isotropy is restored down to the smallest scales shown in

the plot, $\sim 40 h^{-1} \text{Mpc}$, demonstrating that eFAM is indeed able to model dynamics well into the non-linear regime.

A successful reconstruction must remove RSD at all epochs, not only at the redshift of the survey. We check for this behaviour estimating the isotropic 2PCF of the RecL sample at a conveniently high redshift. Here, we stop the reconstruction at $z \sim 40$ corresponding to the highest redshift for which the measured Σ_{\parallel} and Σ_{\perp} have the smallest, non-negative value (see Section 4 of Sarpa et al. (2019) for a detailed discussion on the definition and justification of the maximum redshift of the eFAM reconstruction). The 2PCF wedges and multipoles of the RecL sample are plotted in the bottom panels of Fig. 3. Similarly to the RecZ case, RSD are successfully removed, although some residual anisotropy is seen at very large separations. We interpret this feature as a signature of imperfect correction for external tidal fields. To check this hypothesis, we repeated the test varying the depth of the buffer region d_{Buffer} . The results are shown in Fig. 4. Thick lines represent the average monopole (blue-solid) and quadrupole (red-dashed) moments of the 2PCF measure in 20 independent mock catalogues. Shaded areas represent the standard deviation. The value of d_{Buffer} used in each reconstruction is indicated in each panel. For all the values of d_{Buffer} we testes, we also show the reference case $d_{\text{Buffer}} = 200 h^{-1} \text{Mpc}$ (light blue and light red curves and areas for the monopole and quadrupole moments, respectively). The quality of the reconstruction improves when using larger d_{Buffer} , i.e. when discarding an increasing fraction of objects in the sample. The main effect of the spurious tidal field is a deceptive quadrupole at large separation that is particularly significant for $d_{\text{Buffer}} = 50 h^{-1} \text{Mpc}$, and progressively vanishes when moving towards larger sizes of the buffer region. A second effect is a tiny artificial correlation at the deep's scale in the monopole. Based on this test, we decide to set $d_{\text{Buffer}} = 200 h^{-1} \text{Mpc}$ being confident that with this choice RSD are effectively removed down to $\sim 40 h^{-1} \text{Mpc}$ and the measured monopole signal is robust to less conservative choices of d_{Buffer} .

A more quantitative assessment of the quality of the RSD removal is provided by the likelihood analysis. Table 2 shows the values of the best-fitting parameters to the mean clustering wedges estimated in the 400 mocks and their 1σ uncertainty. The parameters that quantify RSD are the growth rate f and the velocity dispersion Σ_S . Their value can be estimated by measuring the wedges in redshift-space (Obs sample, first row in the Table). The efficient removal of RSD by eFAM is testified by the 1σ compatibility of both parameters with zero the RecZ and RecL samples alike. To perform the likelihood analysis we have considered the range of separation $s = [50, 150] h^{-1} \text{Mpc}$ and the best-fitting curves are drawn in yellow colour in Fig. 3.

As shown by Padmanabhan et al. (2012), another test of the quality of the reconstruction is to compare the relative magnitude of the two damping parameters, $\Sigma_{\perp, \parallel}$, before and after reconstruction. Indeed, to move galaxies to their correct real-space position one should remove all sources of anisotropy, not just the RSD. As result, a successful reconstruction should bring the ratio $\Sigma_{\parallel}/\Sigma_{\perp}$ closer to unity (Ivanov & Sibiryakov 2018). Since eFAM is designed to account for non-linear motions, we also expect that the absolute magnitude of both parameters should be reduced. The contour levels of the marginalized, joint probability distribution of Σ_{\parallel} and Σ_{\perp} shown in Fig. 5 prove that this is indeed the case. The blue-dotted contours show the results obtained before reconstruction for the Obs sample while the large blue-dot show the median values. The magnitude of the two parameters is rather large and their ratio is about 2σ apart from unity (dot-dashed purple line). Dark-green continuous curves (and the green-triangle) show the results after the eFAM reconstruction for the RecZ sample, i.e. after removing

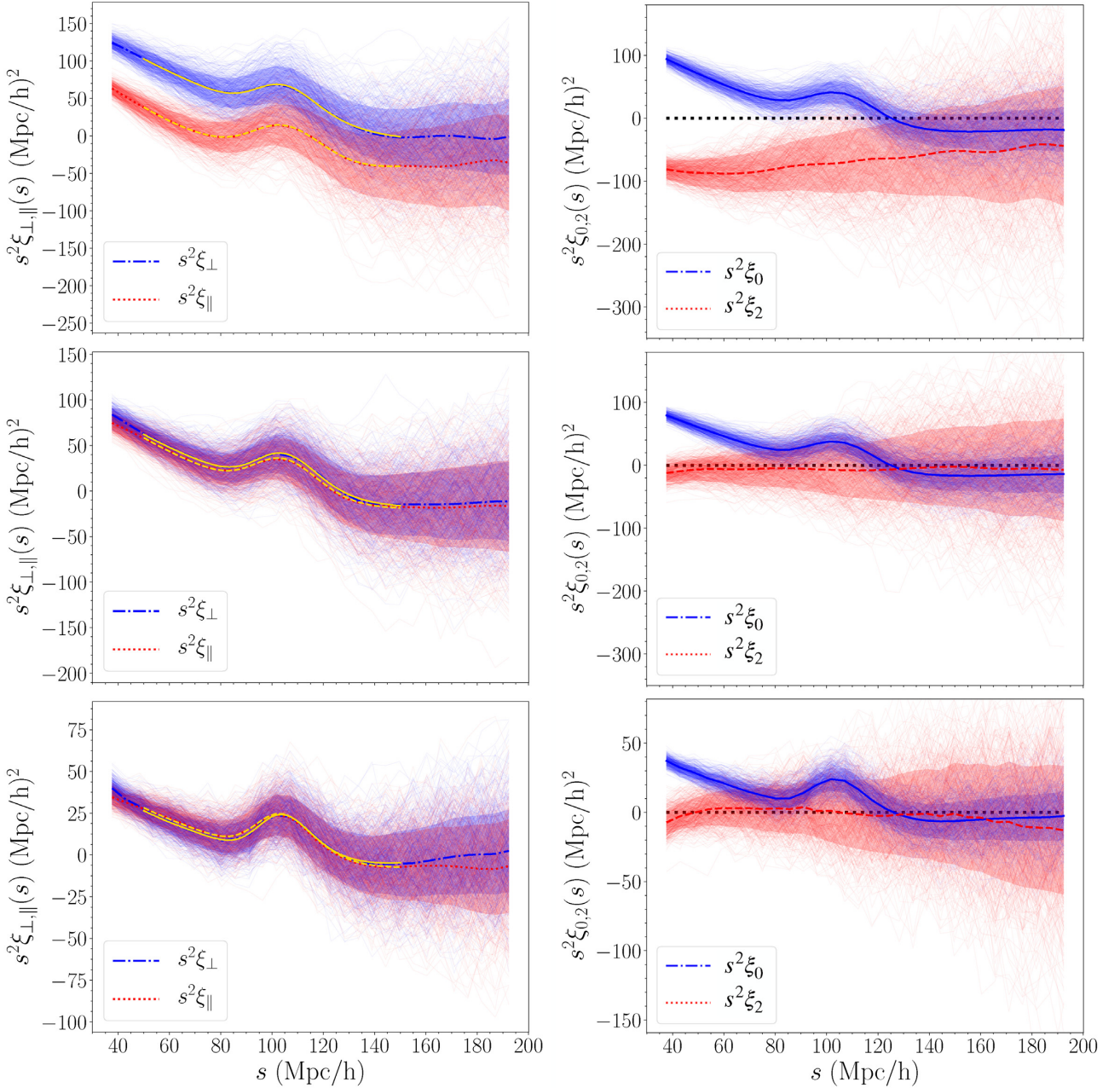


Figure 3. Two-point correlation function wedges (left-hand panels) and multiples (right) estimated from 400 mock catalogues. Top: mock galaxies are placed at their observed redshift positions, Middle: galaxies are placed at the reconstructed real-space positions at the same redshift of the survey. Bottom: galaxies are located at their back-in-time reconstructed real-space positions. Thin curves show the wedges measured in each mock. Thick curves with different line-styles show their average. Shaded regions show the 1σ uncertainty strip. Yellow curves on the right-hand panels are the best-fitting models.

RSD but with non-linear effects still present. The magnitude of the two parameters remains the same but their ratio is closer to unity. Gold-dashed contours show the RecL case, after eFAM back-in-time reconstruction. In this case, also the magnitude of both parameters is significantly reduced. Besides, we notice that the reconstruction significantly diminishes the uncertainty in the measured Σ_{ii} . This improvement indicates that eFAM efficiently lifts the degeneracy between Σ_{ii} and the $A_p(s)$ parameters describing the broad-band shape of the 2PCF while restoring statistical isotropy.

The validation tests performed in this section are designed to quantify the ability of eFAM to remove the anisotropies generated

by linear and non-linear motions rather than the ones arising from an incorrect choice of the fiducial cosmology. For this reason, we assume the same cosmological model as the mock catalogues. To test for the absence of Alcock-Paczynski distortions, we do not set the dilation parameters equal to unity but treat instead α_{ii} and α_{\perp} as free parameters in the fit and check if their best-fitting value is indeed consistent with unity. The results listed in Table 3 show that this is indeed the case. The largest departures are at $\sim 2\sigma$ level for α_{ii} . These results prove that eFAM successfully removes RSD from the 2PCF and allow us to use the real-space template ($f = \Sigma_s = 0$) to model the reconstructed 2PCF of the data in Section 7.

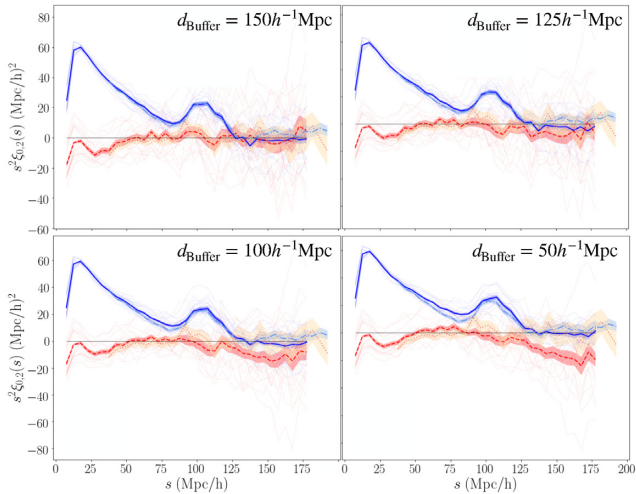


Figure 4. Effect of the external (super-survey) tidal field on reconstruction. Mean monopole (blue-solid lines) and quadrupole (red-dashed lines) of the 2PCF averaged over 20 mocks measured for different choices of the discarded buffer region, d_{Buffer} . Light-blue dot-dashed lines and orange-dotted lines mark the mean monopole and quadrupole as estimated for $d_{\text{Buffer}} = 200h^{-1}\text{Mpc}$.

6.2 Acoustic scale measurements

6.2.1 Alcock-Paczynski distortions

The second part of the validation process focuses on the ability of eFAM to estimate the dilation parameters α_{\parallel} and α_{\perp} . For this purpose, we perform a likelihood analysis in each mock using the clustering wedges measured in the separation range $s \in [50, 150]h^{-1}\text{Mpc}$. In this analysis, the values of f , Σ_{\parallel} , and Σ_{\perp} are kept constant. For the Obs sample they are set equal to their best-fitting values listed in Table 2, while for RecZ and RecL we set them equal to zero. To reduce the uncertainty in the estimated α_{\perp} and α_{\parallel} , we searched for the largest volume of the sample, namely for the minimum value of d_{Buffer} , yielding an unbiased estimation of the real-space density field. We find $d_{\text{Buffer}} = 125h^{-1}\text{Mpc}$ to be a good compromise between sample volume size and systematic errors driven by edge effects. A smaller d_{Buffer} would not bias the position of the BAO peak, albeit it would reduce the accuracy of the RSD modelling for $s > 150h^{-1}\text{Mpc}$, as shown in the top-right panel of Fig. 4. Hereafter, we will only consider the Obs and RecL samples.

The results of our reconstruction of the acoustic feature are shown in Fig. 6 and 7 summarized in Table 3. Each point in Fig. 6 represents the best-fitting value obtained from each mock catalog before (top) and after (bottom) eFAM reconstruction. In both cases, we fit the distribution of points in the $(\alpha_{\perp}, \alpha_{\parallel})$ plane with a bi-variate Gaussian for which we show its maximum (star symbol) and the 68, 95 and 99 per cent confidence levels. The histograms on the top and right side of each panel show the marginalized probability distribution function of α_{\perp} and α_{\parallel} , respectively. Since our fiducial cosmology is set to be one of the mocks, the expected values are $\alpha_{\perp} = \alpha_{\parallel} = 1$.

The distribution of the best fit α_{\parallel} and α_{\perp} in the Obs sample (top panel) is $\gtrsim 30$ per cent broader than in the RecZ sample (bottom panel). Alongside, the central value of the bi-variate Gaussian is more biased pre-reconstruction, ~ 1.7 per cent along the α_{\parallel} direction, than post-reconstruction; see Table 3. Not expecting any Alcock-Paczynski distortion, we interpret both the bias and the strong uncertainty in the pre-reconstruction estimates of α_{\parallel} and α_{\perp} as a sign of the inadequacy of the fiducial RSD model (equation 11) to

account for non-linear effects. Remarkably enough, the fact that our reconstruction significantly decreases both the dispersion and the offset of the dilation parameters indicates that eFAM successfully accounts for both linear and non-linear motions in the range of scales considered in the analysis. The residual bias and dispersion affecting in the post-reconstruction measurements are probably a consequence of the imperfect modelling of the external mass distribution and its tidal field.

To further investigate the sources of uncertainty we have considered the results of the likelihood analysis performed on a single mock in which no clear BAO peak is detected in the clustering wedges before reconstruction (top right panel of Fig. 8). The histograms in the central and right top panels show the one-dimensional posterior probability function for α_{\perp} and α_{\parallel} , respectively, obtained from the Markov Chain procedure. With no BAO feature detected, neither distribution exhibit a well defined, main peak. Instead, $\mathcal{P}_{\alpha_{\perp}}$ (left) has two maxima of comparable amplitude whereas $\mathcal{P}_{\alpha_{\parallel}}$ shows a broad peak extending beyond the range allowed for this parameter. As a result, one obtains a noisy and biased estimate of both parameters. The situation dramatically improves after eFAM reconstruction (bottom panels). A BAO signature is now visible in both wedges (bottom right) and, consequently, the posterior probability distributions of both dilation parameters show a well defined, sharp maximum centered on the expected values $\alpha_{\perp} = \alpha_{\parallel} = 1$.

An alternative way to assess the impact of eFAM in reducing the uncertainty of α_{\perp} and α_{\parallel} is that of considering the errors of these parameters estimated from the clustering wedges measured in a single mock catalog. The scatter-plot in Fig. 7 compares the single-mock uncertainties $\sigma_{\alpha_{\perp}}$ (top panel) and $\sigma_{\alpha_{\parallel}}$ (bottom panel) estimated before the reconstruction (X-axis) to those estimated after the reconstruction (Y-axis). The histograms on top and on the side show the marginalized one-dimensional distributions of both quantities. Before the reconstruction, both distributions are skewed toward large errors. After reconstruction eFAM succeeds in reducing the skewness by ~ 30 per cent.

6.2.2 Robustness to non-linear effects.

To test the robustness of eFAM reconstruction to the inclusion of small scales characterized by non-linear effects, we push the likelihood analysis performed in the previous section down to separations as small as $s_{\text{min}} = 25h^{-1}\text{Mpc}$. The results of this test are summarized in Table 3 where we list the best-fitting values of α_{\perp} and α_{\parallel} and their error for $s_{\text{min}} = 25$ and $40h^{-1}\text{Mpc}$ along with the reference case $s_{\text{min}} = 50h^{-1}\text{Mpc}$. We notice that the bias on α_{\parallel} measured before the reconstruction increases when pushing the correlation analysis down to smaller scales. This is not surprising and confirms the increasing inadequacy of the RSD model in accounting for the effect of non-linear peculiar velocities. The eFAM reconstruction reverse this trend. The systematic error on α_{\parallel} remains the same, irrespective of the s_{min} considered. This result demonstrates the success of eFAM reconstruction and corroborates our conclusion that systematic errors on α_{\parallel} originate from the external tidal field rather than non-linear effects.

7 RECONSTRUCTING SDSS DATA

In this section, we apply eFAM to the SDSS DR12 galaxy sample and analyse the clustering properties before and after eFAM reconstruction.

To perform the reconstruction, we set the fiducial cosmology to the one of the mocks. We weigh the mass of each galaxy by its linear

Table 2. Best fit parameters and their 1σ uncertainties estimated to the mean clustering wedges for the three cases considered: redshift-space sample (Obs), eFAM reconstructed real-space sample at the redshift of the survey (RecZ), and back-in-time reconstructed sample (RecL).

sample	α_{\perp}	α_{\parallel}	$\Sigma_{\perp}(h^{-1}\text{Mpc})$	$\Sigma_{\parallel}(h^{-1}\text{Mpc})$	f	$\Sigma_s(h^{-1}\text{Mpc})$
Obs	0.997 ± 0.003	1.011 ± 0.007	7.37 ± 0.45	10.20 ± 1.82	0.51 ± 0.13	4.43 ± 1.83
RecZ	0.997 ± 0.003	1.011 ± 0.006	7.79 ± 0.36	9.12 ± 0.68	0.07 ± 0.06	1.09 ± 1.18
RecL	0.997 ± 0.002	1.010 ± 0.004	5.21 ± 0.39	6.93 ± 0.53	0.06 ± 0.05	1.02 ± 0.88

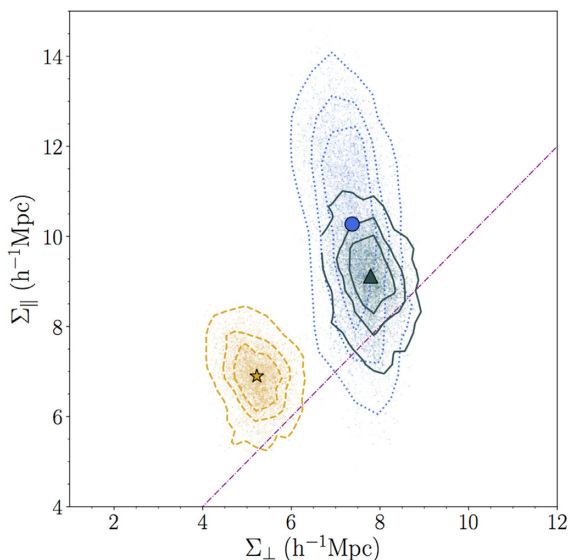


Figure 5. Marginalized probability contours of Σ_{\perp} , Σ_{\parallel} estimated from the fit of the mean clustering wedges of the mocks obtained from the Obs sample (blue dotted contours), the RecZ sample (dark green solid contours) and the RecL sample (gold, dashed contours). The large symbols (circle, triangle and star) show the value of the best-fitting parameters for, respectively, the Obs, RecZ and RecL cases. The contours are drawn at 1, 2, and 3σ confidence levels. Small dots show the values obtained at each step of the Markov-Chain. The purple, dot-dashed line is drawn for reference to indicate $\Sigma_{\parallel} = \Sigma_{\perp}$.

bias, $b = b(z_{\text{gal}})$, estimated interpolating the redshift-to-bias relation of SDSS-III galaxies at the redshift of the galaxy, z_{gal} (Salazar-Albornoz et al. 2017).

For the present study, we consider the Obs and RecL sample. Here, distances are assigned to redshift coordinates using the same (fiducial) cosmological of reconstruction. To avoid to be affected by spurious tidal fields induced by edge effects, we exclude from the analysis all galaxies lying in a buffer region within $d_{\text{Buffer}} = 125h^{-1}\text{Mpc}$ from the sample's edges.

For the two samples, we evaluate the 2PCF of the galaxies using the statistical weights described in Section 3.1 and measure its parallel and perpendicular clustering wedges. Finally, we compare ξ_{\perp} and ξ_{\parallel} with the models presented in Section 5.1.

Similarly to Ross17, we focus the analysis on the estimate of $H(r_s/r_s^{\text{fid}})$ and $D_A(r_s^{\text{fid}}/r_s)$, largely determined from the modelling of the BAO peak. We follow the same procedure adopted in Section 6.2. We model the clustering wedges in the separation range $50 < s < 150$ and fix the values of f , Σ_s , Σ_{\parallel} , and Σ_{\perp} to those measured in the mocks (Table 3). In particular, for the RecL sample, we set $f = \Sigma_s = 0$ after having verified that the effective removal of RSD in the reconstructed data.

In the likelihood analysis, we use the covariance matrix estimated from the mocks, as described in Section 4.3. We note here that,

although eFAM reconstruction is rather insensitive to the background cosmology (Nusser & Colberg 1998), a mismatch between the fiducial and correct values of the cosmological parameters may generate spurious peculiar velocities and residual anisotropies in the reconstructed 2PCF of the SDSS-DR12 galaxies. We search for such anisotropies by comparing the two clustering wedges and proceed with the likelihood analysis only after having excluded their presence.

We illustrate the results of our analysis in Fig. 9 and 10. Top panel of Fig. 9 shows the SDSS-DR12 parallel and perpendicular wedge as measured in redshift-space before eFAM reconstruction. The bottom panel shows the same quantities measured after the reconstruction. Red-triangles and blue-dots show ξ_{\parallel} and ξ_{\perp} , respectively, multiplied by the square of the spatial separation, s^2 . The error-bars are the 1σ uncertainties estimated from the mocks, which corresponds to the diagonal elements of the covariance matrix. The BAO peak is clearly seen in all measurements. To highlight its statistical significance, we over-plot the best-fitting models (equation 13) represented by continuous and dashed, solid curves. Dotted and dot-dashed curves show the best fit clustering wedge model with no BAO feature in it, obtained using the no-wiggle power spectrum P_{nw} of Eisenstein & Hu (1998).

As expected, in redshift-space (top panel) ξ_{\parallel} and ξ_{\perp} are significantly different from each other because of RSD. In the RecL sample (bottom panel), the two wedges are in good agreement, confirming the efficiency of the eFAM at recovering the real-space correlation function and thus justifying our decision to set $f = \Sigma_s = 0$ in the model fitting. From the plots, it is also evident that the eFAM reconstruction increases the amplitude and the sharpness of the acoustic peak.

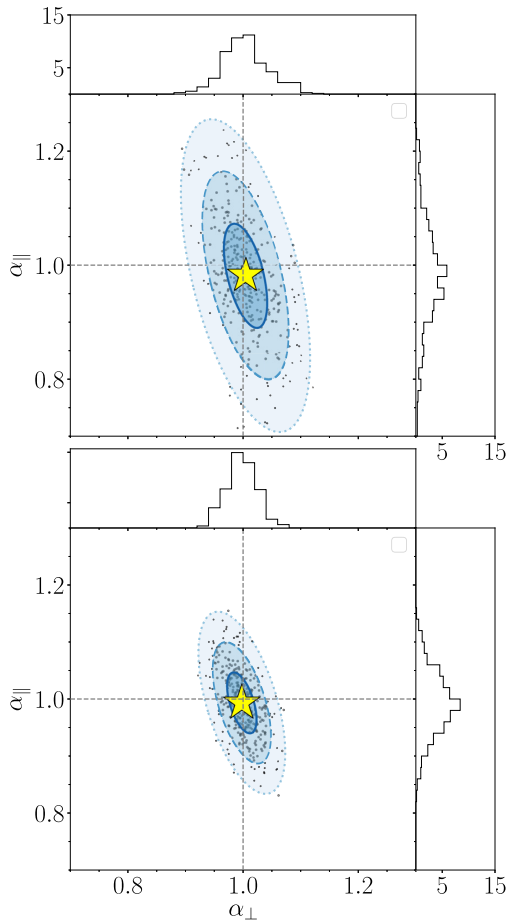
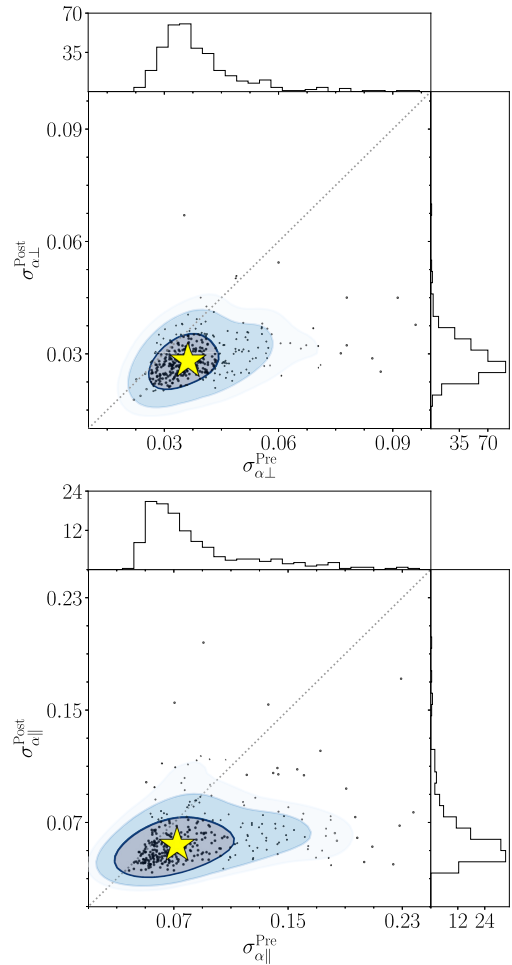
The results of the likelihood analysis corroborate and quantify this visual impression. Fig. 11 show the two-dimensional marginalized maximum likelihood of the dilation parameters α_{\perp} , α_{\parallel} surfaces before (left-hand panel) and after eFAM reconstruction (right-hand panel). Solid surfaces show the likelihood of models that include the BAO peak while transparent surfaces depict the case in which no acoustic feature is included in the model. The comparison between the amplitudes of the peaks in the solid and transparent likelihood surfaces is a proxy of the significance of the BAO peak detection.

Before reconstruction, non-linear RSD undermine the agreement with the fiducial model yielding a modest maximum in the likelihood. Alongside, the limited prominence of the acoustic feature legitimates the model without acoustic oscillations. For eight free-fitting parameters, the maximum $\Delta\chi^2$ between the two surfaces translates into a 2.1σ BAO detection. After reconstruction, the accurate modelling of both RSD and non-linear clustering favours the fiducial template against the no-wiggle model rising the significance of the BAO detection to 4.4σ .

Let us now focus on the estimate of the two parameters H and D_A derived from the dilation parameters. Equations (7, 8) Fig. 10 shows the probability contours level of the bi-variate Gaussian that best fits the marginalized likelihood. Contours and shaded areas represent the 1, 2, and 3σ confidence levels of the Gaussian before (blue) and after (yellow) eFAM reconstructions. Best-fitting values are shown with

Table 3. Best fit versus separation range. Median values and variance of α_{\perp} , α_{\parallel} , $\sigma_{\alpha_{\perp}}$, and $\sigma_{\alpha_{\parallel}}$ estimated from clustering wedges in different separation ranges.

fitting range	sample	$\langle\alpha_{\perp}\rangle - 1$	$S_{\langle\alpha_{\perp}\rangle}$	$\langle\sigma_{\alpha_{\perp}}\rangle$	$S_{\langle\sigma_{\alpha_{\perp}}\rangle}$	$\langle\alpha_{\parallel}\rangle - 1$	$S_{\langle\alpha_{\parallel}\rangle}$	$\langle\sigma_{\alpha_{\parallel}}\rangle$	$S_{\langle\sigma_{\alpha_{\parallel}}\rangle}$
[50, 150] h^{-1} Mpc	Obs	0.002	0.036	0.031	+0.017 −0.004	−0.017	0.081	0.053 −0.005	+0.067
	RecL	−0.002	0.024	0.024	+0.013 −0.002	−0.010	0.056	0.049 −0.006	+0.032
[40, 150] h^{-1} Mpc	Obs	0.002	0.035	0.032	+0.024 −0.004	−0.026	0.088	0.054 −0.004	+0.061
	RecL	−0.002	0.025	0.027	+0.007 −0.004	−0.011	0.051	0.043 −0.004	+0.030
[25, 150] h^{-1} Mpc	Obs	0.004	0.036	0.033	+0.017 −0.004	−0.032	0.091	0.059 −0.007	+0.065
	RecL	0.000	0.023	0.026	+0.013 −0.003	−0.010 −0.004	0.053	0.048	+0.030


Figure 6. Best fit α_{\parallel} , α_{\perp} values estimated from the clustering wedges in the range $s = [50, 150]h^{-1}$ Mpc measured in each MD-Patchy mocks (dots) for the Obs (Top) and RecL (Bottom) cases. The histograms on the top and on the right part of the panels show the one-dimensional distribution functions of α_{\perp} and α_{\parallel} , respectively. Elliptical contours and blue-shaded areas are drawn at the 68, 95, and 99 per cent confidence level of the best-fitting Gaussian bivariate, centered at the starred symbol. Grey-dashed lines show the expected $\alpha_{\perp} = 1$ and $\alpha_{\parallel} = 1$ values.

Figure 7. Error on the perpendicular (top) and parallel (bottom) dilation parameters, $\sigma_{\alpha_{\perp}}$ (top panel) and $\sigma_{\alpha_{\parallel}}$, before and after eFAM-reconstruction estimated from the clustering wedges. Each dot shows the results from each clustering wedges. Shaded areas contain 68, 95, and 99 of the points. The average value of the distribution is represented by a starred symbol. Histograms depict the 1D marginalized distribution of both errors, before (histogram on top) and after (histogram on the side) eFAM reconstruction. Dotted lines show the case in which reconstruction gives no improvement.

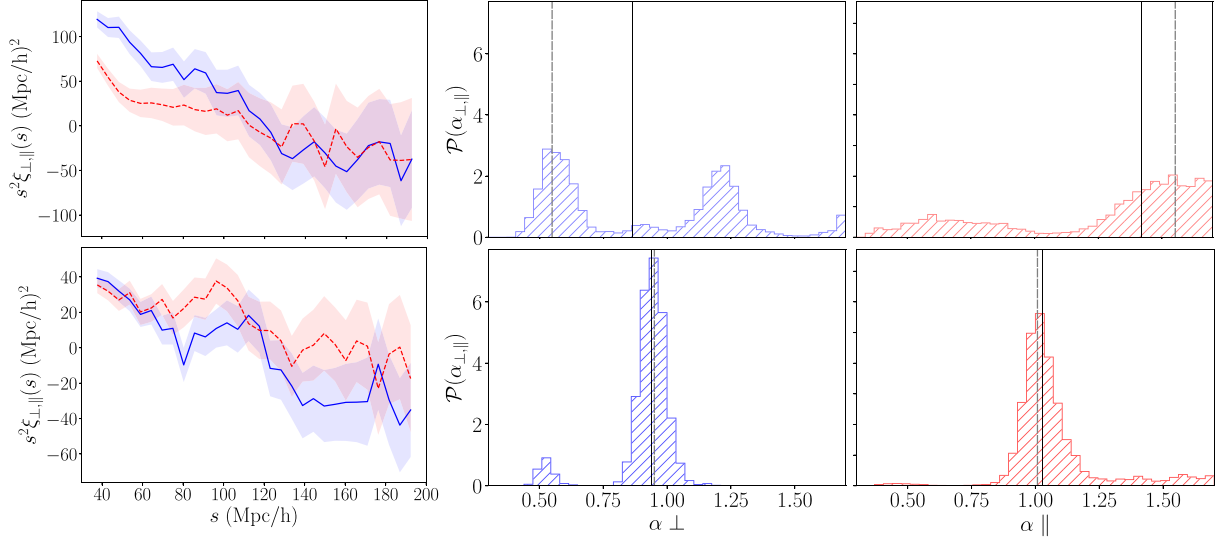


Figure 8. Modelling the clustering wedges in a single mock realization. Left panels: parallel (red-dashed lines) and perpendicular (blue-solid lines) clustering wedges multiplied by the square of the spatial separation s . Shaded areas represent the standard deviation. Central and Right panels: marginalized 1-dimensional probability distribution of $\mathcal{P}_{\alpha_{\perp}}$ and $\mathcal{P}_{\alpha_{\parallel}}$, respectively. Top panels: Results obtained from the Obs sample. Bottom panels: Results obtained from the RecL sample. Solid-black vertical lines mark the median values of the 1-dimensional \mathcal{P} while dashed-grey lines denote the best-fitting values.

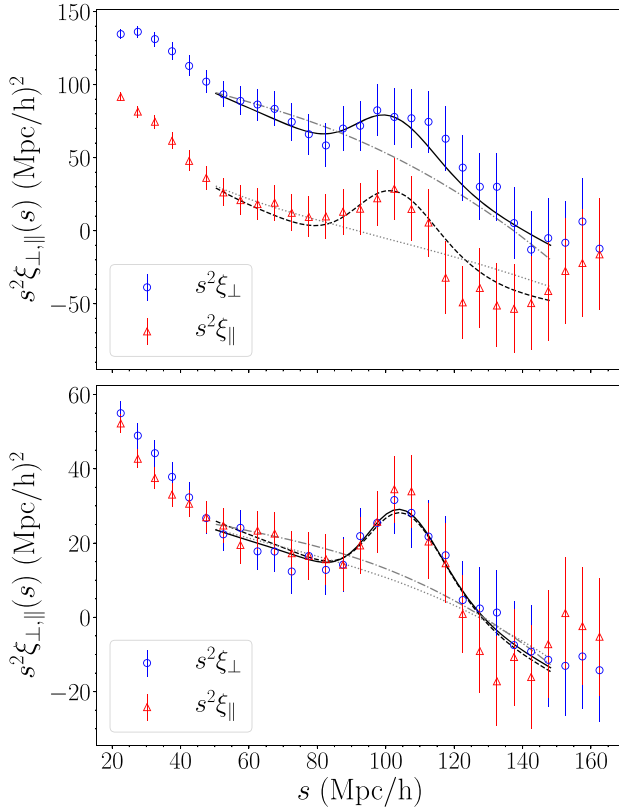


Figure 9. Clustering wedges of the BOSS-DR12 Combined Sample before (top-panel), and after eFAM reconstruction (bottom-panel). Open blue circles show the measured perpendicular wedge. Open red triangles show the parallel wedge. The error-bars are the square root of diagonal of the covariance matrix estimated from the mocks. Best fit clustering wedge models are also shown with black continuous curves. Solid and dashed curves show the best fit model that include the BAO feature. Dot and dot-dashed curves are best-fitting models with no acoustic peak.

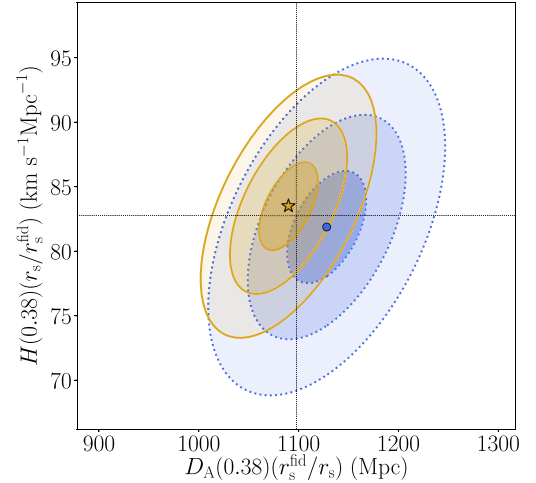


Figure 10. Marginalized likelihood contours for $D_A(r_s^{\text{fid}}/r_s)$ and $H(r_s/r_s^{\text{fid}})$ from the clustering wedges analysis of the SDSS-DR12 Combined Sample in the range $0.2 < z < 0.55$. Parameters are estimated at the effective redshift $z = 0.38$. Contours and shaded areas show 1, 2, and 3σ probability contours of a best-fitting Gaussian bi-variate. Blue contours: results before eFAM reconstruction. Yellow contours: results after eFAM reconstruction. Large blue circle and star symbols show the best-fitting values for the two cases.

a blue circle and yellow star, respectively. In the likelihood analysis, the two parameters $H(r_s/r_s^{\text{fid}})$ and $D_A(r_s^{\text{fid}}/r_s)$, are measured at the redshift of the sample $z = 0.38$. The corresponding fiducial values are $H^{\text{fid}} = 83(\text{km/s})^{-1}(\text{Mpc})^{-1}$ and $D_A^{\text{fid}} = 1098\text{Mpc}$.

Table 4 quantifies these considerations. The best-fitting value of D_A is shifted by ~ 3.5 per cent, comparably to what is seen in the analysis of the mock catalogues (Table 3). There, the effect of eFAM was to increase the magnitude of α_{\parallel} , bringing it into an agreement with the expected value. We thus assume that, similarly, eFAM successfully removes systematic errors on α_{\parallel} , and hence on D_A , from the real data set too. In support of this hypothesis, we note

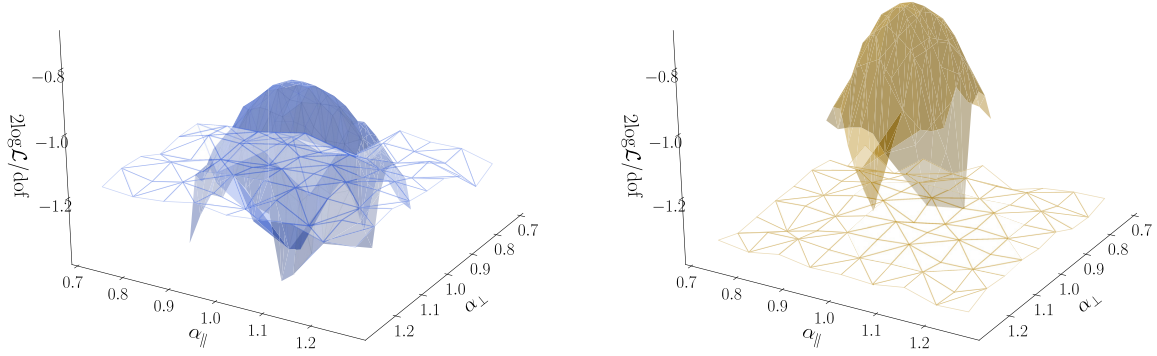


Figure 11. Two-dimensional marginalized maximum likelihood surfaces of α_{\perp} , α_{\parallel} , before (left) and after (right) eFAM reconstruction. Solid surfaces describe the best-fitting model including the acoustic feature, while transparent surfaces depict the best-fitting model without BAO (Eisenstein & Hu 1998).

Table 4. Fit results from the SDSS-DR12 combined sample.

sample	$D_A(0.38)(r_s^{\text{fid}}/r_s)$ (Mpc)	$H(0.38)(r_s/r_s^{\text{fid}})$ (km s $^{-1}$ Mpc $^{-1}$)
Obs	1129 ± 40	82 ± 4
RecL	1090 ± 29	83 ± 3

that Ross17 (see Table 3 in Alam et al. (2017) where $D_M = D_A(1 - z)$) found $H(r_s/r_s^{\text{fid}}) = (81 \pm 2)$ (km s $^{-1}$ Mpc $^{-1}$) and $D_A(r_s^s/r_s) = (1096 \pm 17)$ Mpc in their post-reconstruction analysis, which are in best agreement with our post-reconstruction measurements. On top of this, eFAM also reduces random errors. Relative errors on D_A decrease from $\sigma_{D_A}/D_A \simeq 3.5$ per cent, to $\simeq 2.6$ per cent and statistical uncertainties on H decrease from $\sigma_H/H \simeq 5$ per cent to $\simeq 3.6$ per cent. We further discuss the amplitudes of the parameters uncertainties in Section 8.

8 DISCUSSION AND CONCLUSIONS

In this study, we have applied the eFAM reconstruction algorithm to a subset of the SDSS-DR12 combined sample of galaxies aiming at assessing the performances of eFAM when applied to a real data set characterized by specific selection effects and observational uncertainties.

This work is the follow-up of the study conducted in Sarpa et al. (2019). There, we have appraised the performance of eFAM using a somewhat idealized, simulated data sets with $\mathcal{O}(10^6)$ objects, similar in size to that of current and future spectroscopic redshift surveys. The results of our past analysis have shown that the first implementation of eFAM was indeed able to successfully reconstruct the real-space positions and velocities of each object at any epochs, and specifically at both the observed redshift and at a back-in-time epoch in which density fluctuations were still evolving in the linear regime. Urged by the rising need for accurate reconstruction techniques able to enhance the signal-to-noise of the acoustic feature in the two-point correlation function, we have focused our study on the BAO scale. Despite being of high scientific interest, the application of eFAM to the mere study of the BAO scale is a bit of a limitation. Indeed, eFAM is a non-perturbative, non-linear technique designed to describe non-linear self-gravitating systems, and thus it is able to extract information from scales much smaller than the BAO one.

In the current study, we have repeated the original analysis using a real data set and a suite of realistic simulated catalogues. We chose

the SDSS-DR12 combined sample for three reasons. First of all, its size and characteristics are representative of the state-of-the-art surveys as well as next-generation, wide surveys such as DESI or Euclid. Second, reconstruction techniques based on the Zel'dovich approximation have been already applied to this data set, mainly to extract scientific information from the position and the amplitude of the BAO peak in the two-point correlation function, providing us with a reference for our results. When performing the comparison, we must keep in mind that this constitutes a specific test, limited to quasi-linear scales, where eFAM is supposed to perform just as well as Zel'dovich-based reconstruction algorithms. Third, our choice was encouraged by the public availability of a sufficiently large number of realistic mock catalogues mimicking its clustering properties. A large number of mocks is crucial to both calibrate our reconstruction technique to the specific characteristics of the survey and to estimate errors and their covariance.

To perform this analysis, we modified the eFAM algorithm to be able to deal with all the specific properties of real data sets that are not necessarily present in simulated catalogues. The first one is the galaxy bias. The first implementation of eFAM implicitly assumes that all the mass in the system is associated with discrete, visible objects that maintain their identity when traced back in time. Nusser et al. (1991) have proposed a method to modify this assumption to include the linear bias. Here, we adopted an alternative approach. We associated with each galaxy a statistical weight accounting for both the clustered dark matter component located at the galaxy position and the smooth dark component filling the survey volume. The inclusion of the smooth dark matter field is crucial to match the average mass density within the survey with the cosmic mean and hence to avoid the observed system to be seen as a local under-density. Although we did not perform specific tests to evaluate the goodness of this approach, we did not find any evidence of detectable systematic errors introduced by this biasing scheme in any of the validation tests we performed.

The second aspect we considered in the new implementation of eFAM is the gravitational influence of the unknown mass distribution lying outside the volume of the survey. The problem is general. The lack of clustering information outside the survey volume prevents us to fully reproduce the bulk motions within the survey, while the gravitational potential estimated from a finite and possibly anisotropic volume induces deceptive tidal fields in the proximity of the edges of the survey. In Sarpa et al. (2019), we were able to minimize those effects by considering a spherical geometry and discarding from the post-reconstruction analysis all the objects lying in the

proximity of the spherical surface. However, real surveys are often characterized by more complicated geometry, hence modelling the influence of external matter distribution for real data sets is a serious problem. We proposed here a solution consisting of a two-step procedure. First, we modelled the tidal field and assess their impact under the assumption of a homogeneous isotropic mass distribution outside the sample. Second, as in our original work, we defined a buffer region near the survey's edges. Objects in the buffer region are accounted for in the eFAM reconstruction but excluded from scientific analyses. We run several dedicated tests to validate this procedure and chose the size and shape of the buffer region such as to optimize the compromise between a meaningful statistical sampling and systematic uncertainties induced by tidal effects. We proceeded as follows. eFAM is supposed to displace objects from their observed redshift to their back-in-time real-space position. When it does it successfully, it removes all the anisotropies in the two-point correlation function due to RSD. In our tests, we found that an imperfect removal of tidal effects would manifest itself as a non-zero quadrupole moment on scales larger than the BAO peak. One can then enlarge the buffer region until this spurious signal is completely removed. The availability of realistic mock catalogues is obviously of paramount importance to calibrate the procedure. It is worth stressing that tidal field will become an even more relevant issue with the advent of next-generation wide surveys. Indeed, to minimize evolution effects, typical sub-samples used for scientific analyses will have a depth along the radial direction much shorter than their transverse size. The impact of the tidal field will thus play an important role in optimizing the shape and the volume of the sample to be analyzed.

A distinctive trait of eFAM reconstruction is its ability to reconstruct the full orbit of the mass tracers, as opposed to solely provide their positions at some fixed early epoch. We can thus use eFAM to recover the real-space positions of objects at the observed and high redshift, alike, effectively modelling both redshift distortions and non-linear evolution. This is done by assuming a value for the linear growth rate f . The success of the reconstruction can be then assessed by searching for residual anisotropy in the real-space two-point correlation function of the galaxies. Our tests on the mock reveal a remarkably accurate modelling of RSD. The values of f and Σ_s in the real-space 2PCF measured at the redshift of the sample are consistent with zero. Besides, the eFAM reconstruction reduces by 65 per cent the damping of the acoustic oscillations due to late-time non-linear clustering, $\Sigma_{\parallel, \perp}$. Having successfully removed RSD, we can set the parameters f and Σ_s equal to zero when modelling the 2PCF of galaxies at their back-in-time reconstructed positions, effectively decreasing the number of free parameters in the fit. Furthermore, we found no need to include the additional shape parameters $B_{0,2}$ used in Ross17 to fit the 2PCF. As a result, the back-in-time eFAM reconstructed two-point correlation function can be well fit with an 11-parameter model over the range $[50, 150]h^{-1}\text{Mpc}$, to be compared with the 15-parameter model used to fit the measured 2-point correlation function in redshift-space.

A final, positive characteristic of eFAM worth stressing is its non-linear nature which allows us to push the analyses down to scales that are not accessible to Zel'dovich-based reconstructions. This fact is already evident from the wide separation range, $[50, 150]h^{-1}\text{Mpc}$, used in our analysis. However, our validation tests show that one can safely push the analysis down to separations as small as $25h^{-1}\text{Mpc}$ without introducing systematic errors. Strictly speaking, in this work we only checked this to be true for the estimate of the dilation parameters. However, the visual inspections of the clustering wedges reveal no obvious departures from model predictions on those scales.

In fact, we found that RSD are successfully removed and models fit the reconstructed 2PCF up to $180h^{-1}\text{Mpc}$.

We conclude by summarizing the main results of our analysis:

(i) The tests performed with 400 SDSS-DR12 mock catalogs show that eFAM successfully reconstruct the real-space position of galaxies at the mean redshift of the sample, $z = 0.38$. As a result, redshift-space distortions are removed from the two-point correlation signal in the separation range $[50, 180]h^{-1}\text{Mpc}$. On larger scales, we detect non-zero quadrupole moments that we attribute to the gravitational pull of the inhomogeneous mass distribution outside the survey volume. The impact of these external tidal field can be reduced by excluding objects near the edge of the survey. The size of this exclusion region should be set by compromising between number of objects and range of separation to be used in the clustering analysis. Ultimately it depends on the volume, shape, and redshift of the sample. The lower bound in separation, set to $50h^{-1}\text{Mpc}$, simply represents a conservative choice. In fact, we showed that it can be safely reduced to $25h^{-1}\text{Mpc}$ without compromising our ability to estimate the dilation parameters, $\alpha_{\perp, \parallel}$, and, therefore, the values of H and D_A .

(ii) eFAM reconstruction allows us to increase the precision of both $\alpha_{\perp, \parallel}$ by ~ 3 per cent. This is a significant reduction which, however, is smaller in magnitude than the one obtained by Ross17 using the Zel'dovich approximation to perform the back-in-time reconstruction. This is somewhat surprising since one would expect, based on the results of Sarpa et al. (2019), which the quality of the eFAM reconstruction should be higher than that of Zel'dovich approximation. Part of this discrepancy can be explained by the very nature of the analysis. Here, we are focusing on the $\alpha_{\perp, \parallel}$ parameters that are mainly determined by the reconstruction quality of the BAO peak, and we do not eFAM to perform significantly better than Zel'dovich on these scales. However, we do not expect worse performances either, especially considering that the number of free parameters used in our model for the reconstructed 2PCF is smaller than the one employed in Ross17. We believe that the larger eFAM uncertainty, that are estimated from the mock catalogs, reflect the fewer catalogs (400) and reduced volume (north galactic cap) used to evaluate the covariance matrix with respect to that (1000, both north and south caps) used by Ross17. To corroborate this hypothesis, we notice that Kazin et al. (2013) performed a similar analysis estimating the values of the dilation parameters and their uncertainties using 600 rather than 1000 mocks. The uncertainties presented in their work, $\sigma_H/H = 5.8$ per cent and $\sigma_{D_A}/D_A = 3.1$ per cent (see Table 3), are overestimated with respect Ross17 and comparable with ours.

(iii) From the clustering wedges of the back-in-time reconstructed position of the SDSS-DR12 galaxies, we have estimated the Hubble parameter and the angular diameter distance at the effective redshift of the sample, normalized to their values in the fiducial cosmological model. We found $D_A(r_s^{\text{fid}}/r_s) = 1090 \pm 29$ Mpc and $H(r_s/r_s^{\text{fid}}) = 83 \pm 3(\text{km s}^{-1} \text{Mpc}^{-1})$. The eFAM reconstruction significantly reduces the error in these estimates by 35 and 39 per cent, respectively, compared to the case in which the analysis is carried out in redshift-space, i.e. before reconstruction. These results are in good agreement with those of previous, similar analyses performed on the same data sets (Ross17, Vargas-Magaña et al. (2018)).

The results presented in this work show that eFAM can be successfully applied to current and future spectroscopic data sets with $\mathcal{O}(10^6)$ objects. The enhancement of the clustering signal-to-noise at the BAO peak is comparable to that obtained by the popular reconstruction techniques that assume Zel'dovich approximation.

However, eFAM also succeeds in recovering the correlation signal well into non-linear regime, as demonstrated by its ability to recover the correct shape and amplitude of the galaxy-galaxy correlation function down to separation as small as $25h^{-1}\text{Mpc}$. We conclude that eFAM provides a back-in-time reconstruction method that, can be regarded as a complementary tool to extract scientific information from the numerous small-scale modes that cannot be accessed by standard reconstruction techniques.

ACKNOWLEDGEMENTS

ES thanks A. Longobardi for valuable discussions and precious scientific advice.

The authors thank the anonymous referee for careful reading of the paper and useful comments.

CS and ES acknowledge financial support by Programme National Cosmology et Galaxies (PNCG) of CNRS/INSU with INP and IN2P3, co-funded by CEA and CNES. EB is supported by MUIR/PRIN 2017 ‘From Darklight to Dark Matter: understanding the galaxy-matter connection to measure the Universe’, ASI/INAF agreement number 2018-23-HH.0 ‘Scientific activity for Euclid mission, Phase D’, ASI/INAF agreement number 2017-14-H.O ‘Unveiling Dark Matter and Missing Baryons in the high-energy sky’ and INFN project ‘INDARK’. This research was supported by the Munich Institute for Astro- and Particle Physics (MIAPP) which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC-2094 - 390783311 and was also partially supported by the Programme National Cosmology et Galaxies (PNCG) of CNRS/INSU with INP and IN2P3, co-funded by CEA and CNES.

The massive production of all MultiDark-Patchy mocks for the BOSS Final Data Release has been performed at the BSC Marenostrum supercomputer, the Hydra cluster at the Instituto de Fisica Teorica UAM/CSIC, and NERSC at the Lawrence Berkeley National Laboratory. We acknowledge support from the Spanish MICINN Consolider-Ingenio 2010 Programme under grant MultiDark CSD2009-00064, MINECO Centro de Excelencia Severo Ochoa Programme under grant SEV-2012-0249, and grant AYA2014-60641-C2-1-P. The MultiDark-Patchy mocks was an effort led from the IFT UAM-CSIC by F. Prada’s group (C.-H. Chuang, S. Rodriguez-Torres and C. Scoccola) in collaboration with C. Zhao (Tsinghua U.), F.-S. Kitaura (AIP), A. Klypin (NMSU), G. Yepes (UAM), and the BOSS galaxy clustering working group.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Alam S. et al., 2017, *MNRAS*, 470, 2617
 Alcock C., Paczynski B., 1979, *Nature*, 281, 358
 Anderson L. et al., 2014, *MNRAS*, 441, 24
 Beutler F. et al., 2017, *MNRAS*, 464, 3409
 Blanton M. R. et al., 2017, *AJ*, 154, 28
 Branchini E., Carlberg R. G., 1994, *ApJ*, 434, 37
 Crocce M., Scoccimarro R., 2006, *Phys. Rev. D*, 73, 063519
 Crocce M., Scoccimarro R., 2008, *Phys. Rev. D*, 77, 023533
 Dawson K. S. et al., 2013, *AJ*, 145, 10
 DESI Collaboration, 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
 Drinkwater M., Blake C., Wigglez Team, 2009, *AAONw*, 115, 3
 Eisenstein D. J., Hu W., 1998, *ApJ*, 496, 605

- Eisenstein D. J., Seo H.-J., White M., 2007a, *ApJ*, 664, 660
 Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007b, *ApJ*, 664, 675
 Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
 Green J. et al., 2012, preprint ([arXiv:1208.4012](https://arxiv.org/abs/1208.4012))
 Hada R., Eisenstein D. J., 2018, *MNRAS*, 478, 1866
 Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
 Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, *MNRAS*, 442, 2304
 Ivanov M. M., Sibiryakov S., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 053
 Kaiser N., 1987, *MNRAS*, 227, 1
 Kazin E. A. et al., 2013, *MNRAS*, 435, 64
 Kazin E. A., Sánchez A. G., Blanton M. R., 2012, *MNRAS*, 419, 3223
 Kitaura F.-S. et al., 2016, *MNRAS*, 456, 4156
 Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
 Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
 Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
 Longobardi A., Arnaboldi M., Gerhard O., Hanuschik R., 2015, *A&A*, 579, A135
 Mao T.-X., Wang J., Li B., Cai Y.-C., Falck B., Neyrinck M., Szalay A., 2020, preprint ([arXiv:2002.10218](https://arxiv.org/abs/2002.10218))
 Marulli F., Veropalumbo A., Moresco M., 2016, *Astron. Comput.*, 14, 35
 Nelder J. A., Mead R., 1965, *Computer J.*, 7, 308
 Nusser A., Branchini E., 2000, *MNRAS*, 313, 587
 Nusser A., Colberg J., 1998, *MNRAS*, 294, 457
 Nusser A., Dekel A., Bertschinger E., Blumenthal G. R., 1991, *ApJ*, 379, 6
 Padmanabhan N., White M., 2008, *Phys. Rev. D*, 77, 123540
 Padmanabhan N., White M., Cohn J. D., 2009, *Phys. Rev. D*, 79, 063523
 Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, *MNRAS*, 427, 2132
 Peacock J. A., Dodds S. J., 1994, *MNRAS*, 267, 1020
 Peebles P. J. E., 1989, *ApJ*, 344, L53
 Percival W. J. et al., 2014, *MNRAS*, 439, 2531
 Ragagnin A., Dolag K., Biffi V., Cadolle Bel M., Hammer N. J., Krukau A., Petkova M., Steinborn D., 2017, *Astron. Comput.*, 20, 52
 Rasera Y., Corasaniti P. S., Alimi J. M., Bouillot V., Reverdy V., Balmès I., 2014, *MNRAS*, 440, 1420
 Rodríguez-Torres S. A. et al., 2016, *MNRAS*, 460, 1173
 Ross A. J. et al., 2017, *MNRAS*, 464, 1168
 Salazar-Albornoz S. et al., 2017, *MNRAS*, 468, 2938
 Sarpa E., Schimd C., Branchini E., Matarrese S., 2019, *MNRAS*, 484, 3818
 Schmittfull M., Baldauf T., Zaldarriaga M., 2017, *Phys. Rev. D*, 96, 023505
 Vargas-Magaña M. et al., 2018, *MNRAS*, 477, 1153
 Wang X., Yu H.-R., Zhu H.-M., Yu Y., Pan Q., Pen U.-L., 2017, *ApJ*, 841, L29
 Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, *MNRAS*, 427, 2146

APPENDIX A: CORRECTION FOR SURVEY GEOMETRY: ACCURACY TEST

To assess the validity of the procedure described above, we applied eFAM on a uniform random catalogue with same sky-coverage and redshift proportions of the SDSS-DR12 survey. Fig. A1 depicts the reconstructed velocity field obtained neglecting the external contribution (top panel) and including the correction for the survey geometry (bottom panel). As expected, in the first case the velocity field is dominated by a spurious coherent bulk-flow generated by the non-spherical symmetry of the survey. Adopting the correct estimation for ϕ_{ex} , the reconstructed velocity field is instead characterized by almost vanishing and randomly oriented peculiar velocities. This illustrates the efficiency of the correction in erasing the spurious motions and, more importantly, reassuring from the presence of additional systematics.

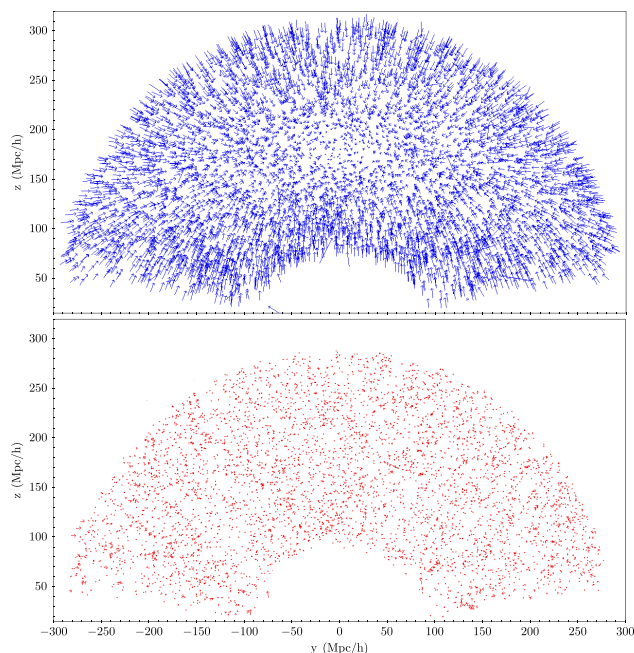


Figure A1. 2D velocity map for the BOSS-like random distribution. Top panel velocity field reconstructed without correcting the external tidal field; the bulk-flow induced by the shape of the survey is dominant. Bottom panel: velocity field reconstructed including the external tidal field; the geometrical bulk flow is successfully removed.

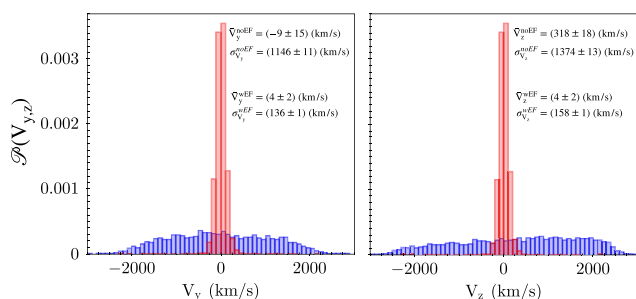


Figure A2. Probability distribution function of two components of the reconstructed peculiar velocities for the BOSS-like random distribution. Before accounting for the external tidal field (top-right) the velocity distribution is characterized by a strong velocity dispersion of about ~ 1400 km/s, describing the spurious in-fall of particles induced by the survey geometry. With the inclusion of the external field (bottom-left) the rms of the velocities is of about ~ 140 km/s, proving the efficiency of the method in removing the bulk-flow.

A more quantitative description of the efficiency of the correction is provided by Fig. A2, which illustrates the PDF of two components of the peculiar velocities. Before applying the correction (blue), the reconstructed PDF is characterized by a wide dispersion accounting for the artificial in-fall of particles and, for the V_z component, by a spurious bulk flow component. With the new method, the rms of the velocities is reduced by 90 per cent and the bulk flow removed, proving the efficacy of the method.

The improvement in the velocity reconstruction due to the external field correction can be more quantitatively assessed through the velocity-velocity comparison illustrated in Fig. A3. Before applying the geometrical correction (left panel), the reconstructed velocities are completely uncorrelated with the ‘true’ ones, showing high-

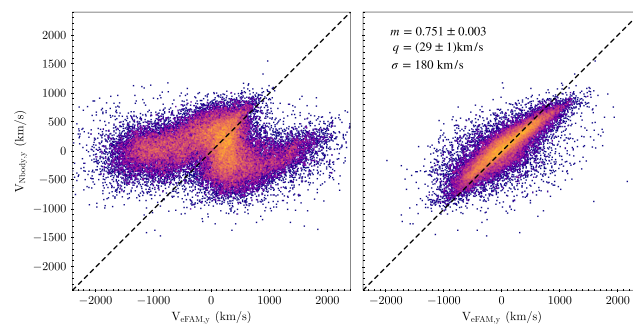


Figure A3. Accuracy tests for eFAM velocity reconstruction applied on a non-spherically symmetric survey. ‘True’ versus reconstructed peculiar velocities for one Cartesian component before (left) and after (right) including the external tidal field computed from a random distribution (results are similar for other components). Along every Cartesian direction, a perfect reconstruction would give a linear regression $V_{\text{Nbody}} = mV_{\text{eFAM}} + q$ with slope $m = 1$, no residual bulk velocity ($q = 0$), and no scatter (solid line). The reconstructed peculiar velocities before including the tidal field are not correlated with the true ones, having a significantly larger velocity dispersion with respect to the N-body velocities. After including the external tidal field, the reconstructed peculiar velocities are instead well-correlated with the true ones.

velocity tails which don’t find correspondence in the N -body field. The correlation is efficiently restored after including the external field (right panel), yielding a precise reconstruction.

APPENDIX B: VELOCITY CHALLENGE: EFAM VERSUS ZEL’DOVICH-LIKE RECONSTRUCTION

To further assess the goodness of eFAM reconstruction compared to the commonly adopted Zel’dovich approximation (ZA), we run an additional, more demanding, test in which we go beyond summary two-point statistics and compare true versus reconstructed velocities on an object-by-object basis. More precisely, we first run eFAM and minimize the action with respect to the full set of orbital coefficients, $\mathbf{C}_{i,n}$. Then we repeat the same exercise but minimize with respect to the subset $\mathbf{C}_{1,n}$, having set $\mathbf{C}_{j,n} = 0$ for $i \neq j$. Since this corresponds to imposing straight line orbits, we dubbed this reconstruction ZA, as opposed to eFAM. We stress that this is not the same as the Zel’dovich approach routinely used for BAO peak reconstructions, mainly because in our case we do not smooth the underlying density field. To run this test, we consider a mock galaxy catalogue extracted from the publicly available *Magneticum* hydrodynamical simulation (Hirschmann et al. 2014; Ragagnin et al. 2017). To match the number density of the SDSS galaxies considered in this work, we have analysed the $z = 0$ snapshot of the `box2b/hr` catalog and selected all objects with stellar-mass above $3.7 \times 10^{10} h^{-1} M_{\text{sun}}$. Redshift-space distortions have been simulated by adding the velocity-induced Doppler shift to the cosmological redshift of each galaxy. The results of these tests are summarized in Fig. B1. The scatter plot in the upper left-hand panel compares the modulus of the ZA-reconstructed velocities to that of the N-body ones. The bottom left panel illustrates the eFAM reconstructed case. Clearly, the ZA-like reconstruction severely overestimates the magnitude of the velocity. We believe that this is the result of having forced objects along straight orbits without smoothing the density field (or without softening the force on conveniently large scales). The largest discrepancies, with ZA velocities as large as ~ 5000 km/s, occur for objects in high-density regions. On the contrary, eFAM performs significantly better. The

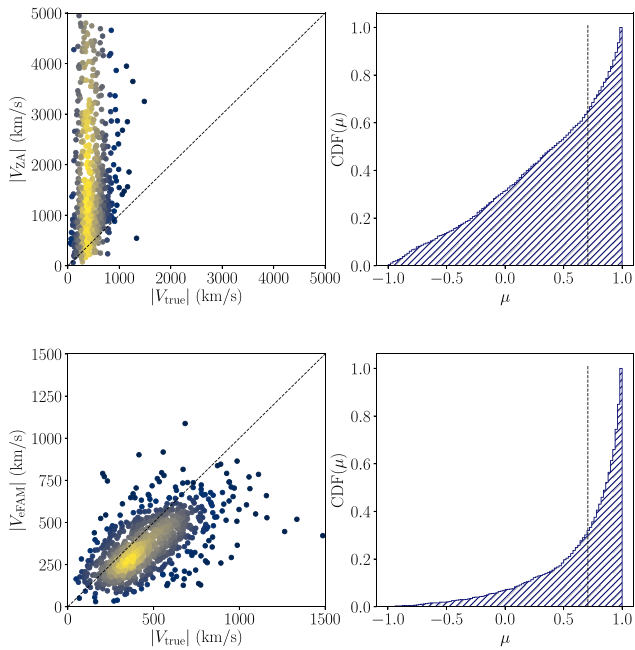


Figure B1. Peculiar velocity reconstruction; Zel'dovich-like approximation versus eFAM. Left-hand panel: individual dots show a point-by-point comparison between the amplitudes of the reconstructed and N-body peculiar velocity vectors. Black-dashed diagonal line shows the case of unbiased reconstruction, $|V_{ZA, eFAM}| = |V_{true}|$. Right: Cumulative probability distribution function of the cosine angle between the reconstructed and N-body velocity vectors, μ . The vertical, black-dashed line is drawn for reference at the mismatch angle 45° s. Top: Zel'dovich like approximation. Bottom panel: full eFAM.

bottom left-hand panel shows no obvious bias in the reconstructed velocities, except for the presence of a few outliers, corresponding to objects with large velocities, presumably in high-density regions too, that eFAM is not able to reproduce. Panels on the right focus on the direction of the velocity vectors. They show the cumulative distribution function of the cosine angle μ between the reconstructed and the N-body velocity; $\mu = 1$ being the case of perfect alignment. The eFAM distribution features a sharp peak at $\mu = 1$ and shows that in 42 per cent of the cases the alignment between the vectors is worse than 45° . On the contrary, in the ZA case this percentage increase to 64 per cent.

We conclude that without a properly tailored, initial smoothing step, the ZA-like method performs poorly in reconstructing velocities, especially when compared to eFAM that, on the contrary, does not seem to introduce any significant bias in the velocity model.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.